

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Signal Analysis and Classification of Audio Samples From Individuals Diagnosed With COVID-19

Permalink

<https://escholarship.org/uc/item/29c2h9p9>

Author

Dubnov, Tammuz

Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Signal Analysis and Classification of Audio Samples From Individuals Diagnosed With
COVID-19

A thesis submitted in partial satisfaction of the
requirements for the degree of Master of Science

in

Engineering Sciences (Mechanical Engineering)

by

Tammuz Dubnov

Committee in charge:

Professor Mauricio de Oliveira, Chair
Professor Thomas Bewley, Co-Chair
Professor Sonia Martinez

2020

Copyright

Tammuz Dubnov, 2020

All rights reserved.

The Thesis of Tammuz Dubnov is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

2020

EPIGRAPH

*The key to artificial intelligence
has always been the representation..*

— Jeff Hawkins

TABLE OF CONTENTS

Signature Page	iii
Epigraph	iv
Table of Contents	v
List of Figures	vii
List of Tables	ix
Acknowledgements	x
Abstract of the Thesis	xi
Chapter 1 Introduction	1
1.1 COVID-19	1
1.2 Current Medical Testing	2
1.3 Overview	3
1.4 Previous Works	3
Chapter 2 Dataset	8
2.1 Audio Data	8
2.2 Data Aggregation	8
2.3 Data Characteristics	10
2.3.1 Gender Ratio	10
2.3.2 Age Distribution	11
2.3.3 Audio Samples	11
Chapter 3 Audio Signal Analysis	14
3.1 Motivation	14
3.2 Signal Analysis	16
3.2.1 Fast Fourier Transformation	16
3.2.2 Mel-Frequency Cepstral Analysis	16
3.2.3 Bicoherence and Biphas	20
Chapter 4 Classification	24
4.1 Methods of Classification	25
4.1.1 Logistic Regression	26
4.1.2 Support Vector Machines	26
4.1.3 Random Forest	27
4.1.4 Multilayer Perceptron	27
4.1.5 Convolutional Neural Networks	28
Chapter 5 Classification Results	31

5.1	Bicoherence	31
5.2	Biphase	31
5.3	Mel-Frequency Cepstrum	32
Chapter 6	Discussion	34
6.1	Bicoherence	35
6.2	Biphase	35
6.3	Mel-Frequency Cepstrum	37
6.4	Conclusion	37
6.5	Future Work	39
Bibliography	41

LIST OF FIGURES

Figure 1.1.	Variations of sound recordings of cough sounds from individuals diagnosed with different conditions, from Korpas et al. [8]	2
Figure 1.2.	Variations of coughs from patients with different pulmonary ailments from Abeyratne et al.[15]	4
Figure 2.1.	Waveform visualization of an audio sample from a healthy (left) and COVID-19 sick (right) coughing three times.	12
Figure 2.2.	Waveform visualization of an audio sample from a healthy (left) and COVID-19 sick (right) saying "Ahhh".	12
Figure 2.3.	Waveform visualization of an audio sample from a healthy (left) and COVID-19 sick (right) saying "Eeee".	12
Figure 2.4.	Waveform visualization of an audio sample from a healthy (left) and COVID-19 sick (right) saying "Oooo".	13
Figure 2.5.	The counts of audio samples across the different prompts for healthy (left) and COVID-19 sick (right).	13
Figure 3.1.	spectrogram am of coughs from a healthy individual (on the left) and a sick individual diagnosed with COVID-19 (on the right).	17
Figure 3.2.	Diagram showing the mapping from frequency to quefrequency in which the individual harmonic structure is mapped to a high quefrequency and the underlying spectral envelope is mapped to a lower quefrequency. [5]	18
Figure 3.3.	Mapping of Frequency in Hz to Pitch in Mel-Scale [16].	19
Figure 3.4.	A visualization of 20 critical bands in the Mel-Scale mapped to the frequencies. [11].	20
Figure 3.5.	Bicoherence of a cough from a healthy individual in a two dimensional representation (left) and three dimensional representation (right)	22
Figure 3.6.	Bicoherence of a cough from a sick individual diagnosed with COVID-19 in a two dimensional representation (left) and three dimensional representation (right)	22
Figure 3.7.	Biphase of a cough from a healthy individual in a two dimensional representation (left) and three dimensional representation (right)	22

Figure 3.8.	Biphase of a cough from a sick individual diagnosed with COVID-19 in a two dimensional representation (left) and three dimensional representation (right)	23
Figure 4.1.	Classification results of Logistic Regression, Support Vector Machine using a linear kernel, Support Vector Machine using an RBF kernel, Random Forest and Multilayer Perceptron for different input data.	26
Figure 4.2.	A diagram of a convolutional neural network	29
Figure 4.3.	A diagram of the architecture of the convolutional neural network	30
Figure 6.1.	A plot of the recall score for various classification method for COVID-19 sick and healthy based on a bicoherence analysis input.	36
Figure 6.2.	A plot of the recall score for various classification method for COVID-19 sick and the overall accuracy based on a bicoherence analysis input.	36
Figure 6.3.	A plot of the recall score for various classification method for COVID-19 sick and healthy based on a biphase analysis input.	37
Figure 6.4.	A plot of the recall score for various classification method for COVID-19 sick and the overall accuracy based on a biphase analysis input.	38
Figure 6.5.	A plot of the recall score for various classification method for COVID-19 sick and healthy based on a Mel-Frequency Cepstrum analysis as input. ..	38
Figure 6.6.	A plot of the recall score for various classification method for COVID-19 sick and the overall accuracy based on a Mel-Frequency Cepstrum analysis as input.	39

LIST OF TABLES

Table 1.1.	Previous cough classification works showing the classification task, results, partially or fully automatic, number of coughs in the dataset, and the methods employed.	6
Table 5.1.	Bicoherence classification results for different classification methods.	32
Table 5.2.	Biphase classification results for different classification methods.	33
Table 5.3.	Mel-Frequency Cepstrum classification results for different classification methods.	33

ACKNOWLEDGEMENTS

A special thanks to my advisor, Professor Mauricio de Oliveira, that has always been supportive and who's diverse scope of expertise was an inspiration and guidance in this interdisciplinary endeavor. With special to dedication to my parents, family, advisor and mentors throughout my journey that has thus far lead me to this masters degree. As the pace of innovation accelerates, I am grateful to those that have shared their knowledge and perspective with me to help shape my view of the world. Special thanks to the professors at UC San Diego that helped expand my horizon of knowledge as it relates to engineering and my academic abilities.

A special thanks to the healthcare providers around the world that are risking their lives to heal the world during this pandemic. And an additional thanks to all of the individuals who worked to collect the audio samples that made this study feasible.

ABSTRACT OF THE THESIS

Signal Analysis and Classification of Audio Samples From Individuals Diagnosed With
COVID-19

by

Tammuz Dubnov

Master of Science in Engineering Sciences (Mechanical Engineering)

University of California San Diego, 2020

Professor Mauricio de Oliveira, Chair
Professor Thomas Bewley, Co-Chair

The COVID-19 pandemic has overwhelmed health care systems around the world as a very serious pulmonary ailment that often leads to coughing. Cough sounds contain underutilized pulmonary health information that can be analyzed. Using signal analysis methods for audio, this paper explores the analysis and classification of cough sounds and audio samples from individuals diagnosed with COVID-19. Using features extracted through signal analysis, a classifier is developed to evaluate whether an audio sample is likely to have a COVID-19 symptoms. This can potentially be used for remote and early diagnostic efforts to help the world initiatives tackling this pandemic.

Chapter 1

Introduction

1.1 COVID-19

COVID-19 is an infectious disease caused by severe acute respiratory syndrome coronavirus that starts by infecting the mucous membranes in the throat and moves down the respiratory tract leading to the lungs, with coughing being a common symptom. Cough sounds contain underutilized pulmonary health information that can be analyzed. This research focuses on the auditory effects symptoms of COVID-19 can have on individuals in which the changes can be detected through analysis of audio recordings of an individual's coughs and speech. COVID-19 is known to result in a dry cough in 67% of cases and result in phlegm production in 33% of cases. Detection for these can be learned using machine learning methodologies trained on cough and speech samples from healthy and sick individuals.

Researchers have hypothesized that analysing the characteristic of a cough's sound can have significant value in prognosis as it can indicate the progress of a disease or the effectiveness of a treatment. Coughs are rapid forced exhalation maneuver that have rich temporal and spectral representations, varying across different diseases. Coughing is a known symptom in over a 100 diseases and are a direct result of pathophysiological changes due to the disease. As Korpas et al. [8] show in Figure 1.1 the signature of the cough sounds vary by disease due to the structural nature of the the tissue and the behavior of the glottis, where the glottis would be affected by the pathological conditions of the disease, that would lead to certain patterns in the cough sound.

Based on this motivation, cough sounds and other vocal samples such as "ahhh" and "eeee" audio samples from people diagnosed with COVID-19 were evaluated.

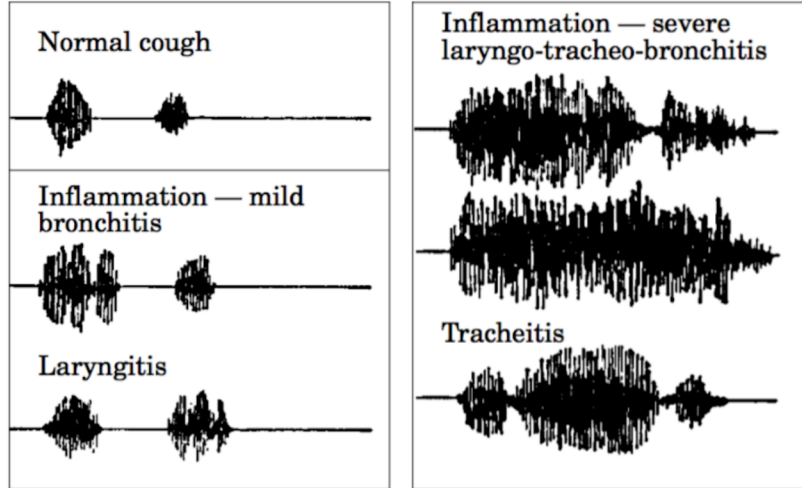


Figure 1.1. Variations of sound recordings of cough sounds from individuals diagnosed with different conditions, from Korpas et al. [8]

As of May 7th, 2020, over 75,000 deaths in the United States alone have been attributed to COVID-19 with over 1.2 million confirmed cases across the nation. Through conversations with medical doctors and certified pulmonologists, it appears that machine-based voice diagnosis technologies are not an accepted technology in the medical field and may be an overlooked tool in the continued effort to combat COVID-19.

1.2 Current Medical Testing

As COVID-19 is a pulmonary ailment, with coughing and shortness of breath among some of its symptoms, it stands to reason that signal analysis of cough sounds from individuals positively diagnosed with the diseases can lead to the discovery of linear and non-linear features unique to COVID-19. By combining these analysis with machine learning approaches, a classifier can be built to help automatically diagnose individuals by analyzing audio samples they provide. This is done with the hopes that it can grow to be a tool for medical professionals in the field that

are overworked with limited resources.

Based on consultations from MD's working in testing centers, the current standard diagnosis method for COVID-19 is using a nasopharyngeal swab and performing a PCR molecular analysis on it to look for COVID-19 markers. It takes between 48 hours to a week to receive and communicate the results to the individual undergoing the diagnosis. Furthermore, the nasopharyngeal swab test is known to have a 30% false negative rate while also putting the doctor conducting the test at risk of contracting the disease. The US is currently estimated to need to double its current rate of testing at around 1.5 million tests per week, a task that seems unreachable in the immediate future with the shortage of testing equipment.

1.3 Overview

This thesis will begin with an overlook on initiatives that collected the data needed for this study and the characteristics of the data collected, then introduce the various signal analysis techniques used in this study and showcase its immediate results on data. We will then continue by introducing machine learning classification algorithm for our needs in chapter four. We will finish with an analysis and discussion of the results of the overall classification algorithm using the signal processing technique listed.

1.4 Previous Works

Numerous research has been conducted over the past few decades to evaluate whether it is possible to classify various attributes of coughs such as whether the cough is wet/dry or the intensity of the cough, through audio analysis of the coughs. Various diseases such as Asthma, Tuberculosis, Bronchitis, among others, can have very prominent effects on the pulmonary system and have been shown to be identifiable through signal processing analysis of the sounds of coughs. Korpas et al. [8] compared cough sounds recorded in a clinical setting with the results of spirometry tests, a common test used to assess a patient's lungs by measuring how

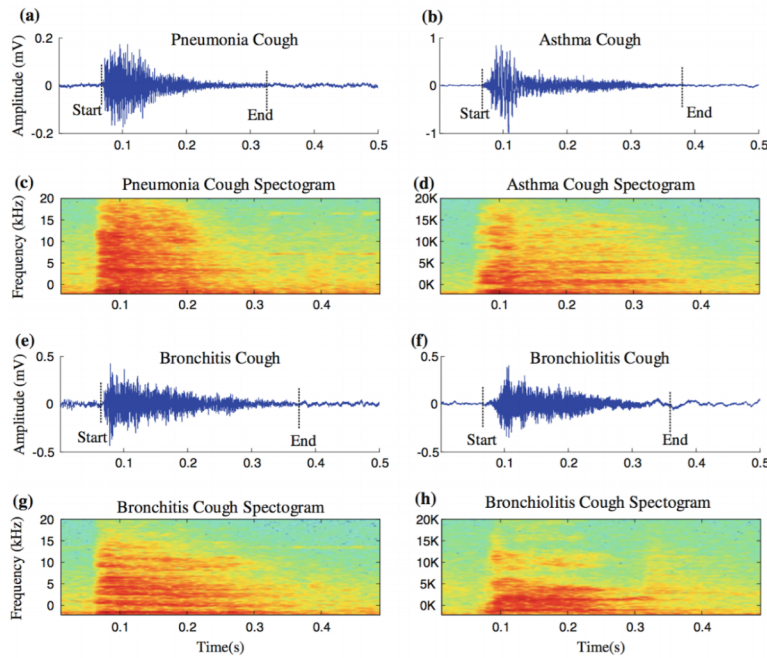


Figure 1.2. Variations of coughs from patients with different pulmonary ailments from Abeyratne et al.[15]

much air is inhaled, is exhaled and how quickly is exhaled. Korpas performed signal processing operations on the cough sounds, inspecting the time domain waveforms and frequency domain periodograms, and concluded that the cough sounds contain separate pieces of information from the spirometry tests. Even more so, Korpas showed that some drugs administered, specifically a bronchodilating drug, to patients would alter the spirometry results but leave the cough sound analysis relatively unchanged.

Abeyratne et al. [15] analyzed and visualized the differences in pneumonia coughs, asthma coughs, and bronchitis coughs as shown in figure 1.2. Abeyratne et al. used a combination of time-series statistics, formant-frequency tracking and general temporal-spectral energy-based features to build a logistic regression model to classify coughs as pneumonia or non-pneumonia. Their model was able to achieve a sensitivity of 94% and specificity of 75% based on parameters extracted from the cough sounds alone through non-contact microphones.

In evaluating cough sounds, a part of the challenge of this problem is the lack of agreement

between professionals as properties of coughs are not well defined. In a study conducted by Swarnkar et al. [17] two domain experts were asked to score coughs as wet or dry, the two experts only agreed on 80% of their cough scores. Note that the accepted approach in research is to ignore the samples in which the annotators are in disagreement and only evaluate the samples in which they are in agreement. Swarnkar et al. use a variety of signal processing methods to look at the cough sounds including analyzing the spectral energy, temporal envelope, and time-independent waveform statistics such as kurtosis. The authors classified wet and dry coughs using a Logistic Regression Model and were able to reach a recall of 55% and specificity of 93% when trained on 536 samples.

A study by Al-khassaweneh and Abdelrahman [1] showed that coughs from asthmatic patients have different energy signature than coughs from non-asthmatic patients. The authors used Wigner Distribution functions and Wavelet Packet Transforms to analyze the time-frequency energy distributions of cough sounds from asthmatic patients and non-asthmatic patients. Their analysis showed that the coughs from asthmatic patients had a measurably higher energy signatures, especially in the low frequency bins, than the coughs from the non-asthmatic patients.

A summary of the results of various cough classification works is presented in figure 1.1. Note that in the table KNN stands for K-nearest neighbors and LRM stands for Logistic regression model. As COVID-19 is known to cause both wet/dry coughs and lead to many cases that are asymptomatic, this research will consider multiple methods including Higher order statistics rather than some of those previously researched methods. For example, spectral and temporal thresholding explored by Chatzarrin [7] will likely not be robust enough to detect COVID-19 coughs. Through a body of works [4] it has been shown that speech and vocal audio have non-linear properties. It was explored in the context of speech synthesis, in which purely linear methods produced intelligible results but were lacking "*naturalness*" in their results [2]; where "*naturalness*" in this context relates to the human-like quality of speech rather than robot-like qualities that are clearly not from a human source. In this research we will be extending the non-linear perspective in analyzing the naturalness of voice to cough sound sample.

Table 1.1. Previous cough classification works showing the classification task, results, partially or fully automatic, number of coughs in the dataset, and the methods employed.

Cough Classification Works					
Publication	Task	Results	Auto	Coughs	Methods
Al-khassaweneh [1]	Asthma	Sensitivity: 88%	Fully	24	Spectral estimation with KNN
Subburaj [12]	Intensity	Accuracy: 98%	Partially	215	Temporal energy-based regression
Swarnkar [17]	Wet/Dry	Sensitivity: 55% Specificity: 93%	Fully	536	A variety of features with a LRM
Chatrzarrin [7]	Wet/Dry	Sensitivity: 100% Specificity: 100%	Fully	8	Spectral / temporal thresholding
Abeyratne [15]	Pneumonia	Sensitivity: 80% Specificity: 73%	Fully	440	A variety of features with an LRM
Botha [6]	Tuberculosis	Sensitivity: 82% Specificity: 95%	Fully	518	Log-spectral bands with an LRM and clinical metrics

Previous work by Shams et al. [13] specifically used higher order statistics to analyse tracheal breath sounds to detect Obstructive Sleep Apnea (OSA). While not applied to coughs, the research was able to observe significant differences in the features observed through bispectral analysis (higher order statistics) between non-OSA individuals and those with sever OSA. The study explains that spectral features fail to account for nonlinear behavior, resulting from the turbulence of air in the trachea, in tracheal breath sounds. Further expanding that those features do not reveal any information on the phase-related characteristics of the signals. The bispectral analysis is a Higher Order Statistic that reveals nonlinear processes in the signal by detecting interactions of correlated harmonics in which difference frequencies are coupled, due to signal deviating from a purely Gaussian model. Bispectral analysis is able to measure both the amplitude and the phase. This has the added benefit of the analysis being blind to additive Gaussian noise recorded alongside the non-Gaussian signal, making the approach more robust

against background noise in the signal.

A work by Poreva et al. [10], uses bispectral analysis to examine chronic obstructive pulmonary disease (COPD) evaluating respiratory sounds. COPD is a widespread occupational disease often due to the exposure to lungs toxic substances in gases and vapors that are in the air and are due to industrial, technical processes and combustion products, substances contained in the smoke. By measure the maximum value of the bicoherence coefficient, the authors were able to perceive a difference between healthy and COPD-sick individuals. Additionally, the authors noted that evaluating the modulus of skewness coefficient can serve as an additional specifying characteristics of COPD-sick respiratory signals.

As discussed above, previous research seems to indicate that higher order statistics are able to capture disease related features in respiratory signals. While bispectral analysis has not yet been used in a cough classification research, to the knowledge of the author, this study will be applying bispectral analysis to cough classification of the COVID-19 pulmonary disease.

Chapter 2

Dataset

2.1 Audio Data

There are a range of sensors utilized by health professionals to measure and extract information related to a disease's symptoms for individuals. An investigation by Drugman et al. [3] explored different sensor options including contact and non-contact microphones, electrocardiography sensors, chest belts, accelerometers and thermistors placed over under the patient's nose. In examining the best sensor to extract information about coughs, Drugman et al determined that a single non-contact microphone contains significantly more information about the cough than other sensors. This has the added benefit of making it usable by anyone with a smartphone as the phone's microphone can serve as the sensor. Presumably, adding additional information such as the individuals age and temperature can improve a classifier's performance but that is beyond the scope of this study.

2.2 Data Aggregation

As this pandemic has reached international levels, there have come together multiple initiatives to collect audio data from individuals diagnosed with COVID-19. Initiatives including Corona Voice Detect and Coughona have built a web landing page in which the page records the individuals' voice as it prompts them to cough or utter specific phrases. It further contains a form in which the individual can report additional metadata about themselves, such as if they

have a history of smoking, while remaining anonymous.

This study utilizes the data collected in Corona Voice Detect in partnership with voca.ai. The data is continually being collected, if you or someone you know is positively diagnosed with COVID-19 please consider filling out the form at <https://voca.ai/corona-virus/>

The data collected consists of:

- a unique ID for the individual and their device
- the individual's age
- the individual's gender
- the individual's height (optional)
- the individual's country of residence (optional)
- the individual's smoking habits (optional)
- the individual's status, are they in isolation or not (optional)
- the individual's diagnosis status (optional)
- the individual's temperature (optional)
- A free-form text of how the individual feels (optional)
- The date and time of the upload (optional)
- Audio recording of the individual coughing three times
- Audio recording of the individual saying "Ahhh" for as long as they can
- Audio recording of the individual saying "Eeee" for as long as they can
- Audio recording of the individual saying "Ohhhh" for as long as they can

- Audio recording of the individual counting from one to twenty
- Audio recording of the individual saying the alphabet from a to z
- Audio recording of the individual reading a segment from a story

The dataset available as of the time of this writing contains 30 positively diagnosed with COVID-19 and 1811 not diagnosed with COVID-19. There are further initiatives within clinical settings in which doctors are keeping close statistics of patients and recording audio samples from patients daily in a very controlled setting. At the time of this writing, that data is not yet available and contains confidential private information, and therefore will not be included in this study. It remains that the results discussed here can be extended or improved by incorporating that data when it becomes available.

2.3 Data Characteristics

The Corona Voice Detect initiative allows individuals to upload their information and record their voice without enforcing a controlled setting or verifying that the individual provides their accurate information. As such, some samples have been discarded as they contain corrupted entries such as entries without any voice recordings or recordings that are purely noise. Furthermore, not all individuals provided voice samples for all audio samples, namely coughing, counting, "ahhh", "eeee", etc., and thus the number of entries for each type of audio sample is disproportional.

2.3.1 Gender Ratio

Researchers have shown that each gender has a different vocal characteristic that is dependent on timber and pitch [9]. As such, it may be that biases in gender ratios within our dataset may hinder our model's performance and remains as an area for further exploration. The dataset contains the following ratios:

Positively diagnosed with COVID-19: Male: 18 , Female: 7

Male to Female ratio: 72% / 28%

Not diagnosed with COVID-19: Male: 1006 , Female: 537

Male to Female ratio: 65.2% / 34.8%

2.3.2 Age Distribution

Research has shown that there are acoustical changes in human voices across different age groups [14], that vary further between genders. As such, the distribution of ages within our dataset may affect the performance of our model across input from different age groups. This should resolve itself with a sufficiently large dataset that contains enough training samples from each age group. This remains as an area for further exploration. The dataset contains the following ratios:

Positively diagnosed with COVID-19:

mean: 27.8 , median: 25.0 , standard deviation: 9.5

youngest age: 6, oldest age: 52

Not diagnosed with COVID-19:

mean: 37.3 , median: 33 , standard deviation: 15.8

youngest age: 5 , oldest age: 89

2.3.3 Audio Samples

As we rely on individuals to record themselves in whatever setting they may be and with no supervision, some of the samples are corrupted and unusable further limiting our training set. Waveforms of both a healthy and COVID-19 sick individual coughing is visible in figure 2.1. Waveforms of both a healthy and COVID-19 sick individual saying "ahhh", "eeee" and "ooo" is visible in figure 2.2, figure 2.3 and figure 2.4 respectively. A count of each audio sample for the healthy and COVID-19 sick audio sets is shown in figure 2.5.

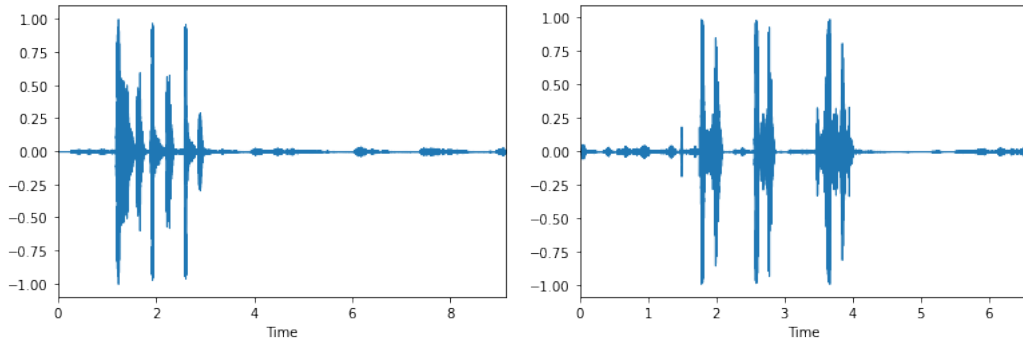


Figure 2.1. Waveform visualization of an audio sample from a healthy (left) and COVID-19 sick (right) coughing three times.

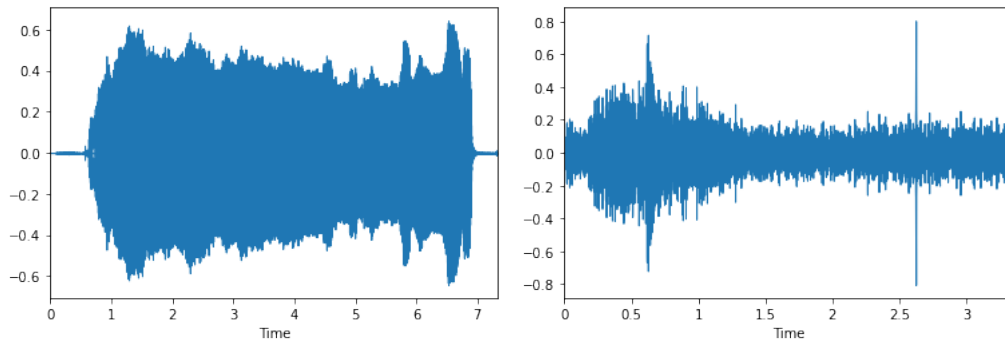


Figure 2.2. Waveform visualization of an audio sample from a healthy (left) and COVID-19 sick (right) saying "Ahhh".

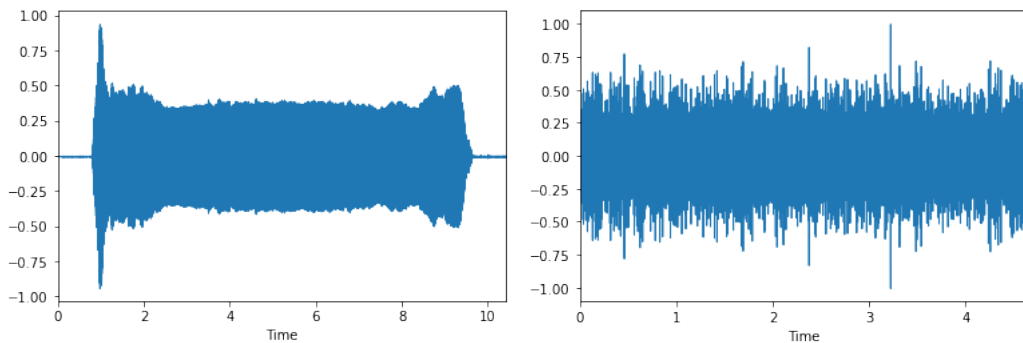


Figure 2.3. Waveform visualization of an audio sample from a healthy (left) and COVID-19 sick (right) saying "Eeee".

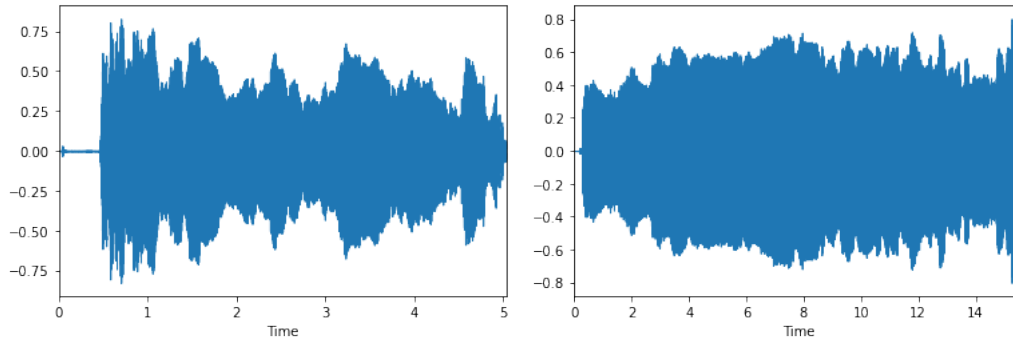


Figure 2.4. Waveform visualization of an audio sample from a healthy (left) and COVID-19 sick (right) saying "Oooo".

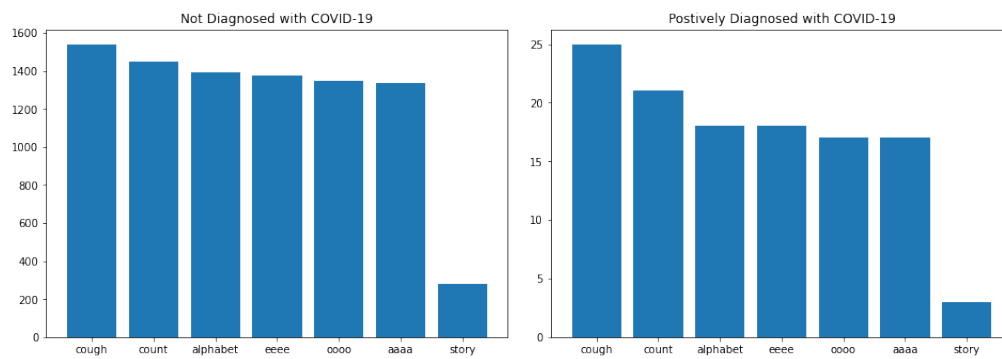


Figure 2.5. The counts of audio samples across the different prompts for healthy (left) and COVID-19 sick (right).

Chapter 3

Audio Signal Analysis

3.1 Motivation

A common method of modeling vocalisation, and specifically speech, is the linear model of speech production known also as a source-filter model. This model has not been applied to cough analysis, as it mostly focuses on the changes in the vocal tract that happen during speech. We will refer to this model as a general motivation for considering the non-linear aspects of audio signal that we are considering as a distinguishing feature between COVID-19 sick and normal general vocal audio samples. In our dataset the sounds include both spoken sounds (alphabet, numbers), single vowels (ahhh, eee, ooo) and coughs. To focus on the sounds from the lungs and trachea, we consider the linear source-filter models as a reason to remove spectral information that is related to resonances of the vocal tract, so we can separate and view the difference in the sounds originating from the lungs and sounds originating from the vocal cords.

Source-filter model separates the vocal sound into two stages: a sound source that represents the air flow from the lungs and the flapping of the vocal cords in the case of voiced speech, which is then shaped or filtered by the resonant properties of the vocal tract. Each element in this two stage model has its own statistical properties, including spectral shape and higher-order statistical fine structure. Since most of the changes in the energy of the different frequency elements is carried out by the vocal tract anterior to the sound source, which in the case of voiced sounds is the entire supra-glottal vocal tract including parts of the oral cavity and

the nasal cavity, we can intentionally remove the spectral aspects by normalization of the signal spectrum.

Common audio analysis methods extract features, such as spectral features, that fail to account for nonlinear behavior. While previous research on cough classification has primarily focus on such features, breath sound research has found that higher order statistics come up with rich features that are also able to detect the phase-related characters. Recent success in speech synthesis using Neural Networks shows that high quality sounds can be generated by using non-linear model that is mapping spectral information to sound waveforms, which means that using higher order statistics we can capture those non-linear properties for a deeper analysis. Bispectral features were suggested as powerful features for analysis of complex systems, such as synthesized voice models and breath sounds, as well as other applications such as EEG. Bispectral analysis has the added benefit of inherently filtering out Gaussian noise from the sample.

Unlike most of the literature on bispectral analysis that only considers the question of non-linear modeling, we do not assume here necessarily that the differential sounds signatures of COVID-19 sick sounds to healthy sounds are differences in their non-linearity. Although intuitively it may be plausible that healthy sounds will have less obstructions in the airflow which would therefore make it less likely for some coupling or statistical dependencies across frequencies and their respective phases to emerge. Our main motivation in looking into bispectrum is that by removing the spectral part of the signal we can focus on more subtle sound signatures, and that the machine learning methods can actually capture the patterns in bispectrum, which otherwise would be missing if one were only to look at power-spectral or the Fourier Magnitude of the signal.

3.2 Signal Analysis

In our approach we consider higher order statistics to evaluate the audio samples to explore if samples from healthy individuals versus those that are diagnosed with COVID-19 have a fundamental difference due to non-linear effects on the airway caused by the symptoms. A core element of this study focuses on the higher order statistical analysis of bispectral. Bispectral analysis reveals coupling between frequencies by measuring their coupling within the audio samples. The bispectral analysis does so by evaluating the higher-order correlations of the signal in the Fourier domain.

3.2.1 Fast Fourier Transformation

In analysing an audio signal $y(k)$, it is first decomposed according to the Fourier transform:

$$Y(\omega) = \sum_{k=-\infty}^{\infty} y(k)e^{-ik\omega}$$

with $\omega \in [-\pi, \pi]$.

The study deals with audio samples with a limited sample rate and as such employs a discrete Fourier Transformation using the Fast Fourier Transformation. We evaluate the Fast Fourier Transformation across time by looking at the spectrogram am of the audio samples. This introduces an additional variable around the window size in the spectrogram am setting the frequency resolution. spectrogram am of coughs from both healthy and COVID-19 sick individuals are visible in figure 3.1.

3.2.2 Mel-Frequency Cepstral Analysis

A common feature in audio analysis is the Mel-Frequency Cepstral Analysis (MFCC). To understand the nature of this representation we need first to explain the concept of cepstrum and then apply the mel-scale transformation to the frequency dimension. The main motivation in

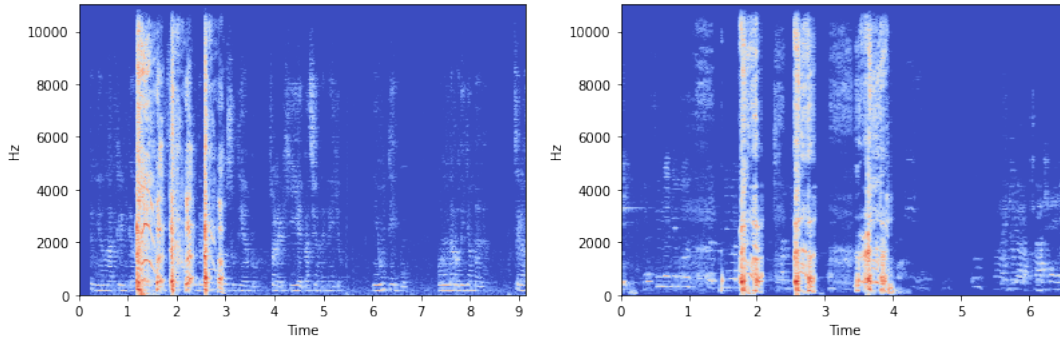


Figure 3.1. spectrogram am of coughs from a healthy individual (on the left) and a sick individual diagnosed with COVID-19 (on the right).

developing this representation is to separate the aspects of vocal tract spectral shape from that of the excitation signal based on the source-filter model discussed earlier. If we assume that the speech or vocalisation signal is produced by a periodic excitation of the glottal pulses passing through a vocal tract, the frequency representation of this model is expressed as multiplication of the frequency shape of the excitation and the vocal tract as follows:

$$Y(f) = H(f)E(f)$$

Where $H(f)$ is the frequency shape of the vocal tract, $E(f)$ is the frequency analysis of the excitation and $Y(f)$ is the frequency analysis of the resulting vocal signal. An underlying assumption in cepstral analysis is that the excitation $E(f)$ contains a lot of spectral details, namely narrow spectral peaks that originate from the periodicity of the glottal pulse. In contrast, the shape of vocal tract $H(f)$ is relatively smooth as it contains broad spectral magnification regions due to the formant structure of human speech. An idea of separation of these two components is known as homomorphic processing and it done by taking the log of the absolute value of the frequency response:

$$\log |Y(f)| = \log |H(f)| + \log |E(f)|$$

$Y(f)$ can then be separate into the two components $E(f)$ and $H(f)$ by performing an additional step of frequency analysis. Based on the previous assumption that $E(f)$ contains a lot of spectral details it will now show through the high frequency elements, whereas $H(f)$ which is relatively smooth will show through the low frequency elements

This is easily visible in figure 3.2 showing the mapping of the individual frequencies, $E(f)$, as the Harmonics structures mapped to a higher quefrequency and the overall trend shape, $H(f)$, as the spectral enveloped mapped to the lower quefrequency. The frequency analysis of the original sound signal is known as spectrum whereas the second application of frequency analysis to the log-spectrum is known as cepstrum. The smooth spectral envelope due to formant structure on the left permeated by the sharp harmonic structure due to the excitation signal. When translated once again into frequency domain after doing the log operation, the fast / sharp harmonic detail are translated to a peak further on the right side of the cepstrum plot, while the smooth envelope is represented by the left part of the cepstral graph.

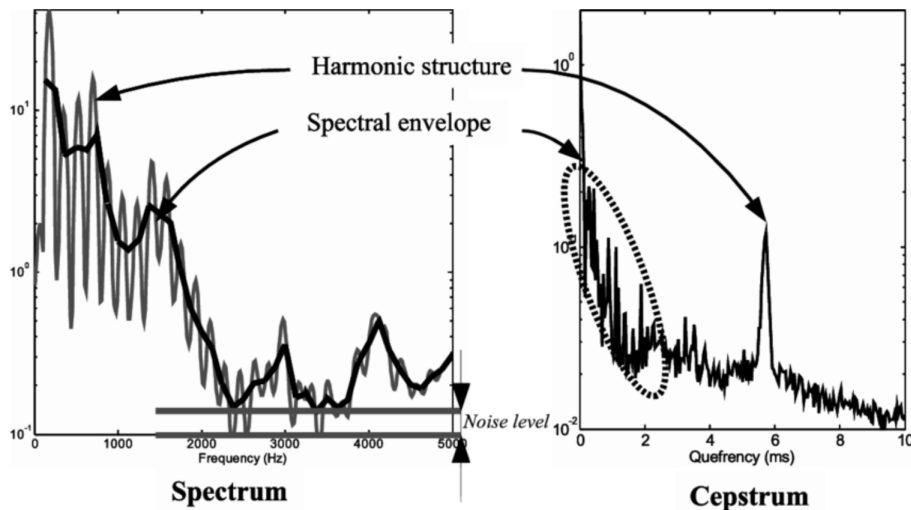


Figure 3.2. Diagram showing the mapping from frequency to quefrequency in which the individual harmonic structure is mapped to a high quefrequency and the underlying spectral envelope is mapped to a lower quefrequency. [5]

Homomorphic processing allows for a simple separation of the vocal tract components from the excitation detail. It is important to note that operation is performed on the absolute value

of the Fourier transform, thus eliminating phase information. The components in cepstral analysis are known as cepstral coefficients, and since the analysis is done using discrete Fourier transform, the cepstral coefficients are discrete as well. Taking a limited number of these coefficients (commonly 20-40 first coefficients) effectively separates the envelope from the excitation.

An additional step in the Mel-Frequency Cepstral Analysis is to show the cepstral analysis on a Mel-Scale. The Mel-Scale is a transformation of the usual frequency unit of Hertz into a partially logarithmic scale, based on psychoacoustic human perception of the sensation of the pitch. Broadly, the Mel-Scale is judged by listeners to be linear up to about 500 Hz, and then logarithmic for higher frequencies. The mapping of Hz to the Mel-Scale is visible in figure 3.3. The mapping of Hz to Mel is demonstrated in the following graph:

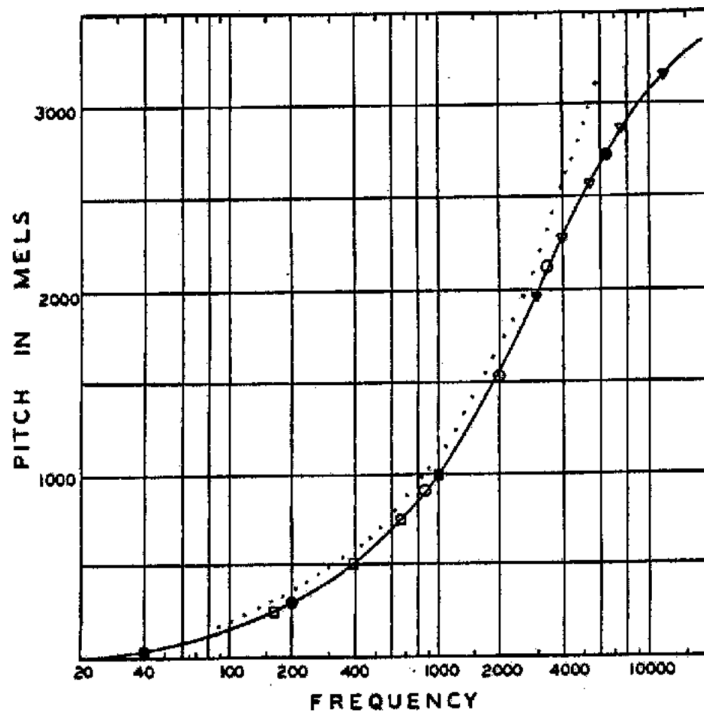


Figure 3.3. Mapping of Frequency in Hz to Pitch in Mel-Scale [16].

In performing a Mel-Frequency Cepstral Analysis on a discrete signal, the quefrequencies are grouped into critical bands. Critical bands are the psychoacoustic phenomena that refers to minimal changes frequency that are perceived by humans as two separate tones. This gives the

Mel warping of the spectral analysis across both the window size and the number of bands when capturing frequencies of up to 4kHz. A visualization of the mapping of the MFCC's bands to the frequencies is visible in figure 3.4.

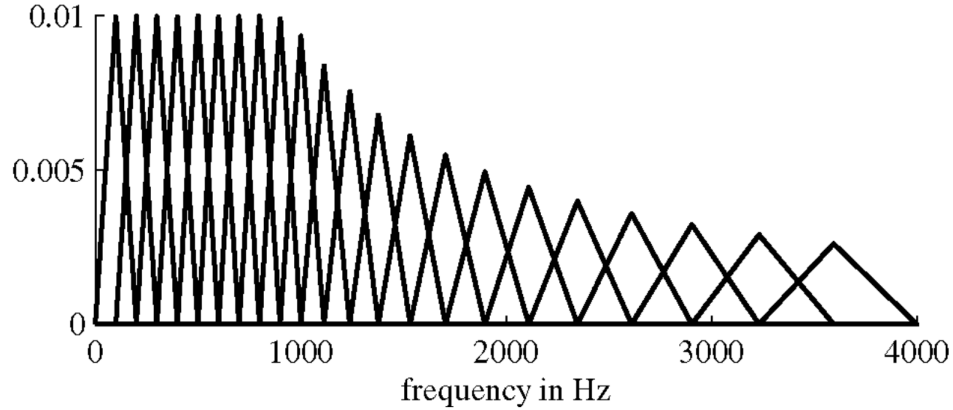


Figure 3.4. A visualization of 20 critical bands in the Mel-Scale mapped to the frequencies. [11]

3.2.3 Bicoherence and Biphas

In moving past the spectral envelope of the audio signal and into the higher order statistics of the sound, bispectrum can capture coupling between frequencies. Bispectrum is used to measure third-order correlations

$$B(\omega_1, \omega_2) = Y(\omega_1)Y(\omega_2)Y^*(\omega_1 + \omega_2)$$

where * denotes the complex conjugate. The bispectral response reveals correlations between the triple of harmonics $[\omega_1, \omega_1, \omega_1 + \omega_1]$, $[\omega_2, \omega_2, \omega_2 + \omega_2]$, $[\omega_1, \omega_2, \omega_1 + \omega_2]$, and $[\omega_1, -\omega_2, \omega_1 - \omega_2]$ [2]. Rather than evaluating the complex bispectrum, one can take the magnitude:

$$|B(\omega_1, \omega_2)| = |Y(\omega_1)| \cdot |Y(\omega_2)| \cdot |Y(\omega_1 + \omega_2)|$$

whereas the biphase evaluates the phase:

$$\angle B(\omega_1, \omega_2) = \angle Y(\omega_1) + \angle Y(\omega_2) - \angle Y(\omega_1 + \omega_2)$$

Finally, the bicoherence is the normalized bispectrum:

$$B_c(\omega_1, \omega_2) = \frac{Y(\omega_1)Y(\omega_2)Y^*(\omega_1 + \omega_2)}{\sqrt{|Y(\omega_1)Y(\omega_2)|^2|Y(\omega_1 + \omega_2)|^2}}$$

where the magnitude is in the range $[0, 1]$.

Bicoherence, $|B(\omega_1, \omega_2)|$, measures the amount of quadratic phase-coupling that occurs in the audio signal between any two of its frequency components, due to their nonlinear interactions. Quadratic phase-coupling is the result of nonlinearity in the signal in which two frequency components ω_1 and ω_2 , with phases ϕ_1 and ϕ_2 respectively, are simultaneously present in the signal along with a frequency at $\omega_3 = \omega_1 \pm \omega_2$ with $\phi_3 = \phi_1 \pm \phi_2$. Two frequency components ω_1 and ω_2 are completely phase-coupled when their bicoherence index is 1 and have no quadratic phase-coupling when their bicoherence index is 0.

Figure 3.5 shows the bicoherence of a cough sound sample from a healthy individual whereas figure 3.6 shows the bicoherence of a cough sound sample from a sick individual diagnosed with COVID-19. Similarly figure 3.7 shows the biphase of a cough sound sample from a healthy individual whereas figure 3.8 shows the biphase of a cough sound sample from a sick individual diagnosed with COVID-19. A noticeable difference is that the sample from the sick individual diagnosed with COVID-19 has a significantly higher maximum value in their bicoherence. While it may be difficult to comment on the biphase, a classification algorithm may be able to learn to distinguish between the biphase of healthy and COVID-19 sick individuals.

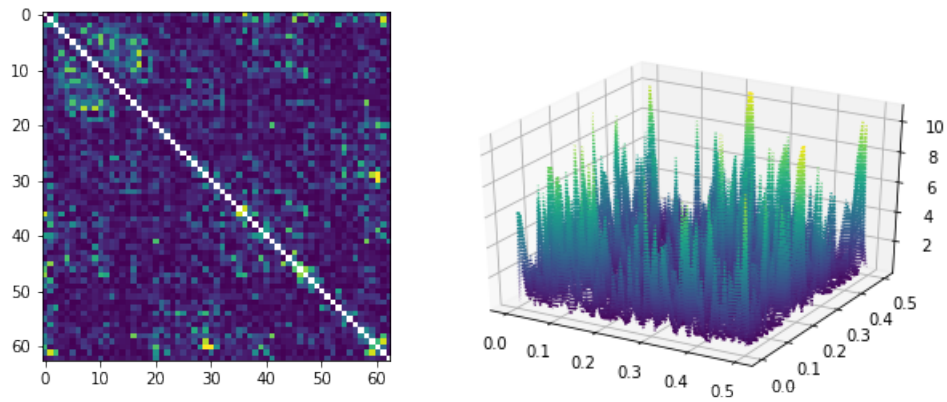


Figure 3.5. Bicoherence of a cough from a healthy individual in a two dimensional representation (left) and three dimensional representation (right)

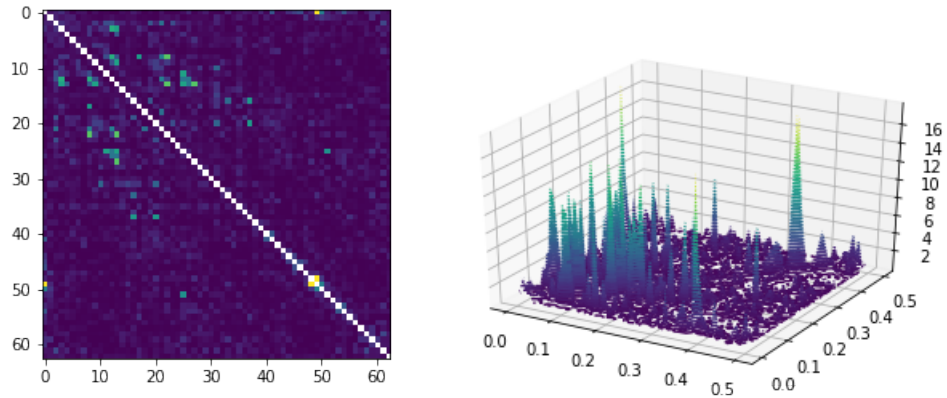


Figure 3.6. Bicoherence of a cough from a sick individual diagnosed with COVID-19 in a two dimensional representation (left) and three dimensional representation (right)

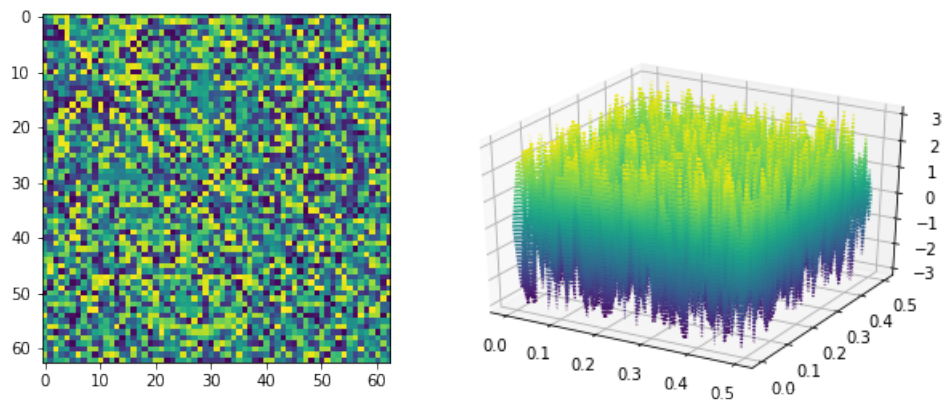


Figure 3.7. Biphasic nature of a cough from a healthy individual in a two dimensional representation (left) and three dimensional representation (right)

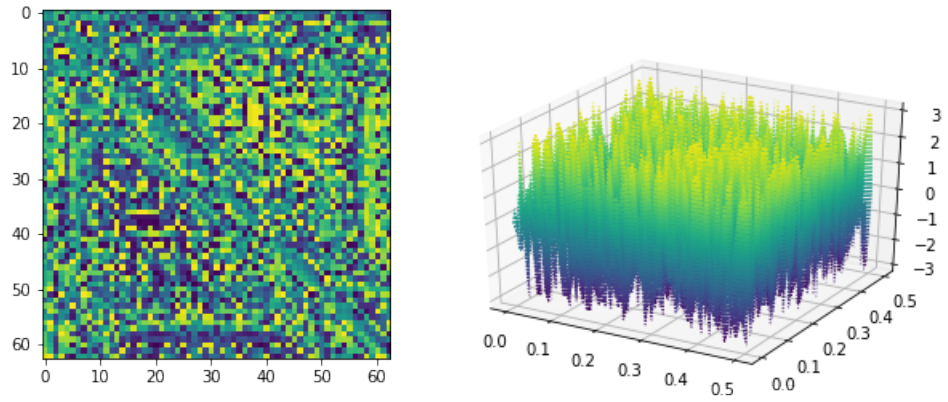


Figure 3.8. Biphase of a cough from a sick individual diagnosed with COVID-19 in a two dimensional representation (left) and three dimensional representation (right)

Chapter 4

Classification

Classification entails labeling an input cough sound as either healthy (negative) or sick with COVID-19 (positive). In assigning the label, one needs to keep in mind that the classification results in *truepositives* and *truenegatives* in which cough sounds are labeled correctly, and *falsepositives* and *falsenegatives* in which a healthy individual is incorrectly labeled as COVID-19 sick and a COVID-19 individual is incorrectly labeled as healthy, respectively. These definitions help us define the following metrics:

$$Accuracy = \frac{TruePositives + TrueNegatives}{TruePositives + TrueNegatives + FalsePositives + FalseNegatives}$$

properly named as it represents us how accurate the labels are.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

representing how trustworthy the positive labels are so if a cough sound is labeled as COVID-19 sick what is the probability that it is indeed correct.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

representing how capable the model is of capturing all of the COVID-19 sick cough

samples, alternatively representing how many COVID-19 sick coughs it mislabels. Lastly, the metric

$$F1 - Score = \frac{2 * Recall * Precision}{Recall + Precision}$$

represents a combination metric of precision and recall, where an *F1Score* of zero indicates that the *Recall* and/or *Precision* is zero and as the *F1Score* nears one it indicates that the classification is nearing perfect *Precision* and *Recall*.

Due to the contagious nature of COVID-19, as the virus spreads by respiratory droplets when someone with the disease coughs, sneezes or talks to another individual six feet away or less, *Accuracy* is not the core metric we need to track. The consequences of mislabeling an individual as healthy can be significant in allowing the disease to continue spreading throughout the population, as such *Recall* needs to be the primary metric we examine. *Recall* can't be the sole metric we examine as a perfect recall is possible by simply labeling all individuals as positive, instead we can use the *Accuracy* as the secondary metric.

4.1 Methods of Classification

There exist numerous methods of classification, with different methods being more or less appropriate depending on the nature of the data. In this study we explore Logistic Regression, Support Vector Machine using a linear kernel, Support Vector Machine using a radial basis function (RBF), Random Forest, Multilayer Perceptron and Convolutional Neural Network. As can be seen in figure 4.1 the proper model depends on the dispersion of the two classes. The bicoherence and biphasic output a 63x63 matrix, as we are using a resolution of 128 in our spectrogram am, the MFCC outputs a 40x301 matrix. The classification models listed, other than the Convolutional Neural Network, do not work with a matrix and instead perform on a multidimensional array, as such we flatten our 63x63 matrix into an array of 3,969 and our 40x301 matrix into an array of 12,040.

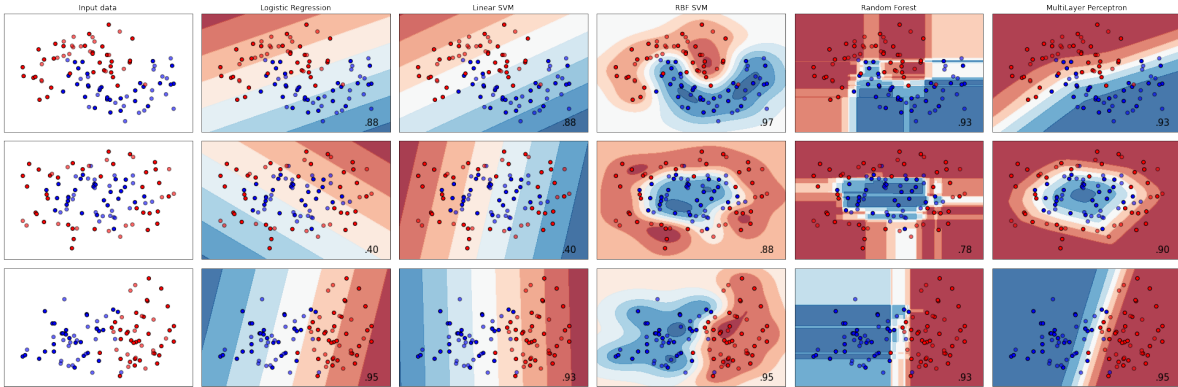


Figure 4.1. Classification results of Logistic Regression, Support Vector Machine using a linear kernel, Support Vector Machine using an RBF kernel, Random Forest and Multilayer Perceptron for different input data.

4.1.1 Logistic Regression

Logistic Regression is a linear model for classification. The linear model solves for coefficients w by minimizing the cost function:

$$\min_{w,c} \frac{1}{2} w^T w + C \sum_{i=1}^n \log(e^{-y_i(X_i^T w + c)} + 1)$$

where y is the label and X is the input. We will be using the scikit-learn implementation of Logistic Regression.

4.1.2 Support Vector Machines

Support Vector Machine (SVM) is a classification method that solves for a hyperplane between the two labels by maximizing the margins between the hyperplane and the nearest samples for each label. As can be seen in figure 4.1, this can behave similarly to Logistic Regression. In SVM, we can apply an additional kernel to project the data into a higher dimension where the hyperplane in the higher-dimensional representation is able to better label

the data. We will be using radial basis function (RBF) kernel in which the kernel is defined as

$$K(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}}$$

where σ is a free variable and x and x' are samples from the training dataset. We will use a scaled σ as $1/(\text{number of features} * \text{variance of } X)$ and an auto σ as $1/(\text{number of features})$. We reference the traditional SVM as Linear SVM and the SVM using the RBF kernel as Non-linear SVM. We will be using the scikit-learn implementation of Support Vector Machine.

4.1.3 Random Forest

Random forest is a classification method that constructs a multitude of decision trees on different subsets of the training dataset. Each tree uses a sub-sample of the dataset and may use a subset of the input features. As decision trees tend to over-fit, Random Forest overcomes this through the randomness of the sub-samples of the training dataset and the randomness of the subset of the input features for each tree and then averages the final results of the many trees in the forest to return the final prediction. We will be using the scikit-learn implementation of Random Forest.

4.1.4 Multilayer Perceptron

The Multilayer Perceptron (MLP) is a an artificial neural networks that adjusts its weights iteratively to learn to classify the training data. The MLP consists of an input layer followed by a variable number of layers, known as the depth of the network, in which every layer consists of a variable number of neurons that are fully connected to the layer before them and after them and lastly an output layer exists that receives the last hidden layer and outputs the final classification. The weights of the connection between the neurons are learned iteratively. Each neuron in the MLP has a non-linear activation function, the study uses a ReLu activation implementation in

which

$$f(x) = \max(x, 0)$$

The neural network iteratively updates its weights by going over all of the training data numerous times, known as an epoch, minimizing its error and back-propagating the adjustments to the weights based on a learning rate. We will be using the scikit-learn implementation.

4.1.5 Convolutional Neural Networks

Recent advances in machine learning have led to deep learning in which model learn parameters in sophisticated architectures. In evaluating an RGB image in a neural network there are specialized approaches that enable the better processing and modeling of the rgb pixel matrix. A convolutional neural network applies filters and max pooling operations that are learned through training to extract features of features required to perform complex operation it was trained to perform. Figure 4.2 shows a visual diagram of the filter convolutional layer and max pooling layers in a convolutional neural network. As can be seen in figure 4.2, the layers are applied to a small subset of pixels from the input image and the operation repeats until the all the pixels in the input image are considered at least once. In this way, the size and resolution of the input image directly affects the time required to send the image through the network to perform the intended operation.

For our application, a convolutional neural network allows us to perform a classification operation without flattening the input data. This let's the neural net utilize the additional localization information across multiple frequency couplings revealed in the bicoherence, biphasic and MFCC, that is lost due to flattening in the other approaches discussed previously. As we only have a single channel, rather than three channels for RGB pixels, the neural network will use a single channel. The architecture of the Convolutional Neural Network we utilize is visible in figure 4.3. The model consists of four Conv2D convolution layers. Each convolutional layer is

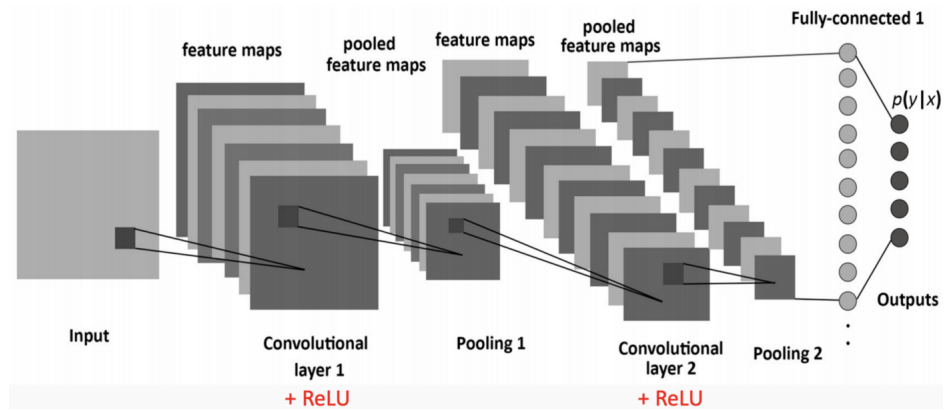


Figure 4.2. A diagram of a convolutional neural network

followed by a pooling layer of MaxPooling2D and the final convolutional layer is followed by a GlobalAveragePooling2D type. The Max Pooling layers take the maximum size for each window and the Global Average Pooling layer takes the average which is suitable for feeding into the final dense output layer. The pooling layers reduce the dimensionality of the model, reducing the parameters needed to be learned, and therefore reducing the computation requirements. The dropout layer further reduces the training time while also limiting overfitting. The output layer has only 2 nodes as it performs the binary classification of COVID-19 sick versus healthy.

Layer (type)	Output Shape	Param #
conv2d_25 (Conv2D)	(None, 39, 300, 16)	80
max_pooling2d_25 (MaxPooling)	(None, 19, 150, 16)	0
dropout_25 (Dropout)	(None, 19, 150, 16)	0
conv2d_26 (Conv2D)	(None, 18, 149, 32)	2080
max_pooling2d_26 (MaxPooling)	(None, 9, 74, 32)	0
dropout_26 (Dropout)	(None, 9, 74, 32)	0
conv2d_27 (Conv2D)	(None, 8, 73, 64)	8256
max_pooling2d_27 (MaxPooling)	(None, 4, 36, 64)	0
dropout_27 (Dropout)	(None, 4, 36, 64)	0
conv2d_28 (Conv2D)	(None, 3, 35, 128)	32896
max_pooling2d_28 (MaxPooling)	(None, 1, 17, 128)	0
dropout_28 (Dropout)	(None, 1, 17, 128)	0
global_average_pooling2d_7 ((None, 128)	0
dense_7 (Dense)	(None, 2)	258

Total params: 43,570
 Trainable params: 43,570
 Non-trainable params: 0

Figure 4.3. A diagram of the architecture of the convolutional neural network

Chapter 5

Classification Results

The classification methods covered in the previous chapter were applied to the COVID-19 cough dataset. The data was split into an 80/20 split, with 80% of the data used in the training set and 20% used in the test set. The classification *accuracy*, *precision*, *recall* and *f1 – score* are indicated for the classification’s performance on the test set not the training set.

5.1 Bicoherence

The results for the classification performances on the bicoherence are visible in table 5.1. The bicoherence data used as input is a 63x63 matrix, using a resolution of 128 in the spectrogram am, that is flattened into an array for Logistic Regression, Random Forest, Multilayer Perceptron and Support Vector Machines. The Convolutional Neural Network uses the 63x63 matrix as the direct input.

5.2 Biphase

The results for the classification performances on the biphase are visible in table 5.2. The biphase data used as input is a 63x63 matrix, using a resolution of 128 in the spectrogram am, that is flattened into an array for Logistic Regression, Random Forest, Multilayer Perceptron and Support Vector Machines. The Convolutional Neural Network uses the 63x63 matrix as the direct input.

Table 5.1. Bicoherence classification results for different classification methods.

Classification Method	Accuracy	Precision (Sick/Healthy)	Recall (Sick/Healthy)	F1-Score (Sick/Healthy)
Logistic Regression	65%	40% / 73%	31% / 80%	40% / 68%
Random Forest	71%	100% / 71%	4% / 100%	7% / 83%
Multilayer Perceptron	70%	0% / 70%	0% / 100%	0% / 82%
Support Vector Machine Linear	64%	39% / 73%	35% / 77%	37% / 75%
Support Vector Machine Non-Linear (Auto Sigma)	71%	55% / 73%	23% / 92%	32% / 81%
Support Vector Machine Non-Linear (Scale Sigma)	65%	42% / 75%	42% / 75%	42% / 75%
Convolutional Neural Network	70%	70% / 71%	96% / 17%	81% / 28%

5.3 Mel-Frequency Cepstrum

The results for the classification performances on the Mel-Frequency Cepstrum are visible in table 5.3. As the lengths of the audio samples varied, all samples were adjusted to a length of seven seconds. Audio samples longer than seven seconds were cropped at seven seconds, whereas those shorter than seven seconds were padded with zeros until the seven second mark. At a resolution of 40 and with a hop length of 512, the MFCC data used as input is a 40x301 matrix that is flattened into an array for Logistic Regression, Random Forest, Multilayer Perceptron and Support Vector Machines. The Convolutional Neural Network uses the 40x301 matrix as the direct input.

Table 5.2. Biphase classification results for different classification methods.

Classification Method	Accuracy	Precision (Sick/Healthy)	Recall (Sick/Healthy)	F1-Score (Sick/Healthy)
Logistic Regression	58%	36% / 72%	44% / 64%	40% / 68%
Random Forest	68%	0% / 49%	0% / 100%	0% / 81%
Multilayer Perceptron	65%	33% / 69%	11% / 90%	17% / 78%
Support Vector Machine Linear	48%	32% / 69%	56% / 46%	41% / 55%
Support Vector Machine Non-Linear	31%	31% / 0%	100% / 0%	48% / 0%
Convolutional Neural Network	64%	67% / 40%	89% / 14%	77% / 21%

Table 5.3. Mel-Frequency Cepstrum classification results for different classification methods.

Classification Method	Accuracy	Precision (Sick/Healthy)	Recall (Sick/Healthy)	F1-Score (Sick/Healthy)
Logistic Regression	77%	39% / 88%	50% / 83%	44% / 86%
Random Forest	84%	67% / 85%	22% / 98%	33% / 91%
Multilayer Perceptron	82%	0% / 82%	0% / 100%	0% / 90%
Support Vector Machine Linear	73%	32% / 83%	44% / 79%	37% / 83%
Support Vector Machine Non-Linear (Auto Sigma)	82%	0% / 82%	0% / 100%	0% / 90%
Support Vector Machine Non-Linear (Scale Sigma)	80%	45% / 89%	50% / 87%	47% / 88%
Convolutional Neural Network	79%	76% / 79%	43% / 94%	55% / 86%

Chapter 6

Discussion

In evaluating the results of the classifiers for the different analytic methods, we need to consider both the recall to understand how many COVID-19 sick diagnosis we are missing (increasing the risk of those individuals spreading the disease further) and the overall accuracy to measure how correct it is. In an ideal scenario both the recall and accuracy will be perfect, but in a less-than-ideal world that we reside in there results of the classifier will need to be properly incorporated into the overall healthcare workflow.

While this approach is not reliable enough to be the sole diagnosis tool, it offers a non-invasive immediate diagnosis method that can reach millions of people. Using it as the sole diagnosis leaves us prone to an individual having a false positive result, self-isolating for two weeks and then having the false notion that he or she has built an immunity to the disease. This would leave the individual vulnerable to contracting the disease later on but not approaching COVID-19 diagnosis methods due to their false notion and therefore not self-isolating further. This would further limit the individual's perception of the reliability of the classifier's diagnosis reliability. If the classifier is instead used as a trigger to get further tests at a hospital or testing center, than it can have a very valuable contribution in helping individuals take knowledgeable steps to protect themselves and society. Vice versa, it could also be used as a tool to limit testing - with a high recall rate for sick diagnosis an individual that is classified as healthy could be guided to not undergo further testing and leave the scarce testing equipment to others that are

more likely to have the disease.

Note that in the figures LR stands for Logistic Regression, RF stands for Random Forest, MLP stands for Multilayer Perceptron, SVM-L stands for Support Vector Machine with Linear kernel, SVM-NLA stands for Support Vector Machine with RBF kernel and auto gamma, SVM-NLS stands for Support Vector Machine with RBF kernel and scaled gamma, CNN stands for Convolutional Neural Network.

6.1 Bicoherence

The Bicoherence analysis using the classification methods in this study, largely tends for a high recall for the healthy set and a low recall for the COVID-19 sick set as can be seen in figure 6.1. The training and testing data are both imbalanced datasets with many more healthy samples than COVID-19 sick samples. This would largely justify why the healthy recall is so much higher as it is a much easier task (as there are many more of them). This further highlights the unique property that the Convolutional Neural Net approach was able to get a flipped result with a higher recall for the sick set rather than the healthy set.

Using our unique criteria, of wanting to maximize the accuracy but prioritizing a higher COVID-19 sick recall due to the contagious nature of the disease, a shared perspective of the two is needed to evaluate which classifier works the best. As such, the Convolutional Neural Net seems to be the best classifier with a high overall accuracy and extremely high COVID-19 sick recall.

6.2 Biphase

The biphase analysis method seems to contain the least usable information as multiple classifiers end up having a recall of 0 or close to 0 meaning that the classifier optimizes to label the entire testing set with the same label. It is interesting to note that overall the Logistic Regression classifier, which is comparatively the simplest classifier, performs the best in maximizing both recalls.

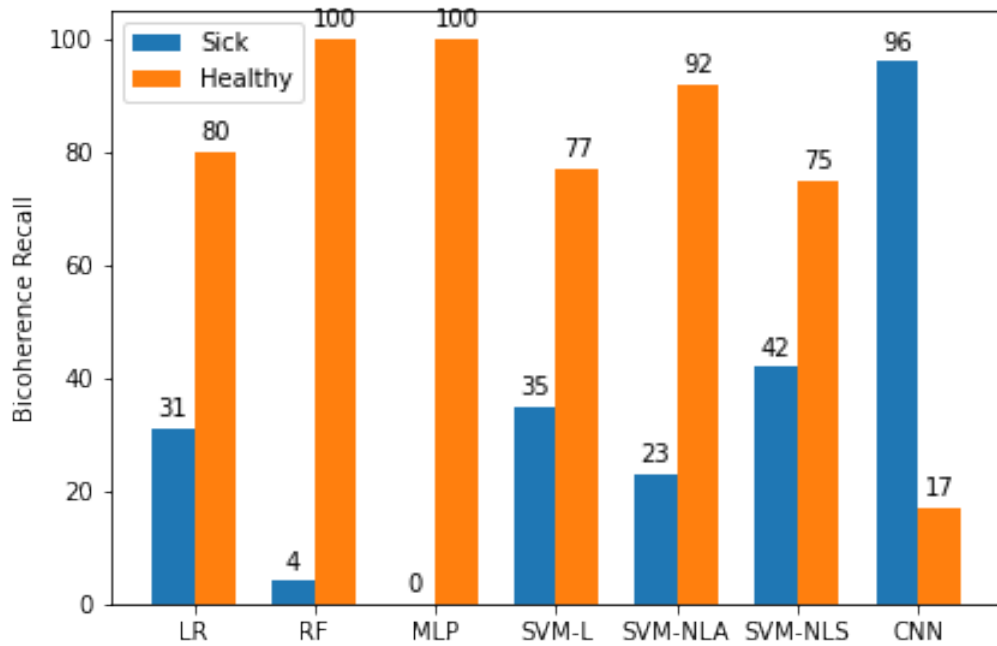


Figure 6.1. A plot of the recall score for various classification method for COVID-19 sick and healthy based on a bicoherence analysis input.

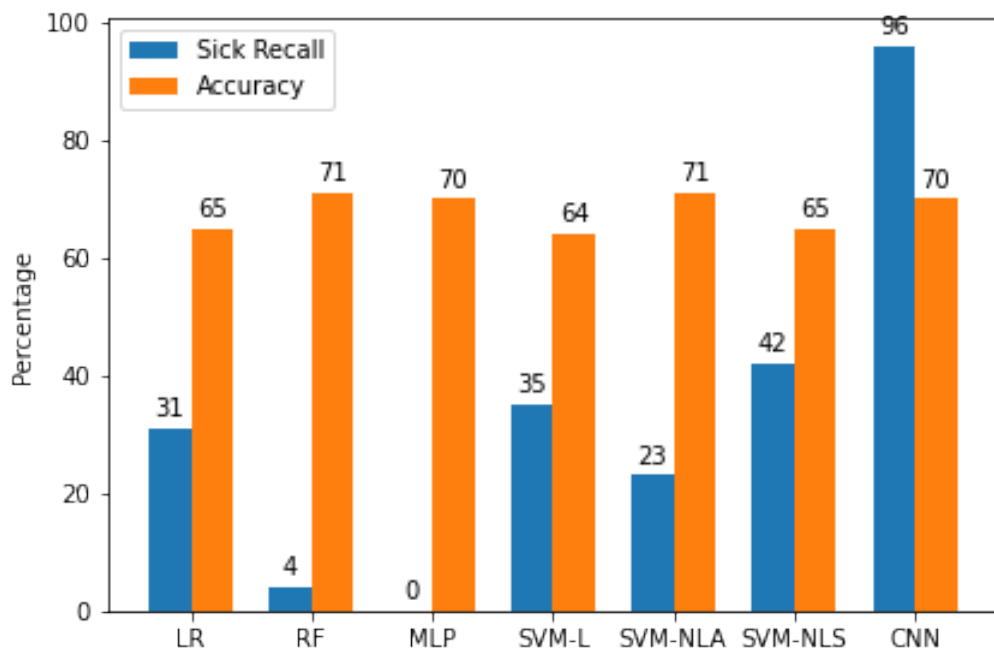


Figure 6.2. A plot of the recall score for various classification method for COVID-19 sick and the overall accuracy based on a bicoherence analysis input.

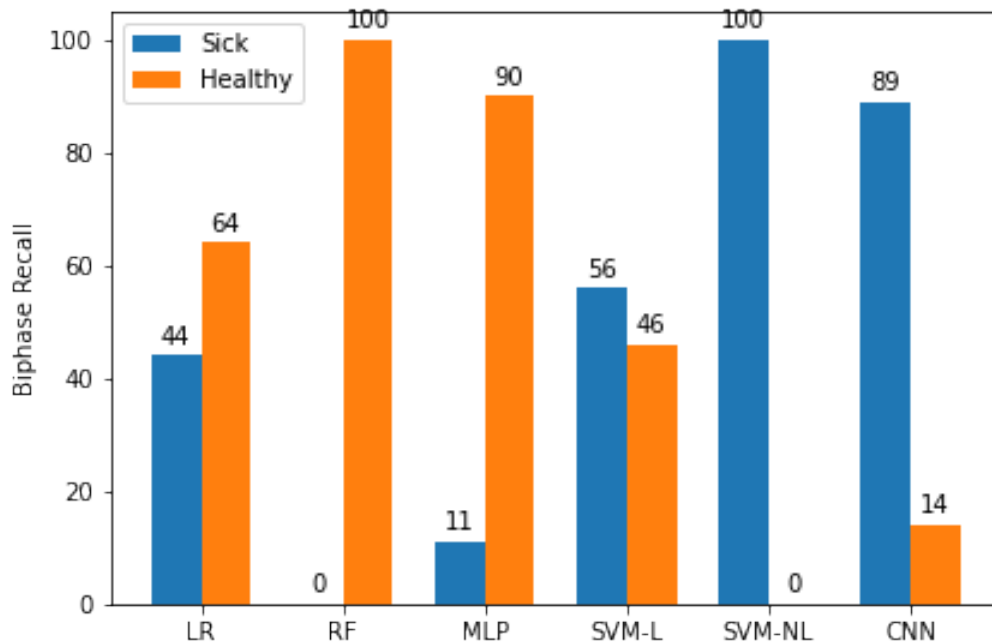


Figure 6.3. A plot of the recall score for various classification method for COVID-19 sick and healthy based on a biphase analysis input.

When keeping the accuracy in mind the Convolutional Neural Network appears to have the highest recall and is one of the highest accuracies.

6.3 Mel-Frequency Cepstrum

The MFCC analysis doesn't seem to have a classifier that performs particularly well for the COVID-19 sick recall.

When further viewing the accuracy alongside the recall, the Support Vector Machine with non-linear kernel and scaled gamma performs the best.

6.4 Conclusion

In comparing and contrasting the different analysis methods, it is difficult to set an objective metric to compare their performance. In trying to evaluate which analysis method had the most information relevant to our classification task we can evaluate the highest average F1

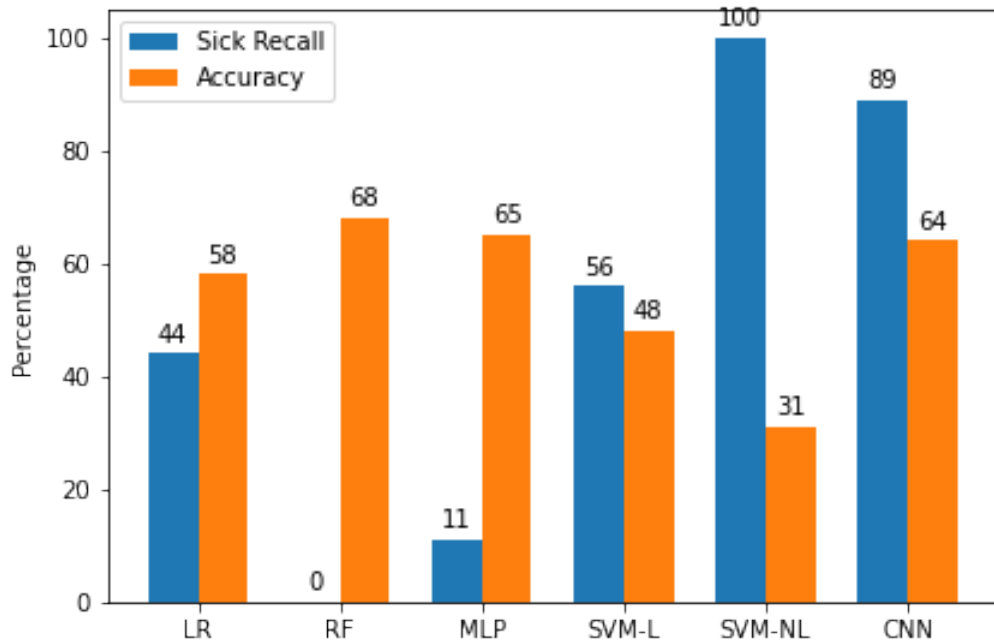


Figure 6.4. A plot of the recall score for various classification method for COVID-19 sick and the overall accuracy based on a biphasic analysis input.

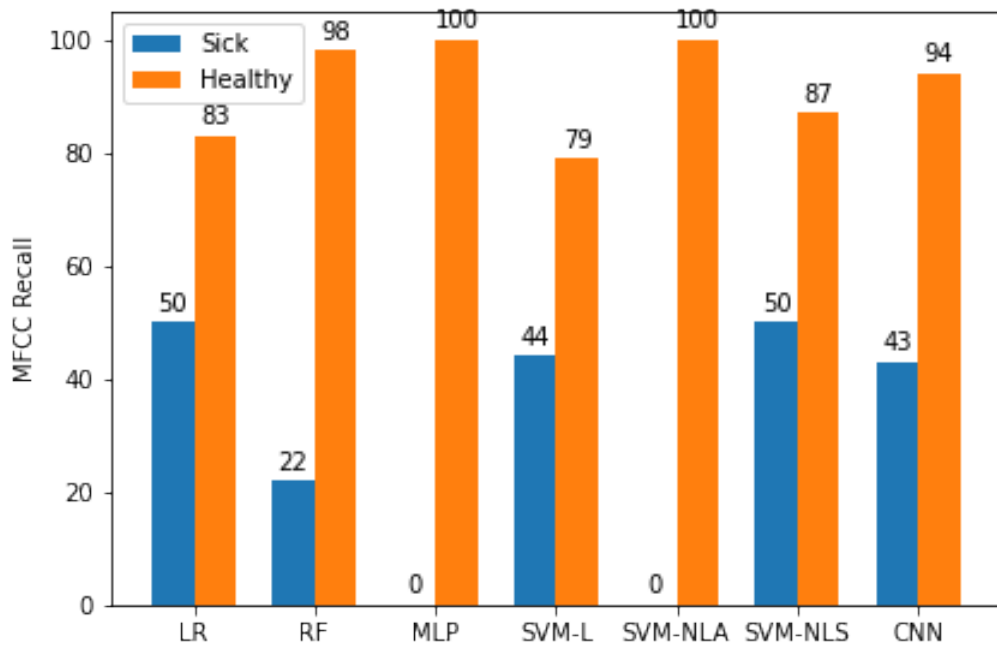


Figure 6.5. A plot of the recall score for various classification method for COVID-19 sick and healthy based on a Mel-Frequency Cepstrum analysis as input.

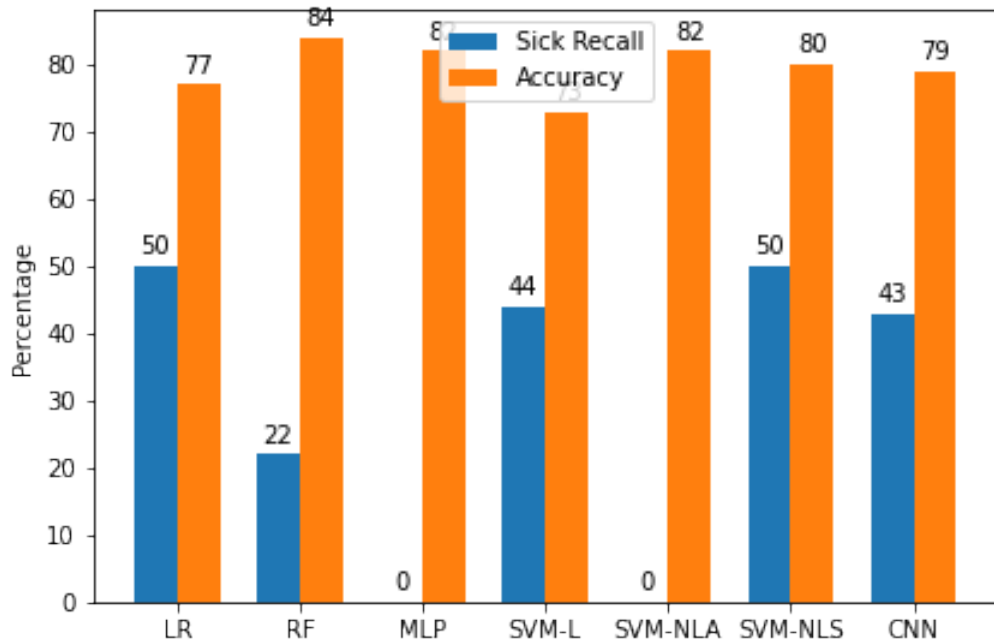


Figure 6.6. A plot of the recall score for various classification method for COVID-19 sick and the overall accuracy based on a Mel-Frequency Cepstrum analysis as input.

score of all the classification methods for the analysis methods. Doing so we get the following results for the highest F1-score:

Bicoherence analysis using Support Vector Machine: 58%

Biphase analysis using Logistic Regression: 54%

MFCC analysis using Convolutional Neural Network: 70%

It appears that the Mel-Frequency Cepstrum contains the most information and is the best sole analysis method to utilize for this classification task.

6.5 Future Work

This work has room for improvements in regards to the amount of data, balancing the datasets and combining the techniques and classifiers to come up with an ensemble method that performs better. A key area that would improve the performance and generalizability of the

classifiers is to increase the dataset to increase the training set. With more data we can also get more meticulous about our training set, narrowing the training to specific audio sample types (such as only coughs or only ahhs) and incorporating more granular data such as the phase of the disease (is the individual newly diagnosed or towards the end of the recovery process). Adjusting for the label imbalance, as there are many more healthy samples than COVID-19 sick samples, can further improve the performance and a method such as random up-sampling can be utilized.

Another area for further research is doing more granular research on the specific classifiers used, such as further optimizing the Support Vector Machine's kernel options, nu values and gamma values, and experimenting with different architectures for the Convolutional Neural Network.

In general, this study has revealed that all of the different analysis methods contain some amount of information towards our target classification task. An improved method can be created by creating an ensemble of the analysis methods (such as using all bicoherence, biphasic and mfcc simultaneously as the input to the classifier) and of the classifiers (such as defining an ensemble method that utilizes multiple classifiers and weights their respective results to come up with the final result).

Bibliography

- [1] Mahmood Al-khassaweneh and Ra'ed Bani Abdelrahman. A signal processing approach for the diagnosis of asthma from cough sounds. *Journal of Medical Engineering Technology*, 37(3):165–171, 2013.
- [2] Ehab A. AlBadawy, Siwei Lyu, and Hany Farid. Detecting ai-synthesized speech using bispectral analysis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.
- [3] T. Drugman, J. Urbain, N. Bauwens, R. Chessini, C. Valderrama, P. Lebecque, and T. Dutoit. Objective study of sensor relevance for automatic cough detection. *IEEE Journal of Biomedical and Health Informatics*, 17(3):699–707, 2013.
- [4] Justin Fackrell. *Bispectral analysis of speech signals*. PhD thesis, 06 1997.
- [5] Rubén Fraile, Nicolas Saenz-Lechon, Juan godino llorente, V Osma-Ruiz, and Corinne Fredouille. Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex. *Folia phoniatica et logopaedica : official organ of the International Association of Logopedics and Phoniatics (IALP)*, 61:146–52, 02 2009.
- [6] R. M. Warren M. Klopper K. Dheda P. D. van Helden G. H. R. Botha, G. Theron and T. R. Niesler. Detection of tuberculosis by automatic cough sound analysis. *Physiological Measurement*, 39(4):045005, 2018.
- [7] R. Goubran H. Chatzarrin, A. Arcelus and F. Knoefel. Feature extraction for the differentiation of dry and wet cough sounds. *2011 IEEE International Symposium on Medical Measurements and Applications*, 2011.
- [8] J. Korpas, J. Sadlonova, and M. Vrabec. Analysis of the cough sound: an overview. *Pulmonary pharmacology*, 9(5):261–268, 1996.
- [9] Cyril R Pernet and Pascal Belin. The role of pitch and timbre in voice gender categorization. *Frontiers in psychology*, 3, 2012.
- [10] A. Poreva, Y. Karplyuk, A. Makarenkova, and A. Makarenkov. Detection of copd's diagnostic signs based on polyspectral lung sounds analysis of respiratory phases. In *2015 IEEE 35th International Conference on Electronics and Nanotechnology (ELNANO)*, pages 351–355, 2015.

- [11] Lawrence R. Rabiner and Ronald W. Schafer. *Introduction to Digital Speech Processing*. Now Publishers Inc., Hanover, MA, USA, 2007.
- [12] L. Parvez S. Subburaj and T. G. Rajagopalan. Methods of recording and analysing cough sounds. *Pulmonary Pharmacology*, 9(5):269–279, 1996.
- [13] Ehsan Shams, Davood Karimi, and Zahra Moussavi. Bispectral analysis of tracheal breath sounds for obstructive sleep apnea. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2012:37—40, 2012.
- [14] Elaine Stathopoulos, Jessica Huber, and Joan Sussman. Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4-93 years of age. *Journal of speech, language, and hearing research : JSLHR*, 54:1011–21, 08 2011.
- [15] Amalia Setyati Udantha R. Abeyratne, Vinayak Swarnkar and Rina Triasih. Cough sound analysis can rapidly diagnose childhood pneumonia. *Annals of Biomedical Engineering*, 41(11):2448–2462, 2013.
- [16] Srinivasan Umesh, L. Cohen, and D. Nelson. Frequency warping and the mel scale. *IEEE Signal Processing Letters*, 9:104–107, 2002.
- [17] Anne B. Chang Yusuf A. Amrulloh Amalia Setyati Vinayak Swarnkar, Udantha R. Abeyratne and Rina Triasih. Automatic identification of wet and dry cough in pediatric patients with respiratory diseases. *Annals of Biomedical Engineering*, 41(5):1016–1028, 2013.