: N° d'ordre
: N° de série

**PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA**
**Ministry of Higher Education and Scientific Research**

**UNIVERSITY OF ECHAHID HAMMA LAKHDAR - EL OUED**

**FACULTY OF EXACT SCIENCES**
**Computer Science Department**

**Thesis**
**submitted in partial fulfilment of the requirements for the degree of**

# ACADEMIC MASTER

Field: **Mathematics and Computer Science**
Option: **Computer Science**
Specialty: **Distributed Systems and Artificial Intelligence**


Presented by :

- **Degaa Manar**
- **Bouzouaid Nour Elhouda**


## Title

# AN AUTOMATIC COUGH CLASSIFICATION AI TOOL FOR PNEUMONIA PATHOLOGY DETECTION BASED MACHINE LEARNING AND DEEP LEARNING MODELS

Defended on June 16 th 2022

**Examination Commitee :**

| | | | |
|---|---|---|---|
| M. | Gharbi Kaddour | MCA | Chair |
| M. | Khabache Mohibeddine | MAA | Examinator |
| M. | **Khouladi Nedjoua Elhouda** | MAA | Supervisor |

Academic year: 2021-2022

# *Dedication*

Thanks to Allah first-then the efforts of my loved ones who helped me a lot with their advice and efforts when we were going to offer this job.

Dedicate this work to

those who set me on the path of life, made me calm, and took care of me until I became old

(B.Soumia)

the owner of a fragrant biography and enlightened thought. for the man who deserves the first credit in attaining higher education (D.Lazher), may God prolong his life.

My brothers who left a great impact on many difficulties,

*(Laid and Firas Sader Eddine)*

I rely on them in every big and small. my only sister

*(Maouhebe)*

And to my sister from another mother

*(G.Sawsen Houria/H.Fatma)*

My colleague in search

*(B. Nour Elhouda)*

To all my best friends

*(Widad/Randa/Zaineb/Hania/Samer/Kawthar/Donia)*

*DEGAA MANAR*

# *Dedication*

Thanks to Allah first-then the efforts of my loved ones who helped me a lot with their advice and efforts when we were going to offer this job.

Dedicate this work to

those who set me on the path of life, made me calm, and took care of me until I became old

(B.Yamina)

the owner of a fragrant biography and enlightened thought. for the man who deserves the first credit in attaining higher education (B.Hocine), may God prolong his life.

My brothers who left a great impact on many difficulties,

*(Mohamed salah and Abd eerouf)*

I rely on them in every big and small. my sisters

*(Hada,Yasmina,Khaoula,Aicha)*

My colleague in search

*(D. Manar)*

To all my best friends

*BOUZOUAID NOUR ELHOUDA*

# Acknowledgements

In the Name of Allah, the Most Merciful, the Most Compassionate all praise be to Allah, the Lord of the worlds; and prayers and peace be upon Mohamed His servant and messenger. First and foremost, I must acknowledge my limitless thanks to Allah, the Ever-Magnificent; the Ever-Thankful, for His help and bless. I am totally sure that this work would have never become truth, without His guidance. I owe a deep debt of gratitude to our university for giving us an opportunity to complete this work. I am grateful to some people, who worked hard with me from the beginning till the completion of the present research particularly my supervisor Dr. Khouladi Nedjoua Elhouda

who has been always generous during all phases of the research, and I highly appreciate the efforts expended by Khammes.A and Badjouti.A

# Abstract

The coronavirus appeared in late 2019, which is represented by a large family of viruses that affect the respiratory system, where its symptoms vary from one infection to another: difficulty breathing,fever and exhaustion, the majority of patients suffer from dry cough. These viruses are easily transmitted from one person to another unlike other viruses, which led to paralysis of the world inits various sectors.

With the advancement of the world and the development of technology in recent years, the field ofautomated media and artificial intelligence has witnessed great progress, as it has contributed to solving many problems in various fields, the most important of which are health and generalmedicine. Etc., with the advent of this deadly epidemic, we have been using deep and machinelearning to reduce and prevent the spread of coronavirus.

That is why we did this research, which classifies cough using deep learning into a positive cough or a negative cough for coronavirus through a cough recorded on a smartphone. This study helps us reduce the spread of this deadly epidemic. After several tests and attempts to use models ANN, we reached a result of 97 %

**Keywords:** Features Extraction - deep learning - Covid-19

# résumé

Le coronavirus est apparu fin 2019, qui est représenté par une grande famille de virus qui infectent le système respiratoire, où ses symptômes varient d'une infection à l'autre : difficulté à respirer, fièvre et épuisement, la majorité des patients souffrent de toux sèche, et ces virus se transmettent facilement d'une personne à une autre contrairement à d'autres virus, ce qui a conduit à la paralysie du monde dans ses différents secteurs.

Avec l'avancement du monde et le développement de la technologie ces dernières années, le domaine des médias automatisés et de l'intelligence artificielle a connu de grands progrès, car il a contribué à résoudre de nombreux problèmes dans divers domaines, dont les plus importants sont la santé et la médecine générale ... Etc., avec l'avènement de cette épidémie mortelle, nous avons utilisé l'apprentissage profond et automatique pour réduire et prévenir la propagation du coronavirus.

C'est pourquoi nous avons fait cette recherche qui classe la toux à l'aide de l'apprentissage profond en toux positive ou en toux négative du coronavirus par le biais d'une toux enregistrée sur un smartphone, car cette étude nous aide à réduire la

propagation de cette épidémie

Après plusieurs tests et tentatives d'utilisation de modèles ANN, nous avons atteint un résultat de 97 %.

**Mots clés:** Caractéristiques d'extraction  Covid-19  l'apprentissage en profondeur

# ملخص

ظهر فيرس كورونا في اواخر عام ٢٠١٩ ، الذي يتمثل في عائلة كبيرة من الفيروسات تصيب الجهاز التنفسي ، حيث تختلف أعراضه من إصابة إلى أخرى : صعوبة في التنفس ، الحمى و الإرهاق ، أغلبية المرضى يعانون من السعال الجاف ، وتنتقل هذه الفيروسات بسهولة من شخص إلى آخر عكس الفيروسات الأخرى ، مما أدى إلى شلل للعالم في مختلف قطاعاته.

مع تقدم العالم و تطور التكنولوجيا في السنوات الأخيرة ،شهد مجال الأعلام الآلي و الذكاء الإصطناعي تقدما كبيرا ،حيث ساهم في حل العديد من المشاكل في مختلف المجالات أهما الصحة والطب العام ..إلخ،و مع ظهور هذا الوباء الفتاكو جعلنا نستخدم التعلم العميق و الآلي في الحد من إنتشار فيروس كورونا و الوقاية منه

. لهذا قمنا بهذا البحث الذي يقوم بتصنيف السعال بإستخدام التعلم العميق إلى سعال إجابي أو سعال سلبي لفيروس كورونا من خلال السعال المسجل على الهاتف الذكي ، حيث تساعدنا هذه الدراسة في الحد من أنتشار هذا الوباء القاتل . وبعد عدة إختبارات و محاولات بإستخدام موديلات ANN وصلنا إلى نتيجة بلغت 97%

**الكلمات المفتاحية** : التعلم العميق ـ كوفيد ١٩ ـ ميزات الاستخراج ـ ANN

# Contents

# List of Figures

# List of Tables

# General Introduction

The SARS-CoV-2 virus causes Coronavirus Illness (COVID-19), an infectious disease. Most people infected with the virus experience mild to moderate respiratory symptoms, which go away without the need for any specific treatment. Some people who become infected, on the other hand, experience severe symptoms and require medical intervention. Those who have underlying disorders such cardiovascular disease, diabetes, chronic respiratory disease, cancer, and other illnesses are more prone to develop severe symptoms of the disease. COVID-19, on the other hand, can cause serious sickness and death in people of any age.

The majority of countries have prepared their healthcare systems to deal with the pandemic following the swift global trigger. International and national planning initiatives have encountered social, institutional, and educational, political, environmental, industrial, and technical obstacles while receiving a lot of attention . A huge disparity in research articles and drugs has prompted a recent initiative to find an adequate diagnosis and develop a new strategy to slow the virus's spread.

In our current era, technology and artificial intelligence have invaded all areas of life, from health, education, commerce, etc., so we suggested, through this research, the employment of machine learning in detecting infection from the Corona virus using the sound of coughing, in order to limit the spread of this virus. It also reduces the material costs, especially with the weak material income and the global financial crisis.

# Chapter 1

# Cought detction and classifiction:an overview

## 1.1 Introduction:

This Chapter will commence with an overview on initiatives collected data needed for this study and the related work of previous research papers working in this topic, then we will introduce the various models proposed by scholars for cough analysis using different datasets, this study will provide several motivations from exposing background . We will carry out this chapter by revealing main machine learning classification pipeline used for cough detection. Finally, we will determine score metrics for assessing these models.

## 1.2 Overview of Cough Detection:

coronavirus Disease of 2019 (COVID-19), a novel pathogen of the severe acuter espiratory syndrome coronavirus 2 (SARS-Cov-2), first appeared in late November 2019 and has since spread around the world, causing a global epidemic problem. [1]

coronavirus Disease, a novel pathogen of the severe acute respiratory syndrome coronavirus from the SARS-Cov-2 pathogen that was discovered in 2019. Since its inception in the late 1990s, it has become a global epidemic problem Since November 2019, it has spread worldwide. [1]

Since the pandemic began in 2019, there have been nearly 150 million confirmed cases and over 3 million deaths, according to the World Health Organization39;s (WHO) April 2021 report [2]. Furthermore, the United States of America (US) has the highest number of With over 32.5 million and 500,000 cases and deaths, respectively.

According to a report published by the World Health Organization (WHO) [2] in April 2021, the pandemic has claimed the lives of approximately 150 million people and caused over 3 million deaths. In the United States (US), there have been over 32.5 million and 500,000 deaths, respectively the highest total number.

These huge numbers have put a strain on many healthcare services, especially given the virus's ability to develop more genomic variants and spread more quickly among people.

Because of the virus's ability to develop more genetic variations and spread faster among people, a large number of patients has put a strain on many healthcare services.

After an explosion of cases due to a new variant of COVID-19, India, which is one of the world's largest vaccine suppliers, is now severely affected by the pandemic. It now has over 17.5 million confirmed cases, placing it second only to the United States worst-affected countries [2] [3]

India, one of the world's largest vaccine suppliers, is now seriously affected by the pandemic, following an explosion of cases caused by a new variant of COVID-19. It now has over 17.5 million confirmed cases, making it the second most affected country After the United States, the "worst-affected country" in the world [2] [3].

COVID-19 patients can be asymptomatic or develop pneumonia, which can lead to death in severe cases. In most cases, the virus is incubating for 1 to 14 days before symptoms of infection appear [4]. Patients who have COVID-19 Cough, shortness of breath, fever, and other common signs and symptoms have been observed ARDS [5] [6], fatigue, and other acute respiratory distress syndromes.

Most infected people experience mild to moderate viral symptoms, but they eventually recover. Patients with severe symptoms such as severe pneumonia, on the other hand, are typically persons over 60 years old who have diabetes, cardiovascular disease (CVD), hypertension, or cancer [4] [5]. COVID-19 is most often diagnosed early, which helps to prevent it from spreading and progressing to severe infection stages. This is normally accomplished by following early patient isolation and contact tracing procedures. Furthermore, prompt medicine and effective treatment lessens symptoms and lowers the

pandemic's fatality rate [7]

The reverse-transcription polymerase chain reaction (RT-PCR) test is the current gold standard for diagnosing COVID-19 [8] [9]. It is the most widely used method for successfully confirming the presence of this viral infection around the world. Furthermore, ribonucleic acid (RNA) analysis in virus-infected patients provides additional information about the infection, but it takes longer to diagnose and is not as accurate as other diagnostic techniques [10]. Another excellent diagnostic method (sensitivity 90%) that often offers extra information regarding the severity and evolution of COVID-19 in the lungs is the integration of computed tomography (CT) screening (X-ray radiations) [11] [12]. CT imaging is not advised for individuals who are asymptomatic or have minimal symptoms in the early stages of an illness. Due to abnormalities in pulmonary tissues and their accompanying activities, it gives important information about the lungs in patients with mild to severe stages [13]

Several research have recently used newly developed artificial intelligence (AI) methods to identify and categorize COVID-19 in CT and X-ray images [14]. In various research (using CT scans as inputs), machine and deep learning algorithms were used to achieve a discriminating accuracy of over 95% between healthy and diseased subjects [15] [16] [17] [18] [19] [20]. The ability of trained models such as support vector machines (SVM) and convolutional neural networks (CNN) to detect COVID-19 in CT images with minimal pre-processing steps is the primary contribution of these

investigations. Furthermore, some research have used deep learning with additional feature fusion approaches and entropy-controlled optimization [21], rank-based average pooling [22], and entropy-controlled optimization . To identify COVID-19 in CT scans, researchers used the pseudo-Zernike moment (PZM) [23] and the internet-of-things (IoT) [24]. Furthermore, substantial research employing X-ray imaging and machine learning for COVID-19 evaluation has been conducted [25] [26] [27]. The majority of current state-of-the-art techniques use two-dimensional (2D) X-ray pictures of the lungs to train neural networks on extracting characteristics and thereby detecting virally infected patients.

Despite the high levels of accuracy reached in the majority of investigations, CT and X-ray imaging involve ionizing radiations, making them unsuitable for routine testing. Furthermore, due to their high prices and increased maintenance needs, these imaging modalities may not be available in all public healthcare systems, particularly in countries hit hard by the epidemic. Recently, researchers have used an ultrasound-based imaging technique to screen lungs for COVID-19 [28] and have reached excellent levels of performance (accuracy ¿89 percent ). When it comes to merging these methodologies with machine learning, researchers' ultimate objective is to identify potential alternatives that are simple, fast, and cost-effective combining these methods with machine learning.

Coughing and breathing noises, which are biological respiratory cues with a direct relationship to the lungs, might be another potential approach to detect the presence of a

viral infection [29]. Respiratory auscultation is a non-invasive and safe method of diagnosing the respiratory system and its related organs. Clinicians often use an electronic stethoscope to hear and record the sound of air flowing within and outside the lungs when breathing or coughing. This allowed for the detection and identification of any pulmonary abnormalities [30] [31] [32]. Lung sounds might provide important information about the viral infection due to the ease with which they can be recorded, and hence could provide an early warning to the patient before proceeding with more comprehensive pharmaceutical procedures. Furthermore, due to its capacity to generalize across a large collection of data with reduced computing cost, integrating basic respiratory signals with AI algorithms might be a key to improving the sensitivity of detection for positive instances [33]

## 1.3 Background and Motivation:

Many research have looked at the information conveyed by respiratory sounds in COVID-19-positive individuals [34] [35] [36]. Furthermore, vocal patterns recovered from COVID-19 patients' speech recordings have been discovered to have indicators for the presence of viral infection [37]. Furthermore, a telemedicine technique was investigated to examine indications of COVID-19 infection-related alterations in respiratory sounds [38]. Recently, AI was used in two studies: one to distinguish COVID-19 in cough signals [39] and another to use voice recordings to assess the severity of patients' sickness, sleep quality, weariness, and worry [40]. Despite the excellent levels of performance attained in the

aforementioned AI-based studies, more research into the capacity of respiratory sounds to carry important COVID-19 information is still needed, especially when incorporated inside the framework of advanced AI-based algorithms. Furthermore, given the surge in confirmed positive COVID-19 cases around the world, it is critical to ensure that a system capable of recognizing the disease in signals recorded using portable devices such as computers or smartphones, rather than traditional clinic-based electronic stethoscopes, is available.

Lungs, larynx, and articulators make up the majority of the human voice generating system. The lungs are thought to be the power source for the voice generating mechanism. Respiratory disorders impair the lungs' ability to function correctly, affecting the human voice generation mechanism. Obstructive and restrictive respiratory disorders are the two basic types of respiratory diseases [41]. Obstructive lung disorders restrict the pulmonary airways, limiting a patient's capacity to completely remove air from the lungs. As a result, a substantial volume of air stays in the lungs at all times. People with restrictive lung disorders, on the other hand, are unable to fully expand their lungs in order to fill them with air. As a result, the lungs are unable to fully expand with air. As a result, the lungs do not fully inflate. Some people have both obstructive and restrictive respiratory illnesses. Cough is a frequent sign of lung illnesses such as obstructive, restrictive, and mixed. As a result, cough sounds are thought to be helpful in diagnosing lung disorders caused by respiratory problems [42]. COVID-19 is also classified as a respiratory infection. COVID-19 may cause the lungs to fill with fluid and become irritated, similar to other respiratory

disorders. As a result, individuals may experience respiratory difficulties and require immediate medical treatment. COVID-19, if left untreated, can escalate to acute respiratory distress syndrome (ARDS), a kind of lung failure [43]. Coughing is a frequent sign of any respiratory disease, including COVID-19, however current research suggests that the COVID-19 cough is dry, persistent, and hoarse in the early stages of coronavirus infection. As a result, COVID-19 patients' cough sound samples differ from those of patients with other respiratory illnesses. As demonstrated in Fig. 1, human cough samples have three phases: explosive phase, intermediate phase, and vocal phase. The glottal airflow variation in the vocal cord is represented by these phases, which vary depending on the pathological circumstances of the patients. To study the differences between the cough sound samples of a COVID-negative (i.e., healthy/control) individual and a COVID-positive patient, two segmented cough sound samples are randomly picked from the Virufy database [44]. Figure 2 shows cough sound samples from a healthy person and a COVID-positive patient. This illustration shows how the healthy sample resembles the usual human cough sound signal seen in Figure 1. However, the COVID-19 patient's cough sound sample differs greatly from a typical human cough sound sample. The intermediate and vocal phases, for example, are longer in the COVID-positive patient than in the healthy person. Furthermore, the COVID-positive patient's signal amplitude during the voiced phase is greater than the healthy subject's. As seen in Fig. 2, the amplitudes of the explosive phase change between these two cough sound samples. Because of the variations

indicated above, the cough sound can be utilized to distinguish the COVID-positive patient from the healthy person. Figure 3 shows the power spectral densities (PSDs) of these two samples. The healthy cough sound has major frequencies of progressively decreasing magnitudes, as seen in the diagram. The COVID-positive cough sound samples, on the other hand, do not have highly identifiable frequencies.

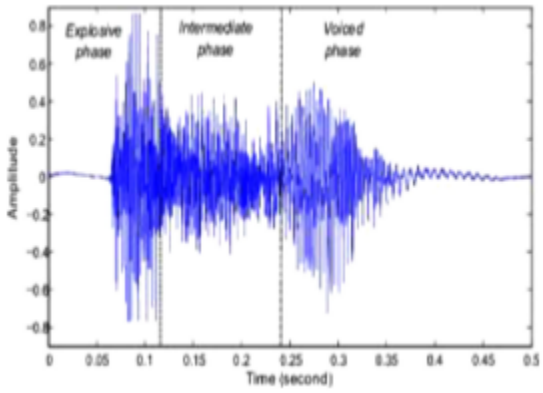



**Figure 1.1:** A typical cough sound signal

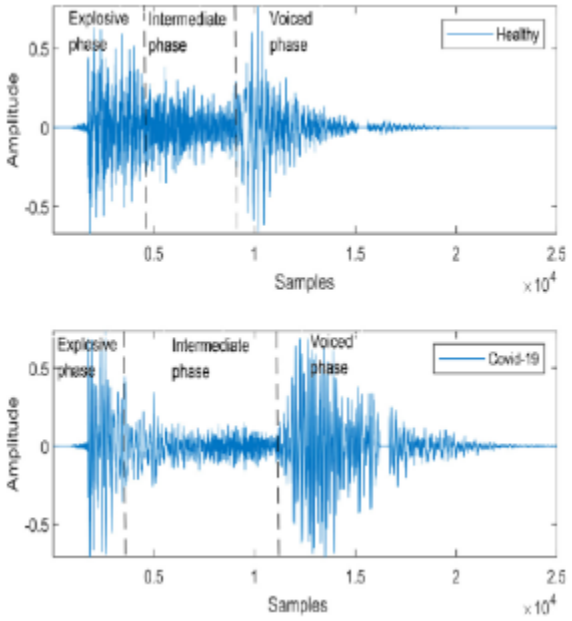



**Figure 1.2:** Comparison of the cough sounds for a healthy subject and a COVID-19patient collected from the Virufy database

Motivated by the foregoing, this work proposes a full deep learning technique for detecting COVID-19 utilizing just breathing noises captured through a smartphone device's microphone (Fig 1). The recommended method is a quick, no-hassle solution.COVID-19 pre-screening technology that is low-cost and readily deployed, especially for developing nations Due to the pandemic's widespread spread, the city is in total lockdown. Despite the present gold price,The standard, RT-PCR, has a high success rate in identifying viral infection and has a variety of applications.The high costs of equipment and chemical agents are among the restrictions.Expert nurses and physicians are required for diagnosis, social separation is violated, and the length of time it takes to

get findings from testing (2-3 days). As a result, the construction of a deep learning model overcomes the majority of these restrictions, allowing for improved resurrection in the healthcare and economic sectors in a number of nations.

## 1.4 Related work:

Cough classification systems try to match cough sounds to illnesses, whereas cough detection systems try to match audio signals to cough or non-cough occurrences. Cough detection systems, like cough classification systems, rely on the analysis of cough audio signals, which is also the backbone of cough detection systems. As a result, a review of cough detection systems was on the horizon. Cough detection systems that have been studied in the literature may be divided into two kinds.

### 1.4.1 Type of detection :

Cough detection is a hot topic of research, with various academics proposing methods for detecting cough noises in audio recordings. These techniques may be classified into three groups:

**1)identification and segmentation of coughs automatically (without classification):**

Martinek et al. [45] collected numerous temporal, frequency, and entropy variables and

utilized a decision tree to distinguish between voluntary cough noises and speech for automated cough segmentation. They used data from 20 people, each of whom coughed 46 times, and reported median sensitivity and specificity values of 100% and 95%, respectively. However, because the participants must cough at the start of each recording to get distinct cough signal patterns, their approach is subject–dependent. Barry et al. [46] created the Hull Automatic Cough Counter by combining linear predictive coding (LPC) coefficients with a probabilistic neural network (PNN) classifier (HACC) However, because the participants must cough at the start of each recording to get distinct cough signal patterns, their approach is subject–dependent. Barry et al. [46] created the Hull Automatic Cough Counter by combining linear predictive coding (LPC) coefficients with a probabilistic neural network (PNN) classifier (HACC). With a sensitivity of 80% and specificity of 96%, this reliably differentiated between cough and non–cough episodes in 33 patients. Tracey et al. [47] created a cough detection system to track patient recovery from TB. They retrieved MFCCs from 10 test individuals' audio signals. With a mix of artificial neural network (ANN) and support vector machine (SVM) classifiers, they were able to detect coughs with an overall sensitivity of 81 percent. The total number of coughs in the dataset was not given in any of these techniques. For cough identification, Swarnkar et al. [48] combined MFCC data with additional spectral parameters such as formant frequencies, kurtosis, and Bscore. For a test dataset of 342 coughs from three people only, they were put into a neural network, which had a sensitivity of 93 percent and a specificity

of 94 percent. Utilizing approximately 1400 cough sounds from 14 participants, Amrulloh et al. [49] developed a cough detector for pediatric wards using ANN classification and a non–contact recording technique, reaching sensitivity and specificity values of 93 percent and 98 percent, respectively. Matos et al. [50] identified thirteen MFCCs and used a Hidden Markov Model to classify them (HMM). Their test dataset included 2155 coughs from 9 participants, and their approach had a sensitivity of 82 percent for cough detection. The overall number of false detections was not given, however the average number of false positives per hour was 7, with a wide range of patients. Liu et al. [51] employed Gammatone Cepstral Coefficient (GMCC) characteristics in conjunction with SVM classification to classify 903 coughs from four participants, with sensitivity and specificity of 91% and 95%, respectively. Lucio et al. [52] employed k-Nearest Neighbor (kNN) classification to obtain 79 MFCC and Fast Fourier Transform (FFT) coefficients. Their program classified 411 cough sounds with 87 percent sensitivity and 84 percent specificity using a sample collected from 50 people. Larson et al. [53] reported a cough detection approach that collected data utilizing the built-in microphone of a mobile phone. Random forest classification with a maximum of 500 decision trees was utilized in their method, which achieved 92 percent sensitivity on over 2500 cough sounds from 17 individuals. All of these approaches try to extract cough segments from audio recordings without being able to categorize them into distinct cough kinds.

**2) Coughs that have already been identified are automatically classified:**

Cough sound classification aids in determining the underlying cause of coughs so that patients can receive the appropriate therapy. Several methods for automatic cough classification have been described to detect distinct cough kinds, however they all need manual segmentation of cough sounds prior to automatic classification. Chatrzarrin et al. [54] investigated the different stages of dry and wet coughs and discovered that the second phase of dry coughs contains less energy than the first phase. They also discovered that most of the signal strength is concentrated between 0-750 Hz for wet coughs and 1500-2250 Hz for dry coughs during this period. They recognized 14 wet and dry coughs with 100 percent accuracy using a simple thresholding procedure. Swarnkar et al. [55] employed a classifier based on the Logistic Regression Model (LRM) to distinguish between dry and wet coughs in pediatric patients with various respiratory diseases. B–score, nongaussianity, formant frequencies, kurtosis, zero crossing rate, and MFCCs were among the characteristics they utilized. They obtained sensitivities of 84 percent and 76 percent for detecting wet and dry coughs, respectively, for a test database of 117 coughs from 18 participants. By analyzing cough sounds and crackles, Kosasih et al. [56] established an algorithm for automated detection of juvenile pneumonia. To distinguish between pneumonia and nonpneumonia cough sounds, they employed MFCCs, nongaussianity index, and wavelet characteristics with an LRM classifier. For a total of 375 cough samples from 25 individuals, their approach had sensitivity and specificity of 81% and 50%, respectively. Parker et al. [57] investigated the performance of three different classifiers

for the categorization of pertuss coughs. They created a dataset of 16 non-pertussis cough signals and 31 pertussis cough signals using audio recordings of pertussis cough noises found on the internet. The cough noises were manually separated and split into three portions from this data, with 13 MFCC characteristics and the energy level extracted. Following that, an ANN, a random forest classifier, and a kNN classifier were used to classify the characteristics. The false positive error for each of these classifiers was 7%, 12%, and 25%, respectively, whereas the false negative error was 8%, 0%, and 0%.

**3) a disease diagnosis based on the sound and kind of cough:**

While there exist algorithms for cough recognition and particular cough classification, none of them do entirely automated cough detection and classification, according to the review above. Furthermore, just one algorithm for pertussis cough classification has been disclosed, and it depends on manual segmentation of cough sounds before categorization. Furthermore, most high-performing cough detectors employ complicated classifiers, rendering them inappropriate for usage on devices with limited resources. Recently, researchers have proposed that cough sounds be used to detect COVID-19 early. Cough is a sign of 30 different diseases [58] [59], thus there are still some obstacles to overcome. As a result, distinguishing the cough sound of COVID-19 patients from that of other patients is quite difficult. The authors looked at three illnesses in [58]: bronchitis, pertussis, and COVID-19. A total of 247 normal cough samples and 296 pathological cough samples were

examined. The authors implemented a binary classifier and a multiclass classifier using a variety of models. The binary classifier distinguishes between pathological and normal cough sounds, whereas the multiclass classifier classifies the diseases into one of three kinds. The authors of a similar study [60] looked at bronchitis, bronchiolitis, and pertussis. To distinguish against these disorders, they employed a CNN.

### 1.4.2 Used Methods for detection:

is important to understand the characteristics of the cough sound before making any attempt to detect it. The cough sequence begins with mechanical or chemical stimuli and ends when unwanted substances are removed from the airways [61]. The acoustic features of the cough sound depend on the airflow velocity as well as the dimensions of the vocal tract and airways [61]. This makes it possible to detect or classify the cough sound based on the acoustic features because these features depend on the cause of the cough.

In order to detect cough from sound signals, special methods have been used by experts and researchers in this field, which are divided into:

### 1.4.3 Models based Deep Learning:

| Authors | Cough data collection | Feature extraction | Models used | Total samples |
|---|---|---|---|---|
| J Liu et al. 1 | Collar fitted mic | MFCC | Neural network pretrained using Deep belief networks | 3873 |
| J Amoh et al. 21 | Wearable acoustic sensor | Spectrogram obtained from Short Time Fourier Transform (STFT) | Two layered CNN model | 627 |
| M Soliński et al. 19 | NHANES database of spirometry curves | Feature extraction from signal processing | Two layered ANN model | 10874 |
| R Xaviero et al. 34 | Multiple repositories 5 | Feature extraction from STFT | Logistic regression model | 1980 |
| A Windmon et al. 22 | Smart phone mic | 13 features extracted using STFT, MFCC | Random Forest model | 165 |

| | | | | |
|---|---|---|---|---|
| J Amoh et al. 6 | Wearable acoustic sensor | Cough audio (RNN), Spectr-temporal image of cough audio using STFT | RNN and 2dCNN model | 627 |
| L Perna et al. 41 | Microphone | 12 MFCC | SVM, KNN, XGBOOST model | 31754 |
| H Hee et al. 23 | Microphone | CQCC and MFCC | GMM-UBM model | 2332 |
| P Kadambi et al. 25 | Microphone | 13 MFCC | ANN model | 5670 |
| E Larson et al. 31 | Smart phone mic | FFT, PCA | RF model | * |
| J Alvarez et al. 42 | Smart phone mic | Spectral features | SVM model | * |
| K Nguyen et al. 15 | Smart phone mic, accelerometer | MFCC, Chroma feature analysis, zero crossing rate | SVM, ANN, RF | 800 |

| | | | | |
|---|---|---|---|---|
| R Sharan et al. 7 | Smart phone mic | MFCC | SVM | 3262 |
| S Matos et al. 43,44 | Microphone | MFCC, linear predictive coding, filter banks output | HMM | 4268 |

**Table 1.1:** Summary of State of the Art techniques in literature

### 1.4.4 Datasets:

**1)Coswara:**

In the Coswara dataset, there were 1171 subjects with "deep cough" records available, of which 92 were positive for COVID-19 and 1079 were healthy. This corresponds to a total of 1.05 hours of cough recordings (after preprocessing) used for the experiment.

**2)Sarcos:**

The Sarcos dataset contains data from a total of 44 subjects, 18 of whom are COVID-19 positive and 26 who are not. This amounts to a total of 2.45 min of cough audio recordings (after pre-processing) that has been used for experimentation. COVID-19 positive coughs are 15%–20% shorter than non-COVID coughs. [62]

| Dataset | Label | Subjects | Total audio | Average per subject | Standard deviation |
|---|---|---|---|---|---|
| Coswara | COVID-19 Positive | 92 | 4.24 min | 2.77 s | 1.62 s |
| | Healthy | 1079 | 0.98 h | 3.26 s | 1.66 s |
| | Total | 1171 | 1.05 h | 3.22 s | 1.67 s |
| Sarcos | COVID-19 Positive | 18 | 0.87 min | 2.91 s | 2.23s |
| | COVID-19 Negative | 26 | 1.57 min | 3.63 s | 2.75s |
| | Total | 44 | 2.45 min | 3.34 s | 2.53s |
| ComParE | COVID-19 Positive | 119 | 13.43 mins | 6.77s | 2.11sec |
| | Healthy | 398 | 40.89 mins | 6.16 s | 2.26sec |
| | Total | 517 | 54.32 mins | 6.31s | 2.24sec |

**Table 1.2:** Taypes of Dataset

## 1.5 Metric for Assessment :

As shown in the table, this matrix is called the confusion matrix and is used to evaluate the

performance of a classification model where true or false values have been applied to a set of test data (Sokolova and Lapalme 2009). The four basic parameters are true positive (TP), true negative (TN), false positive (FP) and false negative (FN). TP consists of multiple anomalies and has been identified as the correct scenario. TN is the regular number of instances of error measurements. FP is a collection of regular instances classified as Anomaly Scenario FP. FN is a list of anomalies observed as common scenarios. begintable[H]

| | Predicated negative | Predicate positive |
|---|---|---|
| Actual negative | TP | FN |
| Actual positive | FP | TN |

**Table 1.3:** confusion matrix.

After calculating the parameters in the confusion matrix, the evaluation metrics can be calculated such as accuracy, precision, recall, and F1-Score as follows:

**1) Accuracy:**

It is the most important metric for efciency. As shown in Eq. (1), it is simply the submission of true positives and true negatives divided by the total values of confusion matrix components. The most reliable model is the best, but it is important to ensure that there are symmetrical datasets with almost equal false positive values and false adverse values Therefore, certain parameters must be determined to determine the quality of our

system.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1.1}$$

**2) Precision:**

It as shown in Eq. (2), it is the relationship between true positive predicted values and full positive predicted values.

$$Precision = \frac{TP}{TP + FP} \tag{1.2}$$

**3) Recall:**

As shown in Eq. (3), recall is the ratio between predicted true positive values and the submission of predicted true positive values and predicted false negative values.

$$Recall = \frac{TP}{TP + FN} \tag{1.3}$$

**4) F1-score:**

It is an overall measure of a model's accuracy that combines precision and recall in that weird way that addition and multiplication just mix two ingredients to make a separate dish altogether. As shown in Eq. (4), the F1-score is twice the ratio between the multiplication

and precision and recall submission.

$$F1 - score = 2 \times \frac{Precision\ddot{O}Recall}{Precision + Recall} \tag{1.4}$$

**5) sensitivity:**

is the rate of correctly predicted positive samples in the positive class samples.

$$sensitivity = \frac{TP}{TP + FN} \tag{1.5}$$

**6) specificity:**

is the ratio of accurately predicted negative class samples to all negative class samples

$$specificity = \frac{TN}{TN + FN} \tag{1.6}$$

## 1.6 Conclusion:

This chapter highlights the main headings extracting from deep review of previous research papers, which have exposed many details as type of cough detection, the proposed methods using different AI tools focalizing on deep learning techniques, the different type of datasets used for training and classifying, finally, the performance metrics that assess the efficiency of designed techniques.

# Chapter 2

# Background and Feature Engineering

## 2.1 Introduction:

This chapter will start with an overview of feature engineering for relevant work from previous papers working on this topic, then will look at extracted features that have been applied to different voices using different datasets, and will also focus on deep features. Finally, we will look at some of the previous work that relied on extracted features in classifying coughs.

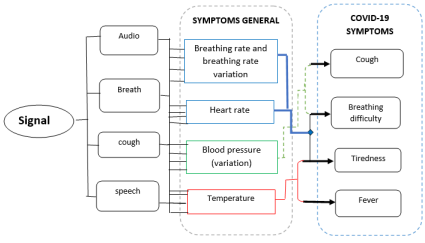



**Figure 2.1:** Caption

## 2.2 Features Engineering:

Feature engineering is a machine learning approach for creating new variables that aren't in the training set using data. It can generate new features for both supervised and unsupervised learning, with the objective of making data transformations simpler and faster while simultaneously improving model accuracy. When dealing with machine learning models, feature engineering is essential. A bad feature, regardless of the data or design, will have a direct influence on your model. [63]

### 2.2.1 Features Extraction:

The technique of extracting features from a data collection in order to uncover meaningful information is known as feature extraction. This reduces the amount of data into manageable numbers for algorithms to handle without altering the original relationships or key information. [63]

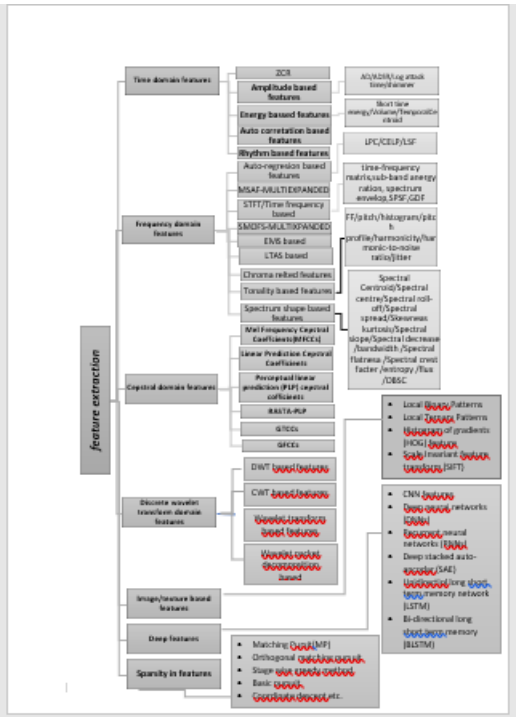



**Figure 2.2:** Taxonomy Audio Features

[]

### 2-2-1-1 Time domain features:

1. Zero crosing rate :

The rate of change of sign of the signal during an audio frame is known as the zero-crossing rate (ZCR). It is defined as the number of times a signal changes its sign from positive to negative and vice versa, divided by the frame duration. ZCR stands for the number of times a signal crosses the zero level in a one-second span. Figure 1 depicts the ZCR

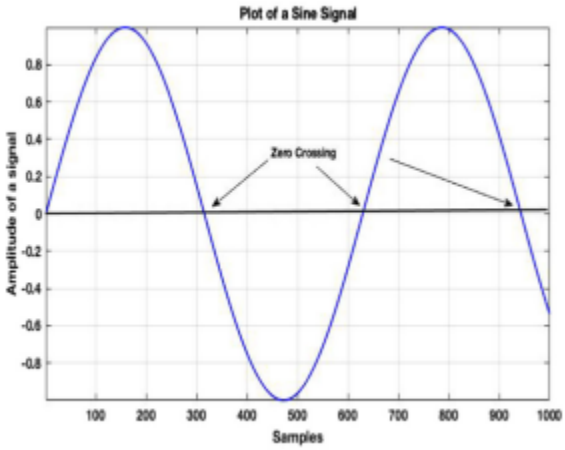
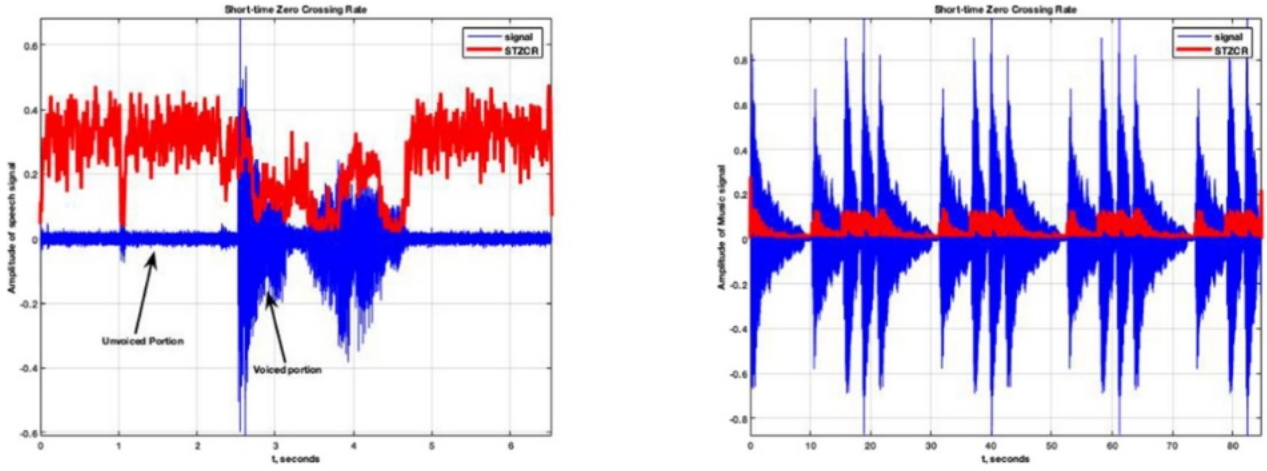



**Figure 2.3:** Zero-Crossing in signal.

ZCR is a fast approach to assess whether a speech frame is voiced, unvoiced, or quiet by detecting voice activity. In comparison to the vocal component of the speech, the ZCR is greater for the unvoiced bits. The voice signal and associated ZCR are shown in Figure 2(a). The ZCR for unvoiced portions is clearly much higher than for voiced segments. Of course, in ideal conditions, the ZCR for the silent section of a clean utterance must be zero.

ZCR is also a technique for estimating the fundamental frequency (FF) of speech [64]. ZCR is equal to the signal frequency multiplied by 2. Therefore, we can claim that ZCR provides indirect information about the frequency of the signal. Therefore, this property can be used to create discriminators and classifiers [65]. The ZCR of the music clip is shown in Fig. 2(b). Notably, it is larger than the ZCR of the speech signal [66]



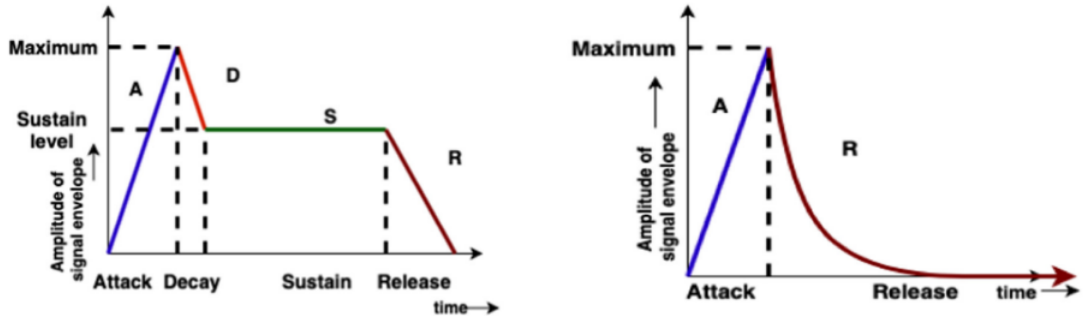

(a) ZCR for Speech Signal

(b) ZCR for Music Signal

**Figure 2.4:** Comparison of ZCR for Speech and Music.

The MATLAB pseudo code for calculating ZCR is given below: [] 2. ADSR envelope detection:

Because the decay and sustain parts of speech and ambient sounds are not clearly apparent, the ADSR envelope function is not attainable for most real-time sounds (only present in music sounds). The AR envelope is a modified envelope based on attack and rest that is used to cope with such an issue. The decay element is absent, and the sustain and release parts have been combined. The ADSR and AR envelopes are utilized in musical instrument timbre analysis [67]. The ADSR and AR envelops for a signal are shown in Figure 3. Algorithm 2 provides the MATLAB pseudo code for the ADSR envelope.




(a) ADSR envelope

(b) AR envelope

**Figure 2.5:** Amplitude based envelope feature.

3. Log attack time: It is the logarithmic (base 10) time interval between the start of the time interval and the end of the time interval. It's been utilized to identify musical onsets [68] and recognize ambient sounds [69] [70] 4. Short time energy calculation: The sound signals are inherently non-stationary. As previously stated, utilizing the framing/windowing approach, non-stationary signals may be turned into tiny sections of quasi-stationary signals. Because the signal's energy is varied throughout, it's impossible to forecast a value. The short time energy, which is the energy from a frame, is estimated for this purpose. Average energy per frame is defined as STE [66]. Unvoiced portions have a low STE, whereas voiced segments have a high STE. Voiced-unvoiced segment detection [71], music onset detection [68], ambient sound detection [72], vowel recognition and analysis [73], and audio-based surveillance systems [74] are all applications of STE .Algorithm 4 contains the pseudo code for computing STE. 5. Volume or loudness of audio signal: One of the most promising characteristics of the human auditory system is the

loudness or loudness of a sound. Loudness is mathematically defined as the root mean square (RMS) of the signal amplitude within a frame. Speech/music discrimination [75], speech segmentation, and acoustic scene classification [76] are among its applications. The pseudocode for estimating the loudness of the audio stream is explained in Algorithm 5. 6. Temporal centroid: The time averaged over the energy envelope is the time center of gravity. Both recognition of ambient noise [72] and classification of acoustic scenes [76] use temporal emphasis. The following is the algorithm of TC:

**2-2-1-2 Frequency domain features:**

1. Extraction of LPC coefficients LPC removes signal redundancy and attempts to predict future values by linearly combining previously known coefficients. LPC is an all-pole filter that uses a linear prediction model to describe the spectral envelope of digital speech in compressed form. LPC coefficients are used for audio segmentation and restoration. The built-in function ”lpc” in MATLAB has the parameters audio signal x and the order of the linear predictor. 2. Peak Frequency: The frequency of maximum power is known as peak frequency. It provides an approximation of the signal's most prominent frequency and aids in the calculation of the signal's fundamental frequency. Peak frequency is an important metric to consider for categorizing music and speech, as well as gender classification. A single-sided FFT spectrum's peak frequency is seen in Fig. 04.

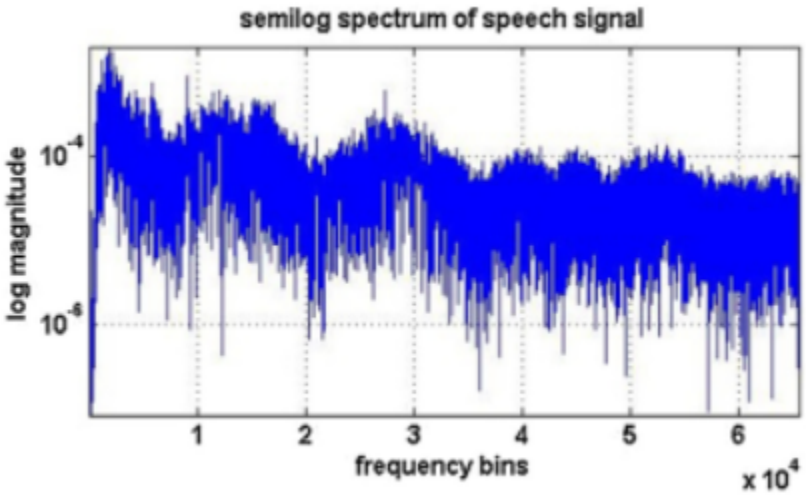



**Figure 2.6:** Peak Frequency of a Signal from single sided spectrum.

3. MSAF-MULTIEXPANDED acoustic features: The Method for Choosing Frequency Amplitudes Hand-crafted multi-expanded (MSAF MULTIEXPANDED) filter-based features are used to identify defects in electric motors used for drilling or grinding [77] or to detect defective commutator motors [78]. The discrepancies between FFT spectra are extracted using this auditory characteristic. The following is the generalized approach for extracting MSAF-MULTIEXPANDED acoustic features: 4. SMOFS-MULTICRAFTED acoustic feature: Shortened method of frequencies selection MULTIEXPANDED (SMOFS-MULTIEXPANDED), which is related to MSAF-MULTIEXPANDED, is another handcrafted feature that is used in industry applications to diagnosis faults in motors [79]. SMOFS-MULTIEXPANDED features are extracted in a different way from MSAF-MULTIEXPANDED features, as seen in the technique below.

5. Envelope modulation energy in desired frequency band:

A signal's time-frequency transform is a method of examining it with time on one axis and frequency on the other. Using a time-frequency distribution, a time-frequency analysis might be obtained (TFD). TFD produces a time-frequency representation (TFR) that is commonly referred to as the time-frequency domain. The time-domain displays changes in signal amplitude over time, whereas the magnitude of the frequency content in the frequency domain only provides frequency information and not time. A time and frequency resolution (TFR) overcomes this gap. The most frequent method of obtaining a TFR is through the STFT. TFR characteristics are also useful for analyzing nonstationary components of a signal, such as trends, discontinuities, and patterns, which are typically overlooked by time or frequency analysis. TFR features are also useful for analyzing nonstationary elements of a signal, such as trends, discontinuities, and patterns, which are often overlooked by time or frequency domain features [80] 6. Fundamental frequency FO : The fundamental frequency of the universe The local normalized spectro-temporal auto-correlation function's initial peak is F0. Fundamental frequency is the lowest frequency of a periodic waveform in basic words. The basic frequency in music is the musical pitch that is heard as the lowest subpart present. Music onset detection [68], ambient sound identification [69] [70], and audio retrieval [81] all employ fundamental frequency.

#### 2-2-1-3 Spectrum shape based features:

1. Spectral centroid of an audio signal: The Spectral Centroid identifies the location of the spectrum's center of mass. It is also known as the bright-ness aspect of a sound since it defines the brightness of a sound signal. It is calculated by treating the spectrum as a distribution, with the frequencies as the values and the normalized amplitude as the odds of seeing them. The spectral centroid is a useful tool for determining the brightness of a sound signal and is used to assess music timbre [82], music categorization [83] [84], and music mood classification [85] [86]. The music's spectral centroid is higher than the speech spectral centroid [66] 2. Spectral Roll-off point of an audio signal: – Spectral Center:
This characteristic is linked to the spectral centroid. It is the measurement of the signal spectrum's median frequency. Because this is a medium frequency, the greater and lower energies are balanced. This capability is used to track musical signals' rhythm [87]
– Spectral roll-off: The frequency below which 95 percent of the signal energy is confined is known as the spectral roll-off point. Speech/music categorization [66], music genre classification [83] [84] [86] [88], musical instrument classification, and audio-based surveillance systems [74] all employ the spectral roll of. 3. Spectral Spread of an audio signal: Spectral Dispersion is another name for it. This characteristic is linked to the signal's bandwidth [87]. It may be defined as the ratemap's average deviation around its centroid. Pure tonal sounds have a narrow spectral spread, but noise-like signals have a high spectral spread. Music and ambient sounds have a large range, but spoken sounds

have a small range [66] 4. Spectral skewness of an audio signal: The symmetry of a spectrum about its arithmetic mean is measured by spectral skewness, a third-order statistical parameter. This feature is zero for soft segments and high for vocal segments. Skewness is a third-order statistical feature. A skewness of zero indicates a symmetrical distribution, a skewness less than zero indicates more energy on the right side of the spectrum, and a skewness greater than zero indicates more energy components on the left side of the spectrum. This property is used for emotion detection [86] [89], music genre classification [84] [90] [91] [92], motor bearing fault diagnosis [93] and Parkinson's disease detection [94] 5. Spectral kurtosis of an audio signal: Kurtosis, on the other hand, is a fourth-order statistical metric that indicates how flat the spectrum is around its mean value. The spectral kurtosis of a gaussian distribution is 0; if the kurtosis is less than 0, we see a flat distribution; if the spectral kurtosis is more than 0, we see a sharp peaked spectrum. Spectral Kurtosis, like spectral skewness, is employed in music genre categorization [84] [90] [92] and mood classification [86] [89], as well as failure identification in electric motor bearings [93] and Parkinson's disease diagnosis from speech [94] 6. Spectral slope of an audio signal: Spectral Slope: It is the slope of the signal's amplitude and is calculated using linear regression. Speech analysis [95] and speaker identification from a speech signal [96] both employ this characteristic. Spectral Decrease: It explains the rate-map representation's average spectral-slope, with a particular emphasis on low frequencies [97] 7. Spectral Flatness: The frequency distribution of the power spectrum's

frequency distribution is measured by spectral flatness. It's determined as the geometric mean divided by the arithmetic mean. It may be used to tell the difference between harmonic and noisy sounds. The spectrum flatness of harmonic sounds is near to zero, whereas the value of spectral flatness for noise-like sounds is close to one. It's used for music onset detection [68], music classification [96], and audio fingerprinting. 8. Spectral Crest Factor: The spectral crest factor, in contrast to spectral flatness, indicates how peaked the power spectrum of the sound stream is. It's also utilized to tell the difference between harmonic/tonal and noise-like sounds. Harmonic/tonal sounds have a greater value, whereas noise-like sounds have a lower value. Audio fingerprinting [96] and music categorization [84] both exploit this characteristic. 9. OBSC extarction algorithm: It's the difference between peaks and troughs measured using octave scale filters in sub-bands. It has been used to classify music [84] and to classify music moods [86] [89]

10. Mel Frequency Cepstral Coefficients (MFCCs):

The cepstral representation of an audio sample is used to generate MFCCs. The discrete cosine transform of log power spectrum on a nonlinear mel scale is used to depict the short-time power spectrum of an audio clip. The frequency bands of MFCCs are evenly spaced on the mel-scale, which closely resembles the human auditory system, making MFCCs a crucial characteristic in a variety of audio signal processing applications. Speech recognition [98], speech augmentation [99], speaker recognition [100], music genre classification [101], music information retrieval [**?**] [102], audio similarity

measurement [101], vowel detection [73], and other applications have all employed MFCCs.

11. Linear Prediction Cepstral Coefficients (LPCCs): The cepstrum provides a variety of benefits, including soucre-filter separation, orthogonality, compactness, and so on. Cepstrum coefficients are resilient and useful for machine learning because of their qualities. However, because linear prediction coefficients (LPC) are very sensitive to numerical precision, it is preferable to convert them to the cepstral domain. The altered coefficients are referred to as LPCCs. LPCs are said to be the source of LPCCs. Speech recognition [103], speech analysis [104], noise reduction [105], music genre categorization [106], and other applications have all employed LPCCs. The processes to calculate LPCCs from an audio signal are described in the algorithm below. 12. PLP cepstral extraction:

PLP coefficients are based on the following concepts: crucial band spectral resolution, equal-loudness curve, and intensity loudness power law [107]. Perceptual processing is used to produce the PLP coefficients from the linear prediction coefficients (LPC) before auto-regressive modeling. The linear coefficients are then transformed to cepstral coefficients after this procedure. Emotion recognition [108], speech recognition [109], infant crying sound analysis [110], animal sound vocalization analysis [111], and other applications have all used it. 13. Relative-spectral PLP (RASTA-PLP) feature:

RASTA is a method that smooths out short-term noise changes and removes any consistent offset caused by static spectral colouring in the speech signal by applying a

band-pass filter to the energy in each frequency sub-band. The RASTA-PLP features are a noise-resistant variant of the PLP characteristics. The RASTA-PLP function has the benefit of attempting to include the human auditory system's noise cancelling feature. Speech recognition [112], gender categorization [113], speaker verification [114], and other applications use this hybrid feature. The algorithm below explains how to extract RASTA-PLP characteristics

14. GFCCs extraction:

As a generalized form of MFCCs, GFCCs [115] were introduced. GFCCs employ mel-scale characteristics that are theoretically well-founded for almost all terrestrial animals and provide good voice representation for nearly all species. GFCCs are derived from the greenwood equation and may be applied with only a rudimentary understanding of a species' lowest and maximum frequency range. For all species, the Greenwood equation approximately maps the cochlear-frequency position. This property is widely used in environmental sound identification, particularly in the classification of animals and birds [116]

15. GTCCs extraction:

Noise reduction is the fundamental issue with automated speech recognition (ASR) systems. In several ASR systems in recent years, the GTCCs have demonstrated noise resilience. GTCCs are built on gammatone filter banks, which provide a cochleagram as an output, which is a frequency-time representation of a sound source. The extraction of

GTCCs is identical to that of MFCCs, with the exception that the melfilter bank is substituted with the gammatone filter bank. GTCCs, like MFCCs, can have extra properties such as delta GTCCs and delta-delta GTCCs, which are the first and second order derivatives of GTCCs, respectively. These cepstral characteristics are employed in sound detection in the environment [117] and speech recognition [118] 16. Local Binary pattern from spectrogram:

Face detection, face identification, object detection, and other computer vision applications employ the local binary pattern. Local spatial information and gray scale contrast are measured by LBPs. The LBPs are extracted from the spectrograms of the signals in audio signal processing and used in audio scene classification [119] [120], depression analysis from speech [121], snore sound discrimination [122], emotion detection [123], and pathological voice detection (Cordectomy and frontolateral resection diseases) detection [124] 17. Local Ternary Pattern Extraction:

LTPs are the LBPs' extensions. LTPs, like LBPs, are derived from a signal's visual description, such as a spectrogram. However, unlike LBPs, it does not filter pixels into a binary pattern of 0s and 1s, instead using a threshold constant to divide pixels into three values: 1, 0 and 1. In this way, each threshold pixel may have any of these three values, and after thresholding, neighboring pixels could be merged to form ternary patterns. The LTPs are employed in audio scene categorization [125], fall detection in older individuals by sound analysis [126], and health care monitoring through speech analysis [127]. The

following is the algorithm for extracting LTPs:

18. HOG descriptor from spectrogram:

HOG features, like LBPs, are used to extract time-frequency information from spectrograms. This property has also been utilized to classify acoustic scenes [120] [128], discriminate snoring sounds [122], and identify emotions [123]

19. SIFT descriptor:

SIFT is a computer vision feature extraction approach that was originally used to detect local information in pictures. This characteristic is also used to detect local information in spectrograms of audio streams. It has been extensively employed in the detection of emotions [123] and the categorization of audio-video concepts [129]

### 2.2.2 Spectogramme:

A spectrogram is a visual representation of the frequency spectrum of a signal as it varies over time. When applied to an audio signal, spectrograms are sometimes called sonographs, voiceprints, or vocograms. When data is represented in a 3D plot, it can be called waterfalls.

**Deep Features:**

Deep learning has been shown to be an effective method for extracting high-level characteristics from low-level data. Deep features are features collected from the hidden layers of various deep learning algorithms. Convolutional neural networks (CNNs), deep

neural networks (DNNs), recurrent neural networks (RNNs), Deep stacked auto-encoder (SAE), unidirectional long short term memory network (LSTM), bi-directional long short term memory (BLSTM), and other similar models might be used to extract deep features. Deep neural networks are used to extract deep features (DNNs). The DNNs are fed the MFCCs or any other relevant audio characteristic as input. The depth of the deep features is determined by the depth of the neural network. The deep characteristics provided by lower layers in a shallow neural network may be regarded of as speakeradapted features. Class-based discriminating traits might be retrieved from the higher levels. A DNN's bottleneck layer might potentially be used to extract deep features. Figure 05 depicts the extraction of deep features from a DNN's bottleneck layer. DAF stands for deep audio features in this diagram.

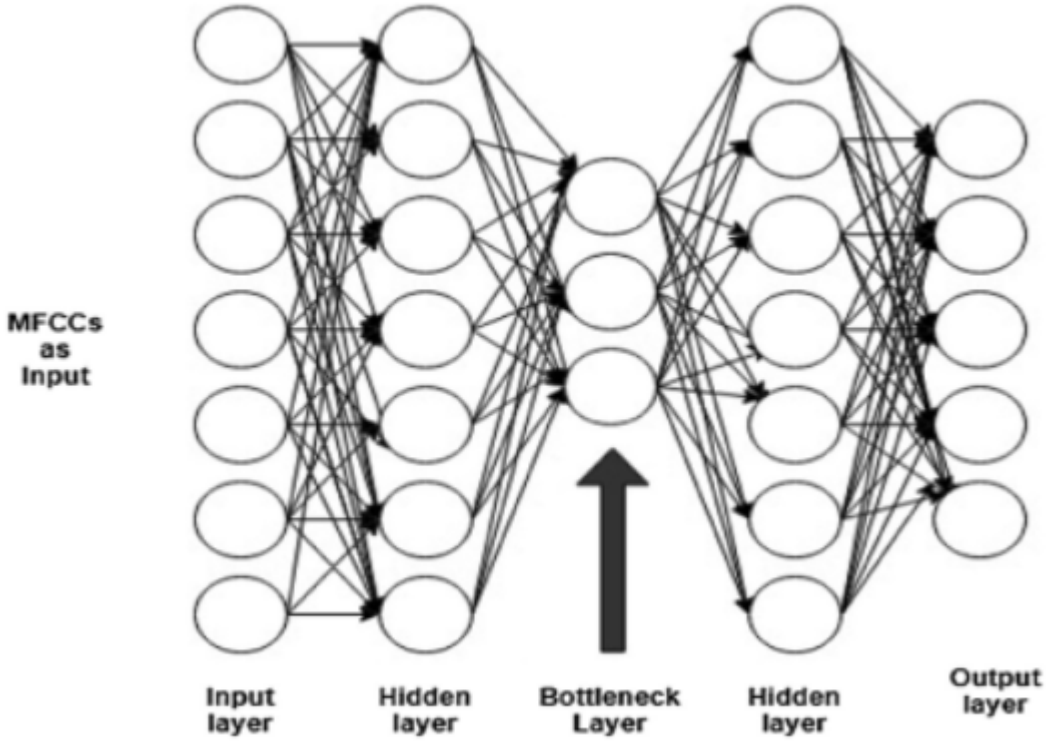



**Figure 2.7:** Deep feature from bottleneck layer.

Convolution layers, pooling layers, and fully linked layers are the three fundamental components of any CNN design, as depicted in Fig. 06. On the spectrogram of an audio stream, convolution layers apply a fixed number of convolutional filters. A feature map is the result of this layer. The dimension of the feature maps created by convolutional layers is reduced by the pooling layer, resulting in a reduction in processing time. The global features of the local feature maps are extracted using fully linked layers. Deep features can be retrieved from any of these levels; for example, the early layers of the CNN provide deep features, which are nothing more than spectrogram pixels and edges. Deep features with

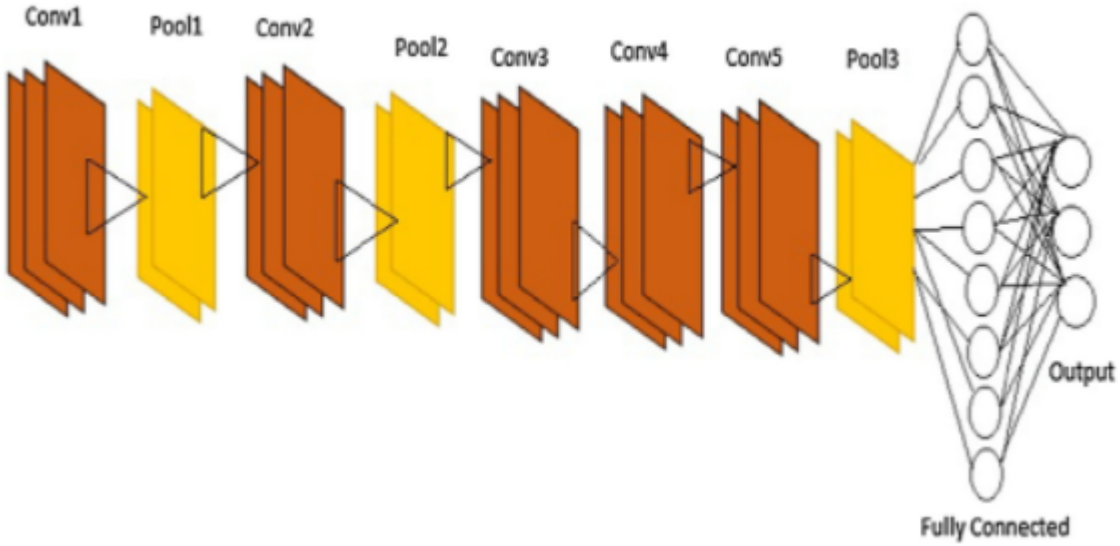



**Figure 2.8:** Basic structure of CNN.

good discrimination are produced by the upper levels. while the extracted deep features From the local feature maps, fully linked layers derive global features. Deep features can be derived from any of these levels; for example, the earliest layers of the CNN provide deep features, which are nothing more than spectrogram pixels and edges. Deep features with great discriminative power are produced by the upper layers. The deep features retrieved from the completely connected layer provide global information, whereas the deep features collected from the fully connected layer provide local information.

The auto-encoder is a form of artificial neural network that is mostly used for dimension reduction and operates unsupervised. The stacked auto-encoder (SAE) is a neural network with many layers of sparse auto-encoders, with the output of each layer being supplied as the input to the next layer. After the SAE has been trained, the deep

features may be recovered from the unsupervised trained SAE's hidden deep layers and used for a variety of applications. RNNs are used to encode the sequential knowledge in the model. RNNs have an advantage over DNNs in that they can process sequential knowledge and can internally memorize information. LSTM is the name of the memory unit. The types of memory element are unidirectional or bidirectional LSTM. In unidirectional LSTM, information flows only in one way, but in BLSTM, one LSTM layer processes data in one direction and another LSTM layer processes data in the opposite direction. Deep features might be recovered from any LSTM-RNN deep layer.

Acoustic scene categorization [130] [131], speaker recognition [132], audio-video analysis [133], gender recognition [134], emotion recognition [135], and spoofing detection [136] have all exploited deep features in audio signal processing.

## 2.3 Related work:

Corona virus disease Covid-19, which was declared a pandemic by the World Health Organization, first appeared in the Chinese city of Wuhan, and spread throughout China and other countries in the world in a short time [135], and this virus causes acute respiratory infection with a high death rate. It poses a serious threat to humans. The most common symptoms of the outbreak of Covid-19 are high fever, dry cough and difficulty breathing, acute respiratory distress syndrome (ARDS) may develop after a while in seriously ill people, and for the detection of Covid-19 disease series of examinations that

must be carried out at the level of hospitals or private laboratories. This leads to overcrowding and spread of infection, as RT-PCR is a time-consuming and costly test with delayed results. To reduce crowding of people within treatment areas and reduce infection between them, some studies have been conducted using the analysis of cough audio data taken from a mobile phone, and an approach called feature extraction from cough audio data with deep learning techniques has been used, facilitating the detection and classification of Covid-19 disease. The table 1 shows some of the studies and results that have been done to classify cough:

| Dataset | Features extraction | Proposed Approach | Metric (Accuracy) |
|---|---|---|---|
| [20] | MFCC | LSTM | 62% |
| [12] | MFCC | LSTM | 71% |
| [20] | MFCC | LSTM | 73% |
| [21] | Spectrogram | ResNet-50 | 88% |
| [21] | MFCC+Spectrogram | ResNet-50+DNN | 89% |
| [21] | Proposed feature set +Spectrogram | ResNet-50+DNN | 94% |

**Table 2.1:** some of the studies and results that have been done to classify cough

## 2.4 Conclusion:

In this chapter, the general features of the sound were studied, including the extracted features: the features from time domain, frequency domain, time-frequency domain, cepstral domain, phase domain,sparse domain, eigen domain, wavelet domain and image/-texture based features,and deep features

# Chapter 3

# Proposed Approche and Discussion

## 3.1 Introduction:

Throughout this chapter we will propose a pipeline for classifying cough sounds and move on to pre-processing for cough sound data analysis. At this point, we will also focus on one of the features extracted in Chapter 2, MFCC, as well as define the model that will be used to classify and define the input dataset for training and testing. We will also discuss the results obtained by creating and training the classification model.

## 3.2 Proposed Pipeline:

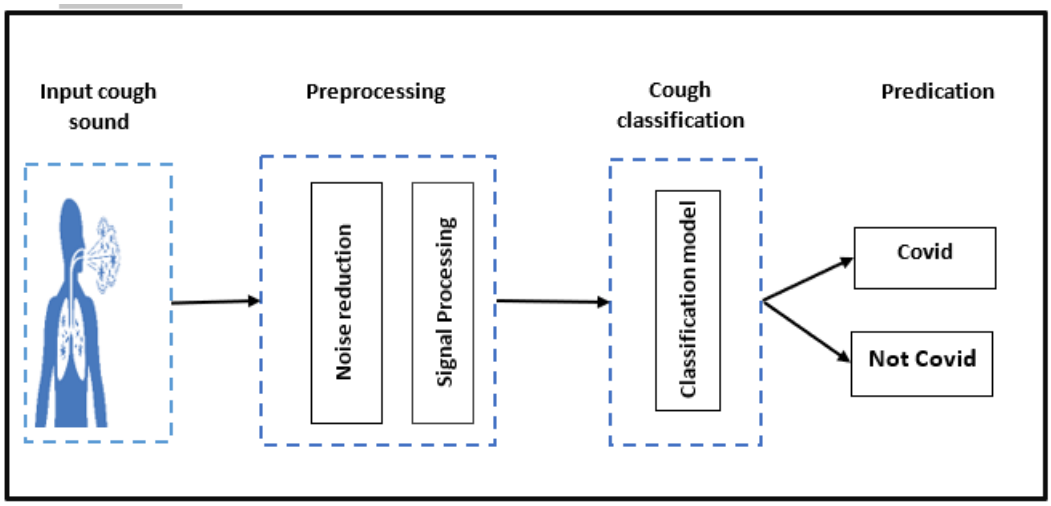



**Figure 3.1:** An overview of a proposed method for classifying COVID-19 using the cough sound.

### 3.2.1 Pre-processing:

- **Input audio file :**

Wav format, single channel, 22050sample rate.

- **Noise reduction :**

Reduces noise in the sound file of a cough using reduce noise function from the noise library in Python.

- **Signal Processing :**

In this part we convert already preprocessed coughs recordings into a feature for each cough.

### 3.2.2 Feature Engineering:

- **Extract Features:**

Here we will be using Mel-Frequency Cepstral Coefficients(MFCC)-40 frequencies that characterizes the overall shape of a spectral envelope from the cough audio samples. the feature were obtained by means of the Python librosa library. is a time series of length compatible with the length of the original cough wave, so the final feature is an average of the corresponding time series, which helps us not only significantly reduce dimensions, but also avoid the problems with the different feature length.

### 3.2.3 Classification Model:

An Artificial Neural Network (ANN) is an information processing model inspired by the brain. ANNs, like people, learn through examples. An ANN is configured for a specific application, such as pattern recognition or data classification (eg, sound classification), through a learning process. Learning largely involves modifications to the synaptic connections between neurons. It has obtained superior results in a variety of applications of voice classification [ ] As reported in the literature, ANN is among the best classifiers.

Based on its great success and why not use it in this study To differentiate between a COVID-19 positive cough and a COVID-19 negative cough. We chose this model for:

-Ability to work with incomplete knowledge

-Having fault to lerance

-Having a distributed memory

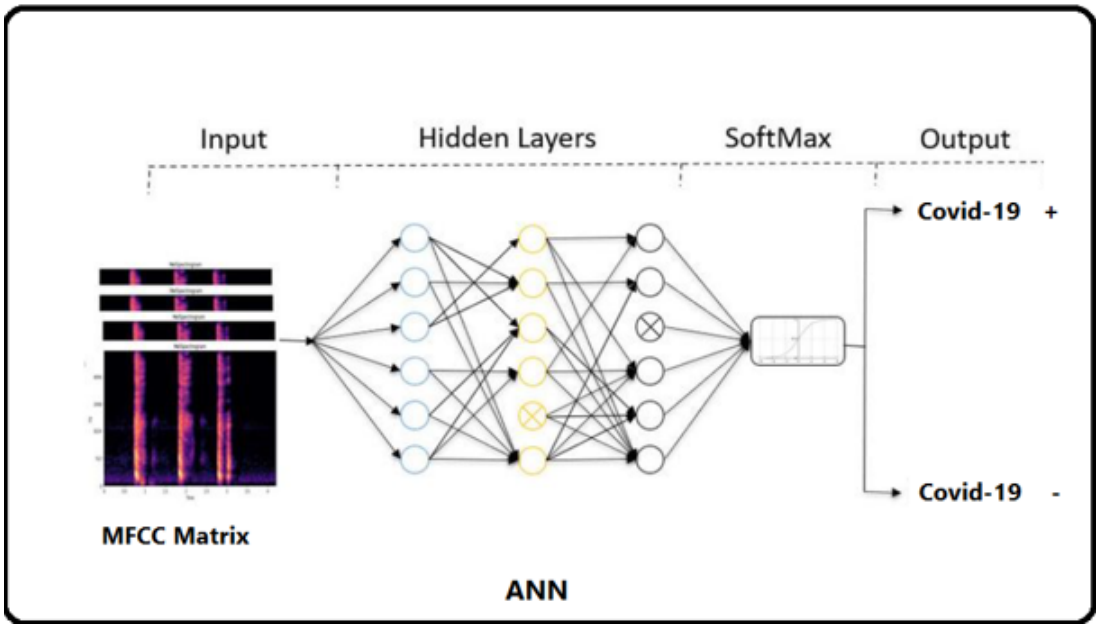



**Figure 3.2:** Artificial neural network architecture

## 3.3 Trained and Tested Dataset:

The dataset contains the sounds signals of people with Covid-19 and people who are not infected, containing 733 coughs from people tested as non-covid with the virus and 17 coughs from people with covid. Cough recordings have the following features: sample rate (22050),

audio format(.wav)

## 3.4 Codes each step:

Before everything, The library and packages necessary to start work must be called, as shown in the following figure:

```
#import library
import os
import pandas as pd
import librosa
import librosa.display
import IPython.display as ipd
from tqdm import tqdm
import numpy as np
from  glob import  glob
import matplotlib.pyplot as plt
import noisereduce as nr
```

**Figure 3.3:** Importing library

We load the database and put it in a variable as shown in the following figure:

```
[ ]  # load audio files with librosa
     cough_files  = glob('/content/drive/MyDrive/trial_covid/*.wav')
     c, sr = librosa.load(cough_files[0])
```

**Figure 3.4:** Load data

We used the mfcc() function Librosa that generates an MFCC from time-series audio data as shown in the following figure:

```
mfcc = librosa.feature.mfcc(y=c, sr= sr, n_mfcc=40)
librosa.display.specshow(mfcc, sr=sr, x_axis='time')
```

**Figure 3.5:** MFCC Extract

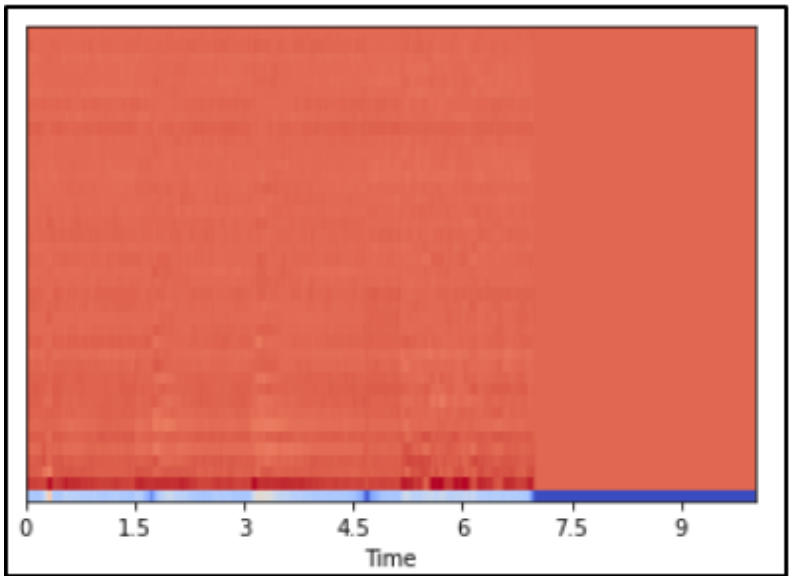



**Figure 3.6:** MFCC Show

Extract the MFCC for each file in the dataset and store it in a Panda Dataframe along with its label as shown in the following figure:

```
import numpy as np
extracted_features = []
for index_num, row in tqdm(metadata_path.iterrows()):
    file_name = str(row['file'])
    final_class_labels = str(row['status'])
    data = feature_extraction(file_name)
    extracted_features.append([data, final_class_labels])
```

```
750it [00:28, 26.47it/s]
```

**Figure 3.7:** Extract the MFCC for each file

```
### converting extracted_features to Pandas dataframe
extracted_features_df=pd.DataFrame(extracted_features,columns=['feature','status'])
extracted_features_df.head()
```

**Figure 3.8:** : Convert extracted features into frames data

This figure shows the code for creating a model by going through several layers to the end.

```
num_labels=y.shape[1]
#create the Network
model=Sequential()
###first layer
model.add(Dense(100,input_shape=(40,)))
model.add(Activation('relu'))
model.add(Dropout(0.5))
###second layer
model.add(Dense(200))
model.add(Activation('relu'))
model.add(Dropout(0.5))
###third layer
model.add(Dense(100))
model.add(Activation('relu'))
model.add(Dropout(0.5))

###final layer
model.add(Dense(num_labels))
model.add(Activation('sigmoid'))
```

**Figure 3.9:** ANN model for Classification

Print the network structure

```
# Print model architecture summary
model.summary()
```

**Figure 3.10:** Print network structure

The result of print network structure shown in figure:

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 100)               4100

 activation (Activation)     (None, 100)               0

 dropout (Dropout)           (None, 100)               0

 dense_1 (Dense)             (None, 200)               20200

 activation_1 (Activation)   (None, 200)               0

 dropout_1 (Dropout)         (None, 200)               0

 dense_2 (Dense)             (None, 100)               20100

 activation_2 (Activation)   (None, 100)               0

 dropout_2 (Dropout)         (None, 100)               0

 dense_3 (Dense)             (None, 2)                 202

 activation_3 (Activation)   (None, 2)                 0

=================================================================
Total params: 44,602
Trainable params: 44,602
Non-trainable params: 0
_________________________________________________________________
```

**Figure 3.11:** The result of print network structure

After that we compiling the model and start the training and the test:

```
#Compile the models
model.compile(loss='categorical_crossentropy',metrics=['accuracy'],optimizer='adam')
```

```
history=model.fit(X_train, y_train, batch_size=num_batch_size, epochs=num_epochs, validation_data=(X_test, y_test), callbacks=[checkpointer]
duration = datetime.now() - start
print("Training completed in time: ", duration)
```

**Figure 3.12:** Code of compiling the model and start train and test

```
Epoch 93: val_loss did not improve from 0.09805
19/19 [==============================] - 0s 5ms/step - loss: 0.1392 - accuracy: 0.9750 - val_loss: 0.1419 - val_accuracy: 0.9800
Epoch 94/100
17/19 [=========================>....] - ETA: 0s - loss: 0.1414 - accuracy: 0.9743
Epoch 94: val_loss did not improve from 0.09805
19/19 [==============================] - 0s 6ms/step - loss: 0.1443 - accuracy: 0.9750 - val_loss: 0.1383 - val_accuracy: 0.9800
Epoch 95/100
19/19 [==============================] - ETA: 0s - loss: 0.1359 - accuracy: 0.9767
Epoch 95: val_loss did not improve from 0.09805
19/19 [==============================] - 0s 6ms/step - loss: 0.1359 - accuracy: 0.9767 - val_loss: 0.1647 - val_accuracy: 0.9800
Epoch 96/100
15/19 [======================>.......] - ETA: 0s - loss: 0.1461 - accuracy: 0.9708
Epoch 96: val_loss did not improve from 0.09805
19/19 [==============================] - 0s 6ms/step - loss: 0.1359 - accuracy: 0.9733 - val_loss: 0.1550 - val_accuracy: 0.9800
Epoch 97/100
19/19 [==============================] - ETA: 0s - loss: 0.1938 - accuracy: 0.9733
Epoch 97: val_loss did not improve from 0.09805
19/19 [==============================] - 0s 5ms/step - loss: 0.1938 - accuracy: 0.9733 - val_loss: 0.1588 - val_accuracy: 0.9800
Epoch 98/100
15/19 [======================>.......] - ETA: 0s - loss: 0.1267 - accuracy: 0.9812
Epoch 98: val_loss did not improve from 0.09805
19/19 [==============================] - 0s 5ms/step - loss: 0.1572 - accuracy: 0.9767 - val_loss: 0.1233 - val_accuracy: 0.9800
Epoch 99/100
18/19 [===========================>..] - ETA: 0s - loss: 0.1773 - accuracy: 0.9740
Epoch 99: val_loss did not improve from 0.09805
19/19 [==============================] - 0s 5ms/step - loss: 0.1719 - accuracy: 0.9750 - val_loss: 0.1525 - val_accuracy: 0.9800
Epoch 100/100
18/19 [===========================>..] - ETA: 0s - loss: 0.1335 - accuracy: 0.9757
Epoch 100: val_loss did not improve from 0.09805
19/19 [==============================] - 0s 6ms/step - loss: 0.1297 - accuracy: 0.9767 - val_loss: 0.1231 - val_accuracy: 0.9800
Training completed in time:  0:00:13.218381
```

**Figure 3.13:** sown the execute training

## 3.5 Discussion:

Cough is one of the most important symptoms in clinical trials of the respiratory system, while objective indicators of cough severity are severely absent. It cannot be measured accurately due to technical conditions.

We wanted to enter into the adventure and try more in this field, despite the difficulties we faced, the failure of implementation due to the lack of Internet flow and lack of time.

For this, we used functions found in library Liberosa to help us extract the features from the cough sound and then enter these features into the ANN model in order to train it to classify the cough between a sick person or not.

After training the model several times until we reach an result, which is 97.6%.

## 3.6 Conclusion :

COVID-19 is a large-scale contagious respiratory disease that has spread across the world in 2020. Therefore, a low-cost, fast, and easily available solution is needed to provide a COVID-19 diagnosis to curb the outbreak. According to recent studies, one of the main symptoms of COVID-19 is coughing. The goal of this research effort is to develop a method for the automatic diagnosis of COVID-19 by detecting cough during recorded conversations.

# Conclusion

Coronaviruses are a broad family of viruses that produce symptoms ranging from the common cold to severe acute respiratory syndrome flu-like strains (SARS). In the Chinese city of Wuhan, a novel coronavirus was discovered in 2019.

The proposed framework has been built by including recorded audios from datasets, which have been sequentially taken as input, then processed and analyzed by a model based AI to classify this audio with illustrating a probability of having Covid-19, then if the patient recorded cough demonstrates a high probability, this patient will be considered a suspect case, and further investigation will be required,the proposed system work as follows:

Request an audio record, Apply the model based AI to analyze the recorded audio, after a probability shows the percentage of covid-19 infection Training phase has 80% audios, Testing phase has 20% audios; The modification of model was elaborated in Preprocessing phase, study was approved by the following datasets: .....for coughs

## 3.7 Limitation of this work

This research has veered from one axe point to another due to a number of challenges that we encountered while working on the initial topic. This last has been committed to collecting data, seeing and observing local factors, and obtaining insights research on one's own data; also, this study aids the healthcare service in reaching peaks, since we are all aware of the low level of hospitals and reanimation rooms. There have been several impediments that have prevented this task from taking place; we list a few of them below:

This investigation, in order to be completed to a high degree, requires resources and time, as well as the removal of several roadblocks that impede the suggested goal work's development. The government has backed the collaboration in order to find a solution to the present pandemic, as Algerian inhabitants have a poor degree of access to healthcare.

In this new sort of virus, which causes a range of symptoms, the correlation and confusion have developed, while there is an asymptomatic derivation.

# Bibliography

[1] G. Sanderson, "Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet infectious diseases. 2020; 20(5):533–534.https://doi.org/10.1016/S1473-3099(20)30120-1PMID: 32087114." `ANitty-GrittyExplanationofHowNeuralNetworksReallyWork`, 2017. Accessed: 2020-04-11.

[2]

[3] "Padma T. India's COVID-vaccine woes-by the numbers. Nature. 2021;. https://doi.org/10.1038/d41586-021-00996-y PMID: 33854229,"

[4] *World health organization (WHO). Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19);. https://www.who.int/docs/default source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.*

[5] *Paules Catharine I and Marston Hilary D and Fauci Anthony S. Coronavirus infections more than just the common cold. Jama. 2020; 323(8):707–708.*

*https://doi.org/10.1001/jama.2020.0757 PMID:31971553.*

[6] *Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nature medicine. 2020; 26(7):1037–1040. https://doi.org/10.1038/s41591-020-0916-2 PMID: 32393804.*

[7] *Liang T, et al. Handbook of COVID-19 prevention and treatment. The First Affiliated Hospital, ZhejiangUniversity School of Medicine Compiled According to Clinical Experience. 2020; 68.*

[8] *Lim J, Lee J. Current laboratory diagnosis of coronavirus disease 2019. The Korean Journal of Internal Medicine. 2020; 35(4):741. https://doi.org/10.3904/kjim.2020.257 PMID: 32668512.*

[9] *Wu D, Gong K, Arru CD, Homayounieh F, Bizzo B, Buch V, et al. Severity and Consolidation Quantification of COVID-19 From CT Images Using Deep Learning Based on Hybrid Weak Labels. IEEE Journal of Biomedical and Health Informatics. 2020; 24(12):3529–3538. https://doi.org/10.1109/JBHI.2020.3030224 PMID: 33044938.*

[10] *Zhang N, Wang L, Deng X, Liang R, Su M, He C, et al. Recent advances in the detection of respiratory virus infection in humans. Journal of medical virology. 2020; 92(4):408–417. https://doi.org/10.1002/jmv.25674 PMID: 31944312.*

[11] *Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology. 2020; p. 200432. https://doi.org/10.1148/radiol.2020200432 PMID: 32073353.*

[12] *Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing in coro-navirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology. 2020; p. 200642. https://doi.org/10.1148/radiol.2020200642 PMID: 32101510.*

[13] *Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, et al. The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society. Chest. 2020;. https://doi.org/10.1016/j.chest.2020.04.003.*

[14] *Mohammad-Rahimi H, Nadimi M, Ghalyanchi-Langeroudi A, Taheri M, Ghafouri-Fard S. Application of machine learning in diagnosis of COVID-19 through X-ray and CT images: a scoping review. Frontiers in cardiovascular medicine. 2021; 8:185. https://doi.org/10.3389/fcvm.2021.638011 PMID: 33842563.*

[15] *O zkaya U, O ztu rk Ş, Barstugan M. Coronavirus (COVID-19) classification using deep features fusion and ranking technique. In: Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach. Springer; 2020. p. 281–295.*

[16] *Wang X, Deng X, Fu Q, Zhou Q, Feng J, Ma H, et al. A weakly-supervised framework for COVID-19classification and lesion localization from chest CT. IEEE transactions on medical imaging. 2020; 39(8):2615–2625. https://doi.org/10.1109/TMI.2020.2995965 PMID: 33156775.*

[17] *Majid A, Khan MA, Nam Y, Tariq U, Roy S, Mostafa RR, et al. COVID19 classification using CT images via ensembles of deep learning models. Computers, Materials and Continua. 2021; p. 319–337. https://doi.org/10.32604/cmc.2021.016816.*

[18] *Khan MA, Kadry S, Zhang YD, Akram T, Sharif M, Rehman A, et al. Prediction of COVID-19-pneumonia based on selected deep features and one class kernel extreme learning machine. Computers Electrical Engineering. 2021; 90:106960. https://doi.org/10.1016/j.compeleceng.2020.106960 PMID:33518824.*

[19] *Akram T, Attique M, Gul S, Shahzad A, Altaf M, Naqvi SSR, et al. A novel framework for rapid diagnosis of COVID-19 on computed tomography scans. Pattern analysis and applications. 2021; p. 1–14.*

[20] *Zhao W, Jiang W, Qiu X. Deep learning for COVID-19 detection based on CT images. ScientificReports. 2021; 11(1):1–12. https://doi.org/10.1038/s41598-021-93832-2 PMID: 34253822.*

[21] *Khan MA, Alhaisoni M, Tariq U, Hussain N, Majid A, Damas˘evičius R, et al. COVID-19 Case Recognition from Chest CT Images by Deep Learning, Entropy-Controlled Firefly Optimization, and Parallel Feature Fusion. Sensors. 2021; 21(21):7286. https://doi.org/10.3390/s21217286 PMID: 34770595.*

[22] *Shui-Hua W, Khan MA, Govindaraj V, Fernandes SL, Zhu Z, Yu-Dong Z. Deep rank-based average pooling network for COVID-19 recognition. Computers, Materials, Continua. 2022; p. 2797–2813.*

[23] *Zhang YD, Khan MA, Zhu Z, Wang SH. Pseudo zernike moment and deep stacked sparse autoencoder for COVID-19 diagnosis. Cmc-Computers Materials Continua. 2021; p. 3145–3162. https://doi.org/10.32604/cmc.2021.018040.*

[24] *Kaushik H, Singh D, Tiwari S, Kaur M, Jeong CW, Nam Y, et al. Screening of COVID-19 patients using deep learning and IoT framework. Cmc-Computers Materials Continua. 2021; p. 3459–3475. https:// doi.org/10.32604/cmc.2021.017337.*

[25] *Wang L, Lin ZQ, Wong A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. Scientific Reports. 2020; 10(1):1–12. https://doi.org/10. 1038/s41598-020-76550-z PMID: 33177550.*

[26] *. Khasawneh N, Fraiwan M, Fraiwan L, Khassawneh B, Ibnian A. Detection of COVID-19 from Chest X-ray Images Using Deep Convolutional Neural Networks. Sensors. 2021; 21(17):5940. https://doi.org/10.3390/s21175940 PMID: 34502829.*

[27] *Jain R, Gupta M, Taneja S, Hemanth DJ. Deep learning based detection and analysis of COVID-19 on chest X-ray images. Applied Intelligence. 2021; 51(3):1690–1700. https://doi.org/10.1007/s10489-020-01902-1.*

[28] *Diaz-Escobar J, Ordo ñez-Guille n NE, Villarreal-Reyes S, Galaviz-Mosqueda A, Kober V, Rivera-Rodri-guez R, et al. Deep-learning based detection of COVID-19 using lung ultrasound imagery. Plos one.2021; 16(8):e0255886. https://doi.org/10.1371/journal.pone.0255886 PMID: 34388187.*

[29] *Brown C, Chauhan J, Grammenos A, Han J, Hasthanasombat A, Spathis D, et al. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. arXiv preprint arXiv:200605919.2020;.*

[30] *Andrès E, Gass R, Charloux A, Brandt C, Hentzler A. Respiratory sound analysis in the era of evidence- based medicine and the world of medicine 2.0. Journal of medicine and life. 2018; 11(2):89. PMID:30140315.*

[31] *Pramono RXA, Imtiaz SA, Rodriguez-Villegas E. Evaluation of features for classification of wheezes and normal respiratory sounds. PloS one. 2019; 14(3):e0213659. https://doi.org/10.1371/journal.pone.0213659 PMID: 30861052.*

[32] *Fraiwan M, Fraiwan L, Alkhodari M, Hassanin O. Recognition of pulmonary diseases from lung soundsusing convolutional neural networks and long short-term*

*memory. Journal of Ambient Intelligence and Humanized Computing. 2021; p. 1–13. https://doi.org/10.1007/s12652-021-03184-y PMID: 33841584.*

[33] *Faezipour M, Abuzneid A. Smartphone-Based Self-Testing of COVID-19 Using Breathing Sounds.Telemedicine and e-Health. 2020;. https://doi.org/10.1089/tmj.2020.0114 PMID: 32487005.*

[34] *hui Huang Y, jun Meng S, Zhang Y, sheng Wu S, Zhang Y, wei Zhang Y, et al. The respiratory sound features of COVID-19 patients fill gaps between clinical data and screening methods. medRxiv. 2020;.*

[35] *Wang B, Liu Y, Wang Y, Yin W, Liu T, Liu D, et al. Characteristics of Pulmonary auscultation in patients with 2019 novel coronavirus in china. Respiration. 2020; 99(9):755–763. https://doi.org/10.1159/000509610 PMID: 33147584.*

[36] *Deshpande G, Schuller B. An Overview on Audio, Signal, Speech, Language Processing for COVID-19. arXiv preprint arXiv:200508579. 2020;.*

[37] *Quatieri TF, Talkar T, Palmer JS. A framework for biomarkers of covid-19 based on coordination of speech-production subsystems. IEEE Open Journal of Engineering in Medicine and Biology. 2020;1:203–206. https://doi.org/10.1109/OJEMB.2020.2998051.*

[38] *Noda A, Saraya T, Morita K, Saito M, Shimasaki T, Kurai D, et al. Evidence of the Sequential Changes of Lung Sounds in COVID-19 Pneumonia Using a Novel Wireless Stethoscope with the Telemedicine System. Internal Medicine. 2020; 59(24):3213–3216. https://doi.org/10.2169/internalmedicine.5565-20PMID: 33132331.*

[39] *Laguarta J, Hueto F, Subirana B. COVID-19 Artificial Intelligence Diagnosis using only Cough Record-ings. IEEE Open Journal of Engineering in Medicine and Biology. 2020; 1:275–281. https://doi.org/10.1109/OJEMB.2020.3026928 PMID: 34812418.*

[40] *. Han J, Qian K, Song M, Yang Z, Ren Z, Liu S, et al. An Early Study on Intelligent Analysis of Speech under COVID-19: Severity, Sleep Quality, Fatigue, and Anxiety. arXiv preprint arXiv:200500096. 2020;.*

[41] *Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology. 2000;. https://doi.org/10. 1152/ajpheart.2000.278.6.H2039 PMID: 10843903.*

[42] *Shannon CE. A mathematical theory of communication. The Bell system technical journal. 1948; 27 (3):379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.*

[43] *Higuchi T. Approach to an irregular time series on the basis of the fractal theory. Physica D: Nonlinear Phenomena. 1988; 31(2):277–283. https://doi.org/10.1016/0167-2789(88)90081-4.*

[44] *Zheng F, Zhang G, Song Z. Comparison of different implementations of MFCC. Journal of Computer science and Technology. 2001; 16(6):582–589. https://doi.org/10.1007/BF02943243.*

[45] *Martinek J, Tatar M, Javorka M. Distinction between voluntary cough sound and speech in volunteers by spectral and complexity analysis. J Physiol Pharmacol. 2008; 59(6):433–440. PMID: 19218667.*

[46] *Barry SJ, Dane AD, Morice AH, Walmsley AD. The automatic recognition and counting of cough.Cough. 2006; 2. doi: 10.1186/1745-9974-2-8 PMID: 17007636.*

[47] *Tracey BH, Comina G, Larson S, Bravard M, Lopez JW, Gilman RH. Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis. In: IEEE EMBC; 2011. p. 6017–6020.*

[48] *Swarnkar V, Abeyratne UR, Amrulloh Y, Hukins C, Triasih R, Setyati A. Neural network based algorithm for automatic identification of cough sounds. In: IEEE EMBC; 2013. p. 1764–1767.*

[49] *Amrulloh YA, Abeyratne UR, Swarnkar V, Triasih R, Setyati A. Automatic cough segmentation from non-contact sound recordings in pediatric wards. Biomed Signal Process Control. 2015; 21:126–136. doi: 10.1016/j.bspc.2015.05.001 .*

[50] *Matos S, Birring SS, Pavord ID, Evans H. Detection of cough signals in continuous audio recordings using Hidden Markov Models. IEEE Trans Biomed Eng. 2006; 53(6):1078–1083. doi: 10.1109/TBME.2006.873548 PMID: 16761835.*

[51] *Liu JM, You M, Li GZ, Wang Z, Xu X, Qiu Z, et al. Cough signal recognition with Gammatone Cepstral Coefficients. In: 2013 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP). IEEE; 2013. p. 160–164.*

[52] *Lucio C, Teixeira C, Henriques J, de Carvalho P, Paiva RP. Voluntary cough detection by internal sound analysis. In: Biomedical Engineering and Informatics (BMEI), 2014 7th International Conference on; 2014. p. 405–409.*

[53] *Larson EC, Lee T, Liu S, Rosenfeld M, Patel SN. Accurate and privacy preserving cough sensing using a low-cost microphone. In: 13th International Conference on Ubiquitous Computing, UbiComp'11 and the Co-located Workshops. Beijing, China; 2011. p. 375–384.*

[54] *Chatrzarrin H, Arcelus A, Goubran R, Knoefel F. Feature extraction for the differentiation of dry and wet cough sounds. In: Medical Measurements and Applications Proceedings (MeMeA), 2011 IEEE International Workshop on. IEEE; 2011. p. 162–166.*

[55] *. Swarnkar V, Abeyratne U, Chang A, Amrulloh Y, Setyati A, Triasih R. Automatic Identification of Wetand Dry Cough in Pediatric Patients with Respiratory Diseases. Ann Biomed Eng. 2013; 41(5):1016–1028. doi: 10.1007/s10439-013-0741-6 PMID: 23354669.*

[56] *Kosasih K, Abeyratne UR, Swarnkar V, Triasih R. Wavelet Augmented Cough Analysis for Rapid Childhood Pneumonia Diagnosis. IEEE Trans Biomed Eng. 2014; 62(4):1185–1194. doi: 10.1109/TBME.2014.2381214 PMID: 25532164.*

[57] *Parker D, Picone J, Harati A, Lu S, Jenkyns MH, Polgreen PM. Detecting paroxysmal coughing from pertussis cases using voice recognition technology. PloS one. 2013; 8(12):e82971. doi: 10.1371/journal.pone.0082971 PMID: 24391730.*

[58] *A. Imran, et al., AI4COVID: AI-enabled preliminary diagnosis for COVID-19 from cough samples via an app, Informatics in Medicine Unlocked 20 (June 2020) 1–13, doi:10.1016/j.imu.2020.100378 .*

[59] *Daniel More, "Causes and Risk Factors of Cough Health Conditions Linked to Acute,Sub-Acute, or Chronic Coughs "available at https://www.verywellhealth.com/causes-of-cough-83024.*

[60] *C. Bales, et al., Can Machine Learning Be Used to Recognize and Diagnose Coughs? in:Proceedings of the IEEE International Conference on E-Health and Bioengineering (EHB),2020 October 29-30, Lasi, Romania, doi: 10.1109/EHB50910.2020.9280115*

*[Rumana Islam, 2022]Islam, Rumana, Esam Abdel-Raheem, and Mohammed Tarique. A study of using cough sounds and deep neural networks for the early detection of Covid-19.Biomedical Engineering Advances (2022): 100025.*

[61] *Thorpe W, Kurver M, King G, Salome C. Acoustic analysis of cough. In: Intelligent Information Systems Conference, The Seventh Australian and New Zealand 2001; 2001. p. 391–394.*

[62] *[COVID-19 cough classification using machine learning and global smartphone recordings, COVID-19DetectioninCough,BreathandSpeechusingDeepTransfer LearningandBottleneckFeatures .*

[63] *https://towardsdatascience.com/what-is-feature-engineering-importance-tools-and-techniques-for-machine-learning-2080b0269f10.*

[64] *Kedem B. Spectral analysis and discrimination by zero-crossings. Proc IEEE 1986;74(11):1477–93.*

[65] *Saunders J. Real-time discrimination of broadcast speech/music. 1996 IEEE international conference on acoustics, speech, and signal processing conference proceedings, vol. 2. IEEE; 1996. p. 993–6.*

[66] *Al-Shoshan AI. Speech and music classification and separation: a review. J King Saud Univ-Eng Sci 2006;19(1):95–132.*

[67] *Burred JJ, Robel A, Sikora T. Dynamic spectral envelope modeling for timbre analysis of musical instrument sounds. IEEE Trans Audio Speech Lang Process 2010;18(3):663–74.*

[68] *Smith D, Cheng E, Burnett IS. Musical onset detection using MPEG-7 audio descriptors. Proceedings of the 20th international congress on acoustics (ICA), Sydney, Australia, vol. 2327. p. 1014.*

[69] *Muhammad G, Alghathbar K. Environment recognition from audio using MPEG-7 features. In: 2009 Fourth international conference on embedded and multimedia computing. IEEE; 2009. p. 1–6.*

[70] *Valero X, Alías F. Applicability of MPEG-7 low level descriptors to environmental sound source recognition. In: Proceedings 1st Euroregio Conference, Ljubjana.*

[71] *Yang X, Tan B, Ding J, Zhang J, Gong J. Comparative study on voice activity detection algorithm. In: 2010 International conference on electrical and control engineering. IEEE; 2010. p. 599–602.*

[72] *Peltonen V, Tuomi J, Klapuri A, Huopaniemi J, Sorsa T. Computational auditory scene recognition. In: 2002 IEEE international conference on acoustics, speech, and signal processing, vol. 2. IEEE; 2002. p. II–1941.*

[73] *Korkmaz Y, Boyací A, Tuncer T. Turkish vowel classification based on acoustical and decompositional features optimized by genetic algorithm. Appl Acoust 2019;154:28–35. https://doi.org/10.1016/j.apacoust.2019.04.027.*

[74] *Rabaoui A, Davy M, Rossignol S, Ellouze N. Using one-class SVMs and wavelets for audio surveillance. IEEE Trans Inf Forensics Secur 2008;3 (4):763–75.*

[75] *Fu Z, Lu G, Ting KM, Zhang D. A survey of audio-based music classification and annotation. IEEE Trans Multimedia 2011;13(2):303–19.*

[76] *Jiang H, Bai J, Zhang S, Xu B. SVM-based audio scene classification. In: 2005 international conference on natural language processing and knowledge engineering. IEEE; 2005. p. 131–6.*

[77] *Glowacz A. Fault detection of electric impact drills and coffee grinders using acoustic signals. Sensors (Basel, Switzerland) 2019;19(2):269. https://doi.org/10.3390/s19020269..*

[78] *Glowacz A. Acoustic-based fault diagnosis of commutator motor. Electronics 2018;7(11):299. https://doi.org/10.3390/electronics7110299.*

[79] *Glowacz A. Fault diagnosis of single-phase induction motor based on acoustic signals. Mech Syst Signal Process 2019;117:65–80. https://doi.org/10.1016/j.ymssp.2018.07.044.*

[80] *Ghoraani B, Krishnan S. Time-frequency matrix feature extraction and classification of environmental audio signals. IEEE Trans Audio Speech Lang Process 2011;19(7):2197–209. https://doi.org/10.1109/TASL.2011.2118753.*

[81] *Wold E, Blum T, Keislar D, Wheaten J. Content-based classification, search, and retrieval of audio. IEEE Multimedia 1996;3(3):27–36.*

[82] *Agostini G, Longari M, Pollastri E. Musical instrument timbres classification with spectral features. EURASIP J Adv Signal Process 2003;2003(1): 943279.*

[83] *Bergstra J, Casagrande N, Erhan D, Eck D, Kégl B. Aggregate features and adaboost for music classification. Mach Learn 2006;65(2–3):473–84.*

[84] *Li T, Ogihara M, Li Q. A comparative study on content-based music genre classification. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. ACM; 2003. p. 282–9.*

[85] *Wang F, Wang X, Shao B, Li T, Ogihara M. Tag integrated multi-label music style classification with hypergraph. In: ISMIR. p. 363–8.*

[86] *Lu L, Liu D, Zhang HJ. Automatic mood detection and tracking of music audio signals. IEEE Trans Audio Speech Lang Process 2006;14(1):5–18.*

[87] *Sethares WA, Morris RD, Sethares JC. Beat tracking of musical performances using low-level audio features. IEEE Trans Speech Audio Process 2005;13(2):275–85.*

[88] *Tzanetakis G, Cook P. Musical genre classification of audio signals. IEEE Trans Speech Audio Process 2002;10(5):293–302.*

[89] *Bhat AS, Amith VS, Prasad NS, Mohan DM. An efficient classification algorithm for music mood detection in western and hindi music using audio feature extraction. In: 2014 fifth international conference on signal and image processing. p. 359–64.*

[90] *Ahrendt P, Meng A, Larsen J. Decision time horizon for music genre classification using short time features. In: 2004 12th European signal processing conference. IEEE; 2004. p. 1293–6.*

[91] *Tzanetakis G, Cook P. Musical genre classification of audio signals. IEEE Trans Speech Audio Process 2002;10(5):293–302.*

[92] *Baniya BK, Lee J, Li ZN. Audio feature reduction and analysis for automatic music genre classification. In: 2014 IEEE international conference on systems, man, and cybernetics (SMC). IEEE; 2014. p. 457–62.*

[93] *Duan Z, Wu T, Guo S, Shao T, Malekian R, Li Z. Development and trend of condition monitoring and fault diagnosis of multi-sensors information fusion for rolling bearings: a review. Int J Adv Manuf Technol 2018;96(1):803–19. https://doi.org/10.1007/s00170-017-1474-8.*

[94] *Tuncer T, Dogan S. A novel octopus based Parkinson's disease and gender recognition method using vowels. Appl Acoust 2019;155:75–83. https://doi.org/10.1016/j.apacoust.2019.05.019.*

[95] *Shukla S, Dandapat S, Prasanna SRM. Spectral slope based analysis and classification of stressed speech. Int J Speech Technol 2011;14(3):245.*

[96] *Murthy HA, Beaufays F, Heck LP, Weintraub M. Robust text-independent speaker identification over telephone channels. IEEE Trans Speech Audio Process 1999;7(5):554–68.*

[97] *Peeters G, Giordano BL, Susini P, Misdariis N, McAdams S. The timbre toolbox: extracting audio descriptors from musical signals. J Acoust Soc Am 2011;130(5):2902–16.*

[98] *Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans Acoust Speech Signal Process 1980;28(4):357–66.*

[99] *Krueger A, Haeb-Umbach R. Model-based feature enhancement for reverberant speech recognition. IEEE Trans Audio Speech Lang Process 2010;18(7):1692–707.*

[100] *Sahidullah M, Saha G. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Speech Commun 2012;54(4):543–65.*

[101] *Müller M. Information retrieval for music and motion, vol. 2. Heidelberg: Springer; 2007.*

[102] *Hu N, Dannenberg RB, Tzanetakis G. Polyphonic audio matching and alignment for music retrieval. In: 2003 IEEE workshop on applications of signal processing to audio and acoustics (IEEE Cat. No. 03TH8684). IEEE;2003. p. 185–8.*

[103] *Bernard A, Alwan A. Source and channel coding for remote speech recognition over error-prone channels. 2001 IEEE international conference on acoustics,speech, and signal processing. Proceedings (Cat. No. 01CH37221), vol. 4. IEEE;2001. p. 2613–6.*

[104] *Kinjo T, Funaki K. On hmm speech recognition based on complex speech analysis. In: IECON 2006–32nd annual conference on IEEE industrial electronics. IEEE; 2006. p. 3477–80.*

[105] *Chen J, Huang Y, Li Q, Paliwal KK. Recognition of noisy speech using dynamic spectral subband centroids. IEEE Signal Process Lett 2004;11(2):258–61..*

[106] *Maddage NC, Xu C, Wang Y. A SVM C based classification approach to musical audio; 2003. .*

[107] *Hermansky H. Perceptual linear predictive (PLP) analysis of speech. J Acoust Soc Am 1990;87(4):1738–52.*

[108] *Glodek M, Tschechne S, Layher G, Schels M, Brosch T, Scherer S, Schwenker F.Multiple classifier systems for the classification of audio-visual emotional states. In: Affective computing and intelligent interaction. Berlin, Heidelberg: Springer; 2011. p. 359–68.*

[109] *Dave N. Feature extraction methods LPC, PLP and MFCC in speech recognition.Int J Adv Res Eng Technol 2013;1(6):1–4.*

[110] *Protopapas A, Eimas PD. Perceptual differences in infant cries revealed by modifications of acoustic features. J Acoust Soc Am 1997;102(6):3723–34..*

[111] *Clemins PJ, Johnson MT. Generalized perceptual linear prediction features for animal vocalization analysis. J Acoust Soc Am 2006;120(1):527–34.*

[112] *Koehler J, Morgan N, Hermansky H, Hirsch HG, Tong G. Integrating RASTA-PLP into speech recognition. Proceedings of ICASSP'94. IEEE international conference on acoustics, speech and signal processing, vol. 1. IEEE; 1994. p. I–421.*

[113] *Zeng YM, Wu ZY, Falk T, Chan WY. Robust GMM based gender classification using pitch and RASTA-PLP parameters of speech. In: 2006 international conference on machine learning and cybernetics. IEEE; 2006. p. 3376–9.*

[114] *Hardt D, Fellbaum K. Spectral subtraction and RASTA-filtering in text- dependent HMM-based speaker verification. 1997 IEEE international conference on acoustics, speech, and signal processing. IEEE; 1997. p. 867–70.*

[115] *Greenwood DD. A cochlear frequency-position function for several species-29 years later. J Acoust Soc Am 1990;87(6):2592–605.*

[116] *Valero X, Alias F. Gammatone cepstral coefficients: biologically inspired features for non-speech audio classification. IEEE Trans Multimedia 2012;14 (6):1684–9.*

[117] *Valero X, Alias F. Gammatone cepstral coefficients: biologically inspired features for non-speech audio classification. IEEE Trans Multimedia 2012;14 (6):1684–9.*

[118] *Yin H, Hohmann V, Nadeu C. Acoustic features for speech recognition based on Gammatone filterbank and instantaneous frequency. Speech Commun2011;53(5):707–15.*

[119] *Abidin S, Togneri R, Sohel F. Spectrotemporal analysis using local binary pattern variants for acoustic scene classification. IEEE/ACM Trans Audio Speech Lang Process 2018;26(11):2112–21. https://doi.org/10.1109/ TASLP.2018.2854861..*

[120] *Yang W, Krishnan S. Combining temporal features by local binary pattern for acoustic scene classification. IEEE/ACM Trans Audio Speech Lang Process 2017;25(6):1315–21. https://doi.org/10.1109/TASLP.2017.2690558.*

[121] *He L, Cao C. Automated depression analysis using convolutional neural networks from speech. J Biomed Inf 2018;83:103–11. https://doi.org/ 10.1016/j.jbi.2018.05.007.*

[122] *Demir F, Sengur A, Cummins N, Amiriparian S, Schuller B. Low level texture features for snore sound discrimination. In: Conference proceedings: annual international conference of the IEEE engineering in medicine and biology society. IEEE engineering in medicine and biology society. Annual conference.p. 413.*

[123] *Sun B, Li L, Zuo T, Chen Y, Zhou G, Wu X. Combining multimodal features with hierarchical classifier fusion for emotion recognition in the wild. In: Proceedings of the 16th international conference on multimodal interaction. ACM; 2014. p. 481–6.*

[124] *Tuncer T, Dogan S, Ertam F. Automatic voice based disease detection method using one dimensional local binary pattern feature extraction network. Appl Acoust 2019;155:500–6. https://doi.org/10.1016/j.apacoust.2019.05.023.*

[125] *Tuncer T, Dogan S. Novel dynamic center based binary and ternary pattern network using M4 pooling for real world voice recognition. Appl Acoust 2019;156:176–85. https://doi.org/10.1016/j.apacoust.2019.06.029.*

[126] *Adnan SM, Irtaza A, Aziz S, Ullah MO, Javed A, Mahmood MT. Fall detection through acoustic local ternary patterns. Appl Acoust 2018;140:296–300.https://doi.org/10.1016/j.apacoust.2018.06.013.*

[127] *Hossain MS. Patient state recognition system for healthcare using speech and facial expressions. J Med Syst 2016;40(12):1–8. https://doi.org/10.1007/ s10916-016-0627-x.*

[128] *Rakotomamonjy A, Gasso G. Histogram of gradients of time-frequency representations for audio scene classification. IEEE/ACM Trans Audio Speech Lang Process 2015;23(1):142–53. https://doi.org/10.1109/ TASLP.2014.2375575.*

[129] *Tuncer T, Dogan S. Novel dynamic center based binary and ternary pattern network using M4 pooling for real world voice recognition. Appl Acoust 2019;156:176–85. https://doi.org/10.1016/j.apacoust.2019.06.029.Jiang W, Cotton C, Chang SF, Ellis D, Loui A. Short-term audio-visual atoms for generic video concept classification. In: Proceedings of the 17th ACM international conference on multimedia. ACM; 2009. p. 5–14.*

[130] *Li Y, Zhang X, Jin H, Li X, Wang Q, He Q, Huang Q. Using multi-stream hierarchical deep neural network to extract deep audio feature for acoustic event detection. Multimedia Tools Appl 2018;77(1):897–916. https://doi.org/ 10.1007/s11042-016-4332-z.*

[131] *Li Y, Li X, Zhang Y, Wang W, Liu M, Feng X. Acoustic scene classification using deep audio feature and BLSTM network. Paper presented at the 371-374; 2018.https://doi.org/10.1109/ICALIP.2018.8455765.*

[132] *Rahmani MH, Almasganj F, Seyyedsalehi SA. Audio-visual feature fusion via deep neural networks for automatic speech recognition. Digital Signal Process 2018;82:54–63. https://doi.org/10.1016/j.dsp.2018.06.004..*

[133] *Takahashi N, Gygli M, Van Gool L. AENet: learning deep audio features for video analysis; 2017. .*

[134] *W.J. Guan, W.H. Liang, Y. Zhao, H.R. Liang, Z.S. Chen, Y.M. Li, X.Q. Liu, R.C. Chen, C.I. Tang, T. Wang, C.Q. Ou, L. Li, P.Y. Chen, L. Sang, W. Wang, J.F. Li, C.C. Li, L. M. Ou, B. Cheng, S. Xiong, Z.Y. Ni, Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis, Eur Respir J., 2020. .*

[135] *Badshah AM, Rahim N, Ullah N, Ahmad J, Muhammad K, Lee MY, Kwon S, Baik SW. Deep features-based speech emotion recognition for smart affective services. Multimedia Tools Appl 2019;78(5):5571–89. https://doi.org/ 10.1007/s11042-017-5292-7.*

[136] *Qian Y, Chen N, Yu K. Deep features for automatic spoofing detection. SpeechCommun 2016;85:43–52. https://doi.org/10.1016/j.specom.2016.10.007.*