



UNIVERSITY OF KASDI MERBAH OUARGLA



**Faculty of New Information Technologies and
Communication**

Department of Computer Science and Information Technology

Master Thesis

Domain: Computer science and Information Technology

Specialty: Informatique Industrielle

Topic

Human Cough sound for identifying COVID-19

Presented by

Chifa Khenfer, Hacini Marwa

In front of the jury composed of :

President	Laallm Fatima. Z	Professor	University of Ouargla
Supervisor	Kahlessenane Fares	M.A.A	University of Ouargla
Examiner	Benkaddour Mohammed. K	M.C.A	University of Ouargla

Class of: 2021/2022

ACKNOWLEDGMENTS

First of all, we thank **ALLAH** and thank Him for having succeeded in reaching this moment and paying it for us throughout our study life

We would like to express our gratitude to our supervisor

Mr.Kahlessenane Fares for his efforts and patience with us to complete this research

We also thank everyone who contributed to the success of this message and achieving those good results, whether from far or near

Dedicates

With the grace of **ALLAH**, first and foremost I dedicate this humble work:

To my Mother:

You have given me life, the tenderness, and the courage to succeed. to thank you for your sacrifices and for the affection with which you have always surrounded me.

To my father:

My shoulder, my strength, my security, the person who deserves all my appreciation and respect.

To my second father 'uncle Muhammad Lahsan' 'may God have mercy on him ' for everything, he gave me

To my brothers my support in life 'Farouk. Samir. Muhammad el-Sheikh'

To my sisters, my happiness in life is 'Hadjer. Hasna'.

To our Supervisor Fares KAHLESSENANE for his time and effort with us

To all my family, loved ones and friends, especially 'Omaima, Marwa, Hinda...'

To all my teachers from the first day of school until today.

To the one who helped us and guided us in spite of all circumstances

Dedicates 2

I extend my thanks to **ALLAH** Almighty first and foremost. I dedicate my graduation to the candle of my life and the secret of my existence to those who left my world and did not leave my heart, the one I always wished to be near my **dear mother**, might God have mercy on her.

To the one whom God has enjoined upon me in righteousness and kindness to the one who taught me all matters of life at the expense of his effort and energy, my **dear father**. I also dedicate my graduation to those who have always supported me in my life, my **brothers** Mourad, El-Hadj Ahmed, Youssef and Muhammad Taha and my **sisters** Zinab, Chaima, Malak and my little girls Noussaiba and djouwayria.

I also dedicate my graduation to all my family and my **dearest friends** "Chifaa, Siham and Hinda", and to **our supervisor**, Fares Kahlesanane for his efforts and hard work with us, to **my fiancé** and all those **who loved** and supported me in this life. And thanks to all of **my teachers, friends** and **colleagues** who stood with me during this long journey of success.

Abstract

Since the first appearance of the Coronavirus 19, the developments are witnessed to this day, it has made great changes in the daily life of humans, especially in health care systems, which in turn have developed more effective ways to confront it. Since most of its symptoms affect the respiratory system in the first place, which often leads to coughing. These contain the latter contains health information that can be exploited for early detection of HIV infection using audio signal analysis methods. This research focuses on the classification of cough sounds and audio samples using features extracted through signal analysis. The classifier model is built by the convolutional neural network to assess whether the audio sample is likely to contain COVID-19 symptoms. For an early remote diagnosis to screen, diagnose, monitor, and spread awareness about COVID-19.

Keywords: COVID-19, Audio Analysis, Artificial intelligence, Cough, CNN, MFCC

ملخص

منذ الظهور الأول لفيروس كورونا 19 والتطورات التي تشهدها حتى يومنا هذا، قام بتغييرات كبيرة في الحياة اليومية للبشر، وخاصة على أنظمة الرعاية الصحية، والتي بدورها طورت طرقًا أكثر فعالية لمواجهتها. نظرًا لأن معظم أعراضها تؤثر على الجهاز التنفسي في المقام الأول، مما يؤدي غالبًا إلى السعال. تحتوي هذه على هذا الأخير على معلومات صحية يمكن استغلالها للكشف المبكر عن عدوى فيروس نقص المناعة البشرية باستخدام طرق تحليل إشارة الصوت. يركز هذا البحث على تصنيف أصوات السعال وعينات الصوت باستخدام الميزات المستخرجة من خلال تحليل الإشارة. تم تصميم نموذج المصنف بواسطة الشبكة العصبية التلافيفية لتقييم ما إذا كان من المحتمل أن تحتوي عينة الصوت على أعراض COVID-19. لتشخيص عن بُعد في وقت مبكر لفحص وتشخيص ومراقبة ونشر الوعي حول

19-Covid

Résumé

Depuis la première apparition du virus Corona 19 et les développements dont il est témoin à ce jour, il a apporté des changements majeurs dans la vie quotidienne des humains, notamment sur les systèmes de santé, qui à leur tour ont développé des moyens plus efficaces pour y faire face. Comme la plupart de ses symptômes affectent le système respiratoire en premier lieu, ce qui conduit souvent à la toux. Ceux-ci contiennent des informations sur la santé qui peuvent être exploitées pour la détection précoce de l'infection par le VIH à l'aide de méthodes d'analyse du signal audio. Cet article se concentre sur la classification des sons de la toux et des échantillons de voix à l'aide de caractéristiques extraites par l'analyse du signal. Le modèle de classificateur est conçu par le Convolutional Neural Network pour évaluer si l'échantillon audio est susceptible de contenir des symptômes de COVID-19. Pour un télédiagnostic précoce pour dépister, diagnostiquer, surveiller et faire connaître le Covid-19

Mots-clés : COVID-19, Analyse audio, Intelligence artificielle, Toux, CNN, MFCC

Table of Contents

ACKNOWLEDGMENTS.....	I
Dedicates.....	II
Abstract.....	IV
LIST OF FIGURES	X
Introduction.....	1
Previous works.....	2
Covid-19.....	4
I.1. Introduction	4
I.2. Definition.....	5
I.3. Symptoms.....	5
I.4. The effect of Covid on the sound	5
I.5. How does it spread?	6
I.6. Artificial intelligence and detection of Covid from sound	7
I.7. Conclusion.....	8
Audio signal	9
II.1 Introduction	9
II.2 Sound wave	10
II.2.1 audio signal.....	10
II.3 Types of audio signal	11
II.4 Analog to Digital conversion	12
II.4.1 Sampling.....	13
II.4.2 Quantization	14
II.4.3 Encoding.....	16
II.5 audio features	16
II.6 Conclusion	18
Artificial Intelligence for sound processing.....	20
III.1 Introduction.....	20
III.2 Artificial Intelligence.....	20
III.3 Types of Artificial Intelligence.....	21
III.3.1 Reactive Machines	21
III.3.2 Limited Memory	22

III.3.3 Theory of Mind	22
III.3.4 Self-Awareness.....	22
III.4 Machine Learning.....	22
III.4.1 How to work machine learning algorithms?.....	23
III.4.2 Machine Learning applications	23
III.4.3 Machine learning types:.....	24
III.5 Deep Learning.....	27
III.5.1 Artificial neural network	27
III.5.2 Structure of deep learning algorithm	28
III.5.3 Types of Deep learning.....	29
III.6 Deep learning vs. machine learning	29
III.7 Convolutional neural networks (CNNs).....	30
III.7.1 Architecture of CNN	30
III.8 Conclusion	33
Materials, Methods & results	34
IV.1Introduction.....	34
IV.2 Implementation	34
IV.3 Audio Data acquisition.....	35
IV.3.1 Collecting Data	35
IV.4 Method.....	36
IV.4.1. Exploratory Data	36
IV.4.2 preprocessing audio.....	37
IV.4.3 Feature extraction and feature selection:	38
IV.4.4 Model architecture	39
IV.5 Code Explaining	40
IV.6 Evaluation and Results	43
IV.6.1 Evaluation data:.....	43
IV.6.2 Evaluation method:.....	43
IV.6.3 confusion matrix.....	45
IV.6.4 Discussion.....	47
IV.6 Conclusion.....	48

Conclusion & perspectives	49
Bibliography	51

LIST OF TABLES

Tableau 1.1 : Structure of confusion matrix.....	43
Tableau1.2: Performance Matrix.....	44

LIST OF FIGURES

FIG.II.1 - Sound wave	9
FIG.II.2 - Human auditory system	10
FIG.II.3 - Analog vs. Digital signals	11
FIG.II.4 - Converting an analog signal to a digital.....	11
FIG.II.5 - Natural sampling	13
FIG.II.6 - Flat top sampling	13
FIG.II.7 - Quantization process	14
FIG.III.1 - Types of Artificial Intelligence.....	20
FIG.III.2 - How to work machine learning algorithms	22
FIG.III.3 - Machine Learning types.....	23
FIG.III.4 - Deep Learning	25
FIG.III.5 - Artificial neural network.....	26
FIG.III.6 - Structure of deep learning algorithm.....	26
FIG.III.7 - Deep learning vs machine learning.....	28
FIG.III.8 -Architecture of CNN.....	28
FIG.IV.1 - The steps for building the proposed model.....	34
FIG.IV.2 - Waveform visualization of an audio sample positive (first) and sample negative (second).....	35
FIG.IV.3 - Steps of MFCC generation.....	37
FIG.IV.4 - Importing library.....	38
FIG.IV.5 - Loading our dataset.....	39
FIG.IV.6 - Preprocessing stage.....	39
FIG.IV.7 - Extract the features.....	40
FIG.IV.8 -: model architecture.....	40

FIG.IV.9 - Percentage of Accuracy.....	42
FIG.IV.10 - loss of the model.....	42
FIG.IV.11 - Accuracy of the model.....	43
FIG.IV.12 - Confusion matrix.	45
FIG.IV.13 - Performance Matrix.....	45

Introduction

The coronavirus pandemic is an ongoing global pandemic of coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2, which has killed more than 6.29 million people and infected more than 531 million others in more than 188 countries and territories to date. Date 4 June 2022.

COVID-19 has become one of the problems that humanity has to live with. Therefore, it is necessary to discover quick and easy ways to discover it. Therefore, that it does not affect the daily course of human life. Detecting Covid 19 by sound signals is considered an innovative method. In this research, we focus on the auditory effects that COVID-19 symptoms can have on individuals, as changes can be detected by analyzing audio recordings of an individual's cough. People without symptoms may differ from healthy people in the way they cough. These differences cannot be deciphered by adopting the human ear, but the differences between them can be captured by artificial intelligence techniques using machine learning and deep learning methodologies trained on cough samples from healthy and sick individuals. Indicates disease progression and treatment efficacy thanks to prompt screening, early remote diagnosis, and limiting its spread by training a classifier based on COVID-19 prediction using cough-based features.

We will start this research in the first chapter by defining the covid and the severity of its general symptoms and the respiratory system in particular, which affects the voice of the patient.

Then, we will talk in the second chapter about the nature of sound and audio signals, their characteristics, and how to represent them digitally, giving a glimpse of the human voice.

The third chapter will be reserved for artificial intelligence and it is sweeping of all areas of life, including machine learning and deep learning that simulates human thought by means of neural networks, where we mentioned CNN and how it works.

Finally, we will present in the fourth chapter the practical side of our research and the stages involved in building a classifier model to predict COVID-19 using cough-based features using the CNN algorithm and then show and discuss the results.

Previous works

As soon as humanity was exposed to this pandemic, the largest companies and universities raced to develop applications and platforms to help detect Covid.

1. Cambridge University

Cambridge University³ provided a web-based platform and an android application to upload three coughs, five breaths, and three speech samples of reading a short sentence, and to report C19 symptoms & status. The crowd sourced data collected come from more than 10 different countries and comprises samples from 6 613 subjects with 235 C19 positive subjects. Note that in the work presented [1], only cough and breathing sounds are considered. With a manual examination of each sample, 141 cough and breathing items of 62 as C19 positive tested users and 298 items from 220 non-C19 users are used for building a binary classification model to distinguish between C19 and non-C19. [1]

2. Coughvid

Coughvid is another app from EPFL (Ecole PolytechniqueFédérale de Lausanne) to tell apart C19 cough from other cough categories such as normal cold and seasonal allergies. This data set has more than 20 000 cough samples [2], all the samples are passed through an open-source cough detection machine-learning model to identify the cough segments. [2]

3.VoiceMed7

VoiceMed7 is another android and web application that captures crowd-sourced speech and cough sounds and returns the C19 infection status on the fly [2]. The different stages in this cloud-based pre-trained CNN-based system comprise pre-processing the collected signal, using a cough detector to identify if it is a cough

signal, and then a C19 cough detector to further detect if the audio signal is a C19 cough. [2] .8 the authors used 900 coughs and 2 000 non-cough audio samples for building the cough detector. Similarly, the authors employed 165 C19 and 613 non-C19 samples for building the C19 cough detector. The accuracy of the cough classifier is reported to be 83.7% and the accuracy of the C19 classifier is reported to be 89.69% using deep spectrograms. [2]

1

Covid-19

I.1. Introduction

At the beginning of 2020, the world first learned about the Coronavirus. In the beginning, Corona was presented as a local or regional epidemic spreading in and around China, but it quickly crossed the borders and turned into a pandemic that threatens public health, and spreads all over the world.

Many countries have taken measures to limit mass contact with the aim of stopping the spread of the virus. The main measures taken to prevent its spread were, it is the temporary closure of places where people are heavy. However, that was not enough, to impose curfews and quarantine measures in most countries of the world, and restrictions were imposed on movement between cities and countries.

Global scientists have also intensified their research efforts on COVID-19 to improve capabilities for effective testing, prevention, control, and treatment of infections, especially since its symptoms, which do not appear until after a period of infection, as well as the danger of the doctors' presence with the injured, require the intervention of technology to reduce direct communication and the role of artificial intelligence in predicting the infection, despite the absence of symptoms in a person.

I.2. Definition

Coronaviruses are a family of viruses that can cause illnesses such as the common cold, severe acute respiratory syndrome (SARS), and Middle East respiratory syndrome (MERS). In 2019, a new type of coronavirus was discovered that caused an outbreak of a disease that originated in China [3].

The virus is known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The resulting disease is called novel coronavirus disease 2019 (COVID-19). In March 2020, the World Health Organization declared the Coronavirus (Covid-19) a global pandemic. [3]

I.3. Symptoms

The World Health Organization has come up with a list of symptoms that can be used to determine the presence of the COVID-19 virus in the human body. Breathing difficulties, [4] cough, muscle aches, chills, loss of taste, sore throat, and bloating. Body temperature is also a common symptom of infection with the virus it takes over an individual's immune system. [4] However, the uniqueness of the virus makes it difficult to ascertain the list of symptoms since the effects of the virus on individuals vary widely. Symptoms appear only after an incubation period of 6 to 7 days. During this time, the individual infected with the virus remains asymptomatic. [4]

I.4. The effect of Covid on the sound

The Coronavirus continues to reveal new symptoms as it spreads around the world, and one of the symptoms that we have recently known is the change in the voice infected with Corona, some COVID-19 patients have reported the voice changing to become "husky" or hoarse, and these symptoms are rooted in other consequences of the Corona virus. "Any upper respiratory infection will cause upper airway inflammation, and that includes the vocal cords," said Dr. Joseph Khabaza, MD, a pulmonologist and critical care physician at the Cleveland Clinic. Laryngitis can occur with any of these viruses. He added, "The Corona virus itself causes

infections, but many secondary symptoms are what exacerbates what is happening."

He continued, "When the corona infection affects the upper respiratory tract the lungs, the coughing will increase more, and you already have a sore throat and vocal cords from the infection, and then the secondary cough that occurs can be violent and more annoying. Specifically, coughing can cause inflammation of the larynx, which is the organ in the throat that contains the vocal cords, which move to allow breathing and vibrate to help you speak.

This inflammation affects the flexibility of those vocal cords, making them swollen and stiff. This means that they cannot vibrate as much. This can affect the tone and depth of your voice, making it sound coarse or even making your voiceless "whisper."

I.5. How does it spread?

COVID-19 is transmitted from person to person in close contact within 6 meters through respiratory droplets when an infected person speaks, sings, breathes, or coughs.[5] Other less common methods of spread are air transport, transmission through contaminated surfaces, and through people in poorly ventilated enclosed spaces. Since infected people do not initially have symptoms and the virus spreads from person to person, it is fatal and needs immediate effective methods to control its spread and treat its symptoms. These doubts and negative effects create the need for effective methods to identify individuals infected with MERS-CoV and to remain asymptomatic during the incubation period of the virus. [5]AI (Artificial Intelligence) has shown some positive signs in detecting asymptomatic individuals and helping isolate them from healthy individuals, thus helping to limit the spread of the virus. [5]

I.6. Artificial intelligence and detection of Covid from sound

Artificial intelligence and machine learning algorithms have helped save lives during the coronavirus pandemic thanks to their rapid ability to analyze health data as it arrives from around the world. [6]

The scientists also helped in analyzing the genetic information (DNA) of this virus very quickly, which enabled scientists to determine the characteristics of the virus. In addition, it did not stop there, but artificial intelligence helped scientists understand how quickly the virus could mutate. He also helped them develop vaccines against the Coronavirus and test the effectiveness of these vaccines. [6]

In addition, AI predicts results in advance if enough data is shown. This is thanks to smart devices and built-in sensors that protect even doctors from direct contact with patients to detect them, along with effective wireless communication technologies to achieve health monitoring at low costs, and thus data collection has become easier than before. More specifically, smartphones have been used to predict the onset of Covid symptoms and among the recently innovative methods is the detection of Covid 19 by audio signals. This is what we aim at in this research, using artificial intelligence techniques to screen, diagnose, monitor, and spread awareness about Covid. Symptoms of this virus directly affect the respiratory system and thus affect the patient's speech and sound signals such as shortness of breath, dry cough ... etc.

The evaluation of human audio signals has advantages: it is non-intrusive, easy to obtain, and both recording and evaluation can be performed almost instantaneously. The human voice signal provides sufficient 'marks' for C19, resulting in insufficient quality classification performance as C19 to be distinguished from other respiratory diseases and from typical subjects showing peculiarities in speech production. Many pieces of research and applications have appeared that work to detect and limit the spread of the virus by sound, and this is what we will discuss in this research. [6]

I.7. Conclusion

In light of this ongoing crisis, countries faced challenges, the most important of which were related to health care: first, how to screen people who have symptoms effectively and on time to avoid transmission of infection in crowded places; And second, how to ensure that patients receive prompt and appropriate treatment amid the rapid spread of the virus and in the face of limited medical resources.

To address these challenges, researchers have rapidly applied their expertise in artificial intelligence, technologies, and related products to support early efforts to prevent and control the epidemic. He used artificial intelligence to develop tests to detect the Coronavirus in a matter of weeks when it would have taken months without employing artificial intelligence.

Among these tests is how the infection with the virus changes the tone of those infected with it, to infer the possibility of infection with the Coronavirus through sound analysis, and if it matches the sound patterns of positive cases, then those infected with the Coronavirus generate their cough with a different echo.

Also, integrating this model with devices such as smartphones will be able to serve as a simple, effective, and free tool for detecting those infected with COVID-19 easily.

2

Audio signal

II.1 Introduction

Sound can be defined as a vibration that originates and is carried over material media (such as air) to move through it, and rarefactions (splitting air particles from each other). When a person speaks, the sound arises from the vibration of the vocal cords of the larynx.

The Audio Set ontology is a collection of sound events organized in a hierarchy. It covers a wide range of everyday sounds, from human and animal sounds to natural and environmental sounds, to musical sounds of things. Whereas human sounds are sounds produced by the human body such as respiratory sounds, are sounds generated by the movement of air through the respiratory system (nose, mouth, trachea, and lungs) like breathing, which is air being moved in and out of the lungs in order to sustain life? Cough is a sudden and often repetitively occurring reflex that consists of inhalation, a forced exhalation against a closed glottis, and a violent release of air following the opening of the glottis.

II.2 Sound wave

Sound is a mechanical vibration of a medium, a wave is one of the forms (propagation patterns) Sound waves move through the air by displacing air particles in a chain reaction. As one particle is displaced from its equilibrium position, it pushes or pulls on neighboring molecules, causing them to be displaced from their equilibrium.[7] As particles continue to displace one another with mechanical vibrations, the disturbance is transported throughout the medium. This particle-to-particle, mechanical vibrations of sound conductance qualify sound waves as mechanical waves. Sound energy, or energy associated with the vibrations created by a vibrating source, requires a medium to travel, which makes sound energy a mechanical wave.[7] the **FIG.II.1** is shown the Sound wave

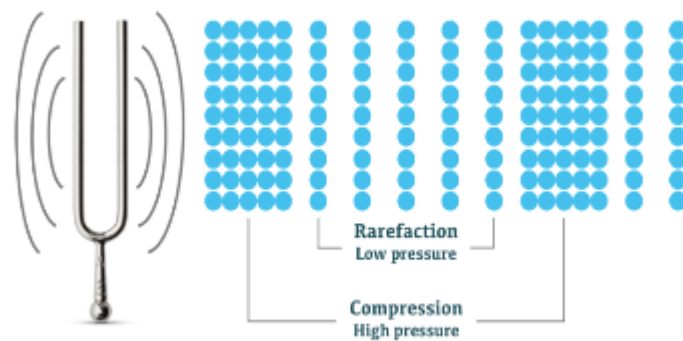


FIG.II.1 - Sound wave

II.2.1 audio signal

An **audio signal** is a representation of sound. It encodes all the necessary information required to reproduce sound in the audible frequency range of roughly 20 to 20,000 Hz (the limits of human hearing), **FIG.II.2** shows an example of this.

Audio signals may be synthesized directly or may originate at a transducer such as a microphone, musical instrument pickup, phonograph cartridge, or tape head. Loudspeakers or headphones convert an electrical audio signal back into a sound of roughly 20 to 20,000 Hz. [7]

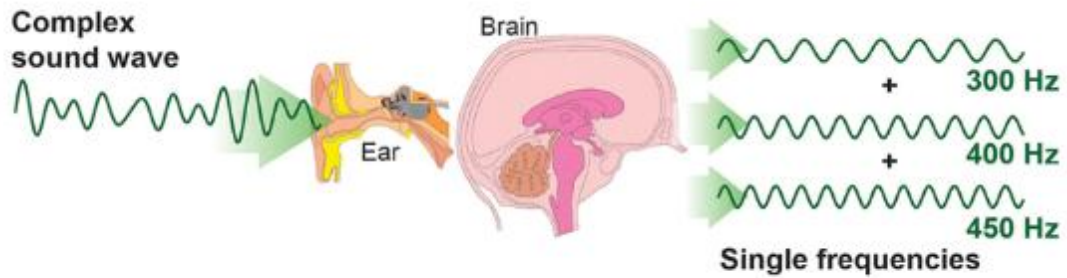


FIG.II.2 - Human auditory system

II.3 Types of audio signal

Analog refers to audio-recorded using methods that replicate the original sound waves. Digital audio is recorded by taking samples of the original sound wave at a specified rate, called the sampling rate. CDs and MP3 files are examples of digital formats. [8]

In the real world, conversions between digital and analog waveforms are common and necessary. ADC (Analog-to-Digital Converter) and the DAC (Digital-to-Analog Converter) are part of audio signal processing and they achieve these conversions (FIG.II.3). [8]

- **Analog Signal:** An analog signal is any continuous signal for which the time-varying feature of the signal is a representation of some other time-varying quantity i.e., analogous to another time-varying signal.[8]
- **Digital Signal:** A digital signal is a signal that represents data as a sequence of discrete values; at any given time, it can only take on one of a finite number of values.[8]

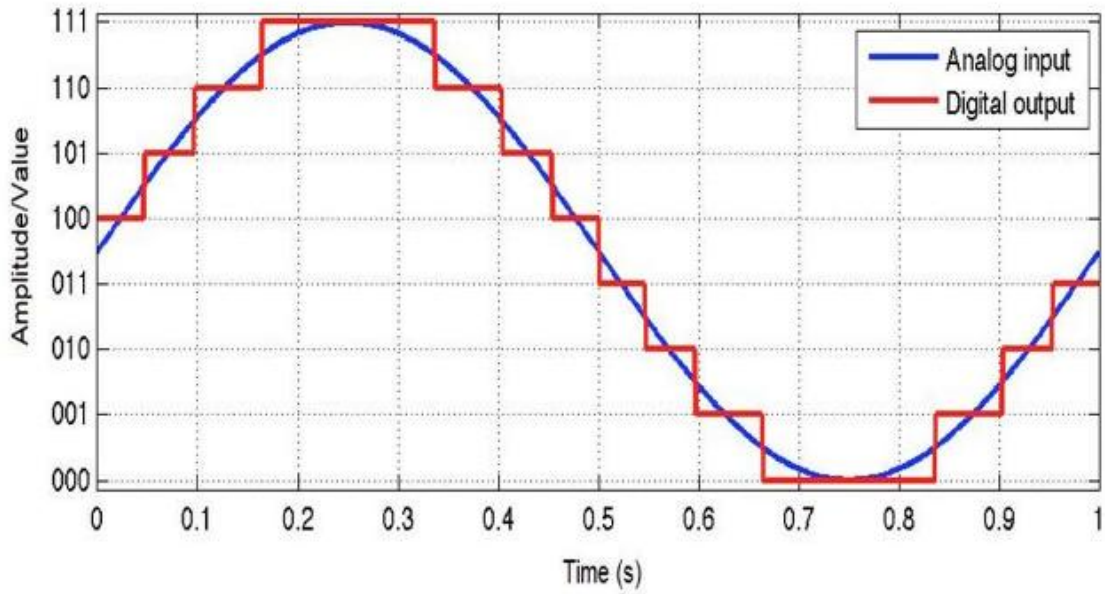


FIG.II.3 – Analog vs. Digital signals

II.4 Analog to Digital conversion

The most common technique to change an analog signal to digital data (FIG.II.4) is called pulse code modulation (PCM). A PCM encoder has the following three processes: [8]

- *Sampling*
- *Quantization*
- *Encoding*

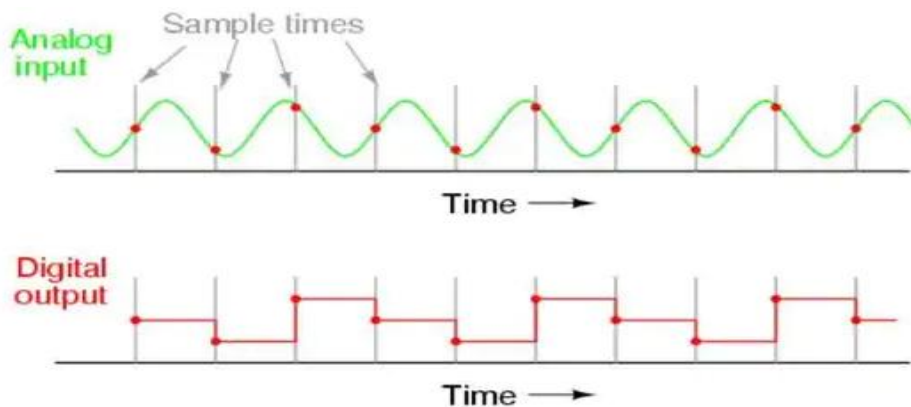


FIG.II.4 – Converting an analog signal to a digital

II.4.1 Sampling

The first step in PCM is sampling. Sampling is a process of measuring the amplitude of a continuous-time signal at discrete instants, converting the continuous signal into a discrete signal.

One important consideration is the sampling rate or frequency. According to the *Nyquist theorem*, the sampling rate must be at least 2 times the highest frequency contained in the signal. It is also known as the minimum sampling rate and is given by:

$$F_s \geq 2 * f_h$$

It is possible to apply a low pass filter to eliminate the high-frequency components present in the input analog signal to ensure that the input signal is sampled is free from the unwanted frequency components. This is done to avoid the aliasing of the message signal. There are three sampling methods:

II.4.1.1 Ideal Sampling

In ideal sampling also known as Instantaneous, sampling pulses from the analog signal are sampled. This is an ideal sampling method and cannot be easily implemented.

II.4.1.2 Natural Sampling

Natural Sampling is a practical method of sampling in which a pulse have finite width equal to T. The result is a sequence of samples that retain the shape of the analog signal.

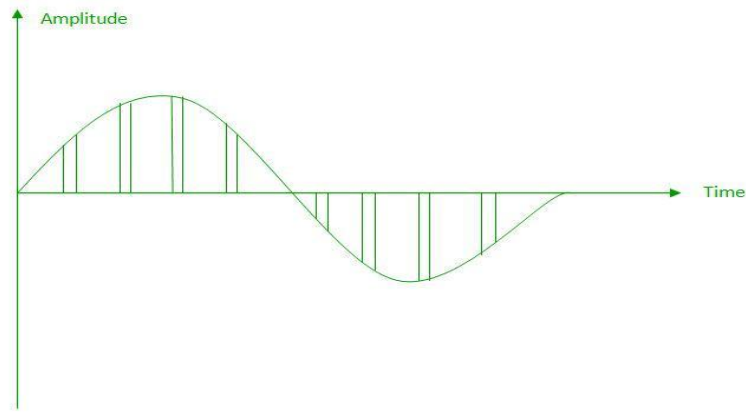


FIG.II.5 - Natural sampling

II.4.1.3 Flat-top sampling

In comparison to natural sampling flattop, sampling can be easily obtained. In this sampling technique, the top of the samples remains constant by using a circuit. This is the most common sampling method used

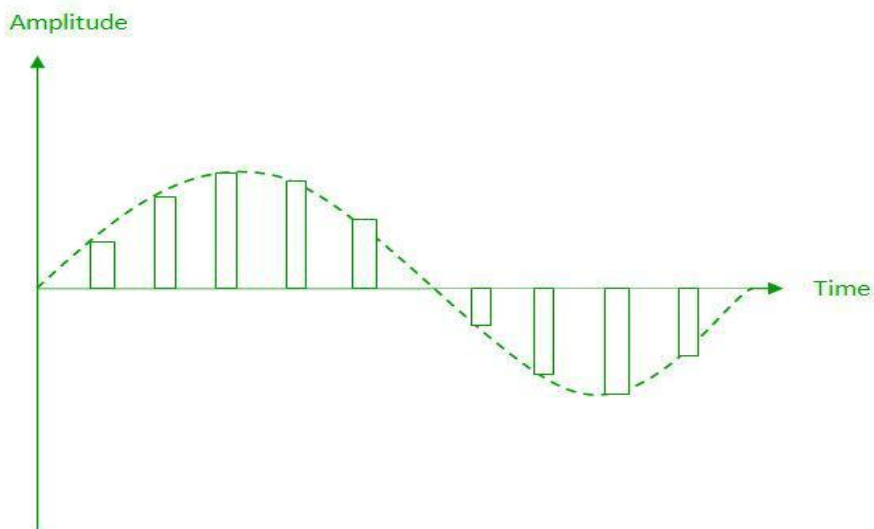


FIG.II.6 - Flat top sampling

II.4.2 Quantization

The result of sampling is a series of pulses with amplitude values between the maximum and minimum amplitudes of the signal. The set of amplitudes can be infinite with non-integral values between two limits.

The following are the steps in Quantization:

- We assume that the signal has amplitudes between V_{max} and V_{min}
- We divide it into L zones each of height d where

$$d = (V_{max} - V_{min}) / L$$

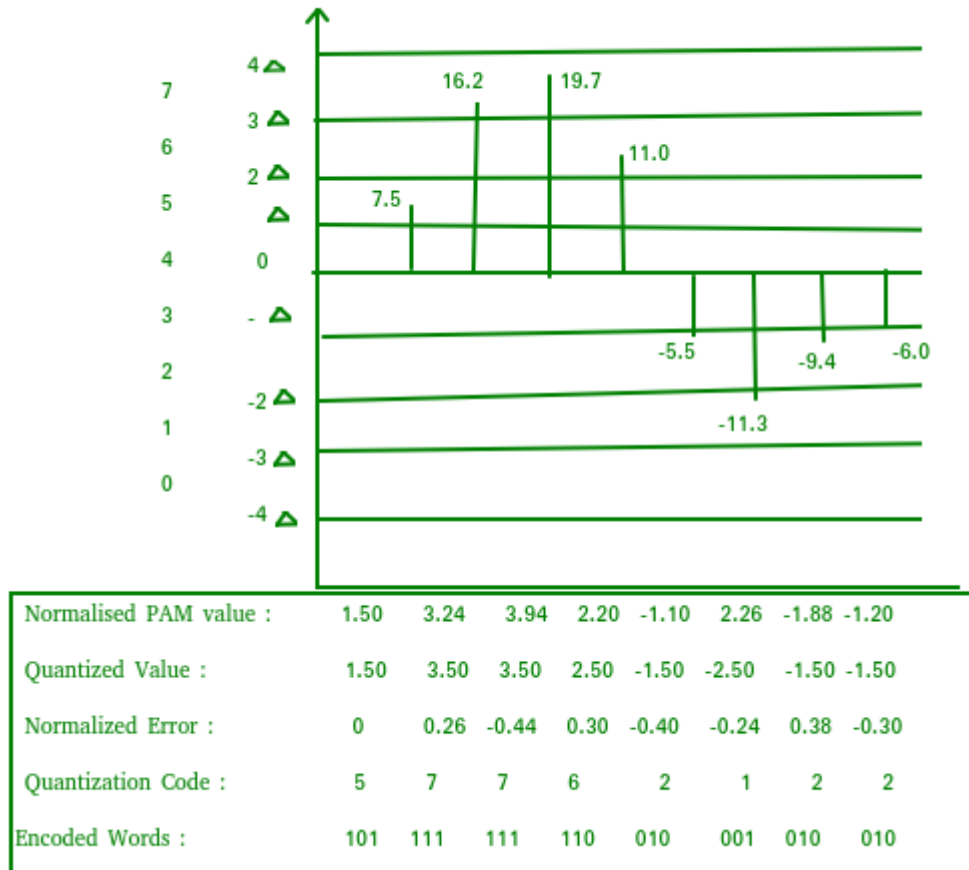


FIG.II.7 - Quantization process

- The value at the top of each sample in the graph shows the actual amplitude.
- The normalized pulse amplitude modulation (PAM) value is calculated using the formula amplitude/d.
- After this, we calculate the quantized value that the process selects from the middle of each zone.
- The Quantized error is given by the difference between the quantized value and normalized PAM value.

- The Quantization code for each sample is based on quantization levels at the left of the graph.

II.4.3 Encoding

The digitization of the analog signal is done by the encoder. After each sample is quantized and the number of bits per sample is decided, each sample can be changed to an n-bit code. Encoding also minimizes the bandwidth used.

II.5 audio features

There are, in general, two types of audio features: the physical features and the perceptual features. Physical features refer to mathematical measurements computed directly from the sound wave, such as the energy function, the spectrum, the cepstral coefficients, the fundamental frequency, and so on. Perceptual features are subjective terms that are related to the perception of sounds by human beings, including loudness, brightness, pitch, timbre, rhythm, etc. [9]

II.5.1 Perceptual features

Perceptual features are those that can be perceived by humans, they are also known as paralinguistic features as they deal with the elements of speech that are properties of large units such as syllables, words, phrases, and sentences. Since they are extracted from these large units, they are long-term features. The most widely used prosodic features are based on fundamental frequency, energy and duration.

II.5.1.1 Frequency (Pitch)

Pitch is the quality that enables us to judge sounds as being “higher” and “lower. It provides a method for organizing sounds based on a frequency-based scale. Pitch can be interpreted as the sound term for frequency, though they are not exactly the same. A high-pitched sound causes molecules to rapidly oscillate, while a low-pitched sound causes slower oscillation. Pitch can only be determined when a sound has a frequency that is clear and consistent enough to differentiate it from noise. Because pitch is primarily based on a listener’s perception, it is not an objective physical property of sound.

II.5.1.2 Amplitude (Dynamics)

The amplitude of a sound wave determines its relative loudness. In music, the loudness of a note is called its dynamic level. In physics, we measure the amplitude of sound waves in decibels (dB), which do not correspond with dynamic levels. Higher amplitudes correspond with louder sounds, while shorter amplitudes correspond with quieter sounds. Despite this, studies have shown that humans perceive sounds at very low and very high frequencies to be softer than sounds in the middle frequencies, even when they have the same amplitude.

II.5.1.3 Duration (Tempo/Rhythm)

In physics, the duration of a sound begins once the sound registers and ends after it cannot be detected. Duration is the amount of time that a cough, it can be described as long, short, or as taking some amount of time. The duration of a cough influences the rhythm of a sound.

II.5.2 Physical features

When sound is produced by a person, it is filtered by the shape of the vocal tract. The sound that comes out is determined by this shape. An accurately simulated shape may result in an accurate representation of the vocal tract and the sound produced. Characteristics of the vocal tract are well represented in the frequency domain .

Spectral features are obtained by transforming the time domain signal into the frequency domain signal using the Fourier transform. They are extracted from speech segments of length 20 to 30 milliseconds that are partitioned by a windowing method.

II.5.2.1 Mel Frequency Cepstral Coefficients (MFCC)

MFCCs have been largely employed in the speech recognition field but also in the field of audio classification, due to the fact that their computation is based on a perceptual-based frequency scale in the first stage (the human auditory model which is inspired on the frequency Mel Scale).[9]

II.5.2.2 Line Linear Prediction Cepstral Coefficients (LPCC)

LPCC can be directly obtained with a recursive method from Linear Prediction Coefficient (LPC). LPC is basically the coefficients of all-pole filters and is equivalent to the smoothed envelope of the log spectrum of the speech (Wong and Sridharan, 2001). Another feature, Log-Frequency Power Coefficients (LFPC), mimics logarithmic filtering characteristics of the human auditory system by measuring spectral band energies using Fast Fourier Transform (Nwe et al., 2003a).ar Prediction Cepstral Coefficients (LPCC).

II.5.2.3 Gammatone Frequency Cepstral Coefficients (GFCC)

Is also a spectral feature obtained by a similar technique of MFCC extraction. Instead of applying a Mel filter bank to the power spectrum, a Gammatone filter-bank is applied. Formants are the frequencies of the acoustic resonance of the vocal tract. They are computed as amplitude peaks in the frequency spectrum of the sound. They determine the phonetic quality of a vowel, hence used for vowel recognition n.

II.5.2.4 Teager energy operator-based features

There are features that depend on the Teager Energy Operator (TEO). It is used to detect stress in speech and has been introduced by Teager and Teager (1990) and Kaiser (1990, 1993). According to Teager, speech is formed by a non-linear vortex-airflow interaction in the human vocal system. A stressful situation affects the muscle tension of the speaker that results in an alteration of the airflow during the production of the sound.

II.6 Conclusion

The conversion cough into a signal is considered the first step in detection covid 19. This step follows signal processing and eliminating noise. The difference in coughing results in a difference in the audio features of the signal, although, humans are aware of only some very clear changes in them, unlike artificial intelligence, which showed the ability to detect changes in audio features with

high accuracy. This is what we discuss in the next chapter of the role of artificial intelligence in detecting Covid 19 by cough.

3

Artificial Intelligence for sound processing

III.1 Introduction

Artificial intelligence technologies have become very popular in the past decade for several reasons, the most important of which are: the power and enormous capabilities of modern computers, which made it possible to implement very complex algorithms that were previously impossible to solve, as well as the widespread use of sensors connected to the Internet and data transmission in a way that was previously impossible. It means that artificial intelligence is changing faster, and its predictions are becoming more accurate. In the following, we will discuss what artificial intelligence is.

III.2 Artificial Intelligence

Artificial intelligence (AI) is a wide-ranging branch of computer science concerned with building smart machines capable of --performing tasks that typically require human intelligence (It is the endeavor to replicate or simulate human intelligence processes by machines).[10] The expansive goal of artificial intelligence has given rise to many questions and debates. So much so, that no singular definition of the field is universally accepted, so we offer definitions of the largest pioneers in this field:

First, in 1955 the computer scientist John McCarthy, coined the term artificial intelligence, or AI. His pioneering work in AI – which he defined as "the science and engineering of making intelligent machines". [10]

Second, Patrick Winston, the Ford professor of artificial intelligence and computer science at MIT, defines AI as «algorithms enabled by constraints, exposed by representations that support models targeted at loops that tie thinking, perception and action together."

Lastly, Russel and Norvig defined AI as "the study of agents that receive precepts from the environment and perform actions." [10]

III.3 Types of Artificial Intelligence

We have four types of artificial intelligence shown in theFIG.III.1

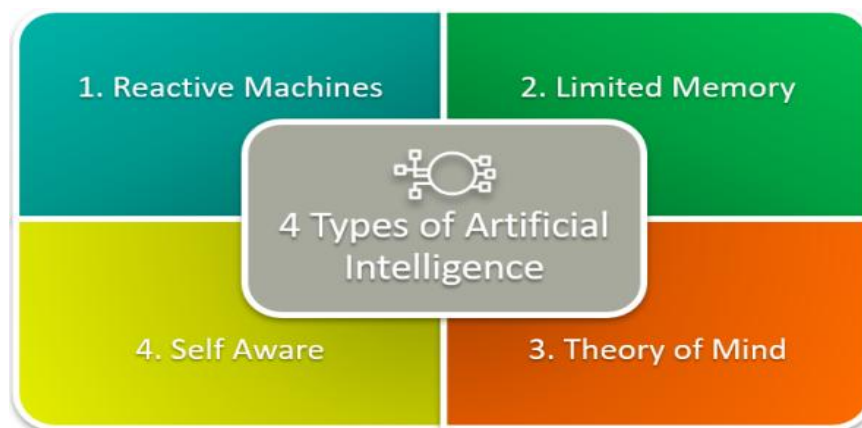


FIG.III.1 - Types of Artificial Intelligence

III.3.1 Reactive Machines

Reactive machines are simple in that they do not store memories' or use previous actions to predict future actions. They simply observe and react to the environment around them. Deep Blue, IBM's chess grandmaster defeating machine, is a reactive machine that observes and reacts to the pieces on a chessboard. It cannot draw on any of its previous experiences, and it cannot get better with practice.

III.3.2 Limited Memory

Machines with limited memory can only store data for a short length of time. They can use this information for a limited time, but they cannot save it to a library of their experiences. Many self-driving cars have Limited Memory technology, which allows them to store data such as the speed of adjacent automobiles, their distance from them, the speed limit, and other information that can assist them in navigating roads.

III.3.3 Theory of Mind

Their thoughts, emotions, memories, and mental models, according to psychology, guide people's behavior. Researchers working on Theory of Mind seek to create computers that mimic our mental models by constructing representations of the world and other agents and entities in it. One of these researchers' goals is to create computers that can interact with humans and understand human intelligence as well as how events and the environment affect people's emotions. While many computers employ models, there is no such thing as a computer with a "mind".

III.3.4 Self-Awareness

Though many AI enthusiasts believe that self-aware machines are the ultimate objective of AI development, they are the stuff of science fiction. Even if a machine can function similarly to a person, such as conserving itself, anticipating its own needs and desires, and relating to people on an equal footing, the question of whether it can become really self-aware, or 'conscious,' is best left to philosophers.

III.4 Machine Learning

The branch of artificial intelligence involved with making computers work without being explicitly programmed is known as machine learning. where systems may "learn" from data, statistics, and trial and error in order to improve processes and innovate more quickly. Machine learning allows computers to learn in the same way that humans do, allowing them to solve some of the world's most difficult issues.

For example, voice recognition is an example of tacit knowledge. We can know a person's voice, but it's difficult to explain how or why we recognize it. We rely on our personal knowledge banks to connect the dots instinctively in order to recognize someone based on their speech.

III.4.1 How to work machine learning algorithms?

It is a set of programs developed in general and with general rules for processing the entered data in all forms and finding relationships and patterns in the data by applying statistical and mathematical equations - where each algorithm has certain characteristics and outputs. So, it can represent data in different ways or predict new data outputs - based on relationships and patterns Inferred from the input data. There are a lot of machine learning algorithms and we will discuss some of them, shown in the FIG.III.2

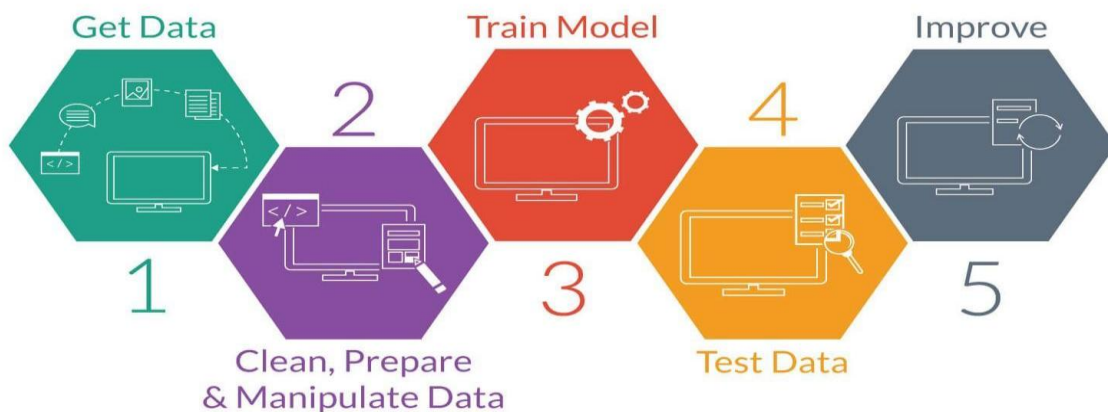


FIG.III.2 - How to work machine learning algorithms

III.4.2 Machine Learning applications

Machine learning algorithms are used in complex problems that are difficult to describe in logic, such as "voice recognition" applications.[11] It is also used with problems whose data is constantly changing, such as systems for predicting the trend of sales of goods. Due to the wide range of use of machine learning algorithms, their applications are endless. Here are some of these applications [11]:

III.4.2.1 Healthcare

Disease identification and prediction have become significantly more accurate and quicker thanks to machine learning. Radiology and pathology departments all over the world are currently using machine learning to analyze CT and X-RAY scans to detect disease.

Can detect and diagnose diseases (such as cancer or viruses) faster than humans. Machine learning has also been used to anticipate deadly diseases such as Covid, Ebola, and Malaria, and the CDC uses it to track flu cases each year.

III.4.2.2 self-driving car

The companies aim to build cars that can drive safely without the need for a driver. The manufacture of these cars relies primarily on machine learning algorithms.

III.4.2.3 Simultaneous translation

Simultaneous translation is based on machine learning algorithms, since its speed and flexibility of translating texts from one language to another smoothly requires fast processing and an efficient model, which cannot be done with traditional algorithms.

III.4.2.4 Recommendation engines

When you search for a movie or a product, you will find advertisements for similar products the next day, or while browsing a social media site, you will find recommendations for movies similar to the movie you searched for before. This is what recommendation engines do. It mainly uses machine-learning algorithms.

III.4.3 Machine learning types:

There are 3 main types, which are shown in the FIG.III.3

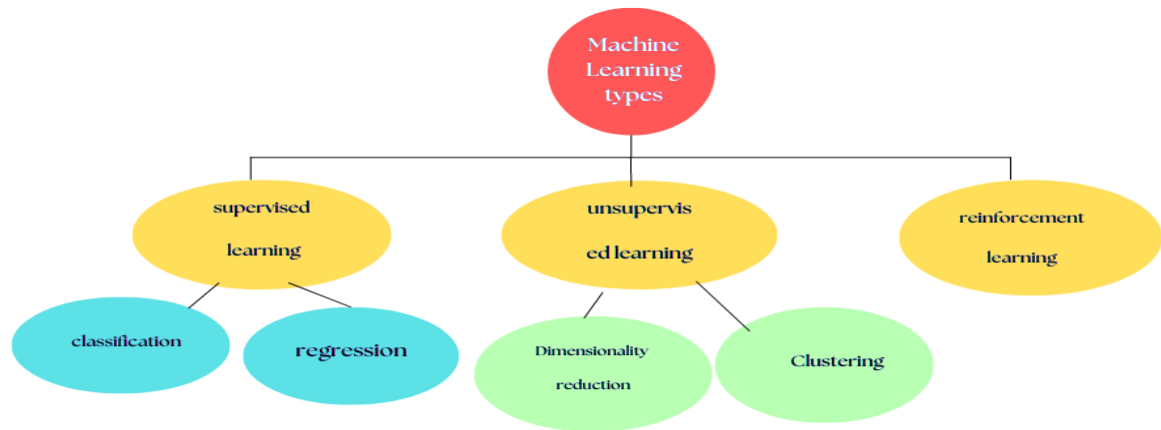


FIG.III.3 - Machine Learning types

III.4.3.1 Supervised learning

This approach is based on machine learning from the training data classified as "inputs and outputs", so it is called supervision, where we supervise the learning by giving the outputs to the data. Through training, the computer builds relationships and patterns between data and outputs, and a model is reached that can predict new data outputs.

The main goal of supervised learning is to scale up data and make predictions based on disaggregated sample data concerning unavailable, future, or unknown data. Supervised machine learning includes two major processes: classification and regression:

Classification is the process of learning from past data samples and manually training the model to predict the essentially binary outcomes (yes/no, true/false, 0/1). For example: whether someone has covid or not etc. The classification algorithm recognizes certain types of objects and categorizes them accordingly to predict one of the two possible outcomes. like Support Vector Machines (SVM)

regression is the process of identifying patterns and calculating the predictions of continuous outcomes. For example: predicting the house rates or next month's sales forecast like Support Vector Regression (SVR).

III.4.3.2 Unsupervised learning

This methodology is used when it is difficult to pre-categorize the entered data. The machine learns from the unsorted data set and categorizes it based on the discovery of similarities and internal differences in the data. Therefore, it is called unsupervised learning, that is, not subject to disaggregated data (Describes data supplied to it by sifting through it and making sense of it).

Unsupervised learning algorithms apply the following techniques to describe the data:

- Clustering is the process of segmenting data into meaningful groups (i.e., clusters) based on internal patterns without having any prior knowledge of group credentials. Individual data objects' similarity as well as features of their dissimilarity from the rest define the credentials (which can also be used to detect anomalies) like K-means clustering.
- Dimensionality reduction: In most cases, the entering data contains a lot of noise. Dimensionality reduction is used by machine learning algorithms to reduce noise while distilling the relevant data. like PCA (Principal Component Analysis).

III.4.3.3 Reinforcement learning

Reinforcement learning is about developing a self-sustaining system that, through contiguous sequences of attempts and failures, improves itself based on the named dataset and interactions with incoming data specific to how the programmer (machine) makes the decision (choice) in an environment in order to maximize overall reward. Reinforcement learning differs from supervised learning in that it does not require any pairs of inputs and outputs, but instead focuses on direct performance that improves cumulative reward.

III.5 Deep Learning

It is a subset of machine learning and artificial intelligence (AI). It is the discipline concerned with the study of "artificial neural networks" that simulate neural networks in the human brain, which are shown in the FIG.III.4. [17]

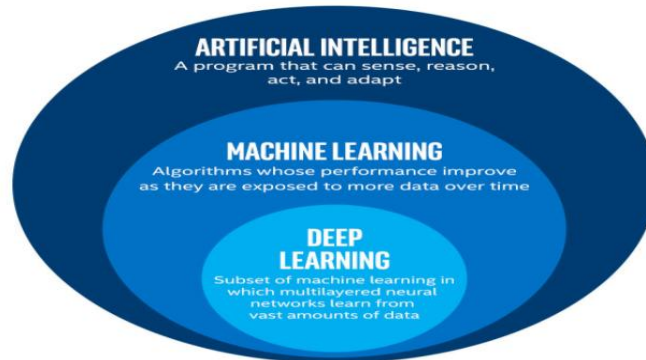


FIG.III.4 - Deep Learning

III.5.1 Artificial neural network

An "artificial neural network" is defined as a piece of computing system designed to simulate the analysis and processing process in the human brain. The basic processing unit in the human brain is the neuron, and the artificial neuron in the machine corresponds to it. [17] An assembly of artificial neurons is known as an artificial neural network, which are shown in the FIG.III.5

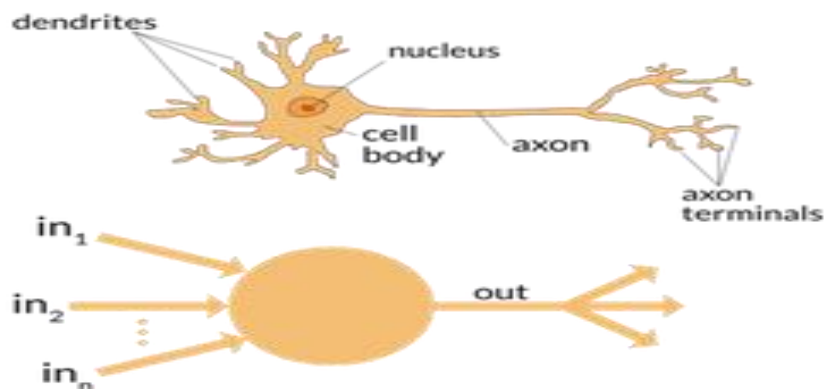


FIG.III.5 - Artificial neural network

III.5.2 Structure of deep learning algorithm

Deep neural networks consist of multiple layers of interconnected nodes. Where the layer on the left end is the input layer, the layer on the right is the output layer, and in the middle are several hidden layers responsible for processing. The architecture of stratified deep learning algorithms enables better data handling and better performance. Which are shown in the **FIG.III.6**

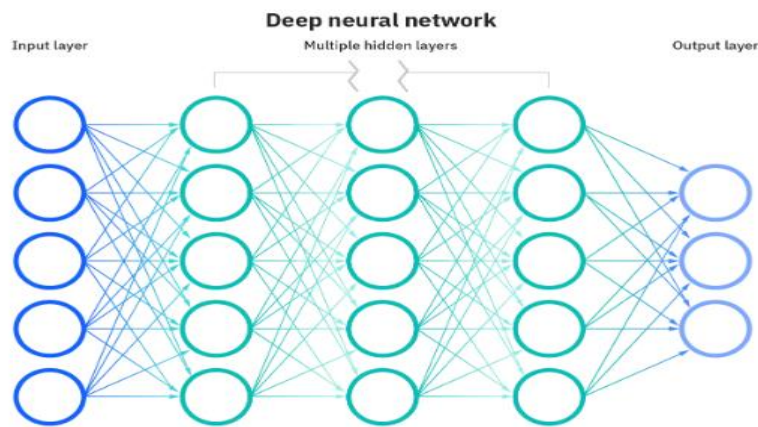


FIG.III.6 - Structure of deep learning algorithm

Each layer builds upon the previous layer to refine and optimize the prediction or categorization. This progression of computations through the network is called forward propagation. The input and output layers of a deep neural network are called visible layers. The input layer is where the deep learning model ingests the data for processing, and the output layer is where the final prediction or classification is made.

Another process called back propagation uses algorithms, like gradient descent, to calculate errors in predictions and then adjusts the weights and biases of the function by moving backwards through the layers in an effort to train the model. Together, forward propagation and back propagation allow a neural network to make predictions and correct for any errors accordingly. Over time, the algorithm becomes gradually more accurate.

III.5.3 Types of Deep learning

The above describes the simplest type of deep neural network in the simplest terms. However, deep learning algorithms are incredibly complex, and there are different types of neural networks to address specific problems or datasets as CNN ,RNN [12]

- Convolutional neural networks (CNNs): which are commonly used in computer vision and image classification applications, can recognize characteristics and patterns within an image, allowing tasks such as object detection and recognition to be accomplished. For the first time in 2015 a CNN bested a human in an object recognition challenge for the first time.[12]
- Recurrent neural networks (RNNs): are typically used in natural language and speech recognition applications as it leverages sequential or times series data.[12]

III.6 Deep learning vs. machine learning

The discipline of deep learning emerged as an extension and evolution of machine learning when traditional machine learning algorithms were unable to perform some complex tasks. Traditional machine learning algorithms require a simplified and tidy set of data to learn from. It generally goes through some pre-processing to get it organized into a structured format. However, it is not capable of learning from large and complex data sets, such as different sound waves, image dimensions, and the number of pixels within them.

So, it uses deep learning algorithms to handle complex data like this. It automates feature extraction, removing some of the reliance on human experts. As in the "voice recognition" applications that Siri uses, it helps Google recognize the voices of its speakers and "image recognition" applications. Image recognition, used by Facebook to recognize people's faces in photos. The difference between them is shown in FIG.III.7

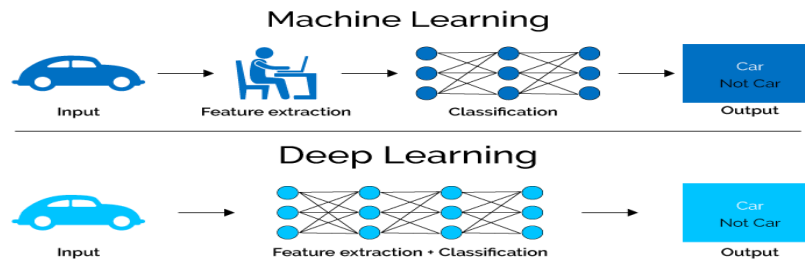


FIG.III.7 - Deep learning vs. machine learning

III.7 Convolutional neural networks (CNNs)

Is a class of deep learning network (DNN) which is widely used for computer vision or NLP. During the training process, the network's building blocks are repeatedly altered in order for the network to reach optimal performance and to classify images and objects as accurately as possible. Convolutional neural networks excel at learning the spatial structure in input data.

III.7.1 Architecture of CNN

There are two main parts to a CNN architecture, a convolution tool that separates and identifies the various features of the image for analysis in a process called as feature extraction, and a fully connected layer that utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages. Which are shown in the FIG.III.8 [13]

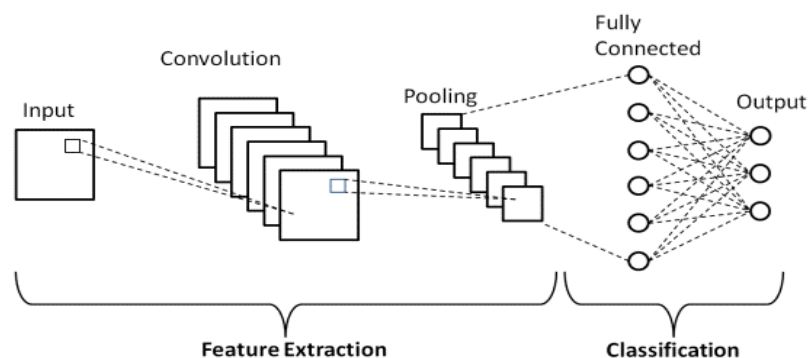


FIG.III.8 -Architecture of CNN

III.7.1.1 Convolution tool

There are three types of layers that make up the CNN which are the convolutional layers, pooling layers, and fully connected (FC) layers. When these layers are stacked, a CNN architecture will be formed. In addition to these three layers, there are two more important parameters, which are the dropout layer and the activation function, which are defined below. [13]

a. Convolutional Layer

This layer is the first layer that is used to extract the various features from the input images. In this layer, the mathematical operation of convolution is performed between the input image and a filter of a particular size $M \times M$. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image with respect to the size of the filter. The output is termed as the feature map which gives us information about the image such as the corners and edges. Later, this feature map is fed to other layers to learn several other features of the input image. [13]

b. Pooling Layer

In most cases, a Convolutional Layer is followed by a Pooling Layer. The primary aim of this layer is to decrease the size of the convolved feature map to reduce the computational costs. This is performed by decreasing the connections between layers and independently operates on each feature map. Depending upon the method used, there are several types of Pooling operations. In Max Pooling, the largest element is taken from the feature map. Average Pooling calculates the average of the elements in a predefined sized Image section. The total sum of the elements in the predefined section is computed in Sum Pooling. The Pooling Layer usually serves as a bridge between the Convolutional Layer and the FC Layer.[13]

III.7.1.2 Fully Connected Layer

The Fully Connected (FC) layer consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. These layers are usually placed before the output layer and form the last few layers of a CNN Architecture. In this, the input image from the previous layers is flattened and fed to the FC layer. The flattened vector then undergoes few more FC layers where the mathematical functions operations usually take place. In this stage, the classification process begins to take place. [13]

a. Dropout

Usually, when all the features are connected to the FC layer, it can cause over fitting in the training dataset. Overfitting occurs when a particular model works so well on the training data causing a negative impact in the model's performance when used on new data. To overcome this problem, a dropout layer is utilized wherein a few neurons are dropped from the neural network during training process resulting in reduced size of the model. On passing a dropout of 0.3, 30% of the nodes are dropped out randomly from the neural network. [13]

b. Activation Functions

Finally, one of the most important parameters of the CNN model is the activation function. They are used to learn and approximate any kind of continuous and complex relationship between variables of the network. In simple words, it decides which information of the model should fire in the forward direction and which ones should not at the end of the network. It adds non-linearity to the network. There are several commonly used activation functions such as the ReLU, Softmax, tanH and the Sigmoid functions. Each of these functions have a specific usage. For a binary classification CNN model, sigmoid and softmax functions are preferred and for a multi-class classification, generally softmax is used. [13]

III.8 Conclusion

After we got to know artificial intelligence in all its branches and its capabilities to solve the most difficult daily problems faster and more accurately. Especially in the health field, "aforementioned," and this is what we targeted in our research through the diagnosis and detection of COVID-19 based on coughing.

4

Materials, Methods & results

IV.1 Introduction

researchers have found that people who are asymptomatic of Covid-19 may differ from healthy individuals in the way that they cough. These differences are not decipherable to the human ear. But it turns out that they can be picked up by artificial intelligence. From an auditory inspection, we can see that it is tricky to auditory the difference between some of the classes. Particularly, the forms of repetitive cough sound for the injured and the healthy are similar. The following will demonstrate how to apply Deep Learning techniques to the classification of human cough sounds, specifically focusing on the identification of whether the person has covid 19 or not.

IV.2 Implementation

We have used Google Colab GPU (Tesla K80 12GB GDDR5 VRAM), Python 3.7 and TensorFlow 2.2.0. For the implementation of CNN, the deep learning library of TensorFlow 2.2.0 is used, and the training and the testing procedures are done in the Google Colab platform.

IV.3 Audio Data acquisition

Health practitioners use a variety of sensors to monitor and extract information about a disease's symptoms for individuals. Different sensor choices were investigated by Drugman , including contact and non-contact microphones, electrocardiography sensors, chest straps, accelerometers, and thermistors put over the patient's nose. Drugman et al discovered that a single non-contact microphone provides substantially more information about the cough than other sensors while looking for the optimal sensor to extract information about coughs.

This has the added benefit of making it usable by anyone with a smartphone as the phone's microphone can serve as the sensor.

IV.3.1 Collecting Data

This part is about the specific dataset and exploratory data analyses in our methodology section. With the onset of COVID-19, the need for rapid analysis and testing of COVID-affected patients has been at the forefront of health research. Massachusetts Institute of Technology (MIT) has created a web interface for the general public to upload their voice samples and answer a questionnaire. This helps self-label the recorded audio sample into COVID-19 and non-COVID-19 categories. The Indian Institute of Science (IISc) also maintains a repository of COVID-19 and non-COVID-19 voices called "*Coswara*".

In this work, we explore "*Coswara*" datasets only Because MIT datasets are not available. These datasets capture human audio in multiple acoustic forms like breathing, cough, alphabets, and vowel pronunciation. For the COVID detection task, we mainly focus on using cough audio data. show that respiratory ailments with cough as one of its symptoms have distinct underlying features, and can be extracted using appropriate signal processing techniques like MFCC.

This Dataset contains the sound signals of Covid-19 positive and negative patients, which can be used for classification. The Dataset is created from the samples collected from "*Coswara*" and "*Virufy*" which are highly reliable. There are 1349

coughs which are of people who tested 786 negatives and 563 coughs are of COVID-19 positive people. we download the ZIP file with .wav from “kaggle”. [14] Then we create a CSV file containing the features extracted from these cough sounds which are negative or positive, or directly import the data to Google Colab.

IV.4 Method

This study demonstrates the feasibility of an alternative form of COVID-19 detection, harnessing digital technology through the use of acoustic biomarkers and deep learning. Specifically, we show that a deep neural network-based model can be trained to detect symptomatic and asymptomatic COVID-19 cases using audio recordings of coughing. Figure IV.1 shows the steps for building the proposed model

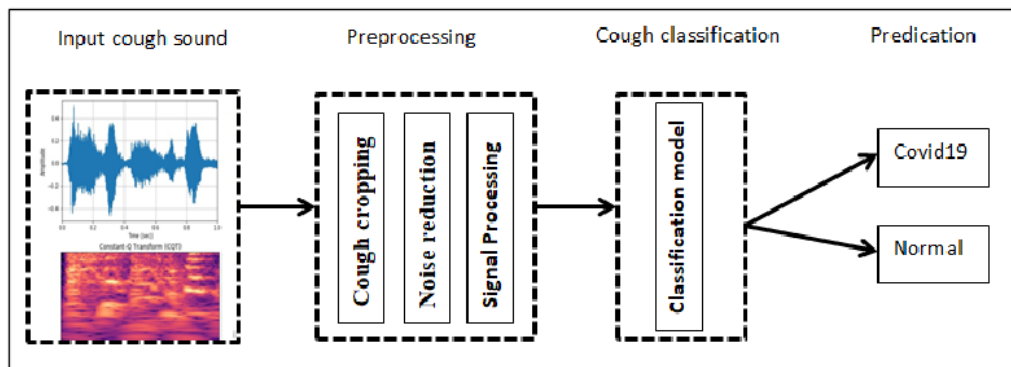


FIG.IV.1 - The steps for building the proposed model

IV.4.1. Exploratory Data

A visual assessment reveals that visualizing the differences between the classes is difficult because the waveforms' sounds (positive and negative) are similar in shape. Waveforms of both a healthy and COVID-19 sick individual coughing are visible in Figure IV.2.

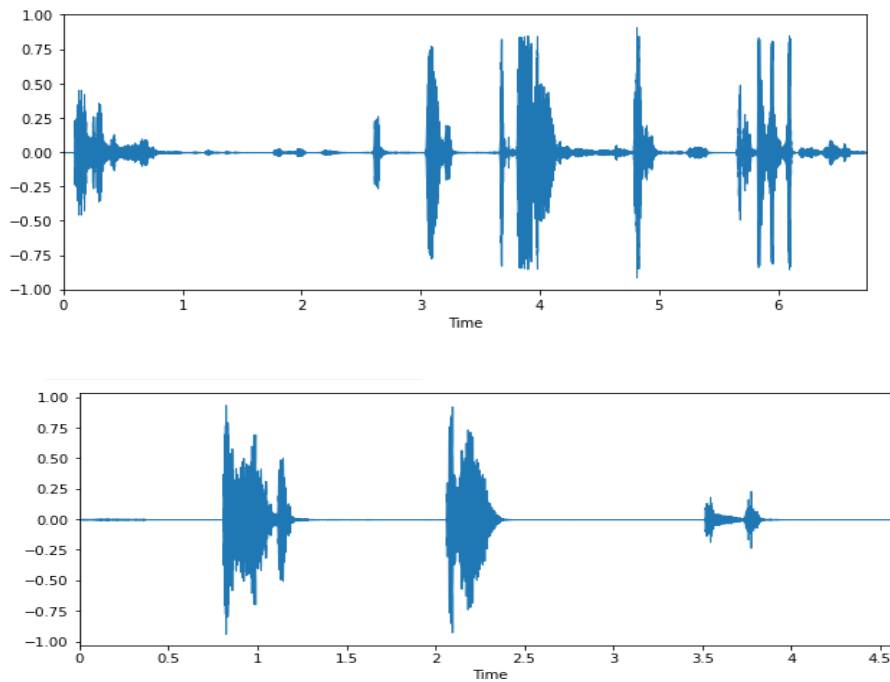


FIG.IV.2 – Waveform visualization of an audio sample positive (first) and sample negative (second)

Next, we will do a deeper dive into each of the audio file properties extracting, the number of audio channels, sample rate, and bit-depth using the following code.

IV.4.2 preprocessing audio

We identified the following audio properties that need preprocessing to ensure consistency across the whole dataset: [15]

- **Audio Channels:** Most of the samples have two audio channels (meaning stereo) with a few with just one channel (mono). The easiest option here to make them uniform was to merge the two channels in the stereo samples into one by averaging the values of the two channels.[15]
- **Sample rate:** We applied a sample rate transformation technique (up-transform or down-conversion) so that we could see a neutral representation of their waveform which allowed us to make a fair comparison[15]
- **Bit-depth:** There is also a wide range of bit depths. We needed to normalize them by taking the maximum and minimum amplitude values for a given bit depth.[15]

Python program Librosa for music and audio processing allows us to input audio as a NumPy array for analysis and editing.

We used Librosa's *load()* method for a lot of the preprocessing, which transforms the sampling rate to 22.05 kHz, normalizes the data so the bit-depth values range between -1 and 1, and flattens the audio channels into mono by default.

IV.4.3 Feature extraction and feature selection:

Feature extraction is an important step toward classifying cough sounds. However, generating a classification model from a high-dimensional dataset takes time and may converge to a local minimum due to the large search space. Therefore, selecting a miniature set of relevant features in an audio sample can significantly improve performance by creating a classification model. There are several techniques for feature selection, where we use Spectral Roll-off Point (SRP), Spectral Centroid (SC), ZCR Crossing Rate (ZCR), and Mel-Frequency Cepstral Coefficients, However, the MFCC method has gained popularity due to its efficiency in analyzing speech and sound signals in general, and thus was chosen in the analysis of cough sounds :

- **Zero-Crossing Rate (ZCR):** The ZCR is the most common type of zero-crossing-based audio feature. It is defined as the number of time-domain zero crossings within a processing frame and indicates the frequency of signal amplitude sign change. ZCR allows for a rough estimation of dominant frequency and spectral centroid.
- **Spectral Roll-off Point (SRP):** The spectral roll-off point is the N% percentile of the power spectral distribution, where N is usually 85% or 95%.The spectral roll-off point is the frequency below which N% of the magnitude distribution is concentrated. It increases with the bandwidth of a signal.
- **Spectral Centroid (SC):** represents the “balancing point”, or the midpoint of the spectral power distribution of a signal. It is related to the brightness of a sound. The higher the centroid, the brighter (high frequency) the sound is. A

spectral centroid provides a noise-robust estimate of how the dominant frequency of the signal changes over time.

- **Mel Frequency Cepstral Coefficients (MFCC):** feature represents the short-term power spectrum of the speech signal. To obtain MFCC, utterances are divided into segments, then each segment is converted into the frequency domain using a short-time discrete Fourier transform. A number of sub-band energies are calculated using a Mel filter bank. Then, the logarithm of those sub-bands is calculated. Finally, the inverse Fourier transform is applied to obtain MFCC. It is the most widely used spectral feature. As summarized in Figure IV.3 [16].

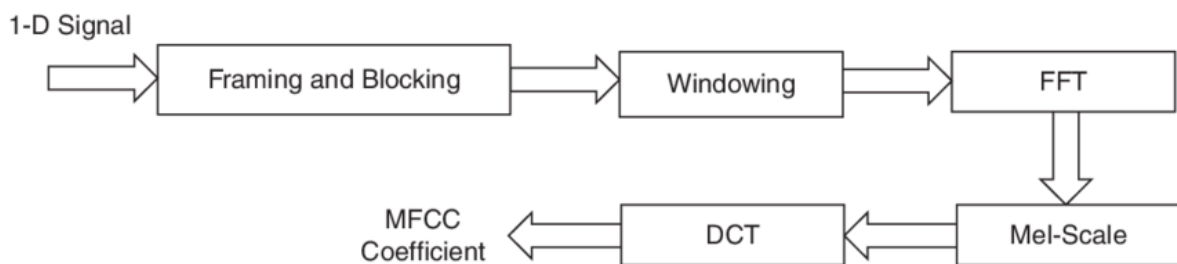


FIG.IV.3 - Steps of MFCC generation.

IV.4.4 Model architecture

We start with constructing a MultiLayer Perceptron (MLP) Neural Network using *Keras* and a *Tensorflow* backend. Starting with a sequential model so we can build the model layer by layer.

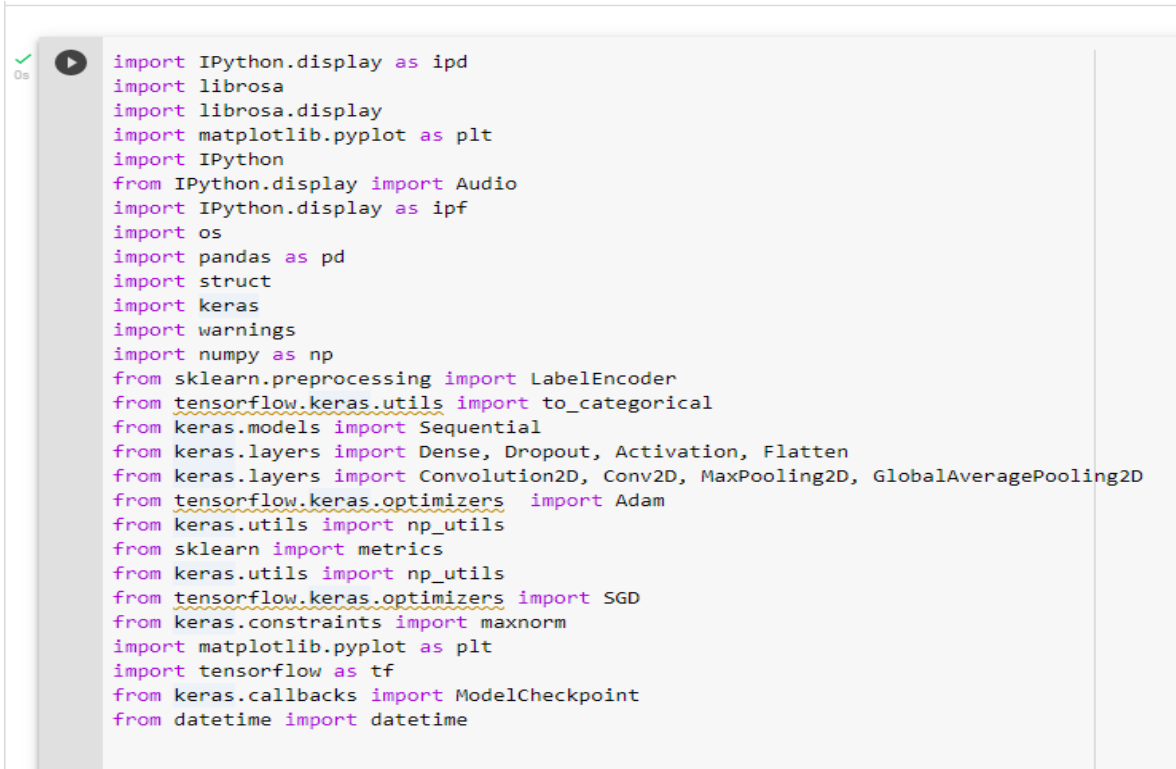
We begin with a simple model architecture, consisting of three layers, an input layer, a hidden layer, and an output layer. All three layers will be of the dense layer type which is a standard layer type that is used in many cases for neural networks. The first layer will receive the input shape. As each sample contains 43 features (or columns) we have a shape of (1x43) this means we will start with an input shape of 43. The first two layers will have 550 nodes. The activation function we are using for our first 2 layers is the *ReLU* or Rectified Linear Activation. This activation function has been proven to work well in neural networks.

We will also apply a Dropout value of 50% on our first two layers. This will randomly exclude nodes from each update cycle which in turn results in a network that is capable of better generalization and is less likely to overfit the training data.

Our output layer will have 2 nodes (num_labels) that match the number of possible classifications. The activation for our output layer is softmax. Softmax makes the output sum up to 1 so the output can be interpreted as probabilities. The model will then make its prediction based on which option has the highest probability.

IV.5 Code Explaining

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



```
import IPython.display as ipd
import librosa
import librosa.display
import matplotlib.pyplot as plt
import IPython
from IPython.display import Audio
import IPython.display as ipf
import os
import pandas as pd
import struct
import keras
import warnings
import numpy as np
from sklearn.preprocessing import LabelEncoder
from tensorflow.keras.utils import to_categorical
from keras.models import Sequential
from keras.layers import Dense, Dropout, Activation, Flatten
from keras.layers import Convolution2D, Conv2D, MaxPooling2D, GlobalAveragePooling2D
from tensorflow.keras.optimizers import Adam
from keras.utils import np_utils
from sklearn import metrics
from keras.utils import np_utils
from tensorflow.keras.optimizers import SGD
from keras.constraints import maxnorm
import matplotlib.pyplot as plt
import tensorflow as tf
from keras.callbacks import ModelCheckpoint
from datetime import datetime
```

FIG.IV.4 – Importing library

Here, we load the database and put it in a variable as shown in the following figure:

```
[3] from google.colab import drive
drive.mount('/content/drive')
!ls ~/content/drive/My Drive

Mounted at /content/drive
/bin/bash: -c: line 0: unexpected EOF while looking for matching `"'
/bin/bash: -c: line 1: syntax error: unexpected end of file
```

```
metadata = pd.read_csv('/content/drive/MyDrive/cleaned_data/COVID_data.csv')
metadata.head(len(metadata))
```

index	slice_file_name	classe_name
775	557_Negative_male_79.wav	Negative
776	54_Negative_male_34.wav	Negative
777	585_Negative_female_53.wav	Negative
778	556_Negative_male_36.wav	Negative
779	584_Negative_male_44.wav	Negative
780	671_Negative_male_30.wav	Negative
781	802_Negative_female_27.wav	Negative
782	844_Negative_male_29.wav	Negative
783	874_Negative_male_37.wav	Negative
784	953_Negative_male_49.wav	Negative
785	983_Negative_male_50.wav	Negative
786	605_Positivo_47_m_8-6-20_1.wav	Positive
787	35_Positivo_50_f_17-6-20_1.wav	Positive
788	609_Positivo_39_m_16-6-20_1.wav	Positive
789	67_Positivo_34_m_21-6-20_2.wav	Positive

FIG.IV.5 – Loading our dataset

Next, preprocessing stage mentioned in section FIG.IV.4.2 is performed as shown in FIG.IV.6.

```
audiodef = pd.DataFrame(audiodata, columns=['num_channels', 'sample_rate', 'bit_depth'])
audiodef.head()
```

	num_channels	sample_rate	bit_depth
0	1	48000	16
1	1	48000	16
2	1	48000	16
3	1	48000	16
4	1	48000	16

FIG.IV.6 – Preprocessing stage

Then, extracting the features of all the sounds in the database, as shown in Figure.IV.7.

```
[11] def extract_features(file_name):
    audio, sample_rate = librosa.load(file_name, res_type='kaiser_fast')
    #mfccs = librosa.feature.mfcc(y=audio, sr=sample_rate, n_mfcc=40)
    zcr = ZCR(c=audio, sr=sample_rate)
    sepec_cent = Sepec_cent(c=audio, sr=sample_rate)
    rolloff = Rolloff(c=audio, sr=sample_rate)
    #spectral_bandwidth(c,sr)
    #chroma_stf(c,sr)
    mfcc = MFCC(c=audio, sr=sample_rate)

    feature_matrix=np.array([])
    # use np.hstack to stack our feature arrays horizontally to create a feature matrix
    feature_matrix = np.hstack((zcr,sepec_cent,rolloff,mfcc))
    return feature_matrix
```

FIG.IV.7 – Extract the features

Now, we build our model by:

- **Convert the data and labels**

We use **sklearn.preprocessing.LabelEncoder** to encode the categorical text data into model-understandable numerical data.

- **Split the dataset**

We use **sklearn.model_selection.train_test_split** to split the dataset into training and testing sets. The testing set size will be 20% and we will set a random state. Then, we develop our model mentioned in section IV.4.4 as shown in figure IV.8.

```
num_labels = yy.shape[1]
filter_size = 2

model = Sequential()
#first layer
model.add(Dense(550, input_shape=(43,)))
model.add(Activation('relu'))
model.add(Dropout(0.5))
#second layer
model.add(Dense(550))
model.add(Activation('relu'))
model.add(Dropout(0.5))

model.add(Dense(num_labels))
model.add(Activation('softmax'))
model.summary()
```

FIG.IV.8 –: model architecture

IV.6 Evaluation and Results

IV.6.1 Evaluation data:

Contrary to the popular opinion that analyzing secondary data is easy or quick, same methodologies are often used in primary research and rely on existing data that may not be ideal. To make an informed decision about data usage, we can go through the following points to improve our database:

- Data collection methods.
- Data file format.
- Data set documentation including variable names and descriptions.
- Data quality including reliability and validity.
- Extent of missing data.

IV.6.2 Evaluation method:

We use the following metrics for assessing the proposed method:

- **Accuracy:** is defined as the degree to which the result of a measurement conforms to the correct value or a standard and essentially refers to how close a measurement is to its agreed value. And it too Proportions correct classifications from the overall number of cases.

$$\text{Accuracy} = \frac{\text{Correct prediction}}{\text{total correct prediction}}$$

Where we got the result of 90% shown in **FIG.IV.9**.

```
# Evaluating the model on the training and testing set
score = model.evaluate(x_train, y_train, verbose=0)
print("Training Accuracy: ", score[1])
```

```
score = model.evaluate(x_test, y_test , verbose=0)
print("Testing Accuracy: ", score[1])
```

Training Accuracy: 0.9064815044403076

Testing Accuracy: 0.855555534362793

FIG.IV.9 - Percentage of Accuracy

- **Loss:** is the result of a bad forecast. The loss is a number that indicates how poorly the model predicted in one example. In Figure IV.10, there are two curves, the first is the curve as a percentage of the loss in training and test. When increasing epochs we notice a decrease in the percentage of loss and the second curve has a percentage of accuracy in training and test when increasing epochs we notice the increase in the percentage of accuracy in both

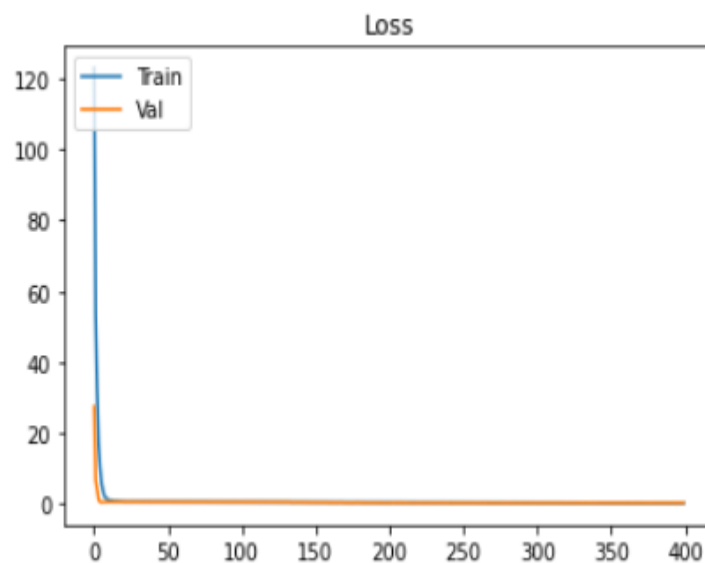


FIG.IV.10 - loss of the model

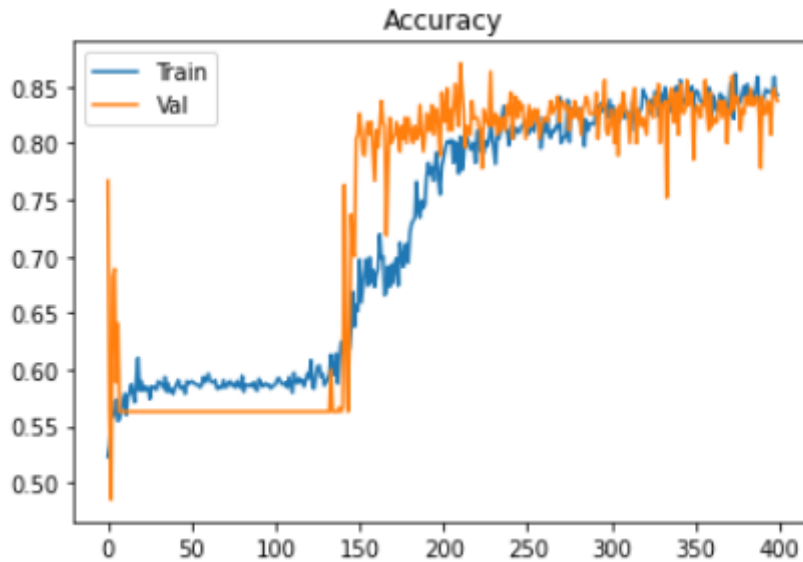


FIG.IV.11 – Accuracy of the model

IV.6.3 confusion matrix

A confusion matrix is a table that is used to define the performance of a classification algorithm. A confusion matrix visualizes and summarizes the performance of a classification algorithm. A confusion matrix is shown in Table IV.1.

Predicted value		
Actual value	Predicate positive	Predicated negative
Actual positive	TP	FN
Actual negative	FP	TN

Table.IV.1 – Structure of confusion matrix.

The confusion matrix consists of four basic characteristics (numbers) that are used to define the measurement metrics of the classifier. These four numbers are :

1. **TP (True Positive):** TP represents the number of patients who have been properly classified to have COVID19.

2. **TN (True Negative):** TN represents the number of correctly classified patients who are healthy.
3. **FP (False Positive):** FP represents the number of misclassified patients with the disease but actually are healthy. FP is also known as a Type I error.
4. **FN (False Negative):** FN represents the number of patients misclassified as healthy but actually have COVID19. FN is also known as a Type II error.

Performance metrics of an algorithm are *accuracy, precision, recall, and F1 score*, which are calculated on the basis of the above-stated TP, TN, FP, and FN as shown in **Table IV.2**.

- **The accuracy of an algorithm** is represented as the ratio of correctly classified patients (TP+TN) to the total number of patients (TP+TN+FP+FN).[17]
- **The precision of an algorithm** is represented as the ratio of correctly classified patients with the disease (TP) to the total patients predicted to have the disease (TP+FP).
- **Recall metric** is defined as the ratio of correctly classified diseased patients (TP) divided by the total number of patients who have actually the disease.
- **The perception** behind the recall is how many patients have been classified as having the disease. The recall is also called sensitivity.
- **The F1 score** is also known as the F Measure. The F1 score states the equilibrium between the precision and the recall.

Precision	Recall	f1-score	Accuracy
0.894	0.73	0.84	0.84

Table IV.2 – Performance metrics

	precision	recall	f1-score	support
COVID+	0.82	0.93	0.87	152
COVID-	0.90	0.74	0.81	118
accuracy			0.85	270
macro avg	0.86	0.84	0.84	270
weighted avg	0.85	0.85	0.85	270

FIG.IV.12- Performance metrics

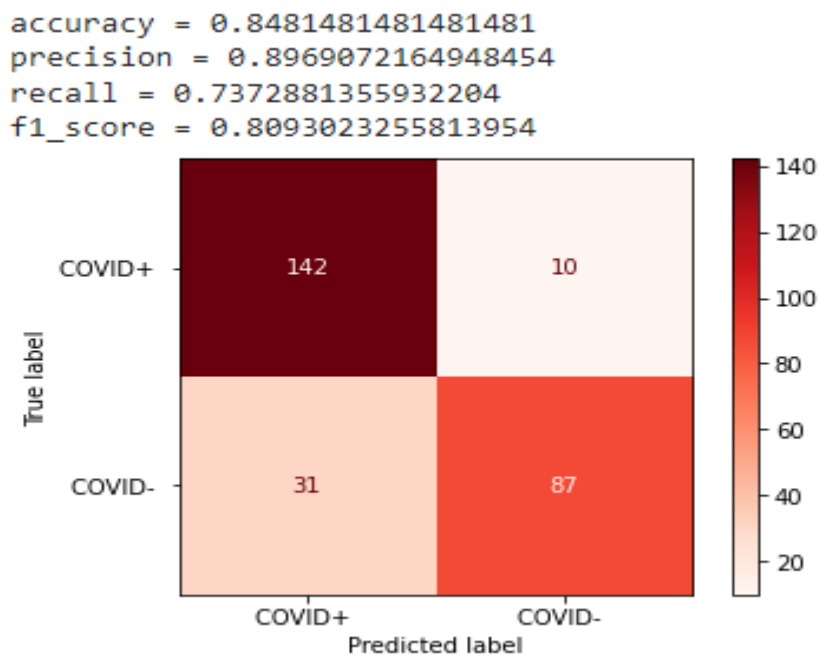


FIG.IV.13 - Confusion matrix

IV.6.4 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 the adventure and try in this field, despite the difficulties we faced due to the lack of a database and the lack of information on the disease, which is unstable and is still developing.

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

After training the model several times and each time we enhanced the value until we reached an acceptable result, which is 90%.

IV.6 Conclusion

We used deep learning and machine learning and used many experiments on a set of models, and after a period of training and testing, we came to an acceptable result for detecting COVID-19 through the cough of a sick person. This project reduces the proportion of people who will come to the hospital, through a person who can test himself while he is at home by processing his coughs through our model. If the prediction rate is above the threshold, he will be called to Hospital, this is to complete treatment.

Conclusion & perspectives

Despite the changes in Covid-19 and the diseases that are emerging from it, the world is still talking due to its rapid spread and to alleviate this spread. In this study, we focused on the auditory effects that Covid-19 symptoms can cause on individuals where changes can be discovered by analyzing audio recordings for coughing. The individual is still late to process the sound compared to the processing of images, but this does not mean superiority in the results, because there are sound studies that have good results. The focus in this study was on the detection of Covid-19 through coughing, with the aim of examining and early diagnosis and limiting its spread.

In this research, we learned about Covid-19 and the severity of its symptoms on the respiratory system, especially, which affects the voice of the injured. Then, we touched on the treatment of the audio signal and extracting each of the following features: Zero-crossing rate. Special Roll-FF Point (SRP) .Spectral Center (SC) and MFCC, which is very popular in the field of sound processing, of which we used 40 features.

In the last, we got 43 features for each sound; we used it as an entrance for CNN to train the classification model, where we got a 90 % training rate that is considered insufficient in the field in which we work, as we found it difficult to find reliable data for the study's work, and also hindered the weakness of the Internet on the outcome of the model.

In future studies, we aim to improve the result to become more accurate and generalize the study on all the symptoms of the respiratory system than speech, breathing, and to include health history. Working to improve sound features using new technologies more powerful for classification.

Bibliography

- [1] Orlandic, L., Teijeiro, T. and Atienza, D., 2021. The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. *Scientific Data*, 8(1).
- [2] Deshpande, G., Batliner, A. and Schuller, B., 2022. AI-Based human audio processing for COVID-19: A comprehensive overview. *Pattern Recognition*, 122, p.108289.
- [3] MayoClinic.org. 2022. مرض فيروس كورونا المستجد 2019 (كوفيد-19) - الأعراض والأسباب - Mayo Clinic (مايو كلينك). [online] Available at: <<https://www.mayoclinic.org/ar/diseases-conditions/coronavirus/symptoms-causes/syc-20479963>> [Accessed 12 June 2022].
- [4] Health. 2022. Could Your Hoarse Voice Be a Symptom of COVID? Here's What Experts Want You to Know. [online] Available at: <<https://www.health.com/condition/infectious-diseases/coronavirus/is-hoarseness-symptom-of-covid>> [Accessed 12 June 2022].
- [5] N, S. (May 10, 2021). Deep Learning Anomaly Detection methods to passively detect COVID-19 from Audio. *A thesis submitted in partial fulfilment for the degree of Master of Science in Data Science*.
- [6] Yang, F., Wu, Q., Hu, X., Ye, J., Yang, Y., Rao, H., ... & Hu, B. (2021). Internet-of-Things-Enabled Data Fusion Method for Sleep Healthcare Applications. *IEEE Internet of Things Journal*, 8(21), 15892-15905.

- [7] 2022. [online] Available at:
<<https://www.pasco.com/products/guides/sound-waves>> [Accessed 12 June 2022].
- [8] Ambardar, A. (1995). *Analog and digital signal processing* (p. 700). BOSTON, MA: PWS.
- [9] Akcay, m. b., & oguz, k. (January 2020). *Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers* (Vol. 116).
- [10] Builtin.com. 2022. *What is Artificial Intelligence? How Does AI Work? | Built In*. [online] Available at: <<https://builtin.com/artificial-intelligence>> [Accessed 12 June 2022].
- [11] Builtin.com. 2022. *What is Artificial Intelligence? How Does AI Work? | Built In*. [online] Available at: <<https://builtin.com/artificial-intelligence>> [Accessed 12 June 2022].
- [12] Education, I., 2022. *What is Deep Learning?*. [online] Ibm.com. Available at: <<https://www.ibm.com/cloud/learn/deep-learning>> [Accessed 12 June 2022].
- [13] upGrad blog. 2022. *Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network | upGrad blog*. [online] Available at: <<https://www.upgrad.com/blog/basic-cnn-architecture/>> [Accessed 12 June 2022].
- [14] Kaggle.com. 2022. *Covid 19 cough sounds*. [online] Available at: <<https://www.kaggle.com/datasets/pranaynandan63/covid-19-cough->

sounds?fbclid=IwAR0iCpcZKjAKsyIy2Sn5ZY2EcG3fWYidH386pSTZ30yO
KtL6]k9obG1XtjQ> [Accessed 12 June 2022].

- [15] GitHub. 2022. *Udacity-ML-Capstone/Report.pdf at master · mikesmales/Udacity-ML-Capstone*. [online] Available at: <<https://github.com/mikesmales/Udacity-ML-Capstone/blob/master/Report/Report.pdf>> [Accessed 12 June 2022].
- [16] Réda, A., & Aoued, B. B. (n.d.). Artificial Neural Network & Mel-Frequency Cepstrum Coefficients-Based Speaker Recognition.
- [17] موقع الأكاديمية بوست. 2022. ما هو التعلم العميق؟ - موقع الأكاديمية بوست [online] Available at: <<https://elakademiapost.com/%D9%85%D8%A7-%D9%87%D9%88-%D8%A7%D9%84%D8%AA%D8%B9%D9%84%D9%85-%D8%A7%D9%84%D8%B9%D9%85%D9%8A%D9%82/>> [Accessed 12