COVID-19 DETECTION THROUGH COUGH RECORDINGS USING ARTIFICIAL INTELLIGENCE

[1]Hadia Nowshad, [2]Hiba Naushad, [3]Jumana Hasin, [4]Wafa Ahamed Sharafudheen, Nikhil Dharman
[1,2,3,4,5]Computer Science and Engineering, College of Engineering Thalassery,
Kerala, India
[1]hadianow@gmail.com, [2]hiba.n789@gmail.com, [3]jumanahasin0@gmail.com,
[4]wafahmed01@gmail.com, [5]nikhildharmanmk@gmail.com

Abstract—The Covid-19 pandemic represents one of the great- est global health emergencies of the last few decades with indelible consequences for all societies throughout the world. The cost in terms of human lives lost is devastating on account of the high contagiousness and mortality rate of the virus. Millions of people have been infected, frequently requiring continuous assistance and monitoring. Smart healthcare technologies and Artificial Intelligence algorithms constitute promising solutions useful not only for the monitoring of patient care but also in order to support the early diagnosis, prevention and evaluation of Covid-19 in a faster and more accurate way. On the other hand, the necessity to realise reliable and precise smart healthcare solutions, able to acquire and process voice signals by means of appropriate Internet of Things devices in real-time, requires the identification of algorithms able to discriminate accurately between pathological and healthy subjects. Several studies report, in fact, significant effects of this virus on voice production due to the considerable impairment of the respiratory apparatus. COVID-19 subjects, especially including asymptomatics, could be accurately discriminated from a forced-cough cell phone record- ing using Artificial Intelligence. Cough recordings selected from the Coswara database, an available crowd-sourced database, have been analysed and processed using ML model to distinguish between healthy and pathological voices. This was used to pre- screen for COVID-19 from cough recordings, longitudinally monitor patients in real-time, non-invasively, and at essentially zero variable cost.

I. INTRODUCTION

The emergence of coronavirus has been considered a major threat to public health in almost all countries around the world during the last year. Millions of lives have been and are cur- rently being disrupted by this pandemic. Strict social measures in combination with existing tests and consequently dramatic economic costs, have proven sufficient to significantly reduce pandemic numbers, but not to the extent of extinguishing the virus. Timely treatment for many patients is needed, as well as early diagnosis and monitoring, with healthcare workers aiming to utilize the limited resources available most effec- tively. On account of its high infection rate, the development

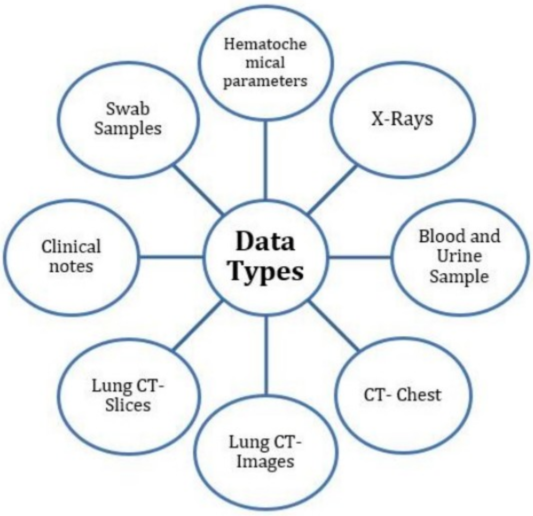
of techniques able to identify the presence of Covid-19 and distinguish it from other forms of influenza and pneumonia, in a fast and reliable way, is crucial. Although, in fact, the World Health Organisation (WHO) currently recommends the diagnosis of Covid-19 using molecular tests in laboratories, the tracking of the virus globally and to diagnosis of the pathology at an early stage could considerably benefit from this solution. It would be ideal for an easy, portable, non-invasive and low- cost mass screening phase since the analysis of the cough recordings can be acquired through a mobile device, such as a smartphone or tablet. The unlimited real-time diagnostic of our tool could help intelligently test individuals and thereby categorizing them as healthy or pathological based on their test results.

II. THEORETICAL BASIS

A. COVID-19 Data

The essential source for the analysis is the variety of datasets accessible in the clinical centers, hospitals, and test labs. These datasets are open-source and are freely accessible for research purposes. Many specialists or research groups have published data and articles on Covid-19 which are useful for the treatment of COVID-19. Swab tests, X-rays, blood [4] and urine tests [5], and other physiological data including CT scans are considered for the dataset Different datasets for COVID-

19. The datasets are partitioned into two sections. i.e., medical images and textual data. The medical images are gathered from X-rays and CT-Scans containing Chest and Lung scans. However, the text data is from social media and case studies. Different works of literature investigated the COVID-19 based on these datasets. The global reporting studies are helpful for the research field to comprehend the seriousness that continues with the Coronavirus. In the proposed system, the vocal aspects of the individual is taken into consideration.



B. Detection of an infection

Asymptomatic people who are infected with Covid-19 ex- hibit, by definition, no discernible physical symptoms of the disease. They are thus less likely to seek out testing for the virus, and could unknowingly spread the infection to others. But it seems those who are asymptomatic may not be entirely free of changes wrought by the virus. MIT researchers have now found that people who are asymptomatic 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. In a paper published recently in the IEEE Journal of Engineering in Medicine and Biology, the team reports on an AI model that distinguishes asymptomatic people from healthy individuals through forced-cough recordings, which people voluntarily submitted through web browsers and devices such as cell- phones and laptops.

C. Vocal sentiments

Prior to the pandemic's onset, research groups already had been training algorithms on cellphone recordings of coughs to accurately diagnose conditions such as pneumonia and asthma. They first trained a general machine-learning algorithm, or neural network, known as ResNet50, to discriminate sounds associated with different degrees of vocal cord strength.

III. METHODOLOGY

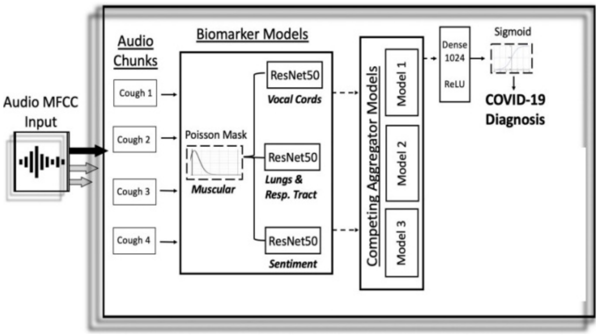
A. Database

Coswara is an available crowd-sourced database accessible on the GitHub platform, realised by the Indian Institute of Science (IISc) Bangalore. The aim of the project, named Coswara, is to collect sound samples to provide a database useful for an evaluation of the reliability of technologies to support the diagnosis of Covid-19. Coughing, breathing and voice sounds were collected from each subject, in addition to data relating to health status, gender, age, certain pre-existing health conditions (i.e. asthma, diabetes) and geographical location. The samples were recorded in all continents, except Africa, with a prevalence of sounds coming from Asia.

All the recordings were sampled at 44.1 KHz and their resolution is 32-bit. An opportune filter was applied to reduce the noise added during the acquisition. Voice sounds partic- ularly corrupted by noise and of a duration of less than one second were excluded. The subject forced-cough audios and diagnostic results were used to train and validate the COVID- 19 discriminator.

B. Data Processing

All the datasets are preprocessed to remove noise, sampling, normalization, etc. Processed data or images are applied to the segmentation processor feature extraction, which leads to testing and training using machine learning and deep learning algorithms, such as, neural networks, classifiers, etc. Later, the classification and performance metrics of all models/ algorithms can be calculated to depict the quality of results.

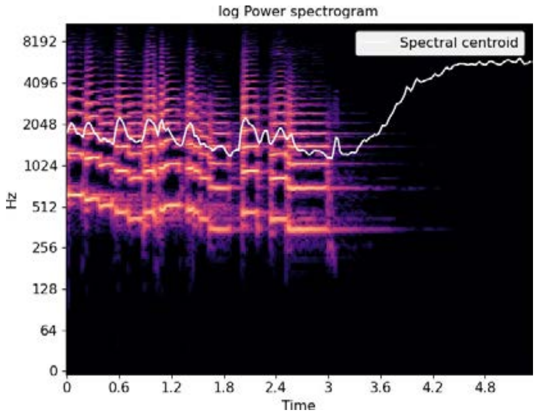


C. Feature Extraction

The choice of the appropriate features that can represent the data in a relevant way is fundamental in that this can have a significant impact on the reliability of the system. Therefore, we evaluated not only the main parameters used in clinical practice to assess voice quality, but also other well-known parameters used in literature for the voice classification, such as Mel-Frequency Cepstral Coefficients (MFCC) or Spectral Centroid or RollOff. The feature extraction stage involves ex- tracting multi-dimensional MFCC feature vectors. Three types of audio samples with two classes each are transformed into the Mel scale for further processing. The Mel scale categorizes pitch where humans can interpret changes in pitch to be equal in length from each other along this scale. It is intended to make changes in frequency, such as with a spectrogram, more closely reflect audible changes. Mel scale provides a higher resolution in lower frequencies and vice versa. Since symptomatic acoustic data are known to have more energy in lower frequencies, the MFCC is a more suitable representation for these sounds. The following features were extracted and fed into the model as part of the training.

• Chroma stft: Computes a chromagram from a wave- form or power spectrogram.

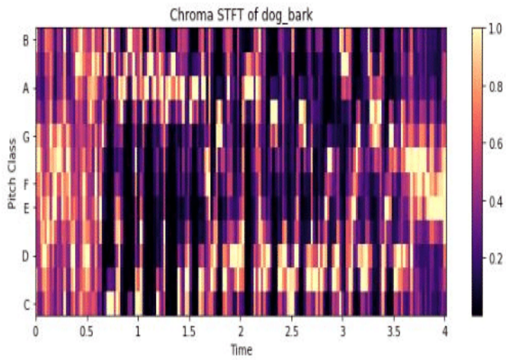
Chroma STFT The Chroma value of an audio basically represent the intensity of the twelve distinctive pitch classes that are used to study music. They can be em- ployed in the differentiation of the pitch class profiles be- tween audio signals. STFT represents information about the classification of pitch and signal structure. It depicts the spike with high values in low values (dark regions).

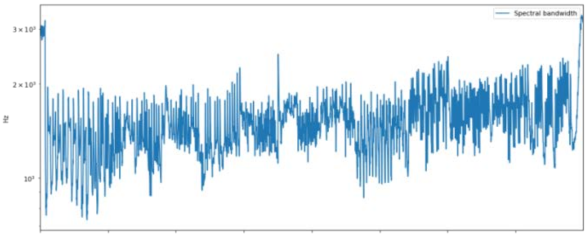


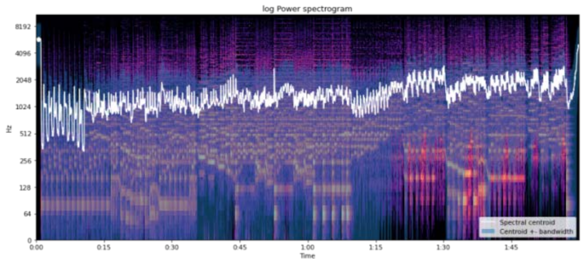
• Spectral centroid:Computes the spectral centroid. The spectral centroid is a measure used in digital signal processing to characterise a spectrum. It indicates where the center of mass of the spectrum is located. Perceptu- ally, it has a robust connection with the impression of brightness of a sound. It is sometimes called center of spectral mass.

• Spectral bandwidth:Computes the p'th-order spectral bandwidth.

Bandwidth is the difference between the upper and lower frequencies in a continuous band of frequencies. As we know the signals oscillate about a point so if the point is the centroid of the signal then the sum of maximum deviation of the signal on both sides of the point can be considered as the bandwidth of the signal at that time frame.

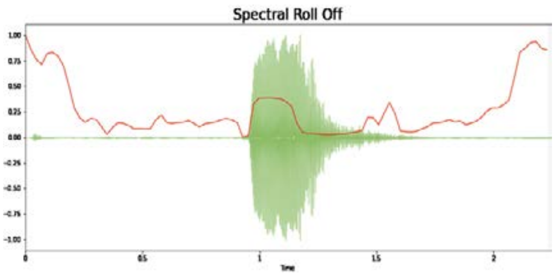






• Roll off:Computes the roll-off frequency.

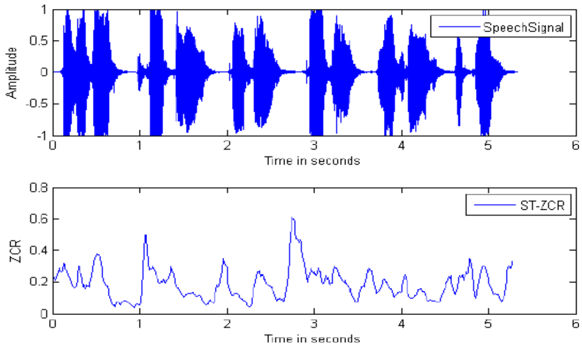
The roll-off frequency is defined as the frequency under which some percentage (cutoff) of the total energy of the spectrum is contained. The roll-off frequency can be used to distinguish between harmonic (below roll-off) and noisy sounds (above roll-off).



• Zero crossing rate:Compute the zero-crossing rate of an audio time series.

The zero-crossing rate (ZCR) is the rate at which a signal changes from positive to zero to negative or from negative to zero to positive. Its value has been widely used in both speech recognition and music information retrieval. Average zero-crossing rate refers to the number of times speech samples change algebraic sign in a given frame. The rate at which zero crossings occur is a simple measure of the frequency content of a signal. It is a measure of number of times in a given time interval/frame that the amplitude of the speech signals passes through a value of zero.

Apart from the aforementioned features which are provided by librosa, twenty mfcc features were also used.



A. Random Forest Classifier

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an en- semble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models. Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piece-wise constant approximation. The proposed system employs random forest classifier to classify the cough recordings of the individuals as either positive or negative. This model was chosen over the CNN model as the former showed better performance with the audio recordings. Neural networks have been shown to outperform a number of machine learning algorithms in many industry domains. They keep learning until it comes out with the best set of features to obtain a satisfying predictive performance. However, a neural network will scale the variables into a series of numbers that once the neural network finishes the learning stage, the features become indistinguishable to us. The fundamental reason to use a random forest instead of a decision tree is to combine the predictions of many decision trees into a single model. The logic is that a single even made up of many mediocre models will still be better than one good model. There is truth to this given the mainstream performance of random forests. Random forests are less prone to overfitting because of this. Random Forest is less computationally expensive and does not require a GPU to finish training. A random forest can provide a different interpretation of a decision tree but with better performance.

Neural Networks will require much more data than an everyday person might have on hand to actually be effective. The neural network will simply decimate the interpretability of the features fed as the input to a model, to a the point where it becomes meaningless for the sake of performance.

B. Librosa

Librosa is a Python package for music and audio analysis. Librosa is basically used when we work with audio data like in music generation(using LSTM's), Automatic Speech Recog- nition. It provides the building blocks necessary to create the music information retrieval systems. Librosa helps to visualize the audio signals and also do the feature extractions in it using different signal processing techniques. For the analysis of the cough recordings librosa has been used.

IV. CONCLUSION

This initiative, though there is still much work to do, shows that artificial intelligence could be used to tackle and track the covid-19 infection. AI techniques can produce a free, non- invasive, real-time, any-time, instantly distributable, large- scale COVID-19 asymptomatic screening tool to augment current approaches in containing the spread of COVID-19. Practical use cases could be for daily screening of students workers, and public as schools, jobs, and transport.

A. Need to Improve the Dataset

Apart from cough recordings, a wider dataset consisting of mere human speech recordings could also be trained to develop an even more resourceful system. This will lead to an improvement of the quality of the dataset which will result in the wider and better use of this tool.

REFERENCES

[1] R. J. Barro, J. F. Ursua, and J. Weng, "The coronavirus and the great influenza pandemic: Lessons from the "spanish flu" for the coronavirus's potential effects on mortality and economic activity," Nat. Bur. Econ. Res., Cambridge, MA, USA, Tech. Rep. w26866, 2020.

[2] World Health Organization. (2020) WHO Coronavirus Disease (COVID- 19) Dashboard. Accessed: Jan. 1, 2021. [Online].

[3] M. Venter and K. Richter, "Towards effective diagnostic assays for COVID-19: A review," J. Clin. Pathol., vol. 73, no. 7, pp. 370–377, Jul. 2020.

[4] J. Laguarta, F. Hueto and B. Subirana, "COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings," in IEEE Open Journal of Engineering in Medicine and Biology, vol. 1, pp. 275-281, 2020, doi: 10.1109/OJEMB.2020.3026928.

[5] L. Verde, G. De Pietro, A. Ghoneim, M. Alrashoud, K. N. Al-Mutib and G. Sannino, "Exploring the Use of Artificial Intelligence Techniques to Detect the Presence of Coronavirus Covid-19 Through Speech and Voice Analysis," in IEEE Access, vol. 9.

[6] N. Mohan, S. Kabeer and N. Nasir, "Artificial Intelligence (AI) in the diagnosis of COVID-19 Detection: A Review," 2021 28th IEEE International Conference on Electronics, Circuits, and Systems (ICECS), 2021, pp. 1-6, doi: 10.1109/ICECS53924.2021.9665470.

[7] McFee B, McVicar M, Balke S, Thome´ C, Raffel C, Lee D, et al. librosa/librosa: 0.6.1; 2018. Available from: https://doi.org/10.5281/zenodo.6097378.