Automated Lung Sound Analysis/Classification

Abstract: Effective management of respiratory disorders depends on patient surveillance and early diagnosis. In clinical settings, lung auscultation—which involves using a stethoscope to hear a person's lungs—is used to identify respiratory disorders. The two types of lung sounds that are often distinguished are normal and adventitious. The most frequent abnormal lung sounds over those that are normal are crackles, wheezes, and squawks, and their presence usually denotes a pulmonary condition. Nevertheless, due to the lengthy process and high level of medical skill required for traditional diagnosis, the rapid advancement of technology has made it possible for data volumes to increase dramatically. Numerous studies have suggested various strategies for the automatic classification of lung sounds in order to solve this challenge. In this report we are focussed on extracting features from an audio which can then be fed into a CNN model for automatic classification of healthy and unhealthy sounds thus reducing the manual drudgery of classification of these audios.

1. Introduction

Over the past few decades, there has been a lot of research and development in the automated classification of respiratory sounds. Automated respiratory sound classification has the ability to identify anomalies in the first phases of a respiratory malfunction and thus improve the trajectory of medical diagnosis.

Mainly there are two types of Lung sounds:

1. **Normal Lung sounds**: Sounds heard over the chest wall with no apparent abnormalities and with a median frequency of 500Hz are classified as Normal Lung sounds.

2. **Adventitious Lung Sounds**: Adventitious sounds, the medical term for abnormal lung sounds, are respiratory noises which might include crackles, rhonchi, and wheezes that are irregular in nature.The noises might be the initial sign of a respiratory condition like pneumonia, COPD or other respiratory diseases. Adventitious noises are broadly categorized as follows:

2.1. **Crackles:** In the lungs, crackles are minute bubbling noises.These are generally musical in nature following a pattern and originate from the base of the lungs. These were earlier called ‘Rales’. These can be bilateral, i.e occurring in both lungs or unilateral i.r occurring in a single lung. These can also be coarse or fine crackles.

2.2. **Rhonchi:** Rhonchi, similar to snoring, are respiratory soundsThese are generally low pitched and occur where airways in the lungs are broader. These are a symptom of liquid or semi liquid mucus accumulation in some part of the lung.

2.3. **Wheezes:** Wheezes produce loud noises. These, Unlike Rhonchi are noises produced in narrow pathways of the lung and are generally high pitched in nature.

A publically available large database of lung sounds, whether normal or adventitious in which new algorithms can be implemented, is thus extremely important. This was made possible in the context of the International Conference on Biomedical and Health Informatics (ICBHI) 1st scientific challenge with a vision of developing Machine Learning algorithms that are able to classify respiratory sound recordings passed through them as either normal or adventitious.

The database was created by two research teams in Portugal and in Greece, and it includes 920 recordings acquired from 126 subjects. A total of 6898 respiration cycles were recorded. The cycles were annotated by respiratory experts as including crackles, wheezes, a combination of them, or no adventitious respiratory sounds. The recordings were collected using heterogeneous equipment and their duration ranged from 10s to 90s. The chest locations from which the recordings were acquired was also provided. Noise levels in some respiration cycles were high, which simulated real life conditions and made the classification process more challenging[1]

The majority of algorithms designed to identify or categorize events contain two phases.The algorithms designed to classify these sounds can usually be broken down in two steps.

1)The extraction of some features or variables that will be used as a metric to

classify the sounds.

2)These features extracted in step 1 are then passed through a certain method subject to the type of variable extracted.

Mel-frequency cepstral coefficients (MFCCs), spectral characteristics, energy, entropy, and wavelet coefficients are some of the most commonly extracted features for this purpose. Support vector machines (SVMs), artificial neural networks (ANNs), Gaussian mixture models (GMMs), k-nearest neighbors (k-NNs), and logistic regression models are some of the machine learning algorithms proposed for automatic classification of lung sounds.

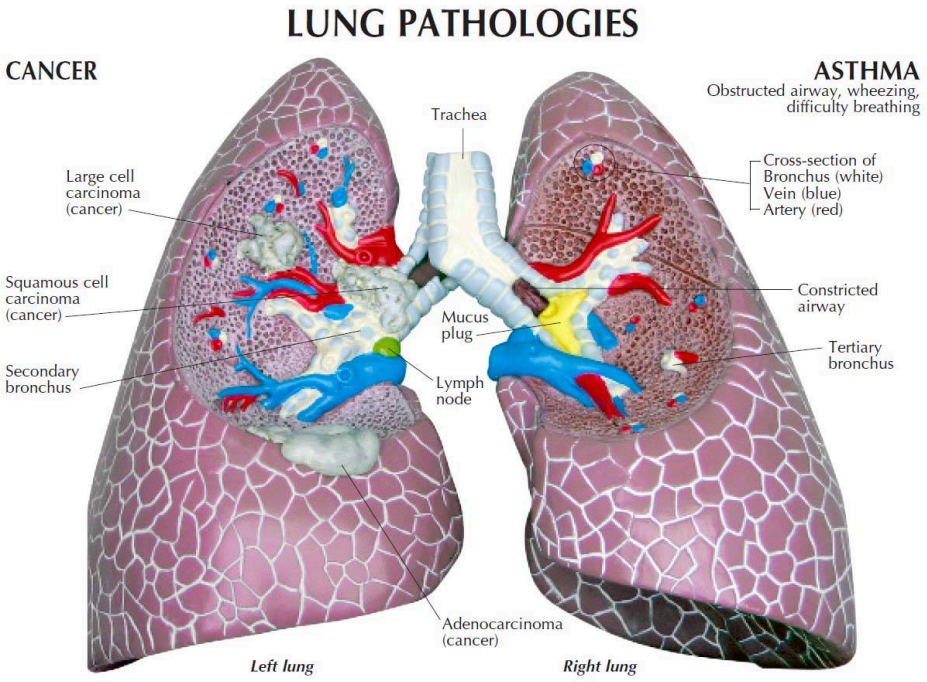


Fig1: Cross section view of Cancerous Lung(Left) and Asthmatic lung(Right)[Taken from AnatomyNow.com]

Feature Extraction

The feature extraction methodologies deployed by us include spectrogram extraction through Short time Fourier Transform(STFT) and Mel Frequency Cepstral Coefficients(MFCCs).

1. *SHORT TIME FOURIER TRANSFORM*

The variation of sinusoidal frequency and phase in short frames of time of a signal are calculated using Short time Fourier transform or STFT. To calculate STFT, we window the signal i.e divide it into short frames of time of equivalent intervals and then separately compute the fourier transform of each window.

In addition to showing the frequency content of the signals, it is also desirable to have an idea of when each frequency content is dominant and this is made possible by STFT .The squared magnitude of this windowed STFT representation, |X(m, ω)| 2 , also known as spectrograms can be used for this purpose.Fig 2 shows the visualizes the process of extracting short time fourier transform

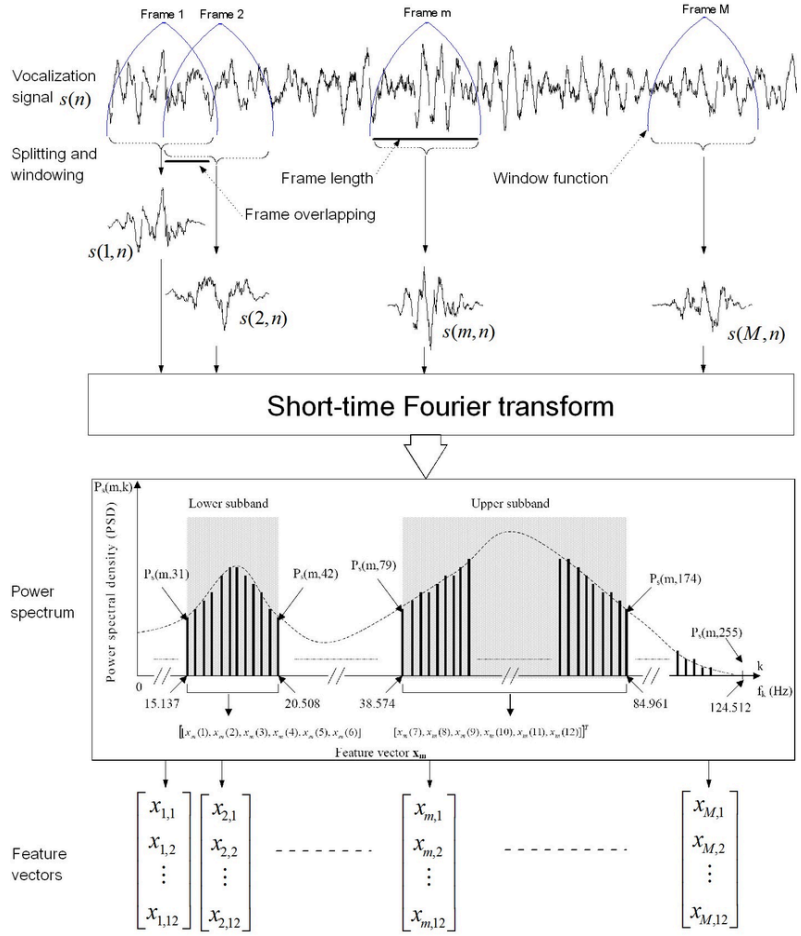


Fig 2: Feature extraction through Short time Fourier transform[2]

2. *SPECTROGRAM*

Visual representation of frequencies of a given signal with time is called Spectrogram. In a spectrogram representation plot — one axis represents the time, the second axis represents frequencies and the colors represent magnitude (amplitude) of the observed frequency at a particular time. Strong frequencies are represented by bright colors.[3]

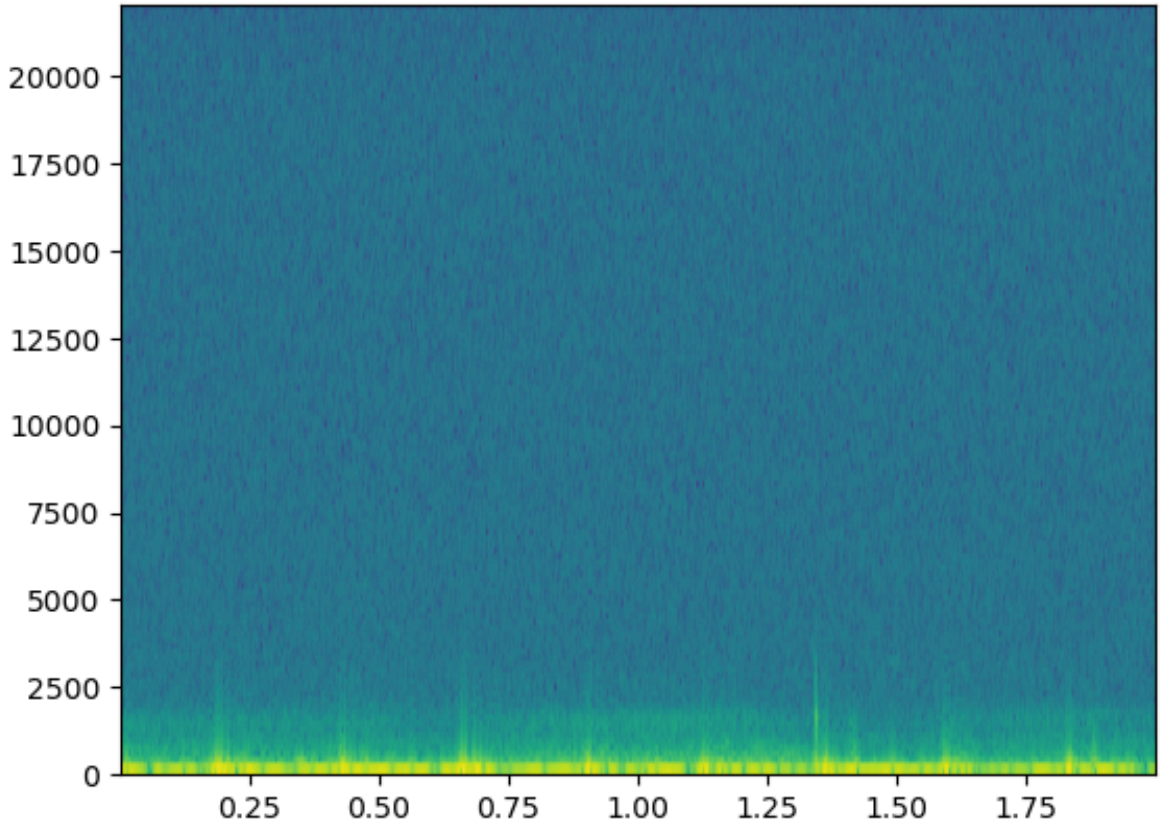


Fig 3: Spectrogram of lung sound with upper respiratory tract infection(URTI)

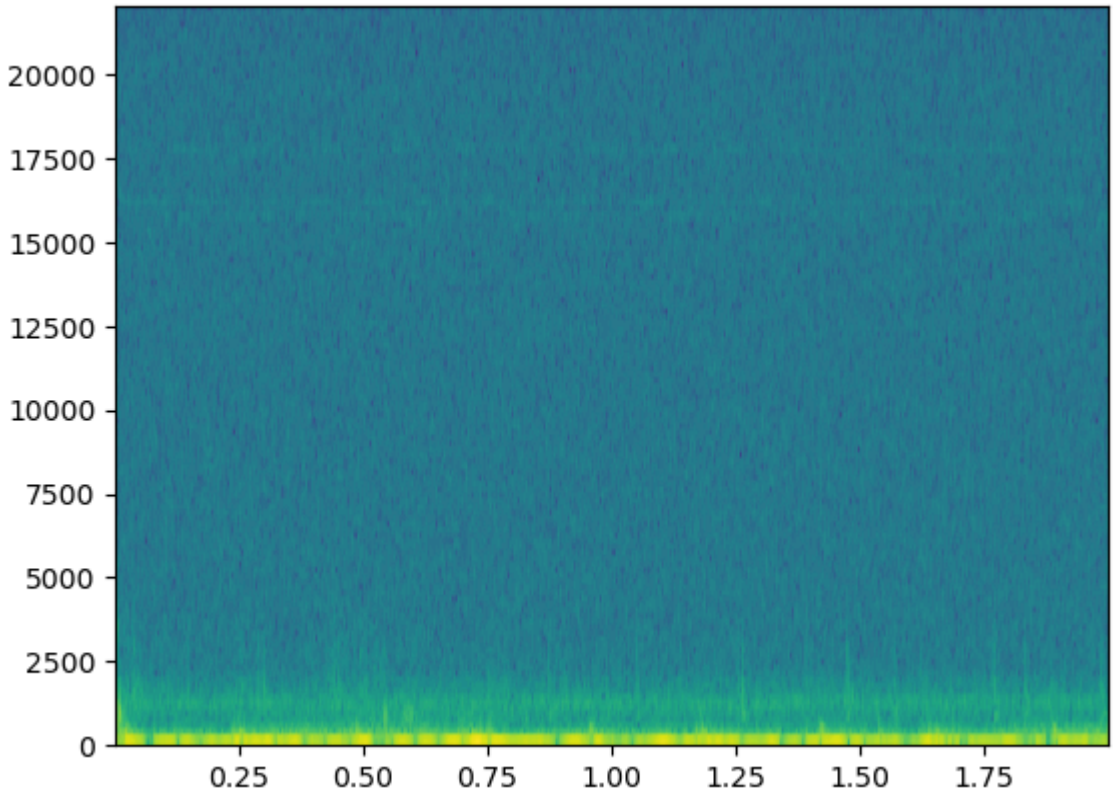


Fig 4: Spectrogram of healthy lung sound

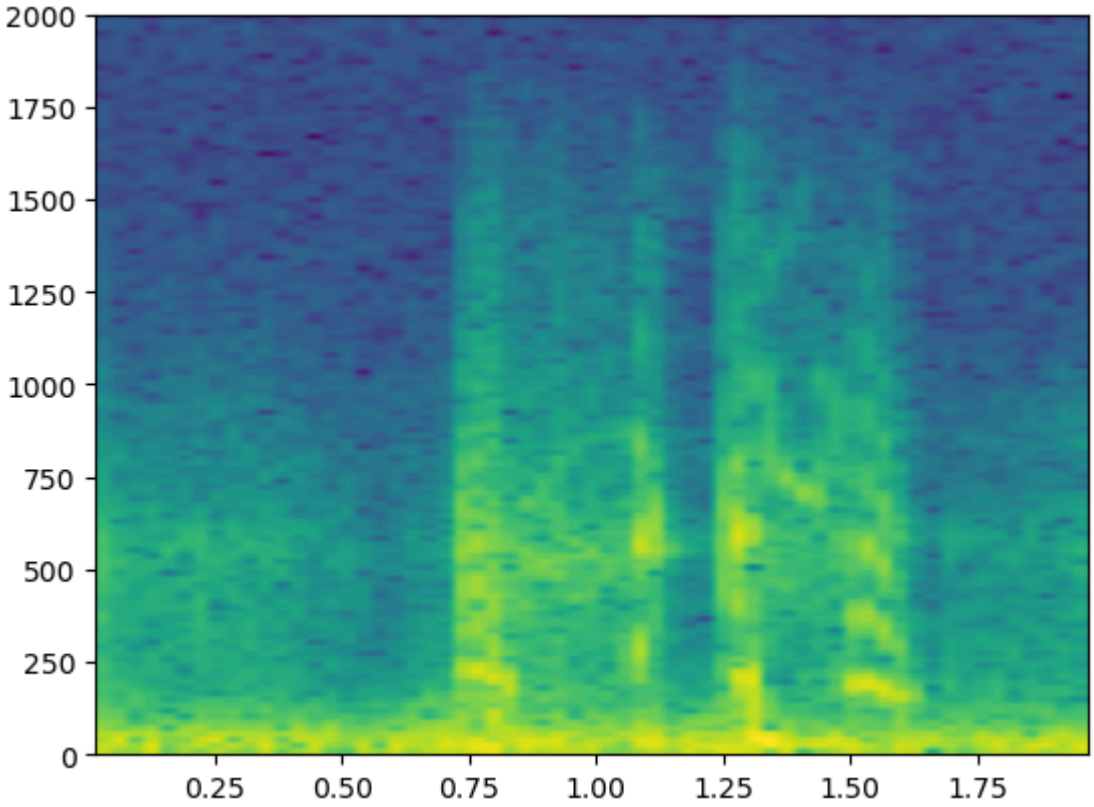


Fig 5: Spectrogram of lung sound with COPD

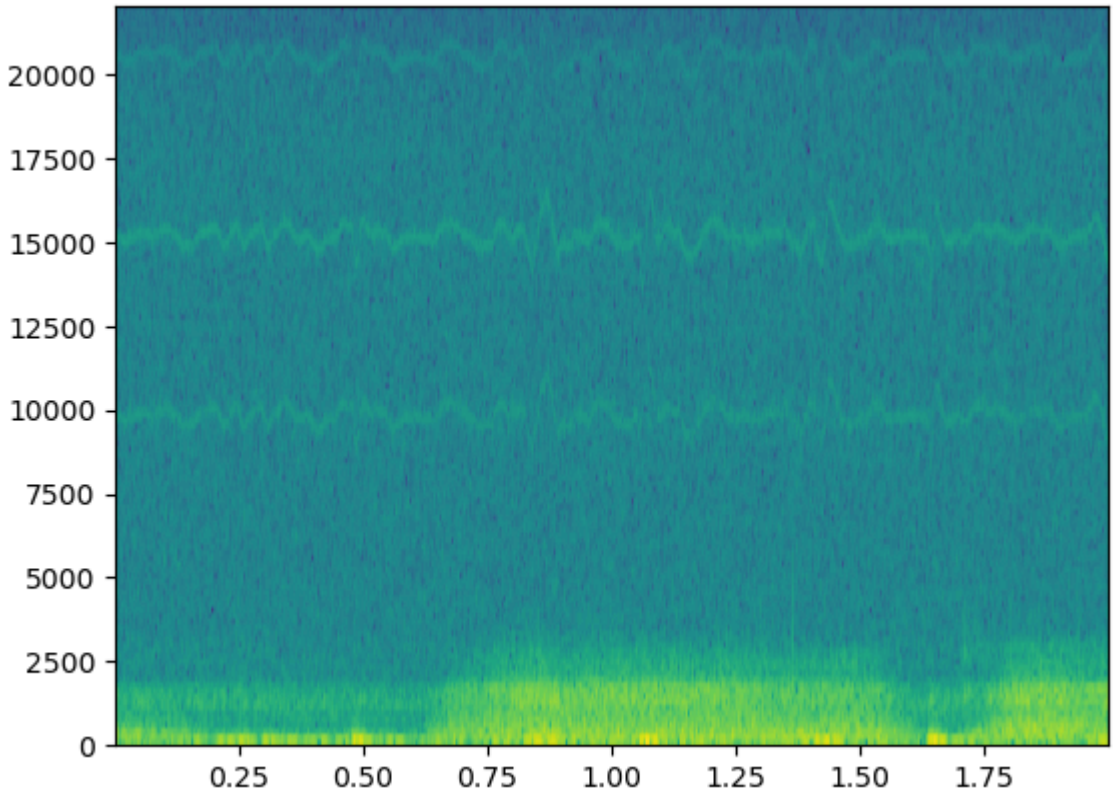


Fig 6:Spectrogram of lung sound with bronchitis

3. *Mel Frequency Cepstral coefficients(MFCCs)*

In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC.[1] They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). The difference between the cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal spectrum. This frequency warping can allow for better representation of sound, for example, in audio compression that might potentially reduce the transmission bandwidth and the storage requirements of audio signals.[4]

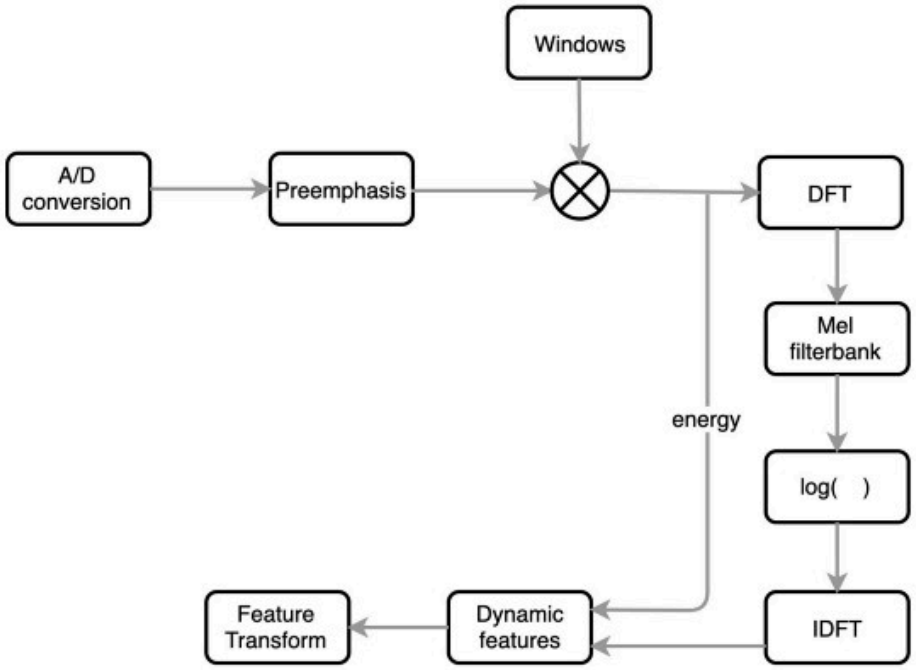


Fig 7: Modular representation of mel-frequency cepstral coefficients (MFCCs) feature extraction process[Taken from AnalyticsVidhya]

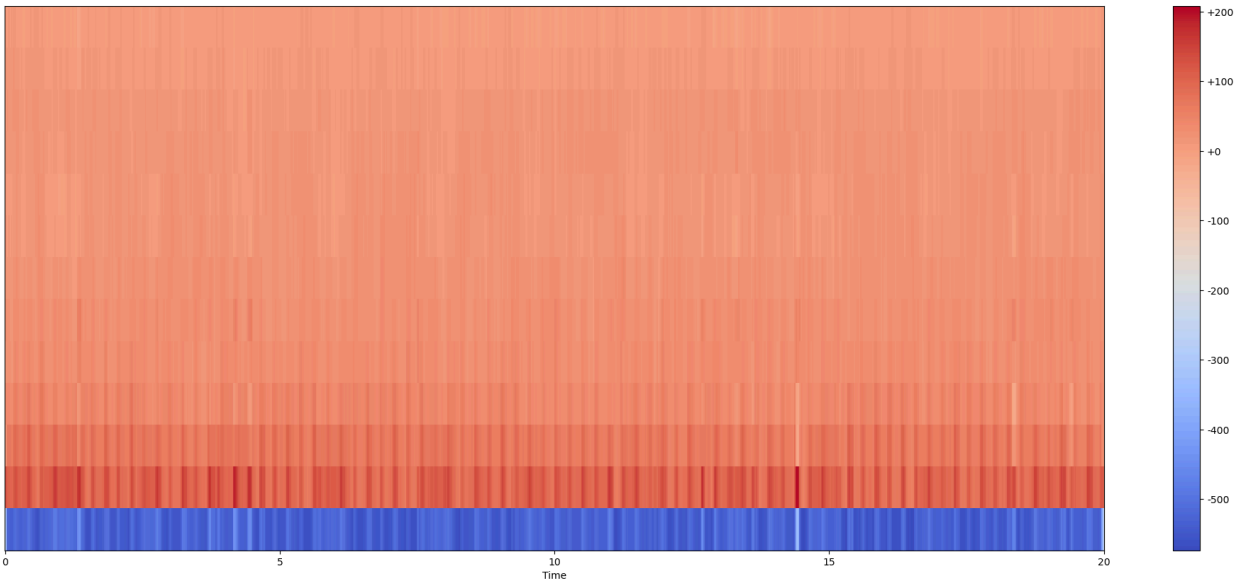


Fig 8: MFCC plot of a lung sound with URTI

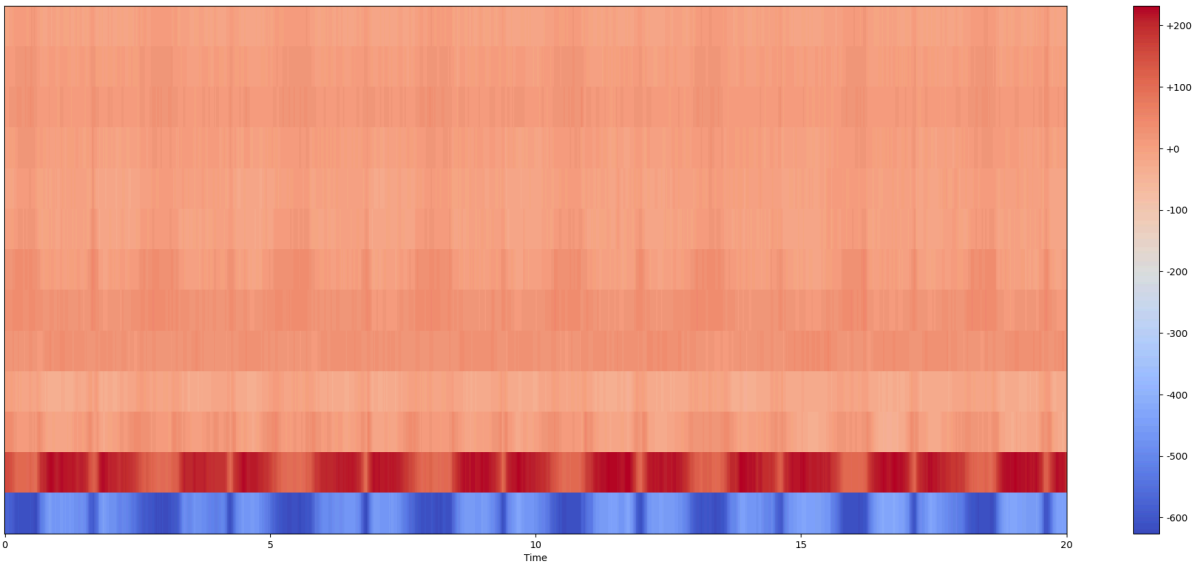


Fig 9: MFCC plot of lung sound with bronchitis

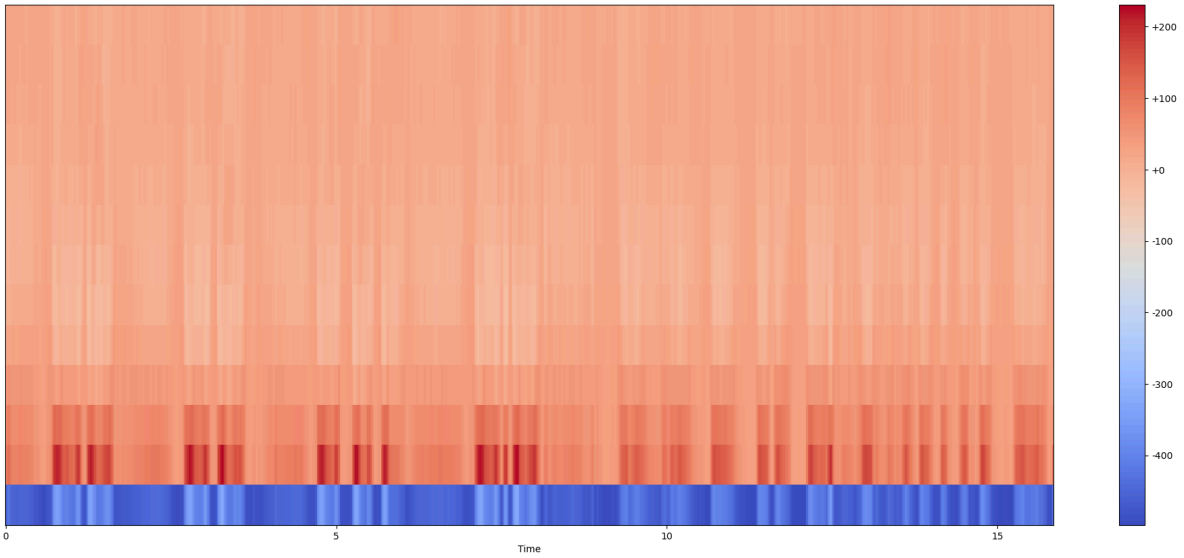


Fig 10: MFCC plot of lung sound with COPD

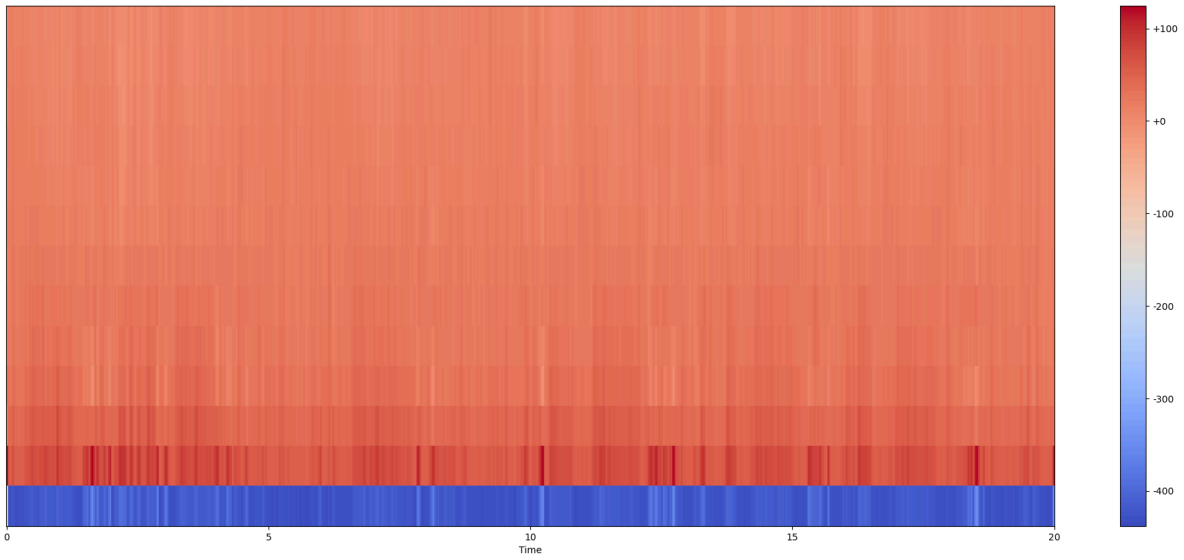


Fig 11: MFCC plot of lung sound with Asthma

Data Preprocessing

Data preprocessing is crucial for successful machine learning models, as the accuracy and usefulness of results depend heavily on the quality of preprocessed data. In audio analysis and modeling, preprocessing is vital for converting complex and noisy raw audio data into a suitable format for further analysis. Techniques such as filtering, normalization, segmentation, feature extraction, and encoding are used to remove noise, extract relevant features, and improve analysis and modeling accuracy. Effective preprocessing is essential for achieving reliable results and enhancing the overall quality and usefulness of audio data analysis and modeling. One of the most important preprocessing steps in machine learning is data resampling. This is the process of either upsampling or downsampling (decreasing the number of samples in a dataset).

When we lower the sample rate of audio by an integer, it is referred to as downsampling. This sample rate is the rate at which we sample our audio when digitizing it from analog to digital. Downsampling is also referred to as decimation. Downsampling gives a kind of compression effect to our waveform/signal.

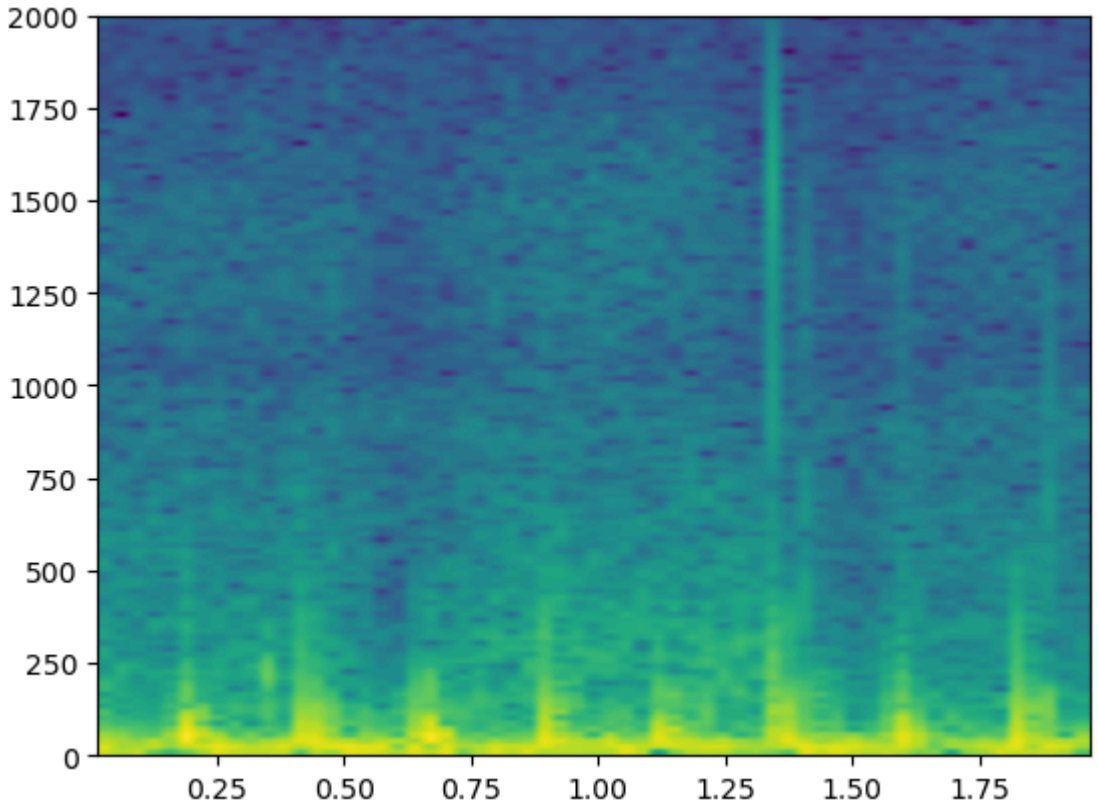


Fig 9: downsampled spectrogram of lung sound with URTI

Convolutional Neural Network (CNN) model architecture

We will use a CNN model architecture.

We'll employ a sequential model with a straightforward model architecture made up of four Conv2D convolution layers, with a dense layer serving as the final output layer.

The purpose of the convolution layers is feature detection. To operate, you slide a filter window over the input, multiply the input by a matrix, and then store the result in a feature map. Convolution is the name of this operation.

The number of nodes in each layer is specified by the filter parameter. The kernel_size argument determines the size of the kernel window, which in this example is 2, resulting in a 2x2 filter matrix. Each layer will grow in size from 16, 32, 64, to 128.

The first layer will be fed an input shape of (40, 862, 1), where 40 is the number of MFCCs, 862 is the number of frames with padding, and 1 indicates that the audio is mono.

ReLU will be the activation function for our convolutional layers. On our convolutional layers, we will utilise a modest Dropout value of 20%.

Each convolutional layer is coupled with a pooling layer of type MaxPooling2D, with the final convolutional layer of type GlobalAveragePooling2D. The pooling layer reduces the dimensionality of the model (by decreasing the parameters and subsequent computing needs), which reduces training time and overfitting. The Max Pooling type uses the maximum size for each window, whereas the Global Average Pooling type uses the average that can be fed into our dense output layer.

Our output layer will contain 6 nodes (num_labels), which corresponds to the number of classifications. Softmax is the activation for our output layer. Softmax makes the output total to one, allowing it to be read as probabilities. After that, the model will forecast which choice has the highest chance.

Layers of CNN Model architecture

Conv2D: This layer performs the convolution operation to extract features from the input images.

- **Filters=16** It specifies the number of filters or feature maps the layer will learn. Each filter is responsible for detecting a specific pattern in the input images.
- **kernel_size=filter_size:** The kernel size determines the spatial dimensions of the filters.
- **activation='relu':** RELU stands for rectified linear unit. This acts as a kind of activation function for introduction of non linearity into our model.This helps in ascertaining the complex patterns arising in the data.

MaxPooling2D: This layer performs downsampling or pooling but also retains all important information

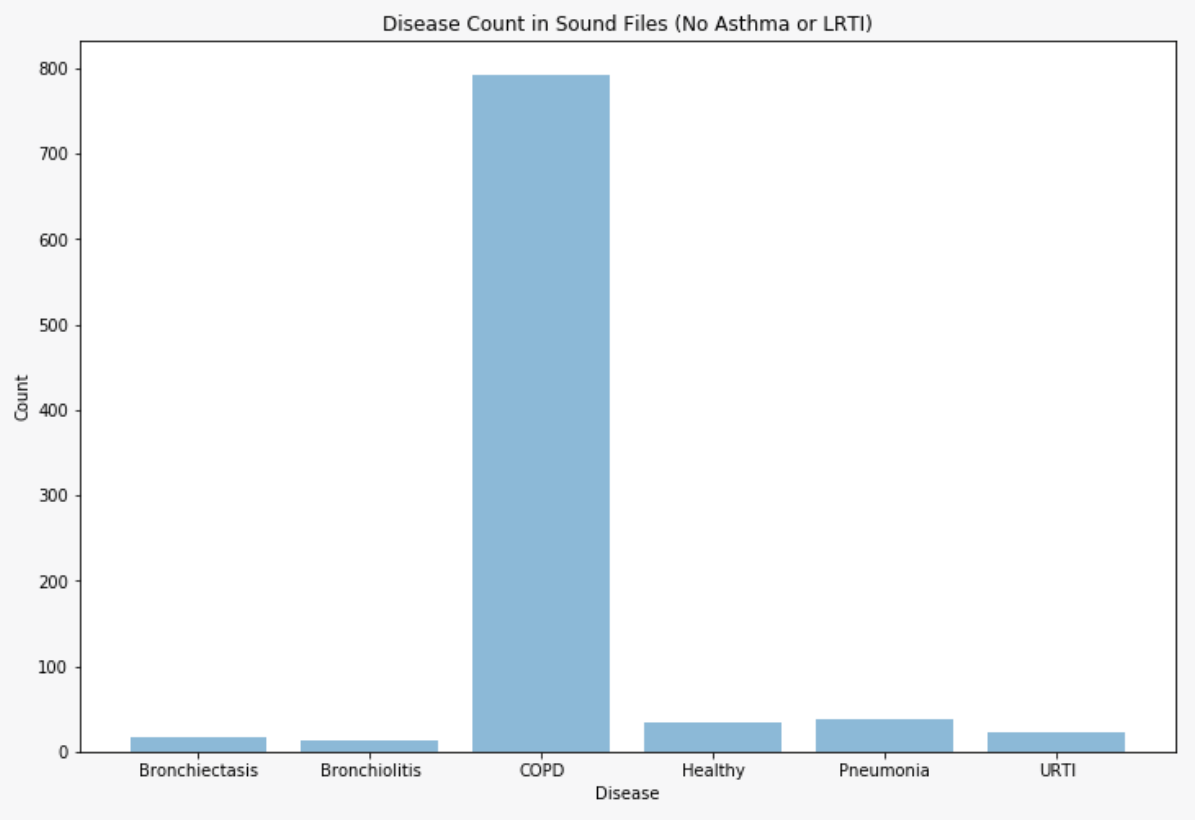
- pool_size=2: It specifies the size of the pooling window. Here, a 2x2 pooling window is used, which means the spatial dimensions will be reduced by half.

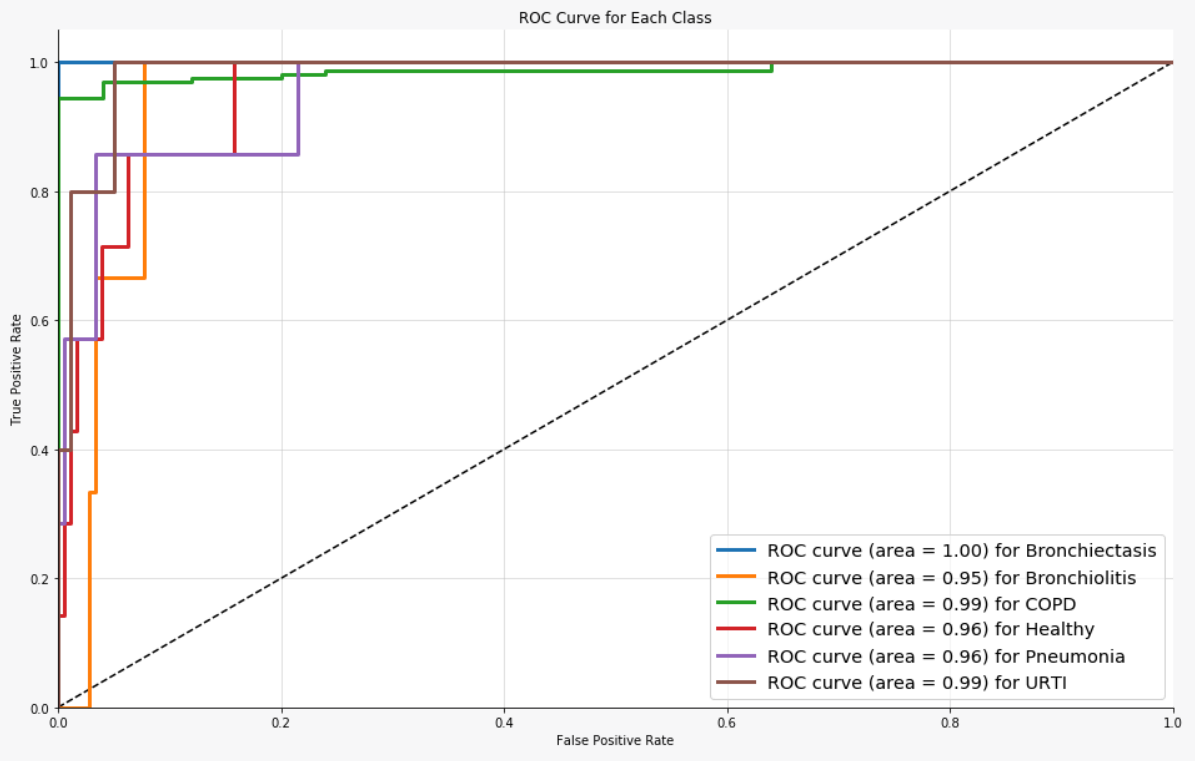
Dropout: This layer applies dropout regularization to eliminate overfitting.

GlobalAveragePooling2D: This layer computes the spatial average of the feature maps, collapsing them into a single vector. It reduces the spatial dimensions to a fixed size, making the model suitable for handling inputs of variable sizes.

Dense: This is a fully connected layer, which takes the features extracted from the convolutional layers and performs classification.

- **activation='softmax':** The softmax activation function is used to convert the raw scores of the neurons into probability scores, indicating the likelihood of each class.





Acknowledgments

My great professor, Dr. Mahesh R. Panicker, has been an outstanding mentor and has provided continuous support over the course of this research, and I am really grateful to him. My work has been profoundly influenced by Dr. Panicker's vast knowledge and competence in the field, which has served as a constant source of inspiration.

I would like to express my sincere gratitude to the Indian Institute of Technology, Palakkad for giving me the chance to complete my internship. Modern facilities and a supportive research atmosphere at the institute have greatly aided my learning and development. In addition, I am incredibly grateful to the Indian Academy of Sciences for giving me the chance to participate in this renowned fellowship. Along with their financial help, their encouragement pushed me to strive for academic excellence. My sincere gratitude is extended to all the professors, associates, and employees at the academy and the institute, whose unwavering support and cooperation have been crucial in enabling this research project.

Finally, I want to thank my family and friends for their continuous support and patience while I was conducting this research. Their support has served as a continual motivator, encouraging me to persevere in the face of difficulties. In closing, I just want to say how grateful I am for everyone's support and encouragement along this journey. They have helped me advance in my academic and scientific endeavors because they have faith in my talents.

References

1. Rocha, B.M. *et al.* (2018). A Respiratory Sound Database for the Development of Automated Classification. In: Maglaveras, N., Chouvarda, I., de Carvalho, P. (eds) Precision Medicine Powered by pHealth and Connected Health. ICBHI 2017. IFMBE Proceedings, vol 66. Springer, Singapore. https://doi.org/10.1007/978-981-10-7419-6_6
2. Bahoura, Mohammed. (2016). FPGA Implementation of Blue Whale Calls Classifier Using High-Level Programming Tool. Electronics. 5. 8. 10.3390/electronics5010008
3. https://en.wikipedia.org/wiki/Spectrogram
4. ://en.wikipedia.org/wiki/Mel-frequency_cepstrum#:~:text=In%20sound%20processing%2C%20the%20mel-frequency%20cepstrum%20%28MFC%29%0is,coefficients%20that%20collectively%20make%20up%20an%20MFC.%20

5. Towardsdatascience.com
6. Sovijärvi A, Dalmasso F, Vanderschoot J, Malmberg L, Righini G, Stoneman S (2000) Definition of terms for applications of respiratory sounds. Eur Respir Rev 10:597–610
7. Alqaderi, Mohammad & Rad, Ahmad. (2018). A Multi-Modal Person Recognition System for Social Robots. Applied Sciences. 8. 387. 10.3390/app8030387
8. https://www.medicalnewstoday.com/articles/adventitious-breath-sounds-types-causes-and-locations#definition