Breath sounds as a biomarker for screening infectious lung diseases

7th International Electronic Conference on Sensors and Applications

Contributed by

Harini S, Parijat Deshpande and Beena Rai

TCS Research, Pune, India

Motivation

- ✔ Regular monitoring of breath sounds.
- ✔ Screening needs to be done remotely and quickly.
- ✔ Affordability.

Hence, we have proposed to give a smartphone-based solution that monitors breath sounds from the user via the in-built microphone in his smartphone and our AI based anomaly detection engine.

Introduction to breath sounds

- ➔ Breath sounds can be broadly classified into normal and adventitious sounds.
- ➔ Adventitious breath sounds are abnormal sounds that occur over the lungs and airways.
- ➔ Abnormal sounds can be further classified into crackles, wheeze, rhonchi and stridor.

https://www.youtube.com/watch?v=zRv5tNCMpyY

Proposed solution

Task 1 :

Pre-processing of the sounds acquired from the user.

Task 2 :

Breathing sound detector: After the sounds are acquired from the user, this module detects whether the recorded sound is breath or not.

Task 3:

Anomaly detection engine: Identifies if any anomaly is detected in the user's breath sound or not.

Signal Processing features - Mel scale

1. **Mel scale:** The scale of pitches judged by listeners to be equal in distance from one another.

Why Mel scale?

- The perception of human perceived frequency is not linear.
- In a given interval, the pitch of higher frequency range sounds seems to be larger than the same interval at lower frequency.
- The Mel scale was defined so that intervals that sound the same also have the same measurement in Mel.

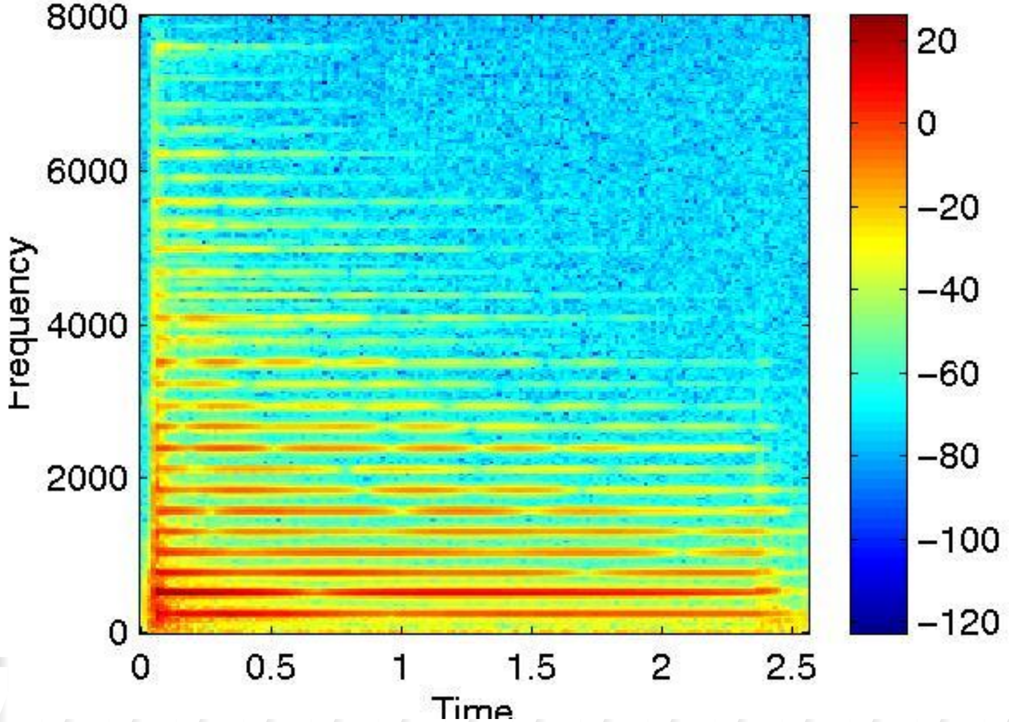
$$\mathrm{Mel}(f) = 2595 \log\left(1 + \frac{f}{700}\right)$$

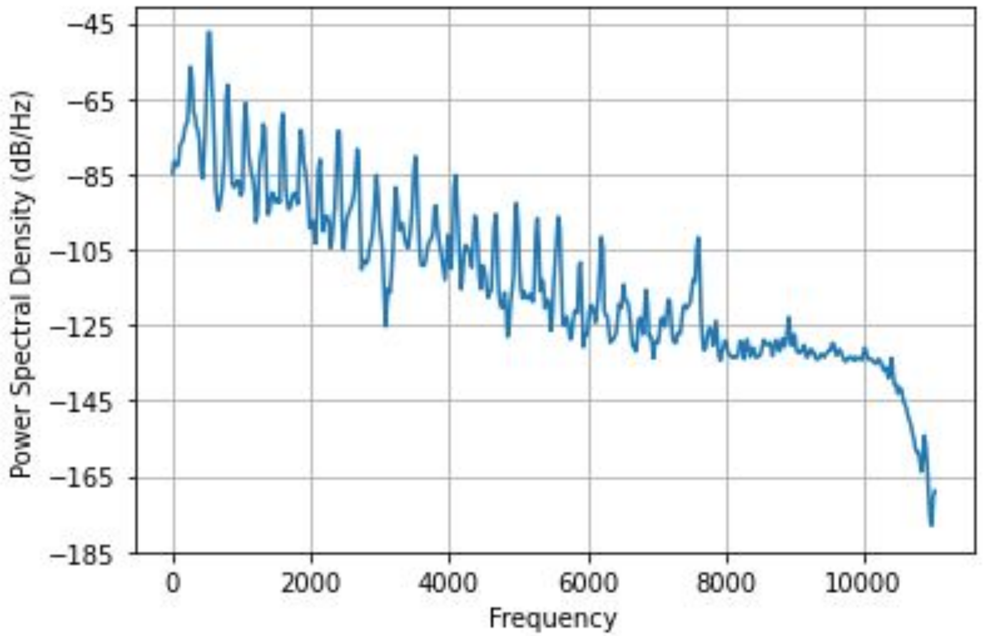
2. **Fourier transform**: Fourier transform is a mathematical concept that can decompose a signal into its constituent frequencies.

Signal Processing features - Spectrum

3. Spectrum: The frequency content present in a sound signal, usually a pictorial representation called *spectrogram*, which displays time vs frequency.

4. Power spectral density (PSD): Shows the strength of the variations of power as a function of frequency. It shows at which frequencies variations are strong and at which frequencies variations are weak.

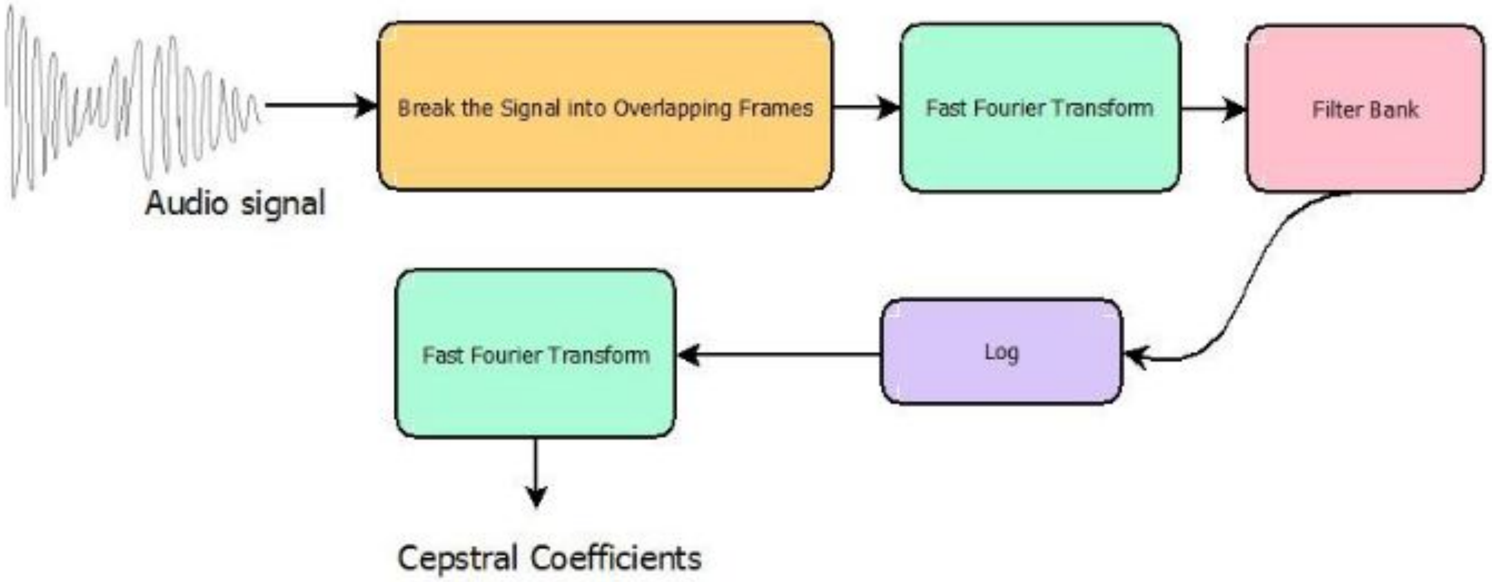




Signal Processing features - Cepstrum

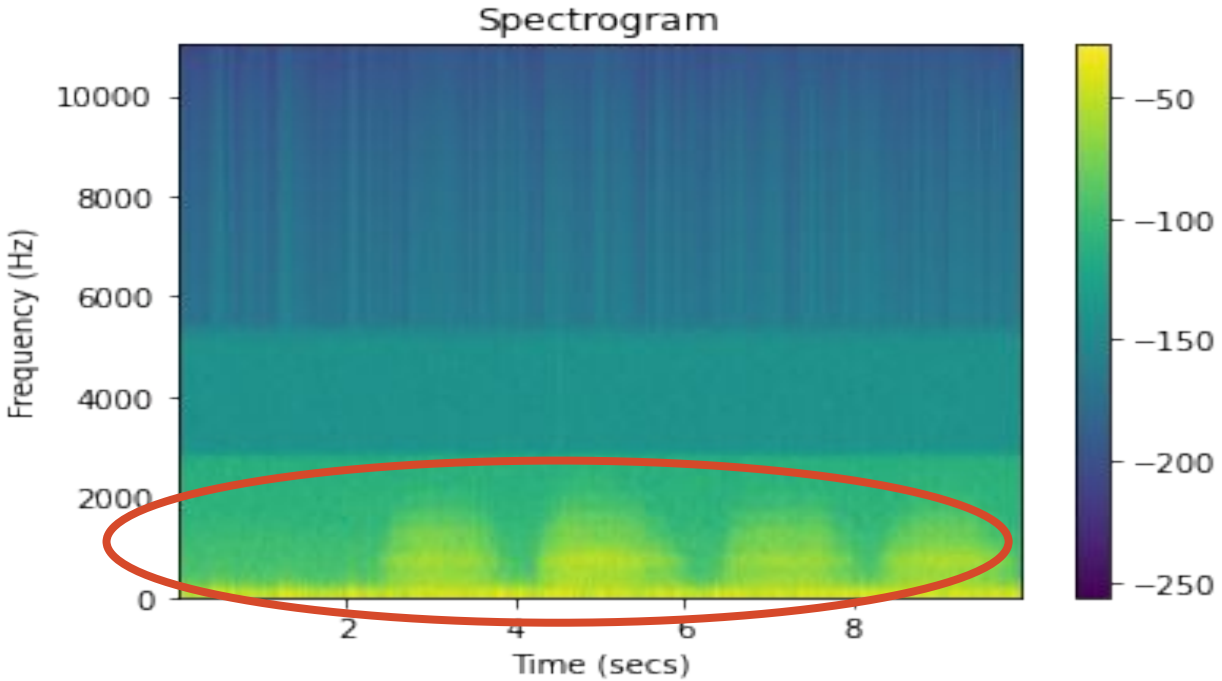
5. Cepstrum: Fourier transform of logarithm of spectrum.
Since we apply a transform on the frequency spectrum itself, the resulting spectrum is neither in the frequency domain nor in the time domain and it is called *quefrency domain.*

6. Mel Frequency Cepstral Coefficients (MFCC): The mel-frequency cepstrum is a representation of the power spectrum of a sound.
Mel-frequency cepstral coefficients are coefficients that collectively make up an MFC.

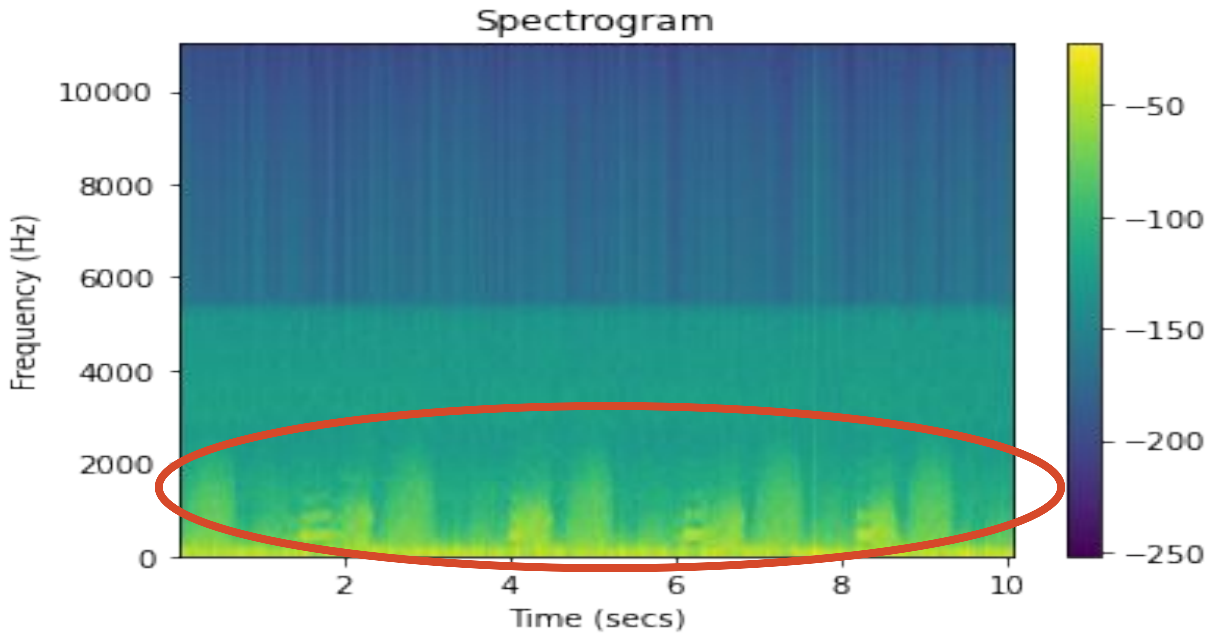


Spectrogram for normal and abnormal sounds

Spectrogram of normal sound

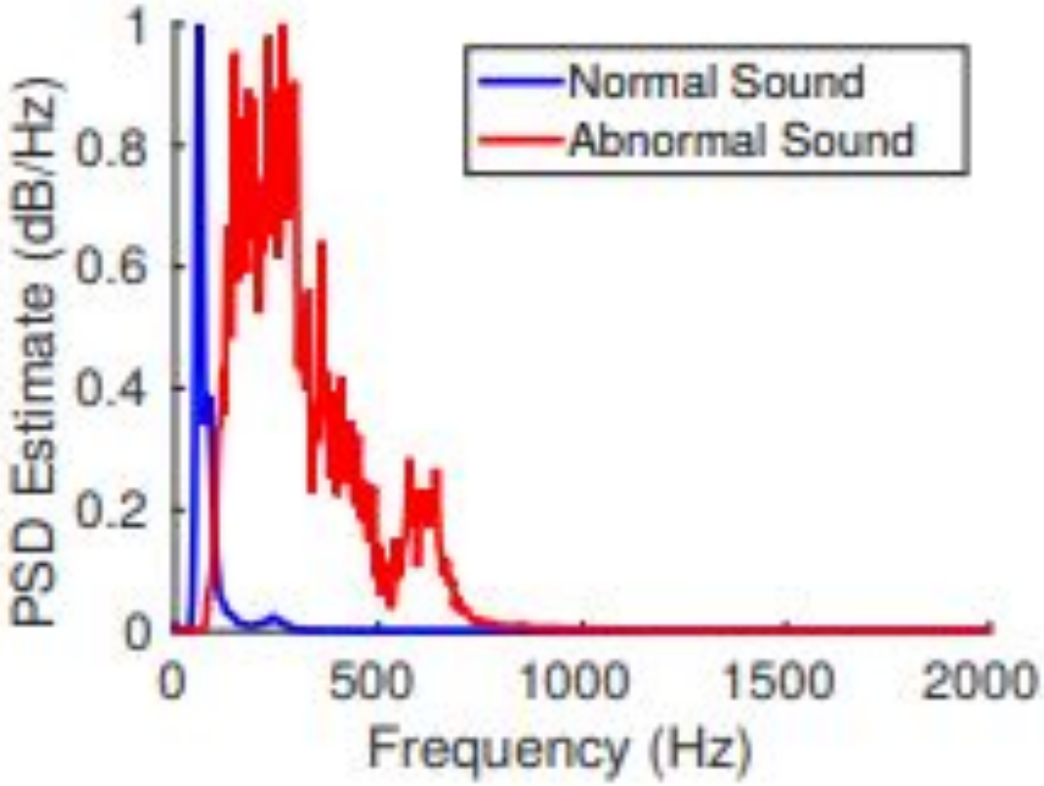


Spectrogram of abnormal sound



Significant differences can be seen in the spectrogram of normal and abnormal breath sound, which we have used as our feature.

PSD estimates for normal and abnormal sound



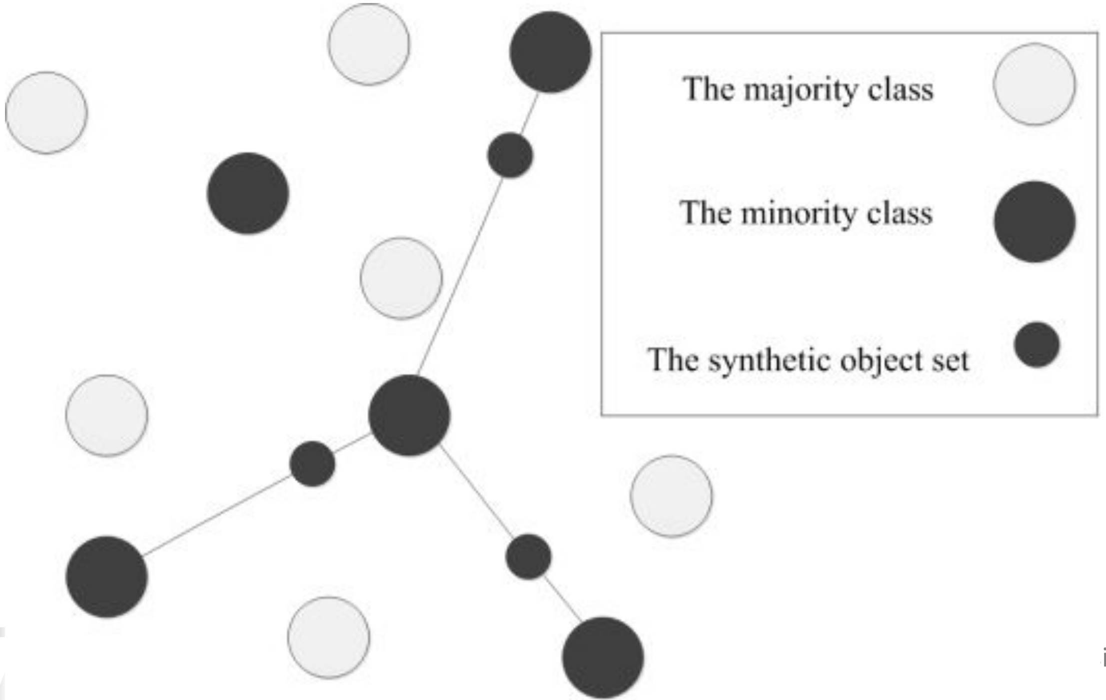
PSD value for normal sound tend to be lower than for abnormal sound, which is a significant feature.

Data Sources

For, the breath detector, we have used the ESC - 50 dataset, which has 50 different classes of environmental sounds, one of them is breath sound.

For anomaly detection engine, we have used the RALE database. The RALE database consisted of wheeze, crackles, rhonchi, stridor sounds etc. which are adventitious.

Since, this dataset was unbalanced, SMOTe algorithm is used.

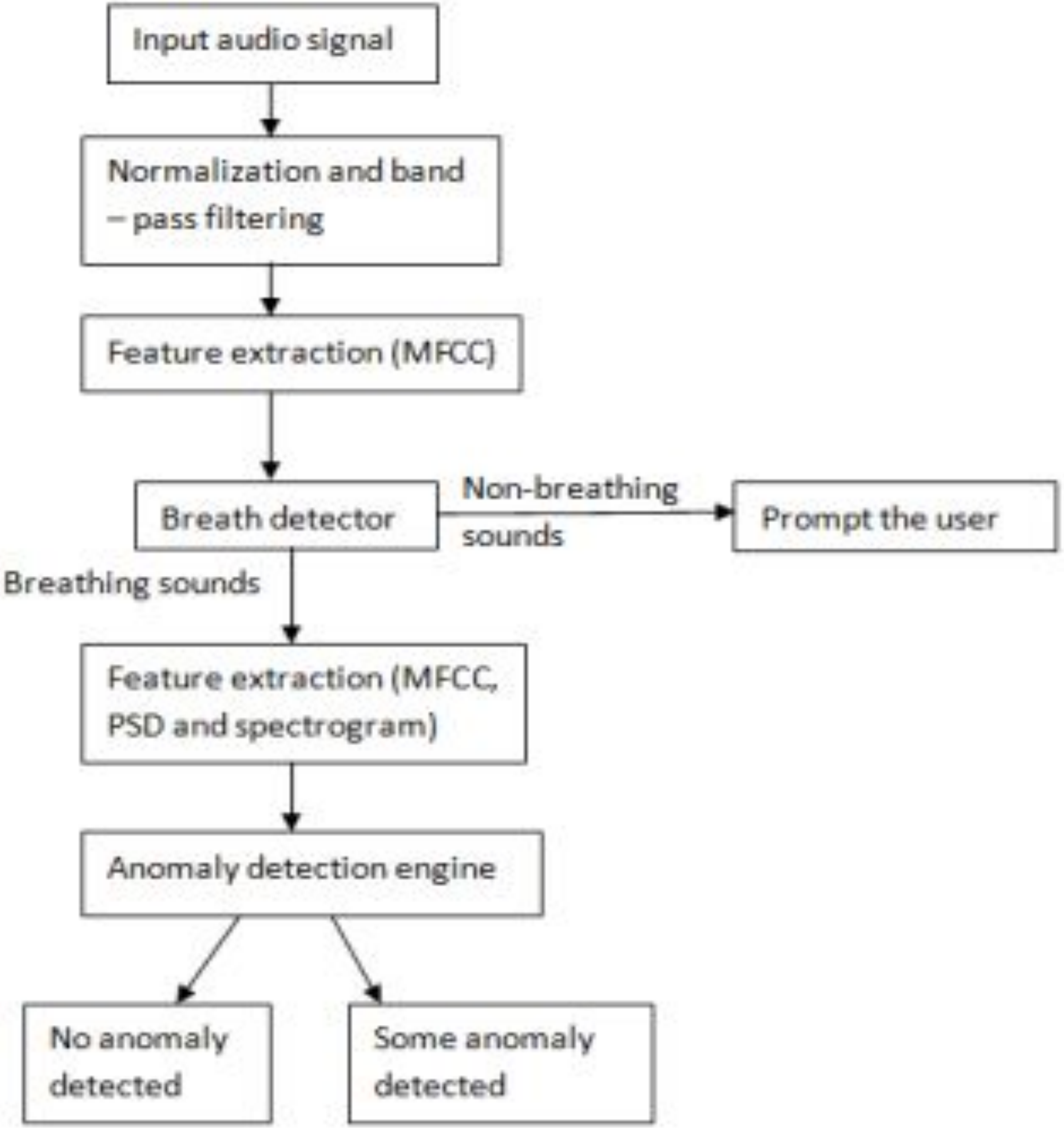


Contd.

<u>Synthetic Minority Over-sampling Technique (SMOTe)</u>

1. For each minority instance, k number of nearest neighbors are found such that they also belong to the same class.
2. The difference between the feature vector of the considered instance and the feature vectors of the k nearest neighbors are found.
3. The k difference vectors are each multiplied by a random number between 0 and 1.
4. Now, the difference vectors, after being multiplied by random numbers, are added to the feature vector of the considered instance at each iteration.

Solution flowchart



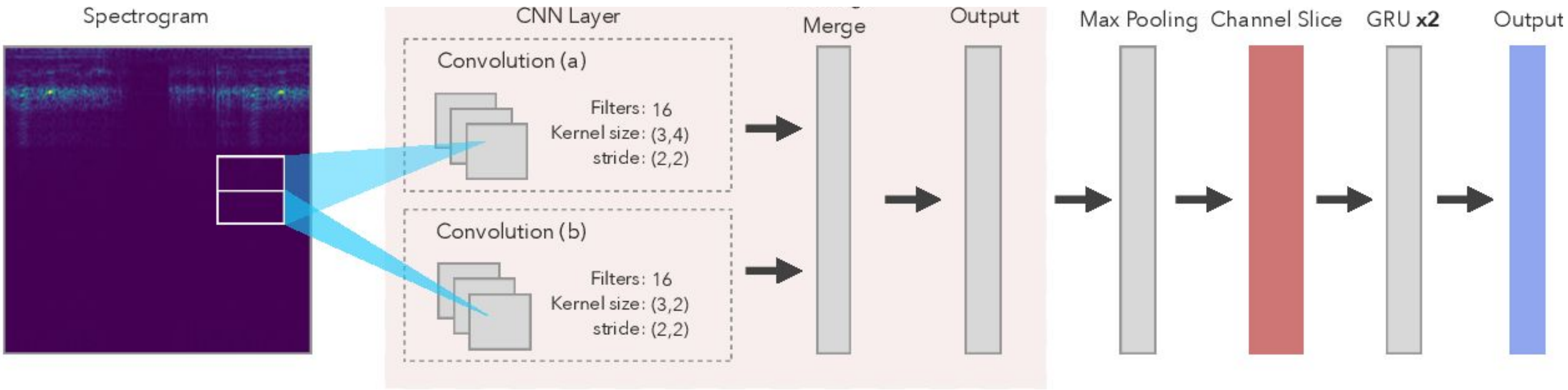
TATA CONSULTANCY SERVICES

Solution approach

Approach 1: *Machine learning based approach:* In this approach, I have extracted Power Spectral Density (PSD) of audio samples along with MFCCs. These are used as input to different machine learning algorithms.

Approach 2: *Deep learning based approach:* In this approach, I have used spectrogram images of the audio samples as feature. Considering the effectiveness of deep learning in classifying images,I have used Convolutional Neural Network (CNN) and CNN-GRU model for classification.

CNN-GRU


Spectrogram
CNN Layer
Convolution (a)
Filters: 16
Kernel size: (3,4)
stride: (2,2)
Convolution (b)
Filters: 16
Kernel size: (3,2)
stride: (2,2)
Merge
Output
Max Pooling
Channel Slice
GRU x2
Output

Results - Breath detector

Using 3 machine learning algorithms, an average accuracy of 98% accuracy is achieved.

Performance metrics

Model	Test accuracy(%)	Precision (%)	Recall (%)
KNN	99.39	98.00	99.00
RF	99.10	98.00	98.00
LR	98.79	98.00	98.00

Results - Anomaly detection engine

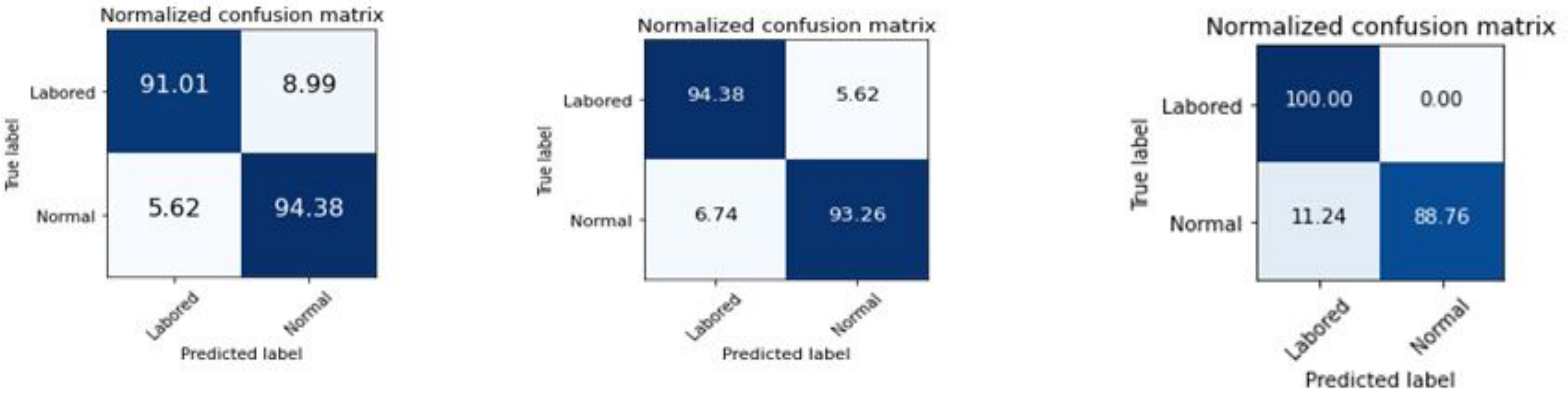
Using the first approach, from 5 different machine learning algorithms, an average of 94% accuracy is achieved.

Performance metrics

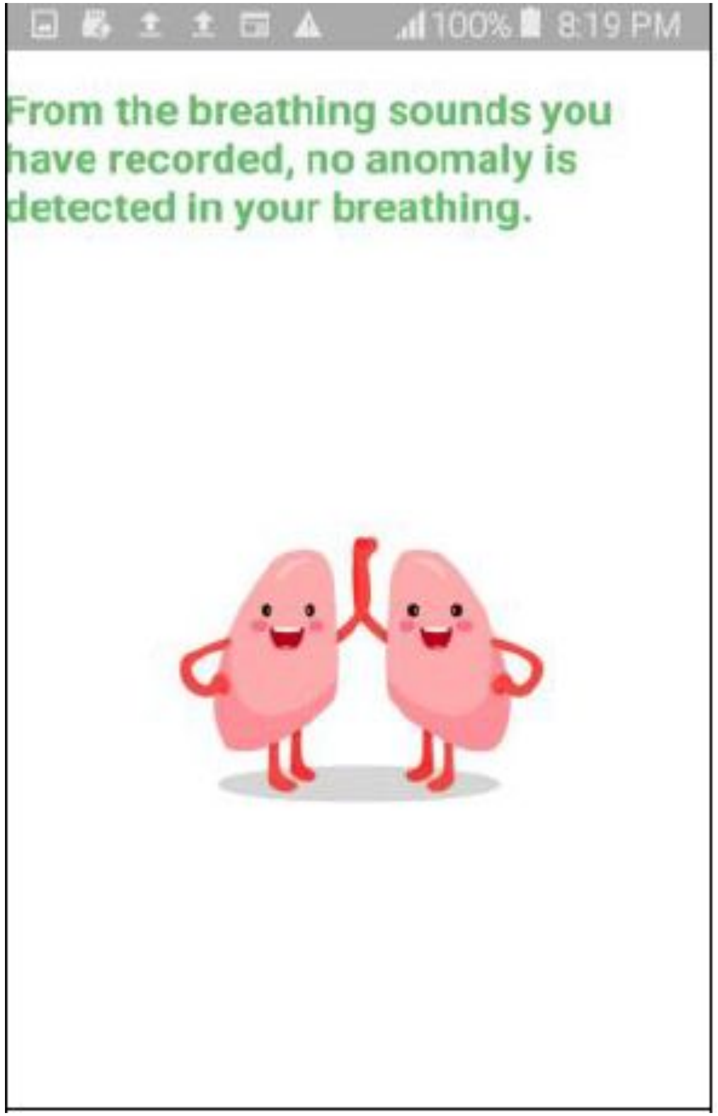
Model	Test accuracy (%)	Precision (%)	Recall (%)
LR	91.35	91.00	91.00
SVM	93.60	93.00	94.00
ANN	94.70	92.00	90.00
RF	91.01	92.00	90.00
KNN	91.50	92.00	90.00

Results (Confusion matrices)

In the second approach, with CNN, an Ensembled CNN, and Gated - CRNN, an accuracy of about 95% is achieved.



Results - User Interface



Future Work

- Fine tuning of the model, by getting feedback from the users.
- On augmenting more data, we wanted to do exact screening of the disease.

THANK YOU

Contact:

Parijat Deshpande - parijat.deshpande@tcs.com
Harini S - 17pd06@psgtech.ac.in

https://sciforum.net/paper/view/conference/8200