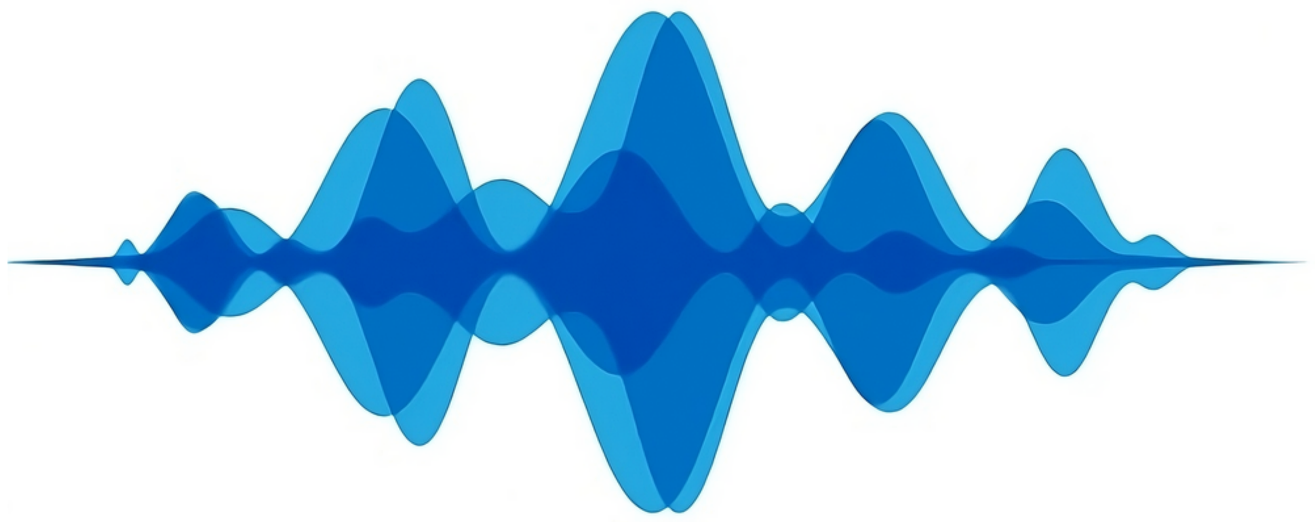
Predicting Respiratory Diseases with Machine Learning

From Respiratory Sound



Hrayr Muradyan, Armen Nalbandyan

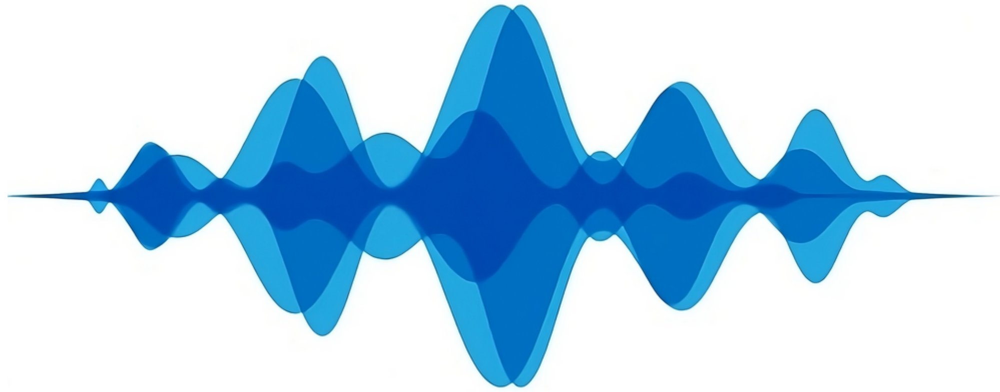
Be a doctor for a minute

Can you find the difference?

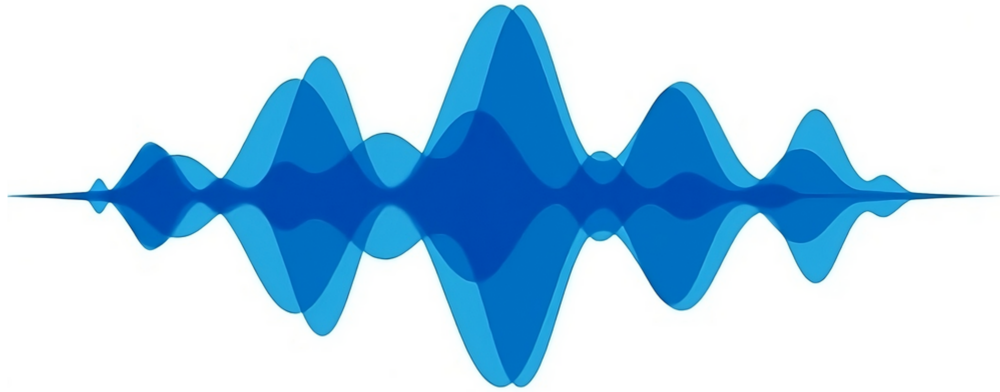
Case 1

Case 2

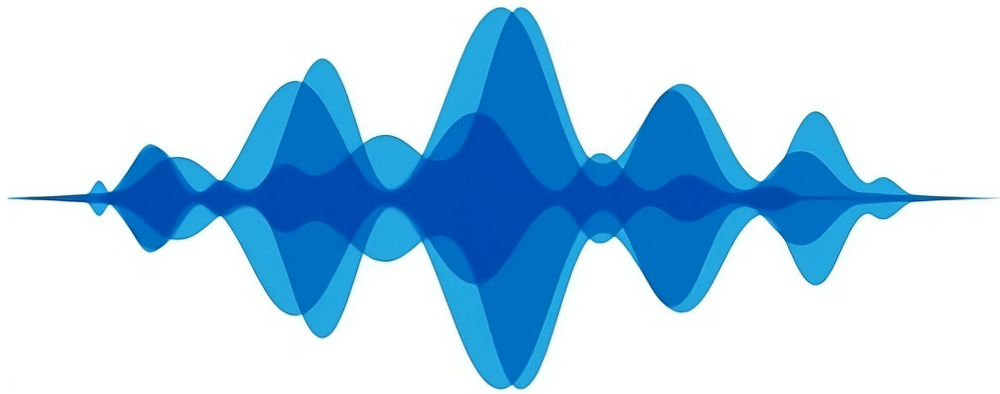
Case 1



Case 1



Case 2



Case 2

Any observations?

Case 1 – ?

Case 2 – ?

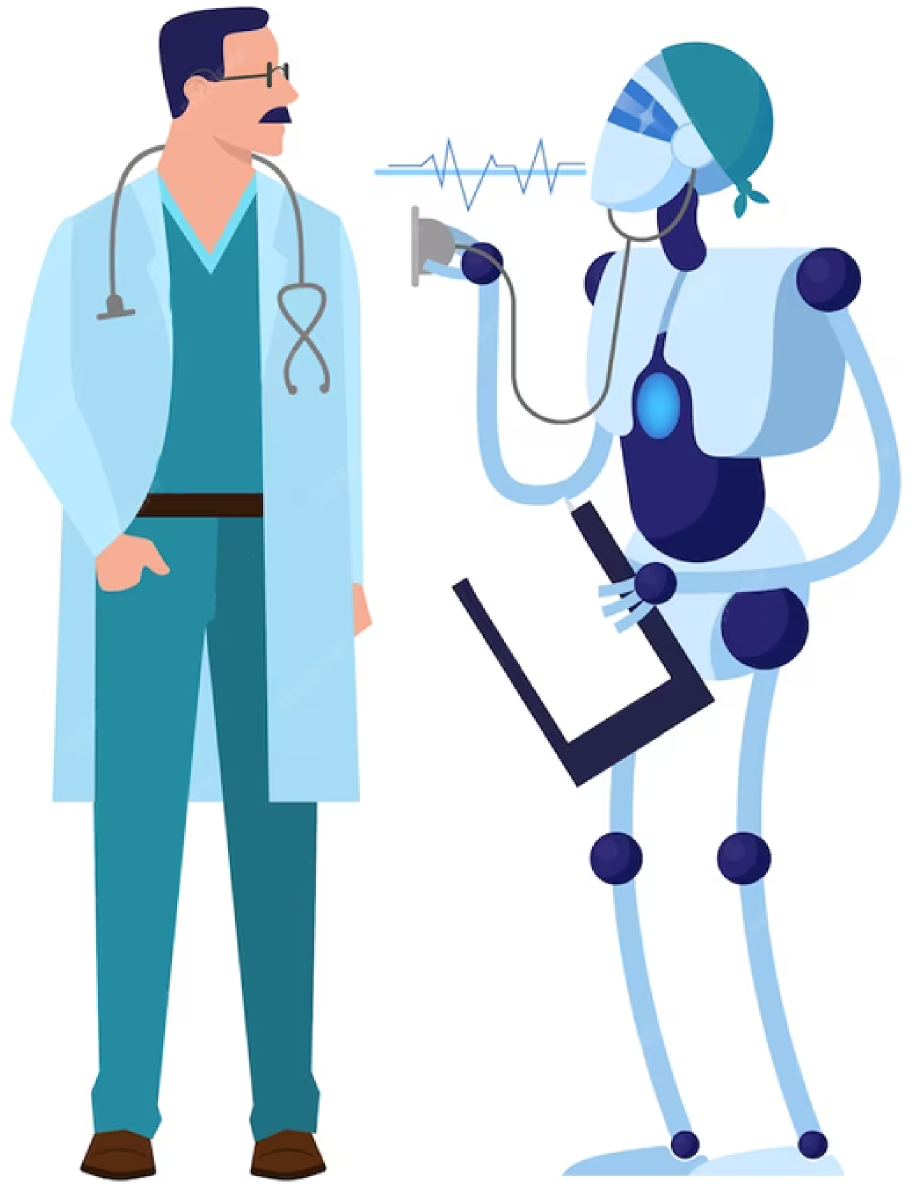
Any observations?

Case 1 - Healthy

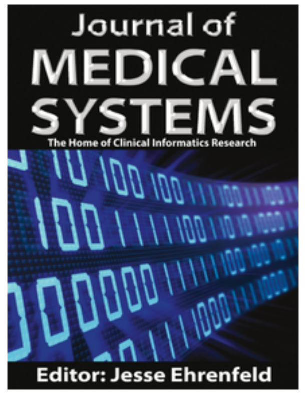
Case 2 - COPD

Motivation

We suggest that while it is hard for humans to identify specific diseases from respiratory sounds, machine learning approaches are very likely to uncover the patterns hidden in the recordings if there are any.



Related Work



Respiratory Sound Based Classification of Chronic Obstructive Pulmonary Disease: a Risk Stratification Approach in Machine Learning Paradigm

Journal of Medical Systems

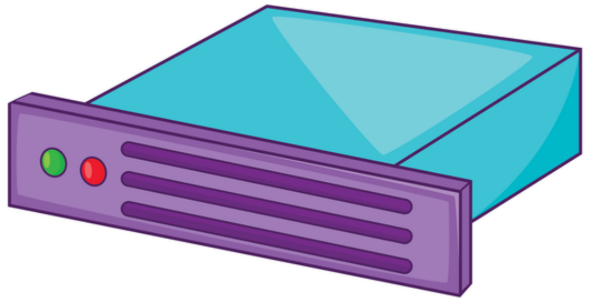
Automatic Crackle Detection Algorithm Based on Fractal Dimension and Box Filtering

Procedia Computer Science

A Respiratory Sound Database for the Development of Automated Classification

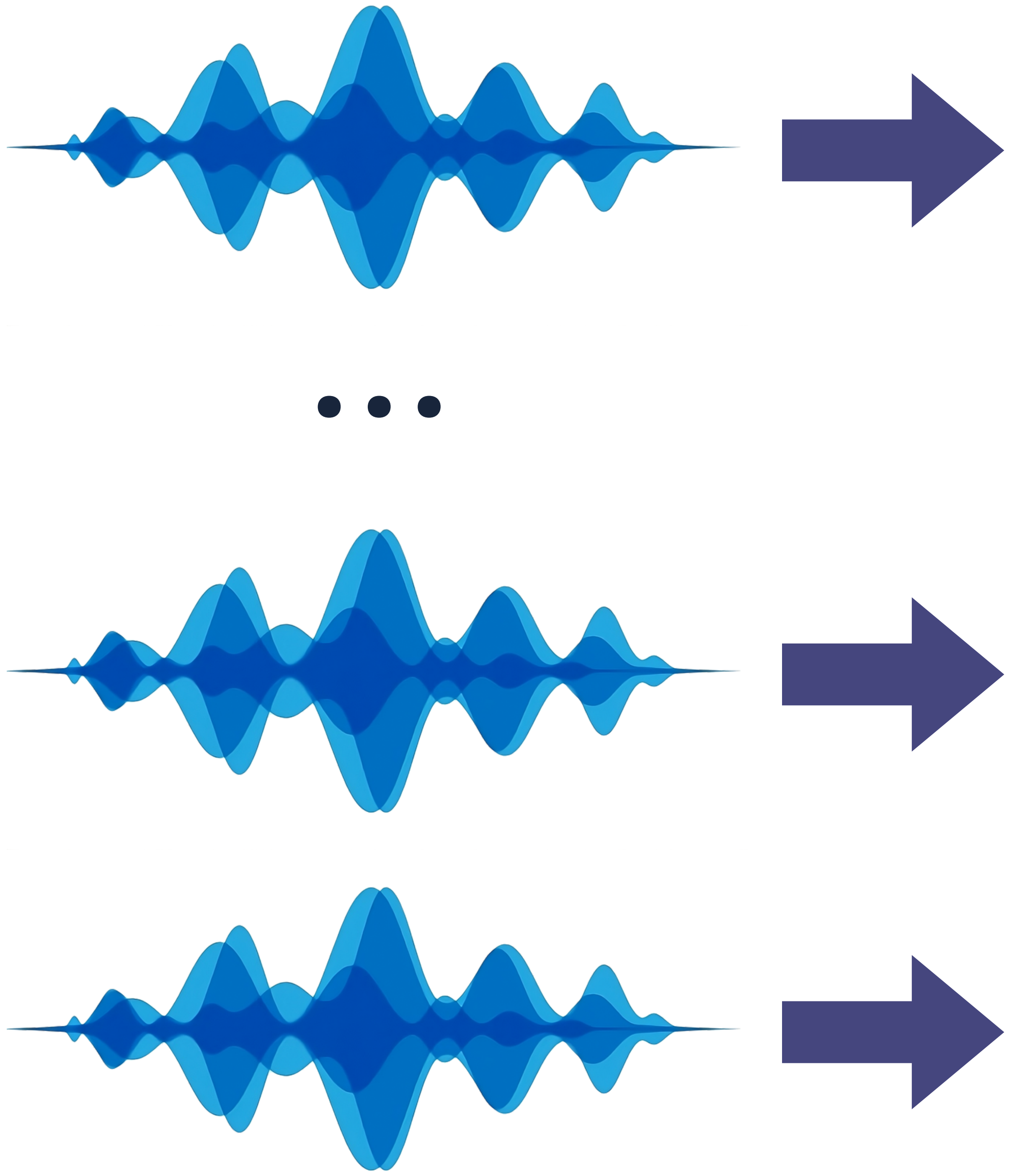
Precision Medicine Powered by pHealth and Connected Health

Data Description

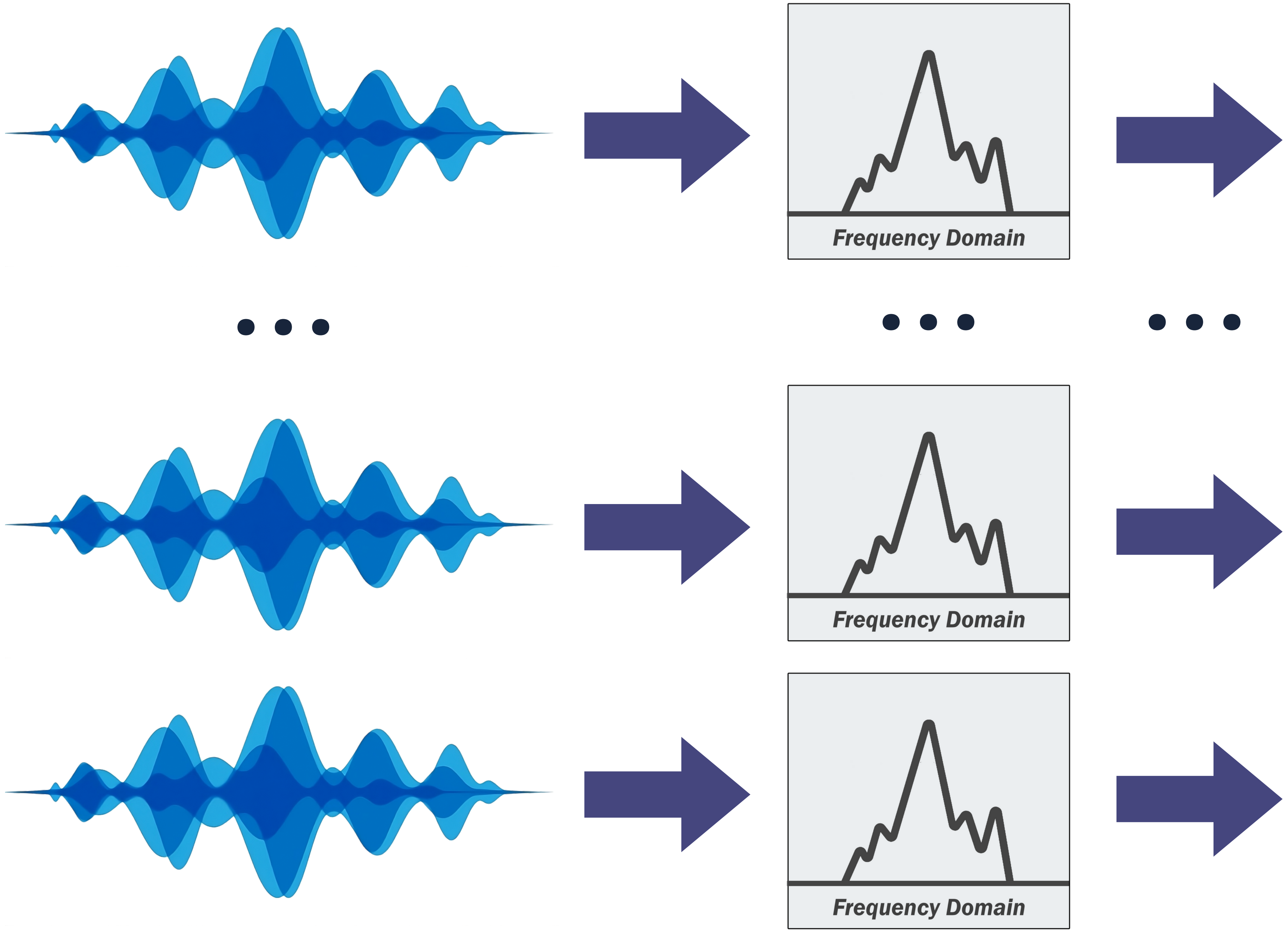


- 920 annotated respiratory recordings of 126 patients taken from digital stethoscopes and other recording techniques,
- Each audio file information includes: Patient number, Chest Location and Recording Equipment (total 4 equipments),
- The annotation file has 4 columns which include: Beginning of respiratory cycle(s), End of respiratory cycle(s), Presence/absence of crackles, Presence/absence of wheezes,
- Separate file for patient diagnosis containing patient number and the respective diagnosis (total 7 diseases and healthy state),
- Separate patient demographic info including Age, Gender, Adult BMI, Child Height (cm), Child Weight (kg).

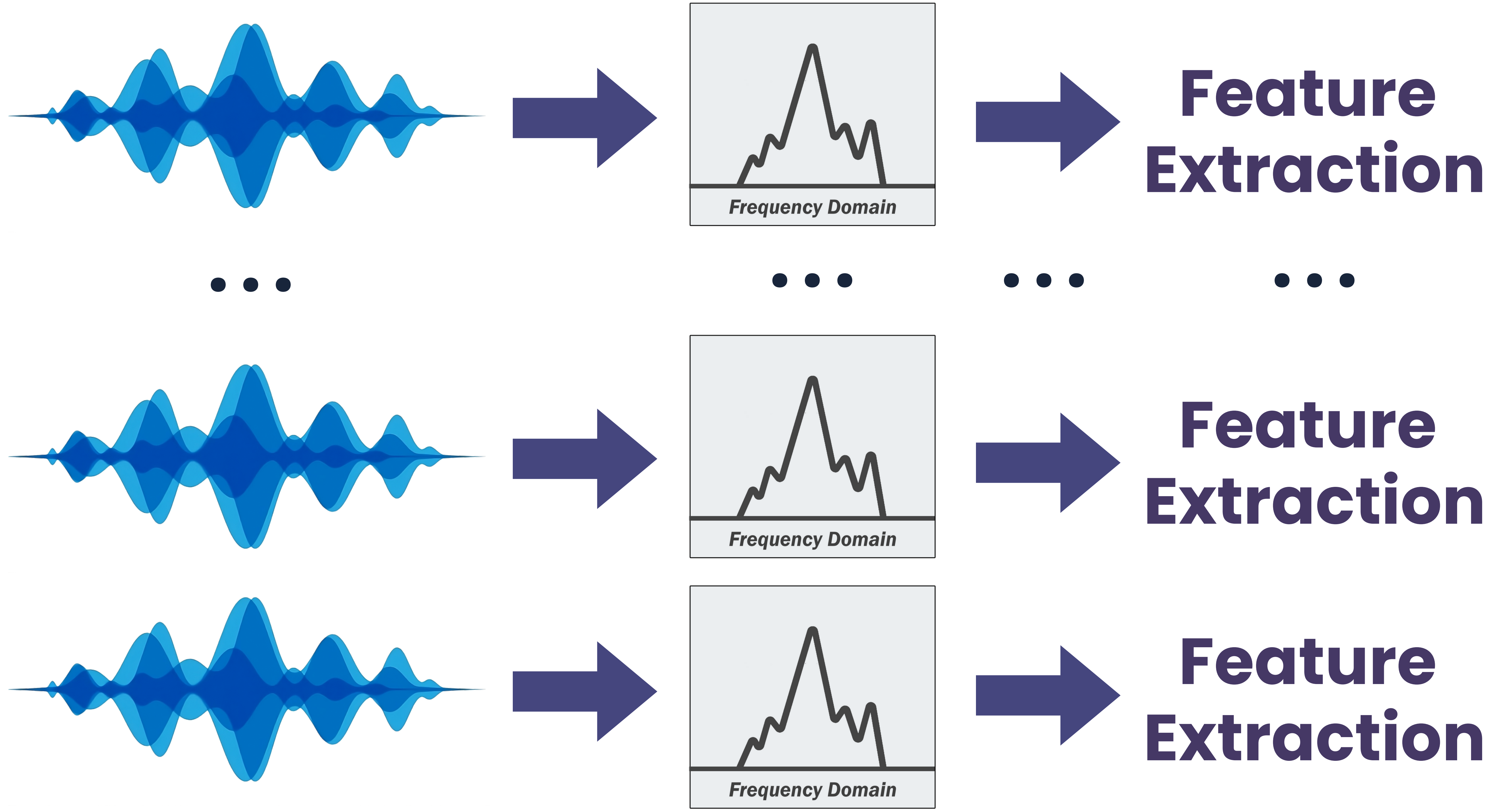
Methodology



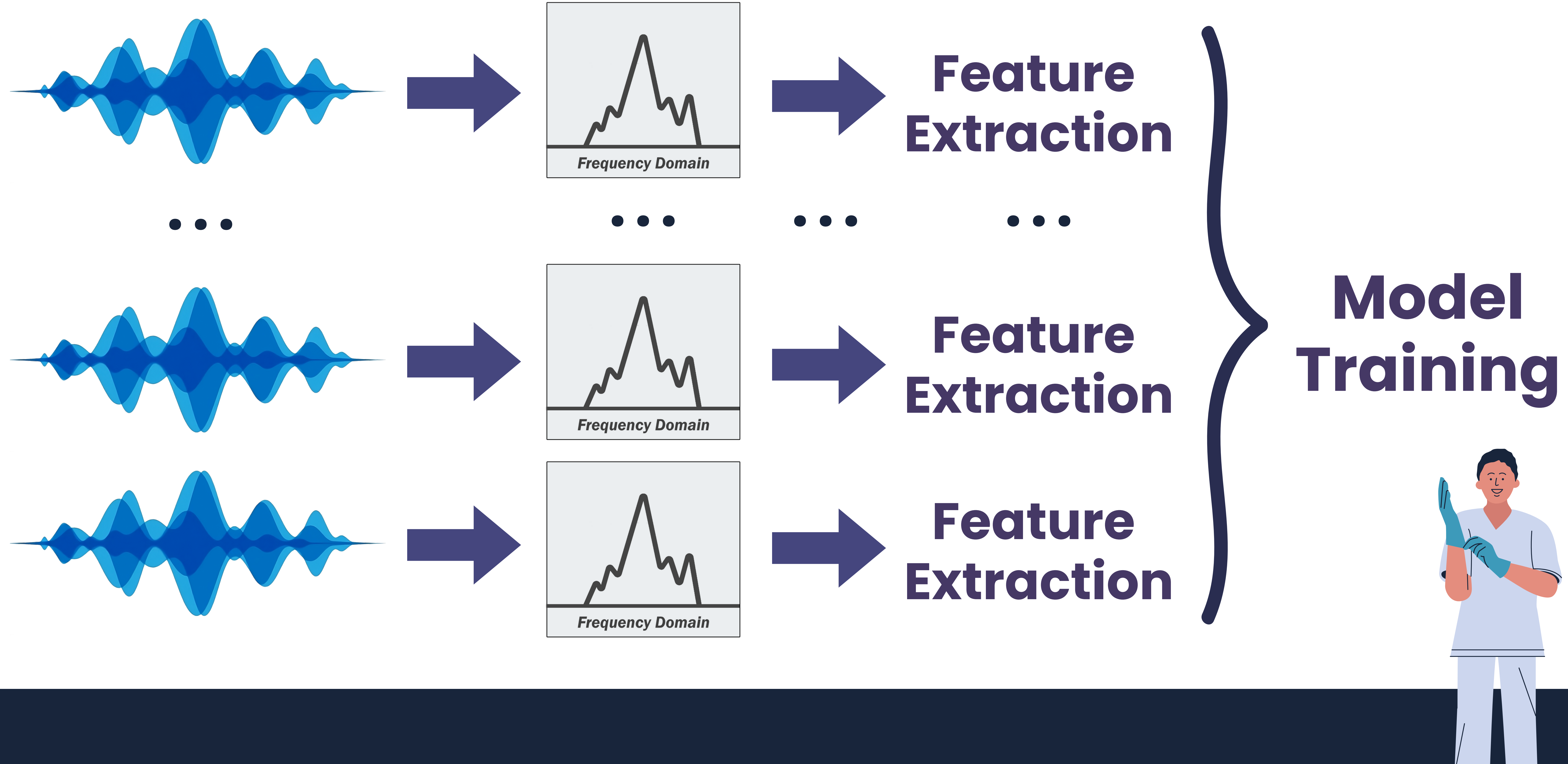
Methodology



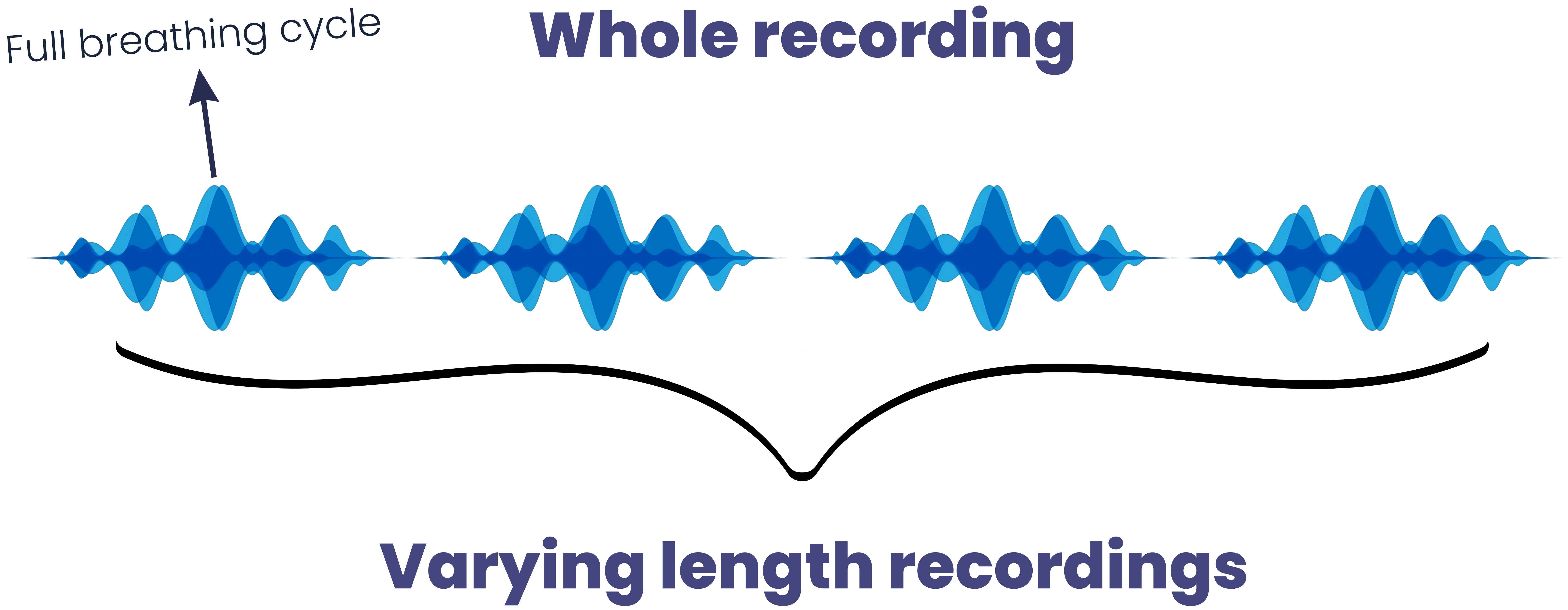
Methodology



Methodology

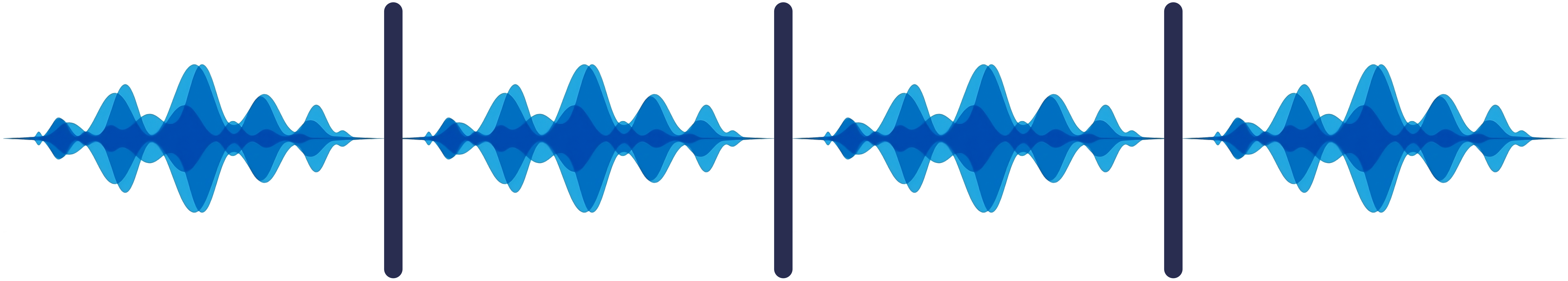


Cycle Extraction



Cycle Extraction

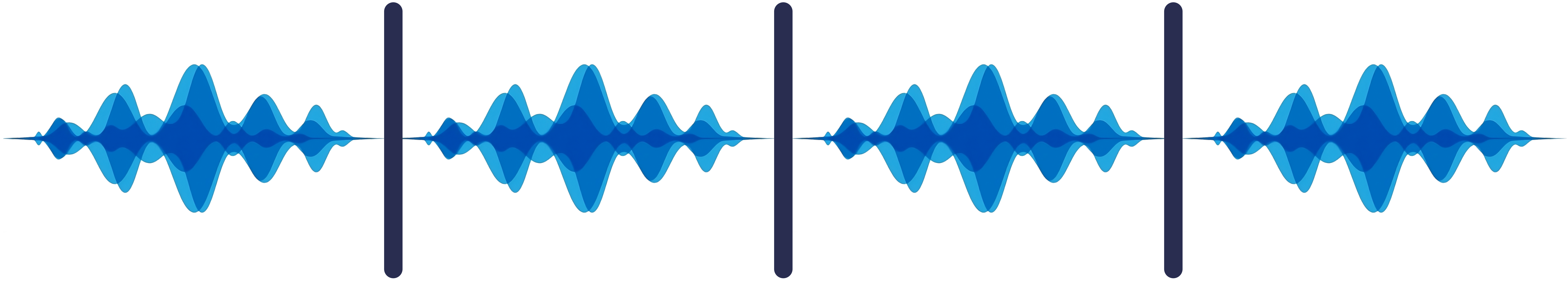
Splitting the whole recording into separate cycles



As mentioned earlier, the annotations data contains each cycle’s starting and ending information

Cycle Extraction

Splitting the whole recording into separate cycles



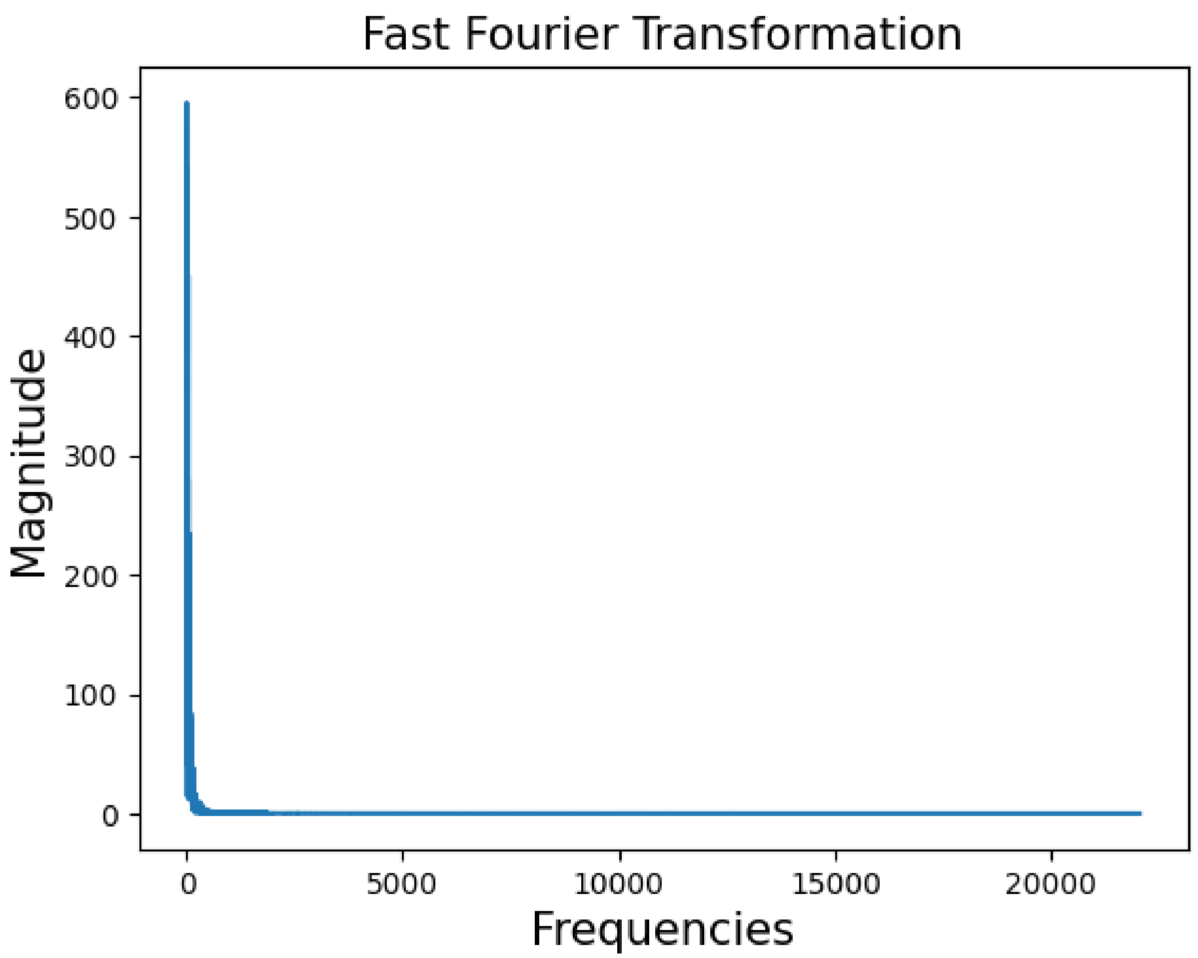
Removing the cycles less than a certain threshold.
(e.g. less than 1 second)

Fast Fourier Transform

1) $$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cdot e^{-j\frac{2\pi}{N}kn}$$

Fast Fourier Transform

1) $$X(k) = \frac{1}{N}\sum_{n=0}^{N-1} x(n) \cdot e^{-j\frac{2\pi}{N}kn}$$

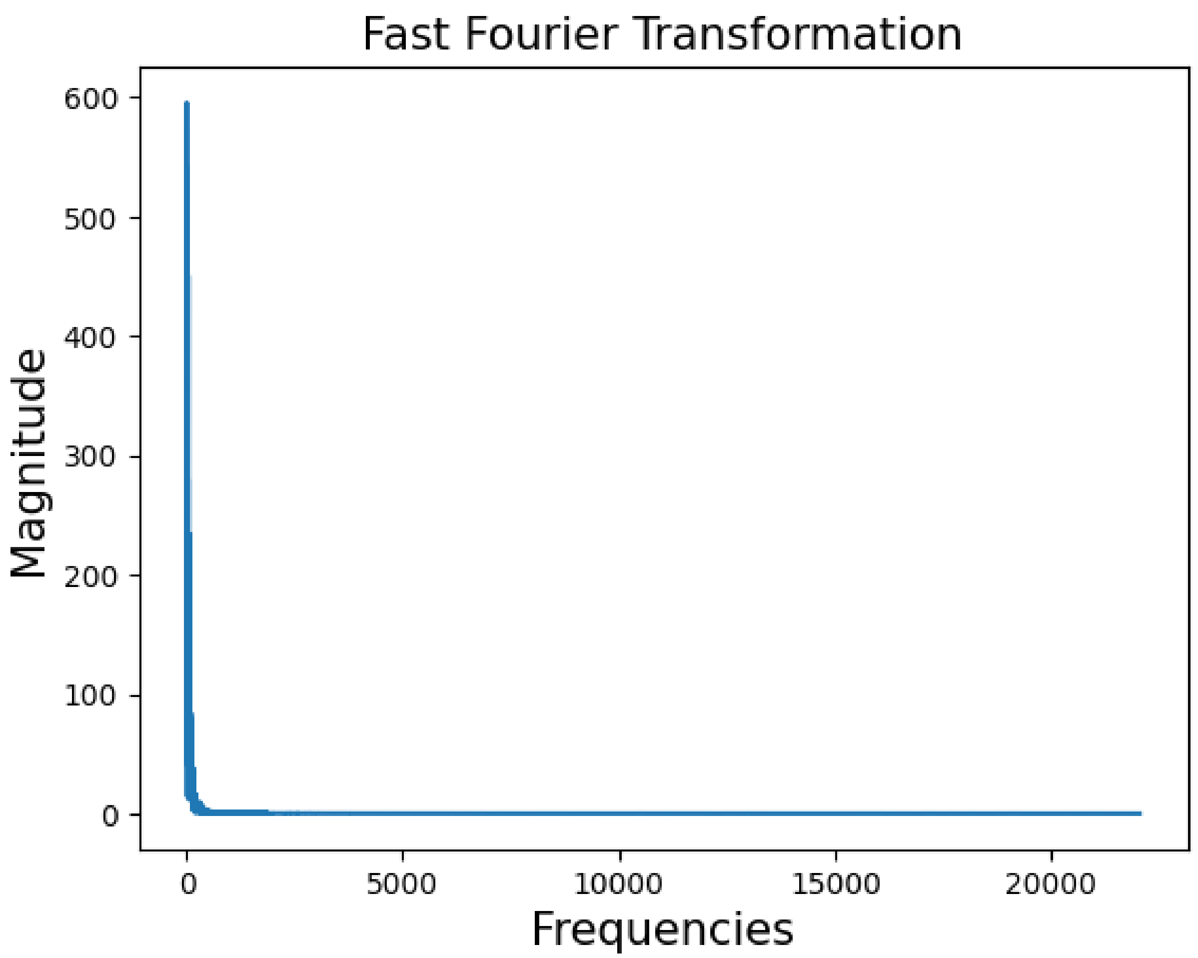
2) 

***Frequencies are calculated up to sampling-rate/2 frequency due to Nyquist theorem**

Fast Fourier Transform

1) $$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cdot e^{-j\frac{2\pi}{N}kn}$$

2)



3) **Magnitudes X Frequencies**

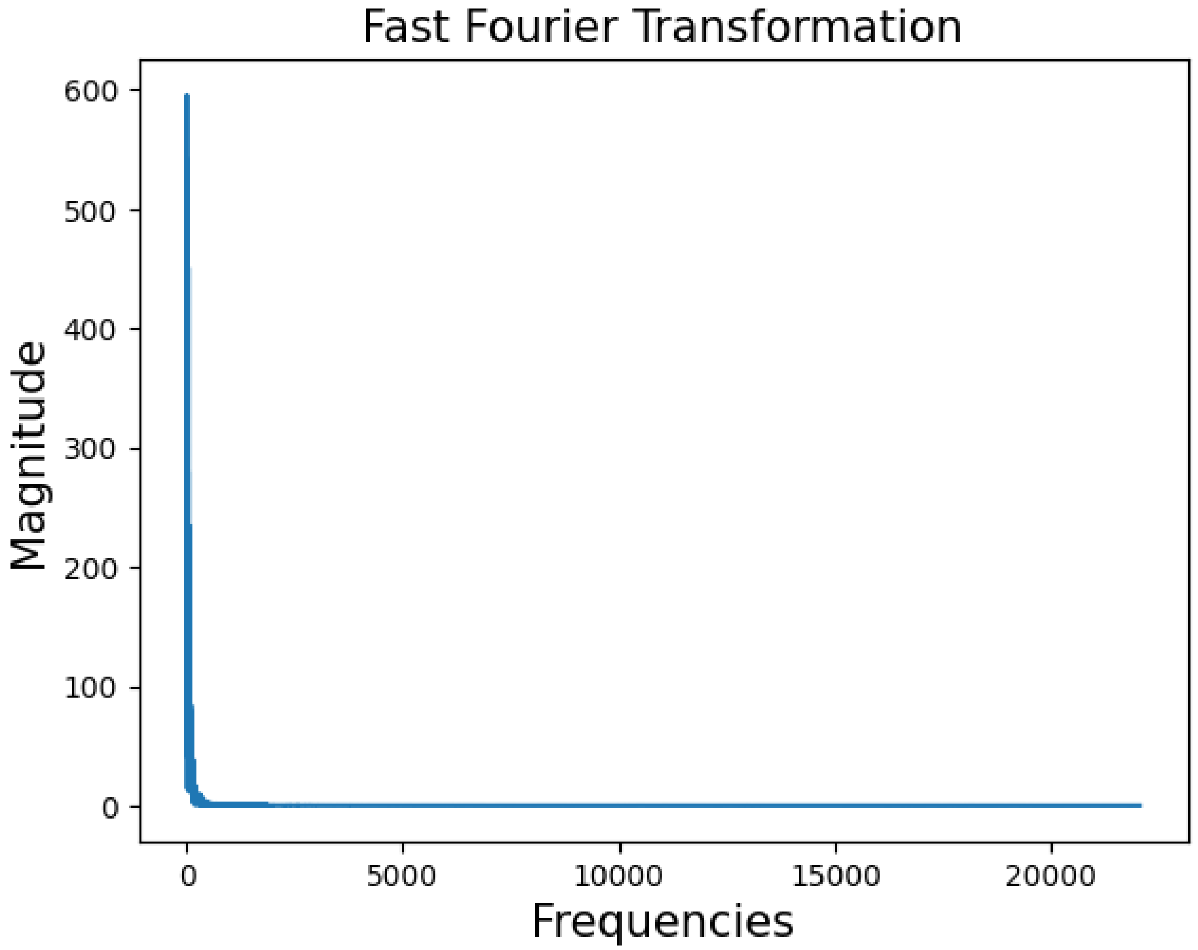
***Frequencies are calculated up to sampling-rate/2 frequency due to Nyquist theorem**

Fast Fourier Transform

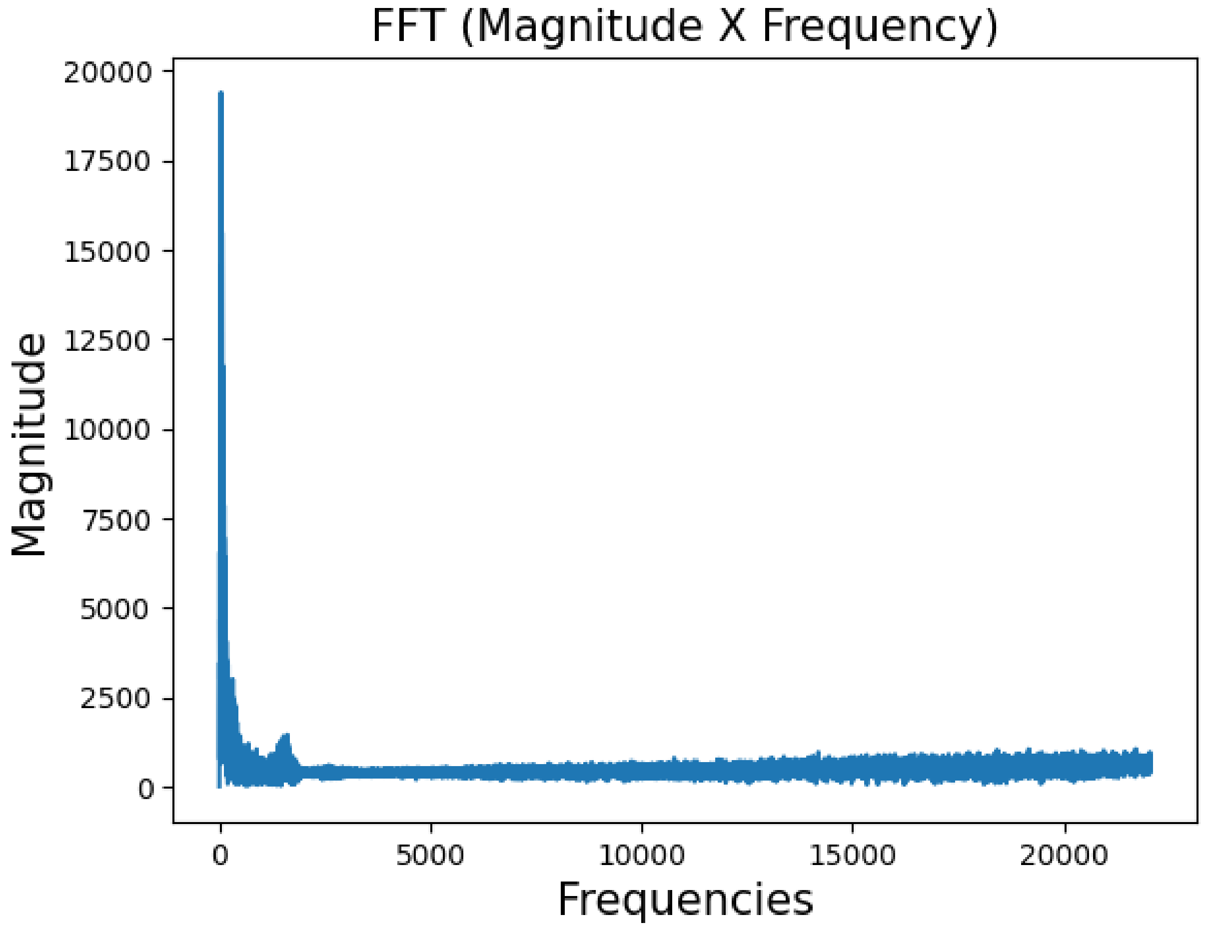
1) $X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cdot e^{-j\frac{2\pi}{N}kn}$

3) Magnitudes X Frequencies

2)

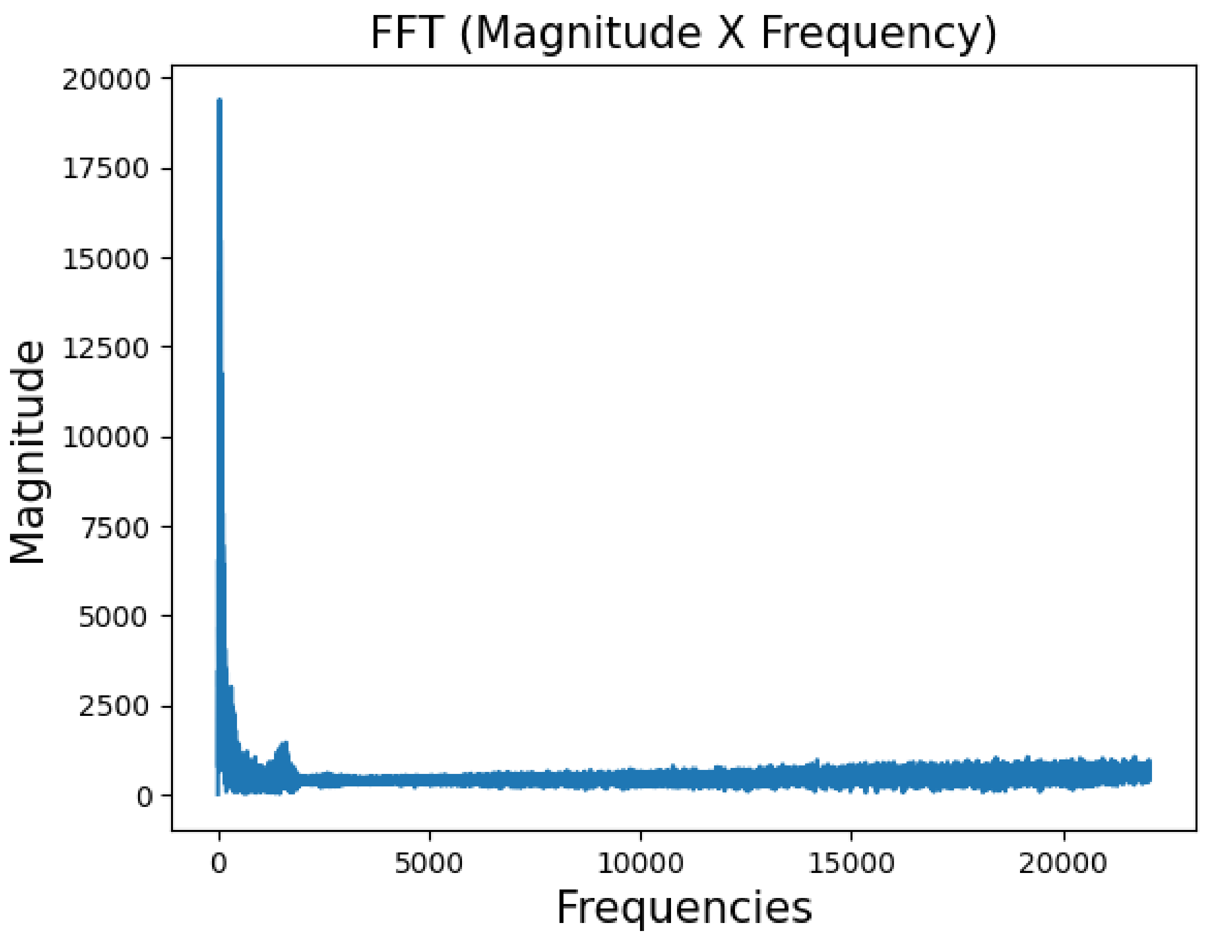


4)



*Frequencies are calculated up to sampling-rate/2 frequency due to Nyquist theorem

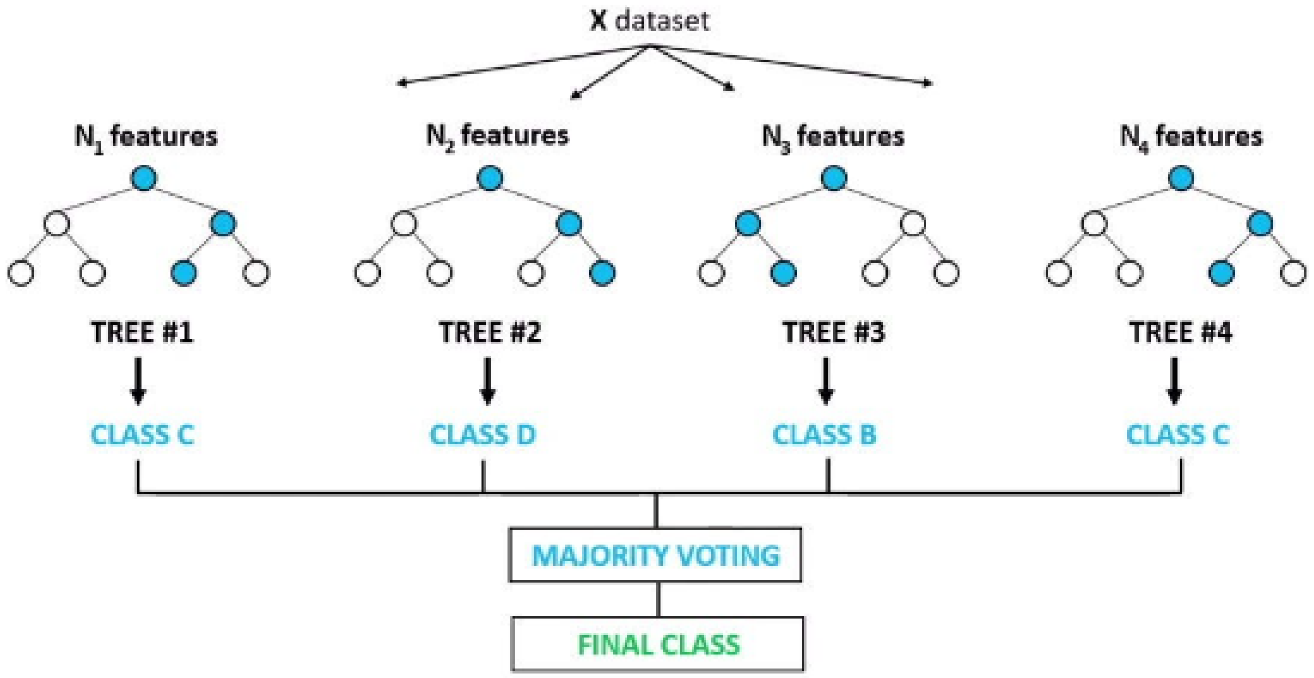
Feature Extraction



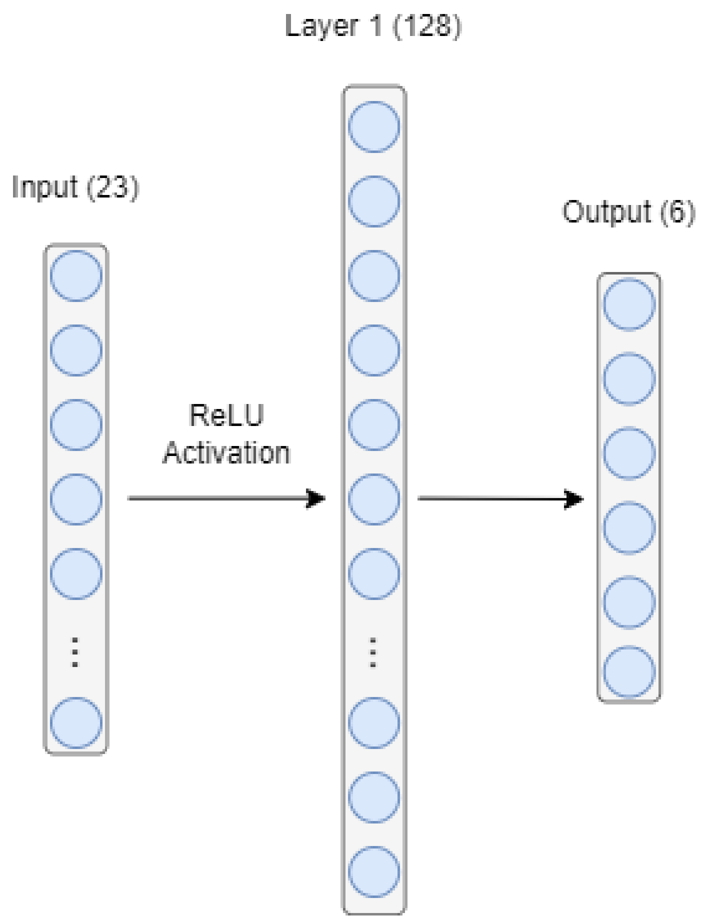
- Maximum
- Maximum Index
- Mean
- Standard Deviation
- Skewness
- Kurtosis

Models Trained

Random Forest Classifier



FC Neural Network

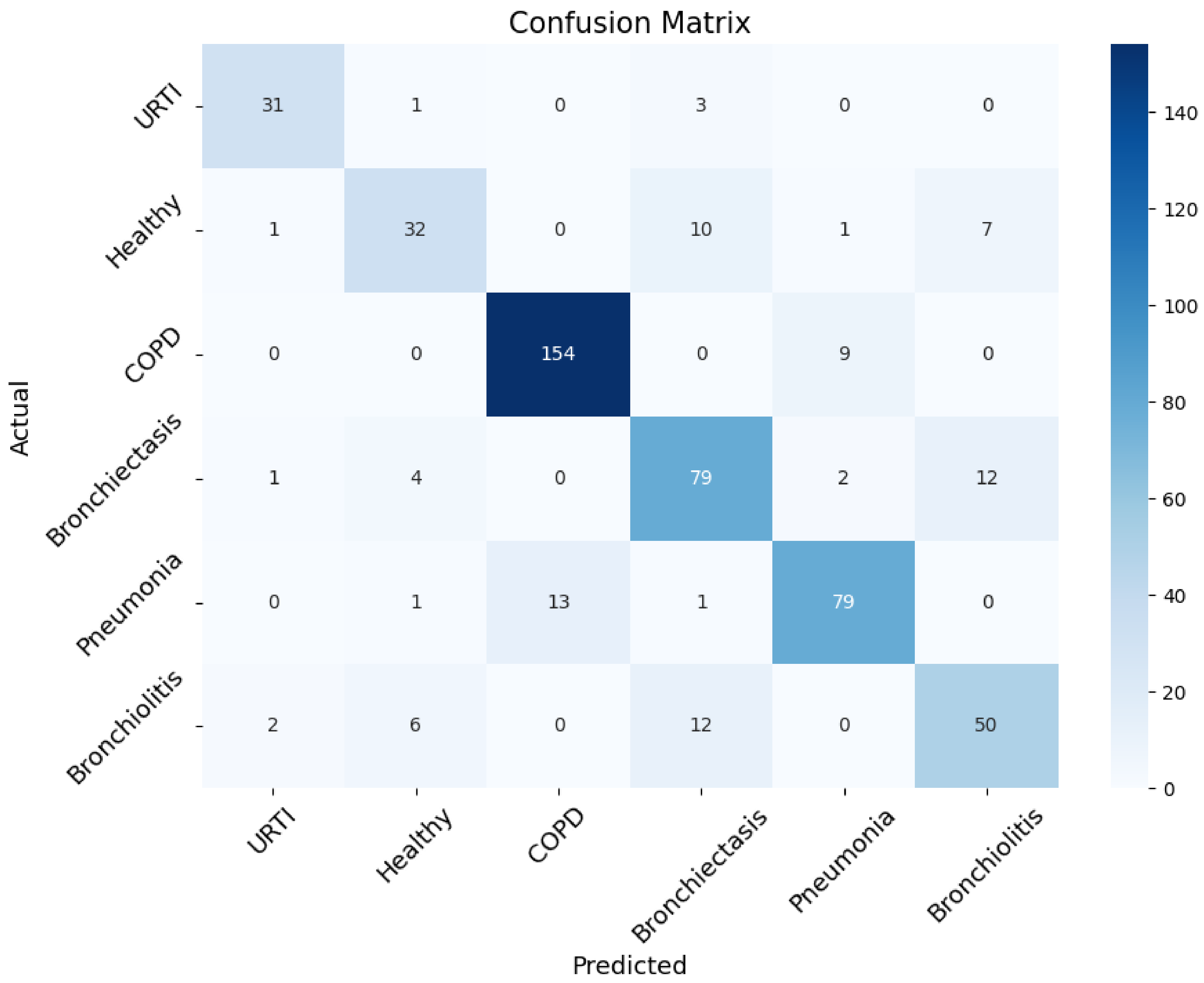


~84%

CORRECT DISEASE CLASSIFICATION RATE

The best results achieved by the FC Neural Network.

Results



Classification Report on test dataset (Macro average)	
Recall	0.803
Precision	0.813
F1-Score	0.807

Machine Learning algorithms proved to work well for respiratory sound disease classification!

Thank you!