

Project Members: Suhani (PES2UG23CS613), Srijan Das (PES2UG23CS600)

Course: Machine Learning

Faculty name: M.Farida Begam

# COVID-19 Cough Detection using Machine Learning

## Problem Statement

COVID-19 detection methods such as RT-PCR are time-consuming and costly. Cough sounds, being a key respiratory symptom, contain identifiable acoustic patterns that can indicate infection. The aim of this project is to build a machine learning based system capable of classifying cough audio samples as COVID-19 positive or healthy, providing a fast, non-invasive pre-screening solution.

## Approach

The project uses the COUGHVID dataset (EPFL, Zenodo), consisting of over 20,000 crowd-sourced cough recordings labeled as COVID-19, healthy, or symptomatic. The approach focuses on extracting relevant acoustic features from cough recordings and applying a supervised learning model to distinguish between COVID-19 and healthy cases.

## Implementation Overview

### 1. Data Preparation

- Downloaded and unzipped the public COUGHVID dataset.
- Converted .webm audio files to .wav using PyDub for compatibility.
- Filtered metadata to include only COVID-19 and healthy samples.

### 2. Feature Extraction

- Extracted audio features using Librosa: MFCCs, Mel-Spectrogram, Spectral Contrast.
- Combined features into numerical arrays using NumPy.

### 3. Model Training and Evaluation

- Applied StandardScaler for feature normalization.
- Trained a Support Vector Machine (SVM) with RBF kernel.
- Evaluated using Accuracy, Precision, Recall, F1-score, and Confusion Matrix.
- Achieved approximately 85-90% accuracy on test data.

## Conclusions and Challenges

The SVM-based approach successfully differentiates COVID-19 coughs from healthy ones using extracted acoustic features. The model performs well despite limited positive samples and noisy real-world audio. Challenges faced included dataset imbalance, environmental noise, and variations in recording quality. Future work includes integrating deep learning models (CNNs, YAMNet) and deploying a mobile app for real-time detection.