Assignment

Develop a signal processing pipeline and algorithm that classifies heartbeat audio recordings into predefined categories using the provided dataset of heartbeat sounds.

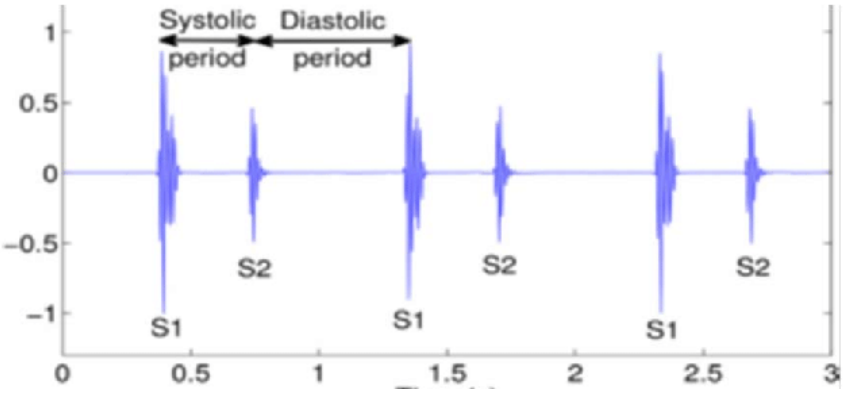
There are three different datasets: healthy, artifact, murmur.

You will have 5 days to complete the assignment from the point of being able to access it. Feel free to submit before. There is no bonus for submitting early.

Objectives

There are two parts to the assignment.

1. Extract features from raw audio files. i.e. a part of your approach should identify cardiac events (S1, S2) in a .wav file.
2. Develop an algorithm or a model that can differentiate between recordings of healthy heartbeats, ones with murmurs and recordings containing artifacts. Clearly document your pipeline, design decisions, and findings.



Deliverables

1. GitHub Repository

We'll provide access to a private GitHub repository. Your submission should include:

- `README.md`:
  - Overview of the repository structure
  - Instructions to set up and run your code
- Code in a language/tools of your choice

2. Technical Report

There's no limit on how long the report has to be. We want to give you an opportunity to convey your thought process to the extent you're satistifed with.

A PDF report that includes:

- Overall approach and design rationale
  - The design rationale is one of the most important sections of this report.
- Data preprocessing steps
- Visualizations of model performance (e.g., confusion matrix, learning curves)
- Key performance metrics (e.g., accuracy, F1-score)
- Challenges encountered and how you addressed them
- Potential improvements and future work
- Any references used

3. Demo Video

- A brief screen recording (max. 5 minutes) showing:
  - Running your pipeline
  - Key output visualizations or algorithm results