Heart Disease Prediction from heart beat audio signals using Machine Learning and Network Analysis

Ligari D. • Alberti A. [1]

[1] *Department of Computer Engineering, Data Science, University of Pavia, Italy*
Course of Advanced Biomedical Machine Learning

Github page: https://github.com/DavideLigari01/advanced-biomedical-project

Date: June 1, 2024

Abstract — —TO BE DEFINED—

Keywords ——TO BE DEFINED—

Contents

1 Introduction 1

2 Dataset 1

3 Dataset Description 1

4 Heart Diseases 2
- 4.1 Normal Category 2
- 4.2 Murmur Category 2
- 4.3 Extra Heart Sound Category 2
- 4.4 Artifact Category 2
- 4.5 Extrasystole Category 2

5 Goals 2

6 Feature Extraction 2
- 6.1 Features Type 2
- 6.2 Methodology 3
- 6.3 Sample Rate and Interval Selection 3

7 Methodology 3
- 7.1 Sampling Rate and Interval Choice 3
  - 7.1.1 Findings on Sampling Rate 3
  - 7.1.2 Findings on Extraction Interval 4
- 7.2 Results 4

8 Feature Selection 4
- 8.1 Optimal number of features for each type of feature 4
- 8.2 Correlation analysis 4
- 8.3 Results 4

9 Model Selection 4
- 9.1 Model on specific classes 4
  - 9.1.1 Explanation of the models 4
  - 9.1.2 Methodology 4
  - 9.1.3 Results 4
- 9.2 Models on all classes 4
  - 9.2.1 Explanation of the models 4
  - 9.2.2 Methodology 4
  - 9.2.3 Results 4
- 9.3 Best model analysis 4

10 Analysis of the Best Model 4

11 Conclusion 4

12 Future Works 4

13 Appendix 6

1. Introduction

2. Dataset

3. Dataset Description

The dataset for this project was obtained from a Kaggle repository titled *Dangerous Heartbeat Dataset (DHD)* [1], which in turn sources its data from the PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) [2]. This dataset comprises audio recordings of heartbeats, categorized into different types of heart sounds. Specifically, the dataset consists of 5 types of recordings: Normal Heart Sounds, Murmur Sounds, Extra Heart Sounds, Extrasystole Sounds, and Artifacts. Data has been gathered from the general public via the iStethoscope Pro iPhone app and from a clinic trial in hospitals using the digital stethoscope DigiScope.

4. Heart Diseases

4.1. Normal Category

In the Normal category, there are normal, healthy heart sounds. These may contain noise in the final second of the recording as the device is removed from the body. They may contain a variety of background noises (from traffic to radios). They may also contain occasional random noise corresponding to breathing, or brushing the microphone against clothing or skin. A normal heart sound has a clear "lub dub, lub dub" pattern, with the time from "lub" to "dub" shorter than the time from "dub" to the next "lub" (when the heart rate is less than 140 beats per minute). Note the temporal description of "lub" and "dub" locations over time in the following illustration:

In medicine, we call the lub sound "S1" and the dub sound "S2". Most normal heart rates at rest will be between about 60 and 100 beats ("lub dub"s) per minute. However, note that since the data may have been collected from children or adults in calm or excited states, the heart rates in the data may vary from 40 to 140 beats or higher per minute. Dataset B also contains noisy_normal data normal data which includes a substantial amount of background noise or distortion. You may choose to use this or ignore it, however, the test set will include some equally noisy examples.

4.2. Murmur Category

Heart murmurs sound as though there is a "whooshing, roaring, rumbling, or turbulent fluid" noise in one of two temporal locations: (1) between "lub" and "dub", or (2) between "dub" and "lub". They can be a symptom of many heart disorders, some serious. There will still be a "lub" and a "dub". One of the things that confuses non-medically trained people is that murmurs happen between lub and dub or between dub and lub; not on lub and not on dub. Below, you can find an asterisk* at the locations a murmur may be.

Dataset B also contains noisy_murmur data murmur data which includes a substantial amount of background noise or distortion. You may choose to use this or ignore it, however, the test set will include some equally noisy examples.

4.3. Extra Heart Sound Category

Extra heart sounds can be identified because there is an additional sound, e.g., a "lub-lub dub" or a "lub dub-dub". An extra heart sound may not be a sign of disease. However, in some situations, it is an important sign of disease, which if detected early could help a person. The extra heart sound is important to be able to detect as

4.4. Artifact Category

In the Artifact category, there are a wide range of different sounds, including feedback squeals and echoes, speech, music, and noise. There are usually no discernible heart sounds, and thus little or no temporal periodicity at frequencies below 195 Hz. This category is the most different from the others. It is important to be able to distinguish this category from the other three categories, so that someone gathering the data can be instructed to try again.

4.5. Extrasystole Category

Extrasystole sounds may appear occasionally and can be identified because there is a heart sound that is out of rhythm involving extra or skipped heartbeats, e.g., a "lub-lub dub" or a "lub dub-dub". (This is not the same as an extra heart sound as the event is not regularly occurring.) An extrasystole may not be a sign of disease. It can happen normally in an adult and can be very common in children. However, in some situations, extrasystoles can be caused by heart diseases. If these diseases are detected earlier, then treatment is likely to be more effective. Below, note the temporal description of the extra heart sounds:

5. Goals

6. Feature Extraction

As demonstrated by [4] and [3], MFCCs are highly effective features for heartbeat classification. In addition to MFCCs, we incorporated other features to capture various characteristics of heart sounds, enhancing the classification accuracy. The features used are explained in the following section.

6.1. Features Type

MFCC

Mel-Frequency Cepstral Coefficients (MFCCs) are representations of the short-term power spectrum of sound. They are derived by taking the Fourier transform of a signal, mapping the powers of the spectrum onto the mel scale, taking the logarithm, and then performing a discrete cosine transform. MFCCs are effective in capturing the timbral texture of audio and are widely used in speech and audio processing due to their ability to represent the envelope of the time power spectrum. In heartbeat classification, MFFCs can reflect the different perceived quality of heart sounds, such as the presence of murmurs or other anomalies.

Chroma STFT

Chroma features represent the 12 different pitch classes of music (e.g., C, C#, D, etc.). They are particularly useful for capturing harmonic and melodic characteristics in music. By mapping audio signals onto the chroma scale, these features can identify pitches regardless of the octave, making them useful for analyzing harmonic content in heart sounds.

RMS

Root Mean Square (RMS) measures the magnitude of varying quantities, in this case, the amplitude of an audio signal. It is a straightforward way to compute the energy of the signal over a given time frame. RMS is useful in audio analysis for detecting volume changes and can help identify different types of heartbeats based on their energy levels. For example, in a given timeframe the RMS may be altered by the presence of a murmur with respect to a normal heart sound.

ZCR
Zero-Crossing Rate (ZCR) is the rate at which a signal changes sign, indicating how often the signal crosses the zero amplitude line. It is particularly useful for detecting the noisiness and the temporal structure of the signal. In heartbeat classification, ZCR can help differentiate between normal and abnormal sounds by highlighting changes in signal periodicity.

CQT
Constant-Q Transform (CQT) is a time-frequency representation with a logarithmic frequency scale, making it suitable for musical applications. Since it captures more detail at lower frequencies, it may be useful for analyzing the low-frequency components of heart sounds.

Spectral Centroid
The spectral centroid indicates the center of mass of the spectrum and is often perceived as the brightness of a sound. It is calculated as the weighted mean of the frequencies present in the signal, with their magnitudes as weights. In heart sound analysis, a higher spectral centroid can indicate sharper, more pronounced sounds, while a lower centroid suggests smoother sounds.

Spectral Bandwidth
Spectral bandwidth measures the width of the spectrum around the centroid, providing an indication of the range of frequencies present. It is calculated as the square root of the variance of the spectrum. This feature helps in understanding the spread of the frequency components in the heart sounds, which can be indicative of different heart conditions.

Spectral Roll-off Spectral roll-off is the frequency below which a certain percentage of the total spectral energy lies. It is typically set at 85% and helps distinguish between harmonic and non-harmonic content. In heartbeat classification, spectral roll-off can be used to differentiate between sounds with a concentrated energy distribution and those with more dispersed energy.

6.2. Methodology

We extracted the features from the audio signals using the *Librosa* library in Python. It is worth to underline four main aspects in the extraction methodology having a direct impact on the results:

- **Normalization**: the audio are loaded using the *torchaudio.load()* function, which normalized the audio signals in the range [-1, 1]. This is important to ensure that the features are on the same scale and to prevent the model from being biased towards features with larger values.
- **Audio Segmentation**: to extract features, we divided the audio recordings into chunks of a given length (in seconds). This segmentation allowed us to analyze the impact of different extraction intervals on model performance, additionally it allow for augmenting the data available.
- **Sampling Rate**: while literature on spoken language often suggests that 16000 Hz is sufficient, it was necessary to assess the best sampling rate for heartbeat sounds specifically. We evaluated two sampling rates to determine the optimal rate for heartbeat sounds.
- **Hop and Window Size**: the hop size determines the number of samples between successive windows, while the window size determines the number of samples considered. Each feature was extracted using the same window length and hop length facilitating a fair assessment of each feature's contribution to the classification task.

The library used for the extraction is *Librosa*, which is a Python package for music and audio analysis.

6.3. Sample Rate and Interval Selection

7. Methodology

7.1. Sampling Rate and Interval Choice

Experimental Setup
We trained various models (Random Forest, Support Vector Machine, Logistic Regression) with different features (Chroma, MFCC_30, MFCC_120, CQT_30, CQT_70), sampling rates (mixed, 4000 Hz), and extraction intervals (0.5s, 1s, 2s, 3s). Features were used as extracted without any additional processing. The models were trained on 80% of the data and tested on the remaining 20%.

7.1.1. Findings on Sampling Rate

Our experiments revealed no clear advantage of using a mix of frequencies over a fixed sample rate, independently of the metric used. Specifically, a fixed sampling rate of 4000 Hz was found to reduce the risk of bias introduction, improve efficiency, and enable the use of a wider variety of features. This sampling rate provided a consistent and reliable basis for feature extraction, as demonstrated in Figure 1.

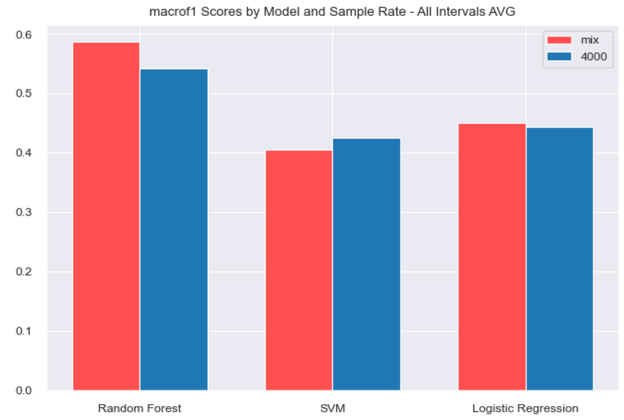

Fig. 1: Comparison of the macro F1 score for different sampling rates.

7.1.2. Findings on Extraction Interval

The choice of extraction interval had a significant impact on the number of samples and the class distribution, as shown in Figure 2. To address the high class imbalance, we used the macro F1 score and Matthews Correlation Coefficient (MCC) as evaluation metrics. Our results indicated that a 2-second interval yielded the best performance. However, this choice also reduced the number of samples, potentially causing issues during training and testing, especially with more complex models.

As a compromise, we recommend using a 1-second interval. This interval offers a good balance between the number of samples and the class distribution, ensuring robust model performance while maintaining a sufficient dataset size. Figures 3 and 4 illustrate the impact of different extraction intervals on the macro F1 score and MCC, respectively.

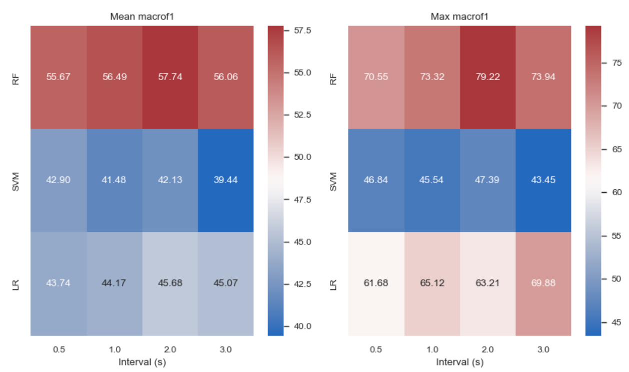

Fig. 3: Comparison of the macro F1 score for different extraction intervals.

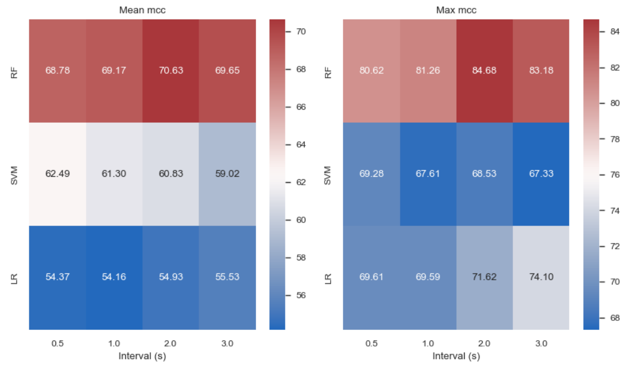

Fig. 4: Comparison of the MCC for different extraction intervals.

7.2. Results

8. Feature Selection

8.1. Optimal number of features for each type of feature

8.2. Correlation analysis

8.3. Results

9. Model Selection

9.1. Model on specific classes

9.1.1. Explanation of the models

9.1.2. Methodology

9.1.3. Results

9.2. Models on all classes

9.2.1. Explanation of the models

9.2.2. Methodology

9.2.3. Results

9.3. Best model analysis

10. Analysis of the Best Model

11. Conclusion

12. Future Works

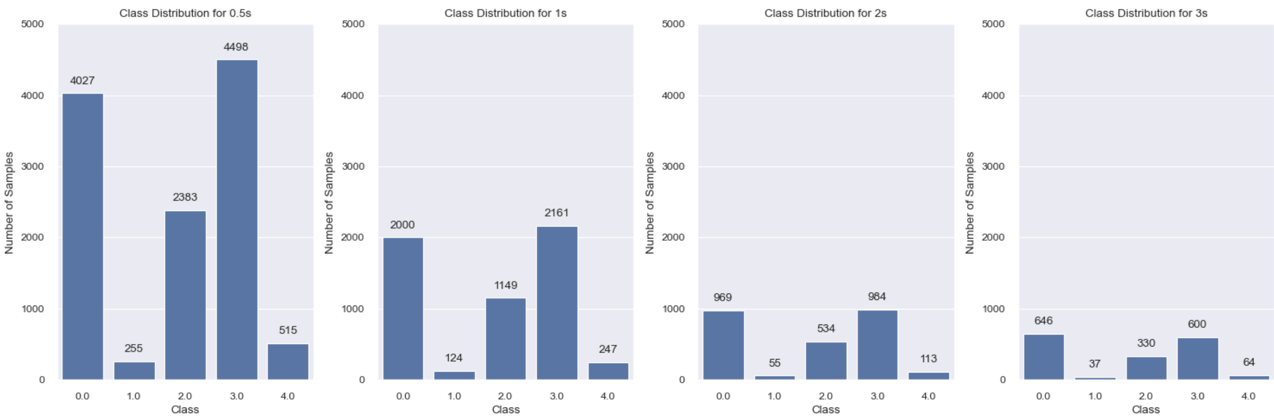

Fig. 2: Class distribution for different extraction intervals.

13. APPENDIX

References

[1] en. URL: https://www.kaggle.com/datasets/mersico/dangerous-heartbeat-dataset-dhd.

[2] P. Bentley et al. *The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results*. URL: http://www.peterjbentley.com/heartchallenge/index.html.

[3] Wei Chen et al. "Deep Learning Methods for Heart Sounds Classification: A Systematic Review". In: *Entropy* 23.6 (May 2021), p. 667. ISSN: 1099-4300. DOI: 10.3390/e23060667.

[4] Ali Raza et al. "Heartbeat Sound Signal Classification Using Deep Learning". en. In: *Sensors* 19.2121 (Jan. 2019), p. 4819. ISSN: 1424-8220. DOI: 10.3390/s19214819.