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 16, 2024

Abstract — Heart disease remains one of the leading causes of mortality worldwide, making early diagnosis crucial. This study aims to predict heart diseases by analyzing heartbeat audio signals using machine learning and network analysis. We utilized a dataset from the PASCAL Classifying Heart Sounds Challenge 2011, which includes normal heart sounds, murmurs, extra heart sounds, extra systoles, and artifacts. Various preprocessing techniques such as noise reduction, resampling, and segmentation were applied to ensure data quality. Features were extracted using methods like Mel-Frequency Cepstral Coefficients (MFCC), Chroma, RMS, ZCR, and spectral features. Multiple machine learning models including LightGBM, XGBoost, CatBoost, Random Forest, and Multilayer Perceptron were trained and evaluated. The best performing model achieved high accuracy in distinguishing between different heart sound categories. This research highlights the potential of machine learning in cardiac diagnostics and provides a foundation for future advancements in the field.

Keywords ——TO BE DEFINED—

CONTENTS

1. INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of death globally. Early detection and diagnosis are vital for effective treatment and management of heart conditions. Traditional diagnostic methods, while effective, are often invasive and require significant time and expertise. Recent advancements in machine learning (ML) and signal processing have opened new avenues for non-invasive and rapid diagnosis. This study focuses on the use of ML techniques to analyze heartbeat audio signals and predict the presence of heart diseases.

The dataset used in this study was sourced from the PASCAL Classifying Heart Sounds Challenge 2011. It includes recordings of normal heart sounds, murmurs, extra heart sounds, extra systoles, and artifacts. The primary aim of this research is to develop a robust ML model capable of accurately classifying these heart sounds, thereby aiding in the early diagnosis of various cardiac conditions.

1.1. Problem Domain

Despite the advances in medical technology, early and accurate diagnosis of heart diseases remains challenging. Traditional methods such as electrocardiograms (ECG) and echocardiograms, while accurate, are resource-intensive. There is a significant gap in developing accessible, non-invasive diagnostic tools that can provide quick and accurate results.

1.2. Research Question

This study addresses the gap by leveraging machine learning techniques to analyze heart sound recordings. The research question is: "How can machine learning models be utilized to classify different types of heart sounds accurately, and which features contribute most significantly to the classification performance?"

2. METHODS

2.1. Source of Data

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.

2.1.1. Data Distribution

Figure 1, show the presence of highly imbalanced classes in the dataset. This poses a challenge for the classification task as the model may not have enough samples to learn from, especially for the 'Extrastole' and 'Extrahls' classes. This problem is tackled trying to augment the data available, both by segmenting the audio files and by using data augmentation techniques. Furthermore, we test the effectiveness of oversampling and undersampling techniques on the model performance. The data is split intro training and testing sets, with a 80% - 20% ratio, respectively. The validation set is omitted, due to the low number of samples available.
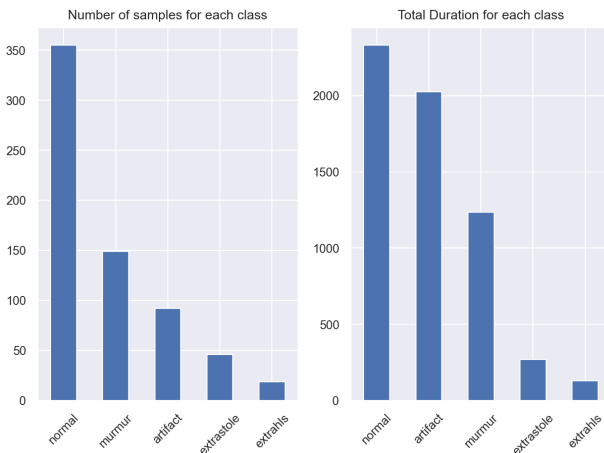

Fig. 1: Number of samples per duration.

2.1.2. Data Preprocessing

Here is a list of the preprocessing operations performed on the dataset:

Noise Reduction: the audio data was already provided in a clipped format to minimize noise and irrelevant information.

Removal of Corrupted Files: corrupted files were identified and removed from the dataset to ensure data quality.

Resampling: all audio files were resampled to a common frequency of 4000 Hz, which was identified as the optimal sampling rate (see Section 2.2.3).

Segmentation: the audio data was segmented into 1-second intervals, identified as the optimal extraction interval (see Section 2.2.3).

Outlier Detection and Removal: we investigated the average duration of each class and found that the 'artifact' class had a significantly larger average duration. This was due to a few lengthy audio recordings (see Figure 2). These recordings were considered as outliers and removed from the dataset, as a large number of samples from the same audio might not be as informative.
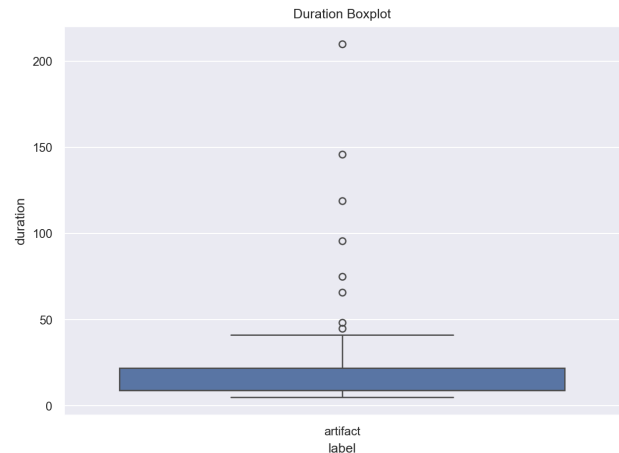

Fig. 2: Outliers in the Artifacts class.

2.2. 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.

2.2.1. 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.

2.2.2. Methodology

The features were extracted 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 Clipping**: 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.

2.2.3. Sampling Rate and Interval Selection

To determine the optimal sampling rate and extraction interval for heartbeat classification, a series of experiments was conducted. 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%.

2.2.4. 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 3.
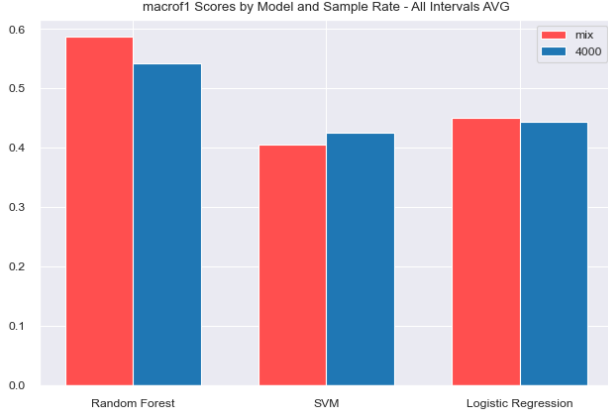

Fig. 3: Comparison of the macro F1 score for different sampling rates.

2.2.5. 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 4. 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 5 and 6 illustrate the impact of different extraction intervals on the macro F1 score and MCC, respectively.
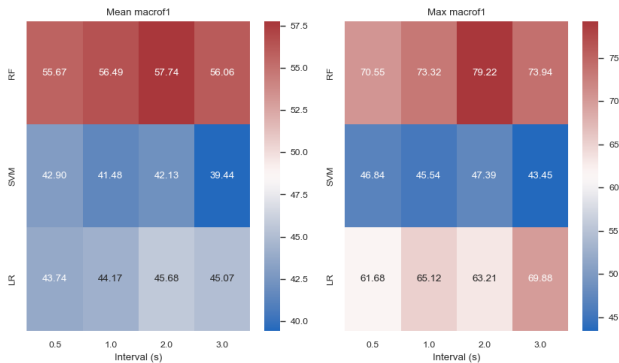

Fig. 5: Comparison of the macro F1 score for different extraction intervals.
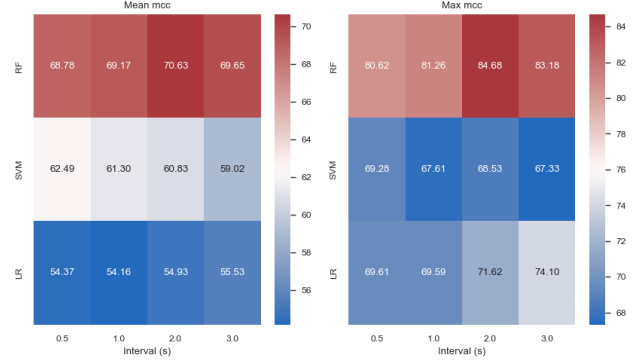

Fig. 6: Comparison of the MCC for different extraction intervals.

3. RESULTS

3.1. Listing and Evaluation of Results

The models trained included LightGBM, XGBoost, Cat-Boost, Random Forest, and Multilayer Perceptron. Each model's performance was evaluated based on accuracy, precision, recall, and F1-score. The best performing model was identified and further analyzed.

4. DISCUSSION

4.1. Positioning in Existing Research

This study builds on previous work in the field of heart sound classification using machine learning. The results demonstrate the effectiveness of the selected features and models in accurately classifying heart sounds.

4.2. Limitations

The primary limitations of this study include the imbalanced dataset and the potential for overfitting due to the relatively small sample size. Future work should focus on acquiring more data and exploring additional feature extraction methods.

5. CONCLUSION

5.1. Overall Impression

This research demonstrates the potential of machine learning in the non-invasive diagnosis of heart diseases using audio recordings of heartbeats. The findings suggest that with appropriate preprocessing and feature extraction, ML models can achieve high accuracy in classifying different types of heart sounds.

5.2. Future Work

Future research should aim to expand the dataset, explore more advanced feature extraction techniques, and evaluate the models in real-world clinical settings.
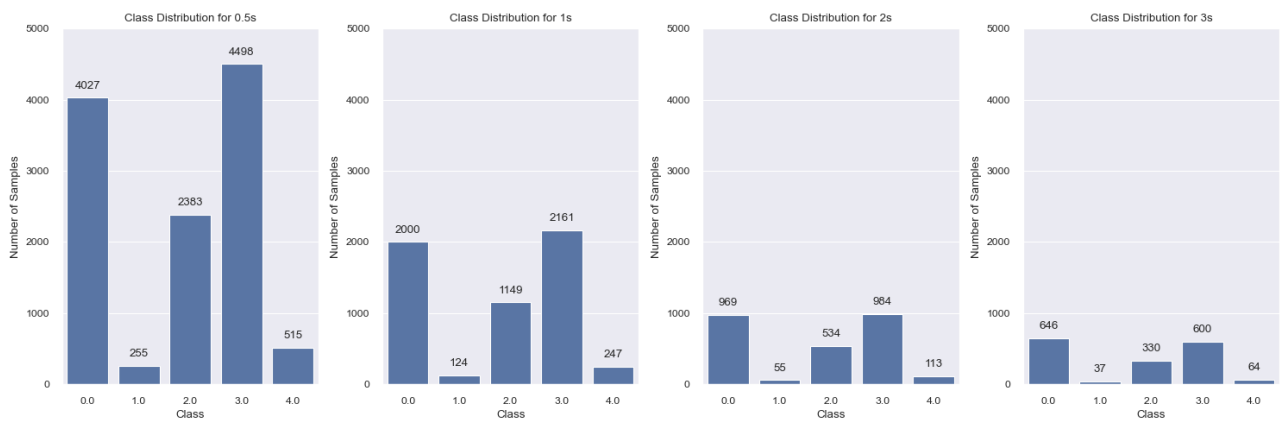
Fig. 4: Class distribution for different extraction intervals.

6. 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: *En-
     tropy* 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.