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 2, 2024

Abstract — —TO BE DEFINED—

Keywords ——TO BE DEFINED—

Contents

1. Introduction

2. 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.

2.1. 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.

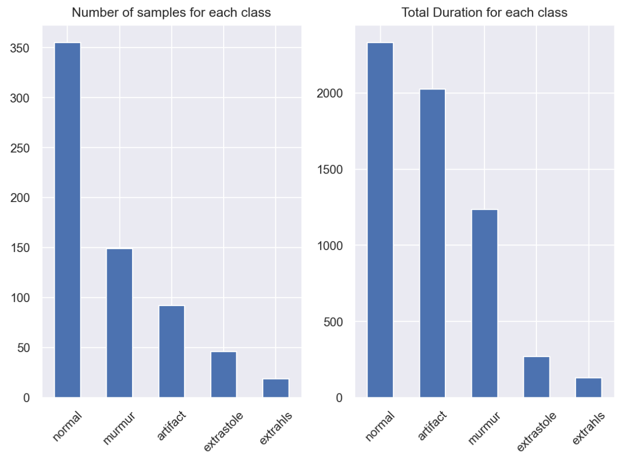

Fig. 1: Number of samples per duration.

2.2. 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 5.3).

Segmentation: the audio data was segmented into 1-second intervals, identified as the optimal extraction interval (see Section 5.3).

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.

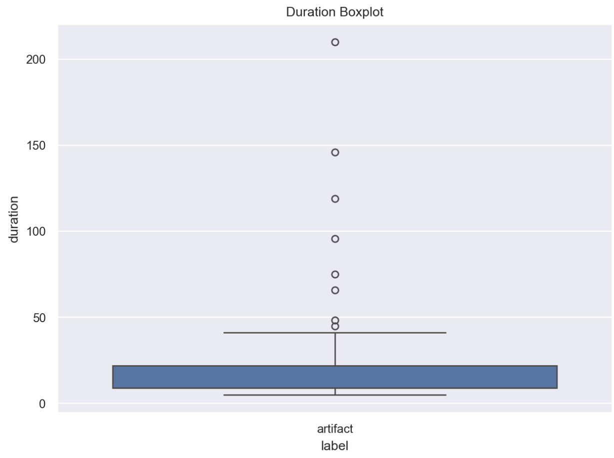

Fig. 2: Outliers in the Artifacts class.

3. Heart Sound Categories

Heart sounds can be categorized into different classes based on their acoustic characteristics and clinical significance. Accurate classification of these sounds is essential for diagnosing and treating a variety of cardiac conditions. The primary categories include Normal heart sounds, Murmurs, Extra Heart Sounds, Artifacts, and Extra Systoles. Understanding the distinct features and clinical implications of each class is a crucial step before building a machine learning model to classify heartbeats. This phase is particularly important for the identification of patterns that are characteristic of specific classes, which in turn guides the selection of features to extract from the audio. This knowledge aids in identifying specific patterns and anomalies within the heart sounds, leading to more precise and reliable model predictions.

3.1. Normal

The Normal category includes recordings of typical, healthy heart sounds. These sounds exhibit the characteristic "lub-dub, lub-dub" pattern, where "lub" (S1) represents the closing of the atrioventricular valves and "dub" (S2) signifies the closing of the semilunar valves. In a normal heart, the time interval between "lub" and "dub" is shorter than the interval from "dub" to the next "lub," especially when the heart rate is below 140 beats per minute. Most normal heart rates at rest fall between 60 and 100 beats per minute, though rates can vary from 40 to 140 beats per minute based on factors such as age and activity level. Recordings may include background noises like traffic or radio sounds and may capture incidental noises such as breathing or microphone contact with clothing or skin. It contains both clean and noisy normal recordings, the latter featuring significant background noise or distortion, which simulates real-world conditions.
Figure 3 shows a sample of a normal heart beat audio. The characteristic "lub-dub, lub-dub" pattern can be observed, where the peaks represent the "lub" (S1) and "dub" (S2) sounds of a healthy heart.

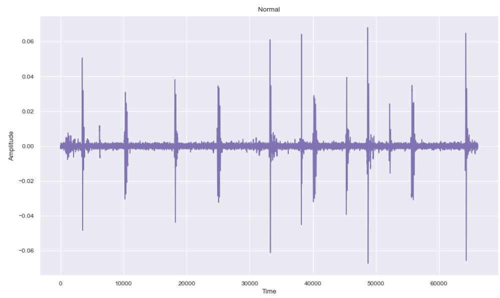

Fig. 3: Sample of normal heart beat audio.

3.2. Murmur

Heart murmurs are abnormal sounds during the heartbeat cycle, such as a “whooshing, roaring, rumbling, or turbulent fluid” noise, heard between the “lub” and “dub” (systolic murmur) or between “dub” and “lub” (diastolic murmur). These murmurs are typically indicative of turbulent blood flow in the heart and can signal various heart conditions, some of which may be serious. It is crucial to distinguish murmurs from the normal “lub-dub” sounds since they occur between the primary heart sounds and not concurrently with them. It also includes noisy murmur data, which mimics real-world recording scenarios by incorporating significant background noise and distortions.
Figure 4 shows a sample of a murmur heart beat audio. The presence of additional sounds between the “lub” and “dub” peaks can be observed, indicating the characteristic “whooshing, roaring, rumbling, or turbulent fluid” noise typical of heart murmurs.

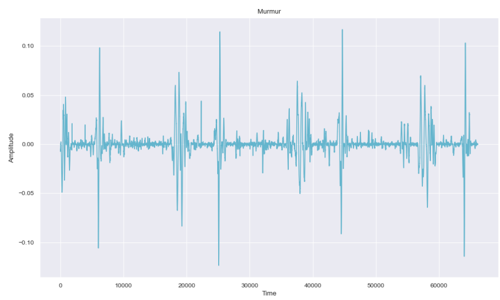

Fig. 4: Sample of murmur heart beat audio.

3.3. Extra Heart Sound

Extra heart sounds are characterized by an additional sound in the cardiac cycle, producing patterns such as “lub-lub dub” or “lub dub-dub”. These sounds can arise from physiological or pathological conditions. For example, a third heart sound (S3) may indicate heart failure or volume overload, while a fourth heart sound (S4) can be associated with a stiff or hypertrophic ventricle. Detecting these extra sounds is important for identifying potential heart diseases early, allowing for timely intervention and management. Figure 5 shows a sample of an extra heart sound audio. The presence of additional peaks within the normal “lub-dub” pattern indicates extra heart sounds, which can be critical for diagnosing various heart conditions.

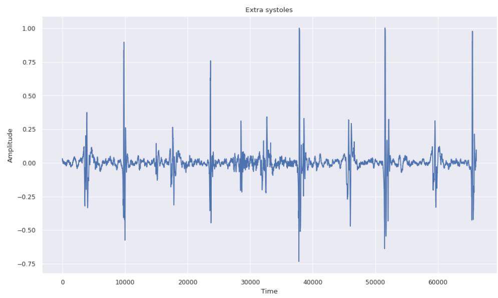

Fig. 5: Sample of extra heart sound audio.

3.4. Artifact

The Artifact category consists of recordings with non-cardiac sounds, including feedback squeals, echoes, speech, music, and various types of noise. These recordings generally lack discernible heart sounds and do not exhibit the temporal periodicity typical of heartbeats at frequencies below 195 Hz. Accurately identifying artifacts is essential to avoid misinterpreting non-cardiac sounds as pathological heart sounds, ensuring that data collection efforts focus on genuine heart sounds. Figure 6 shows a sample of an artifact heart beat audio, there can be observed that there is not a clear pattern in the audio.

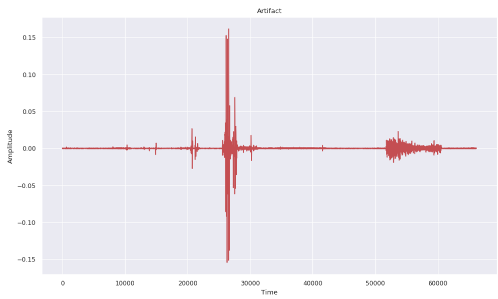

Fig. 6: Sample of artifact heart beat audio

3.5. Extra systoles

Extra systoles refers to extra or skipped heartbeats, resulting in irregular patterns such as “lub-lub dub” or “lub dub-dub”. Unlike the regular extra heart sounds, extra systoles are sporadic and do not follow a consistent rhythm. These premature beats can occur in healthy individuals, particularly children, but they may also be associated with various heart diseases. Identifying extra systoles is crucial as they can be early indicators of cardiac conditions that might require medical attention if they occur frequently or in certain patterns.
In the audio signal depicted in Figure 7, irregularities within the normal “lub-dub” pattern are evident. These irregularities manifest as additional peaks or skipped beats, indicating extra systoles.

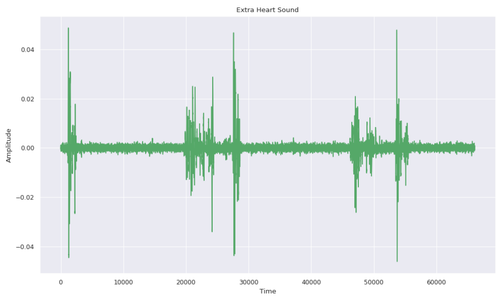

Fig. 7: Sample of extra systoles heart beat audio

3.6. Comparison of Heart Sounds

In Figure 8, a comparison of the different classes of heart sounds can be observed.
As we can see, the "Artifact" signal appears erratic with no consistent pattern, likely representing noise or interference rather than true heart sounds.
The "Murmurs" signal shows irregular fluctuations in amplitude, which could indicate turbulent blood flow typically associated with murmurs.
The signal for "Extra Heart Beat Sound" has occasional spikes in amplitude that stand out from the baseline.
The "Normal" signal appears more uniform and regular compared to the others, reflecting the expected rhythm of a healthy heartbeat. FInally, the signal for "Extra Systoles" shows extra spikes at irregular intervals, indicating unexpected contractions of the heart muscle (systoles) occurring outside the normal rhythm.

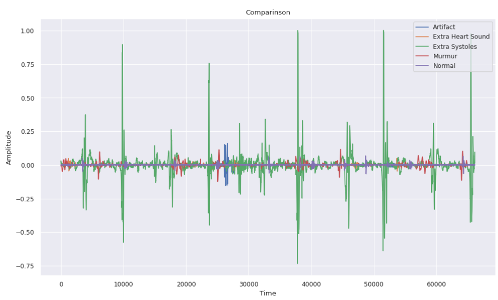

Fig. 8: Comparison of the different classes of heart sounds.

4. Goals

5. 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.

5.1. 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.

5.2. 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.

5.3. 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%.

5.3.1. 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 9.

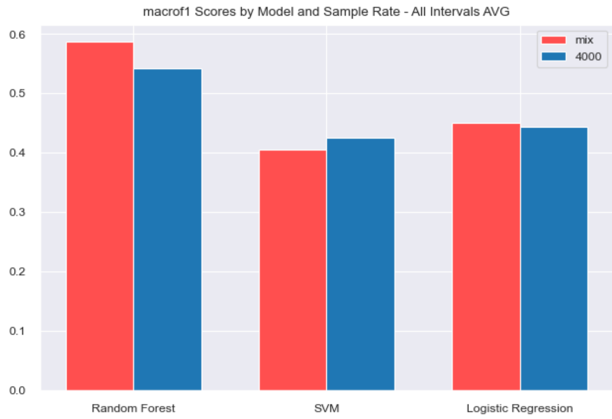

Fig. 9: Comparison of the macro F1 score for different sampling rates.

5.3.2. 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 10. 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 11 and 12 illustrate the impact of different extraction intervals on the macro F1 score and MCC, respectively.

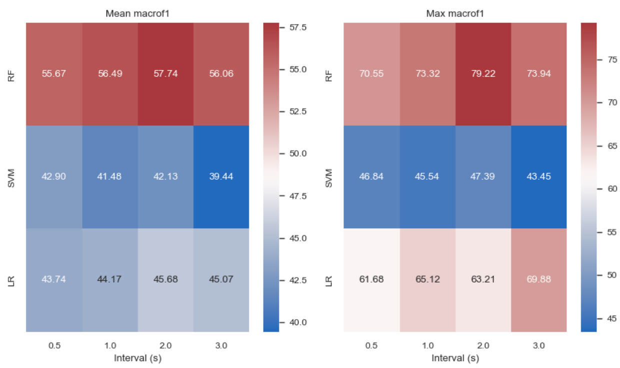

Fig. 11: Comparison of the macro F1 score for different extraction intervals.

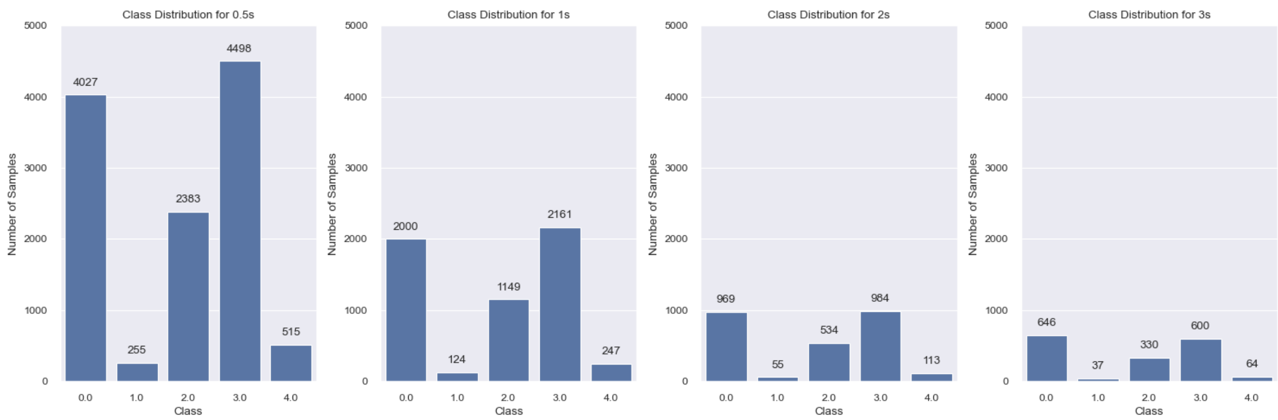

Fig. 10: Class distribution for different extraction intervals.

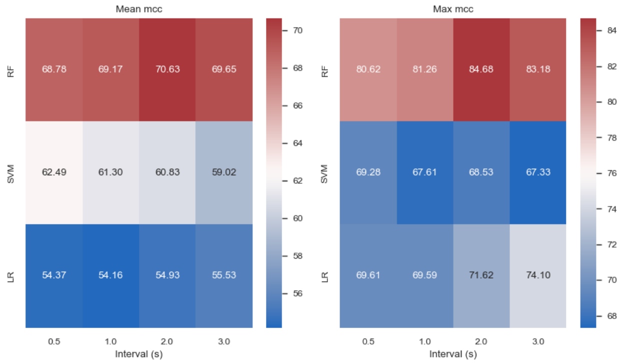

Fig. 12: Comparison of the MCC for different extraction intervals.

5.4. Results

6. Feature Selection

This phase serves a dual purpose. Firstly, since each type of feature (MFCC, Chroma, CQT, etc.) can consist of a varying number of individual features, it is important to determine the optimal number of features for each type to maximize the model's performance. Secondly, this phase aims to identify which specific features are the most relevant and influential for the classification task. By doing so, we can enhance the model's efficiency and accuracy by focusing on the features that contribute the most to the classification process.

6.1. Optimal number of features for each type of feature

Considering that each type of feature can consist of a variable number of individual features, it is necessary to determine the optimal number of features for each type to maximize the model's performance. To achieve this, a 'One Model per Feature' approach was employed. This involved training a separate model for each type of feature, varying the number of features used. A set of 12, 20, 30, 40, 60, 70, 90, and 120 features was extracted for each type.Then, for each feature set, three different models were trained using three classifiers: SVM, Random Forest, and Logistic Regression.

Figure 13 shows the results obtained for each type of feature with the Random Forest model, as it significantly outperformed the other classifiers.

The graph illustrates the F1 scores for different feature types, with varying numbers of features. The MFCC features consistently achieved the highest F1 scores, peaking around 0.7. This indicates that MFCC features are particularly effective for the classification task. Interestingly, the number of features (from 30 to 120) does not drastically affect the performance, suggesting that even a smaller set of MFCC features can be highly informative.

CQT features show moderate performance, with F1 scores around 0.4 to 0.5. The optimal number of features appears to be around 70, beyond which there is no significant improvement. Similarly, RMS features exhibit a range of F1 scores from 0.4 to 0.5, with optimal performance achieved with around 70 features.

For ZCR, SC, SB, and SR features, the F1 scores generally stabilize around 0.4 to 0.5. Increasing the number of features beyond 40 does not result in significant performance gains and can even degrade the model's performance. This suggests that adding too many features, especially those without strong predictive power, can confuse the model and degrade performance.

Overall, the optimal number of features is 30 MFCC,12 Chroma, 70 CQT, 40 RMS, 40 ZCR, 40 SC, 60 SB, and 40 SR.

6.2. Correlation analysis

The previous analysis resulted in 338 features. Given this large number, it is necessary to perform an analysis to identify and remove features that are poorly correlated with the target variable as well as those that are highly correlated with each other. Due to the high number of features, a visual approach, such as a correlation matrix, was not feasible for identifying the features to be removed. Instead, two filters were applied to select the most relevant features.

Since the normality test failed, the Spearman correlation coefficient was used for the analysis.

The first filter is based on the correlation between the features and the target variable. Features with a correlation below a certain threshold with the target variable are removed. The second filter focuses on the correlation among the features themselves. It counts, for each feature, the number of other features with which it has a correlation above a

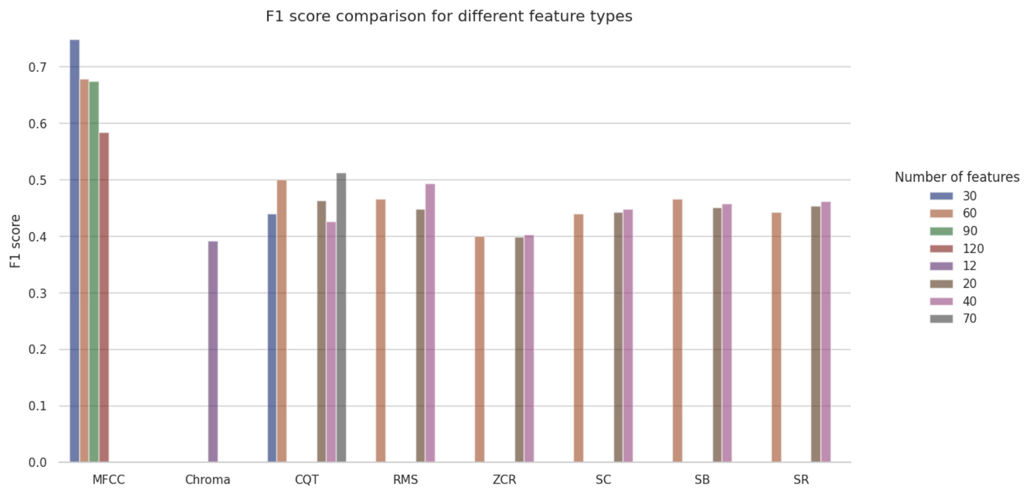

Fig. 13: F1 score per number of features

certain threshold. Features with a number of correlations above a specified threshold are then removed.
The threshold values were chosen empirically. The filters were applied using the combinations of thresholds shown in Table 1.

Threshold	Values
THRESHOLD 1	0 - 0.1 - 0.2 - 0.3 - 0.4 - 0.5
THRESHOLD 2	0.6 - 0.7 - 0.8 - 0.9 - 1
N° FEATURES	5 - 10 - 15 - 20 - 25 - 30 - 40

Table 1: threshold values

Using the filtered data, Random Forest models were trained to evaluate performance, as Random Forest was found to be the best performing model. The combination of thresholds that led to the best performance was selected: threshold 1 = 0, threshold 2 = 0.6, and the number of features = 30.
From these results, it can be observed that with threshold 1 = 0, the filter on the correlation between the features and the target variable is effectively bypassed. However, with threshold 2 = 0.6, a stringent filter is applied on the correlation among the features themselves, as all features with a correlation above 0.6 with at least 30 other features are removed. This indicates that it is more detrimental for the model to have features that are highly correlated with each other than to have features that are poorly correlated with the target variable.
Figure 14 shows the results obtained with the model trained on filtered features compared to the model trained on all features. As can be seen, the model trained on filtered features performs significantly better than the one trained on unfiltered features.

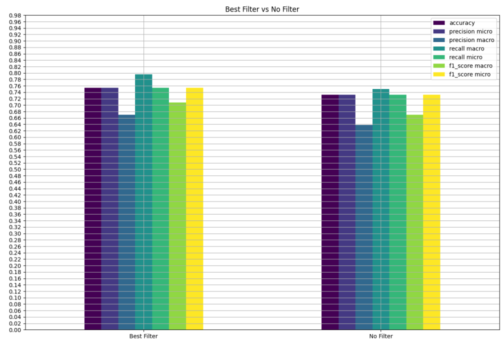

Fig. 14: Comparison of different metrics between the model on all features and the model on the filtered ones

From the entire analysis, 41 features remained: 28 MFCC, 12 Chroma, and 1 ZCR. The correlation matrix of the filtered features is shown in Figure 15. This matrix illustrates the pairwise correlation between the selected features, where the color intensity indicates the strength and direction of the correlation. Dark red cells represent high positive correlations, while dark blue cells indicate high negative correlations.
The matrix demonstrates that the remaining features have low correlations with each other, as evidenced by the predominantly light colors away from the diagonal. This implies that the features are relatively uncorrelated, which helps in preventing multicollinearity issues and enhances the robustness of the model.
The high diagonal values indicate that each feature is perfectly correlated with itself, which is expected. However, the off-diagonal values being close to zero for most feature pairs confirm that the filtering process was effective in selecting features that do not exhibit high inter-correlations.

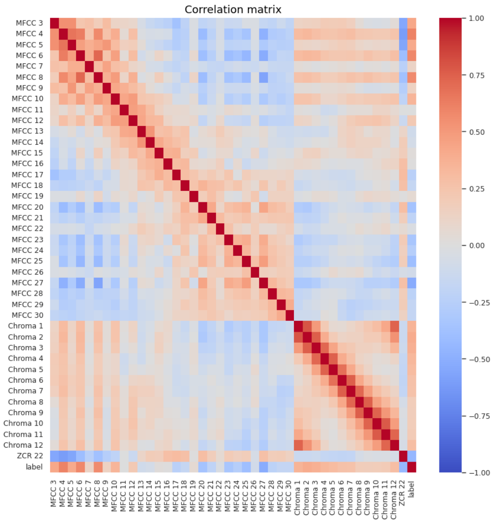

Fig. 15: Correlation matrix of the filtered features

7. Model Selection

7.1. Model on specific classes

7.1.1. Explanation of the models

7.1.2. Methodology

7.1.3. Results

7.2. Models on all classes

7.2.1. Explanation of the models

7.2.2. Methodology

7.2.3. Results

7.3. Best model analysis

8. Analysis of the Best Model

9. Conclusion

10. Future Works

11. 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.