# Goal

The primary objective of this study is twofold: (1) to accurately identify the timing of cardiac events, specifically the S1 and S2 heart sounds, and (2) to classify cardiac signals into one of three categories -normal, artifact, or murmur. The first objective involves detecting S1 and S2 events using features derived directly from the waveform. These event-level features are then incorporated, along with additional signal-based attributes, to perform multi-class classification as part of the second objective.

# Objective 1: To detect S1 and S2 events

## 1.1 Getting an insight on signals

Understanding audio files: The heart sound recordings used in this study are mono-channel signals sampled at 44.1 kHz. To gain an initial understanding of the signal characteristics across different classes-normal, murmur, and artifacts; both time-domain waveforms and their corresponding frequency spectra (via FFT) were visualized and analyzed. This preliminary exploration provided insights into the distinct spectral and temporal features associated with each class.

A line graph of sound waves

AI-generated content may be incorrect.A group of blue lines

AI-generated content may be incorrect.

A line graph of a sound wave

AI-generated content may be incorrect.A close-up of a graph

AI-generated content may be incorrect.

### Figure 1. Representative waveform and corresponding FFT plots for randomly selected audio files from each class (artifact, murmur, normal

# Upon visual inspection of several waveforms across categories, the signals do not exhibit noticeable DC offsets. While there is a possibility of powerline interference, the frequency content, especially in normal signals, shows significantly higher amplitude components than typical powerline artifacts. In artifact recordings, S1 and S2 sounds are not visually distinguishable; some signals are tightly centered around zero, while others exhibit high-amplitude noise. In contrast, murmur signals often show enhanced S1 and S2 peaks, possibly reflecting turbulent blood flow. Frequency analysis (FFT) indicates that normal signals show harmonic spectral content, whereas murmur signals are dominated by energy in the 0 to 200 Hz range.

# Traditional filtering methods (such as bandpass filtering) and down sampling were intentionally avoided. These techniques, while effective for noise suppression, risk degrading the temporal resolution that is critical for accurate event localization. Bandpass filters can introduce phase distortion, which may affect event timing unless compensated using forward and backward filtering. Similarly, down sampling may eliminate high-frequency components where relevant event cues are present. Therefore, to preserve both time and frequency fidelity, minimal preprocessing was used.

## 1.2 Literature suggestion

WHY

Prior literature highlights the effectiveness of Shannon entropy, Hilbert transforms, and wavelet decomposition for detecting cardiac events such as S1 and S2. To evaluate their practical utility, each method was applied to a sample recording. Both Hilbert and wavelet-based approaches outperformed the Shannon envelope in accurately identifying all 8 cardiac events (S1 and S2). While results may vary depending on the specific signal characteristics, this exploratory comparison informed the decision to proceed with the Hilbert transform for event detection. It offered consistent peak identification and allowed parameter fine-tuning to optimize timing precision.

A graph of a graph showing different colored lines

AI-generated content may be incorrect.

### Figure 2. Casual experiments with Shannon, Hilbert and Wavelet Transforms on one of the normal files with event information

## 1.3 Method:

The Hilbert transform was employed to generate an analytic signal by computing a phase-shifted version of the original waveform and combining it with the original signal to form a magnitude envelope. This process enhances transient cardiac events such as S1 and S2 by suppressing zero-mean noise through squaring and summation, thereby preserving temporal structure and improving event visibility.

After constructing the Hilbert envelope, a peak detection algorithm was applied to identify candidate S1 and S2 locations. Key parameters such as inter-peak distance and peak height were tuned to improve detection accuracy. A dynamic thresholding approach was also implemented-automatically adapting the peak detection threshold based on the amplitude and temporal profile of each recording.

To validate this approach, it was tested on normal heart sound recordings with annotated event timings. The method achieved precision and recall values exceeding 80%, with a mean absolute error of 0.0106 seconds in detecting S1 and S2 locations, demonstrating strong performance in temporal localization.

## 1.4 Results:

Accuracy-0.7118

Precision-0.8153 (Out of the total detected events how many of them were true events)

Recall-0.8487 (Out of the true events how many of them were predicted)

F1 score-0.8317

MAE± Std-dev: overall-0.0106± 0.0087 seconds,

MAE± Std-dev: S1-0.0112±0.0089 seconds,

MAE± Std-dev: S2 0.01±0.0084

RMSE overall-0.0137 seconds

Median error overall - 0.0091 seconds

Matched S1-169/195

Matched S2-162/195

A graph of a number of objects

AI-generated content may be incorrect.

### Figure 3. Event detection on normal file

The same Hilbert-based dynamic thresholding method was applied to murmur and artifact recordings to evaluate its generalizability. Visual inspection showed that S1 and S2 event localization remained reliable in murmur signals, likely due to the preserved rhythmic structure despite the turbulent flow. However, in artifact signals, which are characterized by irregular, high-amplitude noise and disrupted cardiac cycles, event detection performance significantly declined. This confirms that artifact signals pose a greater challenge for consistent and accurate event localization.

A graph showing a number of blue and orange lines

AI-generated content may be incorrect.

### Figure 4. Event detection on a Murmur file

A graph showing a waveform

AI-generated content may be incorrect.

### Figure 5. Event detection on Artifact

# Objective 2

## 2.1 Method

The Hilbert envelope provides several useful features for classification, including the location and number of peaks, as well as the mean and standard deviation of envelope values. Additionally, peak prominence, width, and inter-peak distances can offer discriminative insights. For instance, murmurs tend to exhibit higher peak heights and prominence compared to normal signals, reflecting turbulent cardiac flow. In contrast, artifact signals show irregularities where peak-related features may be abnormally low or high, making them distinctly separable.

A graph with green and purple lines

AI-generated content may be incorrect.

### Figure 6. Feature values using Hilbert transforms on a normal case

A graph with green and red lines

AI-generated content may be incorrect.

### Figure 7. Feature values using Hilbert transforms on a murmur case

A graph showing a number of data

AI-generated content may be incorrect.

### Figure 8. Feature values using Hilbert transforms on a normal case

## 2.2 Feature Engineering:

Relying solely on Hilbert transform features for classification would be limiting. To enrich the feature space and improve class separability, a combination of signal processing techniques was employed:

1. Captures the amplitude envelope of the signal, which highlights sudden bursts of energy corresponding to cardiac events like S1 and S2. Extracted features include peak height, prominence, width, and inter-peak intervals.
2. Shannon- Measures the entropy of the signal, capturing its complexity. Higher entropy is expected in murmurs and artifacts compared to normal sounds. Features used include the mean, standard deviation, and peak of the Shannon envelope.
3. Zero crossing rate- Reflects how frequently the signal crosses zero amplitude, indicating the signal’s frequency content. Both mean and standard deviation of ZCR were used. Artifacts, with their higher frequency noise, tend to have a higher ZCR.
4. MFCCs (Mel-Frequency cepstral coefficients)- These mimic humans auditory perception by emphasizing lower frequencies and compressing higher ones. They effectively summarize the shape of the power spectrum. Features include selected MFCC means and their first- and second-order temporal derivatives (deltas), which capture how spectral content evolves over time.
5. Wavelet transforms: Provides multiscale frequency decomposition while preserving time localization. Extracted features include the energy at multiple decomposition levels offering insights into frequency-specific patterns across time

To evaluate the discriminative power of the engineered features across the three classes (normal, murmur, and artifact), a two-step statistical analysis was performed. First, the Kruskal-Wallis test, a non-parametric method, was used to assess whether there were statistically significant differences in feature distributions across any of the three groups. Features with p-values below 0.05 were considered significant. To further identify which specific class pairs differed for each significant feature, Dunn’s post hoc test with Bonferroni correction was applied. This helped isolate the features that maximally contributed to distinguishing all three classes from one another.

Key features that were significantly different in each class were

1. Mean prominence from hilbert
2. Mean and standard deviation of width from hilbert
3. Standard deviation of Shannon envelop
4. MFCC features from 4 to 7
5. Wavelet energy levels from 1 t o 4

## 2.3 Classification algorithm & initial results

The features extracted through the feature engineering pipeline were used to classify the heart sound recordings into three categories: normal, murmur, and artifact. XGBoost was chosen as the classification algorithm due to its ability to capture non-linear relationships more effectively than traditional models like SVM or Random Forest. Additionally, XGBoost is well-suited to handle heterogeneous feature importance and offers built-in regularization mechanisms that help mitigate overfitting.

The dataset was split into training and testing sets using an 80/20 stratified split to preserve class proportions. Labels were encoded using a label encoder, and one-hot encoding was applied specifically for AUC computation in the multiclass setting. The model's performance was evaluated using metrics such as log loss, weighted F1 score, confusion matrix, and ROC-AUC curves for each class. Feature importance was also analyzed to interpret the model's decision-making process.

XGBoost was implemented using the low-level API, which provided greater flexibility in tracking training and validation metrics. The model was trained with a multiclass soft probability objective. Initial results showed signs of overfitting, with high AUC values for artifact (0.97) and murmur (0.94), but suboptimal performance on the normal class. The weighted F1 score, which accounts for both class imbalance and the trade-off between precision and recall, was 0.73. This indicated the need for hyperparameter tuning to improve generalization and enhance overall performance

A graph of a training and validation loss

AI-generated content may be incorrect.A graph with a line graph

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

### Figure 9. Initial model results

## 2.4 Hyper parameter Tuning

Optuna, a hyperparameter optimization framework, was used to fine-tune the XGBoost model. The primary objective was to maximize the weighted F1 score, which provides a balanced measure of precision and recall across potentially imbalanced classes, while also minimizing the multiclass log loss to ensure better generalization. The tuning process involved adjusting several key hyperparameters. These included the maximum depth of the trees, which controls model complexity and the risk of overfitting; the subsample ratio, which defines the percentage of training data used for each boosting round to introduce randomness and reduce overfitting; and the colsample\_bytree parameter, which governs the fraction of features randomly selected for each tree to promote feature diversity. Additionally, both L1 and L2 regularization parameters were tuned. L1 regularization penalizes the absolute magnitude of leaf weights, encouraging sparsity, while L2 regularization penalizes the squared magnitude of weights to provide smoother, more generalizable models.

A graph of a graph of a training

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

### Figure 10. Optimized model results

## 2.4 Final results

The weighted F1 score increased to 0.84, reflecting improved class balance and more reliable classification across artifact, murmur, and normal recordings. The ROC-AUC scores also improved, indicating stronger discriminatory power between the classes.

I should have tried amplitude-based features and Independent Component Analysis (ICA). It is surprising the AUC is not good for normal, I am missing something, but the time is limited.

## 2.5 Conclusion

While these results are promising, validation loss remains an area for improvement. It could be further reduced through enhanced regularization, hyperparameter tuning, or refined feature engineering to promote generalization and minimize residual error. Feature normalization was not applied, given that XGBoost handles varying feature scales inherently well; however, normalization may be incorporated in future work to improve compatibility with other classifiers. Additionally, this analysis used a single stratified train-test split; incorporating k-fold cross-validation in future iterations may offer more robust insights into model generalization across unseen data.

# References:

Shannon: Jasper, Julian, and Khair Razlan Othman. "Feature extraction for human identification based on envelogram signal analysis of cardiac sounds in time-frequency domain." *2010 International Conference on Electronics and Information Engineering*. Vol. 2. IEEE, 2010.

Hilbert: Varghees, V. Nivitha, and K. I. Ramachandran. "Heart murmur detection and classification using wavelet transform and Hilbert phase envelope." *2015 twenty first national conference on communications (NCC)*. IEEE, 2015.

Thalmayer, Angelika, et al. "A robust and real-time capable envelope-based algorithm for heart sound classification: Validation under different physiological conditions." *Sensors* 20.4 (2020): 972.

Wavelet - Golpaygani, Ali Tavakoli, et al. "Detection and identification of S1 and S2 heart sounds using wavelet decomposition method." *International Journal of Biomathematics* 8.06 (2015): 1550078.

Gündüz, Ali Fatih, and Ali Karci. "Heart sound classification for murmur abnormality detection using an ensemble approach based on traditional classifiers and feature sets." *Computer Science* 5.1 (2022): 1-13.

Hosseinzadeh, Mehdi, et al. "Enhanced heart sound classification using Mel frequency cepstral coefficients and comparative analysis of single vs. ensemble classifier strategies." *PloS one* 19.12 (2024): e0316645.

Li, Feng, et al. "Heart sound classification based on improved mel-frequency spectral coefficients and deep residual learning." *Frontiers in Physiology* 13 (2022): 1084420