Heartbeat classification from Phonocardiogram using Deep Learning

Report submitted to

Indian Institute of Technology, Kharagpur

as B.Tech Thesis Project-I

Under the supervision of

Prof. Amardip Ghosh

by

Manish Kumar Roy

20AE30031

DEPARTMENT OF AEROSPACE ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR

November 29, 2023

Certification

This is to certify that this project report entitled "**Heartbeat classification from Phonocardiogram using deep learning**" submitted to Department of Aerospace Engineering, Indian Institute of Technology, Kharagpur, is a bonafide record of work done by **Manish Kumar Roy** (Roll no. **20AE30031)** as his B.Tech thesis project-1 during the Autumn semester, 2023-2024.

Prof. Amardip Ghosh

Assistant Professor Grade - IDepartment of Aerospace Engineering

Indian Institute of Technology, Kharagpur

Date: 29.11.2023

Declaration

I certify that

A. The work contained in this report has been done by me under the guidance of my supervisor.

B. The work has not been submitted to any other Institute for any degree or diploma.

C. I have confirmed the norms and guidelines given in the Ethical Code of Conduct of the Institute.

D. Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given credit to them by citing them in the text of the thesis and givingtheir details in the references.

Manish Kumar Roy
20AE30031

Acknowledgement

I would like to express my sincere gratitude to my superiors Prof. Amardip Ghosh for his invaluable guidance, comments, and suggestions throughout the course of the project. I would also like to thank my parents, my sisters and my friends for their unconditional support.

Table of Contents

List of Figures

Abstract

Cardiovascular diseases are the leading cause of mortality globally, necessitating advancements in early and accurate diagnosis. This project addresses the challenge by employing a deep learning approach to classify heart sound recordings into normal, murmur, and artifact categories. The core of the system is a Bidirectional Long Short-Term Memory (BiLSTM) neural network, leveraging the temporal dynamics of heart sound signals. Trained on a dataset comprising varied heart sound recordings, the model achieved an overall accuracy of 78%, with a weighted average F1-score of 0.78.

The study demonstrated the feasibility of using BiLSTM networks for heart sound classification and highlighted the importance of handling class imbalance through appropriate weighting schemes. Visualizations of the training process indicated stable convergence without overfitting, validating the model's reliability.

Future scope for this project includes the exploration of attention mechanisms to fine-tune model focus, the application of transformer architectures to better capture long-range dependencies, and the potential synergy of CNN and transformer ensembles. Moreover, considering vision transformers to analyze spectrogram images of audio signals could unveil new discriminative features. Expansion of the dataset and testing in clinical environments are also imperative to enhance the model's robustness and practical utility. Ultimately, this project lays the groundwork for further innovation in the application of machine learning to cardiological diagnostics.

1. Introduction

Cardiovascular diseases (CVDs) pose a significant global health challenge, contributing substantially to morbidity and mortality worldwide. As per the World Health Organization (WHO) report, an estimated 17.5 million individuals succumbed to CVDs in 2012, constituting 31% of total global deaths. In the realm of clinical practice, the assessment of the cardiovascular system plays a pivotal role in the identification and treatment of CVDs. An integral component of this evaluative process is the physical examination, encompassing a thorough review of the patient's cardiovascular health. Auscultation of heart sounds, achieved through the use of a stethoscope, stands out as a crucial aspect of this examination, offering invaluable insights into the heart's well-being.
The practice of auscultating heart sounds entails attentive listening to the cardiac cycle's auditory cues. Anomalies detected in these sounds may signify the presence of various pathological cardiac conditions, including arrhythmias, valve diseases, and heart failure, among others. Timely identification of such conditions through auscultation guides subsequent diagnostic procedures and facilitates the prompt management of CVDs.

In the era preceding the 1900s, physicians endeavored to diagnose diseases by directly listening to a patient’s heartbeat, a method known as “immediate auscultation.” Deemed unethical and unscientific, this approach underwent a transformative shift in 1816 when Laennec invented the stethoscope—a now widely embraced medical instrument for detecting heart diseases. During initial screenings, medical professionals employ the stethoscope to discern irregularities in a patient’s heartbeat sounds. However, the interpretation of these sounds remains inherently subjective, demanding specialized training and experience. Moreover, the diagnosis of conditions like heart failure and coronary heart disease proves challenging through stethoscope auscultation, given that accurate interpretation necessitates years of clinical practice. Additionally, the limitations of the human ear, particularly its potential to miss certain sounds—especially when higher-frequency sounds predominate over lower-frequency ones—underscore the complexity of the diagnostic process. Research indicates that while using a stethoscope, medical students and primary care physicians achieve an accuracy rate of only 20-40%, in stark contrast to experienced cardiologists who achieve an 80% accuracy rate.

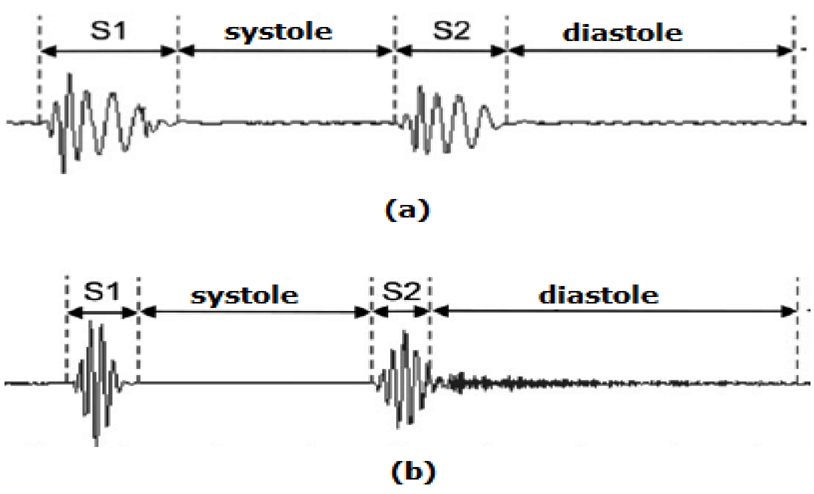


Figure 1. Characteristics of heart sounds obtained from phonocardiogram signals of normal patient (a) and patient with murmur (b)

1.1 Cardiac Cycle or Heart Cycle

The heart undergoes a process known as the cardiac cycle, The cardiac cycle, the rhythmic performance of the human heart from one heartbeat to the next, comprises distinct phases: diastole, where the heart muscle relaxes and fills with blood, and systole, marked by robust contraction and blood pumping. Following emptying, the heart expands to receive blood returning from the body before contracting again to propel blood to the lungs and other systems. A healthy heart, beating at 70 to 75 beats per minute, completes each cardiac cycle in approximately 0.8 seconds.
The heart consists of four chambers—two atria and two ventricles, working in tandem as the left and right hearts. During ventricular diastole, the heart relaxes and both atria fill both ventricles with blood. Towards the end of this phase, atrial systole commences, pumping blood into the respective ventricles. In ventricular systole, the ventricles contract, ejecting blood to the lungs and the body's organs while the atria relax.

The mitral and tricuspid valves, or AV valves, open during ventricular diastole, allowing filling. Atrial systole then forces a final influx of blood into the ventricles. Prompted by signals from the sinoatrial node, ventricular systole begins, closing the AV valves and entering the isovolumic contraction stage.

During ventricular systole, pressures rise, causing the aortic and pulmonary valves to open, leading to blood ejection from the ventricles. As pressures decline, these valves close again, marking the end of systole. Isovolumic relaxation follows, with ventricular pressure decreasing, allowing the atria to refill. The mitral and tricuspid valves reopen as the cycle returns to ventricular diastole.

Throughout the cardiac cycle, blood pressure fluctuates. Specialized pacemaker cells in the sinoatrial and atrioventricular nodes coordinate cardiac muscle movements through electrical impulses. In an electrocardiogram, atrial systole initiates at the P wave, marking the onset of electrical systole and contractions.

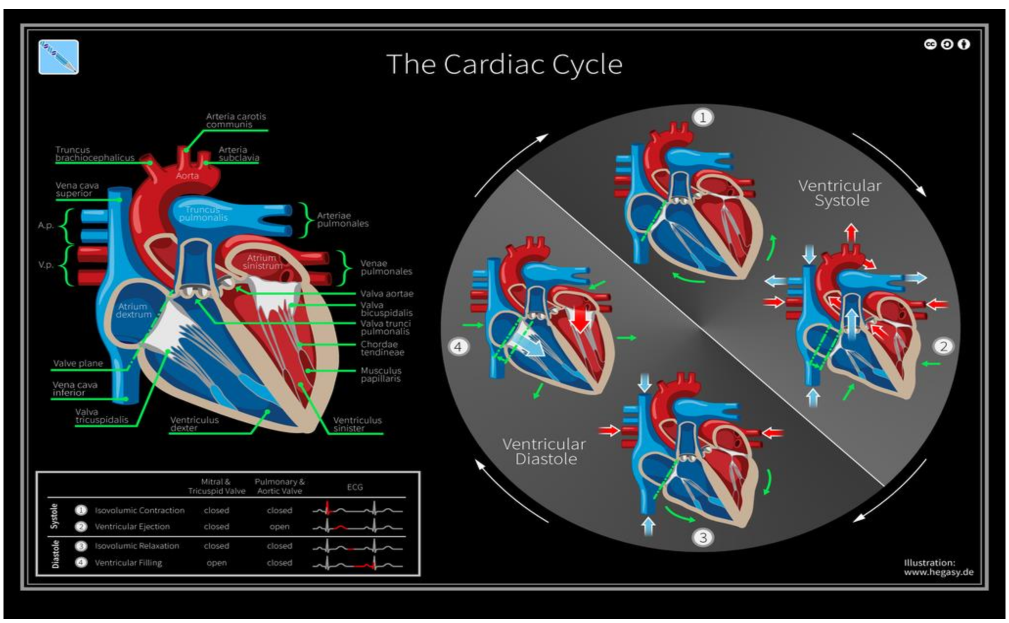


Figure 2. The Cardiac Cycle: Valve Positions, Blood Flow, and ECG

1.2 Heart electrical conduction system

In a well-functioning heart, every activity and repose within each cardiac cycle, commonly known as a heartbeat, is intricately guided by the signals of the heart's electrical conduction system. This system, often likened to the "wiring" of the heart, carries electrical impulses across the cardiomyocytes—specialized muscle cells of the heart. These impulses play a pivotal role in prompting heart muscle contractions, propelling blood from the ventricles into the arteries and throughout the cardiac circulatory system. The system provides a meticulously timed and persistent signaling mechanism, crucial for regulating the rhythmic beating of the heart muscle cells, particularly orchestrating the intricate impulse generation and muscle contractions in the atrial chambers.

The rhythmic sequence, or sinus rhythm, of these electrical signals is masterfully coordinated by two sets of specialized cells: the sinoatrial (SA) node, situated in the upper wall of the right atrium, and the atrioventricular (AV) node, positioned in the lower wall of the right heart between the atrium and ventricle. Often referred to as the cardiac pacemaker, the sinoatrial node serves as the origin point for generating a wave of electrical impulses, igniting atrial contraction by initiating an action potential across the myocardium cells.

Upon reaching the AV node, the impulses experience a deliberate delay, acting as a gate to slow and synchronize the electrical current before it travels beneath the atria. The journey continues through specialized circuits known as the bundle of His and the Purkinje fibers. These components collectively stimulate contractions in both ventricles. The intentional delay at the AV node serves a crucial purpose by allowing sufficient time for blood volume to flow through the atria, filling the ventricular chambers just before the onset of systole (contractions). This strategic timing ensures the ejection of the new blood volume, completing the cardiac cycle. For a visual representation, refer to the Wiggers diagram, particularly examining the "Ventricular volume" tracing (depicted in red) during the "Systole" panel.

1.3 Diastole and Systole in the cardiac cycle

Cardiac diastole represents a pivotal phase in the cardiac cycle, characterized by the heart's relaxation and expansion following contraction. During this crucial interval, the heart replenishes itself by receiving blood returning from the circulatory system. To facilitate this process, both atrioventricular (AV) valves open, enabling the smooth, unpressurized flow of blood directly from the atria into both ventricles. This orchestrated movement ensures that the heart is adequately prepared to collect and pump blood in the subsequent contraction.

1.3.1 Atrial Systole

Atrial systole is a crucial phase in the cardiac cycle marked by the contraction of cardiac muscle cells in both atria. This contraction is triggered by the electrical stimulation and conduction of electrical currents across the atrial chambers, as explained in the earlier discussion on physiology. Although typically categorized as part of the heart's systolic contraction and ejection sequence, atrial systole plays a pivotal role in finalizing diastole, a phase dedicated to filling both ventricles with blood while they are in a relaxed and expanded state. Atrial systole occurs in the latter part of ventricular diastole, specifically in the sub-period known as ventricular diastole–late, as illustrated in the cycle diagram.

During atrial systole, contraction pressure is applied to complete the filling of both ventricles with blood. This phenomenon, often referred to as the "atrial kick," serves as a crucial step in closing the diastole, immediately preceding the onset of the heart's contraction and ejection of blood from the ventricles (ventricular systole) into the aorta and arteries.

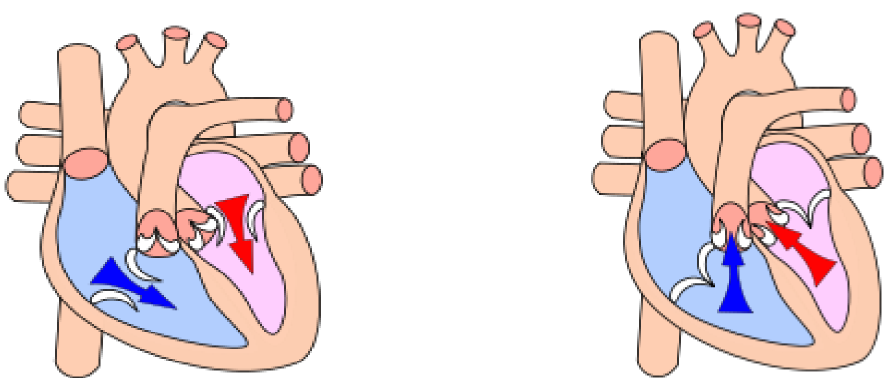
It's noteworthy that the absence or disruption of the atrial kick can occur in cases of abnormal electrical conduction in the heart, such as those seen in conditions like atrial fibrillation, atrial flutter, or heart block. Additionally, any deterioration in the overall condition of the heart, such as in "stiff heart" conditions found in patients with diastolic dysfunction, may result in a degradation of the atrial kick. Understanding the significance of atrial systole and its potential variations is essential for assessing the overall cardiac function and identifying potential abnormalities.

1.3.2 Ventricular systole

Ventricular systole encompasses the coordinated contractions of the ventricular syncytium, a network of cardiac muscle cells within both the left and right ventricles. These contractions, initiated by electrical stimulations, serve distinct yet interconnected functions in the circulatory system.

In the right ventricle, contractions facilitate pulmonary circulation, propelling oxygen-depleted blood through the pulmonary valve and into the pulmonary arteries en route to the lungs. Simultaneously, the left ventricular systole powers systemic circulation, sending oxygenated blood throughout the entire body. This process involves the pumping of blood through the aortic valve, into the aorta, and subsequently through all the arteries that branch out to supply blood to various body systems.

Notably, the larger arteries originating from the left ventricle serve as routine sites for blood pressure measurement during the left ventricular systole. This phase of the cardiac cycle is vital for maintaining both pulmonary and systemic circulation, ensuring that oxygen-depleted blood is efficiently directed to the lungs for replenishment while oxygenated blood is distributed to meet the metabolic demands of the body's diverse systems. Understanding the intricacies of ventricular systole is fundamental to comprehending the dynamics of the circulatory system and its essential role in sustaining overall physiological function.



3.(a) 3.(b)

Figure 3. (a)Cardiac Distole (b)Cardiac(Ventricular) systole

2. Literature Review

In the past few years, the research community has dedicated substantial efforts to conceptualize and develop a range of devices aimed at remotely monitoring patients with Cardiovascular Disease (CVD).

Rachel Hajar explored the transformative journey in our understanding of cardiovascular diseases (CVDs), with a particular focus on the impact of the Framingham Heart Study (FHS). Initiated in response to the death of President Franklin D. Roosevelt in 1945, the FHS became a cornerstone in cardiovascular research. The paper details how the study identified and corrected fallacies in our understanding of CVDs, particularly in relation to risk factors like hypertension, hypercholesterolemia, smoking, diabetes, physical inactivity, and obesity. It highlights the crucial role played by the FHS in dispelling long-standing myths, leading to a paradigm shift in medical practices. The narrative unfolds through a historical lens, showcasing the significant strides made in preventive strategies, including the comprehensive assessment of modifiable risk factors. The study's findings, coupled with current guidelines from organizations like the ACC/AHA, underscore the importance of tailored approaches to reduce cardiovascular morbidity and mortality.

Fatih Demir, Abdulkadir Şengür, Varun Bajaj, Kemal Polat proposed method involves three key stages: spectrogram generation, deep feature extraction using pre-trained convolutional neural network (CNN) models (AlexNet, VGG16, and VGG19), and classification using a support vector machine (SVM) classifier. The evaluation is conducted on two datasets from the Classifying Heart Sounds Challenge, comparing the results with existing methods.

The significance of early detection of heart diseases is emphasized, considering the high global mortality rates associated with cardiovascular diseases. The paper discusses the limitations of traditional auscultation methods and highlights the growing role of machine learning algorithms in automating and improving the accuracy of heart sound analysis.

The datasets used in the study, namely Dataset-A and Dataset-B, are described in terms of their classes and recording sources. The proposed method's effectiveness is evaluated based on precision, sensitivity, specificity, and F-measure values. The experimental results demonstrate that the proposed method, leveraging deep features extracted from pre-trained CNN models and SVM classification, outperforms existing methods in heart sound classification.

The construction of spectrogram images using Short Time Fourier Transform (STFT) is detailed, explaining the rationale for using color spectrogram images as inputs for CNN models. Feature extraction from the CNN models, including AlexNet, VGG16, and VGG19, is outlined, emphasizing the use of the fc6 layer for obtaining deep feature vectors.

The evaluation criteria include total precision, normalized precision, and precision for each class label. The results are presented for both datasets, showcasing the superior performance of the proposed method. Comparative analyses with other methods, such as J48, MLP, and SVM-DM, are provided, highlighting the strengths of the proposed approach.

Ali Raza proposed methodology, integrating a Recurrent Neural Network (RNN) with innovative data framing and down-sampling techniques, showcased remarkable performance in the classification of heartbeat sounds. Outperforming conventional methods such as Decision Tree and Random Forest, the RNN model achieved an impressive accuracy of 80.8% for 12.5-s samples and 77.2% for 27.8-s samples on Dataset-B. Hyper-parameter tuning, particularly with a dropout rate of 0.35, contributed to the model's success, emphasizing its efficiency in accurately detecting and categorizing heartbeat signals.

The robustness of the RNN model was further validated through K-Fold Cross-Validation, yielding a weighted accuracy of 80.45%. This comprehensive evaluation not only affirms the reliability of the proposed approach but also highlights its generalizability across diverse datasets. The study suggests that the integration of deep learning techniques, coupled with thoughtful preprocessing methods, holds great promise for automating heartbeat sound classification. This advancement has significant implications for improving diagnostic capabilities in healthcare applications.

In their research, **Low and Choo** delve into the application of deep-learning neural networks for classifying heart sounds. They suggest a method that involves breaking down heart sounds per heartbeat, converting each segment into an intensity map, and then using Softmax Regression (SMR) and a Convolutional Neural Network (CNN) for classification. The key innovation is that this approach negates the need for manually selecting features for a supervised machine learning method, as the deep neural network can automatically determine relevant features during training. The authors underscore that the CNN algorithm outperformed SMR in heart sound classification. They highlight the ease of implementation and the potential for real-time heart sound classification from mobile phones. Notably, it's mentioned that although the study did not utilize mobile phones for signal capture, the implications for real-time heart sound analysis from such devices are significant. The authors stress the importance of heart sound classification in diagnosing various heart diseases, asserting that automatic analysis can enhance diagnostic accuracy and speed. The study utilizes heart sound recordings from the PhysioNet/Computing in Cardiology Challenge 2016 database, comprising over 3,000 recordings from patients with different heart conditions. These recordings are employed to train and test the proposed deep-learning neural networks. In summary, the paper presents a promising method for heart sound classification through deep learning, with potential implications for quicker and more accurate diagnoses of heart diseases. The prospect of real-time classification using mobile phones holds particular promise for patients in remote or underserved areas lacking access to advanced medical diagnostic tools.

Yaseen, Gui-Young Son, and Soonil Kwon conducted a study focused on classifying heartbeat sounds using different techniques for feature extraction. They used an electronic stethoscope device to capture the digital recording of heart sounds, generating a phonocardiogram (PCG). In their research, the team applied Mel Frequency Cepstral Coefficient (MFCCs) and Discrete Wavelets Transform (DWT) as feature extraction methods. Additionally, they combined these techniques to enhance result accuracy. The study utilized classification algorithms such as Support Vector Machine (SVM), Deep Neural Network, and K-Nearest Neighbor (KNN) for the categorization of heart sounds.

The aim of this research is to capture heart sounds using a mobile phone's microphone, process and filter the signal using signal processing techniques, and finally predict if the heart sound indicates a normal or abnormal function by employing Long Short Term Memory(LSTM) with Mel-frequency cepstral coefficients (MFCCs) as the audio feature.

3. Methodology

To construct a deep learning model for classifying heart sounds from audio data, the initial phase involves data collection and preprocessing. Heart sound audio data is sourced from both public domains and clinical recordings, creating a comprehensive dataset that includes a range of heart sound variations. **Preprocessing** these audio samples is critical for ensuring consistent quality and includes steps like normalizing the audio volume, reducing background noise, and segmenting longer recordings into shorter, more manageable clips. This standardization is crucial for the robust performance of the model across diverse data inputs.

The next step is Exploratory Data Analysis (**EDA**), which serves to unveil the underlying structure and distribution of the audio data. By visualizing different audio features, such as waveforms, spectrums, and spectrograms, insights into the data can be gleaned, which can guide the subsequent feature extraction process. Moreover, EDA helps in identifying class imbalances within the dataset, which can significantly impact the model's performance. To mitigate this, strategies such as adjusting class weights are employed, ensuring that the model does not become biased toward the more prevalent classes.

Feature extraction is then carried out, with a focus on Mel-frequency cepstral coefficients (**MFCCs**), which effectively capture the power spectrum of the audio signals. Following this, the data preparation phase includes the loading of these preprocessed audio files for MFCC extraction, encoding the categorical labels, and splitting the dataset into subsets for training, validation, and testing. This structured approach facilitates the training of the model and its evaluation against unseen data, ensuring a robust assessment of its predictive capabilities.

The model building phase involves designing an **LSTM** (Long Short-Term Memory) neural network architecture that's well-suited to the sequential nature of audio data. The incorporation of layers such as Dense, Dropout, and Bidirectional LSTM layers is aimed at enhancing the learning process and preventing overfitting. Once the architecture is established, the model is compiled with a loss function and optimizer selected for the classification task at hand. The training phase then commences, where the model is fitted to the training data with class weights applied to account for any imbalance present in the dataset. Callbacks like ModelCheckpoint and EarlyStopping are instrumental in monitoring the training process, facilitating the retention of the best model iteration, and preventing overtraining.

Model **evaluation** and **tuning** involve rigorous testing of the trained model on the validation and test sets. Key performance metrics such as accuracy, precision, recall, and F1-score are calculated to assess its classification prowess. Discrepancies between the predicted and actual labels are scrutinized, and the model is fine-tuned accordingly to enhance its generalization capabilities. Once the model demonstrates satisfactory performance, it is saved, paving the way for deployment in real-world applications or for further research and development.

Here's a summarized methodology for our LSTM model to classify heart sounds:

- Data Collection
- Data Preprocessing
- Exploratory Data Analysis (EDA)
- Feature Extraction
- Data Preparation

- Model Building
- Model Training
- Model Evaluation and Tuning
- Prediction and Analysis

3.1 Data Collection

The foundation of our deep learning model is a dataset meticulously compiled from global contributors, aimed at addressing cardiovascular diseases (CVDs)—the leading cause of mortality worldwide. **The "Classifying Heart Sounds Challenge," sponsored by PASCAL and spearheaded by researchers Peter Bentley, Glenn Nordehn, Miguel Coimbra, and Shie Mannor, sought innovative methods to detect cardiac pathologies.** This challenge holds significant promise for medical machine learning applications due to the intricate classification required for audio sample data, often muddled with various background noises.

The dataset comprises recordings gathered from two distinct environments: Dataset A originates from the general public via the iStethoscope Pro iPhone app, while Dataset B is collected through clinical trials using the DigiScope digital stethoscope. Both sets contain heart sounds categorized into 'Normal,' 'Murmur,' 'Extra Heart Sound,' 'Artifact,' and 'Extrasystole,' with the latter two categories exclusive to Dataset A and B, respectively.

Participants of the challenge were tasked with two primary objectives: the segmentation of heart sounds—specifically identifying the S1 (lub) and S2 (dub) sounds within audio data—and the classification of heart sounds into their respective categories. Success in the first task was expected to simplify the second, which is the classification of real heart audio into predefined categories. The winning methodology would be capable of accurately segmenting and classifying a test set of unlabeled data after being trained on the provided datasets.

The audio files in these datasets range from 1 to 30 seconds in length and often include background noise. Notably, most heart sound information lies in low-frequency components, suggesting the application of a low-pass filter at 195 Hz could be beneficial. Furthermore, Fast Fourier transforms are likely to yield valuable insights into volume and frequency variations over time.

Within the 'Normal' category, sounds exhibit a distinct "lub-dub" pattern, while 'Murmur' sounds include additional noise between the 'lub' and 'dub' sounds. The 'Extra Heart Sound' category introduces additional beats, resulting in a "lub-lub dub" or "lub dub-dub" rhythm. 'Artifact' sounds are significantly different, often containing extraneous noises without discernible heart sounds. Lastly, 'Extrasystole' sounds manifest as irregular heartbeats.

This challenge underscores the complexity of classifying heart sounds—a task that demands robust classifiers capable of distinguishing between nuanced audio cues amid a plethora of ambient sounds. To facilitate the development of effective classification methods, segmentation data, and evaluation scripts, detailed metrics and testing procedures were provided. The challenge culminated in a workshop at AISTATS 2012, where the winning methods were presented and discussed.

Dataset	Category	Number of Files	Total File Size
A	Normal	31	14 MB
A	Murmur	34	17.3 MB
A	Extra Heart Sound	19	6.9 MB
A	Artifact	40	22.5 MB
A	Unlabelled Test	52	24.6 MB
B	Normal	320 (includes noisy normal)	13.8 MB
B	Murmur	95 (includes noisy murmur)	5.3 MB
B	Extrasystole	46	1.9 MB
B	Unlabelled Test	195	9.2 MB

The classification of heart sounds is a critical aspect of diagnosing cardiovascular conditions. The heart sounds mentioned in your dataset are characteristic of different cardiac events and are described as follows:

1. **Normal (S1 and S2)**:
    - **Normal** heart sounds are the sounds of a healthy heart and are typically described as a "lub-dub" pattern.
    - The first sound, **S1** or "lub," is created by the closing of the atrioventricular valves (mitral and tricuspid) at the beginning of ventricular contraction, or systole.
    - The second sound, **S2** or "dub," is caused by the closure of the semilunar valves (aortic and pulmonic) at the end of systole, as ventricular pressure falls below the pressure in the aorta and pulmonary artery.
2. **Murmur**:
    - **Murmurs** are sounds produced by turbulent blood flow, which can occur inside or outside the heart. They can be heard in a variety of conditions, both benign and pathological.
    - Murmurs are generally heard during the cardiac cycle's systolic or diastolic phases and are often described using terms like "whooshing" or "swishing."
    - They may indicate various issues, such as valve abnormalities or **defects**, but can also be present in an entirely healthy heart (innocent murmurs).
3. **Extra Heart Sounds**:
    - **Extra heart sounds** are additional noises that can occur within the cardiac cycle, often signifying some **abnormality**.
    - These may include sounds like **S3** and **S4**. The **S3** sound can occur shortly after S2 when the ventricles are resistant to filling. The **S4** sound occurs just before S1 when the atria contract to force blood into a stiff or hypertrophic ventricle.

4. **Artifact**:

    - **Artifacts** are extraneous sounds or noises picked up during auscultation that are not related to the heart's function. These can include environmental noises, movements, talking, or equipment interference.
    - They are considered non-diagnostic and can often obscure or mimic genuine heart sounds, making accurate diagnosis challenging.

5. **Extrasystole**:

    - **Extrasystoles** refer to extra beats or premature heartbeats. They can originate from the atria (atrial extrasystoles) or the ventricles (ventricular extrasystoles).
    - These extra beats can cause an irregular heart rhythm, where the typical "lub-dub" pattern is disrupted by additional or skipped beats.

Each type of heart sound provides valuable information about the heart's condition and function. In a clinical setting, careful auscultation by a trained medical professional can help identify potential heart problems, which can then be further investigated with more sophisticated diagnostic tools if necessary. In the realm of machine learning, being able to classify these heart sounds accurately can greatly aid in the preliminary screening for heart conditions and potentially improve patient outcomes.

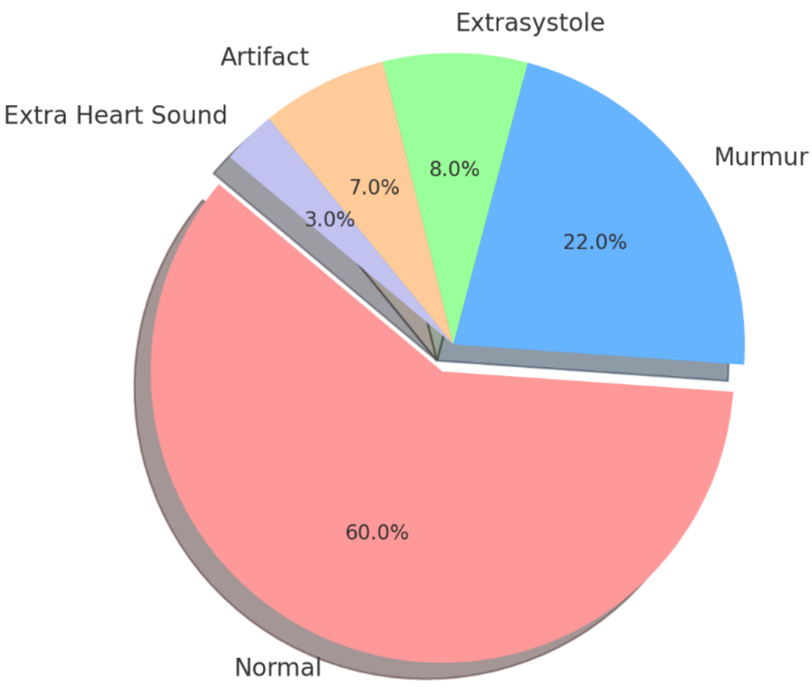


Figure 4. Data Distribution

3.2 Data Visualization

This is essential for understanding the nature of the data you are working with and for identifying patterns that might be useful for classification.

Normal Heart Sound

Waveform Visualization: A plot of the audio signal amplitude over time, which is the most direct representation of the sound. Sound is the pressure of air propagates to our ear. Digital audio file is gotten from a sound sensor that can detects sound waves and converting it to electrical signals. Specifically, it's telling us about the wave's displacement, and how it changes over time.

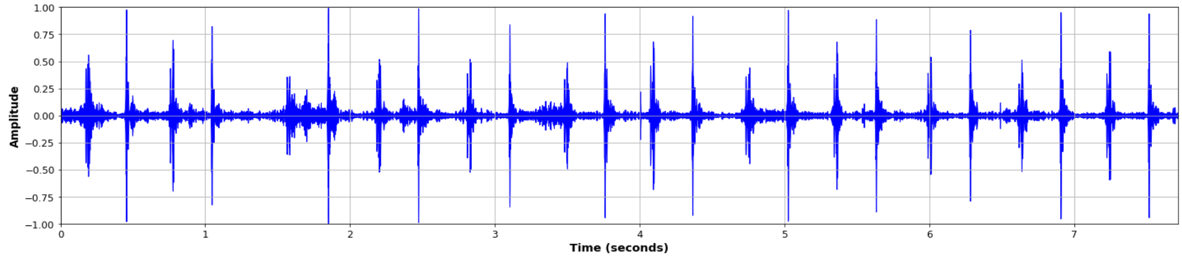


Figure 5. Waveform of a Normal Heart

1. **Regular Heartbeat**: The clear, repetitive peaks and troughs suggest a regular and consistent heartbeat. This is a good indicator of a normal heart rhythm, assuming the recording is from a person at rest.
2. **Heart Rate Calculation**: The distance between the peaks can be used to calculate the heart rate. If the sample rate is known, you can measure the time between the peaks (R-R intervals) to calculate beats per minute (BPM).
3. **Presence of S1 and S2**: The larger spikes likely correspond to the first heart sound (S1 - the "lub"), and the smaller, sharper spikes to the second heart sound (S2 - the "dub"). This "lub-dub" pattern is characteristic of normal heart sounds.
4. **Signal Quality**: The waveform is clean with minimal noise or artifacts, which suggests that the recording quality is high, and the data is suitable for further analysis.
5. **No Abnormal Sounds**: There are no apparent extra sounds like S3 or S4 heart sounds, murmurs, or other irregularities, which would typically be identified as additional peaks or a disruption in the regular "lub-dub" pattern.

Spectrum visualization: It is a technique used to understand the frequency content of an audio signal. It reveals which frequencies are present in the sound and how strong they are (amplitude). To obtain this information from a signal, which is naturally in the time domain (amplitude vs. time), we need to convert it into the frequency domain. This is where the Fast Fourier Transform (FFT) comes into play.

Why Use FFT?

The FFT is an algorithm to compute the Discrete Fourier Transform (DFT) and its inverse. The DFT decomposes a sequence of values into components of different frequencies. This is a fundamental operation in signal processing and analysis because:

1. **Speed**: The FFT is much faster than the direct computation of the DFT, especially for long sequences. It reduces the complexity from $O(N^2)$ to $O(NlogN)$, where N is the number of samples.
2. **Insight into Frequency Components**: Many signals are composed of multiple overlapping frequencies. The FFT helps to disentangle these and analyze the contribution of each frequency to the overall signal.

The Mathematics of FFT

The DFT is mathematically defined as:

$$X[k] = \sum_{n=0}^{N-1} x[n].e^{\frac{-i2\pi}{N}kn}$$

where:

- $X[k]$ is the k-th element of the frequency domain representation.
- $x(n)$ is the n-th element of the time-domain signal.
- N is the total number of samples.
- e is the base of the natural logarithm.
- i is the imaginary unit.

The term $e^{\frac{-i2\pi}{N}kn}$ is a complex exponential, which, when plotted, forms a circle in the complex plane, representing a pure frequency. The DFT sums the product of the original signal and a series of these complex exponentials, which act as "frequency filters".

In practice, the FFT algorithm cleverly reduces the number of computations needed by taking advantage of symmetries and redundancies in the DFT. The fundamental idea is to recursively divide the DFT into smaller DFTs, exploiting the fact that the DFT itself is periodic and symmetrical.

For an audio signal, the FFT output is usually complex, which includes both a real part (cosine component) and an imaginary part (sine component). To get the magnitude spectrum (which represents the amplitude of each frequency), we compute the magnitude of the complex number:

$$Magnitude = \sqrt{Re(X[k])^2 + Im(X[k])^2}$$

And for the phase spectrum (which represents the phase shift of each frequency), we calculate the angle of the complex number:

$$Phase = \tan^{-1}\left(\frac{Im(X[k])}{Re(X[k])}\right)$$

The magnitude spectrum is what is typically plotted in a spectrum visualization, showing the amplitude of each frequency present in the signal. The phase spectrum is often less important in audio processing but can be critical for signal reconstruction and other analyses.

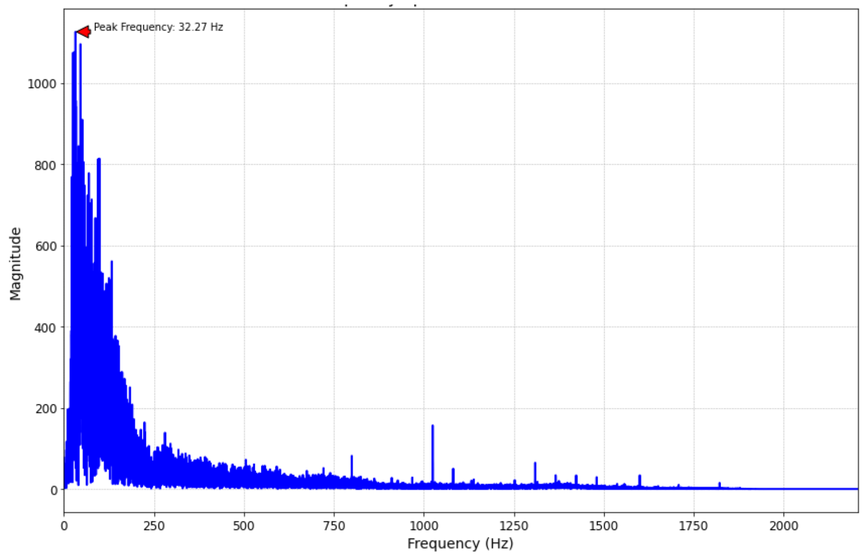


Figure 6. Frequency Spectrum of Heart Sound

Based on the image of the frequency spectrum we got, here are the inferences that can be drawn:

1. **Dominant Low-Frequency Component**: The graph shows a strong peak at a low frequency of around 32.27 Hz. This indicates that the most significant part of the heart sound is at this low frequency, which is common for heart sounds as they are generally low-frequency signals.

2. **Harmonics Presence**: There are other smaller peaks visible in the spectrum, which could represent the harmonics of the heart sound. Harmonics are multiples of the fundamental frequency and are a natural occurrence in most physical signals.

3. **Relative Magnitude of Components**: The magnitude of the fundamental frequency is significantly higher than that of the other frequencies present in the signal. This suggests that the heart sound is relatively clean and has a strong characteristic frequency.

4. **Background Noise**: There is some presence of frequencies across the spectrum, but they are of much lower magnitude compared to the peak. This could be due to background noise or other physiological sounds.

Spectogram Visualization: A spectrogram is a powerful tool for analyzing audio signals. It provides a visual representation of the signal's intensity (or loudness) across various frequencies over time. Humans perceive sound not just in terms of volume (intensity) but also pitch (frequency). The spectrogram reflects this by showing how these two dimensions vary over time.

A spectrogram is a visual way to represent the strength of the signal at various frequencies over time. It's essentially a collection of Fourier Transforms (FT) applied to small segments of the signal, which is why it uses the Short-Time Fourier Transform (STFT).

The STFT divides the signal into small time-based frames and computes the Fourier Transform separately for each frame. This is akin to looking at the signal through a sliding window that moves along with time. Each FT within these small windows transforms the time-domain information into frequency-domain information for that specific segment of time.

The "short-time" aspect is crucial because it allows us to capture how frequencies change over time, which a single Fourier Transform applied to the whole signal could not provide. By using a window (the frame) that is moved across the signal, we obtain a series of frequency spectra over time, which are then plotted on the spectrogram. The x-axis represents time, the y-axis represents frequency, and the color intensity represents the magnitude (or power) of the observed frequency at that time.

This method is particularly powerful for non-stationary signals, where frequency content changes over time, such as music, speech, and, as in your case, heart sounds. It helps to identify the temporal occurrence of various frequency components and is a standard tool in signal processing for analysis and feature extraction purposes.

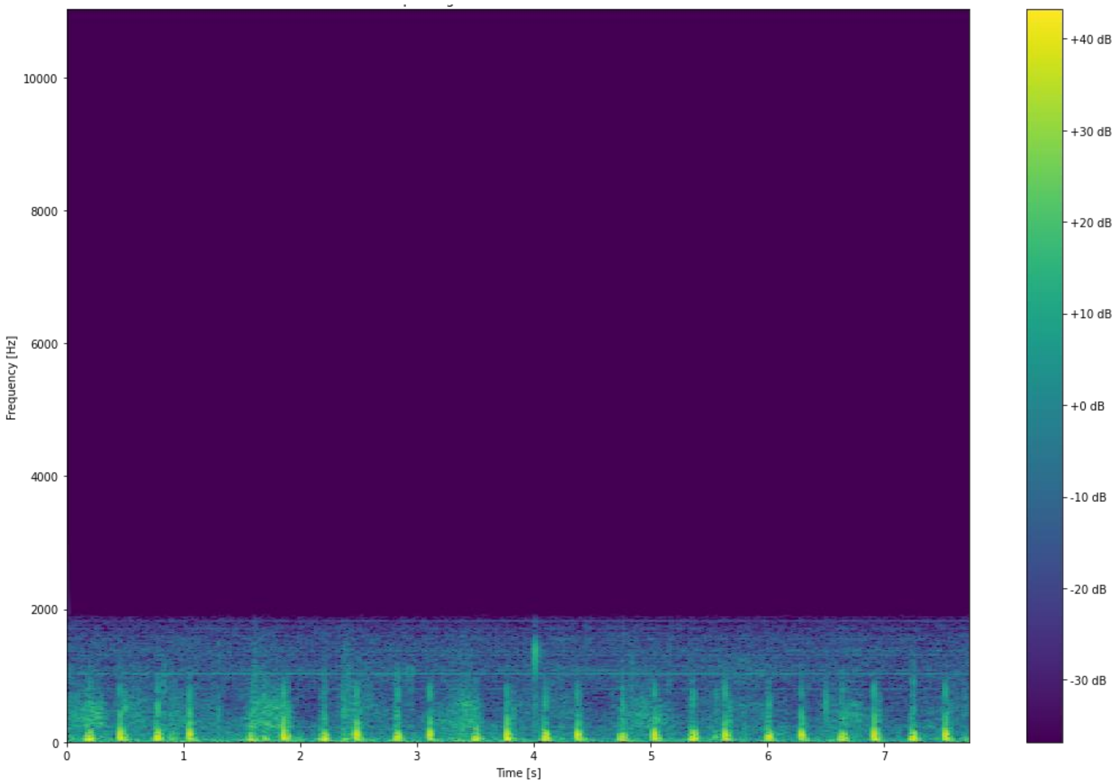


Figure 7. Spectrogram of Normal Heart Sound

1. **Dominant Frequencies**: The bright horizontal bands near the bottom of the spectrogram suggest that there are dominant low-frequency components in the audio signal, which is common in heart sounds due to the nature of the physiological events they represent.

2. **Rhythmic Patterns**: The repeated vertical patterns indicate a regular, periodic nature of the heart sounds, corresponding to the beating of the heart. Each vertical line likely represents a heartbeat.

3. **Intensity Variations**: The color variations from dark to bright represent the intensity (or loudness) of the frequencies at each point in time. The brighter colors in the lower frequencies indicate stronger intensity, which diminishes as you move to higher frequencies.
4. **Background Noise**: The relative absence of bright colors in the upper part of the spectrogram suggests that there is little to no high-frequency noise or that it is of lower intensity compared to the heart sounds.

Mel-Frequency Cepstral Coefficients (MFCCs) Visualization: Mel-Frequency Cepstral Coefficients (MFCCs) are a widely used feature set for sound processing, especially in speech recognition and audio classification tasks. The process of extracting MFCCs involves several steps that convert the raw audio signal into a form that reflects the perceived pitch and timbre of the sound, closely mirroring human auditory perception.

MFCC calculation:

1. **Frame the Signal into Short Windows**:
   - The continuous signal is divided into short frames, typically 20-40 ms long.
   - If $x(t)$ is your signal, it is sliced into frames $x_1(t), x_2(t), \ldots, x_n(t)$

2. **Apply Fourier Transform:**
   - For each frame $x_i(t)$, compute the Discrete Fourier Transform (DFT) to get the frequency spectrum.
   - $X_i(k) = \sum_{n=0}^{N-1} x_i(t). e^{\frac{-i2\pi}{N}kn}$

3. **Map the Powers of the Spectrum to the Mel Scale**:
   - Compute the power spectrum $\mathbf{P_i(k)} = |\mathbf{X_i(k)}|^2$
   - Apply Mel filter banks to the power spectra. The Mel filter bank consists of **m** triangular filters (**m** is typically 26-40) that mimic the human ear's response.
   - Each filter's response $H_m(k)$ is multiplied by the power spectrum and summed to get the Mel power spectrum: $S(m) = \sum_{k=0}^{K} P_i(k). H_m(k)$

4. **Logarithm of Mel Power Spectrum**:
   - Take the logarithm of each Mel power spectrum value to obtain log $(S(m))$
   - This step is applied because human perception of sound intensity is logarithmic.

5. **Discrete Cosine Transform (DCT)**:
   - Apply DCT to the log Mel power spectrum to get the cepstral coefficients.
   - $MFCC(m) = \sum_{k=1}^{M} \log\big(S(k)\big). \cos\,[m\ (k - 0.5)\,\frac{\pi}{M}]$
   - Typically, only the 2nd through 13th coefficients are used as they represent the most significant characteristics of the sound.

6. **Final MFCCs**:
   - The amplitudes of the spectrum after the DCT represent the MFCCs.

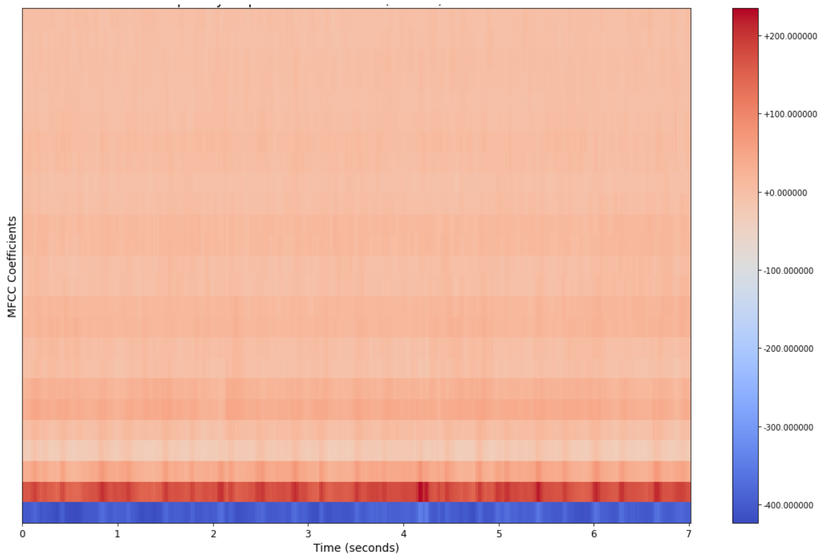


Figure 8: Mel-Frequency Cepsrtral Coefficients (MFCCs) Heart Sound

1. **Low-Frequency Emphasis**: The bottom of the image, which corresponds to the lower MFCC coefficients, shows more intensity (darker colors) compared to the upper part. This suggests that the most significant energy of the heart sound is concentrated in the lower frequencies, which is typical for heartbeats due to the nature of the sounds produced by the heart valves and chambers.

2. **Consistent Patterns**: The uniformity across the time axis indicates that the heart sound is consistent and stable over the time frame visualized. This is expected in a normal heart sound without any arrhythmia or other abnormalities.

3. **No Significant Variations**: The absence of stark changes in the pattern or color intensity suggests that there are no sudden changes in the heart sound characteristics, such as might be caused by murmurs or other pathologies.

4. **Higher Coefficients**: The upper part of the image, representing higher MFCCs, shows less intensity, which indicates that the finer details of the sound (which often include noise and subtle nuances) are not as prominent in the overall sound profile. This is common as the lower MFCCs capture the bulk of the signal's energy and are more relevant for distinguishing between different types of heart sounds.

Figure 9. Comparative Visualization of Normal, Murmur, Extrasystole, Artifact and Extra heart sounds

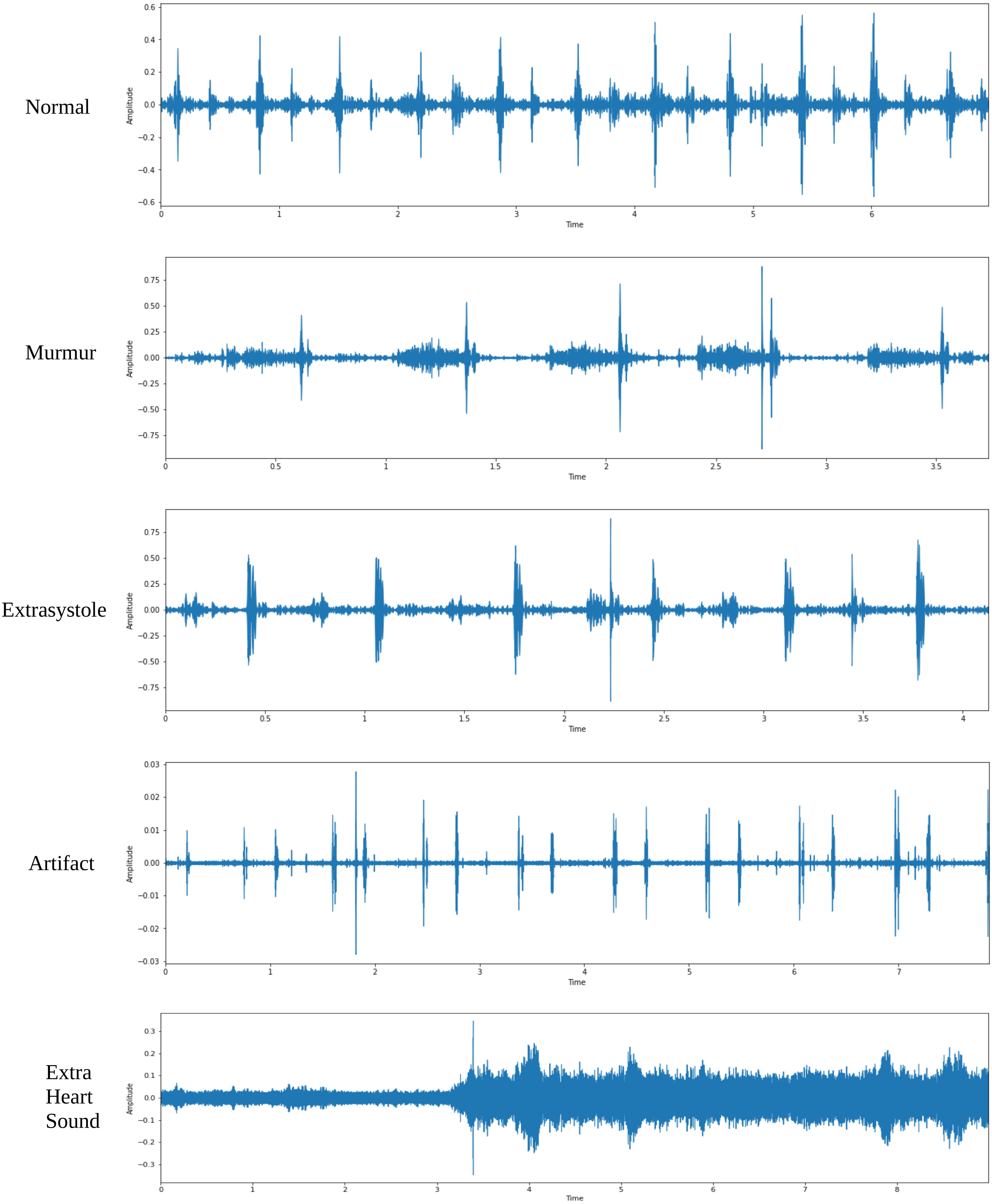
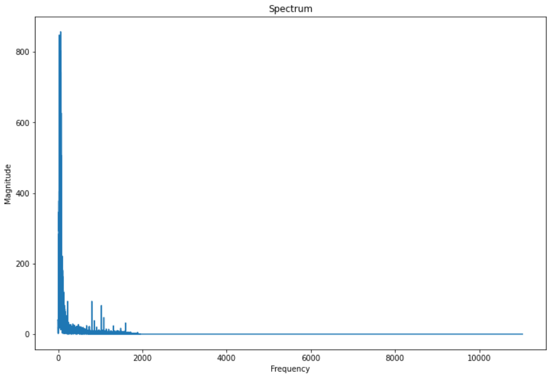
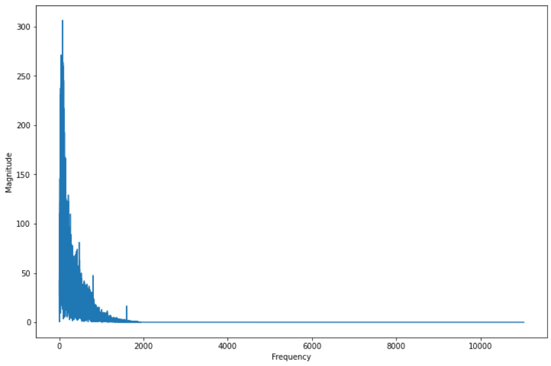


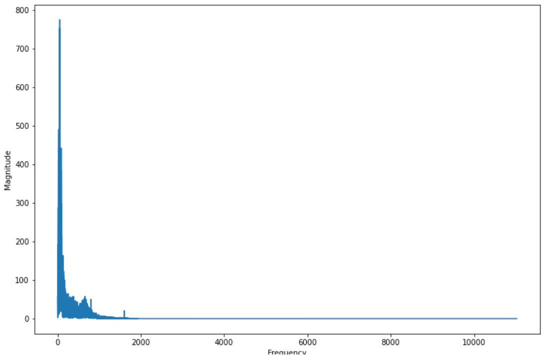
Figure 10. Spectrum of Normal, Murmur, Extrasystole, Artifact and Extra heart sounds



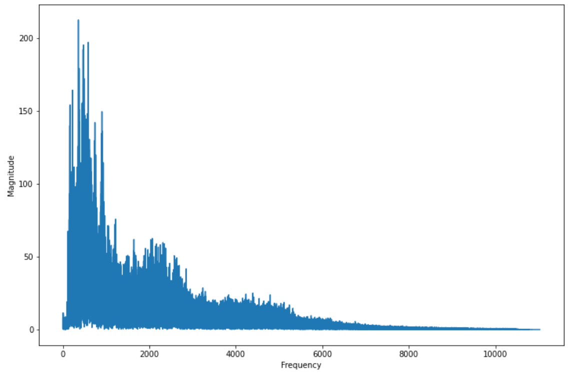
Normal



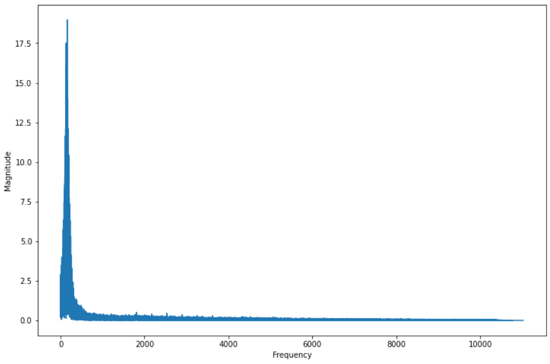
Murmur



Extrasystole

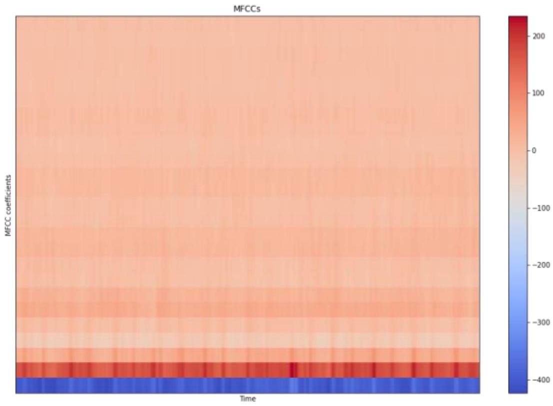


Artifact

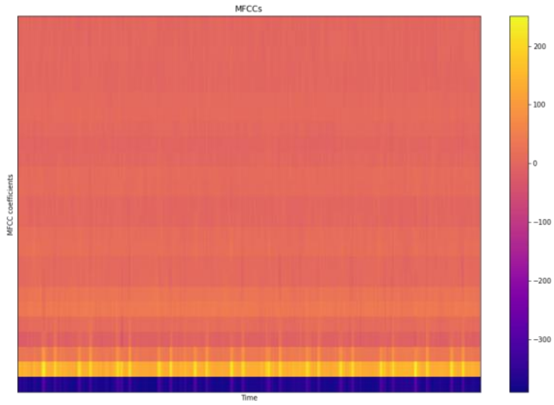


Extra Heart Sound

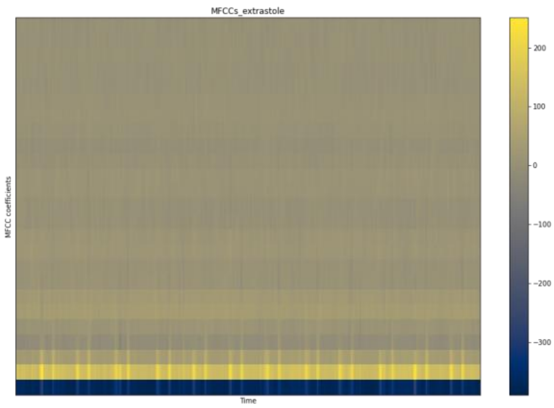
Figure 11. MFCC of Normal, Murmur, Extrasystole, Artifact and Extra heart sounds



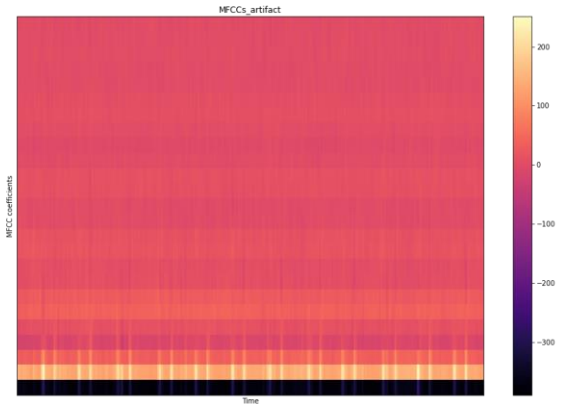
Normal



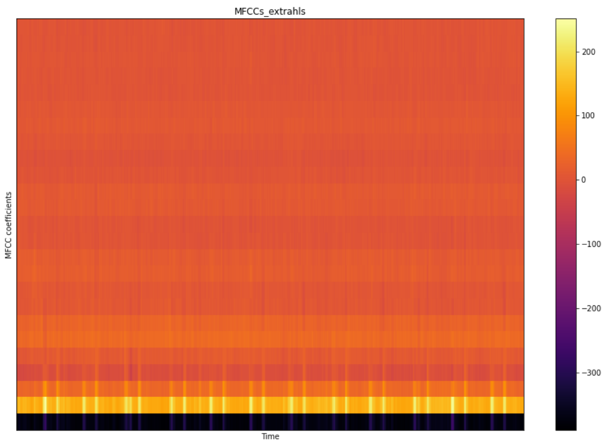
Murmur



Extrasystole



Artifact



Extra Heart Sound

3.3 Data Preprocessing

1. **Encoding Categories**:

    - The heart sound categories are simplified into three types: 'artifact', 'murmur', and 'normal'.
    - The 'normal' category includes sounds labeled as 'extrahls' and 'extrastole', which are types of normal heart sounds with additional characteristics.

2. **Mapping Labels**:
    - A dictionary called **label_to_int** is created to map text labels to integers, which is a common practice in machine learning to convert categorical labels into a format that algorithms can work with.
    - A reverse mapping **int_to_label** is also created to translate back from the integer labels to the original text labels, which is useful for interpreting the model's output.

3. **Sampling Rate:**

    - The audio data is processed at a sampling rate of 22,050 Hz. This is a high-quality sampling rate that captures the necessary range of audible frequencies for heart sounds.

4. **Maximum Clip Duration:**
    - Each audio file is standardized to a maximum duration of 10 seconds. This ensures uniformity in the length of the audio clips being analyzed.

Sound Type	File Pattern	Label
Artifact	'artifact*.wav'	0
Normal	'normal*.wav'	2
Extrahls	'extrahls*.wav'	2
Murmur	'murmur*.wav'	1
Extrastole	'extrastole*.wav'	2

5. **Feature Extraction from audio files using MFCC:**

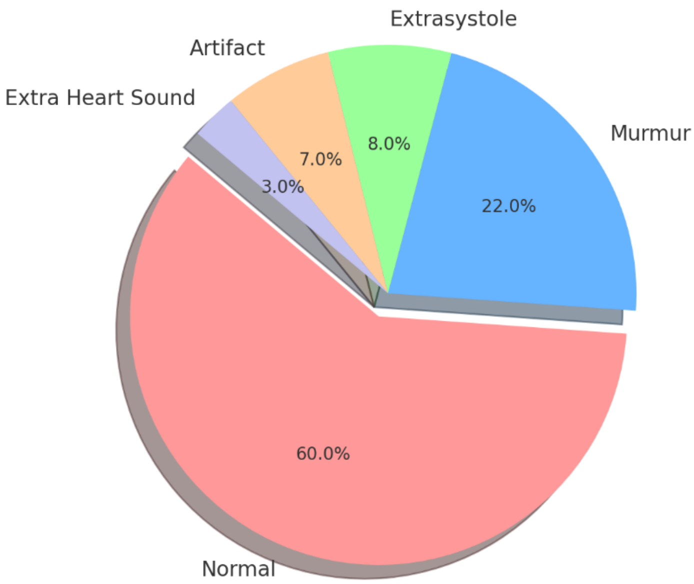
    - **Function Purpose**: Load audio files and extract MFCC features.
    - **Parameters**: **folder** (file directory), **file_names** (list of audio files), **duration** (audio length to process), **sr** (sampling rate).
    - **Process Flow:**
        1. Calculate input_length based on duration and sr.
        2. Initialize data list to store MFCCs.
        3. Loop through file_names to process each audio file.
        4. Load each audio file with librosa.load, applying duration limit.

5. If audio is shorter than duration, pad it to input_length.
6. Compute MFCCs, take the mean across time to form a feature vector.
7. Handle exceptions during loading and feature extraction.
8. Append the feature vector to data.

- **Output**: Returns **data**, a list of numpy arrays containing MFCC features.

**[Note – After extracting the features dataset A and B were concatenated.]*

6. Class Imbalance Handling



Heart Sound Data Distribution

Data Imbalance Issue: The pie chart visualization indicates an imbalanced distribution of classes in the dataset, with 'Normal' sounds being the majority class (60%), followed by 'Murmur' (22%), 'Extrastole' (8%), 'Extra Heart Sound' (3%), and 'Artifact' (7%). Imbalance can bias a model towards the majority class, and it can perform poorly on minority classes.

Calculating Class Weights:

- **TRAIN_IMG_COUNT** is the total number of images or samples in the training dataset.
- **COUNT_0**, **COUNT_1**, and **COUNT_2** represent the number of samples in each class for 'Artifact', 'Murmur', and 'Normal', respectively.

Weight Formula: The weights for each class are calculated using the formula:

- $$weight\ for\ class = \frac{Total\ samples}{(Number\ of\ classes * Samples\ in\ class)}$$

- This formula helps to balance the classes by assigning a larger weight to classes with fewer samples. It's a common technique to compensate for imbalance in datasets.

Assigning Weights:

- **weight_for_0**, **weight_for_1**, and **weight_for_2** are the calculated weights for the 'Artifact', 'Murmur', and 'Normal' classes, respectively.
- These weights are inversely proportional to the class frequencies in the dataset. A smaller class will get a higher weight.

Creating Class Weight Dictionary:

- The **class_weight** dictionary maps each class label to its calculated weight.
- This dictionary is used during the model training process to weigh the loss function, effectively giving more importance to misclassified samples from the minority class.

By applying these class weights during training, the model's learning algorithm will "pay more attention" to the minority classes, helping to improve the model's performance on those classes and aiming for a more balanced classification accuracy across all classes.

3.4 Modelling

We moved forward with **LSTM (Long Short Term Memory)**: The use of LSTMs for classifying heart sounds is motivated by their ability to handle sequential data with long-term dependencies, their robustness to the vanishing gradient problem through their gating mechanisms, and their capacity to model complex relationships within time series data.
Here’s why LSTMs are chosen for tasks like classifying heart sounds:

1. **Sequence Data**: Heart sound recordings are sequential data where the temporal order and context are crucial for accurate classification. LSTMs excel in handling sequences due to their recurrent connections.
2. **Long-term Dependencies**: Heart sounds can have patterns and anomalies that span over varying time intervals. LSTMs are designed to remember information for long periods, which is key in capturing these dependencies.
3. **Avoiding Vanishing Gradient Problem**: Traditional RNNs suffer from the vanishing gradient problem, where gradients shrink and vanish as they are propagated back through time. This makes it hard for the network to learn long-range dependencies. LSTMs overcome this with their gating mechanisms.
4. **Gating Mechanisms**: LSTMs contain three types of gates: input, forget, and output gates. These gates determine what information should be kept or discarded at each step of the sequence, allowing the network to maintain or forget information over long time spans.
5. **Modelling Complex Patterns**: Heart sounds can be quite complex and might contain subtle features that are indicative of specific heart conditions. LSTMs can model these complex patterns and relationships within the audio signals.

6. **Bidirectional Context**: Using a Bidirectional LSTM allows the network to have additional context by processing the data in both forward and backward directions. This can provide a more complete understanding of the temporal dynamics of the heart sounds.
7. **Robustness to Noise**: Heart sound recordings often contain background noise or other artifacts. LSTMs can be more robust to such noise and focus on the relevant sequential patterns that represent the heart sound characteristics.

In summary, the use of LSTMs for classifying heart sounds is motivated by their ability to handle sequential data with long-term dependencies, their robustness to the vanishing gradient problem through their gating mechanisms, and their capacity to model complex relationships within time series data. These characteristics make them particularly suitable for medical signal processing tasks, where the temporal structure of the data carries significant diagnostic information.

LSTM (Long Short Term Memory)

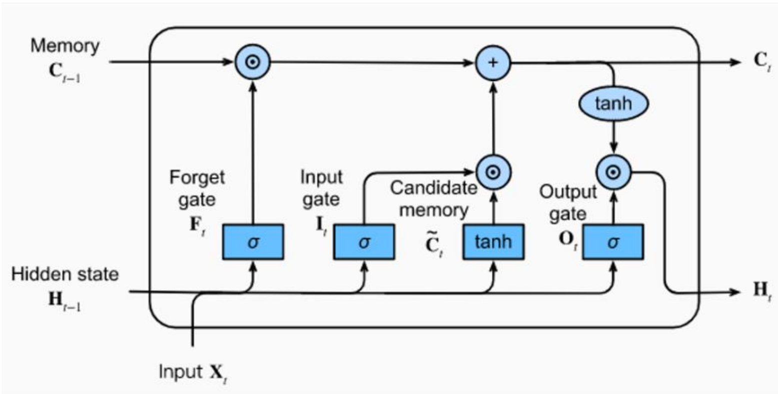


Figure 12. LSTM Model

LSTM networks have a unique structure that allows them to regulate the flow of information using gates. These gates can be thought of as a way to selectively control the access to information, similar to how a lock on a canal controls water flow. Each gate within an LSTM cell performs a specific function:

1. **Input Gate**: It decides what new information is going to be stored in the cell state. It involves a sigmoid activation function that filters the values (between 0 and 1), and a pointwise multiplication with the candidate values (generated by a tanh layer) to update the state.
2. **Forget Gate**: This gate determines what information is going to be thrown away from the cell state. It looks at the previous state h_{t-1} and the current input x_t and passes them through a sigmoid function. If the output of the sigmoid function for a particular cell state value is close to 0, it means forget this information; if it's close to 1, keep this information.
3. **Output Gate**: After the cell state has been updated, the output gate controls what part of the cell state you're going to output. This output will be based on the filtered version of the cell state and will be used for predictions or as a part of a sequence in sequence-to-sequence tasks.

The LSTM gates have weights that are learned during the training process, which allows the network to learn what information is important to keep or discard.

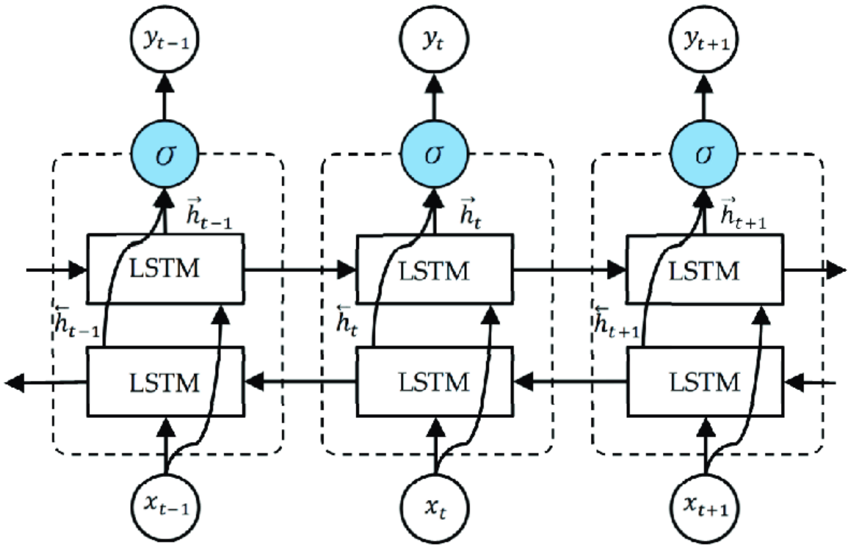
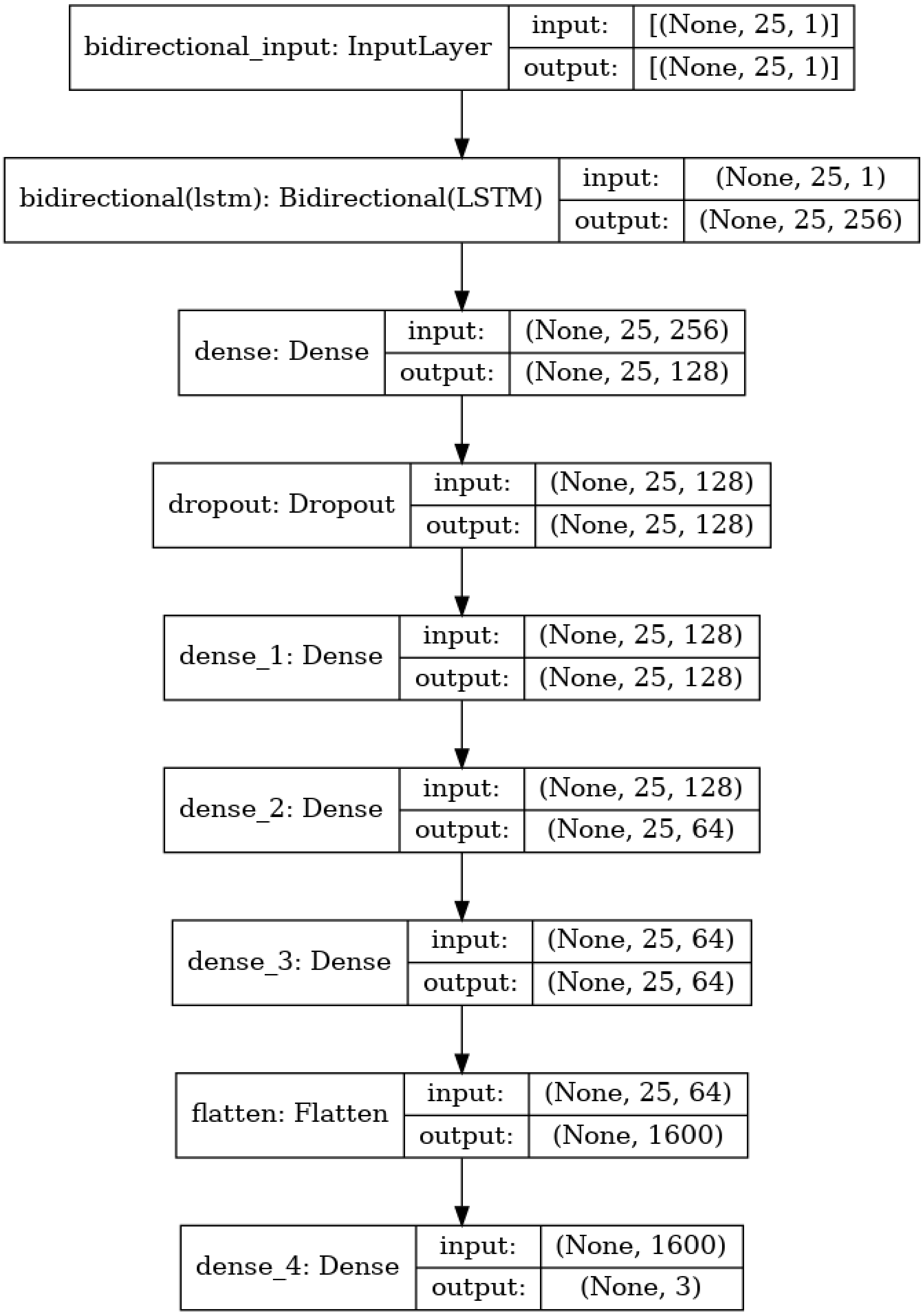


Figure 13. Bidirectional LSTM

A Bidirectional LSTM (BiLSTM) is an extension of the traditional LSTM that can improve model performance on sequence classification problems. In a standard LSTM, the input sequence is processed in a forward direction, i.e., from the first element to the last. However, there can be instances where the context of the entire sequence is needed to make accurate predictions, and this context can be captured in both directions—from beginning to end, and from end to beginning. This is where BiLSTMs come into play.

In a BiLSTM, two LSTM networks are utilized. One processes the data in a forward direction (as normal), and the other processes the data in a reverse direction. The outputs of these two networks are combined at each time step. There are several ways to combine these outputs, such as concatenation or summing.

Figure 15. The complete Architecture



Here is the model architecture summarized in a table format:

Layer Type	Output Shape	Number of Parameters
Bidirectional	(None, 25, 256)	133,120
Dense	(None, 25, 128)	32,896
Dropout	(None, 25, 128)	0
Dense	(None, 25, 128)	16,512
Dense	(None, 25, 64)	8,256
Dense	(None, 25, 64)	4,160
Flatten	(None, 1600)	0
Dense (Output)	(None, 3)	4,803
Total		**199,747**
Trainable Parameters		**199,747**
Non-trainable Parameters		**0**

Training

The key hyperparameters used in the model training:

- **Loss Function**: Categorical Crossentropy

$$Categorical\ cross\ entropy\ loss = -\sum_{i=1}^{N} y_i \log{(\hat{y}_i)}$$

- **Optimizer**: Adam with a learning rate of 0.0001
- **Batch Size**: 3
- **Epochs**: 30

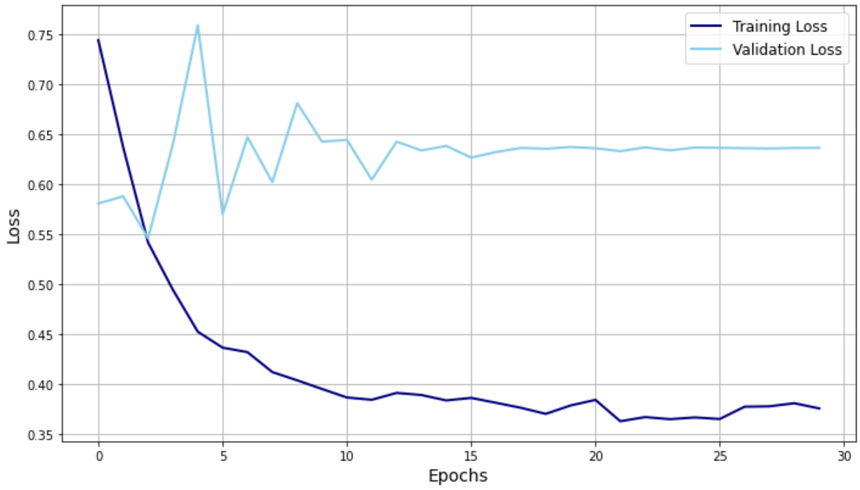


Figure 16. Training and Validation Loss Over Epochs

1. **Initial Learning**: The training loss starts high and shows a significant drop early in the training process, indicating that the model is learning and improving rapidly from its initial state.
2. **Convergence**: As epochs increase, the training loss decreases and starts to plateau, suggesting that the model is converging and making incremental improvements after the initial learning phase.
3. **Validation Loss**: The validation loss decreases alongside the training loss, but it plateaus at a higher value than the training loss. This behavior is typical and indicates that the model is generalizing to new data but is not fitting the validation set as closely as the training set.
4. **Overfitting**: There doesn't appear to be a significant increase in validation loss relative to the training loss as training progresses, which would be a sign of overfitting. Hence, overfitting is not immediately evident from this plot.
5. **Stability**: The validation loss seems relatively stable and does not exhibit high variance, which is a good sign that the model's performance is stable across different subsets of the data.

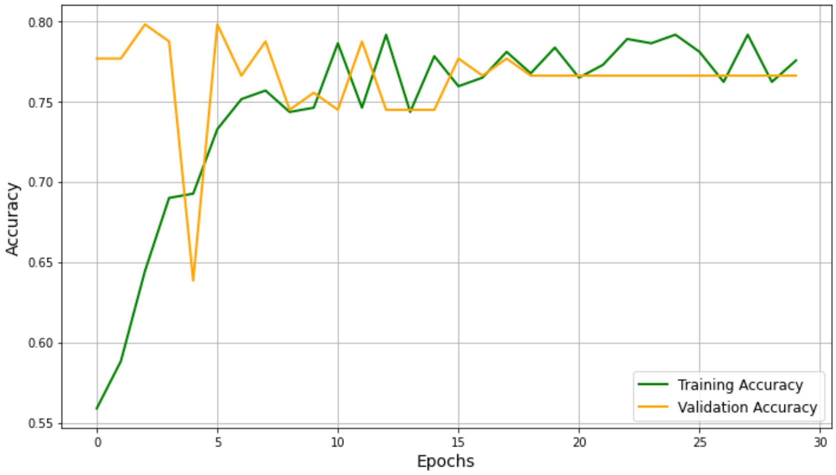


Figure 16. Training and Validation Accuracy Over Epochs

Classification report

Class	Precision	Recall	F1-Score	Support
Artifact	1.00	1.00	1.00	8
Murmur	0.42	0.50	0.46	22
Normal	0.87	0.83	0.85	87
Accuracy			**0.78**	117
Macro	0.76	0.78	0.77	117
Weighted Avg	0.79	0.78	0.78	117

4. Results

Our study employed a deep learning approach to classify heart sound recordings into multiple categories. The model's architecture was a Bidirectional Long Short-Term Memory (BiLSTM) network, which utilized categorical crossentropy as the loss function and the Adam optimizer with a learning rate of 0.0001. The training was conducted over 30 epochs with a batch size of 3. Class weights were applied to address the imbalance in the dataset.

The training process revealed a consistent decrease in loss over epochs, with the training loss starting from a higher magnitude and showing a rapid decline, indicating effective learning. The validation loss decreased concurrently but plateaued at a higher value compared to the training loss, which is a typical behavior suggesting the model's ability to generalize.

Accuracy metrics showed the model achieved an overall accuracy of 78% on the test set. The precision and recall rates for the 'artifact' category were perfect, with both scoring 1.00. The 'murmur' category had a precision of 0.42 and a recall of 0.50, reflecting a need for improvement in distinguishing this particular class. The 'normal' heart sounds were identified with high precision and recall rates of 0.87 and 0.83, respectively.

The F1-scores, which balance precision and recall, were highest for 'artifact' (1.00) and lowest for 'murmur' (0.46), with 'normal' at 0.85. The macro average F1-score across all categories was 0.77, while the weighted average, which accounts for class imbalance, was 0.78, in line with the overall accuracy.

Visual analysis through plotted graphs indicated a stable learning curve without signs of overfitting. The model's performance remained consistent across epochs, which is indicative of a reliable classification process.

These results demonstrate the potential of using BiLSTM networks for heart sound classification tasks. Future work could explore the refinement of model architecture, further tuning of hyperparameters, and the incorporation of additional features to enhance the classification of murmur sounds.

5. Future Work

The current project has established a strong foundation for classifying heart sounds using a BiLSTM model. However, there are several avenues for future exploration that could potentially enhance the performance and robustness of the classification system. The following directions are particularly promising:

1. **Incorporation of Attention Mechanisms**: Integrating an attention layer could allow the model to focus on the most relevant parts of the heart sound sequences, potentially improving its ability to discern subtle differences between classes. This could be

particularly beneficial for distinguishing murmurs, which have lower precision and recall in the current model.

2. **Adoption of Transformer Architectures**: Transformers have revolutionized the field of sequence modeling. Their self-attention mechanism is well-suited for handling long-range dependencies in time-series data, such as heart sound recordings. Applying a transformer model could lead to significant improvements in classification accuracy.

3. **Ensemble Models**: Combining the strengths of CNNs and transformers in an ensemble model could offer complementary perspectives on the data. While CNNs are adept at capturing local patterns in data, transformers excel at understanding the entire context. An ensemble approach could leverage these strengths to improve overall classification performance.

4. **Vision Transformers (ViTs)**: Vision Transformers, which apply transformer models to image classification tasks, can also be adapted for audio classification by converting audio signals into spectrogram images. This approach could unveil new patterns in the frequency-time representations of heart sounds that are not easily captured by traditional methods.

5. **Advanced Preprocessing Techniques**: Exploring advanced signal processing techniques could lead to the extraction of more informative features from the raw audio data, which could, in turn, improve model learning.

6. **Larger and More Diverse Datasets**: Expanding the dataset to include a wider variety of heart sounds from diverse demographics could improve the model's generalizability and robustness.

7. **Real-world Testing and Deployment**: Future work could also focus on deploying the model in real-world clinical settings to validate its efficacy in practice. This could include integrating the model into handheld devices or mobile applications for early screening and diagnosis.

8. **Interpretable AI**: Enhancing model interpretability to understand the decision-making process could build trust in AI-powered diagnostic systems among clinicians.
   The potential extensions outlined above could not only refine the model's predictive accuracy but also broaden the scope of its applicability in the field of healthcare, moving towards a more proactive and patient-centric model of care.

6. Contribution

This project was a collaborative effort between Manish Kumar Roy and Ankit Kumar Prem. The contributions of each member to the project are as follows:

Manish Kumar Roy:

- Conducted initial research on heart sound classification methods and the application of LSTM in this field.
- Led the data collection process, including sourcing and preprocessing the audio files for model training.
- Developed the core LSTM model architecture and implemented the training procedures.
- Managed the debugging and optimization of model parameters to improve accuracy.
- Took the lead in writing the initial draft of the project report, including the literature review and methodology sections.

Ankit Kumar Prem:

- Assisted in the research phase, focusing on the review of LSTM and its efficacy in sequence modeling tasks.
- Contributed to the data preprocessing, particularly in noise reduction and feature extraction from audio data.
- Participated in the model development by fine-tuning the LSTM layers and experimenting with hyperparameters.
- Conducted extensive model evaluations, including performance metrics analysis and validation.
- Contributed to the project report, focusing on the results, discussion, and conclusion sections.

Both of Us were involved in regular meetings to discuss the project progress, challenges encountered, and strategies for moving forward. They also equally participated in the preparation and delivery of the project presentation.

7. References

7.1 Rachel Hajar, M.D "Risk Factors for Coronary Artery Disease" in Proceedings of the National Libraray of Medicine, pp PMC5686931, September 2017

7.2 J. Low and K. Choo, "Classification of Heart Sounds Using Softmax Regression and Convolutional Neural Network," in Proceedings of the International Conference on Communication Engineering and Technology, pp. 18–21, February 2018.

7.3 Yaseen, G.Y.S.; Kwon, S. Classification of Heart Sound Signal Using Multiple Features. Appl. Sci. 2018, 8, 2344

7.4 Fatih Demir, Abdulkadir Şengür, Varun Bajaj, and Kemal Polat "Towards the classification of heart sounds based on convolutional deep neural network" in Proceedings of the National Libraray of Medicine, pp PMC6684704, December 2019

7.5 Pantea Keikhosrokiani,corresponding author A. Bhanupriya Naidu A/P Anathan, Suzi Iryanti Fadilah, Selvakumar Manickam, and Zuoyong Li "Heartbeat sound classification using a hybrid adaptive neuro-fuzzy inferences system (ANFIS) and artificial bee colony" in Proceedings of the National Libraray of Medicine, pp PMC9841877, January 2023

7.6 Detecting Heart Anomalies Using Mobile Phones and Machine Learning

7.7 S. Prasadh, S. Natrajan, and S. Kalaivani, "Efficiency analysis of noise reduction algorithms: Analysis of the best algorithm of noise reduction from a set of algorithms," in Proceedings of the International Conference on Inventive Computing and Informatics, pp. 1137–1140, November 2017

7.8 S. Latif, M. Usman, R. Rana, and J. Qadir, "Phonocardiographic Sensing using Deep Learning for Abnormal Heartbeat Detection," in IEEE Sensors Journal, vol. 18, issue 22, pp. 9393-9400, September 2018.

7.9 Hanna, I.R.; Silverman, M.E. A History of Cardiac Auscultation and Some of Its Contributors. Am. J. Cardiol. 2002, 90, 259–267

7.10 Jiang, Z.; Choi, S. A Cardiac Sound Characteristic Waveform Method for In-Home Heart Disorder Monitoring with Electric Stethoscope. Expert Syst. Appl. 2006, 31, 286–298.

7.11 M. Al-Mardini, F. Aloul, A. Sagahyroon, and L. Al-Husseini, "Classifying Obstructive Sleep Apnea Using Smartphones," in Elsevier Journal of Biomedical Informatics, vol. 52, pp. 251–259, December 2014

7.12 Ali Raza, Arif Mehmood, Saleem Ullah, Maqsood Ahmad, Gyu Sang Choi,and Byung-Won "Heartbeat Sound Signal Classification Using Deep Learning" in Proceedings of MDPI, pp 1424-8220/19/21/4819, October 2019