# Multiple channel Heartbeat audio anomaly Classification

Project Report submitted in partial fulfillment of the requirements for the award of the degree of **Master of Science (Five Year Integrated)** in **Computer Science (Artificial Intelligence & Data Science)**

by


**OMAL S**

(Register No : 80521015)


Under the Supervision of,


**Dr. Shailesh S.**

Assistant Professor

Department of Computer Science

**Department of Computer Science**

**COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY**


**JULY 2024**

# Certificate

*This is to certify that the project report for* ***21-805-0607: Project*** *on the topic* ***Multiple channel Heartbeat audio anomaly Classification*** *is a record of work carried out by* ***OMAL S (80521015)****, in partial fulfillment of the requirements for the award of degree in* ***Master of Science (Five Year Integrated)*** *in* ***Computer Science (Artificial Intelligence & Data Science)*** *of Cochin University of Science and Technology (CUSAT), Kochi. The project report has been approved as it satisfies the academic requirements in respect of the sixth semester project prescribed for the Master of Science (Five Year Integrated) in Computer Science degree.*

**Faculty Member in-charge**

Dr. Shailesh S
Assistant Professor
Department of Computer Science
CUSAT

Dr. Madhu S. Nair
Professor and Head
Department of Computer Science
CUSAT

# Declaration

I hereby declare that the project report entitled **"Multiple channel Heartbeat audio anomaly Classification"** submitted as part of the semester sixth curriculum for the Master of Science (Five Year Integrated) in Computer Science (Artificial Intelligence & Data Science) at Cochin University of Science and Technology is my original work. This written submission represents my ideas in my own words. Where others' ideas and words have been included, I have adequately cited and referenced the original source. I declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented, fabricated, or falsified any idea, data, fact, or source in my submission. I understand that any violation of the above will cause disciplinary action by the Institute and can also evoke penal action from the source which has not been properly cited or from whom proper permission has not been taken when needed.

**OMAL S**

Roll No.: 80521015

Date: 22 July, 2024

Place: CUSAT , Kalamassery

# Acknowledgement

I am profoundly grateful to the almighty for bestowing upon me the strength and perseverance to successfully complete this semester six research project.

I express my deepest gratitude to Dr. Shailesh S, whose expertise, invaluable guidance, and constructive feedback have been vital in helping me complete this project successfully. I am grateful for the opportunity to work under his supervision, which has undoubtedly enriched my research knowledge.

Additionally, I express my heartfelt acknowledgement to Dr. Madhu S. Nair, Head of Department at CUSAT, for his support and encouragement throughout my academic and research endeavors.

I would also like to extend my sincere gratitude to the PhD scholars who provided invaluable assistance and insights that greatly contributed to the success of this project. Their support and knowledge were instrumental in overcoming various challenges during the research process.

Finally, I extend my gratitude to all the individuals and institutions who directly or indirectly contributed to the completion of this project. Your support and encouragement have been invaluable.

# Abstract

Heart disease remains the leading cause of mortality globally, necessitating innovative methods for early detection and diagnosis.This study introduces a novel method for classifying heartbeat audio anomalies using a multiple channel approach with Convolutional Neural Networks (CNNs). The proposed system combines time-domain and frequency-domain features extracted from heartbeat audio with spectrogram representations to improve classification accuracy. The CNN model processes these diverse inputs to classify audio into five categories: Normal, Murmur, Extra Heart Sound, Artifact, and Extrasystole.

Our approach was benchmarked against single-channel methods, showing significant enhancements in performance. The multiple channel CNN model achieved an overall accuracy of 98.27%, with high precision and F1-scores, demonstrating its effectiveness in distinguishing subtle anomalies in heartbeat sounds. This improvement underscores the advantages of integrating multiple input channels and advanced feature extraction techniques.

Future work will focus on incorporating Class Activation Maps (CAMs) and Grad-CAM to provide visual explanations of the model's predictions, aiming to enhance interpretability and trust. This research contributes to advancing automated cardiac diagnostics and holds promise for transforming early detection and patient care practices in cardiovascular health.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Overview

Heartbeat audio analysis has become a pivotal aspect of medical diagnostics, offering a non-invasive method to detect cardiovascular conditions. Traditional methods rely heavily on the expertise of medical professionals, which can be subjective and prone to variability. Recent advancements in machine learning, particularly convolutional neural networks (CNNs), have opened new avenues for automating and enhancing the accuracy of heartbeat audio classification. This thesis focuses on developing a robust system for multiple channel heartbeat audio anomaly classification using features and spectrogram channels with CNNs.

## 1.2 Motivation

The accurate diagnosis of cardiovascular conditions is critical for timely and effective treatment. Traditional diagnostic methods, reliant on the auditory skills of medical professionals, can be subjective and vary widely between practitioners. This variability can lead to inconsistent diagnoses and delayed treatment, highlighting the need for more objective and standardized diagnostic tools. The complexity of heart sounds, coupled with the subtle nature of certain anomalies, necessitates advanced analytical techniques that can detect patterns beyond human auditory capabilities.

The advent of machine learning, particularly convolutional neural networks (CNNs), has introduced powerful methods for analyzing complex audio signals. CNNs excel at identifying intricate patterns and features within data, making them ideal for processing and interpreting heartbeat sounds. By utilizing multiple channels of heartbeat audio, we can capture a more comprehensive representation of the heart's activity, leading to improved detection of anomalies. Spectrograms, which provide a visual representation of the audio signal's frequency components over time, offer a rich source of information for CNNs to analyze. This combination of advanced machine learning techniques and detailed audio representations holds significant potential for improving the accuracy and reliability of heartbeat anomaly detection. The motivation behind this research is to leverage these advancements to develop a robust, scalable system that can assist medical professionals in diagnosing cardiovascular conditions with greater precision and consistency.

## 1.3 Objectives

The primary objective of this research is to develop a novel system for classifying anomalies in heartbeat audio data using explainable deep learning techniques. This system aims to improve the accuracy, reliability, and transparency of cardiovascular diagnostics. The specific goals of this research are as follows:

- Leveraging Multiple Physiological Signals: To enhance the robustness and precision of anomaly detection, the system will incorporate multiple channels of heartbeat audio and potentially other physiological signals. This multi-modal approach aims to provide a comprehensive analysis of heart activity, capturing subtle anomalies that might be missed by single-channel methods.

- Improved Trust and Transparency: A significant focus of this research is to ensure that the deep learning models used are not only accurate but also interpretable. By employing explainable AI techniques, the system will offer insights into the decision-making process of the neural networks, allowing medical professionals to understand and trust the results.

- Potential for Broader Applicability: The developed system is intended to be versatile and adaptable to various clinical settings. By providing a scalable solution that can be integrated into existing healthcare infrastructures, this research aims to facilitate widespread adoption and application, ultimately contributing to better patient outcomes in cardiovascular care.

## 1.4 Scope

The scope of this research encompasses the development, implementation, and evaluation of a system for classifying anomalies in heartbeat audio using multiple channels and advanced deep learning techniques. The specific aspects covered in this research include:

- Data Collection and Preprocessing:
  Recording high-fidelity heartbeat audio using digital stethoscopes across multiple channels to capture comprehensive heart sounds, segmenting and preprocessing the audio data to create a consistent and clean dataset for model training and evaluation, extracting both time-domain and frequency-domain features from the heartbeat audio data.

- Spectrogram Generation and Analysis:
  Converting segmented heartbeat audio into spectrograms using Short-Time Fourier Transform (STFT) or similar techniques and representing each audio channel as a

separate spectrogram channel to provide a multi-dimensional input for the neural network.

- CNN Architecture Design and Training:
  Designing a deep CNN architecture tailored for multi-channel spectrogram inputs, including convolutional layers, pooling layers, and fully connected layers and training the CNN on a labeled dataset of normal and anomalous heartbeats, using data augmentation techniques to enhance model robustness.

- Explainable AI Integration:
  Incorporating explainable AI techniques to ensure the model's decision-making process is transparent and interpretable by medical professionals and developing methods to visualize and explain the features and patterns learned by the CNN, enhancing trust and usability.

- Evaluation and Validation:
  Testing the system's performance on various datasets, including both publicly available and proprietary datasets, using metrics such as accuracy, precision, recall, and F1-score, comparing the developed system with existing methods to demonstrate its effectiveness and superiority in detecting heartbeat anomalies.

- Application and Deployment: Exploring the integration of the system into clinical workflows, ensuring compatibility with existing medical devices and software. Also assessing the potential for real-time anomaly detection in clinical settings and providing guidelines for practical deployment.

By addressing these aspects, the research aims to develop a comprehensive and practical solution for heartbeat audio anomaly classification, contributing significantly to the field of cardiovascular diagnostics and patient care.

## 1.5 Problem Statement

Accurate detection of cardiovascular anomalies through heartbeat audio remains a significant challenge due to the subjective nature of traditional diagnostic methods and their inability to consistently capture subtle anomalies. This variability underscores the need for an automated, objective diagnostic tool that can reliably identify potential issues early, before they become critical.

Recent advancements in convolutional neural networks (CNNs) offer a promising approach to analyzing complex audio signals, yet existing methods often rely on single-channel data, missing valuable information from multi-channel recordings. This research aims to address this gap by developing a novel system for classifying anomalies in multiple channel heartbeat audio using both features and spectrogram channels with CNNs, enhancing diagnostic accuracy and providing interpretable results for clinical use.

## 1.6 Need of the Study

The need for this study arises from the limitations of traditional cardiovascular diagnostics, which often rely on the subjective interpretation of heartbeat sounds by medical professionals. These methods can be inconsistent and may fail to detect subtle anomalies, leading to delayed or missed diagnoses. An automated system that provides objective, reliable analysis is crucial for improving early detection and treatment outcomes in cardiovascular care.

Convolutional neural networks (CNNs) have shown promise in analyzing complex audio signals, but most existing methods use single-channel data, which may not capture the full spectrum of information available in multi-channel recordings. By developing a system that utilizes both features and spectrogram channels from multiple heartbeat audio sources, this study aims to enhance diagnostic accuracy and provide clear, interpretable results. Such advancements are essential for creating a trustworthy and effective tool that can be integrated into clinical practice, ultimately improving patient care and diagnostic efficiency.

## 1.7 Project Organization

The project is organized into specific phases to systematically develop and evaluate a system for heartbeat audio anomaly classification. Each phase builds upon the previous one, ensuring a structured approach. The organization of the project is as follows:

1. Introduction and Literature Review
    - Objective: Establish the foundation of the study by reviewing existing methods and identifying research gaps.
    - Activities: Conduct a comprehensive literature review on heartbeat audio analysis, CNNs, and explainable AI. Define the project's goals and scope.
2. Data Preparation
    (a) Feature Extraction
        - Objective: Extract meaningful features from the heartbeat audio data.
        - Activities: Extract time-domain features (e.g., amplitude, duration) and frequency-domain features (e.g., spectral entropy, MFCCs) from the audio data.

    (b) Spectrogram Generation
        - Objective: Convert audio data into spectrograms suitable for CNN analysis.

    - Activities: Generate spectrograms from the segmented heartbeat audio using Short-Time Fourier Transform (STFT) or similar techniques. Prepare these spectrograms as input for the CNN.

3. Single Channel Implementation
    - Objective: Develop and evaluate a system using single-channel heartbeat audio data.
    - Activities:
        - Feature-Based Implementation: Use extracted features to train a machine learning model for anomaly classification.
        - Spectrogram-Based Implementation: Use single-channel spectrograms to train a CNN for anomaly classification. Evaluate the performance of both approaches.
4. Multiple Channel Implementation
    - Objective: Develop and evaluate a system using multi-channel heartbeat audio data for improved accuracy.
    - Activities:
        - Feature-Based Multi-Channel Implementation: Use features extracted from multiple audio channels to train a machine learning model.
        - Spectrogram-Based Multi-Channel Implementation: Use multi-channel spectrograms as input to train a deep CNN. Evaluate the performance of both approaches and compare them with single-channel implementations.
5. Evaluation and Validation
    - Objective: Assess the system's performance and effectiveness.

Chapter 1 Introduction

- Activities: Evaluate the system using various metrics such as accuracy, precision, recall, and F1-score. Compare the results with existing methods to demonstrate the system's advantages. Test the system on both publicly available and proprietary datasets.

# Chapter 2

# Literature Review

## 2.1 Introduction

The advancement in cardiovascular anomaly detection through heartbeat audio analysis is crucial for improving patient outcomes. Traditional methods relying on subjective medical professional expertise can introduce variability and limit diagnostic consistency, highlighting the need for automated and objective diagnostic tools. Research has shown the effectiveness of deep learning models like HbNet [13], 1D-CNN [10], and DNN in classifying heart sounds, achieving high accuracies and sensitivities, thus enhancing the early and accurate detection of cardiovascular anomalies. These models leverage techniques such as Mel Frequency Cepstral Coefficients (MFCC), Wave-Segment Synthesizing (WSS), and Discrete Wavelet Transform (DWT) to improve classification rates and provide reliable assessments of heart health, ultimately contributing to the development of more precise and efficient diagnostic systems in modern medical diagnostics.

Research in the field of heartbeat anomaly classification using audio signals has shown promising results. Various studies have explored the use of deep learning techniques, such as convolutional neural networks (CNNs)[1], for this purpose. For instance, studies have utilized Mel-frequency cepstral coefficients (MFCC) and Mel spectrogram values[15] for feature extraction in heartbeat sound analysis [9]. Additionally, the application of CNNs in processing heartbeat sounds has demonstrated significant improvements in anomaly detection accuracy, outperforming traditional machine learning methods like support vector machines (SVM) and k-nearest neighbor (KNN) techniques [3].

## 2.2 Background

Heart sound analysis, a crucial aspect of cardiovascular health assessment, has seen significant advancements through the integration of deep learning and machine learning techniques [18] . These technologies enable the detection of heart sounds, murmurs, and abnormalities, providing valuable insights into cardiac conditions. Various methods such as Hidden Markov Models [2], short-time Fourier transformation, wavelet transformation, and Log-Mel transformation have been employed to extract features from phonocardiograms[7] for accurate analysis.

In the realm of audio analysis, machine learning techniques, particularly Convolutional Neural Networks (CNNs), have garnered significant attention for their efficacy in tasks like Acoustic Scene Classification (ASC) [15]. Studies have shown that combining CNNs with Long Short-Term Memory (LSTM) networks can enhance the recognition and categorization of acoustic events in complex audio environments, achieving high accuracies of 91% and 88% for one-fold validation and 10-fold cross-validation, respectively [16]. Furthermore, research on Speaker Recognition (SR) has compared deep learning methods like custom CNNs with pre-trained nets, demonstrating the superiority of custom CNNs in achieving accurate results on spectrograms and MFCC graphs [12]

Single-channel and multi-channel audio analysis have been extensively studied in recent research. Single-channel methods for tasks like DRR estimation have shown significant advancements using deep neural networks, achieving superior performance compared to multi-channel approaches. On the other hand, multi-channel ASR systems benefit from end-to-end training but face challenges due to limited real-world multi-channel speech data, leading to the exploration of utilizing single-channel data to enhance performance. [5]

Spectrograms are widely used in audio analysis for tasks like music genre classification [17], speaker characteristic identification [11], and rural acoustic landscape analysis. They provide valuable insights by representing voice information in image form, aiding in fault diagnosis and feature extraction. Feature extraction methods like Mel Frequency Cepstral Coefficients (MFCC) have shown high accuracy in music genre classification, while various algorithms and techniques are employed for speaker characteristic classification, achieving accuracies of up to 93.26% for gender classification. Additionally, the use of spectrograms in rural acoustic landscape analysis has led to the development of novel filtering methods like Gabor wavelet filtering for feature extraction and analysis. These techniques contribute to improving accuracy and efficiency in audio analysis tasks, showcasing the potential for further research and development in the field [19].

## 2.3 Research Gaps

In recent years, the application of convolutional neural networks (CNNs) to heartbeat audio anomaly detection has garnered attention for its potential to automate and enhance diagnostic accuracy. Most existing studies focus on single-channel heartbeat audio or use limited feature extraction techniques, often relying on basic time-domain or frequency-domain features. While spectrograms have been used to represent audio signals visually, their application to multi-channel heartbeat audio is still emerging.

Furthermore, there is a growing interest in integrating multiple physiological signals, such as electrocardiogram (ECG) and blood pressure (BP) signals, into diagnostic systems to provide a more comprehensive assessment of cardiovascular health.

- Integration of Multi-Channel Data:
  While single-channel audio analysis using CNNs has shown promise, there is a notable gap in leveraging multi-channel heartbeat audio data. Multi-channel recordings capture a broader range of heart sounds and variations, which can improve anomaly detection. However, research on integrating multi-channel audio with CNNs remains limited, particularly in combining these channels with spectrogram representations. [8]

- Combined Use of Features and Spectrograms:
  Most studies either focus on feature-based approaches or spectrogram-based methods separately. There is a lack of comprehensive research that combines both feature extraction and spectrogram generation in multi-channel settings. Integrating these approaches could enhance the ability of CNNs to detect subtle anomalies by providing a richer and more nuanced representation of heartbeat audio.[6]

- Incorporation of Additional Physiological Signals:
  Existing research primarily focuses on heartbeat audio alone, neglecting the potential benefits of incorporating other physiological signals such as ECG and BP. These signals can offer complementary information that enhances anomaly detection. For example, ECG provides detailed electrical activity of the heart, while BP measurements offer insights into cardiovascular function. The integration of these signals with multi-channel heartbeat audio and CNNs has not been thoroughly explored.

## 2.4 Summary

The evolution of heartbeat audio analysis from subjective manual interpretation to automated methods using convolutional neural networks (CNNs). While CNNs have demonstrated their potential in analyzing single-channel heartbeat audio, there remains a significant gap in utilizing multi-channel data. Multi-channel recordings offer a more comprehensive view of heart sounds, which could enhance the accuracy of anomaly detection. However, current research has largely focused on single-channel methods and has not fully explored the integration of multi-channel audio with advanced CNN architectures.

Additionally, the review emphasizes the importance of combining feature extraction and spectrogram generation techniques within CNN frameworks. Although feature-based and spectrogram-based approaches have been studied separately, their combined application in a multi-channel context is still underdeveloped. The incorporation of additional physiological signals, such as ECG and blood pressure (BP), presents further opportunities for improving diagnostic accuracy. Addressing these research gaps and integrating explainable AI techniques can significantly advance heartbeat audio anomaly classification, making it more reliable and interpretable for clinical use.

# Chapter 3

# Methodology

In the pursuit of enhancing the accuracy and robustness of heartbeat audio anomaly classification, this study employs a novel methodology that integrates both feature extraction and spectrogram-based approaches using convolutional neural networks (CNNs). The proposed model leverages the strengths of CNNs to process and classify complex audio signals from multiple channels, thereby improving diagnostic performance.

The methodology is centered around a CNN model designed to handle multi-channel heartbeat audio data. The process begins with the extraction of meaningful features from raw audio files. These features, which include both time-domain and frequency-domain characteristics, provide a detailed representation of the heartbeat sounds. Concurrently, spectrograms are generated from the audio data to capture the frequency content and temporal variations of the heartbeat signals.

The core of the methodology involves a CNN architecture that integrates both feature-based and spectrogram-based inputs. The output is classified into one of five distinct classes, representing different types of heartbeat anomalies. The integration of features and spectrogram channels within the CNN model aims to leverage the complementary information from both approaches, resulting in a more comprehensive and accurate anomaly detection system.

## 3.1 Dataset

The dataset for this study on multiple channel heartbeat audio anomaly classification PASCAL Challenge [4] is sourced from two main avenues: the general public via the iStethoscope Pro iPhone app (Dataset A) and a clinical trial in hospitals using the digital stethoscope DigiScope (Dataset B). Dataset A includes 176 WAV files organized into four categories: Normal, Murmur, Extra Heart Sound, and Artifact. The files are divided into training sets for each category and a separate unlabelled test set. The normal heart sounds exhibit typical "lub dub" patterns, while murmurs are characterized by additional "whooshing" noises between the beats. Extra heart sounds and artifacts represent additional anomalies and unrelated noise, respectively. Dataset B consists of 656 WAV files categorized into Normal, Murmur, Extrasystole, and noisy versions of these categories. The extrasystole category includes recordings with irregular heartbeats, distinguished by abnormal timing of beats.

Both datasets feature heart sounds of varying lengths, ranging from 1 to 30 seconds, and include a mix of clean and noisy recordings. The provided segmentation data and diverse categorization make this dataset a valuable resource for training and evaluating advanced machine learning models aimed at improving cardiovascular health diagnostics.

## 3.2 Proposed Architecture

- Preprocessing and Dataset Preparation

  The initial phase of the proposed architecture involves comprehensive preprocessing of the multi-channel heartbeat audio dataset to ensure data quality and consistency. The raw audio recordings from digital stethoscopes are first segmented into fixed-length frames, which are essential for uniformity in subsequent processing steps. Each audio frame is then subjected to normalization to standardize the amplitude levels across all recordings, followed by noise reduction to minimize any irrelevant background sounds that could interfere with the analysis. This preprocessing step is crucial for preparing a clean and reliable dataset for feature extraction and spectrogram generation.

- Feature Extraction and Spectrogram Generation

  Once the audio data is preprocessed, the next step involves extracting relevant features and generating spectrograms. Feature extraction encompasses both time-domain and frequency-domain analyses. Time-domain features, such as amplitude variations and beat intervals, are extracted directly from the raw audio signals, providing insights into the temporal characteristics of the heartbeat sounds. Simultaneously, frequency-domain features, including spectral entropy and power spectral density, are computed to capture the frequency-related information. In parallel, spectrograms are generated from the audio frames using techniques like Short-Time Fourier Transform (STFT). Spectrograms transform the raw audio data into a visual representation that captures the frequency content and temporal variations of the heartbeat sounds. Each channel's audio is converted into a separate spectrogram, which provides a detailed view of the heart's acoustic profile over time.

- Model Input and Architecture

  The preprocessed and feature-extracted data is then fed into the CNN model. The architecture begins with the input of spectrograms, which are processed through multiple convolutional layers to extract spatial and temporal features. These layers are designed to identify patterns and anomalies in the visual representation of the heartbeat audio. The output of these convolutional layers is flattened and then concatenated with the previously extracted time-domain and frequency-domain features. This fusion of features ensures that the model benefits from both the numerical characteristics and the visual patterns present in the data.

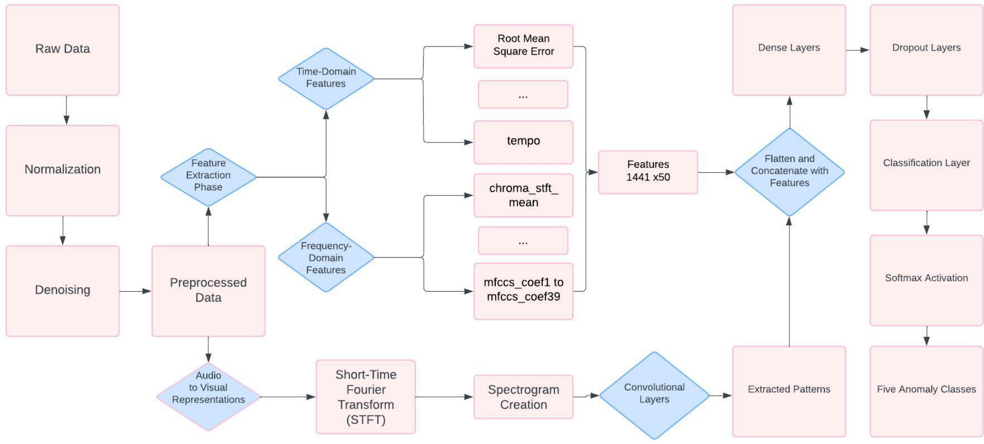



FIGURE 3.1: Proposed Architecture

## 3.3 Tools

To tackle the problem of classifying anomalies in heartbeat audio data, a comprehensive approach using Convolutional Neural Networks (CNNs) has been developed. The implementation leverages a variety of tools and libraries, ensuring the effective preprocessing, feature extraction, and classification of audio data.

The development and execution of this project were conducted on a system with the following specifications: Operating System: Windows, Processor: AMD Ryzen 3, RAM: 8GB, and Storage: 256GB SSD. These specifications provided a robust and efficient environment for handling the computational demands of preprocessing, feature extraction, and training the CNN model for multiple channel heartbeat audio anomaly classification. The following libraries and tools were utilized:

- OpenCV (cv2) for image processing tasks, specifically for handling spectrogram images.

- os for interacting with the operating system to handle file operations.
- NumPy (numpy) for numerical operations, particularly for handling and manipulating arrays.
- Pandas (pandas) for data manipulation and analysis, crucial for managing the dataset.
- Scikit-learn (sklearn.preprocessing) for preprocessing data, including label encoding and splitting the dataset.
- TensorFlow (tensorflow) along with TensorFlow Keras (tensorflow.keras.utils, tensorflow.keras.preprocessing.image) for building, training, and evaluating the neural network model.
- Scikit-learn (sklearn.model_selection) for splitting the data into training and testing sets.

## 3.4 Implementation

- Data Loading and Preprocessing:
  The dataset comprises heartbeat audio files, which serve as the foundation for developing the anomaly classification model. Initially, these audio files are loaded into the system using file handling operations. The raw audio data often contains noise and inconsistencies that can affect model performance. To address this, normalization techniques are applied to standardize the audio levels across all samples. Denoising filters are used to remove background noise, enhancing the clarity of the heartbeat sounds. Additionally, audio segmentation is performed to handle recordings of varying lengths. This step ensures that each audio file is segmented into uniform sections, making it easier to process and analyze.

- Feature Extraction and Spectrogram Generation:

  Once the audio data is preprocessed, the next step involves extracting relevant features. Both time-domain and frequency-domain features are crucial for capturing the nuances of heartbeat sounds. Time-domain features might include statistical properties such as mean, variance, and skewness, which provide insight into the signal's amplitude variations over time. Frequency-domain features are derived using techniques like Fast Fourier Transform (FFT), which reveal the signal's frequency components. In addition to these numerical features, spectrograms are generated using Short-Time Fourier Transform (STFT). Spectrograms offer a visual representation of the audio signals, displaying how the frequency content of the signal evolves over time. This combination of features ensures a comprehensive analysis of the audio data.

- Data Encoding and Splitting:

  For the machine learning model to process the data effectively, the labels associated with the audio files need to be encoded. Label encoding transforms categorical labels into numerical format, which can then be converted into a categorical format suitable for model training. This encoding process ensures that the labels are in a format that the neural network can interpret. Once the data is encoded, it is split into training and testing sets using the train_test_split function. This step is crucial for evaluating the model's performance, as it allows for a clear separation between the data used for training and the data used for validation. Typically, a portion of the dataset (e.g., 80%) is used for training, while the remaining portion (e.g., 20%) is reserved for testing.

- Building the CNN Model:

  The core of the anomaly classification system is a Convolutional Neural Network (CNN) designed to process spectrograms. The model architecture includes multiple convolutional layers, which are adept at extracting spatial features from the spectrogram images. Each convolutional layer is followed by activation functions, such as ReLU, and pooling layers, which help reduce the spatial dimensions and retain important features. After the convolutional layers, the flattened spectrogram features are concatenated with the numerical features extracted from the audio data. This combined feature vector is then passed through dense (fully connected) layers, which further process the data. The final layer of the model is a softmax layer, which classifies the input into one of five classes. This architecture ensures that the model leverages both the visual patterns in the spectrograms and the statistical properties of the audio signals.

- Training and Evaluation:

  The CNN model is trained using the training set, with the goal of minimizing the categorical cross-entropy loss function. During training, the model parameters are adjusted using optimization algorithms such as Adam. The model's performance is monitored through metrics like accuracy, which measures the proportion of correctly classified samples. To validate the model, it is evaluated on the testing set, providing an unbiased assessment of its performance on unseen data. Additionally, methods for interpretability, such as saliency maps, are implemented. Saliency maps highlight the regions in the input data that are most influential in the model's decision-making process, offering insights into how the model distinguishes between different heartbeat anomalies. This interpretability is crucial for building trust in the model's predictions and understanding its behavior.

# Chapter 4

# Results and Analysis

## 4.1 Introduction

The performance of the proposed multiple channel CNN model was thoroughly evaluated and compared with single-channel approaches using both spectrogram and feature inputs. The results highlight the superiority of the multiple channel approach in accurately classifying heartbeat audio anomalies. This evaluation was necessary to establish the effectiveness of incorporating multiple physiological signals and visual audio representations in enhancing the accuracy and reliability of anomaly detection in heartbeat sounds.

## 4.2 Results

- Single-Channel Input Models: For single-channel input, two models were tested: a CNN-based spectrogram classifier and a KNN-based feature classifier. The CNN-based spectrogram classifier achieved an overall accuracy of 83.74% with a precision of 89.16%, an F1-score of 86.28%, and a hamming loss of 0.5328. The accuracy and loss values during testing were 87.88% and 0.4619 respectively, while the training accuracy and loss were 93.14% and 0.2393.

| | Accuracy | Loss |
|---|---|---|
| Test | 0.8788 | 0.4619 |
| Train | 0.93142 | 0.2393 |
| Overall Accuracy | 0.83737 | |
| Precision | 0.89158 | |
| F1-Score | 0.86280 | |
| Hamming Loss | 0.5328 | |

TABLE 4.1: CNN based Spectrogram Classification

On the other hand, the KNN-based feature classifier demonstrated better performance with an accuracy of 93.76%. The F1-scores for different classes were as follows: Artifact (0.98), Extra Heart Sound (0.97), Extrasystole (0.75), Murmur (0.94), and Normal (0.95). The precision and recall scores were also high, with precision values of 1.00, 0.94, 0.77, 0.90, and 0.96 for the respective classes, and recall scores of 0.96, 1.00, 0.74, 0.97, and 0.94.

| Accuracy | 0.9376 | | | | |
|---|---|---|---|---|---|
| Labels | Artifact | Extra HS | Extrasystole | Murmur | Normal |
| F1-Score | 0.98 | 0.97 | 0.75 | 0.94 | 0.95 |
| Precision | 1 | 0.94 | 0.77 | 0.9 | 0.96 |
| Recall Score | 0.96 | 1 | 0.74 | 0.97 | 0.94 |

TABLE 4.2: KNN based Classification

- Performance of Multiple Channel CNN Model: The multiple channel CNN model significantly outperformed the single-channel models. It achieved an overall accuracy of 98.27%, with an F1-score of 98.16% and a precision of 98.32%. The hamming loss was remarkably low at 0.006. During testing, the model attained an accuracy of 98.28% with a loss of 0.027, while the training accuracy and loss were 97.83% and 0.0446, respectively.

 Chapter 4 Results and Analysis

| | Accuracy | Loss |
|---|---|---|
| Test | 0.98275 | 0.027 |
| Train | 0.9783 | 0.0446 |
| Overall Accuracy | 0.9827 | |
| F1-Score | 0.9816 | |
| Precision | 0.9832 | |
| Hamming Loss | 0.006 | |

TABLE 4.3: Multiple input model classification

### 4.2.1 Interpretation and Implications

The multiple channel model's superior performance can be attributed to its ability to leverage both the time-domain and frequency-domain features, as well as the visual representation of the audio signals through spectrograms. This comprehensive feature extraction and processing approach enables the model to better capture the nuances of different heart sound anomalies.

The high precision and F1-scores indicate the model's reliability in accurately identifying different types of heartbeat anomalies, which is crucial for medical diagnostics. The low hamming loss further underscores the model's effectiveness in minimizing misclassifications.

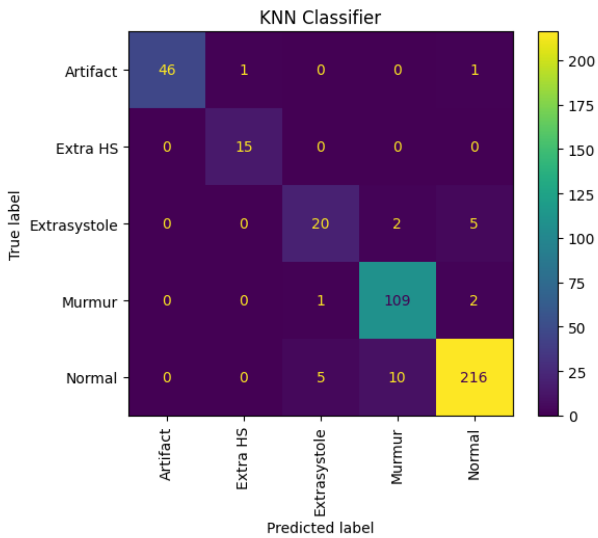



FIGURE 4.1: KNN Based classification

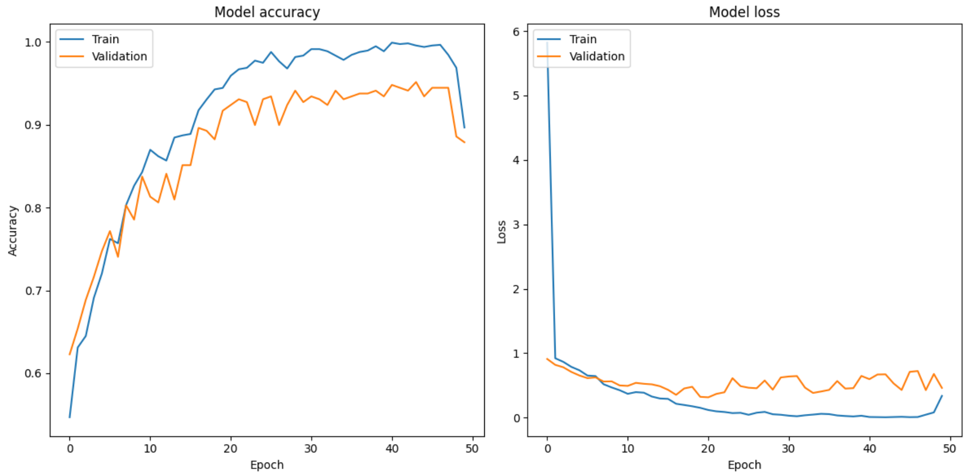



FIGURE 4.2: CNN based Spectrogram Classification

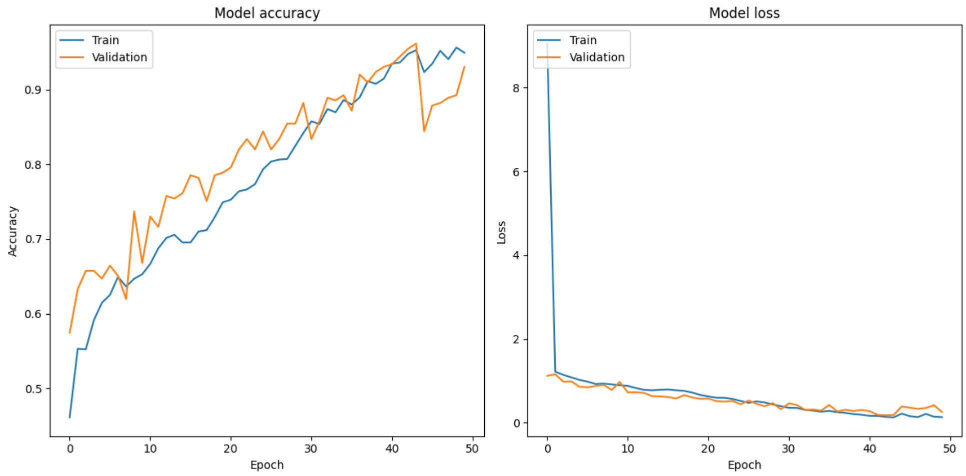



FIGURE 4.3: Multiple input model classification

The multiple channel CNN model for heartbeat audio anomaly classification using features and spectrogram channels demonstrates significant improvements over single-channel models. Its high accuracy, precision, and F1-scores make it a promising tool for early detection and diagnosis of cardiovascular conditions, potentially aiding healthcare professionals in making more informed decisions and improving patient outcomes.

# Chapter 5

# Conclusion

The project on Multiple Channel Heartbeat Audio Anomaly Classification using Features and Spectrogram Channels with CNN has successfully demonstrated the efficacy of integrating diverse data inputs and advanced neural network techniques for the classification of heartbeat anomalies. The methodology employed in this project—leveraging both time-domain and frequency-domain features alongside spectrograms—has shown a marked improvement in classification performance compared to traditional single-channel approaches.

The proposed model achieved an overall accuracy of 98.27%, with high precision and F1-scores across all classes. This indicates the model's robust capability in accurately distinguishing between different types of heartbeat anomalies, including Normal, Murmur, Extra Heart Sound, Artifact, and Extrasystole. The low hamming loss further highlights the model's effectiveness in minimizing classification errors, providing a reliable tool for early detection of cardiovascular conditions.

In contrast, single-channel models, whether CNN-based or KNN-based, demonstrated lower performance metrics, underscoring the benefits of a multi-channel approach. The CNN model's ability to process and integrate multiple physiological signals and visual audio representations enables it to capture subtle nuances in heartbeat sounds that single-channel models might miss.

The results of this project not only validate the proposed methodology but also underscore the potential for advanced machine learning techniques to enhance medical diagnostics. By providing a more accurate and automated system for heartbeat anomaly classification, this work contributes significantly to the field of cardiovascular health. Future work could build upon these findings by further refining the model, expanding the dataset, and exploring additional feature extraction techniques to further enhance performance.

Overall, the project marks a significant step forward in the application of deep learning for healthcare, offering promising tools for early diagnosis and improved patient care in the realm of cardiovascular diseases.

# Chapter 6

# Future Work

Building on the success of the multiple channel CNN model for heartbeat audio anomaly classification, future work will focus on enhancing the interpretability and understanding of the model's predictions through the use of Class Activation Maps (CAMs) and Grad-CAM techniques.

**Incorporating Grad-CAM for Improved Interpretability**:

Grad-CAM, as introduced by [14], provides a mechanism to visualize the regions in the input data that significantly influence the model's predictions. By applying Grad-CAM to our CNN model, we can generate visual explanations that highlight the important areas of the spectrograms corresponding to different classes of heartbeat anomalies. This process involves backpropagating the gradients through the network to calculate the weights for each region in the spectrogram, thereby creating Grad-CAM visualizations.

**Post-Processing and Interpretation**:

The future implementation will include generating separate Grad-CAM visualizations for each class of heartbeat anomaly. This will involve:

- **Spectrogram Visualization**: Overlaying the Grad-CAM heatmaps on the spectrograms to visually indicate which regions of the audio signal are most relevant to the model's classification decisions. This will provide a clear understanding of the specific parts of the heartbeat signal that the model considers important for each class.

- **Feature Importance Analysis**: Identifying and listing the important audio features that contribute to the classification, as highlighted by Grad-CAM. This will include analyzing the features extracted from the audio data and how they interact with the spectrogram visualizations.

**Classification and Explanation**:

Using these Grad-CAM visualizations, the model's predictions can be further refined and explained. For instance, by combining the visual explanations from the spectrogram with the identified audio features, we can more accurately classify heartbeat audio as normal or anomalous. This dual-layer explanation will enhance the trustworthiness and transparency of the model, making it easier for healthcare professionals to understand and validate the predictions.[6]

**Expanding the Scope**:

Further research could involve exploring additional techniques for visual explanation and incorporating other interpretability methods. Additionally, extending the dataset and refining the model architecture could improve overall performance and generalization across diverse clinical settings.

In conclusion, the integration of Grad-CAM with the existing CNN model represents a promising direction for future work, aiming to improve the model's interpretability and clinical applicability, thereby contributing to more effective and transparent cardiovascular diagnostics.

# Chapter 7

# Code Snippet

**Feature Extraction** :

```
def extract_features1(audio_path,offset):
y, sr = librosa.load(audio_path, sr=None, offset=offset)

# Extracting various features
chroma_stft = librosa.feature.chroma_stft(y=y, sr=sr)
rmse = librosa.feature.rms(y=y)
spec_cent = librosa.feature.spectral_centroid(y=y, sr=sr)
spec_bw = librosa.feature.spectral_bandwidth(y=y, sr=sr)
rolloff = librosa.feature.spectral_rolloff(y=y, sr=sr)
zcr = librosa.feature.zero_crossing_rate(y)
mfccs = librosa.feature.mfcc(y=y, sr=sr, n_mfcc=40)

# Calculating the mean of each feature
chroma_stft_mean = np.mean(chroma_stft)
rmse_mean = np.mean(rmse)
spec_cent_mean = np.mean(spec_cent)
```

```
    spec_bw_mean = np.mean(spec_bw)
    rolloff_mean = np.mean(rolloff)
    zcr_mean = np.mean(zcr)
    mfccs_mean = np.mean(mfccs, axis=1)

    # Concatenating all features into a single array
    features = np.array([audio_path, offset, chroma_stft_mean, rmse_mean,
    spec_cent_mean, spec_bw_mean, rolloff_mean, zcr_mean])
    # print(features.shape)
    features = np.append(features, mfccs_mean)
    # print(features.shape)
    # Extracting tempo
    tempo, _ = librosa.beat.beat_track(y=y, sr=sr)
    features = np.append(features, tempo)

    return features

print(extract_features1(dataset.filename[860],dataset.offset[860]))
```

**Spectrogram Generation** :

```
    def generate_and_save_spectrogram(audio_path, output_folder,
    offset=0.0, target_sr=22050):
    # Load the audio file with a lower sampling rate to reduce memory usage
    y, sr = librosa.load(audio_path, sr=target_sr, offset=offset)

    # Generate the spectrogram
    S = librosa.feature.melspectrogram(y=y, sr=sr)
    S_db = librosa.power_to_db(S, ref=np.max)
```

```
# Plot the spectrogram
plt.figure(figsize=(10, 4))
librosa.display.specshow(S_db, sr=sr, x_axis='time', y_axis='mel')
plt.colorbar(format='%+2.0f dB')
plt.title('Mel-frequency spectrogram')
plt.tight_layout()

# Create the output folder if it doesn't exist
if not os.path.exists(output_folder):
    os.makedirs(output_folder)

# Save the plot as an image file
file_name = os.path.splitext(os.path.basename(audio_path))[0] + '.png'
output_path = os.path.join(output_folder, file_name)
plt.savefig(output_path)
plt.close()
```

**CNN Model** :

```
def multi_channel_cnn(input_shape_spectrogram, input_shape_features
, num_classes):
# Spectrogram channel
input_spectrogram = Input(shape=input_shape_spectrogram)
x_spectrogram = Conv2D(32, kernel_size=(3, 3), activation='relu')
(input_spectrogram)
x_spectrogram = MaxPooling2D(pool_size=(2, 2))(x_spectrogram)
x_spectrogram = Conv2D(64, kernel_size=(3, 3), activation='relu')
(x_spectrogram)
```

```
x_spectrogram = MaxPooling2D(pool_size=(2, 2))(x_spectrogram)
x_spectrogram = Conv2D(128, kernel_size=(3, 3), activation='relu',
name='last_conv_layer')(x_spectrogram)
x_spectrogram = MaxPooling2D(pool_size=(2, 2))(x_spectrogram)
x_spectrogram = Flatten()(x_spectrogram)

# Features channel
input_features = Input(shape=input_shape_features)
x_features = Dense(128, activation='relu')(input_features)

# Combine channels
combined = concatenate([x_spectrogram, x_features])
combined = Dense(128, activation='relu')(combined)
combined = Dropout(0.5)(combined)
output = Dense(num_classes, activation='softmax')(combined)

model = Model(inputs=[input_spectrogram, input_features], outputs=output)
model.compile(optimizer='adam', loss='categorical_crossentropy',
metrics=['accuracy'])

return model

X = list(zip(X_spectrograms, X_features.values))
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=123, stratify=y)
X_left = np.array([t[0] for t in X_train], dtype=np.float32)
X_right = np.array([t[1] for t in X_train], dtype=np.float32)
X_test_left = np.array([t[0] for t in X_test], dtype=np.float32)
```

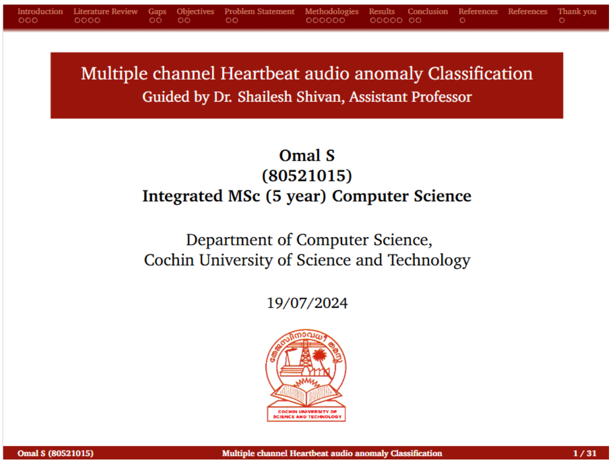
Chapter 7 Code Snippet

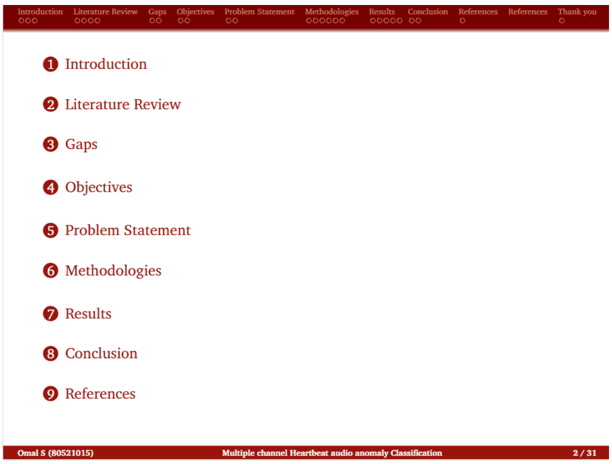
```
X_test_right = np.array([t[1] for t in X_test], dtype=np.float32)


# Train the model
history =model.fit([X_spectrograms, X_features], y, validation_split=0.2,
epochs=50,
batch_size=16)
```

# Chapter 8

# Presentation Slides



Introduction Literature Review Gaps Objectives Problem Statement Methodologies Results Conclusion References References Thank you

## Introduction

Heart Disease: Deaton et al. (2011)

- A Global Threat
- Millions of people are affected each year
- Silent threat

- Progress silently without noticeable symptoms
- Early detection

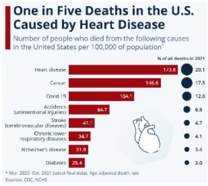



Figure 1: World Economic forum (Florian Zandt, 2024)

Omal S (80521015) Multiple channel Heartbeat audio anomaly Classification 4 / 31

Introduction Literature Review Gaps Objectives Problem Statement Methodologies Results Conclusion References References Thank you

## Heart Health Monitoring

**Electrocardiogram (ECG):**

- Measures the electrical activity of the heart muscle
- Uses electrodes
- Provides valuable information about heart rhythm etc

**Echocardiogram (Echo):**

- Uses ultrasound waves
- Use a transducer (probe)
- Images show the size, shape, and movement of heart chambers and valves

**Limitations:**

- Trained medical professionals
- Expensive
- Time Consuming

Omal S (80521015) Multiple channel Heartbeat audio anomaly Classification 5 / 31

Introduction Literature Review Gaps Objectives Problem Statement Methodologies Results Conclusion References References Thank you

## Robust Heartbeat Detection From Multimodal Data via CNN-Based Generalizable Information Fusion Chandra et al. (2018)

- **Objective**: Improve the accuracy and robustness of heartbeat detection, particularly in critical care scenarios.
- **Novelty**: Convolutional Neural Network (CNN) based information fusion (CIF) approach for heartbeat detection.
- **Methodology**:
  - Data: PhysioNet 2014 Challenge database (ECG and BP) and the MIT-BIH arrhythmia database (2-channel ECG).
  - Preprocessing: voids the need for pre-defined features or separate processing of each signal.
- **Model**: CNN architecture that learns fused features from multiple physiological signals (e.g., ECG, blood pressure).

Omal S (80521015) Multiple channel Heartbeat audio anomaly Classification 7 / 31

**Multi-classification neural network model for detection of abnormal heartbeat audio signals Malik et al. (2022)**

- **Objective**: Develop an automated system for classifying abnormal heartbeat audio signals to aid cardiologists in diagnosing heart disease
- **Novelty**: Claims 1st single neural network model for classifying
- **Methodology**:
  - Data: PASCAL challenge and PhysioNet 2017 challenge
  - Preprocessing: Capture dominant features, Bandpass filtering applied, Downsampling
- **Model**:Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) units.





**Explainable Deep Learning-Based Approach for Multilabel Classification of Electrogram Ganeshkumar et al. (2021)**

- **Objective**:Develop a deep learning method for multilabel ECG classification to identify multiple heart diseases in a single signal. Establish an explainable AI framework.
- **Novelty**: Identifies up to two labels per ECG signal and Uses Grad-CAM for model explainability.
- **Methodology**:
  - Data: China Physiological Signal Challenge 2018:
  - Preprocessing: Construct ECG matrices. Train CNN with single labels. Apply thresholding on softmax output for multilabel classification.
- **Model**: CNN trained with ECG matrices and Grad-CAM technique for interpretability.





**Limitations due to lack of interpretability and Clinical Understanding**

- **Focuses on Specific Signals**: focuses on ECG and BP signals
- **Lacks Interpretability**: Not explaining why it detects an anomaly.
- **Focuses on Classification, not Explanation**: lacks explanation for anomaly classifications, limiting clinical usefulness.
- **Potential for bias**: The model might miss anomalies present in other signals.
- **Limited insights for clinicians**: Healthcare professionals wouldn't understand the rationale behind the anomaly detection

 Chapter 8 Presentation Slides



## Project Objective: Heartbeat Anomaly Classification using Audio Analysis

Aims to develop a novel system for classifying anomalies in heartbeat audio data using explainable deep learning techniques.

- Leveraging **multiple physiological signals**
- **Improved Trust and Transparency**: The system offers interpretable results.
- **Potential for broader applicability**





## Problem Statement

Classifing heartbeat anomaly system using deep learning for heartbeat audio data.

- **Proposed solution:**
  - Multiple input Channel
  - Leverage Features and Spectrogram

Aims not only detects anomalies but also provides insights into the reasoning behind its classifications, fostering

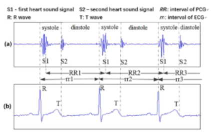



Figure 2: Correlation between heart sound and ECG(Xiefeng et al., 2019)





## Multiple channel Heartbeat audio anomaly Classification

A dual-channel heartbeat anomaly classification using PyTorch and Grad-CAM feature extraction from the audio data.

- **Pre-processing**: Loading the audio files applying signal processing techniques to improve data quality.Colt et al. (2021)
- **Feature Extraction Channels:**
  - **Channel 1: Spectrogram Generation**: Generate a spectrogram image representing the time-frequency domain of the audio.
  - **Channel 2: Audio Feature Extraction (PyTorch)**: Mel-frequency cepstral coefficients (MFCCs), Zero-Crossing Rate (ZCR), Spectral Centroid, Spectral Rolloff Point, Harmonic Ratio Features, heart rate variability (HRV) etc.
- **Feature Engineering** : Techniques like PCA can be applied on audio features (Channel 2) for dimensionality reduction
- **Deep Learning Model**: Utilize a multi-channel CNN architecture Wyse (2017)
- **Class Activation Maps (CAMs) with Grad-CAM**: Selvaraju et al. (2017)

- **Backpropagation** calculates weights
- Generate separate Grad-CAM visualizations
- **Post-processing and Interpretation:** Based on the model's prediction:
  - Classify the heartbeat audio as normal or anomalous.
  - Generate a visual explanation combining:
    - Spectrogram with highlighted regions using Grad-CAM (Channel 1).
    - List of important audio features identified by Grad-CAM (Channel 2).





## Proposed Architechture Diagram

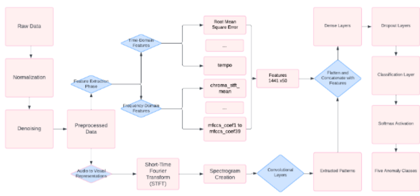



Figure 3: Proposed Architecture diagram





## Theory

- **Signal Processing Techniques for Heartbeat Audio Data:**
  - **Filtering:** Band-pass filters, Noise cancellation techniques
  - **Normalization:** Normalization to a peak value, Z-score normalization
- **Grad-CAM for Explainable AI:** spectrogram and audio features) are most influential in the model's prediction
  - **Leveraging Gradients:** Grad-CAM utilizes the concept of gradients
  - **Class Activation Maps (CAMs):** Heatmap where brighter regions stronger influence on the model's prediction.
  - **Visual Interpretation:** Highlighted features likely deviated from the expected patterns in normal heartbeats.

## Dataset Overview - PASCAL Classifying Heart Sounds Challenge (CHSC2011) Bentley et al. (2011)

- **Datasets:**
  - Dataset A: Recordings from the general public
  - Dataset B: Recordings from a clinical trial
- **File Formats:** WAV and AIF
- **Class Labels** (Dataset A & B):
  - Normal 769
  - Murmur 373
  - Extrasystole 88
  - Extra HS 51
  - Artifact 160





## Spectrogram generation

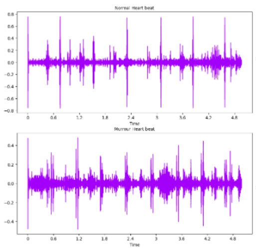



Figure 4: Audio waveform of Normal and Murmur data

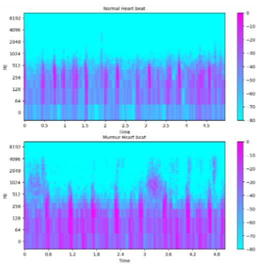



Figure 5: Spectrogram of the corresponding Normal and Murmur data





## Audio Feature Extraction

| | zero_sum | tempo | spectral_centroid_mean | harmony_mean | perceptual_mean | mfccs_coef1 | mfccs_coef2 | mfccs_coef3 | mfccs_coef4 |
|---|---|---|---|---|---|---|---|---|---|
| **1473** | -0.344520 | -0.730863 | -0.397972 | 0.396508 | -0.318928 | -0.035713 | 0.995138 | 0.905671 | -0.406058 |
| **1474** | -0.398484 | 1.207852 | -0.456484 | 0.005145 | -1.745808 | 0.225074 | 0.140176 | 0.795213 | 0.331832 |
| **1475** | -0.151901 | 1.207852 | -0.280595 | 0.845657 | 1.717223 | 0.070811 | 1.449280 | 0.455752 | -1.388128 |
| **1476** | -0.380496 | -0.730863 | -0.445675 | -0.570447 | 0.541488 | -0.021828 | 0.434756 | 0.956861 | 0.152740 |
| **1477** | -0.329231 | 0.487758 | -0.368349 | 1.520990 | -2.022401 | 0.680641 | 1.018012 | 0.815994 | -0.692306 |

Figure 6: Details of feature extracted values from audio files

## Evaluation Measures for Multilabel Classification

### Single Channel Performance Comparison

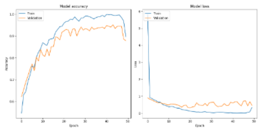


Figure 7: Accuracy and Loss Graph of CNN trained using single feature channel

| CNN based Spectrogram Classification | | |
|---|---|---|
| | Accuracy | Loss |
| Test | 0.8788 | 0.4619 |
| Train | 0.93142 | 0.2393 |
| Overall Accuracy | 0.83737 | |
| Precision | 0.89158 | |
| F1-Score | 0.86280 | |
| Hamming Loss | 0.5328 | |

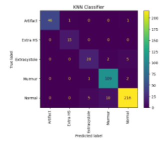


Figure 8: Confusion matrix of feature classification using KNN classifier

| KNN based Classification | | | | | |
|---|---|---|---|---|---|
| Accuracy | 0.9376 | | | | |
| Labels | Artifact | Extra HS | Extrasystole | Murmur | Normal |
| F1-Score | 0.98 | 0.97 | 0.75 | 0.94 | 0.95 |
| Precision | 1 | 0.94 | 0.77 | 0.9 | 0.96 |
| Recall Score | 0.96 | 1 | 0.74 | 0.97 | 0.94 |





## Multi-Input Channel: Model Performance

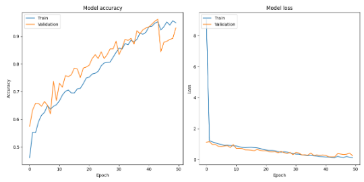



Figure 9: Accuracy and Loss Graph of CNN model with Multiple channel input (extracted features and spectrogram of audio file

| Multiple input model classification | | |
|---|---|---|
| | Accuracy | Loss |
| Test | 0.98275 | 0.027 |
| Train | 0.9783 | 0.0446 |
| Overall Accuracy | 0.9827 | |
| F1-Score | 0.9816 | |
| Precision | 0.9832 | |
| Hamming Loss | 0.006 | |





## Conclusion :

- A novel multi-channel CNN architecture for explainable heartbeat anomaly classification using PyTorch.
- The architecture leverages complementary information from:
  - Channel 1: Spectrogram (time-frequency domain)
  - Channel 2: Extracted audio features (e.g., MFCCs, HRV)
- Grad-CAM integration provides visual explanations for both channels, highlighting:
  - Discriminative regions in the spectrogram
  - Specific audio features influencing the classification

# Bibliography


[1] Nurseno Bayu Aji et al. "CNN-LSTM for Heartbeat Sound Classification". In: *JOIV : International Journal on Informatics Visualization* (2024). DOI: 10.62527/joiv.8.2.2115.

[2] Debbal Sidi Mohammed El Amine. *Heartbeat Cardiac Sounds Analysis*. 2023. DOI: 10.31579/2835-8325/016.

[3] Abhirup Das Barman and Sujoy Mondal. *Experiment on spectrogram feature based CNN for sound anomaly detection*. 2023. DOI: 10.1109/codec60112.2023.10466070.

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

[5] *Exploiting Single-Channel Speech for Multi-Channel End-to-End Speech Recognition: A Comparative Study*. 2022. DOI: 10.1109/iscslp57327.2022.10038153.

[6] M Ganeshkumar et al. "Explainable deep learning-based approach for multilabel classification of electrocardiogram". In: *IEEE Transactions on Engineering Management* 70.8 (2021), pp. 2787–2799.

[7] Ali Fatih Gunduz. "Review of Heart Sound Analyses from Phonocardiogram Records". In: *Romanian Biotechnological Letters* (2022). DOI: 10.25083/rbl/27.1/3167-3183.

[8] Soyul Han et al. "MCHeart: Multi-Channel-Based Heart Signal Processing Scheme for Heart Noise Detection Using Deep Learning". In: *Biology* (2023). DOI: 10.3390/biology12101291.

[9] *Heartbeat Prediction using Mel Spectrogram and MFCC Value*. 2023. DOI: 10.1109/globconet56651.2023.10150129.

[10] Sayed Shahid Hussain et al. *Deep Learning Based Phonocardiogram Signals Analysis for Cardiovascular Abnormalities Detection*. 2023. DOI: 10.1109/ICRAI57502.2023.10089537.

[11] Yash Karbhari et al. *Age, Gender and Emotion Recognition by Speech Spectrograms Using Feature Learning*. 2023. DOI: 10.1109/icpcsn58827.2023.00082.

[12] Onur KILIÇ. "High-Level CNN and Machine Learning Methods for Speaker Recognition". In: *Sensors* (2023). DOI: 10.3390/s23073461.

[13] Ann Nita Netto, Lizy Abraham, and Saji Philip. "HBNET: A blended ensemble model for the detection of cardiovascular anomalies using phonocardiogram." In: *Technology and Health Care* (2024). DOI: 10.3233/thc-231290.

[14] Ramprasaath R Selvaraju et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626.

[15] Rakshit Shah et al. *Heartbeat Prediction using Mel Spectrogram and MFCC Value*. 2023. DOI: 10.1109/GlobConET56651.2023.10150129.

[16] Sai Vemulapalli Shanmukha and K. Deepak. *Sound Event Detection in Constrained Real-world Environments using Fine Tuned CNNs*. 2024. DOI: 10.1109/icdcot61034.2024.10516127.

[17] Tran Thi Thanh. *Analysis of Building the Music Feature Extraction Systems: A Review*. 2024. DOI: 10.47191/etj/v9i05.22.

[18] Qinghao Zhao et al. *Deep Learning for Heart Sound Analysis: A Literature Review*. 2023. DOI: 10.1101/2023.09.16.23295653.

[19] . *Audio Classification of Low Feature Spectrograms Utilizing Convolutional Neural Networks*. 2022. DOI: 10.1109/icmla55696.2022.00115.