COMPUTER VISION LAB

PROJECT REPORT

On

Two Class classification of Paediatric Heart Sound signals using the Short-Term Fourier Transform Features

Submitted By:
Jayesh Nayak (121CS0195)
Ashish Padhy (121CS0196)
Armaan Paikaray (121CS0197)
Animesh Panda (121CS0198)

Submitted To:
Dr. Puneet Kumar Jain

Department of Computer Science and Engineering
NATIONAL INSTITUTE OF TECHNOLOGY ROURKELA

4 April, 2024

TABLE OF CONTENTS

ABSTRACT

Paediatric heart sound classification is pivotal in the timely diagnosis and management of congenital heart diseases (CHD). This research presents a Random Forest Classifier along with a convolutional neural network (CNN) model for the two-class classification of paediatric heart sound signals into normal and abnormal categories.Utilizing a dataset encompassing heart sound recordings from children aged 1 day to 14 years, the proposed model aims to automate diagnostic processes, reducing the inherent subjectivity of traditional auscultation methods. Employing feature extraction techniques from the short-term Fourier transform (STFT), Mel Spectrogram, MFCC, and Chromagram plots facilitates a comprehensive representation of paediatric heart sound acoustic patterns. Our approach holds promise for enhancing paediatric cardiac diagnostics, providing healthcare practitioners with a reliable tool for early CHD detection and improving patient outcomes.

1. Introduction

1.1 Introduction

Heart sounds are vital diagnostic tools used by physicians to assess heart health. By analysing the characteristic sounds produced during cardiac contraction and relaxation, abnormalities indicative of potential heart disease or other underlying conditions can be identified. This project focuses on the development of an automated classification system for paediatric heart sounds.

1.2 Motivation

Traditional auscultation, while valuable, presents limitations. Relying on the expertise of individual physicians can lead to inconsistencies in interpretation due to subjectivity and experience level. Additionally, the time-consuming nature of traditional analysis can impact diagnostic efficiency. Furthermore, access to skilled paediatric cardiologists may be limited in certain healthcare settings.

Machine learning offers a compelling solution to address these limitations. By leveraging automated classification systems, we can potentially achieve:

- Enhanced Objectivity: Machine learning models can be trained on vast datasets, reducing the influence of human bias and leading to more consistent diagnoses.
- Improved Efficiency: Automated systems can analyse heart sounds rapidly, potentially streamlining diagnostic workflows and reducing turnaround times.
- Increased Accessibility: Machine learning-based tools could be deployed in resource-constrained settings, expanding access to accurate paediatric heart sound analysis and potentially facilitating earlier detection of cardiac issues.

1.3 Problem Statement

This project aims to develop a machine-learning model capable of automatically classifying paediatric heart sound recordings into two categories: normal and abnormal. The model will utilise a dataset of heart sound recordings collected from children using a smart stethoscope. The model will utilise a 1D/2D Convolutional Neural Network (CNN) architecture trained on a dedicated paediatric heart sound dataset. By extracting informative features using Short-Term Fourier Transform (STFT) and employing a 1D/2D Convolutional Neural Network (CNN) architecture, the model will learn to differentiate between normal and abnormal heart sound patterns, assisting medical professionals in the early detection of potential paediatric heart conditions. This project seeks to address the limitations of traditional auscultation methods by creating a robust and efficient tool to support medical professionals in the early detection of paediatric cardiac conditions.

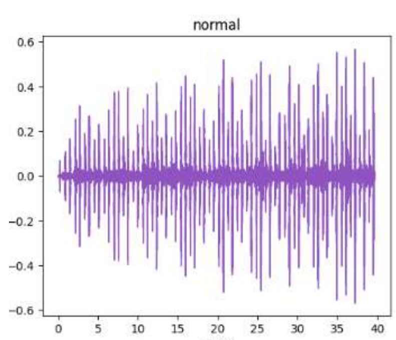
2. Literature Review

Previous research has focused on classifying normal heart sounds and murmurs using the grey-level difference matrix (GLDM) feature extracted from the short-time Fourier transform (STFT) plot. These studies typically employ support vector machine (SVM) classifiers for classification. For instance, a study achieved an accuracy of 83% by extracting GLDM characteristics from the STFT plot and utilising an SVM classifier. In contrast, our research endeavours to advance upon existing methodologies by employing a comprehensive feature extraction approach. We extract features not only from the STFT plot but also from Mel Spectrogram, MFCC, and Chromagram plots. This expanded feature set aims to capture a broader range of acoustic patterns inherent in paediatric heart sounds, thereby enhancing classification performance. Our model achieves a notable accuracy of 86.77%, demonstrating the efficacy of our approach in paediatric heart sound classification.

3. Proposed Methodology

Data Loading and Exploration

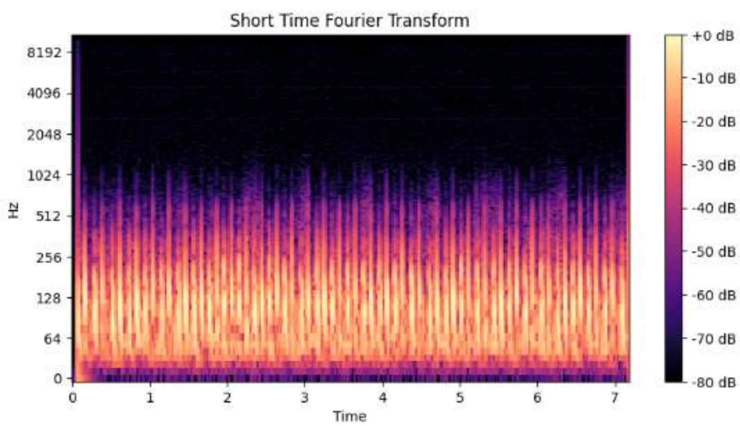
The first step involves loading the heart sound data. We have utilised the ‘soundfile’ library to read the audio files directly from the data directory. Labels associated with each recording are extracted based on the directory structure. Next, we will iterate through all the files in the data directory. To gain a basic understanding of the data, we visualise a few of these audio files. This was helpful in identifying any initial patterns or anomalies.



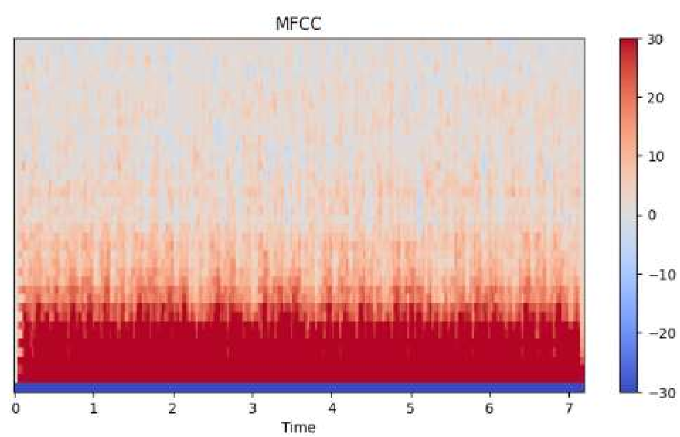
Feature Extraction

Feature extraction plays a crucial role in training machine learning models for accurate classification. We have employed three primary techniques to extract informative features from the heart sound recordings:

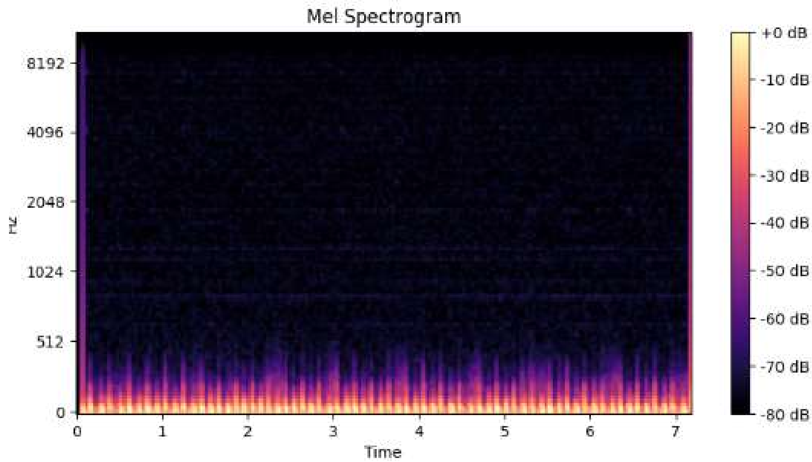
Short-Time Fourier Transform (STFT): This technique decomposes the heart sound signal into its constituent frequencies over time, providing a time-frequency representation. This allows us to capture how the frequency content of the signal changes over time.



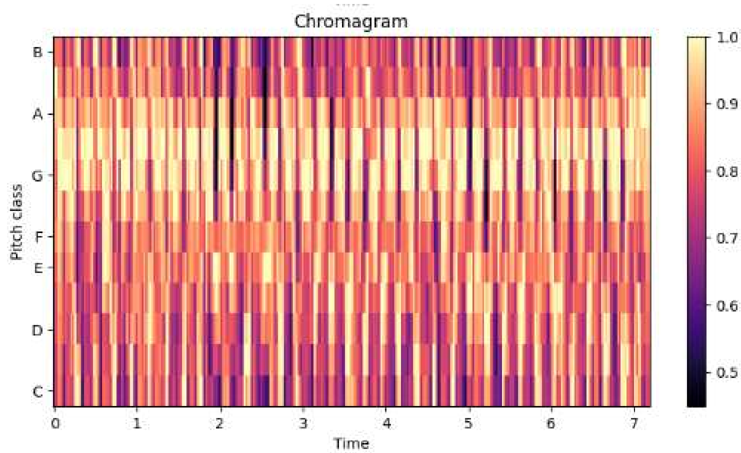
Mel-Frequency Cepstral Coefficients (MFCC): Inspired by human auditory perception, MFCCs focus on frequencies most relevant to human hearing. This approach aims to mimic how humans perceive and distinguish between different heart sounds.



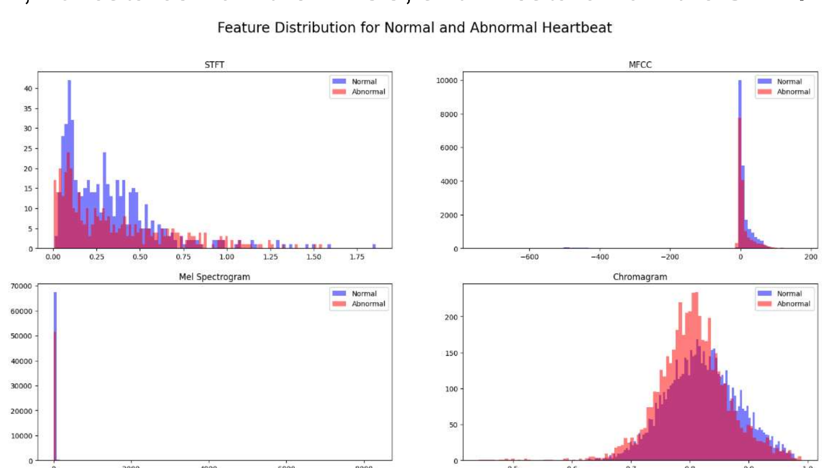
Mel Spectrogram: This technique visually depicts the distribution of sound energy across different frequencies over time. It provides a helpful visualisation tool to understand the frequency content of the heart sounds.



Chromagram: This feature captures the tonal characteristics of the heart sound signal, offering insights into the harmonic content that may be informative for classification.



In total, we extract 128 features from the Mel spectrogram, 12 features from the Chromagram, 40 features from the MFCC, and 1 feature from the STFT.



Data Preprocessing
Prior to feeding the extracted features into our machine learning models, we perform some preprocessing steps. This ensures all features are on a similar scale and improves model performance.

First, we calculated statistical parameters (mean, standard deviation, minimum, and maximum) for each extracted feature set (STFT, MFCC, Mel Spectrogram, Chromagram).

Following this, we used two different scaling techniques: **StandardScaler** and **MinMaxScaler**. These techniques will transform all features to a common range, typically between 0 and 1 or with a mean of 0 and a standard deviation of 1. This standardisation step helps prevent certain features from dominating the model's learning process due to differences in their initial scales.

3.1 Using Classifiers

This approach involves comparing the performance of different machine learning classifiers on the preprocessed data. We trained and evaluated various classifiers and reported the accuracy achieved by each on the heart sound classification task. Additionally, we utilised **Grid Search** to optimise the hyperparameters of these classifiers for improved performance.

We used the following classifiers:

1. **RandomForestClassifier**
2. **SVC**
3. **AdaBoostClassifier**
4. **DecisionTreeClassifier**
5. **KNeighborsClassifier**
6. **SVC RBF kernel**
7. **QuadraticDiscriminantAnalysis**
8. **GaussianNB**

3.2 Using a Convolutional Neural Network

The second method leverages a 1D Convolutional Neural Network (CNN) architecture specifically designed for processing sequential data like audio signals.

The CNN has the following architecture:

Conv1D Layer: Extracts local features using a variable number of filters (32 to 256) with a kernel size of 2, followed by a ReLU activation.

MaxPooling1D Layer: Downsamples the data while preserving important features.
Flatten Layer: Reshapes the output into a 1D vector for dense layers.
Dense Layer 1: Introduces non-linearity with ReLU activation and learns higher-level features with a variable number of units (32 to 256).
Dropout Layer: Prevents overfitting by randomly dropping out 50% of neurons during training.
Output Layer: Predicts the probability of a heart sound belonging to the abnormal class (sigmoid activation).
Hyperparameter tuning optimises the number of filters and dense layer units for best performance.

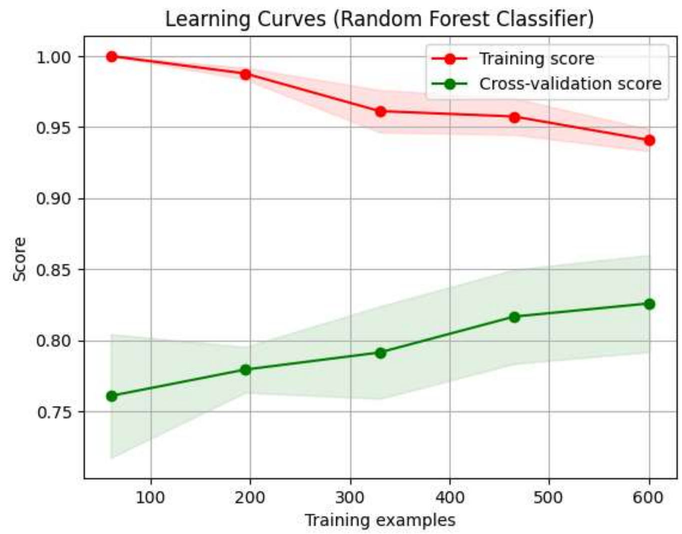
4. RESULTS AND DISCUSSION

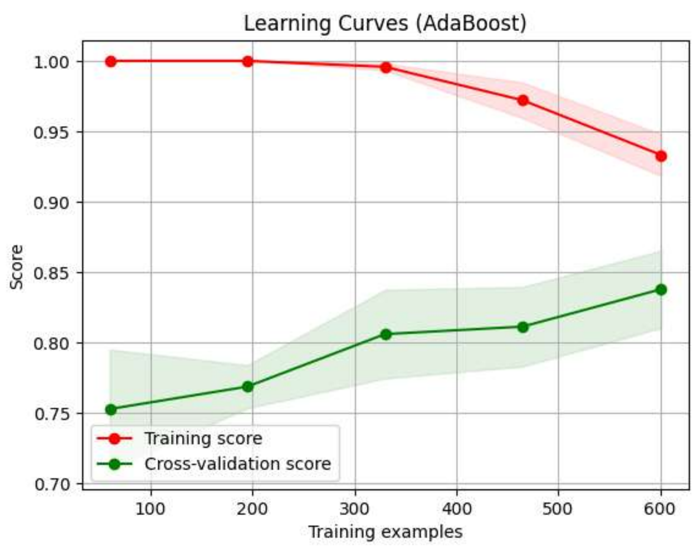
4.1. Experimental setup

This experiment investigated the efficacy of machine learning techniques for classifying paediatric heart sound recordings into normal and abnormal categories. Two primary approaches were explored:

Classifiers: We compared the performance of various machine learning classifiers on the preprocessed data. These classifiers were trained and evaluated to determine their accuracy in classifying heart sounds. Grid Search was employed to optimise the hyperparameters of each classifier, aiming to achieve the best possible performance.

Classifier	Accuracy Score
Random Forest Classifier	87.83%
SVC	84.13%
AdaBoost Classifier	81.48%
Decision Tree Classifier	76.72%
K Neighbors Classifier	75.13%
SVC RBF Kernel	65.61%
Quadratic Discriminant Analysis	48.15%
GaussianNB	40.74%





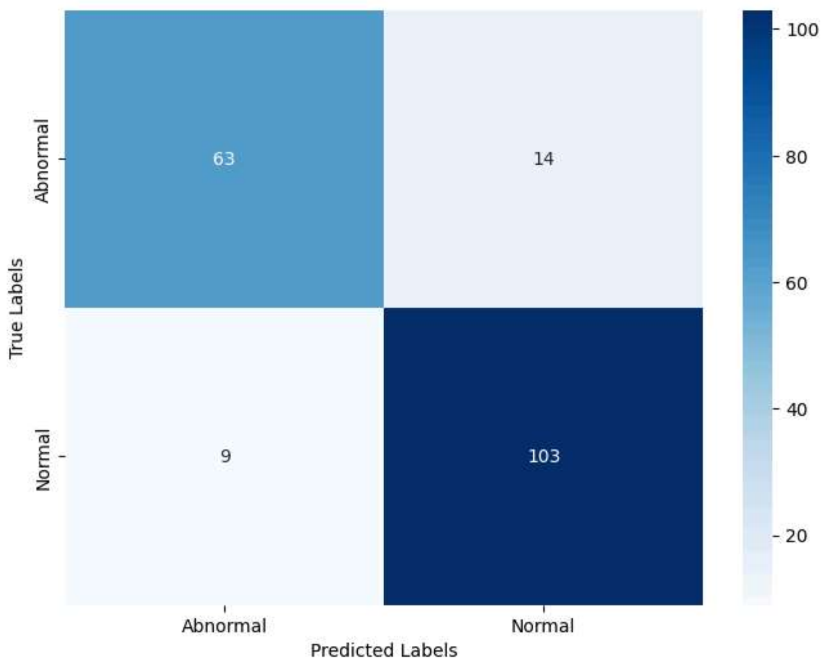
Convolutional Neural Network (CNN): A 1D CNN architecture specifically designed for processing sequential data like audio signals was implemented. The CNN was trained to learn features directly from the heart sound recordings and classify them as normal or abnormal. Hyperparameter tuning using Hyperband was utilised to identify the optimal configuration for the CNN model.

4.2. Dataset Description

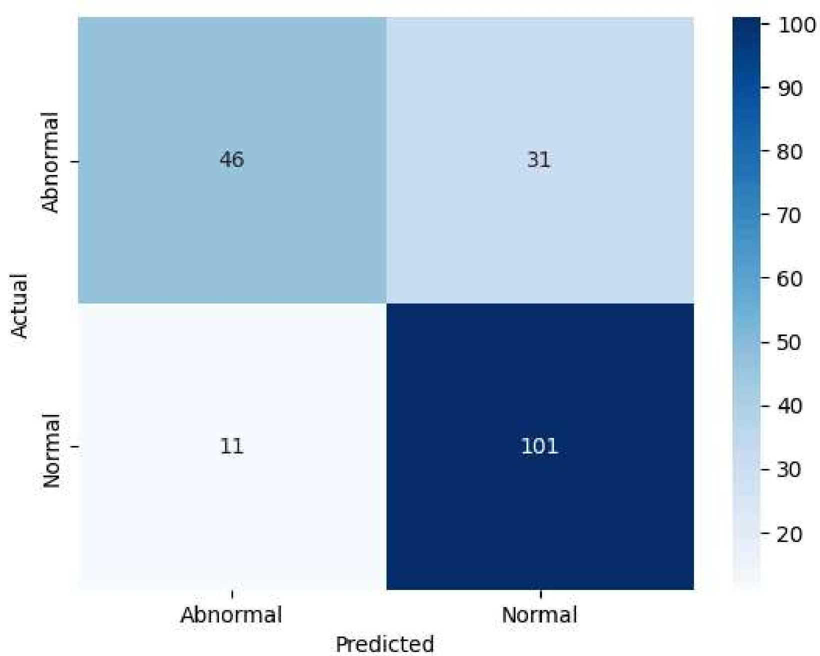
The dataset employed for this study comprised paediatric heart sound recordings collected using a smart stethoscope. The data was meticulously divided into two distinct classes: normal and abnormal heart sounds. Each recording was rigorously labelled based on diagnoses confirmed by paediatric cardiologists. This stringent labelling process ensured the accuracy of the classification task and the reliability of the data for training and evaluating the machine learning models.

4.3. Performance Metrics

Confusion Matrix for Random Forest Classifier



Confusion Matrix for CNN Model



4.4. Results and Discussion

Our experiments evaluated two machine learning approaches for classifying paediatric heart sounds. Classifiers achieved promising results, with the **Random Forest Classifier reaching the highest accuracy of 87.83%**. The Support Vector Classifier (SVC) followed closely with an accuracy of 84.13%.

The CNN model achieved an accuracy of 76.61% before hyperparameter tuning. Hyperparameter tuning led to a modest improvement, with the final accuracy reaching 77.78%.

5. CONCLUSION

This study explored the application of machine learning for classifying paediatric heart sounds into normal and abnormal categories. We investigated two approaches: traditional machine learning classifiers and a Convolutional Neural Network (CNN).

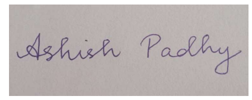
Traditional classifiers, particularly the Random Forest, achieved high accuracy (87.83%), demonstrating their effectiveness in this task. The CNN model, though achieving a lower initial accuracy (76.61%), showed improvement through hyperparameter tuning (77.78%). These results suggest promise for deep learning techniques in heart sound classification.

Future work could explore more sophisticated CNN architectures, feature engineering methods, and larger datasets to potentially achieve even greater accuracy and robustness. This research contributes to the development of automated tools that can assist medical professionals in early detection of paediatric cardiac conditions.

6. REFERENCES

- **Google Tensorflow Blog for audio classification**
- **Librosa**
- **ZCHSound: Open-Source ZJU Paediatric Heart Sound Database With Congenital Heart Disease**
- **Heart Sounds Classification Using Short-Time Fourier Transform and Gray Level Difference Method**

7. Contributions

Name	**Worked On**	**Signature**
Ashish Padhy	Code	
Jayesh Nayak	Report	
Animesh Panda	Presentation	
Arman Paikaray	Presentation