MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
UNIVERSITY OF ABDELHAMID IBN BADIS - MOSTAGANEM

Faculty of Exact Sciences and Computer Science

Department of Mathematics and Computer Science

Major: Computer Science

Thesis of Masters's Degree in Informatics

Option: **Artificial Intelligence for the Internet of Things (IA4IoT)**

Presented by:

BENKEDADRA FATIMA ZOHRA

Theme:

Heart Sound Classification using Convolutional Neural Network

Presented on: 28 May 2024

Supervisor: Mr. MOUMENE Mohammed elamine

University year 2023 - 2024

Abstract

Cardiovascular Diseases (CDV) is a term that groups the disorders related to the heart and blood vessels. One way to diagnose the CDV is using the heart sound where an abnormal sound is heard which indicates a problem. In this work we perform a heart sound classification using Convolutional Neural Network (CNN) and log mel spetrogram where we test different models on the available datasets. The experiments included in this work are a 2D-CNN model and adaptations of ResNet-18 and VGG-11 architectures. Results were evaluated based on accuracy, precision, recall, and F1-score metrics, with the pre-trained ResNet-18 model demonstrating superior performance, achieving an accuracy of 86% on the PASCAL dataset and 70% accuracy on the Physionet datasets of 2016 and 2022.

Keywords: CVDs, heart, sound, murmur, Deep learning, datasets, PhysioNet, data, augmentation, CNN, phonocardiogram classification, normal, abnormal, waveform

Résumé

Les maladies cardiovasculaires (MCV) regroupent les troubles liés au cœur et aux vaisseaux sanguins. Un moyen de diagnostiquer les MCV consiste à utiliser le son cardiaque, où un son anormal indique un problème. Dans ce travail, nous effectuons une classification des sons cardiaques en utilisant des CNN et des spectrogrammes de mel logarithmiques, où nous testons différents modèles sur les ensembles de données disponibles. Les expériences incluses dans ce travail comprennent un modèle CNN 2D et des adaptations des architectures ResNet-18 et VGG-11. Les résultats ont été évalués en fonction de la précision, de la rappel et du score F1, avec le modèle ResNet-18 pré-entraîné démontrant des performances supérieures, atteignant une précision de 86% sur l'ensemble de données PASCAL et une précision de 70% sur les ensembles de données PhysioNet de 2016 et 2022.

Mots-clés: MCV, cœur, son, souffle, apprentissage profond, ensembles de données, PhysioNet, augmentation des données, CNN, classification phonocardiogramme, normal, anormal

ملخص

إحدى طرق تشخيص أمراض القلب والأوعية الدموية هي عن بالاستماع إلى صوت القلب بسماعة الطبيب حيث يُسمع صوت غير طبيعي في صوت القلب مما يدل على وجود مشكلة مفيدة من هذه الطريقة أنها ميسورة التكلفة. يمكن أن يحفز هذا على دمج هذه الطريقة مع طرق التعلم العميق . قامت PhysioNet و باسكال بجمع إحدى مجموعات البيانات الغنية بصوت القلب. في هذا في العمل، نقدم طرق المعالجة المسبقة لإشارة صوت القلب لتحسين أداء نموذج الشبكة العصبية للوصول إلى تصنيف دقة صوت القلب إلى طبيعي وغير طبيعي بنتائج ٠٧٪ و ٦٨٪ على بيانات PhysioNet و باسكال على ترتيب. حيث من بين النماذج المستعملة حقق ResNet 81 هذه النتائج.

كلمات مفتاحية: تصنيف، صوت القلب، أمراض قلبية، تصنيف الإشارات، شبكات عصبية، تعلم عميق، ذكاء إصطناعي

Acknowledgements

I would like to express my deepest gratitude to my supervisor, MR.MOUMENE mohammed elamine, for their unwavering support, guidance, and encouragement throughout the journey of completing this thesis. Their expertise, valuable insights, and constructive feedback have immensely contributed to the refinement of this project.

I am also grateful for the academic environment cultivated by the institution and the professors who have played a pivotal role in shaping my research skills and intellectual growth. Their dedication to fostering learning and critical thinking has been invaluable to my academic journey.

List of Figures

List of Tables

Abbreviations

AV Aortic valve

CNN Convolutional Neural Network

CDV Cardiovascular Diseases

DCT discrete cosine transform

ECG Electrocardiography

HSMM Hidden Semi-Markov Model

MFCC Mel frequency cepstrum coefficients

MFSC Mel domain filter coefficients

MV Mitral valve

PCG Phonocardiogram

PV Pulmonic valve

STFT Short-Time Fourier Transform

TV Tricuspid valve

Contents

Introduction

Cardiovascular diseases CDV are the leading cause of death globally, and the early diagnosis of CDV which groups all types of heart diseases is very important. Usually, the doctor listens to the heart sound using a stethoscope to find any abnormality. For recording the heart sound a digital stethoscope is used. One of the benefits of this method is that it's affordable. This can motivate adapting this method with deep learning methods.

Among the various datasets available, those provided by PhysioNet in 2016 and 2022 hold significance. PhysioNet, a repository of freely available medical research data, offers a wealth of resources for researchers. Leveraging these datasets, our project focuses on detecting abnormalities in heart sounds, a fundamental aspect of AI-driven healthcare. The other dataset is the PASCAL dataset from the PASCAL challenge for heart sound classification.

Abnormal heart sound detection presents a formidable challenge in the realm of Artificial Intelligence (AI) for healthcare. Our project delves into waveform processing techniques, feature extraction methodologies, and sound classification using Convolutional Neural Networks (CNNs) to distinguish between normal and abnormal heart sounds.

This report is structured as follows: Chapter One provides an overview of the domain, data characteristics, and classification objectives. Chapter Two delineates the methodologies that underpin our research. Chapter Three elucidates the datasets employed and the associated data preprocessing techniques. Subsequently, Chapter Four details the experimental setup, while the final chapter delves into the results and ensuing discussions.

Chapter 1
Case study: Heart Sound

1.1 Introduction

The heart is the main part of the body it pumps blood to the rest of the body parts this mechanical movement of the heart muscles results in different sounds which can be heard using a stethoscope a medical device designed for this specific purpose. The heart contains four valves that control the blood flow, the valve opens and closes to let the right amount of blood through. Sometimes the valve is damaged which leads to abnormal blood circulation this results in an extra sound heard with the normal heart sound [4].

To understand the objective and the data we need to understand the heart sound signal this chapter explains the heart sound signal, murmurs, signal characteristics, and finally classification of the heart sound signal.

1.2 Heart sounds, murmurs and PCG

The two main heart sounds are Lub and Dub or S1 and S2 in a full normal cardiac cycle there are four states of the heart sound signal S1, systole, S2, and diastole. The cardiologist uses a stethoscope to listen to the heart sound recording the sound an electronic stethoscope is used. The signal is then plotted as a PCG to identify the segments S1, S2, systole, and diastole. The cardiologist listens to four main places illustrated in (Figure 1.1).

In some cases, cardiologists record an Electrocardiography (ECG) which is an electronic signal that also helps diagnose the CDV. The ECG signal recorded simultaneously with the PCG helps interpret the PCG. A cardiologist knows these sounds and can diagnose the problem just by listening to the heart. The heart sound can be saved in a waveform and plotted for better understanding as a PCG.

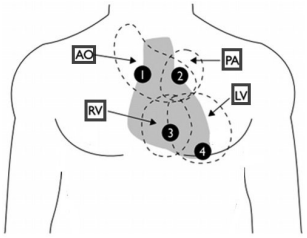


FIGURE 1.1 – Cardiac auscultation spots [1]; AO = aortic area; LV = left ventricle; PA = pulmonary area; RV = right ventricle; 1 = right second intercostal space; 2 = left second intercostal space; 3 = mid left sternal border (tricuspid); 4 = fifth intercostal space, midclavicular line

A murmur is an abnormal sound heard in the heart sound which indicates a heart problem. There are four main murmurs named after the valve location or the location where it's heard from. Also, the types of murmur are defined by the murmur characteristics which are the timing related to the cardiac cycle, the pitch, the shape, and the location.[1]

1.3 PCG automatic segmentation

The PCG signal is divided into four main segments illustrated in Figure 1.2 The segmentation of the PCG helps us calculate the number of beats and thus the segmentation of the audio file for optimization and input reshaping. Without automating the segmentation it can be very time-consuming because it helps experts easily locate the heart signal characteristics by providing suggestions and alternatively helps us use the segments without really understanding the signal. The most used methods try to find the two segments S1 and S2 only and conclude the rest, Some of the methods used are statistical methods based on Hidden Semi-Markov Model (HSMM) [2][5] and machine learning methods (CNN, clustering, etc) [1][6].

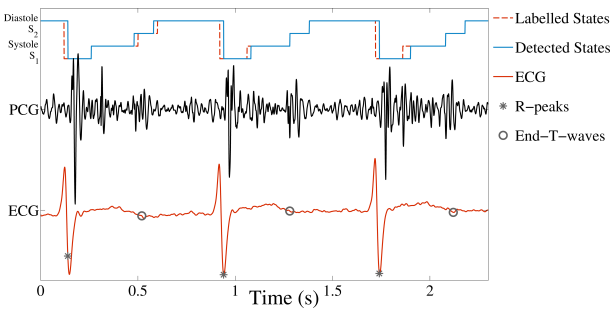


FIGURE 1.2 – Segmented, noisy, PCG using the LR-HSMM method with a clean ECG for reference [2]. The four segments S1, systole, S2 and Diastole

1.4 Signal quality and noise

One challenge faced when listening to the heart sound and analyzing the signal is the noise caused by other body parts the noise can be the lung sound, the abdominal sounds, or any other body parts sounds that collapse with the same frequency range of the heart sound. Other sounds are the device's movement or the patient's movement especially with children. Finally, the environment sounds such as humans talking, room echo, and others. These sounds are recorded unintentionally with the heart sound which can make the murmur sound hardly distinguished or unseen.

Some noise removal techniques use a high-pass, low-pass filter, or band-pass filter. We will explore the implementation of a low pass filter in chapter three where Figure 3.4 demonstrate the results of applying a low pass filter on a heart sound signal. [7]

1.5 Heart sound classification

The problem is taking the heart sound as input and classifying it. The classification can be binary where classes are normal, abnormal [7]. Multi-class classification with three classes normal, abnormal, and unsure or unknown [1] for when the signal is too noisy to identify or with the classes being normal and the murmur type Aortic Stenosis, Mitral Regurgitation, Mitral Stenosis, and Murmur in systole [8].

The choice of the classes depends on the dataset and the objectives. One advantage of binary classification is that the data will be more than in multi-class classification since all the abnormal cases are grouped in one class. One other reason to adapt binary classification is that since the classes are related to the auscultation area the abnormal class can be known only by the device placement. The third option of the unsure of unknown sound for when the signal is disfigured is not needed if the segmentation of the signal is done for beat identification.

Classification using machine learning methods for example Gaussian mixture models. Although machine learning methods are fast and work with small datasets they require complicated data prepossessing it also can be difficult to generalize the problem, unlike deep learning methods which can be adapted to any type of classification [3].

The process of heart sound classification usually starts with data acquisition where a

stethoscope is used to collect heart sounds from different groups (children, pregnant, healthy, unhealthy ... etc). The records need to be standardized and have the same sample rate, saved as waveform additional data can be collected such as the murmur characteristics and patient data. After we can process. The data processing can include frequency resampling, filtering, normalization, and segmentation to a standard optimal audio length[3].

1.6 Objectives

The objective of this project is to implement a group of processing and feature extraction methods on the Physionet/Cinc heart sound datasets of 2016 and 2022 and PASCAL challenge heart sounds dataset to perform a heart sound classification using colvolutional neural network.

1.7 Conclusion

The heart sound is a signal with certain characteristics, the signal can be found mixed with other sounds. Thus it requires cleaning before identifying normal signals and abnormal or murmur which can be a challenge itself. As for the segments they give us a hint on the number of the heartbeats in a recording. The next chapter dives into some suggested solutions for heart sound classification.

Chapter 2
State of the Art

2.1 Introduction

The integration of neural networks in heart sound classification gain popularity with the release of the PASCAL and Physio-net/CinC challenges when more than 10 works were published [3]. In this chapter, we will take a walk to discover the recent methods applied in previous research which helped connect this project.

2.2 Log-mel spectrogram for feature extraction

Spectrograms are a grid-like representation of the audio signal used for training a CNN and also for audio analyzing. The mel-spectrogram, derived from the mel-frequency scale that mimics human auditory perception, offers enhanced resolution particularly for lower frequencies compared to the standard spectrogram [9]. The log mel spectrogram is a mel spectrogram with a log function applied. Figure 3.4 is an example of a signal with log mel spectrogram applied to it. This equations of computing the log mel spectrogram from the raw audio waveform using the given functions are as follows.

Compute the Short-Time Fourier Transform (STFT): When calculating the Short-Time Fourier Transform (STFT) a windowing method is applied for example hann window to prevent spectrum leakage before calculating the Fourier transform.

$$S = \text{STFT}(x,\ \text{n\_fft},\ \text{hop\_length})$$

Compute the power spectrum:

$$P = |S|^2$$

Apply Mel filterbank to the power spectrum:

$$M = \text{mel\_basis} \cdot P$$

Apply logarithmic transformation with offset ε:

$$\text{log\_mel\_spectrum} = \log_{10}(M + \varepsilon)$$

Where:

- x is the input audio signal,
- S is the spectrogram obtained from the STFT,
- P is the power spectrum obtained from the spectrogram S,
- mel_basis is the Mel filterbank applied to the power spectrum P,
- M is the Mel spectrogram obtained from applying the Mel filterbank,
- ε is a small offset added to avoid taking the logarithm of zero,
- log_mel_spectrum is the output log mel spectrogram.

A log mel spectrogram typically has only one channel. It represents the intensity or power of frequencies across time, where each pixel in the spectrogram corresponds to the energy or magnitude of a specific frequency band at a particular time frame. Therefore, it is a two-dimensional representation of the audio signal, with frequency on the y-axis and time on the x-axis. [9]

2.3 Audio augmentation methods

Audio data augmentation techniques generate new audio sounds from the original audio sounds. Some of the audio data augmentation methods are time stretching, noise injection, and signal horizontal inverting. Figure 2.1 demonstrate these methods on a random signal from the Physionet dataset.

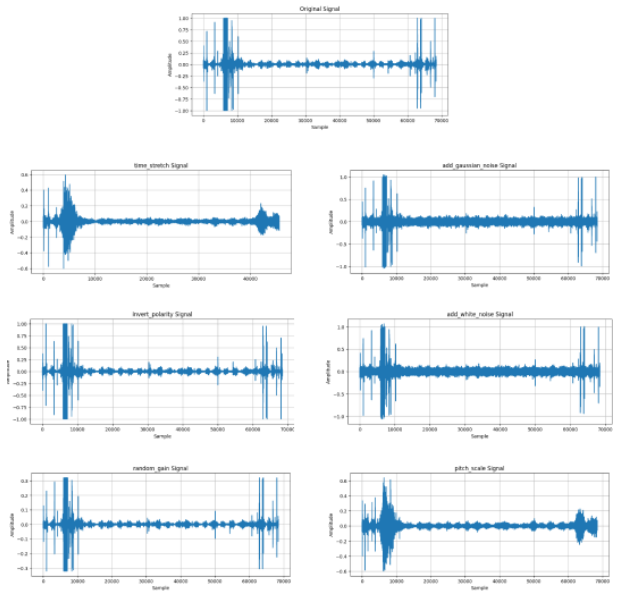


FIGURE 2.1 – Diffrent audio augmentation functions applied on a heart sound audio data with random values for parameters.

2.4 CNN for classification

A CNN, also known as a grid-like topology, is a specialized type of neural network for processing both time-series data and image data. Figure 2.2 convolution applied to a 2D tensor. The kernel unit network of a CNN is a convolution network that uses a specialized kind of linear operation instead of general matrix multiplication in more than one layer. In the case of a 2D CNN, the input x and the kernel are 2D matrixes. The CNN has a three-part Input layer to define the input shape, a convolutional layer for feature extraction, and a fully connected layer or classification layer.

The kernel unit network of a CNN is a convolution network that uses a specialized kind of linear operation instead of general matrix multiplication in more than one layer. When the input is a one-dimensional vector, the output feature map y can be calculated by a discrete convolution operation, which is typically expressed as

$$y(n) = x(n) * (n) = \sum_{m=-\infty}^{\infty} x(m)(n-m) \quad [3]$$

where $*$ denotes a convolution operation and is a convolution kernel. This is usually applied to a 1D CNN. In the case of a 2D CNN, the input x and the kernel are 2D matrices, and

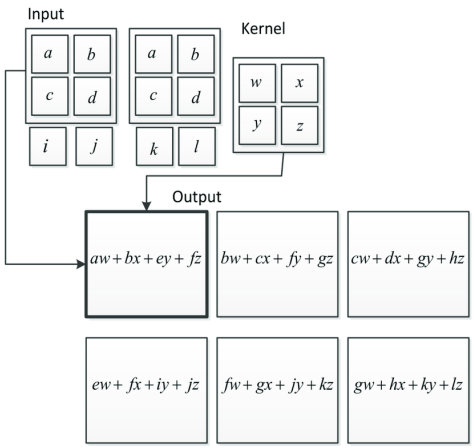


FIGURE 2.2 – 2D convolution operation in the architecture of CNNs [3]

the output feature map $Y(i,j)$ can be computed as

$$Y(i,j) = X(i,j) * (i,j) = \sum_m \sum_n X(m,n)(i-m, j-n) \quad [3]$$

To perform a classification using CNN we first create the architecture of the network, we choose the optimizer, learning rate, batch size, epochs, and the loss function.

One suggested optimizer in the learning phase changes the attributes of your neural network such as weights and learning rate and reduces the losses. The Adam short for Adaptive Moment Estimation optimizer can be used when training a CNN. in 2014, this approach boasts computational efficiency, minimal memory usage, resilience to diagonal rescaling of gradients, and suitability for handling large-scale problems with extensive data and parameters [10].

Two-dimensional CNN-based methods for heart sound classification leverage the success of CNNs in computer vision. They convert 1D heart sound signals into 2D feature maps, often using Mel domain filter coefficients (MFSC), Mel frequency cepstrum coefficients (MFCC), and spectrograms. A proposed 2D CNN approach for automatic recognition, achieving a 72.78% accuracy in a challenge [3]. When combined with MFCCs and spectrograms, the network reached an 81.1% accuracy [3]. Maknickas et al [11]. developed a 2D CNN with MFSCs, achieving an 86.02% accuracy and using Inception and ResNet networks for further improvement. MFCC features, obtained through discrete cosine transform (DCT), have also been utilized in CNNs for heart sound classification. [3]

2.5 Conclusion

The previous works included some methods on different datasets. In this work, we adopt a combination of spectrogram and convolutional neural network for the heart sound classification as a solution.

Chapter 3
Datasets

3.1 Introduction

The selected datasets are from the Physionet challenges [7] [1] from 2016 and 2022. One reason to use these datasets is their diversity and it contains recordings from different groups. The 2022 dataset adds a small missing age group 'children' to the 2016 dataset. The two datasets were used in both training and validation, two sub-datasets from the 2016 dataset were selected for testing the f datasets. The third datset is the PASCAL challenge dataset. [12]

3.2 The PhysioNet/CinC Challenge 2016 Dataset v1.0.0

The PhysioNet/CinC Challenge 2016 dataset includes sourced heart sounds globally from both clinical and nonclinical settings, encompassing healthy subjects and patients with cardiac conditions. The dataset comprises five databases with 3,126 recordings, categorized as normal or abnormal. Recordings vary in length and are resampled to 2,000 Hz. The challenges this dataset presents are the complexity of the real-world recordings and the presence of noise. [7]

3.3 The CirCor DigiScope Phonocardiogram 2022 Dataset v1.0.3

The heart sound dataset was collected during two mass screening campaigns in Northeast Brazil in 2014 and 2015, approved by the institutional review board. Over 2,000 participants attended. Participants underwent clinical examinations, nursing assessments, and cardiac investigations, with data analyzed by expert pediatric cardiologists. Quality assessment

ensured accurate data collection.

The dataset comprises 63% children and 20% infants. Additionally, acknowledging the complexities of caring for individuals with complex Congenital Heart Disease (CHD), a small number of young adults, who volunteered for screening, were also included for examination.

Audio samples were collected from four auscultation points (areas) Figure 1.1 the areas are coded as follows in the recording's names Aortic valve (AV), Pulmonic valve (PV), Tricuspid valve (TV), Mitral valve (MV) [1]:

- Aortic valve (area 1): second intercostal space, right sternal border;
- Pulmonary valve (area 2): second intercostal space, left sternal border;
- Tricuspid valve (area 3): left lower sternal border;
- Mitral valve (area 4): fifth intercostal space, midclavicular line(cardiac apex).

The used device is a Littmann 3200 stethoscope with DigiScope Collector technology [13], sampled at 4 kHz and normalized between -1 and 1. The PCG files from the CC2014 and CC2015 campaigns had an average duration of 28.7 seconds and 19.0 seconds, respectively, with a significant difference confirmed by a Mann-Whitney U test. [1]

The Logistic regression-HSMM-based [2], Adaptive sojourn time HSMM [5] algorithms and a CNN were used to segment recordings into fundamental heart sounds (S1 and S2), with manual annotation required for discrepancies. A cardiac physiologist then blindly classified murmur events in the recordings, which included noise from various sources like stethoscope rubbing and background sounds. [1]

3.4 The PASCAL challenge dataset 2011

This dataset was released for the challenge "Classifying Heart Sounds Challenge" It includes only audio recordings and no information about the patients. The dataset is grouped into two groups A with 31 normal and 34 murmur sounds and B with 320 normal and 95 murmur sounds. Other sounds are provided of artifacts and uncertain conditions but we are not interested in that. The audio files vary in length, from 1 to 30 seconds, with some clipped for noise reduction. The categories are as follows.

Normal Category: Typical healthy heart sounds with a clear ”lub dub” pattern, sometimes with background noise. Heart rates can vary (60-100 bpm), with occasional noise disruptions.

Murmur Category: Abnormal heart sounds characterized by ”whooshing” between ”lub” and ”dub” or vice versa. Murmurs can signal various heart disorders and occur alongside regular heart sounds.

Extra Heart Sound Category (Dataset A): Additional sounds like ”lub-lub dub” or ”lub dub-dub.” These may indicate disease or occur normally. They’re important to detect as they’re not easily identified via ultrasound.

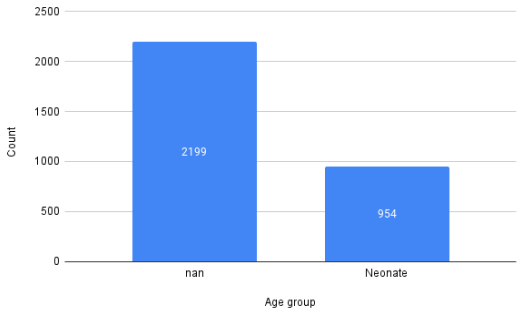
Artifact Category (Dataset A): Diverse sounds like feedback, speech, or music, lacking discernible heart sounds. Distinguishing artifacts is crucial for data accuracy.

Extrasystole Category (Dataset B): Irregular heartbeats with extra or skipped beats, producing irregular heart sounds. While common, they can indicate heart diseases and early detection aids effective treatment.

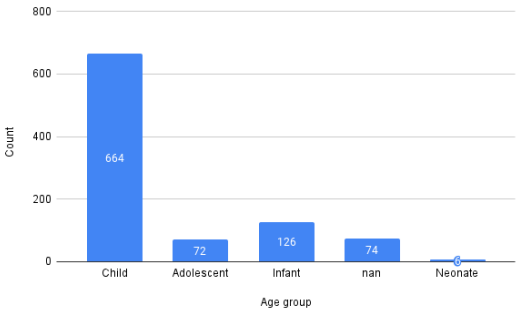
3.5 Exploratory data analysis

In this section, we explore the data from the 2016 and 2022 datasets. The age groups from each dataset are presented in Figure 3.1 The age group is correlated with the noise in the signal the audio recorded from children is more noisy than that of any age group where Neonates including birth to 27 days old Infant includes 28 days old to 1 year old Children include 1 to 11 years old Adolescents include 12 to 18 years old Young Adults are 19 to 21 years old and nan is missing data [1]. The CirCor dataset includes several children’s records which represents a challenge.

The PhysioNet/CinC Challenge 2016 Dataset v1.0.0 is not Figure 3.2 where from 3153 patients only 21.1% have CDV. The CirCor DigiScope Phonocardiogram 2022 Dataset v1.0.3 has 942 patients with 48.4% having CDV. The number of recordings in the 2022 dataset is 1517 abnormal recordings with 397 from AV, 353 from PV, 429 from MV, and 338 from TV. 1621 normal recordings. 396 from AV, 411 from PV, 388 from TV and 426 from MV. The total number of recordings from both datasets is 6291. The murmur is mostly audible in one location.

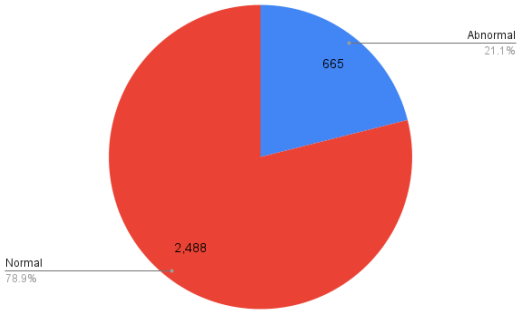


(A) Age groups in the PhysioNet/CinC Challenge 2016 Dataset v1.0.0

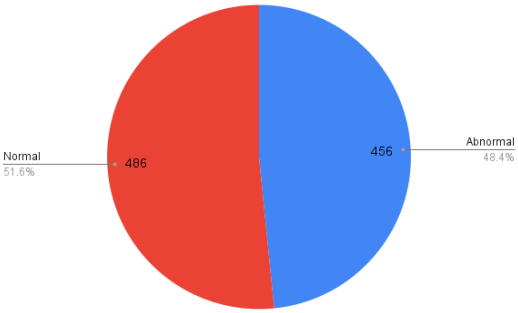


(B) Age groups in the CirCor DigiScope Phonocardiogram 2022 Dataset v1.0.

FIGURE 3.1 – Age groups included in each dataset

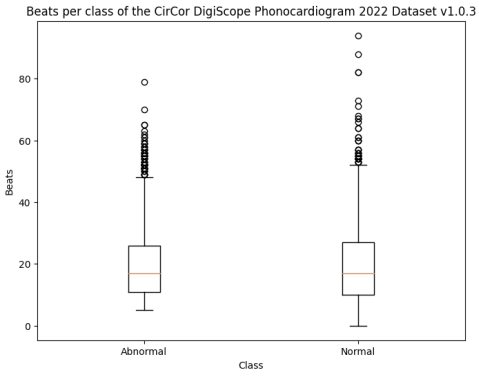


(A) The PhysioNet/CinC Challenge 2016 Dataset v1.0.0

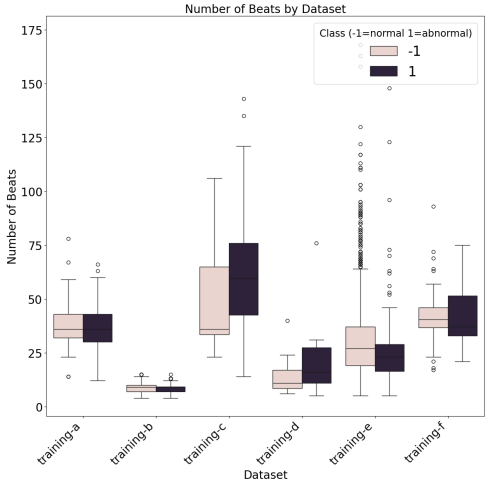


(B) The CirCor DigiScope Phonocardiogram 2022 Dataset v1.0.3

FIGURE 3.2 – Count of records per class in each dataset



(A) Beats in the CirCor DigiScope Phonocardiogram 2022 Dataset v1.0.3



(B) Beats per sub-datasets (a, b, c, d, e, f) from the PhysioNet/CinC Challenge 2016 Dataset v1.0.0

FIGURE 3.3 – Beats in each dataset

Figure 3.3b represents the number of beats in each sub-dataset in the 2016 dataset. The order of the sub-datasets is from left to right (a, b, c, d, e, f) the right box is the abnormal class and left box is the normal class.

3.6 Data preprocessing

The data was passed through a preprocessing pipeline for applying a Butterworth filter to prevent anti-aliasing and for noise reduction, Sample the audio to 4000 Hz and 8000 Hz if not already next the audio was segmented into 2 and 1 second long segments and removed the first and last segments and finally transformed into a Log Mel spectrum 2.2 Normalization is done with the Librosa normalization method min-max method), and the resampling is also done with the Librosa resample method. The Python application of the Butterworth filter is as follows:

```
# Low-pass filter parameters
cutoff_frequency = 100
order = 5  # Filter order

# Design a low-pass Butterworth filter
def butter_lowpass(cutoff, fs, order=5):
    nyquist = 0.5 * fs
    normal_cutoff = cutoff / nyquist
    b, a = butter(order, normal_cutoff, btype='low', analog=False)
    return b, a

def apply_lowpass_filter(data, cutoff, fs, order=5):
    b, a = butter_lowpass(cutoff, fs, order=order)
    y_filtered = filtfilt(b, a, data)
    return y_filtered
```

Here's a walk through the code. The order determines the sharpness of the filter's frequency response. The normal_cutoff is the normalised cutoff. The Nyquist frequency is defined as half of the sampling rate (fs). This relationship arises from the Nyquist-Shannon sampling theorem, which states that to accurately reconstruct a continuous-time signal from its sampled version, the sampling rate must be at least twice the highest frequency component

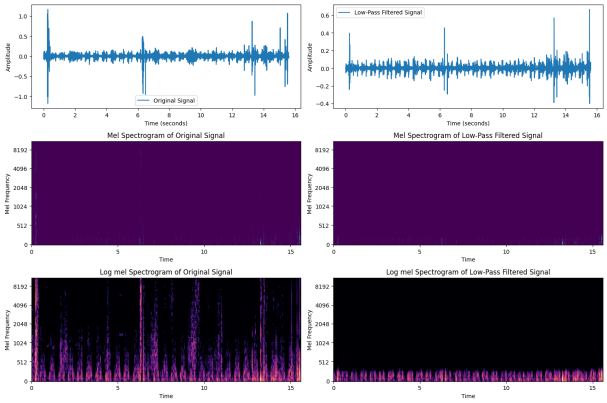


FIGURE 3.4 – The implementation of a log mel spectrogram and mel spectrogram on filtered and not filtered heart sound signals. Left signal: not filtered, Right signal: filtered with a Butterworth filter of order 5 and cutoff 100

present in the signal.

3.7 Conclusion

We took a look at the used data characteristics and problems and the use of log mel spectrograms for the feature extraction and audio analysis including how it shows the role of filters and their effect on noise. The next chapter will discuss the experiments we conducted on the processed data.

Chapter 4
Experimentations and Results

4.1 Introduction

In this chapter, we explore the model architecture and results of the test data. Multiple experiments were done on several models we present only the models with good results. Experimentation parameters such as segment length, applied filtering, augmentation functions, and hyperparameters were carefully selected and varied across experiments.

4.2 Tools and Methods

Python 3 programming language is used for this project. PyTorch and TensorFlow v2.16.1 [14] are used for model training and librosa v0.10.2 [15] for data processing.

The training method is demonstrated in Figure 4.1. The training methodology involved the use of datasets from 2016 and 2022, with evaluation conducted on the training-f dataset. Data preprocessing included segmenting, filtering, and augmentation techniques. The training data is 80% and the validation data is 20%.

The data is passed through a processing pipeline after splitting into train/validation and applied 2 random augmentation methods out of the previously mentioned 2.3 on the abnormal data to balance the classes for training only or no augmentation methods are applied. Multiple models were trained on different data and with different data processing pipelines. The augmentation and data balancing are optional.

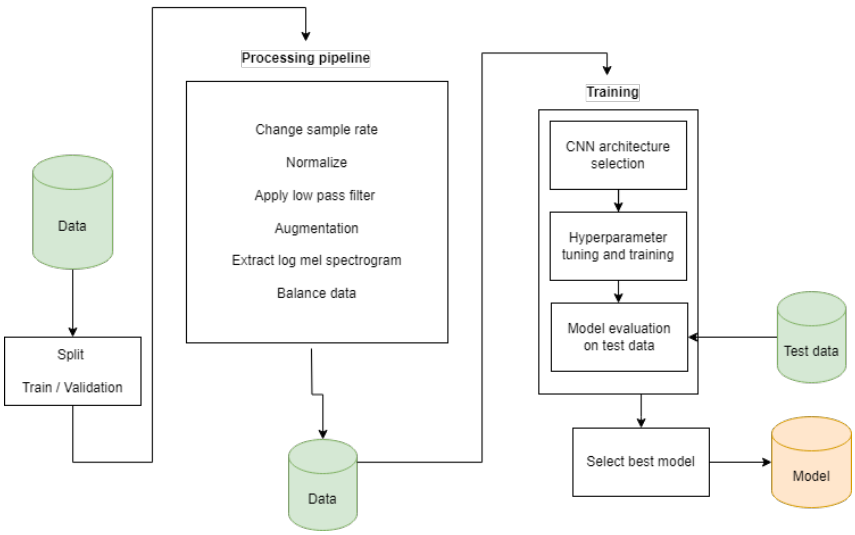


FIGURE 4.1 – Training workflow

4.3 2D-CNN

First, we had to make sure that the processed data was good and that the processing pipeline was valid data using a simple CNN we got the results in the table results. The used data is the 2016 section 3.2 and 2022 section 3.3 datasets and for evaluation, the training-f sub-dataset is used.

Experimentation parameters:

- Segment length: 2 seconds
- Sample rate: 4 kHz
- Applied filtering: Butterworth filter
- Augmentation functions used: all 6 section 2.3
- Segmentation method: All segments
- Model hyperparameters: 20 epochs, 0.001 learning rate, Adam optimizer, Binary Cross-Entropy Loss function.
- Model architecture Figure 4.2

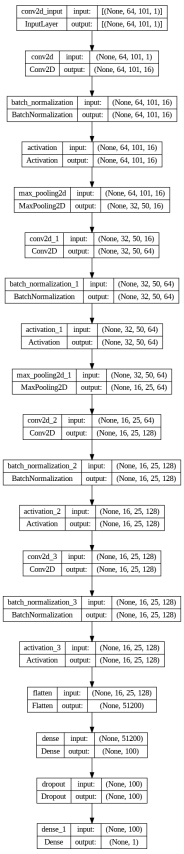


FIGURE 4.2 – Suggested CNN model

4.4 ResNet-18

ResNet18 is a convolutional neural network architecture featuring 18 layers, known for introducing residual connections to address the vanishing gradient problem, enabling the training of deeper networks more effectively. Its design includes residual blocks with skip connections, facilitating efficient information flow and improved performance in image classification tasks. [16]

Experimentation parameters:

- Segment length: 1 second
- Applied filtering: None
- Sample rate: 8 kHz
- Augmentation functions used: addGaussianNoise
- Segmentation method: One segment from each audio

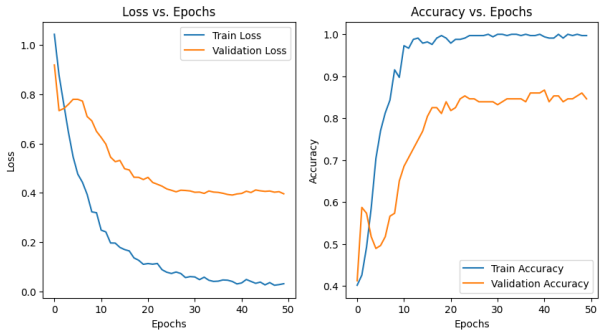


FIGURE 4.3 – Training ResNet18 on PASCAL data

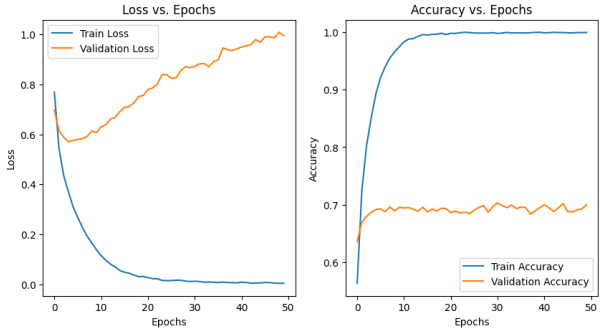


FIGURE 4.4 – Training ResNet18 on Physionet 2016/2022 data

- Model hyperparameters: 50 epochs, 1e-5 learning rate, Adam optimizer, Cross Entropy Loss function.

4.5 VGG-11

VGG11 is a CNN with 11 layers, including 8 convolutional and 3 fully connected layers. Its simple yet deep architecture makes it a fundamental model for image classification tasks. [17]

Experimentation parameters:

- Segment length: 1 second
- Sample rate: 8 kHz
- Applied filtering: None
- Segmentation method: One segment from audio
- Augmentation functions used: None

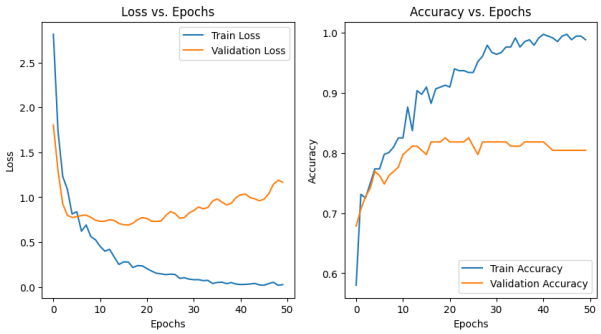


FIGURE 4.5 – Training VGG-11 on PASCAL data

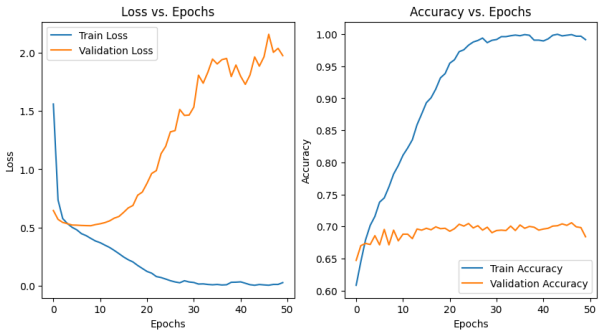


FIGURE 4.6 – Training VGG-11 on Physionet 2016/2022 data

- Model hyperparameters: 50 epochs, 1e-5 learning rate, Adam optimizer, Cross Entropy Loss function.

4.6 Evaluation criteria

The evaluation criteria are used to monitor the model's performance during the training phase and to evaluate the model's performance on the test data.

- The Area Under the Curve (AUC): Is a widely used metric in machine learning, particularly in binary classification tasks, to evaluate the performance of a model's prediction probabilities across different decision thresholds.

- Precision: Ratio of correctly predicted positive observations to the total predicted positive observations. Useful when the cost of false positives is high.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- Accuracy: Ratio of correctly predicted observations to the total observations. Used for evaluating classification models when classes are balanced.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}}$$

- Recall (Sensitivity): Ratio of correctly predicted positive observations to all observations in the actual class. Important when the cost of false negatives is high.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- F1 Score: Harmonic mean of precision and recall. Provides a balance between precision and recall, especially when classes are imbalanced.

$$\text{F1 Score} = \frac{2\ \text{Precision}\ \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.7 Results and Discussions

Results for experimentation 1 section 4.5 to evaluate the augmentation and processing pipeline in Table 4.1 with model AUC increasing by 16%. indeed, the model performance is not better than a random guessing but it's not the point. The point is to test the processing pipeline and the implemented methods for the augmentation.

The pre-trained ResNet18 gave the best results compared to the other networks. Also, the choice of the dataset has a role. since the PASCAL data is not as critical as the Physionet dataset training the model on the PASCAL dataset gave better results.

Processing techniques	Accuracy	Recall	Precision	F1 Score	AUC
No processing	0.53	0.64	0.31	0.42	0.43
Processed without augmentation	0.43	0.75	0.28	0.41	0.53
Processed with augmentation	0.58	0.56	0.33	0.42	0.59

TABLE 4.1 – Results of classification on the training-f dataset using 2D-CNN section 4.3

Dataset	Accuracy	Precision	Recall	F1-score
PASCAL	0.8601	0.8554	0.8601	0.8569
Pysionet datasets 2016 / 2022	0.7075	0.7003	0.7075	0.7019

TABLE 4.2 – Results of classification using ResNet1-8 section 4.4

Dataset	Accuracy	Precision	Recall	F1-score
PASCAL	0.8042	0.8036	0.8042	0.7910
Pysionet datasets 2016 / 2022	0.6839	0.6957	0.6839	0.6877

TABLE 4.3 – Results of classification using VGG-11 section 4.5

4.8 Conclusion

In conclusion, the experiments demonstrate the importance of data processing techniques, such as augmentation, and the choice of model architecture in achieving high classification accuracy. ResNet-18 shows promising results and could be further explored for audio classification tasks.

Conclusion

In conclusion, one problem faced when handling heart sound signals is signal quality, feature extraction method, and network architecture selection. In this project, we used a low pass filter for noise reduction, log mel spectrogram for feature extraction, and CNN for abnormal sound detection. The study compared the performance of a 2D-CNN model and adaptations of ResNet-18 and VGG-11 architectures for image classification. The findings suggest that ResNet-18 is a promising architecture for image classification tasks.

One of the solutions that could be considered in the future is the use of a multi-model architecture for the three phases noise reduction, feature extraction, and classification.

Bibliography

[1] J. Oliveira, F. Renna, P. Dias Costa, D. Nogueira, C. Oliveira, C. Ferreira, M. Alipio, S. Mattos, T. Hatem, T. Tavares, A. Elola, A. Bahrami Rad, R. Sameni, G. Clifford, and M. Coimbra, "The circor digiscope dataset: From murmur detection to murmur classification," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–1, 12 2021.

[2] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-hsmm-based heart sound segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 822–832, 2016.

[3] W. Chen, Q. Sun, X. Chen, G. Xie, H.-Q. Wu, and C. Xu, "Deep learning methods for heart sounds classification: A systematic review," *Entropy*, vol. 23, p. 667, 05 2021.

[4] S. Ghosh, P. R N, and R. Tripathy, *Heart Sound Data Acquisition and Preprocessing Techniques: A Review*, p. 300. 02 2020.

[5] S.-Z. Yu, "Hidden semi-markov models," *Artificial Intelligence*, vol. 174, no. 2, pp. 215–243, 2010. Special Review Issue.

[6] S. Das, S. Pal, and M. Mitra, "Deep learning approach of murmur detection using cochleagram," *Biomedical Signal Processing and Control*, vol. 77, p. 103747, 08 2022.

[7] C. Liu, D. Springer, Q. Li, B. Moody, R. Abad, F. Chorro, F. Castells Ramon, J. Roig, I. Silva, A. Johnson, Z. Syed, S. Schmidt, C. Papadaniil, L. Hadjileontiadis, H. Naseri, A. Moukadem, A. Dieterlen, C. Brandt, H. Tang, and G. Clifford, "An open access database for the evaluation of heart sound algorithms," *Physiological Measurement*, vol. 37, pp. 2181–2213, 11 2016.

[8] Yaseen, G.-Y. Son, and S. Kwon, "Classification of heart sound signal using multiple features," *Applied Sciences*, vol. 8, no. 12, 2018.

[9] D. de Benito, A. Lozano-Diez, D. Toledano, and J. Gonzalez-Rodriguez, "Exploring convolutional, recurrent, and hybrid deep neural networks for speech and music detection in a large audio dataset," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2019, 06 2019.

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[11] V. Maknickas and A. Maknickas, "Recognition of normal-abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients," *Physiological Measurement*, vol. 38, 06 2017.

[12] P. Bentley, G. Nordehn, M. Coimbra, and S. Mannor, "The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results." http://www.peterjbentley.com/heartchallenge/index.html.

[13] D. Pereira, F. Hedayioglu, R. Cruz-Correia, T. Silva, I. Dutra, F. Almeida, S. Mattos, and M. Coimbra, "Digiscope - unobtrusive collection and annotating of auscultations in real hospital environments," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2011, pp. 1193–6, 08 2011.

[14] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[15] L. docs, "`https://librosa.org/doc/latest/index.html`," 2024.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[17] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.