FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

A Deep Learning approach to infer morphological characteristics of the heart from cardiac sound analysis

Luís Francisco Torres Andrade

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Rui Camacho

Second Supervisor: Jorge Oliveira

October 31, 2022

Resumo

Sendo a maior causa de mortalidade a nível global, as doenças cardiovasculares continuam a tirar a vida a cerca de 17.9 milhões de pessoas por ano [1]. O resultado de investigações contínuas focadas em novas tecnologias e metodologias permitiram a aquisição e inspeção de dados na maioria dos países desenvolvidos, no entanto, na maioria dos países em desenvolvimento, devido à pobreza e à comum escassez de recursos, isto não se verificou. A utilização da auscultação cardíaca representa uma das técnicas de deteção e identificação de doenças cardíacas com maior custo-benefício. Neste trabalho, dados cardiovasculares adquiridos através do processo de auscultação cardíaca serão analisados, de modo a tentar detetar sopros cardíacos, usando métodos de inteligência artificial.

O ponto principal deste trabalho consiste no desenvolvimento de uma abordagem inovadora em Deep Learning para análise das componentes de frequências dos sons cardíacos, também conhecido como o "pitch", e o estudo sobre a relevância que esta característica tem no processo de classificação e deteção da presença de certas doenças cardíacas em pacientes.

Esta pesquisa pode servir de suporte para técnicas de "screening" usadas em locais remotos ou em países em desenvolvimento onde doenças como a febre reumática constituem um problema extremamente sério no que toca a saúde pública, como por exemplo, nos estados mais rurais do Brasil.

[1] https://www.who.int/health-topics/cardiovascular-diseases

Abstract

As the major cause of deaths worldwide, cardiovascular diseases continue to take the lives of about 17.9 million people per year [2]. Continuous research on new technology and methodologies allowed the acquisition and examination of data in most developed countries around the world, however, in developing countries, due to poverty and common scarcity of resources, this has not been verified. The usage of cardiac auscultation represents one of the most cost-effective techniques used to detect and identify many heart conditions. In this work, cardiovascular data acquired using cardiac auscultation is going to be analysed, aiming to detect cardiac murmurs through artificial-intelligent algorithms.

The project's focal point will revolve around the development of an innovative Deep Learning approach for the frequency components of heart sounds, therefore a pitch analysis, and how this sound characteristic may be relevant in helping to detect the presence of certain heart diseases in patients.

This research might support screening tools, namely in remote and low-income countries where diseases such as rheumatic fever impose a severe public health issue, e.g., rural Brazil.

[2] https://www.who.int/health-topics/cardiovascular-diseases

Acknowledgments

Firstly, I would like to thank my supervisors, Professor Rui Camacho and Professor Jorge Oliveira, for all the guidance and help provided throughout the development of this dissertation.

I am beyond thankful to my parents and my grandmother for all of the extraordinary patience, motivation, strength and belief passed to me during this period. I would also like to particularly thank my grandfather, despite his absence, for everything that he passed to me throughout the years. Additionally, I would also like to thank Balu for making me smile at all times, despite all the work and pressure, whatever the hour.

Words cannot express my gratitude to my very special person, Beatriz. I thank her with all my heart for all the unconditional patience, support, reassurance and belief and am beyond grateful to have shared this, of many journeys, with her.

I would also like to thank Zé Diogo for making sure to keep me under his radar and for helping me at all times.

Lastly, to all my friends and family, thank you for constantly inspiring me and always keeping my spirits high.

Francisco Andrade

"Sometimes the difference between the unfinished painting and the finished painting is as simple as finding the right frame"

Frederick Jay Rubin

Contents

List of Figures

List of Tables

Abbreviations and Symbols

ANN	Artificial Neural Networks
AV	Aortic auscultation point
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EWT	Empirical Wavelet Transform
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
IIR	Infinite Impulse Response
MV	Mitral auscultation point
PCG	Phonocardiogram
PV	Pulmonary auscultation point
RNN	Recurrent Neural Network
S1	First Heart Sound
S2	Second Heart Sound
S3	Third Heart Sound
S4	Fourth Heart Sound
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TV	Tricuspid auscultation point

Chapter 1

Introduction

In this first chapter, a brief introduction of the dissertation's problem is presented, along with its actual context, while also exposing all contributions regarding its development. Lastly, the structure of the dissertation is presented.

1.1 Motivation

Cardiovascular diseases remain, after years and years, the leading cause of death in the world. Not only taking responsibility for 31% of the global death rate [2], these diseases also pave the way for a decreased quality of life attached to morbidity and disabilities, which can also lead to a perfectly understandable development of a sense of fear towards the behaviour of the condition itself. As shocking as cardiovascular deaths data is in developed countries, everything worsens when poverty comes into the equation. The lack of infrastructures, medical assistance and resources all together cause a major delay in the diagnosis of these diseases, which shortens the life expectancy period, causing early deaths.

Taking this into account, it is known that technological discoveries are usually not available due to their expensive acquirement and maintenance, so there is a need to use cheap yet still reliable solutions to guarantee the assistance needed. This role seems to be filled by cardiac auscultation, which still remains one of the most important and cost-effective tools for monitoring cardiovascular diseases through the analysis of sounds generated by several mechanical activities of the heart [2]. Despite its ordinarily easy access, cardiac auscultation is a challenging skill to master, requiring continuous and intense training, which medics from underdeveloped countries usually lack due to the adverse living conditions of their respective country.

This reality has been well present in the state of Pernambuco, in the northeast region of Brazil. As presented in [7], from 1996 to 2016, there were recorded a total of 3,584 deaths from cardiovascular malformations in this area, with a staggering 81.94% concentrated in children aged under one-year [7]. Due to this, the "Caravana do Coração" project organizes cardiac screening campaigns every year to monitor these locals' cardiac health. All participants undergo clinical examination, nursing assessment and cardiac investigation before proceeding with the electronic

auscultation procedures, whose audio sample results are then stored and annotated. With the data gathered, a dataset for further investigation was developed to improve detection regarding this problem that affects so many regions across the world [2]. The resulting dataset was named "CirCor DigiScope Dataset" and consists of the largest pediatric heart sounds dataset available, gathering audio data and manual annotations from a total of 1,568 patients and about 215,780 heart sounds. A total of 647 single or multiple diagnoses were confirmed, with simple congenital cardiopathy topping the total percentage with about 30.2% [2] while also detecting murmur presence in about 305 patients.

Despite not always directly representing the presence of heart diseases since they can sometimes be harmless, heart murmurs constitute one of the main indicatives in predicting heart diseases. In the CirCor DigiScope Dataset, all heart murmurs detected were annotated by their respective characteristics, such as timing, pitch, grading, among others. The analysis of each characteristic can aid in the accurate detection of heart conditions.

1.2 Work Proposal

With the last mentioned point in mind, the main proposal of this dissertation relies on the development of a deep learning based system whose primary task consists in specifically analyzing one of the latter heart murmur characteristics in order to understand its corresponding relevancy when it comes to the early detection of a cardiac murmur and how its classification can be used to detect cardiac valve diseases. Following the work done by [8], who studied the importance of the number of auscultation points in detecting heart murmurs for several different algorithms, the chosen characteristic for this dissertation's analysis consists of the frequency components of a heart murmur; therefore, the pitch. It is important to mention that the classification of this sound characteristic associated to the prevention and detection of heart diseases has never been explored before. This work will be developed by using the CirCor DigiScope Dataset, which, as mentioned previously, includes specific pitch-type values for each of the detected murmurs, as well as analysis from timing, grading, shape and quality perspectives that could be used to cement several decisions by adding more information to the input data.[2].

1.3 Structure of the Dissertation

Apart from the current introductory chapter, this final report is composed by the following chapters:

- Background (Chapter 2): the main focus of this chapter is to present the basic and fundamental theory behind the heart's functioning, its cardiac cycle, heart murmurs, insights on the cardiac auscultation procedure and heart diseases.

- State of the art (Chapter 3): in this chapter, the state of the art review is presented in order to mention some of the already existing technologies when it comes to audio pre-processing, feature extraction and machine learning models.
- Methodology (Chapter 4): in this chapter, the used dataset, CirCor DigiScope Dataset, is minutely explained, followed by a step-by-step presentation and description of all procedures taken throughout the development of the project and a description of the performance metrics used.
- Experimental results and discussion (Chapter 4): in this chapter, all results obtained for each procedure are exposed, discussed and compared.
- Conclusion (Chapter 5): at last, the main focus of this chapter consists in presenting conclusions regarding all of the registered experiments and enumerating several improvements and suggestions for future works.

Chapter 2

Background

In this chapter, several fundamental theoretical concepts such as the heart's physiology, heart sounds, the cardiac cycle and its correspondent stages and heart diseases will be presented.

2.1 Theory of the Heart

Heart sounds are mainly generated due to the vibration caused by the opening and closing of the cardiac valves during the cardiac cycle, as blood flows through the heart chambers at a particular turbulence that, if high, translates to an also higher number of vibrations obtained [9].

2.1.1 Physiology of the Heart

The heart can be divided into four different chambers: the left atrium, the left ventricle, the right atrium and the right ventricle. There are also two distinct types of cardiac valves - semilunar valves and atrioventricular valves - existing two different valves for each type: the pulmonary valve and the aortic valve, and the mitral valve and the tricuspid valve, respectively. The two valves that comprise the first pair are located between the ventricles and the arteries to create a separation between the two [9]. The pulmonary valve is responsible for separating the pulmonary artery from the right ventricle. It is located at the second intercostal space on the border of the left sternal [2]. In contrast, the aortic valve, located in the same intercostal space except on the right border side of the sternal [2], separates the opposite side ventricle from the aorta. The atrioventricular valves, on the other side, separate both atriums from both ventricles, being the right atrium and right ventricle separation assigned to the tricuspid valve, which is located in the left lower part of the sternal border [2], while the mitral valve separates the left atrium from the left ventricle and is located on the cardiac apex [2], [9]. The correct understanding of these four separate valves' functionalities and locations is crucial since they represent the four optimal cardiac auscultation positions when collecting heart sound data.

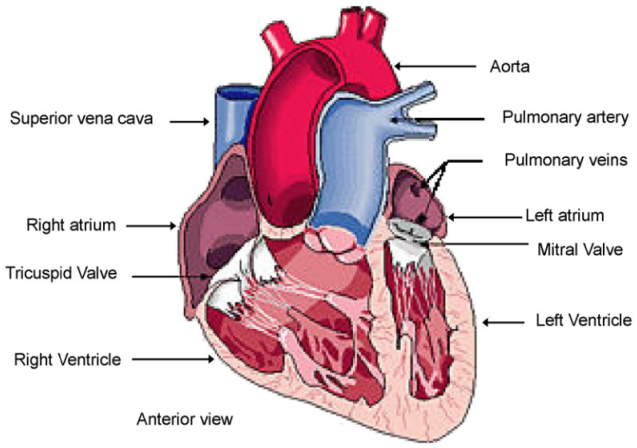

Figure 2.1: Section explanation of a normal human heart [1]

2.1.2 Cardiac Cycle and the First and Second Heart Sounds

Briefly introduced in the first paragraph of this section, the cardiac cycle can, in fact, be divided into two different time spans: systole and diastole. Systole corresponds to the ventricle's contraction period, while diastole corresponds to the heart relaxation period [10]. These periods repeatedly alternate, creating the cardiac cycle as we know it.

The beginning of diastole is marked by the closing of the aortic valve and its ending with the closing of the mitral valve, causing, due to the resulting vibration by the closure of atrioventricular valves, the first heart sound S1 [2], [10]. The systole starts when the mitral valve closes and ends when the aortic valve closes. Due to this vibration caused by the semilunar valves, the second heart sound S2 is produced at the beginning of diastole, concluding the full circle process [2], [10]. The presence of both S1 and S2 causes a *lub-dub* sound [3].

These heart sounds are composed of two components, each one for each valve that caused the vibration in the first place. For S1, the mitral and tricuspid components take place at a very similar point in time which causes differentiation between the two to be more difficult, despite the mitral component being slightly louder [2]. However, in S2, the aortic and pulmonary components can be more easily dismantled due to the earlier occurrence of the aortic component [2].

Both S1 and S2 heart sounds are not related to any pathologies and can be seen as boundaries between systole and diastole [3]. While systole occurs between the appearance of S1 and S2, by this order, diastole occurs after S2 and before S1 [3]. Both of these fundamental heart sounds are generally in the 20 to 200 Hz frequency range and have the highest amplitude values of the PCG signal analyzed [3]. Their standard duration can vary from 70 ms to 140 ms [11].

2.1.3 Gallop Rhythms

Following the explanation of how the S1 and S2 heart sounds are generated, it is also extremely important to mention gallop rhythms, also known as the S3 and S4 heart sounds.

These heart sounds may or not be present or represent signs of heart disease for a certain patient. The time location of both of these sounds is in the diastole phase. However, it is known that S3 appears after S2, while S4 appears before S1. These sounds are named "gallop rhythms" due to the fact that, instead of the *lub-dub* sound mentioned in the previous subsection, an additional sound can be heard, resulting in a sound sequence that is similar to a gallop - *lub-dub-ta* or *ta-lub-dub* [3]. S3's origin is thought to be the result of the blood flow inside the ventricle, which causes tension and the vibration of the chordae. It is commonly a synonym of a heart failure-related pathology when detected in patients other than children or young adults. When it comes to S4, the detection of the sound always relates to the presence of a particular pathological situation.

These gallop rhythms have frequencies between 15 and 65 Hz, low amplitude values, and duration from 40ms to 60ms [11].

2.1.4 Murmurs

In addition to the presence of the heart sounds that were previously explained, there are also additional heart sounds, such as murmurs, that can appear due to alterations in blood flow, for example, and that usually have a longer duration when compared to the latter [3]. Although murmurs can be benign, this being considered normal in children and during pregnancy periods, they can also be related to several pathological diseases, so, for that reason, it is essential to detect and analyze this behaviour as soon as possible [12]. Murmur description is based on several factors of the sound itself, such as timing, quality (intensity, pitch and profile) and location, as already mentioned in the paragraphs above. The remainder of the process would then consist on the analysis of these values for each entry in the dataset, their comparison to the known common values presented by each of the conditions and the making of a trustworthy decision that would ensure an early diagnosis of the respective patient.

2.1.5 Cardiac auscultation

Cardiac auscultation is one of the most used techniques to hear the heart sounds mentioned in the previous paragraphs. The data acquisition process used in the dataset analyzed in this dissertation was cardiac auscultation due to it being a more accessible and cheaper option when compared to other methods like the ECG (Electrocardiogram), for example. Besides being a well-known method, cardiac auscultation is probably one of the most challenging methods to master in heart sound extraction, requiring intensive training and experience to successfully identify and detect heart sound abnormalities.

When it comes to the ideal auscultation locations, as briefly mentioned in a previous section, it is known that [2]:

- The Aortic Valve is better listened to at the right second intercostal space;
- The Pulmonary Valve is better listened to at the left second intercostal space;
- The Tricuspid Valve is better listened to at the left lower border of the sternal;

- The Mitral Valve is better listened to at the fifth intercostal space or midclavicular line (cardiac apex);

A representation of these four different auscultation points can be seen in Figure 2.2.

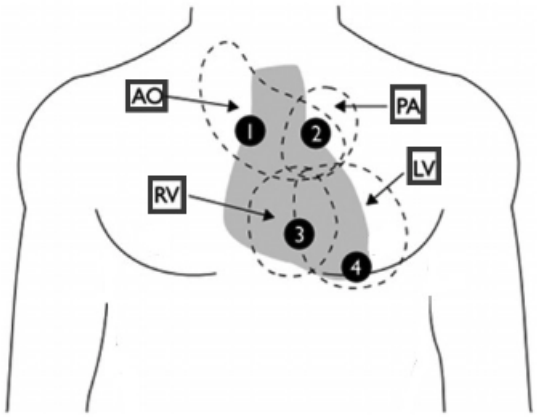

Figure 2.2: Optimal auscultation locations. AO = Aortic Area; RV = Right Ventricle; PA - Pulmonary Area; LV = Left Ventricle; 1 = Aortic Valve optimal auscultation point; 2 = Pulmonary Valve optimal auscultation point; 3 = Tricuspid Valve optimal auscultation point; 4 = Mitral Valve optimal auscultation point [2]

2.1.6 Common heart diseases

As mentioned above, the description of a certain murmur can lead to the discovery of a valve disease in a patient. In the list below, several common valve diseases will be explained, as well as the murmurs related to each one of them [2].

1. Aortic stenosis - Appearing commonly as a result of the calcification of the valves, aortic stenosis corresponds to the narrowing of the aortic valve. It creates a crescendo-decrescendo systolic murmur.

2. Aortic regurgitation - Aortic regurgitation can be described as the result of a bad closure of the aortic valve, which allows turbulent blood flow into the right ventricle. The originated murmur can be described as a decrescendo blowing murmur in the diastole.

3. Pulmonary stenosis - Consisting in the constriction of the pulmonary valve, pulmonary stenosis can be detected through auscultation due to the presence of a crescendo-decrescendo systolic ejection murmur.

4. Tricuspid stenosis - The narrowing of the tricuspid valve is what defines tricuspid stenosis. This type of valve disease originates a diastolic murmur, best heard at the left lower border of the sternal.

5. Tricuspid regurgitation - Tricuspid regurgitation's type of murmur can be described as a systolic one, and it can be described as a degeneration of the leaflets, which consequently allow blood to flow back into the right atrium.

6. Mitral stenosis - Mitral stenosis consists in the increase of difficulty for blood to flow between the left atrium and left ventricle due to the narrowing of the mitral valve, generating a diastolic murmur best heard at the cardiac apex.

7. Mitral regurgitation - Resulting in a cardiac apex best heard systolic murmur, mitral regurgitation represents the allowance of blood flow back to the left atrium.

8. Mitral valve prolapse - Mitral valve prolapse can occur when the leaflets of the mitral valve prolapse into the left atrium as the heart contracts. It typically originates an early systolic click followed by a systolic murmur.

9. Septal defects (Atrial and Ventricular) - Present from birth, septal defects are basically ruptures that allow blood to flow freely between the interatrial septum and the interventricular septum, meaning that we then have atrial septal defects and ventricular septal defects. The first one normally causes a wide and loud S1 heart sound while having a fixed split S2. On the other hand, the ventricular septal defect often corresponds to a holosystolic murmur.

10. Hypertrophic obstructive cardiomyopathy - As its name indicates, hypertrophic obstructive cardiomyopathy consists of a myocardial condition in which the myocardium undergoes several increases in the size of cells and tissues. When auscultating, it presents a systolic ejection murmur, being this component best heard between the cardiac apex and the left sternal border.

11. Patent ductus arteriosus - Patent ductus arteriosus consists of an additional blood vessel found in newborns. When this channel fails to close after birth, it allows blood located in the aorta to flow back to the lungs through the pulmonary artery, which creates a constant "machine-like" murmur possible to be heard through cardiac auscultation.

Following this, in Figure 2.3 we can see several examples of sound waves associated to some of the valve diseases mentioned previously.

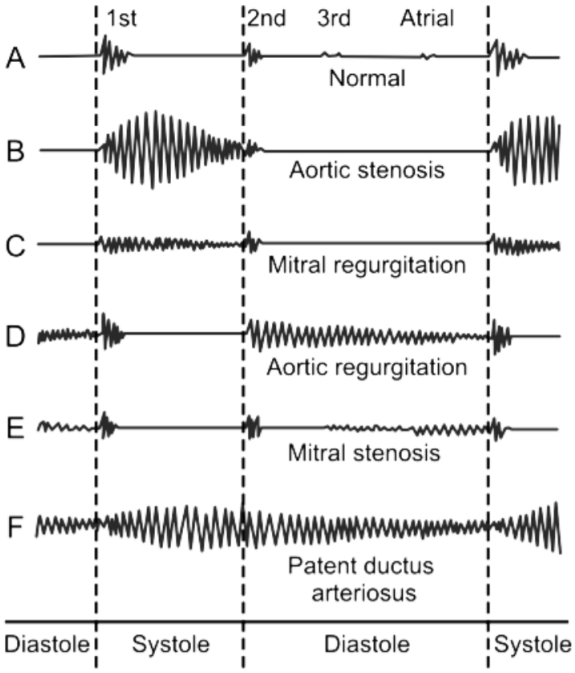

Figure 2.3: Examples of a normal cardiac cycle and systolic, diastolic and continuous murmurs and their respective common heart diseases [3]

Chapter 3

State of the art

The main focus of this chapter relies on the presentation of the dissertation's state of the art, therefore, it will discuss and present information found in other reports or studies that may play an essential role in the development of the proposed project. A heart sound processing chain diagram will also be exposed and decomposed into several different stages, whose joint functioning is required to reach the project's end goal.

3.1 Heart Sound Processing Chain

In order to be able to accomplish a ponderable and sensible decision when it comes to the early prediction of heart diseases, the heart sound processing chain must be thoroughly followed. There are generally three main stages in total:

1. Denoising and Signal Enhancement: responsible for removing and attenuating background noise and other organ sounds from the analyzed heart sound audio signal. It eliminates certain unimportant ranges of low or high frequencies and also reconstructs the samples in the amplitude domain, preparing the audio for the remaining stages.

2. Segmentation: responsible for the division into cardiac cycles through the analysis of the audio in question after filtering. In this stage, the localization of each cardiac cycle and its respective content are extracted. As mentioned in the background section, each cardiac cycle at its base is composed by a systolic sound (S1) followed by a diastolic sound (S2).

3. Detection, Feature Extraction and Classification: responsible for detecting the presence of murmurs in the analyzed audio and extracting certain features of the latter while also doing their respective classification. This classification procedure will focus on a pitch perspective, such as fundamental frequency analysis, to classify the audio and try to understand the relevancy of this characteristic when it comes to predicting the appearance of valve diseases in a particular patient.

In Figure 3.1, we can see the three mentioned steps of the processing chain that have been briefly described in the previous paragraphs:

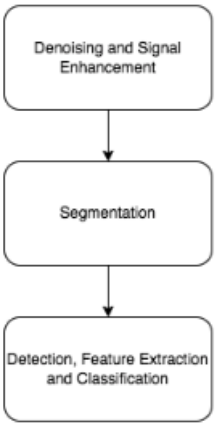

Figure 3.1: Diagram regarding the processing chain for a heart sound

Following the brief presentation of the process, the following subsections will now target each of the stages, explaining them in detail and describing several of the already existing methods found through the reading of references and articles on the proposed problem.

3.1.1 Denoising and Signal Enhancement

The correct detection or prediction of heart valve diseases consists in the difficult to achieve climax resulting from the correct function of several different stages. Firstly, to correctly set up the subsequent phases, we need to de-noise and enhance the provided audio signal. Due to the process and situation involving the extraction of heart sound audio signals, they tend to have several relevant noisy components such as chest movements; breathing sounds or overall organ sounds; shear noises, which consist of the noise created by the rubbing of the recording device on the patient's skin; or even external ambience noise [13]. There is a need to eliminate or reduce the presence of this unwanted information so that the following stages' success rate can, theoretically, be increased.

Following the investigation and reading of several research papers, it can be noted that several different methods may be used in order to successfully denoise heart sound audio signals.

Firstly, and probably the most used method is the application of digital filters in order to eliminate these unwanted sounds. Either being low-pass, high-pass or band-pass filters (these being normally IIR - Infinite Impulse Response - or FIR - Finite Impulse Response), the main objective relies on eliminating specific ranges of frequencies from the frequency spectrum of the analyzed audio [13]. Although the filter cutoff frequency values are empirically determined, several heart sound frequency ranges provided in certain research papers should be considered since there are certain types of murmurs whose frequencies tend to be much higher than the normal frequencies for the S1 and S2 heart sounds. It is known that the significant energy concentration for the S1 and S2 heart sounds is generally below 150 Hz. However, the filtering tends to have ranges from 50 Hz to, in some cases, 1000 Hz in order to cover the high-frequency murmur values. For example,

in [14] a band-pass zero-phase filter with cutoff frequencies of 50-800 Hz is used to remove the unwanted noise, while in [15] a band-pass filter between 20 and 1000 Hz is formed by applying a low-pass filter and a high-pass filter to each of the respective frequencies. Moreover, in [16], a fourth-order Butterworth band-pass filter between 50 Hz and 950 Hz is also applied.

Another commonly used method consists of the application of Wavelet transforms. This transform was developed in order to obtain a simultaneous high resolution both in time and frequency domain [4], being that it can present detailed information on this type of content from the signal analyzed. The two main types of wavelet transforms are: discrete and continuous. For example, DWT (Discrete Wavelet Transform), which can be described as a tree composed of a succession of low and high pass filters, as seen in Figure 3.2, was used by [4] in order to decompose the primary signal into several frequency components so that it could be more efficiently represented by having fewer parameters and less computation time.

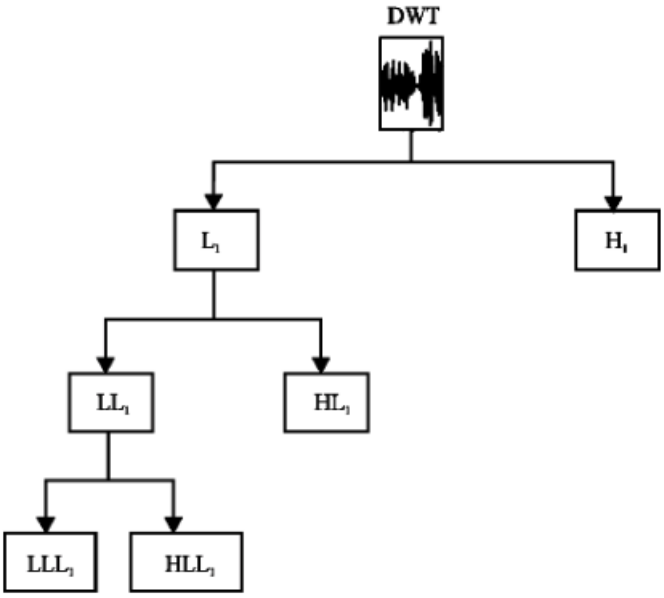

Figure 3.2: Representation of filter tree regarding the application of the Discrete Wavelet Transform [4]

According to [17], the calculated DWT coefficients when it comes to noise are much smaller than the actual heart sound coefficients, meaning that it is possible to eliminate these noisy components if we only retain the significant calculated coefficients.

In [16] we can see a different approach to the latter by using STFT - Short Time Fourier Transform - to pre-process the noisy parts of the audio signal by flagging different segments where this noise seems to be more relevant.

At last, in [5] we can see the usage of an adaptive filtering method to remove unwanted noise in the audio signal, similar to the one shown in Figure 3.3. This method is presented as a better alternative than the simple band-pass filtering since, when using band-pass filters, several frequency bands of the useful heart sound signal may overlap with the noise bands, resulting in worse quality

than the first ones [5]. It also mentions that the useful frequency bands also tend to have noisy components. Therefore, the noise cancellation process should improve overall by using adaptive filtering instead of band-pass filtering.

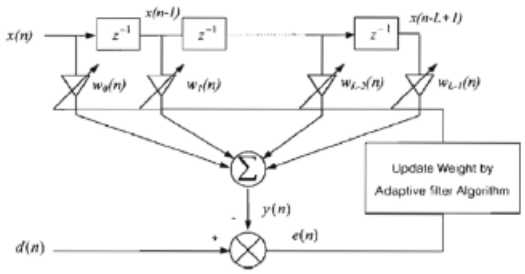

Figure 3.3: Adaptive Noise Cancellation Filter Structure [5]

3.1.2 Segmentation

The common second stage of the processing chain consists of segmenting the heart sound audio signal. The Segmentation stage represents the usage of a certain method to find the location of several sound peaks in order to flag where the fundamental heart sounds appear in the audio signal. This audio segmentation still stands as a relevant component in order to be able to make a correct diagnosis regarding the presence of cardiac valve diseases in an examined patient since, by discovering them, we can divide the heart sounds into systolic and diastolic regions.

As in the previous topic, after examining several research papers, it is possible to conclude that there are multiple types of methods when it comes to this stage.

The usage of an envelope analysis method stands as one of the preferred methods in several research papers. This method is used to identify the location of the fundamental heart sound's mentioned, both S1 and S2, through the application of different approaches such as the Shannon energy envelope, the Hilbert transform or the Stockwell Transform.

In [6], for example, we can see a detailed explanation of how to identify S1 and S2 peaks using the Shannon energy envelope (Figure 3.4) through a comparison of the value for a particular audio peak with a threshold within a defined window. If this value is larger than the defined threshold value, then this value is a peak. By discovering the S1 and S2 peaks in the sample sequence, we can identify the systole period, which happens between S1 and S2 peaks, and a diastole period, which happens between S2 and S1 peaks, respectively [6].

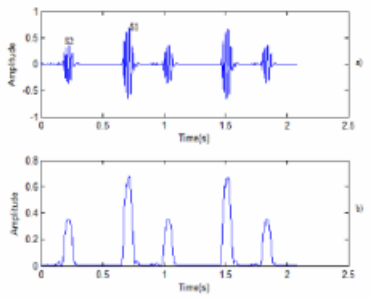

Figure 3.4: Shannon Energy Envelope on a heart sound audio signal [6]

Other studies rely on the usage of time-frequency analysis methods such as the Fourier Transform or the Wavelet Transform. In [18], the authors present the benefits of using EWT (Empirical Wavelet Transform) as a segmentation method, for example. This transform is described to estimate the frequency components of a specific signal and filter the respective audio based on the resulting spectral boundaries. Moreover, [4] presents another segmentation approach using DWT (Discrete Wavelet Transform).

The usage of the Fourier Transform can be seen in [19] where an FFT (Fast Fourier Transform) based method is applied to analyze the audio signals' frequency spectrum peaks so that location predictions of the fundamental heart sounds can be made.

Several research papers like [13] use Markov Models as their primary segmentation method and are able to acquire successful results in a noisy environment analysis.

Despite still being used regularly by several researchers in the community, these methods that comprise the segmentation process will most likely become less common due to the introduction of a deep learning approach. For example, in [14], the segmentation stage that generally appears between the pre-processing and the feature extraction stages is no longer a part of the processing chain due to the fact that, when using a deep learning approach like the one presented, it is not needed.

3.1.3 Detection, Feature Extraction and Classification

Following the bibliography review of the pre-processing and segmentation stages, this final stage appears as a more complex one since it englobes three different tasks that, in an optimal scenario, can all be executed in the same component or block. These three different tasks are detection, feature extraction (which for this dissertation will be focused on the audio signal's pitch characteristic) and, finally, classification.

The process starts with detecting any murmurs present in the analyzed heart sound. Through this analysis, a filtering of the patients can be executed so that the studied population is divided into two big groups: patients with normal heart sounds and patients with abnormal heart sounds. By detecting and dividing the samples into these two groups, we can set aside certain patients from the beginning and classify them as unproblematic since there are no signs of eventual cardiac valve diseases.

After this division is settled, feature extraction methods should be used to extract several characteristics of the detected murmurs. In this thesis, this extraction will base itself on a frequency analysis since our main focus is to study the importance of the pitch values in several heart murmurs and how they can help predict cardiac valve diseases.

Now that these features have been extracted, we then enter the final stage of the processing chain, which is the classification of the respective sounds. This classification method is responsible for dividing each of the analyzed signals into different clusters, classifying them according to whatever rules or methods are used.

Despite preferring to join these three stages into a unique one, the bibliography review for the Feature Extraction, Detection and Classification will be separated into different sub-sections so that each of the methods presented is easier to comprehend, not only when it comes to how it is used but also to what specific task they are applied to.

Pitch-Oriented Feature Extraction

Since the central theme of this project revolves around the pitch of a specific heart sound, the bibliography review for the feature extraction stage will focus on discovered attempts at analyzing several frequency components that may be present in the audio signal. Firstly, we can present the concept of heart sound spectral display, or heart sound spectrography. Initially presented in [20] in 1955, the heart sound spectral display adds the frequency or pitch dimension to the standard PCG signal display. Heart sound spectrography has been considered a growing approach by the research community, resulting in the exhibition of several different methods by multiple published articles.

Firstly, it is possible to verify a typical usage of Fourier Transforms so that the audio signal's spectrogram can be obtained. As known, this acquired spectrogram basically consists of a visual representation of the spectrum of frequencies of the analyzed audio signal as it progresses over time. Adding to this, its importance is also elevated since it offers additional insight into the murmur frequency changes over time. For example, in [21] we can verify the usage of the FFT in order to extract the pitch from the pre-processed audio. Moreover, the STFT was used in [22] so that the respective spectrogram could be extracted prior to scaling it for later comparison, as described in the paper.

However, as mentioned in [23], the usage of the Fourier Transform brings a problem to the table - the "quantum uncertainty theorem". According to the reference, this theorem states that a certain signal and its correspondent Fourier Transform can not both have small support and that neither frequency nor time can be determined to arbitrary precision, which basically results in a non-unique and low fidelity image. Also, because heart sounds are not linear, exponential or sinusoidal, it has been proved that the corresponding usage of Fourier Transforms is not effective [23].

Considering this, there has been a growing usage of Wavelet Transforms over Fourier Transforms since the first ones tend to have a faster computation time and better frequency resolution results [23]. In [24], for example, the authors chose to use the CWT (Continuous Wavelet

Transform) for each of their segmented intervals. By using this transform, a scaleogram (visual Representation of a wavelet transform [1] is obtained.

Another interesting approach can be seen in [23] where the authors present a method based on the Wigner-Ville distribution, which they call HES (Heart Energy Signature). This Heart Energy Signature basically consists of a high-resolution two-dimensional image of the correspondent heart sound audio signal.

Classification

The last stage of the presented processing chain is classification, which consists in the phase where the features extracted are fed to a certain chosen classifier so that the latter can organize the data into different chosen categories. Despite verifying that there has been an everyday use of several machine learning methods over the years, the specific interest in deep learning has been rising due to its several advantages when compared to the latter.

Artificial Neural Network (ANN) is widely known as one of the most commonly used methods for heart sound classification [13]. Some of the advantages of using ANN's consist in their structure's simplicity for physical implementation and how easily they can map complex class distributions, for example. In [25], after using Wavelet transforms to segment the heart sound audio signal, the authors use a Grow and Learn (GAL) network and a Linear Vector Quantization (LVQ) network in order to be able to classify seven different heart sounds.

Another classifier that was presented in several research papers was K-Nearest Neighbours. For example, in [26], a K-Nearest Neighbours classifier was used to evaluate several different spectral features and how they would contribute to the correct detection of the presence of heart murmurs.

SVM, or Support Vector Machine, consists of another successful classification method used, for example, in [22]. Following the extraction and scaling of the spectrograms mentioned in the paper, the authors mention that the SVM classifier would be the suitable choice, ignoring the decision tree based method option, since the scaled spectrogram features are difficult to extract when using the latter method compared to when using SVM's.

As mentioned in the last paragraph, due to the situation of the reference presented, decision trees are also another classifier that is used in several research papers. A decision tree was constructed in [27] by using a featured training dataset and a certain classifying attribute. The results showed that decision trees could help young or inexperienced doctors by providing them with a highly successful decision support system.

Already briefly introduced in the previous sections, Hidden Markov Models can also be used in the classification phase. For example, a Discrete Hidden Markov Model (DHMM) was used in [28] after reducing the audio signals' features acquired by a DFT by using Principal Component Analysis (PCA).

[1] https://handwiki.org/wiki/Scaleogram

Finally, when it comes to deep learning, approaches are mainly divided into Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and hybrid networks.

For example, in [29], the authors use a 2D-CNN network, whose input features are spectrograms and Mel-frequency cepstrum coefficients, and obtain positive results with a final F1-score of 0.885. Moreover, in [30], the usage of a modified AlexNet model having spectrograms as the main input feature has an accuracy of about 97.05% on the PhysioNet/CinC Challenge 2016 Dataset. The usage of several features can also be seen in [31], as the input feature vector corresponds to a concatenation of the respective audio's chromagram, spectral centroid, spectral bandwidth, zero crossing rate and MFCC's. Using these features and a network with several consecutive Dense layers, the author reaches an accuracy value of about 98%.

RNN methods are also present in research papers like [32] and [14]. [32] makes several attempts through the usage of a Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (BLSTM), a Gated Recurrent Unit (GRU) and a Bidirectional Gated Recurrent Unit (BiGRU) with the first of the latter achieving an impressive result of about 99.95%. [14] not only uses an RNN but compares its results to those of Decision Tree and Random forest methods for heart disease classification, with the RNN surpassing the latter methods by about 15%.

Another method that has been steadily becoming more common in several heart sound classification problems consists in the usage of pre-trained models either during the feature extraction stage or the final classification stage. This process is called Transfer Learning. In [33], for example, the authors used three transfer learning models - AlexNet, VGG16 and VGG19 - to extract features from spectrograms. These features will then be concatenated and inputted onto an SVM classifier to divide the sounds into four classes: Normal, Murmur, Extra sound and Artifact. All results are then compared for each of the models and even combinations of the three models, with accuracies increasing compared to other research papers. [34] uses three other transfer learning models on the classification stage - Resnet50, VGG16 and VGG19 - with spectrograms as the feature input for binary classification.

Chapter 4

Methodology

In this chapter, not only will we conduct a detailed presentation of the dataset used, the CirCor DigiScope Dataset, but also a general overview of the techniques and methods used for each section of the overall project. We will first explain the data pre-processing methods, followed by their respective feature extraction and normalization components. Subsequently, we will explain the neural network preparation, including the training, testing, and validation divisions, its correspondent input sections, and, finally, the decision-making function that provides the patient with its correspondent and final prescription.

4.1 CirCor DigiScope Dataset

The CirCor DigiScope Dataset is the result of two mass screening campaigns that occured in 2014 and 2015 in the Paraíba and Pernambuco states, Brazil. Referred to as "Caravana do Coração", these campaigns consisted of the performance of heart screenings and specific exams by cardiologists on several participants from the respective population of the previously mentioned states in order to discover and study the presence of heart pathologies.

After volunteering to be a part of the campaign, all participants completed a socio-demographic questionnaire and also underwent clinical examination (anamnesis and physical), nursing assessment (physiological measurements), and cardiac investigation (electrocardiogram, echocardiogram and chest radiography) [2]. It is also important to mention that all participants under 21 years of age required a signed consent form in order to participate. Following the previously mentioned procedures, researchers executed a data quality assessment in order to filter any incorrectly entered or measured values, inconsistent data, or outliers.

Following this, electronic auscultation was performed for each patient in order to extract audio samples from the four typical auscultation points - AV, PV, MV and TV - while also having two independent cardiac physiologists test the extracted audio for signal quality to guarantee that all audio data allows for a safe and trustworthy murmur characterization and description. The latter process resulted in removing data from 119 participants due to not reaching the required signal quality standards. [2]

The collected dataset includes 1,568 participants, spread over the two screening campaigns, with about 116 patients having attended the "Caravana do Coração" both years. This population size resulted in a total number of 215,780 heart sounds. In Table 4.1, we can see the number of recordings that were collected for each auscultation point for each one of the campaigns:

Table 4.1: Number of audio recordings extracted for each auscultation point, during each of the screening campaigns [2].

	CC2014	**CC2015**
AV	540	817
MV	603	812
PV	497	793
TV	461	754
Unreported	5	1
Total	**2,106**	**3,177**
Average	**3.2 per patient**	**3.5 per patient**

As seen in Table 4.1, we have a total of 2,106 and 3,177 recordings for each campaign, respectively, which amounts to an absolute number of 5,283 recordings, with a mean of 3.2 recordings per patient in CC2014 and 3.5 recordings per patient in CC2015.

Researchers gathered the recordings using an electronic stethoscope and later sampled the signal at 4kHz with a 16-bit resolution while also normalizing it within the [-1;1] range. [2] The average recording length for each campaign stood at 28.7 seconds for CC2014 and 19.0 seconds for CC2015.

The dataset contains information regarding the heartbeat of 1,568 patients, being that 1,144 (73.0%) possess a normal heartbeat, while in 305 (19.5%) of the cases, there is a presence of a heart murmur. The remaining 119 patients (7.6%) can be identified as low quality.

In the presence of murmurs, each case was classified according to several labels such as:

- **Murmur Timing** - Early, Mid and Late Systolic/Diastolic, Holosystolic and Holodiastolic.
- **Murmur Pitch** - Low, Mid and High.
- **Murmur Grading** - I/VI, II/VI, III/VI (Systolic) and I/IV, II/IV, III/IV (Diastolic)
- **Murmur Shape** - Crescendo, Descendo, Diamond and Plateau.
- **Murmur Quality** - Blowing, Harsh and Musical.
- **Highest Murmur Intensity Location and Murmur per location** - AV, MV, PV and TV.

4.2 General Overview

4.2.1 Pre-processing

The first stage of the procedures taken during this work relied on the pre-processing of the audio signals provided in the CirCor Digiscope Dataset in order to better prepare the data that will be input to the next stages of the system.

4.2.1.1 Segment division

As mentioned when presenting the CirCor DigiScope Dataset, there are 2,106 heart sound recordings for CC2014 and 3,177 for CC2015, summing up to a total of 5,283 recordings, of various lengths each, to analyze. We know that, when training neural networks, the bigger the data provided, the better the results may be. With this in mind, a decision was made not to use the recordings in their original form, with their respective lengths, but to divide them into 2-second segments. The primary purpose of this division stands on the previously mentioned point, the augmentation of data for analysis, while still having segments large enough to correctly analyze the presence of murmurs, due to their periodic behaviour. Instead of having a 20-second heart sound recording associated with a respective pitch label, we will then have ten heart sound recording segments, each of them associated with the pitch label of the original recording. Due to this, the number of audio samples increased substantially, resulting in about 14,543 heart sound segments relative to the CC2014 campaign and 15,239 heart sounds segments relative to the CC2015 campaign.

It is also relevant to mention that, since the audio cropping process started at the beginning of the audio and most of the recordings' duration value did not represent a possible divisible by 2 integer, several resulting ending segments under 2 seconds (scraps of audio) were discarded due to not having the exact length required. An example illustrating this division can be seen in Figure 4.1.

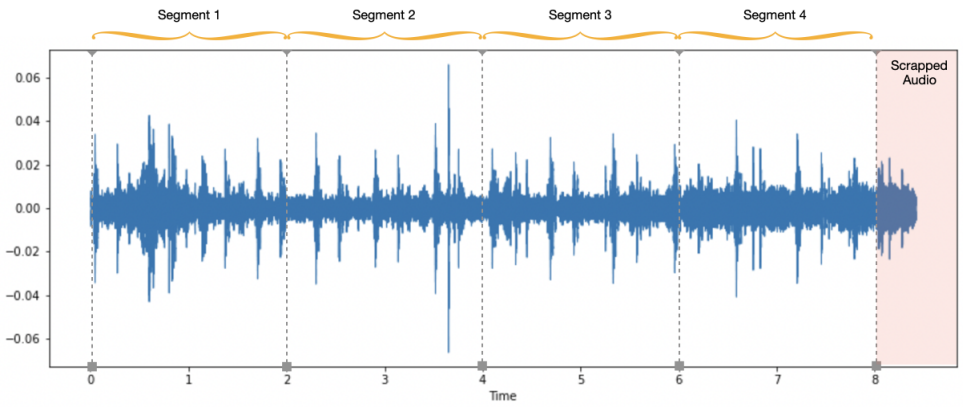

Figure 4.1: Illustration of the segment division procedure on the 49577_AV audio file from the CC2014 campaign.

4.2.1.2 Audio filtering

Following the division mentioned in the previous subsection, we advanced to the filtering stage, which is commonly used and much needed in heart murmur-related works. Heart murmur recordings, as expected, contain a relatively high amount of noise that can, and should, be discarded to increase the success when it comes to extracting features [3]. Although humans can hear from about 20Hz to 20000Hz, most heart sounds appear in the 20-500 Hz range, being therefore irrelevant to keep the frequencies above 500 Hz present in the signal [35].

As mentioned before, the focus of this dissertation revolved around making a pitch analysis of the heart sounds provided. As [35] said, the frequency/pitch of heart sounds or murmurs can be divided into three categories: low frequency, medium frequency, and high frequency. Low-frequency heart sounds are present in the 20 to 100 Hz range, while high-frequency heart sounds appear from the 400 to 500 Hz range. The medium frequency band consists of the remaining frequencies, from 100 Hz to 400 Hz [35].

With this in mind, the audio filtering stage consisted in the application of three different filters to each of the audio signals present in the dataset:

- Low-Frequency range filter - from 20 to 100 Hz;
- Medium-Frequency range filter - from 100 to 400 Hz;
- High-Frequency range filter - from 400 to 500 Hz;

The filters tested consisted of third-order Butterworth filters, respectively associated with the cut-off frequencies mentioned above. After applying these filters, we could then advance to the next phase: feature extraction and normalization.

4.2.2 Feature extraction and normalization

Having finished every component of the pre-processing stage, we advanced to the feature extraction and normalization stage. Feature extraction consists in the process of taking heart sound audio signals and converting them into low-dimensional features to allow and facilitate the further analysis that will be executed in the neural network stage of the process [36].

For this dissertation, the chosen feature was the spectrogram. This feature seemed ideal for the problem presented since spectrograms allow users to analyze the loudness of an audio signal, for each of its respective frequencies, throughout its entire length. A spectrogram was extracted for each of the filtered 2-second segments of data, meaning that the final result per 2-second segment consists of three distinct spectrograms, one for each of the frequency bands. Each of the extracted spectrograms' shapes equaled to (129, 63), which was then reshaped into (129, 63, 1) so that the three spectrogram stacking could be correctly done. After stacking, we cropped the output, removing non-present, thus irrelevant, frequencies from the image. We also vertically flipped the plot so that the low frequencies appeared at the bottom end of the result. The extracted heart sound

spectra resulted from the usage of the STFT - Short Time Fourier Transform - through *librosa*, a Python-based library which is heavily used when it comes to reading, writing and analyzing audio.

Each of the spectrograms was also normalized by using the z-score normalization method, as seen below:

$$z = \frac{x_i - \mu}{\sigma}$$

Where:

- x_i represents the original value.
- μ represents the mean.
- σ represents the standard deviation.
- z represents the normalized value.

The z-score normalization consists of the process of normalizing all values presented such that the resulting mean and standard deviation equal 0 and 1, respectively. Normalized values (z) represent the number of standard deviations that the original value distances from the mean. This process can play a relevant part in fitting the neural network since it reduces the impact of outliers for the presented dataset.

Lastly, we exported the final output as a PNG file, resulting in a final data shape of (68, 63, 3). These values respectively relate to the height, width and depth of the correspondent file. The specific depth value equal to 3 relates to an RGB image, ready to be imported and used as the input of the Neural Network in the next stage of the system. An example of an exported PNG file can be seen in Figure 4.2.

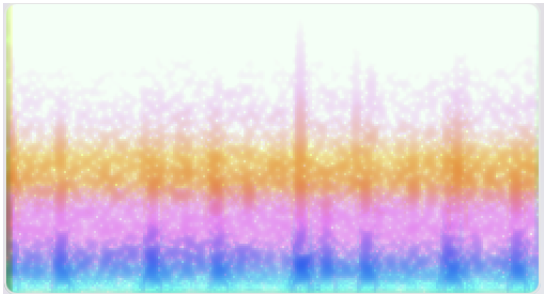

Figure 4.2: Example of the feature extraction output for a segment of the 49577_TV audio file from the CC2014 campaign.

As we can see in Figure 4.2, there are several colors for each of the filtered bands, since it is an RGB image. These colors will relate to different pixel values, helping the model when it comes to the frequency band analysis and pitch prediction.

4.2.3 Neural Network

4.2.3.1 Data preparation

After completing all pre-processing and feature extraction methods, the main focus turned to the actual development of the neural network. Firstly, before inputting the data onto the neural network, the needed data was extracted from the two . CSV files available, CC2014 and CC2015, and inserted into a pandas Dataframe. From the entire file, we imported the corresponding columns:

- `Patient_ID`: the ID's of patients;
- `Murmurs`: equals 1 if the respective patient has been associated with a heart murmur, 2 if a patient has a low-quality audio signal, or 3 if no heart murmur has been detected;
- `Murmur1_Pitch`: presents the pitch value of the systolic murmur, if present, NaN if no heart murmur has been detected, or 0 if the patient has a low-quality audio signal;
- `Murmur1_Grading`: presents the grading of the systolic murmur, if present, NaN if no heart murmur has been detected, or 0 if the patient has a low-quality audio signal;

Having these data points organized in a pandas Dataframe, we could separate our primary X values, the `Patient_ID` column, which will be associated with each patient's respective extracted features for each of the divided segments, and also our y values, the `Murmur1_Pitch` column.

Additionally, the `Murmurs` column is used to divide the patients' IDs' into three different sets, relating to the three value possibilities mentioned in the previous list, for the X and y variables: *murmursPresent*, *murmursAbsent* and *murmursOthers*, respectively. This division represents a significant step prior to the train, test and validation splits since the latter process always needs to be balanced in order not to, for example, risk the probability of having fewer patients with present heart murmurs in a particular set when compared to another. The labels arrays were also One-Hot encoded, converting all values from integers to binary.

Following this, the process firstly consisted of the division of each one of these by an 80/20 (train/test) ratio using the `train_test_split` function present in the *sklearn* library, creating the final test set and a temporary train set. Secondly, this temporary train set was divided once more by an 80/20 ratio in order to create the final training and validation sets. Following this division, the file paths regarding each of the segments for the respective patient's ID were retrieved and passed through the `im_read` function of the *cv2* library (a commonly used library for image-based deep learning problems). The pixel values were inserted in three different arrays, one for each set, while the pitch value for each of the segments was inserted in a labels array, each segment with an associated value equal to the respective patient's value.

It is also important to mention that throughout the project's development, to avoid running the entire code from the beginning, although necessary at times, the *pickle* library was used to store the train, test, and validation sets easily.

After organizing and dividing all features and labels into their respective sets, another normalization was executed by taking each feature value inside the array, that initially corresponded to an image pixel value and ranged from 0 to 255, and normalizing it by dividing it by 255, in order to change all values from their initial range, to a range of 0 to 1.

Dealing with class imbalance also revealed to play an essential factor before inputting data to the neural network [1]. When the data was extracted from the .CSV file, the difference between the number of patients with a normal heart condition versus patients with heart murmurs or a low-quality audio signal was reasonably significant, with a total of 1,167 out of 1,568 patients having no heart murmurs whatsoever. To deal with this imbalance between classes while training the network, each of the class weights was computed using the `compute_class_weight` function present in the *sklearn.utils* library. This function calculates the influence of each class in the complete training set, ensuring that all classes are equally important when training, regardless of the number of samples present. The latter prevents the model from only predicting the most frequent class due to it being more common in the set. This function, when in the selected 'balanced' mode, uses the following formula to return each of the classes' weights:

$$weight = \frac{n\_samples}{n\_classes * n\_samples\_class}$$

Where,

- $n\_samples$ represents the total number of samples in the training set;
- $n\_classes$ represents the number of classes.
- $n\_samples\_class$ represents the total number of samples for the respective class.
- $weight$ represents the weight for the respective class.

As seen in the image above, the more samples are present for a specific class, the lower its weight will be, therefore balancing the set. Having finished this stage, it was now possible to advance and define the actual model that will be used to train and test the data.

4.2.3.2 Transfer Learning

Instead of training the model from scratch through supervised learning, as commonly done for several other projects, this dissertation was marked by the usage of transfer learning. Transfer learning consists in using a pre-trained model, that was initially trained for other problems and has its respective pre-trained weights and biases, on a different problem. By doing so, the lack of enough data that regularly affects the accuracy of machine learning models during training can be overlooked since we are acquiring and transferring knowledge between similar experiments in order to potentiate the outcome of one another. However, it is also important to mention that

[1] https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/class-weights

the transferred models should not be solely used since it is not likely for them to learn any new information if in that situation. Therefore, it is crucial to add new layers to the model [34].

In this study, three different models have been used and tested: Resnet50, VGG16 and VGG19. Resnet50, short for Residual Networks[34], consists of a convolutional neural network with a depth of 50 layers [2] that is able to successfully train deep neural networks due to its feature of skipping layers [37]. This latter feature represents the main strength of a Resnet model since, by doing so, it prevents overfitting and solves the vanishing gradient problem while also helping higher layers perform at a similar level as lower layers [34]. The VGG16 model, developed at the University of Oxford [38], supports 16 convolutional layers and replaces the large-kernel sized filters present in several other models with successive smaller-kernel sized 3x3 filters [38]. Lastly, the VGG19 represents an improvement on the VGG16 model since it increases its convolutional layer count to 19 [38]. All models have been trained on more than one million high-resolution images, being able to classify about 1000 classes from the Imagenet database.

Despite already having reports that the Resnet50 model appears to translate in a higher accuracy in binary heart murmur detection problems when compared to the two other pre-trained models VGG16 or VGG19 as seen in [34], all three models were still tested since the current classification problem is multi-class and not binary.

After transferring the models to the project through the use of the `applications` library in *keras*, all models were then connected to a Flatten layer, collapsing the input dimensions into one dimension, and repeatedly passed through alternating BatchNormalization and Dense layers with "relu" activation function and halving descending units from 1024 to 16. Finally, the models were then connected to the final classification layer composed of five units, corresponding to the five classes present, with a "softmax" activation function. After defining each model, we could then advance to its compilation, where an "adam" optimizer was passed as an argument for several possible learning rates, and also to the definition of two specific callbacks to better the model's training process: ModelCheckpoint, whose primary purpose was to save checkpoints of the model within a certain epoch number period, if any improvement was to be seen, and also EarlyStopping monitoring either the validation accuracy or the validation loss, in order to prevent overtraining the model. These callbacks were then passed onto the fit command, along with the chosen batch size, validation_data, epoch number, and the previously computed class weights.

To monitor the model's performance while training, the "accuracy" and "loss" metrics are used in the training and validation sets so that we can evaluate the training procedure and be aware of any overfitting signs. Overfitting consists of a common phenomenon that happens in the training of neural networks where the model starts trying to fit the training data entirely by memorizing patterns in the set. By doing so, the model then fails to generalize and perform well in unseen data, such as the validation and test sets [39]. A common symptom of overfitting consists of, for example, the gradual increase of the validation loss value while the training loss keeps decreasing. The previously mentioned EarlyStopping callback function works to aid this problem by stopping the training early in order to try to prevent this event from happening. As this study's problem

[2] https://www.mathworks.com/help/deeplearning/ref/resnet50.html

is multi-class, the loss function chosen was "categorical_crossentropy", which expects a one-hot encoded label format, and can briefly be described as the sum of errors made by the model when comparing the predicted value to the true value [3][34]. The formula for this loss function can be seen below:

$$loss = -\sum_{c=1}^{M} y_{o,c} log(p_{o,c})$$

Where,

- M represents the number of classes.
- $y_{o,c}$ represents a binary number (0 or 1) if the class label c is the correct classification for the observation o
- $p_{p,c}$ represents the predicted probability that the observation o is of class c
- $loss$ represents the final calculated loss;

Lastly, when the model finalized its training stage, the test features and labels were then passed in order to obtain a predictions array for the mentioned model. It is, however, important to mention that each of these predictions corresponds to an individual segment from a particular patient, meaning that these results are not final. We then need to organize the segment predictions by patient and input the results into a decision-making algorithm to provide the final diagnostic to the studied patients.

4.2.4 Final decision-making function

In order to organize the segment predictions outputted by the previously described model, an array containing all file directories per patient was saved. Since the number of directories per patient equals the number of segments per patient, both the predictions and the true values were divided into arrays, one for each patient, in order to facilitate the comparison between the two.

After the described process, several algorithms responsible for collecting these segment prediction per patient arrays and outputting a single final diagnostic were created:

- **BiggestCount:**

  Final diagnostic equals the class with the highest number of predictions in the array.

- **OneTimeAppearance:**

  Checks for the presence of a certain class and, if present, returns that class as final diagnostic, if not, moves to another class and repeats the same logic in a "High", "Medium", "Low", "LowQuality" and "Absent" order. If there are any "High" predictions, then the final diagnostic is "High". If not, and there are "Medium" predictions, then the final diagnostic is "Medium", etc.

[3] https://gombru.github.io/2018/05/23/cross_entropy_loss

- **AbsentPercentage:**

  As the Absent segments percentage is higher than the rest of the classes, this algorithm inserts a probability in order to filter that specific class. If 60% or more of the entire array consists of "Absent" predictions, then the final diagnostic equals "Absent". If not, final diagnostic equals the class with the highest number of predictions.

- **AbsentPercentageModified:**

  Since the lower the frequencies in a certain class, the harder to differ between an "Absent" diagnostic, with normal heart sounds, and an abnormal diagnostic, this algorithm inserts a probability for the "Absent" class, as the prior, while increasing value for some classes when compared to others, trying to fight against any wrongly defined segments. The "Low" segment count is multiplied by three, while the "Medium" segment count is multiplied by two. The logic follows the AbsentPercentage algorithm mentioned previously.

All of the above functions were executed and all results were collected in order to advance to the final stage of the process which consists in evaluating the performance of the entire system.

4.3 Performance Evaluation

There are several metrics available to correctly evaluate the performance of a certain machine learning model. The values of these respective metrics refer to how good or how bad the classification process of a certain model is, through the comparison between the predicted values and ground truth [40]. In this section all of the used metrics will be presented, starting with the confusion matrix.

4.3.1 Confusion Matrix

The confusion matrix, despite not being an actual metric, is fundamental to the calculation of the remaining metrics [40]. This matrix is a tabular visualization of the ground-truth labels, which are assigned to each row, versus the predicted values, assigned to each of the present columns [40]. Each of the matrix's cells represents a certain evaluation factor:

- **True Positive (TP)**: represents how many positive samples the model has predicted correctly;
- **True Negative (TN)**: represents how many negative samples the model has predicted correctly;
- **False Positive (FP)**: represents how many negative samples the model has predicted as positive, therefore an incorrect prediction;
- **False Negative (FN)**: represents how many positive samples the model has predicted as negative, therefore an incorrect prediction;

With these definitions in mind, a binary class confusion matrix would look similar to the one in Table 4.2:

Table 4.2: Example confusion matrix for binary classification

		Predicted values	
		Yes	No
True values	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

However, due to the fact that this study's problem is multi-class, there are not any positive or negative classes. This consequently means that, to obtain the previously mentioned values, a calculation for each of the individual classes has to be executed. By doing so, we then obtain all resulting values that can be used to calculate the remaining metrics presented in the following subsections.

4.3.2 Accuracy

Seen as probably the simplest metric to use, accuracy represents the fraction of total samples that were correctly classified by the model [41]. It is defined as the number of correct predictions divided by the total number of predictions, as seen in the equation below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{number\_of\_correct\_predictions}{total\_number\_of\_predictions}$$

Despite being one of the fundamental metrics used either for binary as well as multi-class classification, it may at times provide inaccurate results due to the fact that it associates each class with the same weight [42]. For imbalanced datasets, such as the one used in this dissertation, this is not ideal since highly populated classes will arise over minority classes. In order to combat this we can calculate the Balanced Accuracy metric, that will be explained in the next subsection [42].

4.3.3 Balanced Accuracy

The balanced accuracy metric basically consists of an average of recalls [42]. Due to this, a weighting is applied to each class due to the calculated recall making sure that every class has the same importance in the calculation, despite having a reduced size in the dataset [42]. In imbalanced datasets this metric serves its purpose perfectly, while in balanced datasets its value should approximate the one of the Accuracy metric.

$$Balanced\_Accuracy = \frac{\frac{TP}{Total\_row1} + \frac{TN}{Total\_row2}}{N\_classes}$$

4.3.4 Precision

Another of the most commonly used metrics is precision. Precision represents the ratio of true positives and total positives predicted[40].

$$Precision = \frac{TP}{TP + FP}$$

This metric focuses specifically on Type-1 errors[40], therefore, false positives.

4.3.5 Recall

Recall, also known as sensitivity or hit-rate, can be defined as the fraction of samples that were correctly predicted as positive by the model [41] [43].

$$Recall = \frac{TP}{TP + FN}$$

This metric focuses specifically on Type-II errors [40], therefore, false negatives.

4.3.6 F1-score

The F1-score metric constitutes a combination between the recall and precision metrics, basically representing an harmonic mean of the two [40].

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN}$$

4.3.7 Custom Weighted Accuracy

Lastly, a custom weighted accuracy metric was created, based on the one proposed for the George B. Moody Physionet Challenge [44]. This metric assigns more importance to a certain murmur prediction, based on its pitch. Since the low class only comprises of about 80Hz (from 20 to 100Hz) and shares its frequencies with the fundamental and normal heart sounds, its placed weight is higher, since it is more difficult for the model to make correct predictions. The medium class is assigned to an intermediate weight value while the high class has the lowest weight of the three since it is the most distanced one from the normal heart sounds' frequency range.

$$weighted\_accuracy_{custom} = \frac{4 * nTrueLow + 3 * nTrueMedium + 2 * nTrueHigh + nTrueAbsent + nTrueLowQ}{4 * nLow + 3 * nMedium + 2 * nHigh + nAbsent + nLowQ}$$

Chapter 5

Experimental results and discussion

In this chapter, a presentation and discussion of the experimental results obtained throughout the development of this dissertation will be conducted. Firstly, the results regarding the multi-class classification of the audio segments, therefore the direct output of the model, are presented for each transfer learning model, and the respective class. Despite having described several function attempts in the previous chapter, the results regarding the final diagnostic prediction function presented belong to the highest performing attempt, therefore the "absentPercentage" function. The initial results relate to the usage of the complete CirCor DigiScope dataset. Following these, the same metrics are presented for a partition of the complete dataset - the Grading > I dataset - which will be explained later in the chapter.

The segment multi-classification metric results for every class regarding each of the transfer learning models for the complete CirCor DigiScope Dataset, can be seen from Table 5.1 to Table 5.5, respectively. It is also important to mention that all values were acquired through a 5-Fold Cross Validation method, for each of the models, with the presented values corresponding to the mean and standard deviation obtained for each metric. An EarlyStopping callback was used on all three models, monitoring the val_loss values with a patience of 30 epochs, and all models comprised of a batch size equal to 128.

Table 5.1: Segment prediction resulting metrics regarding the Low class, for each of the transfer learning models on the complete CirCor DigiScope dataset.

	All Segments				
	Low				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.803 ± 0.029	0.514 ± 0.009	0.168 ± 0.062	0.122 ± 0.012	0.138 ± 0.026
VGG16	0.822 ± 0.0123	0.517 ± 0.004	0.143 ± 0.028	0.128 ± 0.008	0.134 ± 0.015
VGG19	0.813 ± 0.011	0.516 ± 0.003	0.156 ± 0.011	0.127 ± 0.005	0.139 ± 0.002

Table 5.2: Segment prediction resulting metrics regarding the Medium class, for each of the transfer learning models on the complete CirCor DigiScope dataset.

	All Segments				
	Medium				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.892 ± 0.009	0.591 ± 0.022	0.254 ± 0.054	0.239 ± 0.052	0.243 ± 0.044
VGG16	0.892 ± 0.005	0.589 ± 0.031	0.243 ± 0.062	0.236 ± 0.068	0.234 ± 0.057
VGG19	0.893 ± 0.009	0.592 ± 0.022	0.253 ± 0.068	0.241 ± 0.051	0.242 ± 0.048

Table 5.3: Segment prediction resulting metrics regarding the High class, for each of the transfer learning models on the complete CirCor DigiScope dataset.

	All Segments				
	High				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.952 ± 0.007	0.622 ± 0.021	0.341 ± 0.039	0.264 ± 0.040	0.297 ± 0.039
VGG16	0.954 ± 0.008	0.635 ± 0.015	0.361 ± 0.064	0.291 ± 0.029	0.319 ± 0.028
VGG19	0.954 ± 0.005	0.636 ± 0.019	0.370 ± 0.068	0.292 ± 0.037	0.326 ± 0.046

Table 5.4: Segment prediction resulting metrics regarding the Absent class, for each of the transfer learning models on the complete CirCor DigiScope dataset.

	All Segments				
	Absent				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.702 ± 0.029	0.602 ± 0.018	0.772 ± 0.053	0.831 ± 0.010	0.799 ± 0.027
VGG16	0.719 ± 0.025	0.611 ± 0.018	0.802 ± 0.048	0.830 ± 0.011	0.814 ± 0.022
VGG19	0.712 ± 0.008	0.605 ± 0.011	0.787 ± 0.021	0.832 ± 0.008	0.808 ± 0.011

Table 5.5: Segment prediction resulting metrics regarding the Low Quality class, for each of the transfer learning models on the complete CirCor DigiScope dataset.

	All Segments				
	Low Quality				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.942 ± 0.005	0.529 ± 0.007	0.102 ± 0.061	0.084 ± 0.017	0.091 ± 0.011
VGG16	0.943 ± 0.0129	0.521 ± 0.014	0.082 ± 0.028	0.069 ± 0.031	0.072 ± 0.034
VGG19	0.942 ± 0.004	0.531 ± 0.0149	0.111 ± 0.010	0.087 ± 0.032	0.097 ± 0.035

By analyzing the presented data, from Table 5.1 to Table 5.5 it is relatively easy to see that the Precision, Recall and F1-score values regarding the "Low", "Medium", "High" and "Low-Quality" classes are not extremely convincing. The low values for these metrics convey either a low capability of the network when it comes to the prediction of the pitch and low-quality classes (minority classes) or the fact that, due to the segment division method, the segment cut window did not include a murmur whatsoever, since its appearance is periodic. This inability is present across all three models. With this, as expected, due to being the majority class, the "Absent" classification metrics are satisfactory.

It can also be noted that all models' performance metrics improve as the pitch type arises. This phenomenon makes sense since the lower frequencies commonly appear where the normal heart sounds commonly appear, causing the neural network to mistake low-frequency murmurs for a typical heart sound. As the pitch value increases, the easier it gets for the model to predict the murmur presence correctly.

Following this per class result presentation, in Figure 5.1 and Figure 5.2 the segment prediction testing accuracy and loss are presented, respectively, regarding each of the transfer learning models on the complete CirCor DigiScope dataset.

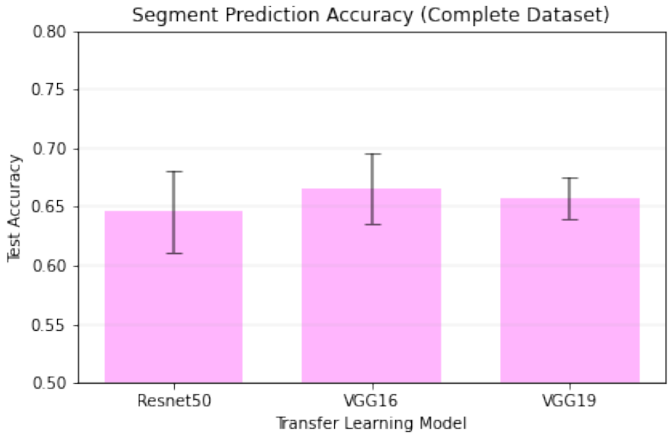

Figure 5.1: Segment prediction testing accuracy for each of the transfer learning models on the complete CirCor DigiScope dataset.

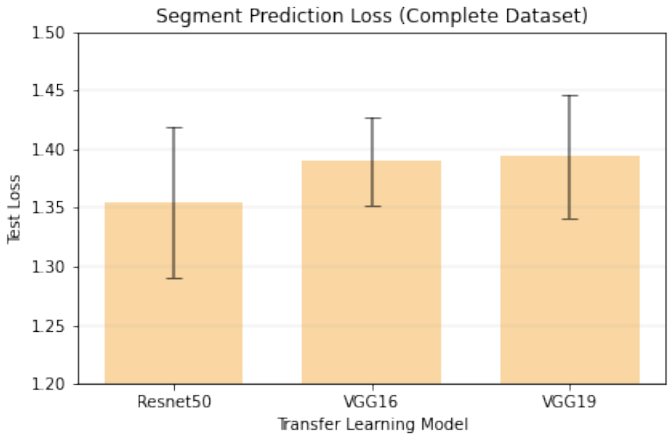

Figure 5.2: Segment prediction testing loss for each of the transfer learning models on the complete CirCor DigiScope dataset.

All three models share similar accuracy values, around 65%, and similar loss values, between 1.35 and 1.40, therefore reaffirming the robustness of the process, regardless of the chosen transfer learning model.

These values will influence and limit the final diagnostic prediction performance results, as expected. From Table 5.6 to Table 5.10, the final diagnostic function's performance metrics per class, regarding the complete CirCor DigiScope dataset can be observed.

Table 5.6: Final diagnostic function resulting metrics regarding the Low class, for each of the transfer learning models on the complete CirCor DigiScope dataset.

	Final Diagnostic Function				
	Low				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.870 ± 0.041	0.512 ± 0.048	0.085 ± 0.120	0.112 ± 0.092	-
VGG16	0.890 ± 0.010	0.475 ± 0.041	0.028 ± 0.057	0.040 ± 0.080	-
VGG19	0.892 ± 0.011	0.568 ± 0.085	0.078 ± 0.088	0.223 ± 0.168	-

Table 5.7: Final diagnostic function resulting metrics regarding the Medium class, for each of the transfer learning models on the complete CirCor DigiScope dataset.

	Final Diagnostic Function				
	Medium				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.935 ± 0.009	0.690 ± 0.057	0.360 ± 0.107	0.419 ± 0.109	0.384 ± 0.101
VGG16	0.938 ± 0.008	0.705 ± 0.041	0.360 ± 0.048	0.449 ± 0.080	0.399 ± 0.059
VGG19	0.940 ± 0.005	0.727 ± 0.018	0.414 ± 0.121	0.488 ± 0.041	0.433 ± 0.055

Table 5.8: Final diagnostic function resulting metrics regarding the High class, for each of the transfer learning models on the complete CirCor DigiScope dataset.

	Final Diagnostic Function				
	High				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.966 ± 0.007	0.721 ± 0.066	0.334 ± 0.137	0.465 ± 0.129	0.386 ± 0.132
VGG16	0.965 ± 0.003	0.721 ± 0.026	0.432 ± 0.138	0.462 ± 0.050	0.441 ± 0.088
VGG19	0.967 ± 0.004	0.745 ± 0.045	0.445 ± 0.159	0.509 ± 0.090	0.459 ± 0.094

Table 5.9: Final diagnostic function resulting metrics regarding the Absent class, for each of the transfer learning models on the complete CirCor DigiScope dataset.

	Final Diagnostic Function				
	Absent				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.783 ± 0.024	0.755 ± 0.053	0.937 ± 0.062	0.805 ± 0.012	0.865 ± 0.022
VGG16	0.793 ± 0.012	0.774 ± 0.040	0.961 ± 0.026	0.801 ± 0.011	0.873 ± 0.009
VGG19	0.797 ± 0.004	0.772 ± 0.020	0.955 ± 0.023	0.809 ± 0.017	0.875 ± 0.001

Table 5.10: Final diagnostic function resulting metrics regarding the Low Quality class, for each of the transfer learning models on the complete CirCor DigiScope dataset.

	Final Diagnostic Function				
	Low Quality				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.923 ± 0.003	-	0.025 ± 0.121	-	-
VGG16	0.922 ± 0.007	-	0.017 ± 0.057	-	-
VGG19	0.925 ± 0.004	-	0.059 ± 0.088	-	-

As expected, despite having some of its metrics improved when compared to the segment prediction values, the final diagnostic results follow the same trend as the prior ones, with the "Absent" class being the most well-predicted and the "Low" and "Low Quality" classes on the other end of that prediction spectrum due to their low f1-score values, which relate to a NaN mean.

In Figure 5.3, all final diagnostic accuracies, for each of the models, are presented.

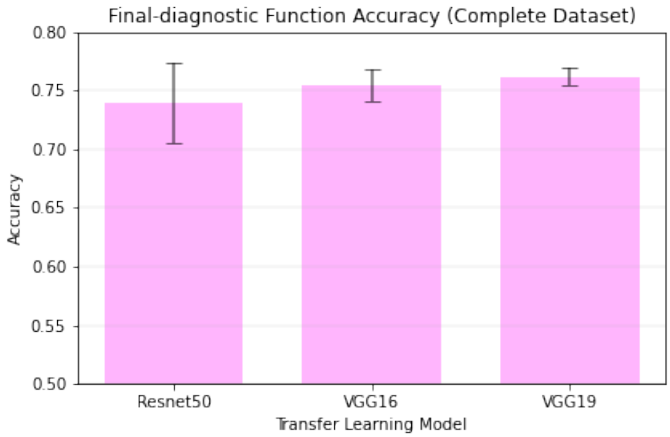

Figure 5.3: Final-diagnostic accuracy for each of the transfer learning models on the complete CirCor DigiScope dataset.

The VGG19 model appears as the highest accuracy one. However, all models share similar accuracies, around 75%.

Having tested all three models for the complete CirCor DigiScope dataset resulting in only satisfactory results, a conditional partition of the dataset was made in order to create the "Grading > I only" dataset. This subdivision was made with the purpose of analyzing each model's performance metrics behaviour when being exposed only to higher grading murmur samples. These higher grading murmurs should be more accessible to detect due to their higher intensity. The Grading > I dataset consisted of data from 1,393 patients, representing 26,669 audio segments. After executing a train, test and validation split, the test data served as input on all previously extracted models. All performance metrics were gathered and used to calculate the mean and standard deviation of the results. These performance metrics regarding the segment prediction on the Grading > I filtered dataset can be seen, per class, from Table 5.11 to Table 5.15. The test segment prediction accuracy and loss obtained for this dataset can also be observed in Figure 5.4 and 5.5, respectively.

Table 5.11: Segment prediction resulting metrics regarding the Low class, for each of the transfer learning models on the Grading > I filtered Dataset.

	All Segments				
	Low				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.932 ± 0.043	0.632 ± 0.056	0.891 ± 0.097	0.266 ± 0.111	0.399 ± 0.149
VGG16	0.956 ± 0.014	0.653 ± 0.042	0.859 ± 0.133	0.309 ± 0.083	0.451 ± 0.101
VGG19	0.950 ± 0.014	0.644 ± 0.055	0.884 ± 0.098	0.291 ± 0.109	0.432 ± 0.136

Table 5.12: Segment prediction resulting metrics regarding the Medium class, for each of the transfer learning models on the Grading > I filtered Dataset.

	All Segments				
	Medium				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.969 ± 0.007	0.837 ± 0.054	0.838 ± 0.093	0.684 ± 0.109	0.746 ± 0.079
VGG16	0.971 ± 0.007	0.854 ± 0.062	0.837 ± 0.093	0.718 ± 0.128	0.762 ± 0.088
VGG19	0.969 ± 0.006	0.835 ± 0.047	0.838 ± 0.096	0.681 ± 0.096	0.745 ± 0.078

Table 5.13: Segment prediction resulting metrics regarding the High class, for each of the transfer learning models on the Grading > I filtered Dataset.

	All Segments				
	High				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.985 ± 0.007	0.866 ± 0.047	0.884 ± 0.064	0.735 ± 0.095	0.795 ± 0.043
VGG16	0.987 ± 0.007	0.882 ± 0.048	0.901 ± 0.071	0.768 ± 0.097	0.824 ± 0.065
VGG19	0.987 ± 0.002	0.882 ± 0.041	0.880 ± 0.079	0.767 ± 0.084	0.813 ± 0.036

Table 5.14: Segment prediction resulting metrics regarding the Absent class, for each of the transfer learning models on the Grading > I filtered Dataset.

	All Segments				
	Absent				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.873 ± 0.054	0.763 ± 0.054	0.885 ± 0.068	0.964 ± 0.008	0.921 ± 0.037
VGG16	0.900 ± 0.034	0.797 ± 0.048	0.918 ± 0.042	0.963 ± 0.006	0.940 ± 0.022
VGG19	0.893 ± 0.023	0.779 ± 0.033	0.908 ± 0.026	0.965 ± 0.006	0.936 ± 0.015

Table 5.15: Segment prediction resulting metrics regarding the Low Quality class, for each of the transfer learning models on the Grading > I filtered Dataset.

	All Segments				
	Low Quality				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.957 ± 0.004	0.580 ± 0.019	0.102 ± 0.097	0.189 ± 0.040	0.128 ± 0.013
VGG16	0.958 ± 0.008	0.573 ± 0.041	0.082 ± 0.133	0.176 ± 0.086	0.104 ± 0.051
VGG19	0.958 ± 0.008	0.590 ± 0.028	0.111 ± 0.098	0.208 ± 0.059	0.139 ± 0.037

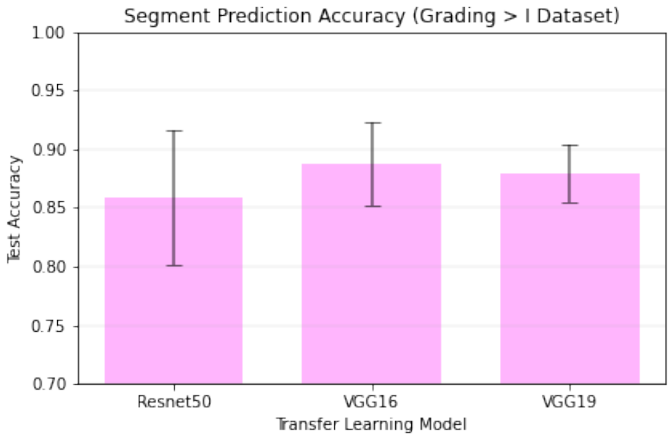

Figure 5.4: Segment prediction testing accuracy for each of the transfer learning models on the Grading > I dataset.

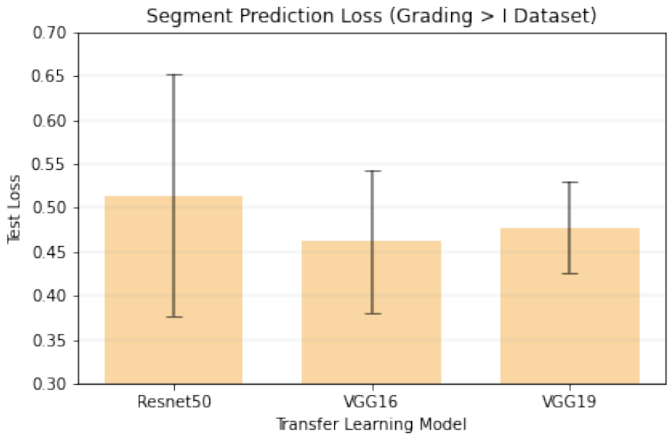

Figure 5.5: Segment prediction testing loss for each of the transfer learning models on the Grading > I dataset.

By observing Table 5.11 to Table 5.15, Figure 5.4 and Figure 5.5, the improvements regarding the previous results relating to the complete CirCor DigiScope dataset are immensely noticeable. All performance metrics increase and improve with more certain and more accurate predictions when the murmurs are of higher intensity.

It is still possible to observe, through each class' f1-score values, that the correct prediction rate improves as the pitch value rises.

One exception regarding these results is the recall metric for the "Low" class. As mentioned in the previous chapter, this metric focuses on false negatives, therefore, segments whose actual value belongs to another class but whose model prediction was the "Low" class. To better understand this recall behaviour, the segment prediction confusion matrix from one of the Cross Validation runs of the Resnet50 model is presented in Figure 5.6.

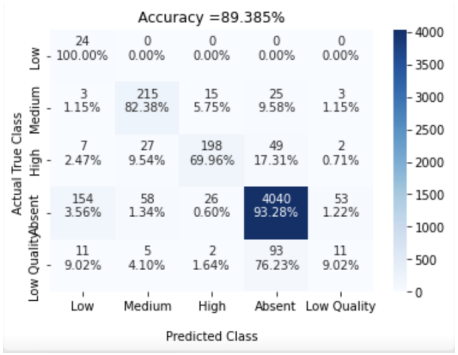

Figure 5.6: Segment prediction confusion matrix for the Resnet50 model on the Grading > I dataset.

As seen in Figure 5.6, the "Low" class holds a low recall value due to the fact that a total of 154 actual "Absent" segments are predicted as being "Low" class segments. This error is seemingly simple to understand due to the fact that low murmurs and normal heart sounds share a similar range of frequencies in the frequency spectrum, which likely causes the model to mistake a typical heart sound for a low murmur sample.

Regarding the accuracy and loss seen in Figure 5.4 and Figure 5.5, all models have their values increased from about 65%, in the complete CirCor DigiScope dataset tests, to about 85%.

Lastly, despite the enhancement of most results, the "Low Quality" class still consists of the most significant prediction struggle.

With the proven improvement regarding the segment prediction process, the final diagnostic function should follow the same trend. Below, from Table 5.16 to Table 5.20, all performance metrics related to the final diagnostic function on the Grading > I dataset are exposed, per class. As observed in the previous segment prediction table, most values thoroughly increase compared to the complete CirCor DigiScope dataset, with still only the "Low Quality" class having no predictions due to its low valued metrics. One important fact to mention is the high values regarding the accuracy, balanced accuracy, precision, recall and f1-score, with some metrics approximating the 100% mark.

Table 5.16: Final diagnostic function resulting metrics regarding the Low class, for each of the transfer learning models on the Grading > I filtered Dataset.

	Final Diagnostic Function				
	Low				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.967 ± 0.043	0.759 ± 0.111	0.860 ± 0.195	0.521 ± 0.222	0.605 ± 0.239
VGG16	0.985 ± 0.012	0.826 ± 0.141	0.809 ± 0.185	0.655 ± 0.281	0.691 ± 0.210
VGG19	0.987 ± 0.010	0.814 ± 0.121	0.860 ± 0.115	0.631 ± 0.240	0.712 ± 0.203

Table 5.17: Final diagnostic function resulting metrics regarding the Medium class, for each of the transfer learning models on the Grading > I filtered Dataset.

	Final Diagnostic Function				
	Medium				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.987 ± 0.002	0.908 ± 0.035	0.845 ± 0.079	0.823 ± 0.071	0.828 ± 0.040
VGG16	0.989 ± 0.003	0.944 ± 0.048	0.845 ± 0.079	0.893 ± 0.098	0.861 ± 0.048
VGG19	0.986 ± 0.004	0.903 ± 0.057	0.863 ± 0.081	0.811 ± 0.117	0.826 ± 0.043

Table 5.18: Final diagnostic function resulting metrics regarding the High class, for each of the transfer learning models on the Grading > I filtered Dataset.

	Final Diagnostic Function				
	High				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.992 ± 0.003	0.936 ± 0.055	0.900 ± 0.049	0.876 ± 0.111	0.881 ± 0.045
VGG16	0.992 ± 0.004	0.921 ± 0.064	0.925 ± 0.061	0.845 ± 0.130	0.875 ± 0.059
VGG19	0.992 ± 0.004	0.937 ± 0.073	0.900 ± 0.093	0.878 ± 0.149	0.873 ± 0.052

Table 5.19: Final diagnostic function resulting metrics regarding the Absent class, for each of the transfer learning models on the Grading > I filtered Dataset.

	Final Diagnostic Function				
	Absent				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.891 ± 0.031	0.872 ± 0.083	0.966 ± 0.050	0.909 ± 0.009	0.936 ± 0.020
VGG16	0.906 ± 0.009	0.909 ± 0.045	0.987 ± 0.015	0.908 ± 0.004	0.946 ± 0.006
VGG19	0.911 ± 0.008	0.904 ± 0.039	0.986 ± 0.011	0.914 ± 0.004	0.949 ± 0.005

Table 5.20: Final diagnostic function resulting metrics regarding the Low Quality class, for each of the transfer learning models on the Grading > I filtered Dataset.

	Final Diagnostic Function				
	Low Quality				
	Accuracy	Balanced Accuracy	Precision	Recall	F1-Score
Resnet50	0.916 ± 0.002	-	0.025 ± 0.195	-	-
VGG16	0.916 ± 0.002	-	0.016 ± 0.185	-	-
VGG19	0.919 ± 0.005	-	0.058 ± 0.115	-	-

In Figure 5.7, it is possible to observe the overall testing accuracy of each model for the Grading > I dataset with, once again, all models sharing high results, near the 90% value.

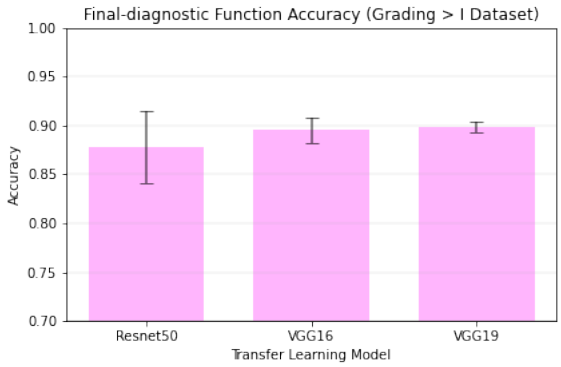

Figure 5.7: Final diagnostic testing accuracy for each of the transfer learning models on the Grading > I dataset.

In order to facilitate the visualization of all accuracy and loss improvements when comparing the attempts on the complete CirCor DigiScope dataset versus the Grading > I filtered dataset, the figures seen below were created.

Figure 5.8 contains all segment prediction accuracies for each of the transfer learning models, regarding the two analyzed datasets. As it is possible to see, across all three models, there is a mean improvement of about 20% when compared to the initial value.

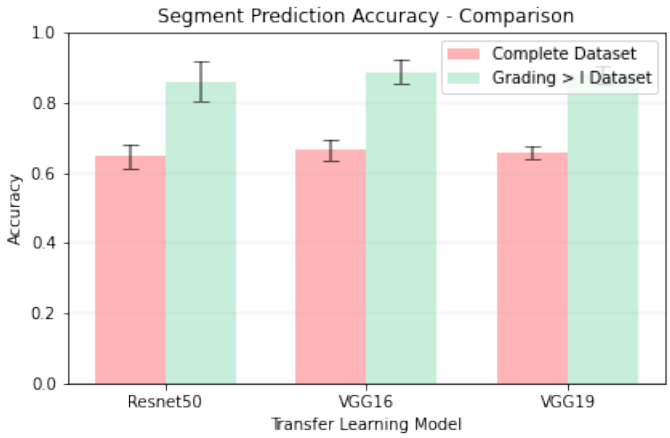

Figure 5.8: Comparison between the segment prediction testing accuracy of each of the transfer learning models, for each of the datasets.

Figure 5.9 contains all segment prediction loss values for each of the transfer learning models, regarding the two analyzed datasets.

These loss values fall drastically for the Grading > I dataset, accompanying the increase of the accuracy values. For example, the Resnet50 mean loss value decreases from 1.355 to 0.514.

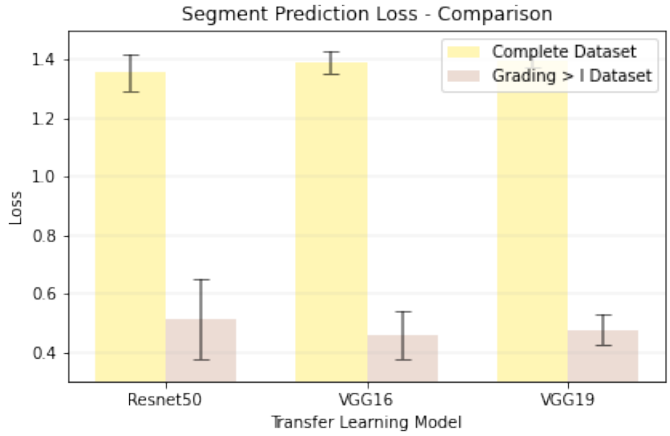

Figure 5.9: Comparison between the segment prediction testing loss of each of the transfer learning models, for each of the datasets.

In Figure 5.10 it is possible to see all final diagnostic accuracy values for each of the transfer learning models regarding the two analyzed datasets. Following the trend of the previous figures, all values increase when using the Grading > I as the testing set for the model.

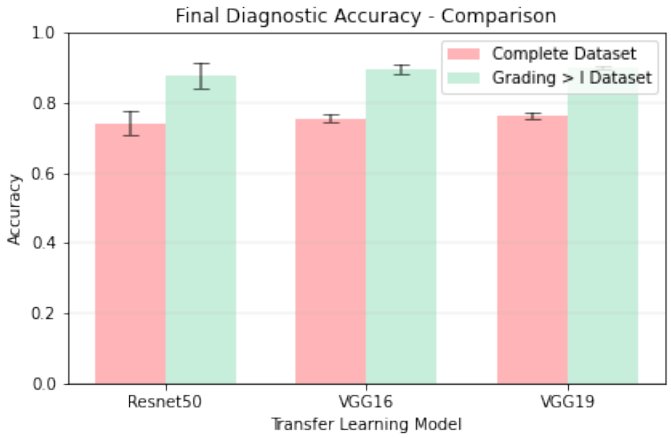

Figure 5.10: Comparison between the final diagnostic accuracy of each of the transfer learning models, for each of the datasets.

Lastly, in Table 5.21 and 5.22, the custom weighted accuracy metric values regarding the segment prediction and final diagnostic function are presented, for the complete CirCor DigiScope Dataset and the Grading > I subset, respectively.

Table 5.21: Custom Weighted Accuracy resulting metric values regarding each of the transfer learning models on the complete CirCor DigiScope dataset.

	Custom Weighted Accuracy	
	Segment Prediction	Final Diagnostic Function
Resnet50	0.471 ± 0.044	0.645 ± 0.079
VGG16	0.501 ± 0.037	0.686 ± 0.031
VGG19	0.486 ± 0.021	0.693 ± 0.034

Table 5.22: Custom Weighted Accuracy resulting metric values regarding each of the transfer learning models on the Grading > I dataset.

	Custom Weighted Accuracy	
	Segment Prediction	Final Diagnostic Function
Resnet50	0.736 ± 0.105	0.825 ± 0.101
VGG16	0.792 ± 0.063	0.871 ± 0.040
VGG19	0.772 ± 0.044	0.869 ± 0.020

As expected, there is an fair increase of the accuracy values from one set to another, accompanying the rise seen while comparing other metrics, for the same exact situation.

Chapter 6

Conclusions and Future Work

This dissertation purposed the creation of an innovative deep learning model for pitch-oriented heart sound multi-classification in order to understand the importance of this sound characteristic in the detection of heart murmurs to further aid the diagnostic of heart diseases. Using the heart sound data provided in the CirCor DigiScope dataset, every sound was divided into several two-second audio segments. Butterworth filters were used to filter the unwanted noise in each audio and separate three frequency ranges individually, regarding three pitch types: Low, Medium and High. This audio data would later be transformed into three different spectrograms that were stacked, being this the main and only feature input of the models. With the intent of getting the best results, instead of training the network from its core, three pre-trained transfer learning models were tested: Resnet50, VGG16 and VGG19. Each of the models' results was then analyzed from a segment prediction standpoint and then forwarded to the final diagnostic prediction function, whose purpose was to examine all predicted values for each of the two-second segments and present a final diagnosis, ready to be communicated to the patient. All mean and standard deviation values were extracted using a 5-Fold Cross Validation method.

The extracted values for each transfer learning model shared similar ranges, with no model essentially more optimal than another, representing the robustness of the entire process. It can also be concluded that the closer the frequency range between a particular pitch-type murmur and normal heart sounds, the harder it is for the model to correctly predict the murmur's presence. Due to this, higher pitched murmurs appear as the most accessible type of murmurs to detect, while low pitch murmurs sit on the other end of that same scale. As the murmur pitch-type value rises, the easier it is for the model to detect its presence. However, it was also clearly observed that when using samples regarding murmurs with a Grading superior to "I", therefore murmurs with higher intensity, all models increased their metrics by a relatively high factor due to the fact that these murmurs are of easier detection, regardless of their frequency range.

Finally, it is possible to conclude that the usage of deep learning algorithms, and specifically transfer learning models, revealed itself as a successful choice for this project since the initial goal

of this dissertation was reached. However, despite obtaining very positive results, there is always room for improvement. Firstly, other deep learning model structures, such as Recurrent Neural Networks, can be tested, as well as other transfer learning models with, for example, a larger required input shape to increase the data's feature definition. Moreover, instead of uniquely using spectrograms as the main features for the provided audio, other features can also be extracted, such as MFCC's, and then concatenated onto the feature vectors to acquire more frequency information from the audio samples. Additionally, experimentations with ordinal regression and classification can also represent interesting trials due to the explicit order in the "Low", "Medium", and "High" classes present in the data. Lastly, other final diagnostic functions could also be explored to better the output accuracy of the system while relating to a heart disease prediction function further down the system's backbone.

References

[1] Koksoon Phua, Jianfeng Chen, Tran Huy Dat, and Louis Shue. Heart sound as a biometric. *Pattern Recognition*, 41:906–919, 3 2008. doi:10.1016/J.PATCOG.2007.07.018.

[2] Jorge Oliveira, Francesco Renna, Paulo Dias Costa, Marcelo Nogueira, Cristina Oliveira, Carlos Ferreira, Alípio Jorge, Sandra Mattos, Thamine Hatem, Thiago Tavares, Andoni Elola, Ali Bahrami Rad, Reza Sameni, Gari D Clifford, and Miguel T Coimbra. The circor digiscope dataset: From murmur detection to murmur classification. 2021. doi:10.1109/JBHI.2021.3137048.

[3] João Manuel and Patrício Pedrosa. Heart sound analysis for cardiac pathology identification: Detection and characterization of heart murmurs. 2013.

[4] Jamal Al-Nabulsi Jalel Chebil. Classification of heart sound signals using discrete wavelet analysis. 2007. URL: https://www.researchgate.net/publication/287592634_Classification_of_heart_sound_signals_using_discrete_wavelet_analysis.

[5] N. Jatupaiboon, S. Pan-Ngum, and P. Israsena. Electronic stethoscope prototype with adaptive noise cancellation. *undefined*, pages 32–36, 2010. doi:10.1109/ICTKE.2010.5692909.

[6] Praveen Kumar Sharma, Sourav Saha, and Saraswati Kumari. Study and design of a shannon-energy-envelope based phonocardiogram peak spacing analysis for estimating arrhythmic heart-beat. *International Journal of Scientific and Research Publications*, 4, 2014. URL: www.ijsrp.org.

[7] João Victor Batista Cabral, Aline Luzia Sampaio Guimarães, Dário Celestino Sobral Filho, and Ana Célia Oliveira dos Santos. Mortality due to congenital heart disease in pernambuco from 1996 to 2016. *Revista da Associacao Medica Brasileira (1992)*, 66:931–936, 2020. URL: https://pubmed.ncbi.nlm.nih.gov/32844925/, doi:10.1590/1806-9282.66.7.931.

[8] Marisa Oliveira. Análise de som cardíaco pediátrico para identificação de sopro. 2020.

[9] Sean Dornbush and Andre E. Turnquest. Physiology, heart sounds. *StatPearls*, 7 2021. URL: https://www.ncbi.nlm.nih.gov/books/NBK541010/.

[10] Joshua D. Pollock and Amgad N. Makaryus. Physiology, cardiac cycle. *StatPearls*, 10 2021. URL: https://www.ncbi.nlm.nih.gov/books/NBK459327/.

[11] H. Naseri and M. R. Homaeinezhad. Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric. *Annals of Biomedical Engineering*, 41:279–292, 2 2013. URL: https://link.springer.com/article/10.1007/s10439-012-0645-x, doi:10.1007/S10439-012-0645-X/FIGURES/11.

[12] Seth L. Thomas, Joseph Heaton, and Amgad N. Makaryus. Physiology, cardiovascular murmurs. *StatPearls*, 7 2021. URL: https://www.ncbi.nlm.nih.gov/books/NBK525958/.

[13] Jorge Oliveira. Subject-driven supervised and unsupervised hidden markov models for heart sound segmentation in real noisy environments. URL: https://sigarra.up.pt/fcup/en/pub_geral.pub_view?pi_pub_base_id=291526.

[14] Ali Raza, Arif Mehmood, Saleem Ullah, Maqsood Ahmad, Gyu Sang Choi, and Byung Won On. Heartbeat sound signal classification using deep learning. *Sensors (Switzerland)*, 19, 11 2019. doi:10.3390/S19214819.

[15] Kausik Basak, Subhamoy Mandal, M. Manjunatha, Jyotirmoy Chatterjee, and Ajoy K. Ray. A comparative study of phonocardiogram analysis techniques based on mixed signal processor. *Proceedings of the 2010 Annual IEEE India Conference: Green Energy, Computing and Communication, INDICON 2010*, 2010. doi:10.1109/INDCON.2010.5712668.

[16] Mohammad K. Zia, Benjamin Griffel, and John L. Semmlow. Robust detection of background noise in phonocardiograms. *undefined*, pages 130–133, 2011. doi:10.1109/MECBME.2011.5752082.

[17] Sheila R Messer, John Agzarian, and Derek Abbott. Optimal wavelet denoising for phonocardiograms. *Microelectronics Journal*, 32:931–941, 2001. URL: https://www.academia.edu/269111/Optimal_Wavelet_Denoising_for_Phonocardiograms.

[18] V. Nivitha Varghees and K. I. Ramachandran. Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope. *IEEE Sensors Journal*, 17:3861–3872, 6 2017. doi:10.1109/JSEN.2017.2694970.

[19] Ajit P. Yoganathan, Ramesh Gupta, Firdaus E. Udwadia, J. Wayen Miller, William H. Corcoran, Radha Sarma, John L. Johnson, and Richard J. Bing. Use of the fast fourier transform for frequency analysis of the first heart sound in normal man. *Medical Biological Engineering*, 14:69–73, 1 1976. doi:10.1007/BF02477093.

[20] V. A. McKUSICK, G. N. WEBB, J. O. HUMPHRIES, and J. A. REID. On cardiovascular sound. *Circulation*, 11:849–870, 1955. URL: https://www.ahajournals.org/doi/abs/10.1161/01.cir.11.6.849, doi:10.1161/01.CIR.11.6.849.

[21] Farshad Arvin, Shyamala Doraisamy, and Ehsan Safar Khorasani. Frequency shifting approach towards textual transcription of heartbeat sounds. *Biological Procedures Online*, 13:7, 2011. URL: /pmc/articles/PMC3396354//pmc/articles/PMC3396354/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3396354/, doi:10.1186/1480-9222-13-7.

[22] Wenjie Zhang, Jiqing Han, and Shiwen Deng. Heart sound classification based on scaled spectrogram and tensor decomposition. *Expert Systems with Applications*, 84:220–231, 10 2017. doi:10.1016/J.ESWA.2017.05.014.

[23] Vladimir Kudriavtsev, Vladimir Polyshchuk, and Douglas L. Roy. Heart energy signature spectrogram for cardiovascular diagnosis. *BioMedical Engineering Online*, 6:1–22, 5 2007. URL: https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/1475-925X-6-16, doi:10.1186/1475-925X-6-16/FIGURES/15.

[24] B. Tovar-Corona and J. N. Torry. Time-frequency representation of systolic murmurs using wavelets. *Computers in Cardiology*, 0:601–604, 1998. doi:10.1109/CIC.1998.731945.

[25] Tamer Ölmez and Zümray Dokur. Classification of heart sounds using an artificial neural network. *Pattern Recognition Letters*, 24:617–629, 1 2003. doi:10.1016/S0167-8655(02)00281-7.

[26] A. F. Quiceno-Manrique, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez. Selection of dynamic features based on time-frequency representations for heart murmur detection from phonocardiographic signals. *Annals of Biomedical Engineering*, 38:118–137, 1 2010. doi:10.1007/s10439-009-9838-3.

[27] Sotiris A. Pavlopoulos, Antonis C.H. Stasis, and Euripides N. Loukis. A decision tree – based method for the differential diagnosis of aortic stenosis from mitral regurgitation using heart sounds. *BioMedical Engineering OnLine*, 3:21, 6 2004. URL: /pmc/articles/PMC481080//pmc/articles/PMC481080/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC481080/, doi:10.1186/1475-925X-3-21.

[28] Rdvan Saraçoglu. Hidden markov model-based classification of heart valve disease with pca for dimension reduction. *Engineering Applications of Artificial Intelligence*, 25:1523–1528, 10 2012. URL: http://dx.doi.org/10.1016/j.engappai.2012.07.005, doi:10.1016/J.ENGAPPAI.2012.07.005.

[29] Tanachat Nilanon, Jiayu Yao, Junheng Hao, Sanjay Purushotham, and Yan Liu. Normal / abnormal heart sound recordings classification using convolutional neural network; normal / abnormal heart sound recordings classification using convolutional neural network, 2016. doi:10.23919/CIC.2016.7868810.

[30] Juan P. Dominguez-Morales, Angel F. Jimenez-Fernandez, Manuel J. Dominguez-Morales, and Gabriel Jimenez-Moreno. Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors. *IEEE Transactions on Biomedical Circuits and Systems*, 12:24–34, 2 2018. doi:10.1109/TBCAS.2017.2751545.

[31] Luca Brunese, Fabio Martinelli, Francesco Mercaldo, and Antonella Santone. Deep learning for heart disease detection through cardiac sounds. *Procedia Computer Science*, 176:2202–2211, 1 2020. doi:10.1016/J.PROCS.2020.09.257.

[32] Siddique Latif, Muhammad Usman, Rajib Rana, and Junaid Qadir. Phonocardiographic sensing using deep learning for abnormal heartbeat detection. 1 2018. URL: http://arxiv.org/abs/1801.08322.

[33] Fatih Demir, Abdulkadir Şengür, Varun Bajaj, and Kemal Polat. Towards the classification of heart sounds based on convolutional deep neural network. *Health Information Science and Systems*, 7:16, 12 2019. URL: /pmc/articles/PMC6684704//pmc/articles/PMC6684704/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6684704/, doi:10.1007/S13755-019-0078-0.

[34] Omair Rashed Abdulwareth Almanifi, Ahmad Fakhri Ab Nasir, Mohd Azraai Mohd Razman, Rabiu Muazu Musa, and Anwar P.P. Abdul Majeed. Heartbeat murmurs detection in phonocardiogram recordings via transfer learning. *Alexandria Engineering Journal*, 61:10995–11002, 12 2022. doi:10.1016/J.AEJ.2022.04.031.

[35] Steven McGee. Auscultation of the heart. *Evidence-Based Physical Diagnosis*, pages 327–332.e1, 2018. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780323392761000391, doi:10.1016/B978-0-323-39276-1.00039-1.

[36] Wei Chen, Qiang Sun, Xiaomin Chen, Gangcai Xie, Huiqun Wu, and Chen Xu. Deep learning methods for heart sounds classification: A systematic review. *Entropy*, 23, 6 2021. doi:10.3390/e23060667.

[37] Priya Dwivedi. Understanding and coding a resnet in keras. URL: https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33.

[38] Srikanth Tammina. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, 9:p9420, 10 2019. URL: https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/, doi:10.29322/IJSRP.9.10.2019.P9420.

[39] Michael A Nielsen. *Neural networks and deep learning*, volume 25. Determination press San Francisco, CA, USA, 2015.

[40] Aayush Bajaj. Performance metrics in machine learning. 2022.

[41] Joydwip Mohajon. Confusion matrix for your multi-class machine learning model. 2020.

[42] Margherita Grandini, Enrico Bagli, and Giorgio Visani. Metrics for multi-class classification: an overview. 8 2020. URL: https://arxiv.org/abs/2008.05756v1, doi:10.48550/arxiv.2008.05756.

[43] Kunumi. Métricas de avaliação em machine learning: Classificação. 2020.

[44] Matthew A. Reyna, Yashar Kiarashinejad, Andoni Elola, Jorge Oliveira, Francesco Renna, Annie Gu, Erick A. Perez Alday, Nadi Sadr, Ashish Sharma, Sandra Mattos, Miguel T. Coimbra, Reza Sameni, Ali Bahrami Rad, and Gari D. Clifford. Heart murmur detection from phonocardiogram recordings: The george b. moody physionet challenge 2022. 2022. URL: https://moody-challenge.physionet.org/2022/#rules.