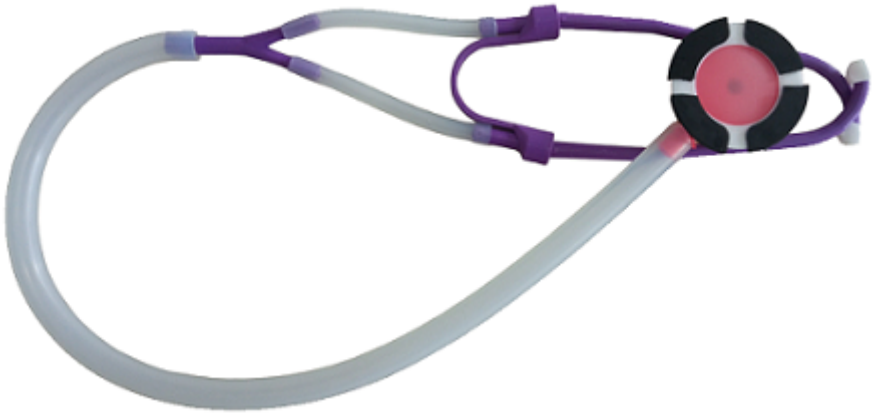
TÉCNICO LISBOA



A Novel System for Cardiovascular Screening Based on Simultaneous Phonocardiography and Electrocardiography During Auscultation

Sofia Margarida Marques Monteiro

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisors: Prof. Hugo Humberto Plácido da Silva
Prof. Ana Luísa Nobre Fred

Examination Committee

Chairperson: Prof. João Miguel Raposo Sanches
Supervisor: Prof. Hugo Humberto Plácido da Silva
Member of the Committee: Prof. Francesco Renna

December 2022

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Preface

The work presented in this thesis was performed at Instituto de Telecomunicações (IT) in Lisbon, during the period from March 2022 to October 2022, under the supervision of professor Hugo Humberto Plácido da Silva. The thesis was co-supervised at Instituto Superior Técnico by professor Ana Luísa Nobre Fred. The work also had the support from doctor Sérgio Laranjo, cardiologist at Centro Hospitalar Universitário de Lisboa Central (CHULC).

Acknowledgments

I would like to first express my deepest gratitude to Professor Hugo Silva and Professor Ana Fred for their constant guidance and encouragement, and for all the opportunities and challenges they have provided me and that made this work possible.

I would also like to thank Afonso Raposo and Arlène Robic for their help and knowledge, that proved invaluable throughout the development of this work. My acknowledgments also extend to Instituto de Telecomunicações (IT), for providing all the necessary tools for this project.

To my family, parents and brother, I am forever thankful for your unparalleled support and patience, and for giving me the opportunities and experiences that have made me who I am.

Finally, to my great friends and honorary family: Francisco, Joaquim, Luís, Madalena, Mariana, Miguel, Tiago, and Vladlen - thank you for keeping my head out of the research and for all the wonderful times we shared.

This work was partially funded by Fundação para a Ciência e Tecnologia (FCT)/MCTES through national funds and when applicable co-funded by EU funds under the project UIDB/50008/2020, by the European Regional Development Fund (FEDER) through the Operational Competitiveness and Internationalization Programme (COMPETE 2020), by National Funds through the FCT under the LISBOA-01-0247-FEDER-069918 “CardioLeather”, and by the IT, whose support is greatly acknowledged.

Abstract

The combination of the Phonocardiogram (PCG) and the Electrocardiogram (ECG) allows the evaluation of the electromechanical condition of the heart and could greatly improve the accuracy of an initial cardiovascular disease diagnosis.

In this work, the advantages of combining synchronous PCG and ECG were analyzed in depth, resulting in the development of deep learning models for abnormality detection and in a novel prototype of an electronic stethoscope that combines PCG and ECG sensors.

The best performing models include a recurrent neural network for murmur detection from heart sounds, which achieved a sensitivity of 82.7% and specificity of 80.1%, and a hybrid neural network based on convolutional and recurrent neural networks for abnormality detection from multi-modal data, which reached a sensitivity of 89.9% and specificity of 78.7%.

The developed prototype was evaluated experimentally to assess the sensors, the morphology of the acquired signals, and the effect of the rotation of the stethoscope head, which further extends the state-of-the-art. It was verified that the ECG lead voltages at different angles of rotation follow Kirchhoff's law and, by comparing the ECG waveforms between a reference and the experimental device, the best case reached a Pearson's correlation coefficient of 0.928.

The results demonstrate the validity of the device and the potential of combining the two signals, establishing a foundation for further development. The accessibility and low cost of the device, combined with automatic techniques for disease screening, should significantly facilitate its integration in healthcare facilities and telemedicine, as well as in research and education.

Keywords

Electrocardiogram (ECG); Phonocardiogram (PCG); Auscultation; Electronic Stethoscope; Multimodal Signals

Resumo

A combinação do fonocardiograma (PCG) e do eletrocardiograma (ECG) permite avaliar a condição eletromecânica do coração, e poderia aumentar significativamente a precisão do diagnóstico precoce de doenças cardiovasculares.

Este trabalho explora as vantagens de combinar o PCG e ECG síncronos, tendo resultado no desenvolvimento de modelos de aprendizagem profunda para deteção de anormalidades e num novo protótipo de um estetoscópio eletrónico que combina sensores de PCG e ECG.

Os modelos com melhor desempenho incluem uma rede neuronal recorrente usada para deteção de sopros em sons cardíacos, com uma sensibilidade de 82.7% e especificidade de 80.1%, e uma rede neuronal híbrida baseada em camadas convolucionais e recorrentes para a deteção de anormalidades em dados multimodais, com uma sensibilidade de 89.9% e especificidade de 78.7%.

O protótipo foi avaliado experimentalmente para analisar os sensores, a morfologia dos sinais adquiridos, e o efeito da rotação da cabeça do estetoscópio, que expande o estado da arte. Verificou-se que as derivações de ECG adquiridas com diferentes ângulos de rotação seguem a lei de Kirchhoff e, ao comparar a morfologia dos ECGs adquiridos com o dispositivo com a de um sinal de referência, obteve-se um coeficiente de correlação de Pearson de 0.928 para o melhor caso.

Os resultados demonstram a validade do dispositivo e o potencial de combinar os dois sinais. O seu baixo custo e elevada acessibilidade, em conjunto com técnicas automáticas para rastreio de doenças, deverão facilitar significativamente a sua integração em estabelecimentos de saúde e em telemedicina, bem como em investigação ou educação.

Palavras Chave

Eletrocardiograma (ECG); Fonocardiograma (PCG); Auscultação; Estetoscópio Eletrónico; Sinais Multimodais

Contents

List of Figures

List of Tables

Acronyms

A-V	Atrioventricular
BiLSTM	Bidirectional Long Short-Term Memory
CinC	Computing in Cardiology
CNN	Convolutional Neural Network
CT	Computed Tomography
CVD	Cardiovascular Disease
dB	Decibel
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DCT	Discrete Cosine Transform
ECG	Electrocardiogram
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
fNIRS	Functional Near-Infrared Spectroscopy
GRU	Gated Recurrent Unit
Hz	Hertz
HSMM	Hidden semi-Markov Model
LR	Logistic Regression
LSTM	Long Short-Term Memory
MFCC	Mel-Frequency Cepstral Coefficient
MITHSDB	MIT Heart Sounds Database
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
PLA	PolyLactic Acid
PCC	Pearson's Correlation Coefficient
PCG	Phonocardiogram

PPG	Photoplethysmography
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SA	Sinoatrial
SNR	Signal-to-Noise Ratio
SGD	Stochastic Gradient Descent
WHO	World Health Organization

Chapter 1

Introduction

This chapter summarizes the context behind the motivation of this work (Sections 1.1 and 1.2), establishes the main objectives (Section 1.3), and presents a summary of the main contributions (Section 1.4) and of the outline of the remainder of the thesis (Section 1.5).

1.1 Overview

Cardiovascular Diseases (CVDs) are the leading cause of death globally and one of the main causes of increased morbidity, which lowers the quality of life of the patients and leads to incremental long-term healthcare costs. An early diagnosis of both congenital and acquired heart conditions, together with regular heart monitoring, is essential to reduce the burden of these diseases and prevent premature deaths [1].

Techniques such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and echocardiography can provide high-quality images and a detailed characterization of both the physiologic function and anatomy of the heart. However, these have a very high cost, requiring complex instrumentation and specialized personnel, reason for which such techniques are uncommon for a first line of screening [2]. Additionally, they are typically only affordable in large hospitals in high-income countries, while, according to the World Health Organization (WHO), low- and middle-income countries have the highest prevalence of CVDs [3]. Even with the development of more advanced cardiac imaging exams, cardiac auscultation remains the most common primary screening method [4]. It is very simple, quick, and inexpensive, and by listening to the heart sounds doctors can diagnose several cardiac conditions related with the mechanical condition of the heart [5].

The digital recording of heart sounds is named Phonocardiogram (PCG), and it can be easily obtained with a chest-placed microphone and used for further analysis and processing [3]. Digital stethoscopes, which can record and store PCG signals during the auscultation routine, are paving the way for the development of automatic disease screening systems.

The Electrocardiogram (ECG), on the other hand, records the electrical activity of the heart and is also one of the most common primary methods of disease screening. It is used to evaluate the electrical condition of the heart and is a very powerful, non-invasive, inexpensive, and easy to use technique. The ECG and the PCG are concurrent phenomena, since they result from the same events in the cardiac cycle [6]. Thus, their simultaneous acquisition and analysis comes as a natural step. If the two signals can be recorded simultaneously during a routine auscultation exam, it is possible to inspect the electromechanical condition of the heart and obtain a faster diagnosis in a very straightforward way. Additionally,

there are many cases where the PCG and ECG contain mutually exclusive information about the heart's condition, so a combination of these two signals could also greatly improve the accuracy of an initial screening [7].

1.2 Motivation

The simultaneous acquisition of the PCG and ECG with highly portable cardiac sensing systems is already a reality and could revolutionize how we address public health policies for cardiology. Current devices are mostly based on integrating ECG electrodes in an electronic stethoscope so the two signals can be acquired at the same time. However, most still present some limitations, which have prevented their widespread implementation in clinical practice, mainly related with inadequate form factors, high costs, or proprietary software and hardware. Advanced signal processing and machine learning techniques for automatic disease detection have also been successfully implemented for both signals, but most research still focuses on exploiting the PCG and the ECG separately [8], typically with PCG used for an early stethoscope-based screening and ECG for a more advanced diagnosis [7].

Therefore, the goal of this work was to perform an in-depth research on the advantages of combining synchronous PCG and ECG for the automatic detection of heart disease during auscultation. Through the integration of deep learning techniques with a hardware architecture for signal acquisition, we hope to contribute to the development of a robust system for the pre-screening of CVD, that can improve the accuracy of an initial diagnosis and thus better inform and support the decisions of medical professionals.

1.3 Objectives

This work performed an analysis of the advantages of using simultaneous electrocardiography and phonocardiography during auscultation for CVD screening, where several objectives can be outlined:

- Analyze the current systems for multi-modal acquisition and classification of cardiac signals, and explore the main difficulties and challenges in implementing them in clinical practice.
- Design and implement deep learning based algorithms for classification of biosignals acquired during auscultation, namely the PCG and simultaneous PCG and ECG.
- Develop a novel hardware prototype for synchronous PCG and ECG acquisition that overcomes the limitations of current devices.
- Create a database with signals acquired using the proposed hardware in order to perform an experimental evaluation of the sensors and the prototype's design.
- Discuss the applicability of the proposed methods in real clinical applications and identify possible improvements and the next steps for future work.

1.4 Contributions

This work has resulted in several contributions in the fields of cardiac signals acquisition and classification:

- The development of a module for PCG processing and analysis, made available as open-source in the BioSPPy Python toolbox [1].
- A paper accepted at the 2022 Computing in Cardiology (CinC) conference, as part of the George B. Moody 2022 challenge for murmur detection from heart sound recordings [9].
- A "Best Oral Presentation" George B. Moody Physionet Challenge award at CinC 2022.
- An article that studies the use of Bidirectional Long Short-Term Memory (BiLSTM) neural networks for the segmentation of PCG signals, which is currently in the second stage of the review process in the *Biomedical Signal Processing and Control* journal.
- The development of a novel open-source and low-cost stethoscope prototype for synchronous multi-modal ECG and PCG acquisition.
- A study of the effect of the rotation of the stethoscope head in the acquired ECG leads, which fills an existing gap in the state-of-the-art.
- The integration of two ECG sensors on the stethoscope head for 2-lead ECG acquisition, allowing the estimation of additional leads at any angle and a full view of the heart's electrical activity in the frontal plane.
- The experimental evaluation of a novel PCG sensor to be integrated in the ScientISST SENSE sensor portfolio.

1.5 Outline

In this chapter the context and motivation behind the work were presented, together with the main contributions and objectives. In Chapter 2 the relevant background is introduced, with the main physiological principles behind the two signals of interest and key concepts related with neural networks and the main classification metrics. Chapter 3 summarizes the state-of-the-art in the automatic analysis of uni-modal ECG and PCG and of the multi-modal signals, and also provides a review of the current systems for the simultaneous acquisition of the two signals. Chapter 4 explores the development of classification methods based on neural networks for heart disease detection from signals acquired during auscultation. Chapter 5 details the development of a novel electronic stethoscope for simultaneous electrocardiography and phonocardiography and the experimental evaluation of the device and sensors. Chapter 6 then outlines the main conclusions and proposes the future work.

[1] https://github.com/scientisst/BioSPPy/blob/main/biosppy/signals/pcg.py

Chapter 2

Background

This section introduces the necessary background, describing the main physiological principles behind electrocardiography (Section 2.1) and phonocardiography (Section 2.2), which are two non-invasive modalities for the acquisition of cardiac signals. The relationship between the two signals and the potential of their combination are also explored (Section 2.3). An introduction to neural networks and the most common architectures for the classification of these signals (Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)) are also presented (Section 2.4), along with an insight into the standard classification metrics (Section 2.5).

2.1 Electrocardiogram (ECG)

The Electrocardiogram (ECG) is a recording of the electrical signals of the heart, and its waveform results from a series of coordinated electrical events across a specialized conduction system. The relevant anatomy and the structure of the ECG waveform are presented in Figure 2.1.

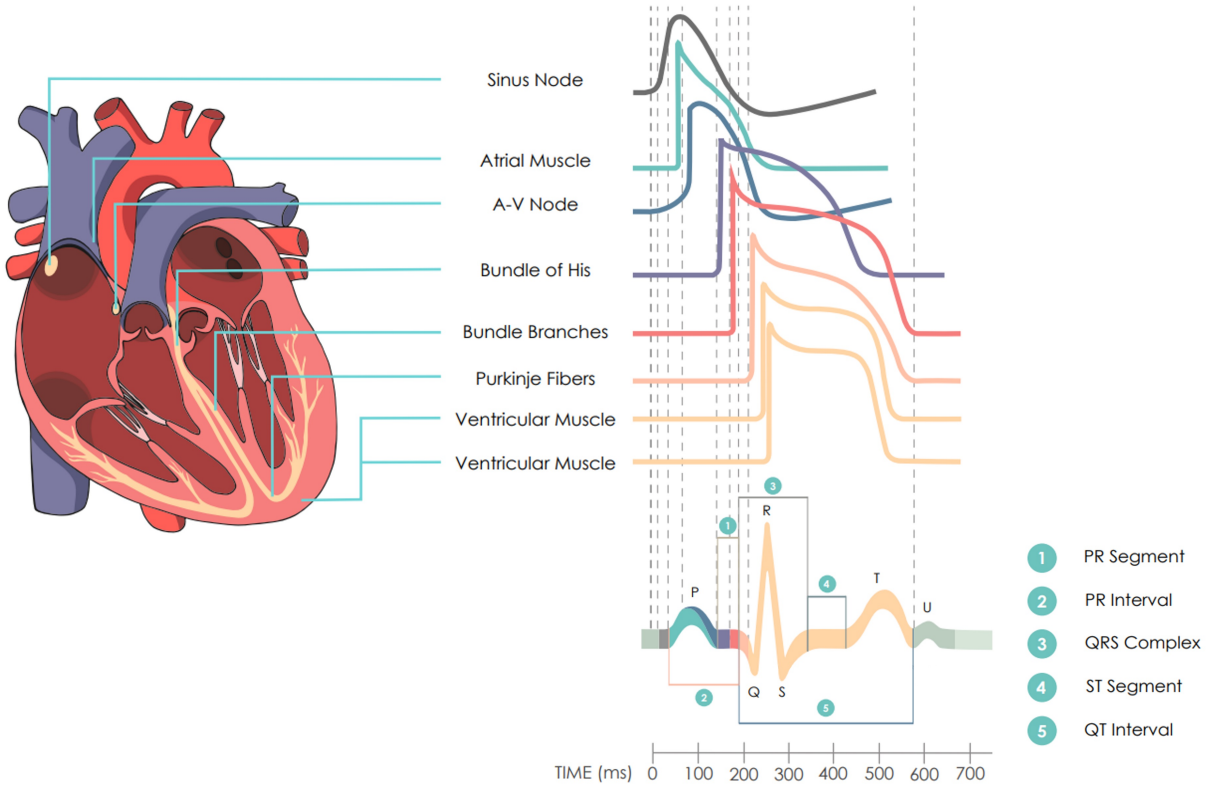


Figure 2.1: Heart conduction system, heartbeat complexes, and ECG waveform [10].

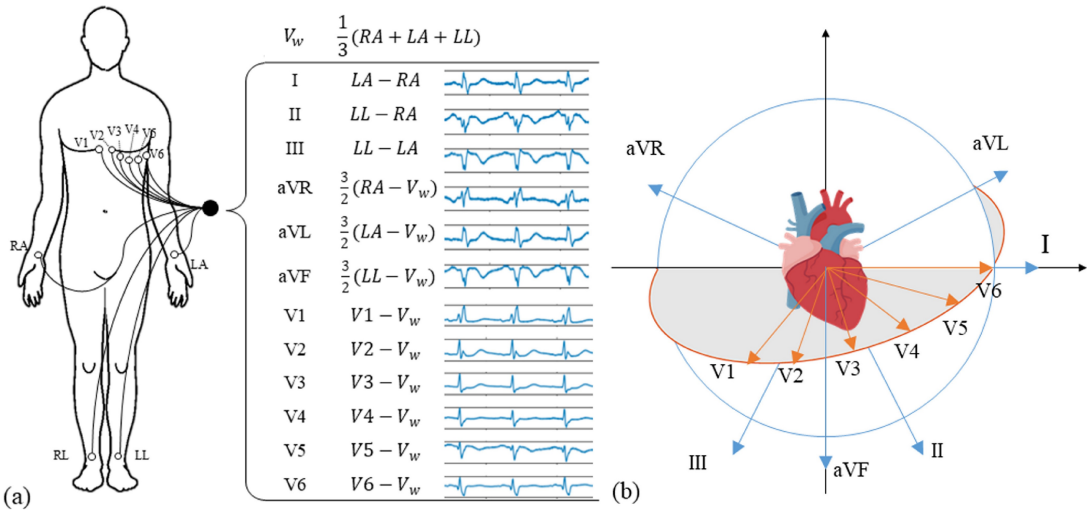


Figure 2.2: Standard placement of electrodes for 12-lead ECG system. a) Spatial placement of 10 electrodes used by 12-lead ECG systems. b) Electrical potential differences from different views of the heart [15].

The action potentials are generated at the Sinoatrial (SA) or sinus node, the natural pacemaker of the heart. When the SA node fires, the electrical activity propagates through the atrial muscle and causes the slow depolarization of the atria, which results in the P wave. Then, there is a propagation delay at the Atrioventricular (A-V) node while the blood is transferred from the atria to the ventricles, resulting in an isoelectric segment. When the A-V node fires, the bundle of His, the bundle branches, and the Purkinje fibers rapidly spread the stimulus to the ventricular muscle, which causes a rapid depolarization. This corresponds to the QRS complex. There is another isoelectric segment (ST segment) during the plateau of the action potential, and then the repolarization of the ventricles causes the T-wave [11].

Any disturbance in this regular rhythmic activity of the heart can be a sign of a pathology. For example, the QRS waveform can be affected by conduction disorders, and the ST segment can be elevated or depressed due to myocardial ischemia or infarction. The ECG is extremely useful for the diagnosis of cardiovascular diseases and continuous monitoring of the heart rate and rhythm [11].

Numerous studies have firmly established the clinical ECG as a standard procedure, and ECG signal acquisition in clinical practice is done using the 12-lead ECG system (represented in Figure 2.2). It consists of three limb leads (Lead I, Lead II, and Lead III) and three augmented leads (aVR, avF, and aVL), which record the electrical activity in the frontal plane, and the six precordial leads (V1-V6), which record the electrical activity in the horizontal plane [12]. In theory, to evaluate most of the heart's electrical activity it would be sufficient to measure only three of its independent components; for example, two of the limb leads to reflect the frontal components, and one precordial lead to reflect the components in the anterior-posterior regions [13]. Even though the 12-lead system has some redundancy [11], using several leads lets us view the cardiac electrical activity from different perspectives and also increases the probability of obtaining suitable signals for diagnosis [14].

Nevertheless, to increase the adoption of this technique, recent work has focused on devising more practical acquisition methods. Single-lead ECGs are ECG recordings with one lead obtained using only three electrodes or two electrodes in a virtual ground configuration [14]. They have been increasingly

used for monitoring of cardiac arrhythmias in both clinical and non-clinical environments and can be recorded using handheld devices that are widely available for both doctors and patients. The more recent devices even allow visual assessment during or immediately after acquisition, and easily sharing the signal between doctors and patients through wireless communication [16].

Even though single-lead ECGs don't carry as much information as 12-lead ECGs, they are a great alternative for screening purposes. The exam is less time-consuming and can even be done by patients in their homes, removing accessibility barriers. It also removes some of the logistical challenges of doing an ECG exam in primary care, improving the early diagnosis of arrhythmias and other cardiac abnormalities [17].

However, single-lead ECGs cannot be used to detect certain conditions, such as acute coronary syndrome, myocardial infarction, or left ventricular hypertrophy [16]. Furthermore, a good characterization of the measured ECG lead is essential for a correct interpretation. For example, it is known that handheld devices record an ECG signal similar to Lead I from the 12-lead ECG [18], but if the electrodes are placed in a different region of the body the signals will reflect a different view of the electrical axis of the heart, altering the morphology of the ECG waveform [14].

Nonetheless, the value of single-lead ECG for clinicians has already been established and it can be used on its own to diagnose many cardiac diseases (such as atrial fibrillation and flutter, left and right bundle branch blocks, or premature ventricular contraction) [16], both through automatic methods and visual assessment. Due to the recent development of portable and low-cost ECG monitoring systems for both long-term and short-term recordings, there has been a high influx of data that needs to be interpreted, which has prompted the rapid implementation of automatic techniques for ECG classification and disease screening [19].

2.2 Phonocardiogram (PCG)

A Phonocardiogram (PCG) is an audio recording of the heart sound signal and it is related to the contractile and mechanical functions of the heart [11]. Heart sounds are generated by the beating heart and the blood flow through it [3], and result from the turbulence created by the closure of the heart valves during the cardiac cycle. Figure 2.3 shows a representation of a phonocardiogram with the four main types of heart sounds and their typical timings during a cardiac cycle.

A normal cardiac cycle has two main components that are always present: the first heart sound (S1) and the second heart sound (S2), which together are called the fundamental heart sounds. S1 results from the closing of the mitral and tricuspid valves after the atrium contracts to move blood towards the ventricles, and it marks the beginning of ventricular systole. S2 results from the closing of the aortic and pulmonary valves after the ejection of the blood from the ventricles, and marks the beginning of the diastole [11]. The dominant frequencies of S1 and S2 are between 20 and 175 Hz, but S1 peaks can have a bandwidth from 10 to 140 Hz, and S2 peaks from 10 to 400 Hz [21]. These heart sounds produce the characteristic *lub-dub* sounds that are always audible during auscultation, with S1 forming the *lub* and S2, which is generally louder, shorter, and sharper than S1, forming the *dub*. They also create the

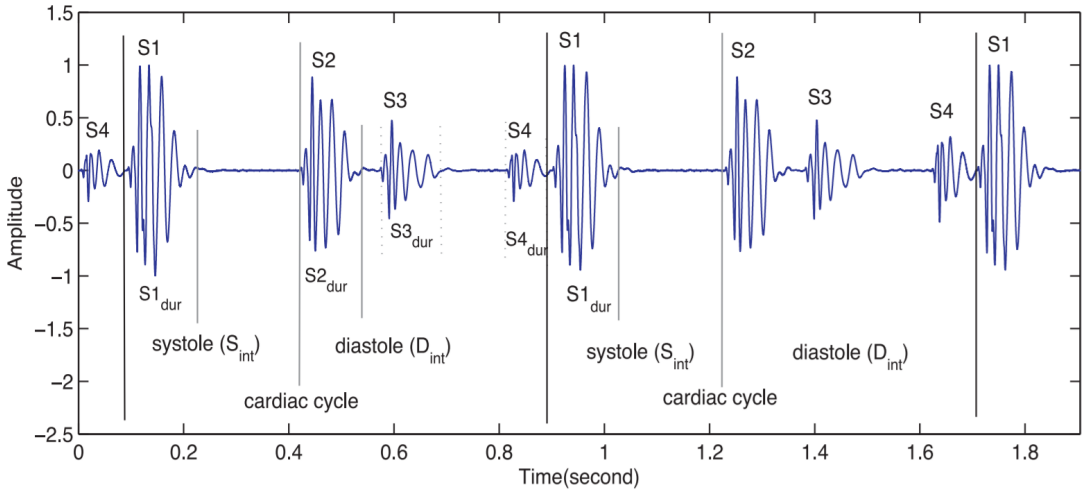


Figure 2.3: Illustration of the PCG signal including the heart sounds S1, S2, S3, and S4, alongside the systolic and diastolic periods [20].

framework for the analysis of all the other heart sounds [22].

Two additional heart sounds are only present in some cases: the third heart sound (S3) and the fourth heart sound (S4). Both originate in the ventricle and are caused by rapid diastolic filling, although they differ in timing and clinical significance. S3 occurs due to a sudden termination of the ventricular rapid-filling phase during a part of diastole, where the ventricular walls are relaxed, and thus has a very low frequency [11]. It may be physiologic in persons younger than 40 years of age, but in older patients its presence can indicate a severe systolic dysfunction or valvular regurgitation. S4 occurs in late diastole due to the displacement of blood into abnormally stiff ventricles by atrial contractions. It is always considered pathological and can indicate ventricular hypertrophy or fibrosis [11,22]. Both S3 and S4 are very low sounds, usually observed in the frequency range of 20 to 70 Hz [21].

In addition to these heart sounds, clicks or heart murmurs can also occur, which result from turbulent blood flow that is strong enough to produce audible noise [11]. Heart murmurs can have very different morphologies and are classified using several characteristics, including timing (systolic, diastolic, or continuous), location, intensity, and pitch (low, high, or a mixture of frequencies) [22]. They can be benign or a sign of cardiovascular defects or diseases, such as valvular stenosis or regurgitation (Figure 2.4), septal defects, or even some types of cardiomyopathy [4]. Placing the stethoscope at the different points of auscultation, that are generally correlated with the location of cardiac valves (Figure 2.5), may reveal the heart murmurs associated with those valves. The frequencies of these sounds are usually between 200 and 600 Hz [20], with the aortic regurgitation murmur containing the highest dominant frequency of approximately 400 Hz [22].

These various features of the heart sounds can be a key tool for the diagnosis of heart abnormalities. However, the PCG is a much more complex signal than the ECG, and so the interpretation of PCG waveforms hasn't been established in clinical practice. Visual analysis usually only allows the identification of coarse features, such as the main heart sounds or the presence of murmurs. Even the visual identification of S1 and S2 becomes complicated if there are murmurs between the sounds or if the heart rate is not low enough such that the diastole (S2-S1 interval) is longer than the systole (S1-S2 interval) [11].

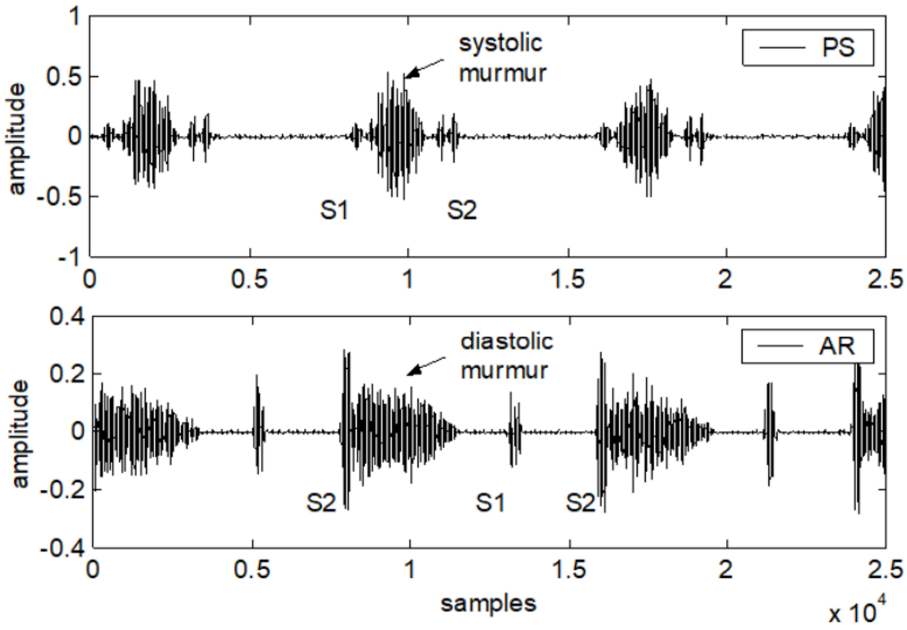


Figure 2.4: PCGs of pulmonary stenosis (PS) with systolic murmur, and of aortic regurgitation (AR) with diastolic murmur [23].

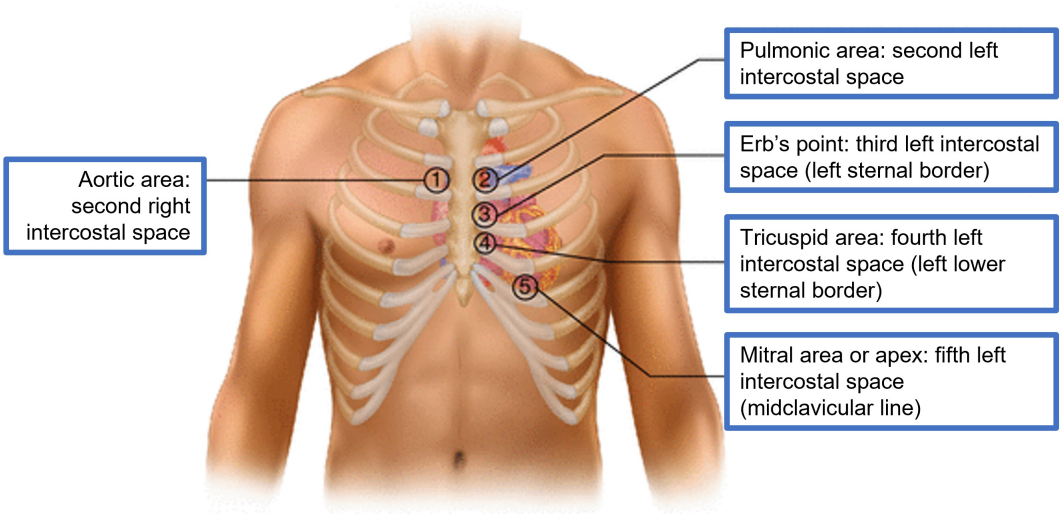


Figure 2.5: Five standard points of auscultation. Image from [24].

The main advantage of the PCG is that it can be obtained by auscultation [11], but the heart sounds can also be difficult to identify because they are faint and appear closely together in time [5], with frequencies at the lower end of our audible range, that aren't as well perceived by the human ear [22]. Additionally, cardiac auscultation with acoustic stethoscopes can only be performed by trained medical professionals that are capable of properly understanding and interpreting the heart sounds, which is a limitation.

These difficulties, together with the recent advances in areas such as machine learning and artificial intelligence [25], have prompted the development of automatic techniques for the analysis of PCG signals. This automatic analysis has three major steps: pre-processing, segmentation, and classification. In pre-processing, the signal is filtered, its quality is assessed, and relevant features are extracted. In segmentation, the objective is to decompose the signal and delineate the start and end of its components [5]. Most of the time, the signal is only segmented into its fundamental heart sounds, that is, S1, systole, S2, and diastole. In classification, the objective is to classify the heart sounds into normal or abnormal, or to identify and characterize pathologies, using the features of each component of the signal [26].

2.3 Multimodal Cardiac Signals

Many bodily processes are not independent of one another. In fact, they are often bound together by the same physical or physiological phenomena and are highly correlated. By analyzing a single biosignal without considering the other components of the system we can only gain a partial understanding of the process, so for a deeper analysis of the underlying mechanisms of the system of interest we should integrate multiple sources of information [11].

The PCG and ECG are multimodal cardiac signals, and thus their simultaneous analysis comes as a natural step. They are concurrent phenomena, seeing that the PCG results from the mechanical operation of the heart, which in turn relies on its electrical activation (expressed by the ECG) [6]. Their main features result from the same events in the cardiac cycle, and thus the two signals are time-locked with each other, as shown in Figure 2.6.

The QRS complex is directly related to the contraction of the ventricles, which seals the atrioventricular valves shut [11]. There is a slight delay between the electrical activation and mechanical contraction of the heart, and thus S1 occurs just after the QRS onset [28]. Ventricular systole ends with the closing of the semilunar valves, and this end of contraction is also marked by the T-wave [11]. S2 occurs approximately at the end of the T-wave.

Even though the two signals are synchronized, evidence of pathology in a PCG signal does not imply the existence of the same problems in the corresponding ECG [28]. For example, defective heart valves often can't be detected from the ECG, but they manifest as heart murmurs and clicks in the PCG. In the same way, conduction disorders are easily detected in the ECG, but are more difficult to identify using the heart sounds. Thus, combining the various features of the heart sounds with the ECG can provide a richer insight into the function of the heart and improve the diagnosis of heart disease.

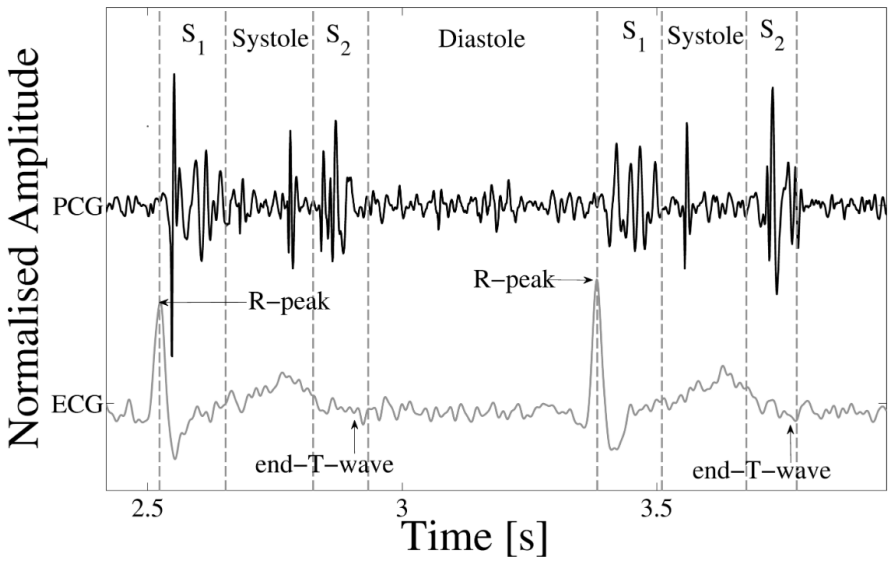


Figure 2.6: Synchronized PCG and ECG signals, with segmentation of the fundamental heart sounds. Mid-systolic clicks, typical of mitral valve prolapse, can be seen [27].

The combination of the two signals can also enhance the segmentation of the PCG into the various heart sounds. Through R-peak and T-wave detectors and prior information about state duration averages [27], the locations of S1 and S2 can be easily inferred even in low quality PCG signals.

2.4 Neural Networks

The term "neural networks" encompasses a very large number of models and learning methods [29]. The simplest unit in a neural network is called a neuron, represented in Figure 2.7, in which the output corresponds to the application of a non-linear activation function to a weighted sum of the inputs.

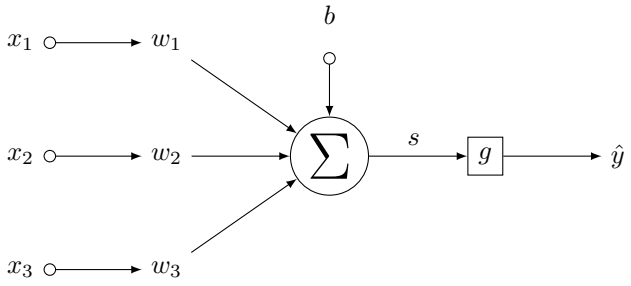


Figure 2.7: Single neuron diagram with three inputs and a bias term.

It receives a set of values $x_i \in \mathbb{R}^n$ and outputs a real value $\hat{y} \in \mathbb{R}$. The first operation corresponds to a weighted sum of the inputs:

$$s = w_1 x_1 + ... + w_n x_n + b = x^T w + b \tag{2.1}$$

where w is the vector of the trainable weights w_i and b is the scalar bias term. The second operation corresponds to the application of a non-linear activation function g, which will restrict the value of the

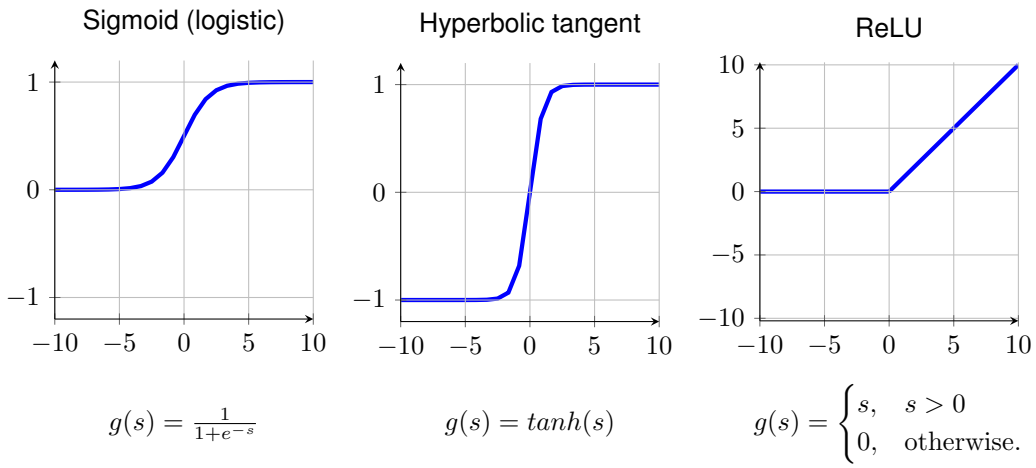


Figure 2.8: Sigmoid, hyperbolic tangent, and ReLU activation functions.

output from the neuron to a certain limit:

$$\hat{y} = g(s) \tag{2.2}$$

Activation functions are what allows neural networks to learn complex patterns in the data. Unlike other classifiers, neural networks can learn the most appropriate nonlinear transformation for each task, which will map the inputs to a space where they are linearly separable [30]. The most common activation functions are the sigmoid, hyperbolic tangent, and Rectified Linear Unit (ReLU), whose definitions are presented in Figure 2.8.

For most tasks, these neurons are organized into more complex models. One example is the multilayer perceptron (Figure 2.9), which is a neural network architecture with multiple units organized in fully connected layers, where each neuron in one layer is connected to all the neurons in the next layer. It contains hidden layers in between the input and output layers that are not observable.

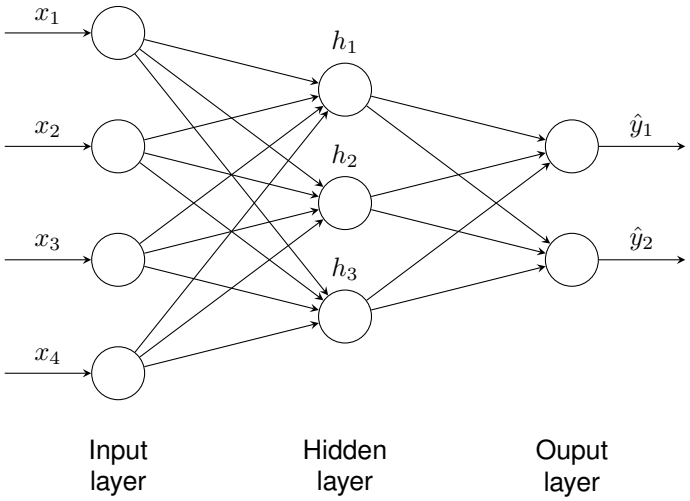


Figure 2.9: Multilayer perceptron with an input layer with 4 units, a single hidden layer with 3 units, and an output layer with 2 units. The weights and activation functions are not represented.

These more complex architectures can be trained based on the minimization of a cost function. The goal is to learn all the weights given a training set with known labels $\mathcal{T} = \{(x^{(k)}, y^{(k)}), k = 0, ..., n\}$. This is done by minimizing the cost function, that is, the total loss. In regression tasks the network is typically trained with the quadratic loss, while in classification tasks it is trained with negative log-likelihood (cross-entropy) loss. Using the quadratic error as the loss function:

$$\mathcal{C} = \sum_{k=1}^{n} L^{(k)} = \sum_{k=1}^{n} ||y^{(k)} - \hat{y}^{(k)}||^2 \tag{2.3}$$

The weights are randomly initialized, and then the minimization of the cost function is generally done using gradient descent. At each step, the weights are updated according to the equation:

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial \mathcal{C}}{\partial w_{ij}} \tag{2.4}$$

where w_{ij} is the weight connecting the input of unit i to unit j and η is the learning rate that determines the step size while moving towards the minimum of the cost function. To obtain the gradient of the loss function L that is necessary to train the weights, the chain rule of differentiation is generally used. Considering a network trained with a single example and starting from the output layer, we get:

$$\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial s_j}\frac{\partial s_j}{\partial w_{ij}} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial s_j}\frac{\partial s_j}{\partial w_{ij}} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial s_j} z_i = \epsilon_j z_i \tag{2.5}$$

where z_i is the output of the previous unit i, s_j is the input of the activation function in the current unit j and $\hat{y}$ is the only output of the network. To calculate the weights for the rest of the network, the same method is applied to the hidden layers, accounting for the fact that the output of each unit in the hidden layer contributes to the output (and therefore loss) of multiple units:

$$\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial s_j}\frac{\partial s_j}{\partial w_{ij}} = \frac{\partial L}{\partial s_j} z_i = \sum_{\substack{k \in \text{next} \\ \text{layer}}} \frac{\partial L}{\partial s_k}\frac{\partial s_k}{\partial z_j}\frac{\partial z_j}{\partial s_j} z_i = \sum_{\substack{k \in \text{next} \\ \text{layer}}} w_{jk}\epsilon_k \frac{\partial z_j}{\partial s_j} z_i \tag{2.6}$$

The update of the weights is done from the last layer to the first layer, and so this process is called backpropagation [31]. To achieve faster convergence and reduce the computational cost, the weights are most often updated using the Stochastic Gradient Descent (SGD) algorithm, which is an extension of the gradient descent of Equation 2.4 that calculates an expectation of the gradient using a smaller set of samples (minibatch) instead of the whole training set. There are other strategies to accelerate the learning process even more and prevent the algorithm from being stuck in suboptimal points in the cost function, including the addition of momentum or the use of other optimizers with adaptive learning rates, such as Adam, AdaGrad or RMSprop [32].

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) process and classify data that can be arranged in a grid-like topology. For example, they can process image data, which is a 2D grid of pixels, or signals (like the ECG and PCG used in this work), which are arranged into a 1D grid of evenly spaced samples [32].

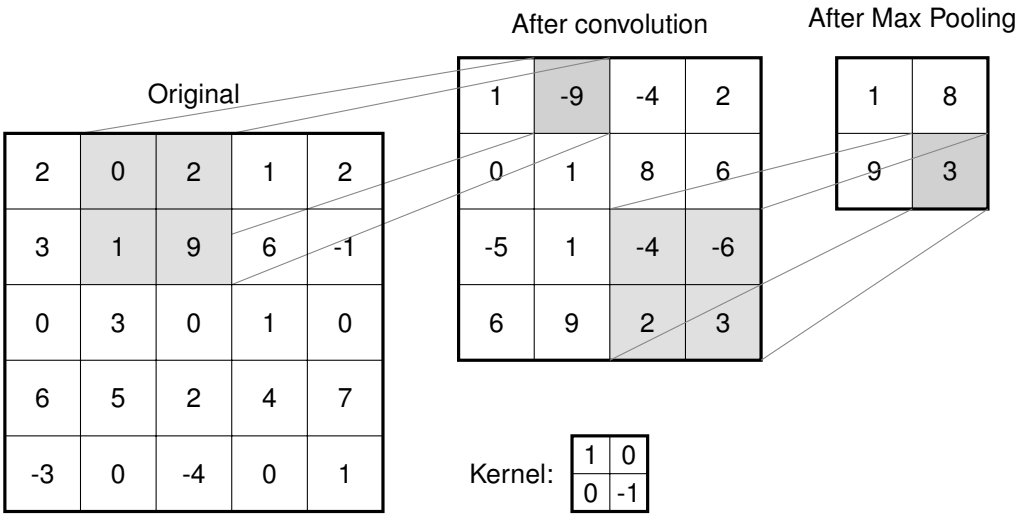


Figure 2.10: Convolution with a 2x2 kernel of stride 1, and max pooling with 2x2 kernel of stride 2.

In the 2D case, CNNs typically receive an image as input and predict either another output image or a label by extracting the most relevant features through a sequence of several layers. These layers include convolutional layers, pooling layers, which reduce the dimensionality, and fully connected layers in the case of classification tasks.

A convolutional layer applied to a 2-D data representation (e.g. an image) receives a 3D input (a collection of 2D arrays called channels), filters it by convolution with a set of kernels, applies an activation function, and then outputs a feature map. The same reasoning can be applied to 1-D data (e.g. an audio signal), the main difference being that the convolution is done over the time axis. As shown in Figure 2.10, a kernel is a matrix which is slid across and multiplied with the input to extract relevant features. The values in the kernel matrix, that determine which parts of the output will be enhanced, are learned during the training of the network [32]. CNNs can therefore automatically learn the most useful features for each task.

A pooling layer downsamples the feature map, reducing dimensionality and making the representation more robust to changes in the position of the features in the image. The most common types of pooling are max pooling, in which the feature maps are divided into rectangular blocks of the same size and only the maximum value is preserved, and average pooling, in which the average value in each block is preserved.

Unlike multilayer perceptrons which use fully connected layers, CNNs typically have sparse interactions, since most of the times the kernel is smaller than the input. The network needs to learn a small amount of parameters for each feature map, which greatly reduces the memory requirements and improves efficiency [32]. This ability to learn complex patterns in the data with simple parameters and filters also prevents overfitting.

Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are a family of deep learning neural networks designed for the processing of sequential data. They can process long sequences of the form $x^{(1)}, ..., x^{(t)}$ and retain past or future information. Most can also process sequences of variable length. A RNN model can be observed in Figure 2.11 as an unfolding of recursive networks into a graph that has a repetitive structure with modules organized in a chain [32]:

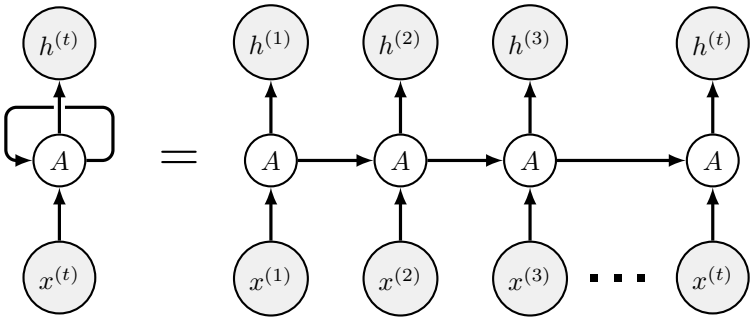


Figure 2.11: Recurrent Neural Network diagram.

RNNs allow information to flow in the sequence from one sample to the next. This information is called the hidden state, and its shape and dimension are hyperparameters. For each sample $x^{(t)}$ in the sequence, the network returns a hidden state $h^{(t)}$ that is passed to the next module. The hidden state at each time step is given by:

$$h^{(t)} = g(W_h \cdot [h^{(t-1)}, x^{(t)}] + b_h) \tag{2.7}$$

where W is the weight matrix, b is the bias term, and g is the nonlinear activation function (most commonly the hyperbolic tangent). It contains information both from the current time step and from the previous time steps. RNNs can produce a single output after reading the entire sequence or, in the case of sequence segmentation tasks, produce an output for each time step [32]. The output at time t is computed from the hidden state as:

$$\hat{y}^{(t)} = g(W_{\hat{y}} \cdot h^{(t)} + b_{\hat{y}}) \tag{2.8}$$

This RNN model has a "causal" structure, meaning that each output only depends on the current time step and on the previous timesteps. However, often it is interesting to have a network that can exploit both past and future information for tasks where the correct interpretation of the data also depends on the future values in the input sequence [32]. This is particularly relevant in the ECG and PCG, as it could better retain the temporal dependence between different structures in the signals (which is also related with disease manifestations). Bidirectional RNNs address this problem by processing the sequence in both directions with separate hidden layers, one forwards and one backwards (Figure 2.12). The output of each time step is obtained by combining the hidden states of these two layers. Assuming the hyperbolic tangent as the activation function for the hidden states:

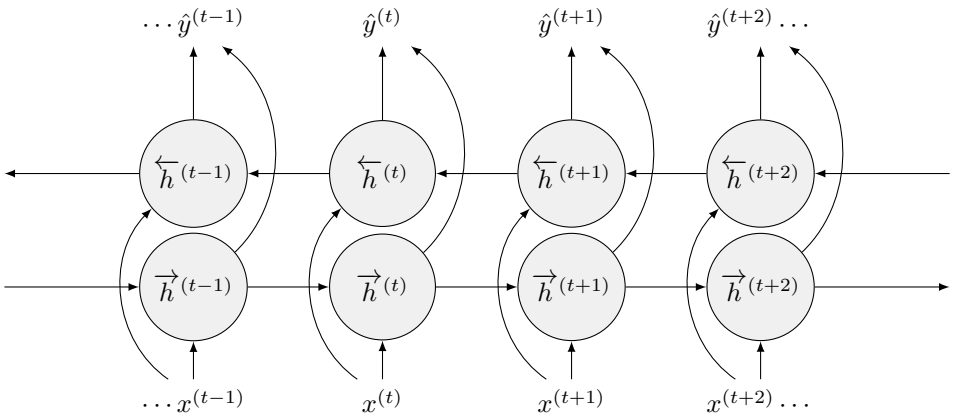


Figure 2.12: Bidirectional Recurrent Neural Network architecture. The hidden states from the forward and backwards layers are concatenated before computing the final output.

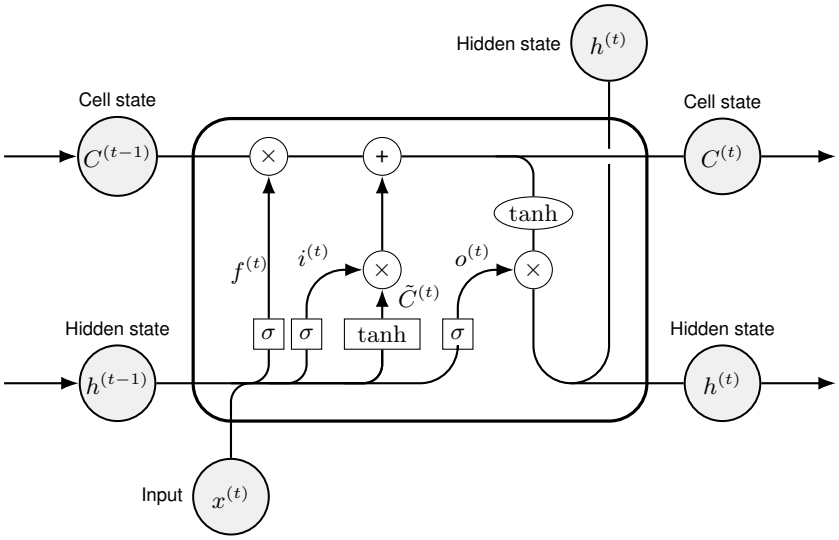


Figure 2.13: Long Short-Term Memory cell.

$$
\begin{aligned}
\overrightarrow{h}^{(t)} &= \tanh(W_{\overrightarrow{h}} \cdot [\overrightarrow{h}^{(t-1)}, x^{(t)}] + b_{\overrightarrow{h}}) \\
\overleftarrow{h}^{(t)} &= \tanh(W_{\overleftarrow{h}} \cdot [\overleftarrow{h}^{(t-1)}, x^{(t)}] + b_{\overleftarrow{h}}) \\
\hat{y}^{(t)} &= g(W_{\overrightarrow{h}\,\hat{y}} \cdot \overrightarrow{h}^{(t)} + W_{\overleftarrow{h}\,\hat{y}} \cdot \overleftarrow{h}^{(t)} + b_{\hat{y}})
\end{aligned}
\tag{2.9}
$$

For the training of the networks, gradients are calculated in the same way as feed-forward networks with the backpropagation algorithm. However, this leads to an issue known as short-term memory where RNNs can't learn the long-term dependencies in the data. Gradients propagated over the many layers in the sequence tend to exponentially shrink and vanish [32], which makes it so RNNs have difficulty in carrying information from the earlier time steps in the sequence to the later ones [33].

A solution to this is the use of gated RNNs, namely Long Short-Term Memory (LSTM) networks. They include gates that regulate the flow of information, and can not only accumulate information over a long duration, but also learn which old states to forget. They have a chain-like structure similar to the one of RNNs, but with repeating cells like the one in Figure 2.13.

Each of these cells contains a cell state $C^{(t)}$ which will remember values over arbitrary periods of time

and is regulated by three gates: the forget gate $f^{(t)}$ and the input gate $i^{(t)}$, which decide, respectively, which information to remove and add to the cell state; and the output gate $o^{(t)}$, that determines the value of the output of each cell (the hidden state) [34]. They are calculated as:

$$\begin{aligned} f^{(t)} &= \sigma(W_f \cdot [h^{(t-1)}, x^{(t)}] + b_f) \\ i^{(t)} &= \sigma(W_i \cdot [h^{(t-1)}, x^{(t)}] + b_i) \\ o^{(t)} &= \sigma(W_i \cdot [h^{(t-1)}, x^{(t)}] + b_o), \end{aligned} \tag{2.10}$$

where W and b are, respectively, the learned weights and bias of the network; σ and $\tanh$ are, respectively, the sigmoid and hyperbolic tangent activation functions; $h^{(t-1)}$ is the hidden state of time $t-1$, at the output of the previous LSTM cell; and $x^{(t)}$ is the input of the cell at the current time step t. The updated cell state and hidden state are then calculated as:

$$\begin{aligned} \tilde{C}^{(t)} &= \tanh(W_C \cdot [h^{(t-1)}, x^{(t)}] + b_C) \\ C^{(t)} &= f^{(t)} C^{(t-1)} + i^{(t)} \tilde{C}^{(t)} \\ h^{(t)} &= o^{(t)} \tanh(C^{(t)}) \end{aligned} \tag{2.11}$$

The hidden states at the output of each cell are used for the classification of each time step. They contain information of the current time step (from the output gate) and of the most relevant previous time steps (from the cell state). Both LSTMs and Bidirectional Long Short-Term Memory (BiLSTM) networks have achieved great performance in multiple application domains, including handwriting and speech recognition, handwriting generation, and machine translation [32].

2.5 Classification Metrics

To properly interpret the usefulness and impact of automatic classification methods in a clinical setting we must have a thorough understanding of their performance. This evaluation should be performed on a held-out representative set of the data, which is used for testing after the training and tuning processes of the model are completed [35]. In the case of a binary classification problem, we can calculate the relevant metrics based on four different quantities, which are represented in Figure 2.14. True Positives (TP) and True Negatives (TN) represent the number of correctly classified samples for the positive and negative class, respectively. In the same reasoning, False Positives (FP) and False Negatives (FN) represent the number of samples that were misclassified as belonging to the positive and negative classes, respectively. Using these values, several metrics can be calculated. Some have different interpretations in multi-class classification problems, but the underlying principles still apply [35]:

- Accuracy, defined as the ratio between the number of correctly classified samples and the total number of samples. This is the most commonly used metric in machine learning applications, but assessing the method's performance based only on the accuracy can be misleading when there are rare outcomes (for example, when the goal is to detect a disease with low prevalence). In this case, a high accuracy could be achieved by a simply predicting the majority class for all samples,

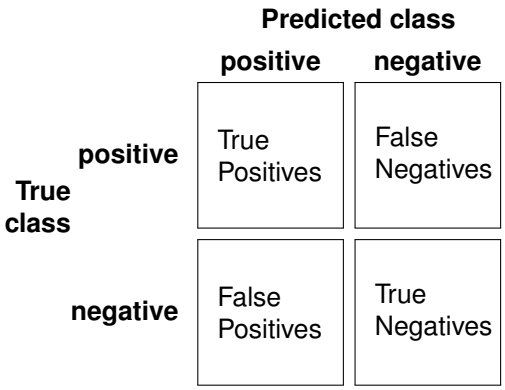


Figure 2.14: Example of confusion matrix for binary classification problem.

which obviously isn't useful in any clinical application [36].

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{2.12}$$

- Sensitivity (also known as recall or true positive rate), defined as the ratio of positive samples which were correctly classified. This is one of the most important metrics in medical applications, since typically the goal is to miss as few positive cases as possible.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{2.13}$$

- Specificity (also known as true negative rate), defined as the ratio of negative samples which were correctly classified.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{2.14}$$

- Precision (or positive predictive value), which calculates the ratio between the positive samples which were correctly classified and all the samples assigned to the positive class. This metric is impacted by the prevalence of the condition.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2.15}$$

- F1-score, calculated as the harmonic mean between sensitivity and precision. It is particularly useful when there is a class imbalance, and it is asymmetric: if there is a large positive class to which the classifier is biased, the F1-score will be high; on the other hand, if the negative class is the majority and the classifier is biased, the F1-score will be low [35].

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}} \tag{2.16}$$

To have a correct interpretation of a classifier's performance, a machine learning study should report

a sufficient number of metrics [35]. For example, a classifier can have high accuracy, high sensitivity, and high specificity, but still have low precision if the disease has a very low prevalence (resulting in a high number of false positives). Whether the goal is to minimize false negatives or false positives, there is always a balance, and it depends on the application. In some situations it is acceptable to have a high number of false positives (for example, in cancer screenings), and thus it is more important to maximize the sensitivity. In other cases, when the costs or risks of follow-up therapy are too high and surpass the potential benefits, it is more important to have a high precision.

Chapter 3

State of the Art

In this section a review of the state of the art in automatic signal classification based on the ECG (Section 3.1) and PCG (Section 3.2) and on the multimodal signals (Section 3.3) is provided. Given our focus on auscultation, a review of the available datasets with heart sounds data (Section 3.4) and of the current devices for the simultaneous acquisition of both signals (Section 3.5) is also presented.

3.1 ECG Automatic Screening

The ECG is one of the most used diagnostic tools, and as such its automatic analysis is already a mature field. With the recent development of portable and low cost ECG monitoring tools for personal use, there has been a high influx of data that called for the use of deep learning methods for classification; these usually show superior performance to traditional machine learning methods [37] and don't require expert knowledge to extract handcrafted features or high quality data [38].

ECG disease and abnormality detection can be divided into different tasks. Arrhythmia detection is one of the most common, where automatic methods can identify atrial flutter, supraventricular tachyarrhythmia, or ventricular trigeminy, for example. Atrial fibrillation is also an arrhythmia but is generally regarded as a special case [38] due to its high prevalence. Other conditions such as congestive heart failure [39] or acute myocardial infarction [40] can also be automatically detected from the ECG using deep learning. These classification tasks focus on the detection of abnormal, irregular beats, and can use binary labels (such as normal of abnormal beats) or multi-class labels to diagnose specific diseases [38], although the latter often requires multi-channel data so the model can interpret the overall behavior of the ECG.

The most commonly used neural network architectures for ECG classification are CNNs, RNNs, or hybrid models. The systematic review article in [38] collected 191 papers that used deep learning strategies for several ECG classification tasks. CNNs comprised the majority of the methods for disease detection, and two types of CNN were used: 1D CNNs, which can apply the kernel directly on the signal along the temporal dimension, and 2D CNNs, that have to apply the kernel on transformed ECG data. For example, on the MIT-BIH arrhythmia database, Huang et al. [41] used a 2D CNN to classify five types of arrhythmias based on time-frequency representations of ECG signal segments, reaching sensitivities above 98% for all the classes; and Li et al. [42] proposed a 2D CNN to classify heartbeats that uses common features such as morphology, RR intervals and beat-to-beat correlation of ECG segments structured as 3D inputs, achieving an overall accuracy of 91.44% in the inter-patient case. In the case of 1D CNNs, for example, in [43] a 1D CNN was applied on denoised ECG signals to classify the

different types of arrhythmias with an overall accuracy of 97.5%, and in [44] the authors developed an end-to-end 1D CNN to build an atrial fibrillation detection system. While 1D CNNs don't require as much pre-processing, they still have the disadvantage of needing to use signal segments with a fixed length.

RNNs process sequential data and are a natural choice for ECG classification, since they can capture temporal dependencies more efficiently than other types of neural networks [45] and process signals of variable length. They also have the ability to use the raw ECG signals as input without preprocessing [46]. Saadatnejad et al. [45] developed two small LSTMs for patient-specific monitoring that can be used in real-time and identify seven arrhythmia classes. Li et al. [47] combined a BiLSTM with an attention mechanism for arrhythmia classification from ECG beats and RR intervals information, achieving an accuracy of 99.49% and sensitivities above 90% for all classes. The top-performing algorithm in the Physionet/Computing in Cardiology Challenge 2017 for atrial fibrillation classification from single-lead ECG recordings [37] also used an RNN with stacked LSTM layers fed with both global and beat features [48], reaching an average F_1 score of 83.1%.

Hybrid models combine CNN and LSTM layers and are particularly useful to process long signals with variable lengths and multi-channel inputs with more than one lead. They are typically an end-to-end approach, where CNN layers extract the local ECG features along the sequence and then the RNN layers process them along the temporal dimension to produce global features [38]. He et al. [49] combined stacked residual convolutional layers to compress the ECG signal into a feature vector, and then used BiLSTM layers to obtain a global feature vector for classification, and achieved an overall F_1 score of 80.6% on the China Physiological Signal Challenge (CPSC) dataset; Hong et al. [50] developed a hybrid CNN and RNN with a multi-level attention mechanism that uses medical domain knowledge features combined with ECG data to increase the interpretability of the model, and Guo et al. [51] used DenseNet, a densely connected CNN with a Gated Recurrent Unit (GRU) with an attention mechanism. There are other neural network architectures that have been used for ECG classification, on their own or combined with other neural network layers, such as autoencoders [52–54] and generative adversarial networks [55–57]

3.2 PCG Automatic Screening

PCG analysis and classification is a less developed field than in the case of ECG, and it is also more difficult, since PCG signals are often highly corrupted by in-band noise and artifacts from the surrounding environment (such as ambient speech, motion artifacts, intestinal or breathing sounds, etc.). Furthermore, the signals may contain other physiological sounds beyond the fundamental heart sounds (such as S3, S4, split heart sounds, murmurs, etc.) that complicate the analysis [27].

They are high-dimensional audio signals and most classification methods cannot process them directly, so they generally need to be converted into lower dimensional features [33]. The feature extraction stage is therefore one of the most important steps for a good classification performance [58]. One of the most common feature extraction methods involves the use of Mel-Frequency Cepstral Coefficients (MFCCs). They are coefficients of the Mel-frequency cepstrum and one of the most used features

for automatic speech recognition and audio classification. As the name suggests, they are based on a cepstrum with a non-linear frequency axis that follows the Mel scale [59]. Other frequency domain features include spectrograms based on both the short-time Fourier transform and the discrete Wavelet transform. Together with the MFCCs, they were the most used features in the Physionet/CinC Challenge 2016, that had the objective of developing algorithms to automatically identify whether a heart sound recording is normal or abnormal [58]. Time-domain features typically focus on the extraction of lower dimension envelopes from the 1D time series [33].

Most PCG classification methods only perform a binary classification between normal and abnormal signals due to the shortage of available datasets (which will be explored in Section 3.4), with deep learning approaches focusing on CNNs, RNNs, or hybrid methods, similarly to the ECG. Most deep learning-based methods do not apply a segmentation algorithm to identify the fundamental heart sounds, but can achieve a good performance nonetheless [33].

Typically, 2D CNNs are applied to time-frequency feature maps extracted from the signals. Rubin et al. [60] used a deep CNN and time-frequency maps based on MFCCs of 3 second-segments starting at S1 (after applying a segmentation algorithm). In [61] the authors used a CNN with 5-second windows of power spectral density features organized in a spectrogram. These methods reached the 7th and 11th place in the challenge, respectively, with balanced accuracy scores of 83.99% and 81.11%. Maknickas et al. [62] used Mel-frequency spectral coefficients, which are derived from MFCCs, as features of a CNN and reached a balanced accuracy of 84.15% on the test set of the challenge. On a different dataset, Demir et al. [63] used pre-trained CNN models (AlexNet, VGG16, and VGG19) to identify whether heart sound recordings have murmurs, extra systoles, or artifacts. The signals were converted into spectrogram images with three channels.

The feature maps used with 2D CNNs allow an accurate representation of the acoustic properties of the heart sounds, but they require an additional pre-processing step and the selection of proper hyperparameters. 1D CNNs can be applied with fewer parameters and without complex pre-processing [33]. In [64], Xu et al. applied a 1D deep CNN with bidirectional connections that classifies raw PCG patches and scored 90.46% on the challenge test set with much fewer parameters. In [65] the authors developed a 1D CNN that classifies PCG signals into normal or abnormal, based on deep features extracted by a denoising autoencoder and obtained a superior performance than methods which used MFCCs.

Since PCG signals are a type of sequential data with a strong temporal correlation, they can also be effectively processed by RNNs, which have achieved great success in the modeling of temporal data even in the presence of noise [66]. RNNs, LSTMs, and GRUs have all been used in heart sounds classification, with LSTMs being the most common. Yang et al. [67] were the first to use a RNN to classify heart sounds. They used GRU layers with frequency domain augmented features and reached a score of 80.18% on the hidden test set of the 2016 Physionet/CinC challenge. Latif et al. [66] evaluated different types of RNNs that used MFCCs from PCG segments as input, including LSTM, GRU, and their bidirectional versions. BiLSTMs achieved the best score with a balanced accuracy of 98.61% on a subset of the training data in the same dataset. In [68] a LSTM was used to distinguish between normal signals, signals with murmurs, or with extra systoles. They used downsampled versions of the signals

as input and achieved an accuracy of 80.80%.

Most hybrid methods combine CNNs and RNNs: convolutional layers are more effective at feature extraction, while recurrent layers can take advantage of the high temporal correlation in the signals to increase their predictive power [33]. In [69] a 1D CNN-BiLSTM network was used to classify valvular heart diseases in PCG recordings, and allowed the detection of both the spatial and temporal information in the signals. It achieved a good performance both on the PhysioNet/CinC 2016 challenge dataset with an accuracy of 87.31% and on the dataset described in [70], with an accuracy of 99.50%, sensitivity of 97.56% and specificity of 100%. Deng et al. [71] applied parallel CNN and BiLSTM blocks that used MFCCs as input. This way, the neural network could take advantage of both the spatial characteristics extracted by the CNN and the temporal characteristics learned by the RNN, and it achieved a balanced accuracy above 98% in the PhysioNet/CinC 2016 challenge dataset with cross-validation on the training set. Noman et al. [72] combined a 1D CNN with a 2D CNN as a way to learn multiple levels of representations from the heart sounds: the 1D CNN captures the temporal and morphological features from the raw waveform, while the 2D CNN captures the spectral features from the MFCCs.

Deep learning approaches have been shown to be more efficient and accurate than traditional machine learning methods, but they are also more affected by limited datasets. There is always the risk of overfitting, since variations in recording equipment, environmental noise, and auscultation locations generate signals with vastly different morphologies and amounts of noise contamination [33], which can lead to a decrease in classification accuracy.

3.3 Simultaneous ECG and PCG Analysis

There are very few studies that explore the performance improvements achieved by combining the PCG and ECG in a simultaneous analysis, and most are based on traditional machine learning methods. As was mentioned in Section 2.3, these signals are complementary and may contain mutually exclusive information about the heart [7], so by combining them it is conceptually possible to improve automatic disease screening. A summary table with the highest scores obtained for each method is presented Table 3.1.

Chakir et al. [8] implemented traditional machine learning methods (naive Bayes, K-Nearest neighbors, support vector machine, random forest, and linear discriminant analysis) for the automatic detection of cardiac abnormalities from simultaneous PCG and ECG data. They demonstrated that there is a significant increase in performance when combining the multimodal signals comparatively to using just the PCG. Singh et al. [73] also tested traditional machine learning classifiers using R-R intervals and R peak values as features for the ECG and the homomorphic and Hilbert envelopes combined with power spectral density features for the PCG. When testing support vector machines, K-nearest neighbors, and an ensemble classifier, they found that the classification based on the combination of the PCG and ECG features outperforms the other tested methods that use only one modality.

Hettiarachchi et al. [7] used a transfer learning approach to deal with the issue of the limited publicly available datasets with simultaneous PCG and ECG signals. They developed a deep dual CNN, with

Table 3.1: Comparison of performance metrics of methods for simultaneous ECG and PCG classification in the state-of-the-art. All used the same dataset, from the Physionet/Cinc challenge 2016.

	Data	Accuracy	Sensitivity	Specificity	Positive Predictive Value	F_1 score
Chakir et al. [8]	100 samples. 60% training.	92.5	92.31	92.86	-	-
Singh et al. [73]	342 samples. 70% training.	93.13	94.44	90.00	-	-
Hettiarachchi et al. [7]	405 samples. 60% training.	90.41	94.74	75.00	-	-
Li et al. [74]	407 samples. 60% training.	86.42	84.96	93.10	98.26	91.13

two distinct CNN blocks being trained on individual PCG and ECG signals from large datasets. The learned parameters were then transferred to an integrated CNN for multi-modal classification. They used scalograms as features for both signal modalities and demonstrated a good performance, on-par with other deep learning studies for uni-modal classification.

To our knowledge, the most recent work was developed by Li et al. [74], where a support vector machine was used to classify the multi-modal signals, followed by the application of Dempster–Shafer theory [75] to fuse the classification results based on the posterior probability outputs. They compared the performance metrics of the classification based on each individual signal and based on the multi-modal signals, and found that the latter were superior.

3.4 Datasets

Deep learning models with deeper layers tend to exhibit higher performance and are more flexible since they can automatically learn the features that are useful for the signal classification. However, the training of these models requires high volumes of data, that should be comprehensive and representative enough to avoid overfitting. One of the major challenges in PCG analysis is indeed the lack of comprehensive annotated datasets that encompass a wide enough variety of signals with most of the commonly encountered pathologies [25].

One of the first public heart sound databases was the PASCAL Heart Sound Challenge dataset, made available in 2011[1]. It contains 496 recordings with segmentation annotations from two different sources: from the general public via a stethoscope iPhone app, and from a clinical trial in hospitals using the digital stethoscope DigiScope. They are divided into normal, murmur, extra heart sound, artifact, and extra systole classes [76]. Although the number of the recordings is relatively large, this dataset is considered very challenging because some of the signals are highly corrupted by noise, and because the data was collected from both adults and children, with high varying heart rates (from 40 to 140 beats per minute, or even higher [76]) and very different durations for systole and diastole. Additionally, the signals have a limited frequency range (up to 195 Hertz (Hz)) due to the application of a low-pass filter,

[1] www.peterjbentley.com/heartchallenge

which removes some of the useful heart sound features for diagnosis [58].

The most widely used datasets are the ones from the Physionet CinC Challenge 2016[2]. This challenge assembled several databases from different sources within the research community, gathered from both clinical and non-clinical environments. The heart sounds were obtained from four different auscultation locations (aortic, pulmonary, tricuspid, and mitral), and from both healthy individuals and patients with a wide range of pathologies. It contains six different training sets, labeled from ”A” to ”F”, comprising a total of 3153 heart sounds [26]. The test sets are not publicly available.

They are labeled as “normal”, or “abnormal”, with an optional “unsure” class for poor quality signals. The training sets are unbalanced (with more normal than abnormal recordings). Reference annotations for the segmentation of the fundamental heart sounds (S1, systole, S2, and diastole) were obtained by applying Springer's state-of-the-art Logistic Regression (LR)-Hidden semi-Markov Model (HSMM) segmentation algorithm, followed by manual review and correction of the beat labels [26]. However, even after manual correction, these ground truth annotations remain biased towards Springer's algorithm [77]. The dataset used in [27] was also made publicly available during this challenge for testing of the LR-HSMM algorithm and contains segmentation annotations based on the analysis of synchronous ECG recordings. The data distribution of the datasets from Physionet is presented in Figure 3.1. These datasets focus mostly on binary classification and have the limitation of having multiple heart sound recordings from the same patients without any identification. This makes it so signals from the same patient are treated as independent samples when training the algorithms [78].

A new Physionet Challenge was launched in 2022 with the goal of detecting the presence or absence of murmurs from multiple heart sound recordings and multiple auscultation locations. The challenge data comes from the CirCor DigiScope dataset [3], which aimed to address the limitations of the previous datasets by collecting a large range of heart sounds and characterizing them using scales and parameters often used in clinical practice [4, 80]. It includes 5272 recordings from a pediatric population of

[2] https://physionet.org/content/challenge-2016/1.0.0/
[3] https://physionet.org/content/circor-heart-sound/1.0.3/

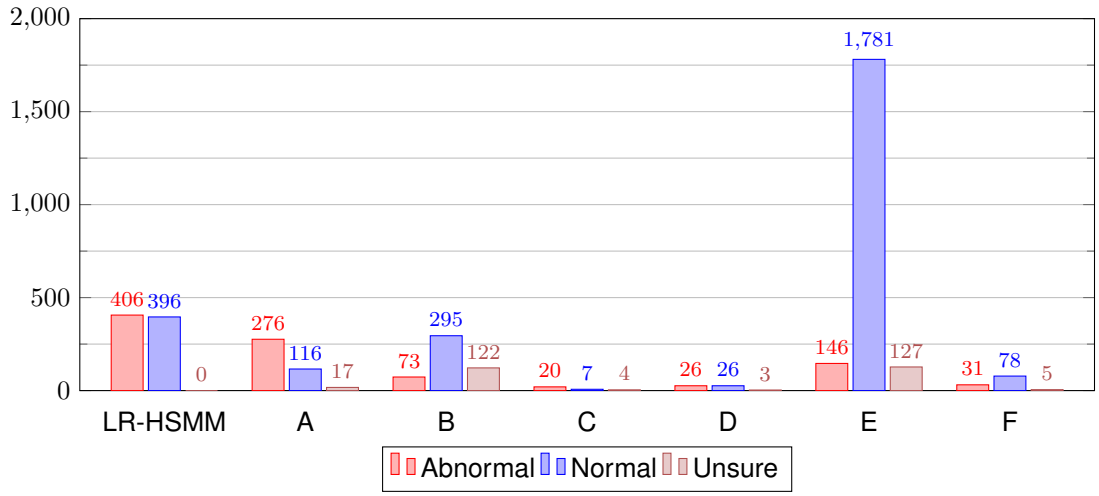


Figure 3.1: Distribution of data from the publicly available PCG Physionet datasets across the different classes. Data from [58, 79].

1568 patients, that were automatically segmented using three state-of-the-art algorithms which identified the fundamental heart sounds. These results were then provided as annotation recommendations to two independent cardiac physiologists, who inspected the annotations and re-annotated the wrong labels [4].

The CirCor DigiScope dataset is now the largest publicly available heart sound dataset and includes a very detailed characterization and classification of murmurs, including timing, pitch, grading, shape, quality, and auscultation location. The ability to use multiple heart sound recordings from the same patient also provides an opportunity to improve automatic CVD diagnosis.

Despite these recent developments in the availability of PCG auscultation data and the known potential of using multi-modal cardiac signals, the available datasets for simultaneous PCG and ECG analysis remain very limited. To our knowledge there are currently only two public datasets.

One is the training set "A" from the aforementioned Physionet CinC Challenge 2016 dataset, which contains 409 PCG recordings with simultaneous single-lead ECG from subjects with a variety of conditions: with murmurs related to mitral valve prolapse, with aortic disease, with innocent and benign murmurs, and also a normal control group. It allows an evaluation of the synchronous signals, but it still suffers from the limitations that were mentioned.

The second one resulted from the electro-phono-cardiogram (EPHNOGRAM) project for the development of a low-cost and low-power device that records simultaneous three-lead ECG and PCG data [81]. The current database [4] contains 69 simultaneous ECG and PCG recordings from 24 healthy adults with ages between 23 and 29. While it provides a great insight into the inter-relationships between the two cardiac signals, it only includes data from healthy subjects and doesn't provide any other annotations.

Dataset scarcity is an even bigger problem when taking into account specific diseases [33]. Most datasets are unbalanced among the different classes and only provide a general classification of the signals as normal/abnormal. They only include the signals, with no mention of relevant clinical information such as gender, age, history of illness, or a characterization of the heart sounds or ECG, which are essential for doctors to perform their assessment. This hinders the development and implementation of computer-aided decision systems for CVD screening in real clinical applications [4, 33].

3.5 ECG and PCG Acquisition Systems

The potential of the complementarity between the two modalities has prompted the development of specialized medical devices for the synchronous acquisition of PCG and ECG, enabled in part by recent developments in mobile health and telemonitoring [81]. Current devices are mostly based on integrating ECG electrodes in an electronic stethoscope with single-lead or three-lead ECG, enabling the two signals to be acquired at the same time. One example is the Eko DUO system [82], which can analyze heart sounds and ECG waveforms while also working as a stethoscope. It records a single-lead ECG using two stainless steel electrodes attached to the stethoscope head. Another example is the CardioSleeve, which works as an add-on device to acoustic stethoscopes and enables the recording of heart sounds

[4] https://physionet.org/content/ephnogram/1.0.3/

Figure 3.2: Commercially available devices for simultaneous ECG and PCG acquisition. From left to right: Eko DUO [87], Coala Heart Monitor [84], Rijuven CardioSleeve [88].

and a 3-lead ECG [83].

Both these devices have wireless transmission for real time signal visualization, to enable telehealth applications, and were designed so the ECG is measured on the patient's chest during the auscultation procedure. However, their form factor is still very different from the one of acoustic stethoscopes, which makes it difficult for doctors to use them in standard medical practices. Also, the ECG leads that are acquired along the auscultation points are poorly characterized [14].

The Coala Heart Monitor is a patch-like device, designed specifically for remote patient monitoring. It records the heart sounds and two ECG leads (one recorded on the thumb, and another on the chest) and can detect atrial fibrillation and other signs of heart disease [84]. The Coala, Eko DUO, and CardioSleeve devices are represented in Figure 3.2.

Klum et al. [85] also presented a wearable patch stethoscope for long-term monitoring that records one ECG lead and the PCG. They were able to estimate the Einthoven leads I and II and respiratory parameters from the multimodal signals. Lin et al. [86] developed a device that can visualize and measure heart sounds and the ECG. Unlike other devices, the ECG and heart sounds measuring parts are separate, and are placed on different regions in the patient's chest.

Other devices can also be found in literature, such as the SmartHeart device, which integrated custom ECG electrodes on the diaphragm of the 3M Littmann 3200 electronic stethoscope, to enable the simultaneous acquisition of the ECG and the PCG. It has the advantage of not altering the conventional stethoscope form factor, but it still relies on a high-end stethoscope that has since been discontinued. Furthermore, given that each modality is acquired by a separate device, synchronization between the ECG and PCG signals is a challenge. In this work, the authors found that ECGs measured at auscultation points further away from the heart correlate more with the precordial leads V1 and V2, while ECGs measured closer to the heart correlate more with leads V4, V5 and V6. Even though the waveforms were different than the ones obtained with a 12-lead ECG, they were considered similar enough to be meaningful [14].

Chapter 4

Heart Disease Detection with ECG and PCG

Given the success of deep learning approaches for PCG and ECG classification, this chapter focuses on the application of deep neural networks, namely CNNs, RNNs, and hybrid models, to two different tasks: detection of heart murmurs and clinical outcome from PCG signals (Section 4.1), as part of the George B. Moody Physionet Challenge 2022, and abnormality classification based on simultaneous PCG and ECG data (Section 4.2). A discussion of the results is then presented in Section 4.3.

4.1 Heart Murmur Detection and Outcome Prediction using PCG

The goal of the George B. Moody Physionet Challenge 2022 was to develop an automatic method that could identify the presence, absence or unsure cases of murmurs and the normal vs. abnormal clinical outcome from heart sound recordings collected from multiple auscultation locations. It used the CirCor DigiScope dataset, which consists of 5272 PCG recordings from 1568 patients in a pediatric population. This dataset was divided into training, validation, and test sets, with 60%, 10%, and 30% of the data, respectively. The training set was publicly released, while the validation and test sets were hidden and used to evaluate the entries of the challenge [89].

The PCG signals were acquired using the 3M Littmann 3200 stethoscope in the four main cardiac auscultation spots (aortic valve, pulmonary valve, tricuspid valve, mitral valve) in a sequential manner, and with a sampling frequency of 4000 Hz. However, some of the patients have recordings from fewer auscultation locations, and some have more than one recording per location. The duration of the recordings also varies between patients and locations, spanning from 4.8 seconds to 80.4 seconds [4].

The murmur labels (Present, Absent, and Unknown) were classified by a single cardiopulmonologist independently of other clinical annotations. They indicate whether the expert was able to detect the presence or absence of a murmur in a patient, or whether they were unsure [89], due to the recordings not meeting the required signal quality standards [80]. The identified murmurs were also characterized according to their timing (early-, mid-, and late systolic/diastolic), shape (crescendo, decrescendo, diamond, plateau), pitch (high, medium, low), quality (blowing, harsh, musical), and grade according to Levine's scale [4] (which defines murmur intensity). It should also be noted that the murmur labels correspond to the patient, and not to the recordings. Thus, it is possible that a patient in the Present class has some recordings where no murmurs have been detected. The auscultation locations where at least

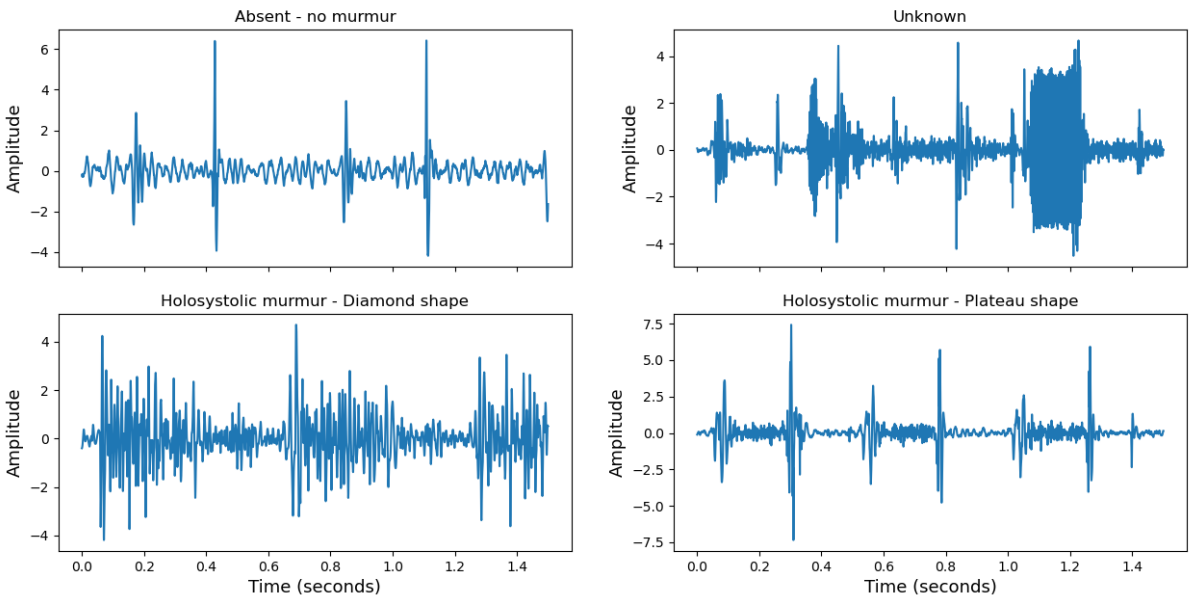


Figure 4.1: Examples of recordings with Absent, Present, and Unknown labels in the CirCor DigiScope dataset. The signals were filtered and normalized to have zero mean and unit variance.

one murmur has been heard are included in the provided annotations. Example signals for the three classes are presented in Figure 4.1.

The clinical outcome annotations indicate the normal or abnormal clinical outcome diagnosed by a different medical expert, who did not have access to the digital auscultations or the murmur labels. It is important to note that this diagnosis results from an overall assessment of the patient's condition based on multiple examinations (such as clinical history, physical examination, analog auscultation, echocardiogram, etc.) and as such does not necessarily imply that the expert identified a murmur [80]. Of the patients that are labeled as having an abnormal clinical outcome in the training set, 32.9% had an audible murmur, 57.7% didn't have any audible murmur, and 9.4% belonged to the Unknown class. For the patients with a normal clinical outcome these percentages are, respectively, 6.0%, 88.9%, and 5.1%. While it is more likely that a patient with an abnormal outcome has a murmur, using the presence of murmurs to predict the clinical outcome would lead to poor results.

A noteworthy aspect is that murmur classes are highly imbalanced, which presents a challenge for the training of deep learning models. The training set contains data from 942 patients; of those approximately 73.8% belong to the Absent class, 19% belong to the Present class, and only 7.2% belong to the Unknown class. On the other hand, the outcome classes are balanced between Normal (51.6%) and Abnormal (48.4%).

To evaluate the performance of the algorithms, two metrics were introduced in the challenge. In the murmur detection task, this evaluation was done with a weighted accuracy metric that assigns more weight to patients that have or potentially have murmurs than to patients that do not have murmurs, indicating that a missed diagnosis is worse than a false positive. Samples belonging to the Present class have five times the weight of samples belonging to the Absent class, and samples in the Unknown class have three times the weight of samples belonging to the Absent class. In the outcome detection

task, a cost-based metric was used. It reflects the cost of human screening, treatment, and of missed diagnosis [89], and it is defined as:

$$C = \frac{f_{algorithm}(n_{patients}) + f_{expert}(n_{TP} + m_{FP}, n_{patients}) + f_{treatment}(n_{TP}) + f_{error}(n_{FN})}{n_{patients}} \tag{4.1}$$

$$f_{algorithm}(s) = 10s, \qquad f_{treatment}(s) = 10000s, \qquad f_{error}(s) = 50000s, \tag{4.2}$$

$$f_{expert}(s,t) = (25 + 397\frac{s}{t} - 1718\frac{s^2}{t^2} + 11296\frac{s^4}{t^4})t \tag{4.3}$$

The submitted algorithms were evaluated using these metrics, with repeated scoring on the hidden validation set and one-time scoring on the hidden test set. This is now the largest publicly available heart sound dataset, and includes a very detailed characterization of murmurs, as well as several auscultation locations per patient. Although it doesn't include recordings of simultaneous ECGs, it could help significantly enhance the characterization and classification of the heart sounds, and consequently improve automatic CVD diagnosis based on auscultation.

4.1.1 Data Preprocessing

As was mentioned, feature extraction algorithms are generally necessary to process PCG recordings because the raw audio signals are high-dimensional sequence data, and most methods cannot process them directly. Two types of features were extracted from the heart sound recordings: spectral features, namely MFCCs (which have been established as being able to efficiently capture the frequency content of the signals for classification), and morphological and temporal features based on signal envelopes, that capture the variability in morphology that may be indicative of abnormalities.

Initially, all signals were filtered using a 2nd order Butterworth bandpass filter with cutoff frequencies of 25 and 400 Hz and went through a spike removal process, to remove unwanted noise and friction spikes [90]. For the spectral features, the static, delta, and delta-delta MFCCs were extracted. This cepstral transformation allows the identification of periodic structures in the audio and the separation of signal components combined by convolution in the frequency space. The Mel scale is a measure that models the subjective pitch and frequency content of audio signals [59], as perceived by humans. It is calculated as

$$Mel(f) = 2595 \log_{10}(1 + \frac{f}{700}) \tag{4.4}$$

where f is the normal frequency scale. MFCCs are calculated by applying the Discrete Cosine Transform (DCT) to the mapping of the Fourier transform power coefficients into the Mel scale (which is obtained by using a triangular bandpass filter bank). This scale is approximately linear for frequencies below 1 kHz and logarithmic for frequencies above 1 kHz [91] (similarly to the human ear), which is useful for

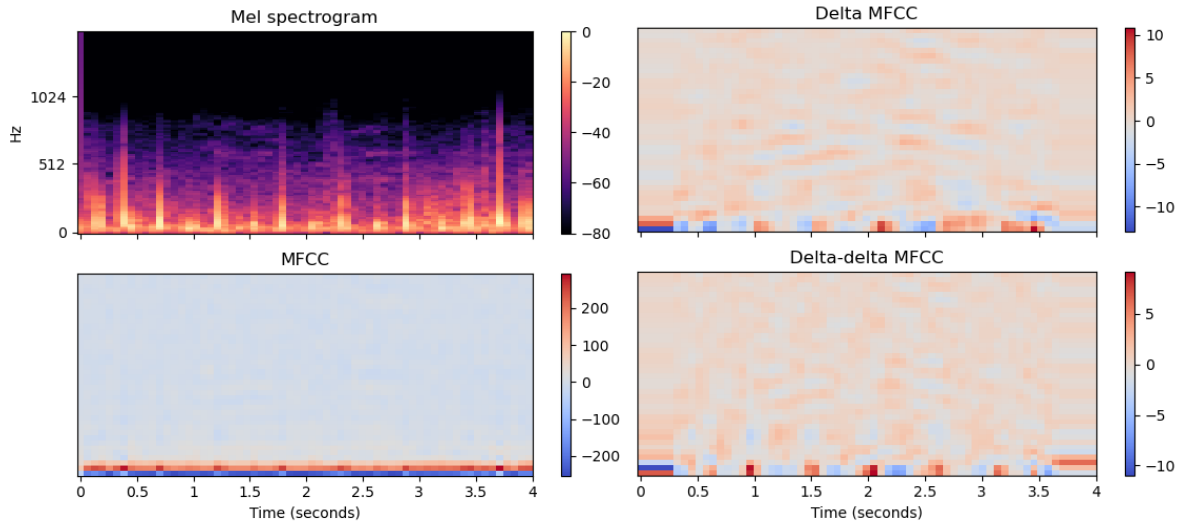


Figure 4.2: Mel spectrogram, MFCCs, delta, and delta-delta coefficients for PCG signal without murmur.

the analysis of PCG signals, since heart sounds and murmurs are concentrated in the lower frequency bands. By using this time-frequency representation the dimensionality of the PCG audio signals is reduced, while retaining the most relevant information for pattern recognition.

To calculate MFCCs for the heart sounds, the librosa 0.9.2 Python library [92] was used, with the following method:

1. Calculate the spectrogram of the signal using a Hann window, with length of 128 milliseconds and overlap of 64 milliseconds.

2. Apply a Mel filterbank to project the FFT bins into the Mel-frequency bins and sum the energy in each filter (to obtain the Mel spectrogram).

3. Convert the Mel spectrogram to a log-scale and apply the DCT type-2 to produce a set of cepstral coefficients.

4. Select the first 40 coefficients and discard the rest.

These coefficients are referred to as static features because they contain information from a single frame. To also include the dynamic information of the signals, the delta and delta-delta coefficients (that correspond, respectively, to the first and second order derivatives of the MFCCs [93]) were computed. In Figures 4.2 and 4.3 we can see the Mel spectrogram, MFCCs, and the delta and delta-delta coefficients for example 4-second segments of a normal signal and of a signal with a murmur. In the normal signal, the high-energy peaks corresponding to S1 and S2 can be clearly identified, while in the signal with the murmur the high energy regions are more spread out throughout the duration of the signal.

For the temporal features, the signals were downsampled to 1000 Hz prior to the extraction of four different envelopes from the PCG:

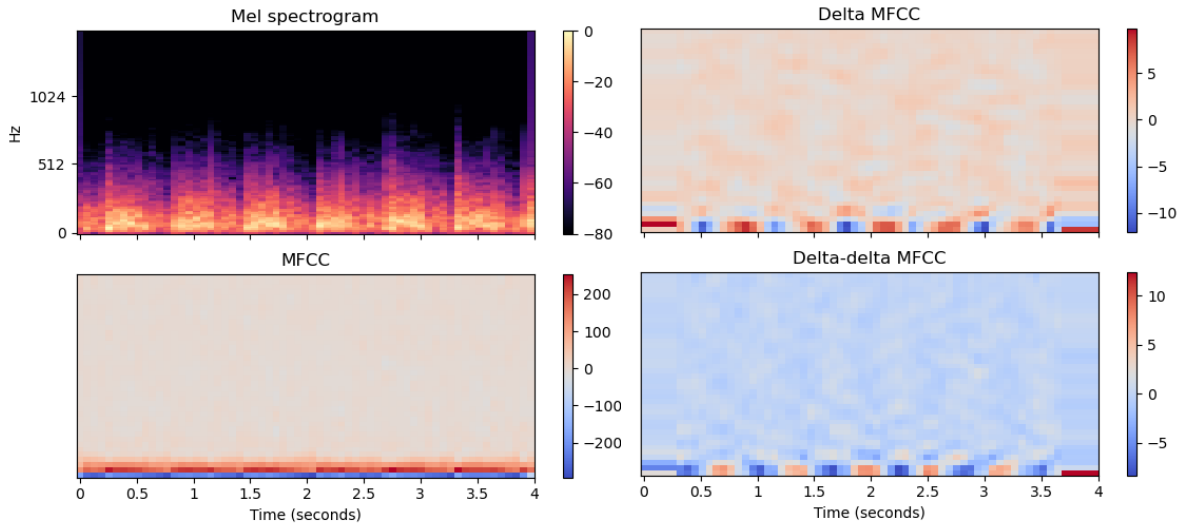


Figure 4.3: Mel spectrogram, MFCCs, delta, and delta-delta coefficients for PCG signal with murmur.

1. Homomorphic envelope, obtained by calculating the exponential of the low-pass filtered natural logarithm of the signal [90]. A Butterworth low-pass filter with a cutoff frequency of 8 Hz and order 1 was used. It is highly effective in the removal of high frequency noise and has been used in numerous studies for the localization of S1 and S2 [94].

2. Hilbert envelope, calculated as the absolute value of the Hilbert transform of the PCG.

3. Power Spectral Density envelope, calculated by finding the mean power spectral density over the frequency range of interest (between 40 and 80 Hz, based on the frequency content of S1 and S2) [27]. A Hamming window with a length of 25 milliseconds and 50% overlap was used.

4. Wavelet envelope, calculated by taking the absolute value of the wavelet coefficients, which are obtained by applying the discrete wavelet transform at decomposition level 3 using reverse biorthogonal wavelets.

The envelopes were then downsampled to 50 Hz and normalized to have zero mean and unit variance (Figure 4.4). This feature extraction method is the one developed by Springer et al. [27], and was implemented in Python based on the open-source code available online[1]. The envelopes form a 4-dimensional multivariate time series for each heart sound segment and, as it can be seen in Figure 4.4, have different trade-off levels between noise rejection and amplitude resolution [5]. They can not only provide a more compact description of the signals [95], but also reduce noise and other effects specific to the recording environment [96].

Prior to feature extraction, each recording was decomposed into smaller fixed-length segments of 4-seconds. These segments still contain multiple cardiac cycles and enough information for the model to learn, and at the same time are small enough to allow a significant increase in the amount of training

[1] https://github.com/davidspringer/Springer-Segmentation-Code

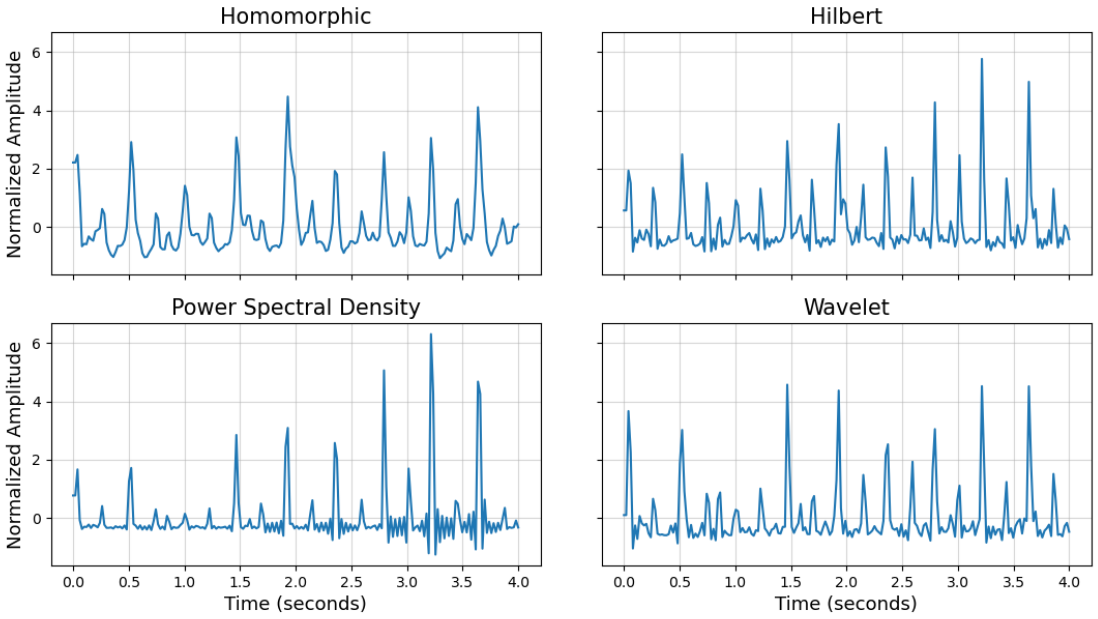


Figure 4.4: Homomorphic, Hilbert, power spectral density, and wavelet envelopes of a 4-second segment of PCG signal.

data to build a more robust model [66]. To deal with the class imbalance in the murmur detection task, instead of removing samples from the majority class (which are an important source of variability to train the models), the minority classes (Present and Unknown) were oversampled by extracting their segments with 75% overlap. Recordings from patients in the Present class without an audible murmur were excluded from training, to avoid biasing the neural networks to false positive errors.

4.1.2 Neural Network Architectures

Three different neural network architectures were employed for the heart sounds classification: a CNN, which receives the static, delta, and delta-delta MFCCs as input, organized into an array with three channels; a RNN, which receives the four envelopes of the PCG signal as input; and a hybrid neural network obtained by combining the CNN and RNN as parallel blocks. With this multi-input hybrid model, the two blocks can complement each other, with the CNN extracting the most relevant spectral features and the RNN extracting the morphological and temporal information. A diagram of the models is presented in Figure 4.5.

The CNN consists of two convolutional layers, each followed by a max pooling layer for downsampling and a dropout layer for regularization to avoid overfitting. The RNN consists of two stacked BiLSTM layers. Unlike standard RNNs, LSTMs can remember or forget values over arbitrary periods of time, and thus are able to process the long distance dependencies in the PCG data. Nonetheless, it is important to consider that, while LSTMs can process longer sequences, their performance decreases on signals longer than 1000 time steps [33]. This is why it was necessary to use signal envelopes with a lower sampling frequency: processing the raw heart sounds directly would not only incur a very high computational cost, but probably also lead to vanishing gradients. The bidirectional layers also

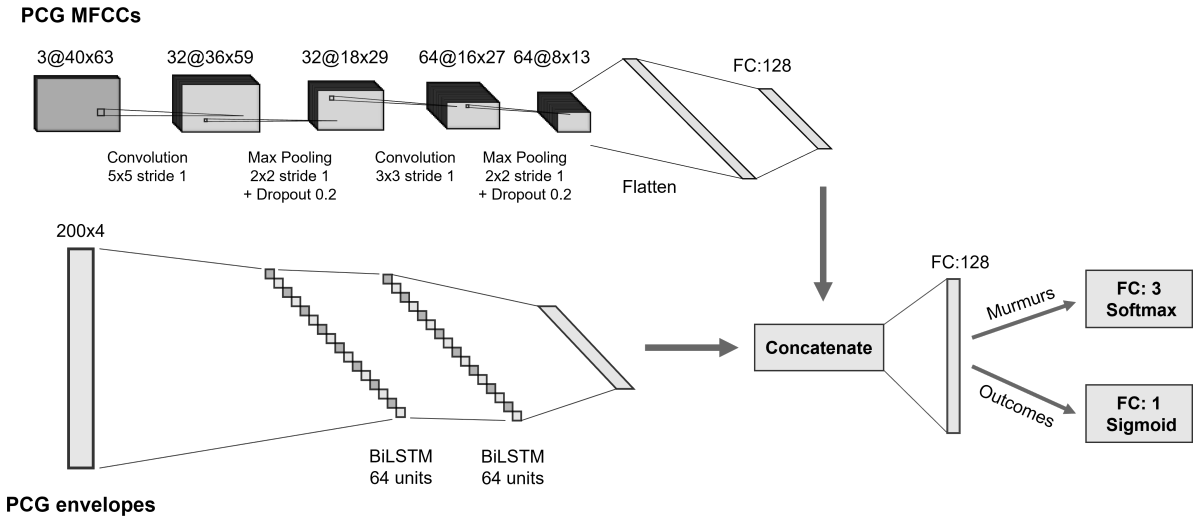


Figure 4.5: Hybrid CNN and LSTM architecture used for PCG classification.

allow the processing of the envelopes in both directions to exploit both the past and future information of the signal. The BiLSTM will read the entire sequence and then yield a single output, obtained by concatenating the hidden state of the last sample in the forward layer and the hidden state of the first sample in the backward layer.

When the two blocks are combined in the hybrid neural network, the feature vectors extracted by each of the blocks are concatenated before going through fully connected layers for classification into one of the available classes. In the last layer, the softmax activation function was used for the murmur multiclass classification and the sigmoid activation function for the outcome binary classification [32]. Table 4.1 shows the chosen parameters of the initial neural network implementations for both tasks.

CNN	
Convolution kernel	5x5 stride 1
Max pooling kernel	2x2 stride 1
Dropout rate	0.2
Activation function	ReLU
BiLSTM	
Hidden state dimension	64 units
Input segment size	200 samples
Nr. BiLSTM layers	2
Activation function	tanh
Network Parameters	
Optimizer	SGD
Learning rate	10^{-3}
Momentum	0.9
Batch size	512
Max. number of epochs	300

Table 4.1: Selected parameters of the neural networks used for PCG murmur and outcome detection.

4.1.3 Training and Evaluation

The evaluation of the performance of the models was done using 5-fold cross validation, which splits the training data into 5 subsets with approximately the same number of samples, and then uses one fold for testing, and the others for training. The test fold rotates 5 times, and then the performance metrics are obtained by averaging over all the folds, which helps reduce the bias in the model's evaluation process. The data was split in a way that ensured that recordings from the same patient don't appear in two different folds, to avoid overfitting, and stratified for each of the tasks, to guarantee an even class distribution over the different folds.

The training was done using categorical cross-entropy as a loss function for murmur detection and binary cross-entropy for outcome prediction [32]. Given the class imbalance in the murmur detection task, a weighing was applied to the loss function to make the model pay more attention to the under-represented classes: each instance of the classes Present and Unknown is worth two instances of the class Absent.

To avoid overfitting, early stopping was used to monitor the loss of the model on a validation set, obtained by randomly dividing the training data in each fold into 90% for training and 10% for validation. The networks were implemented with the Keras submodule from Tensorflow 2.6.1[2].

In order to obtain the final labels, it was necessary to take into account that the murmur detection and outcome prediction tasks are both cases of multiple-instance classification, given that each patient is represented by a set of instances (i.e. the recordings from the different auscultation locations), but it is only the patient that carries the label [97]. The approach was to train the model on individual recordings and then combine the instance-level decisions for the final patient label.

The neural networks use 4-second segments as input so, to obtain one classification per recording, each signal was split into 4-second segments with 50% overlap. The predictions were then combined by averaging the probabilities for each class and afterwards selecting the class with the highest probability.

For the final patient labels, it was assumed that a positive label contains at least one positive instance. In the murmur classification task, the final label and confidence scores for each patient were therefore generated by selecting the recording with the highest probability for the Present class, since it is only necessary that the murmur is audible in one location to confirm its presence. Similarly, in the clinical outcome classification task, the final label and confidence scores were generated by selecting the recording with the highest probability for the Abnormal class.

4.1.4 Results

In Tables 4.2 and 4.3 the standard performance and challenge metrics are presented for the murmur detection and outcome prediction tasks using the different neural network architectures.

The architecture that reached the best performance in murmur detection was the BiLSTM, with all the metrics being superior to the ones of the other models. Adding the spectral information of the MFCCs did not improve the results. A confusion matrix obtained for the best fold using this model is shown in

[2] https://github.com/tensorflow/tensorflow/tree/v2.6.1

Table 4.2: Performance metrics from multiclass PCG murmur classification obtained for each neural network architecture. The positive predictive value and the F_1 score refer to the Present class. The mean and standard deviation are in percentage (%). Best results are in bold.

Model	Accuracy	Sensitivity (Present)	Sensitivity (Unknown)	Specificity	Positive Predictive Value	F_1 score	Weighted Accuracy
CNN	68.1±6.2	82.1±4.4	14.7±6.7	69.8±8.5	46.0±8.7	59.5±7.6	58.3±3.4
BiLSTM	**77.1±1.9**	**82.7±3.7**	**31.2±11.9**	**80.1±3.1**	**58.4±4.3**	**68.3±3.0**	**65.25±4.3**
Hybrid	72.2±5.2	81.0±3.1	29.3±9.9	78.1±1.7	57.9±4.3	67.4±2.8	63.45±3.7

Table 4.3: Performance metrics from PCG clinical outcome classification obtained for each neural network architecture. The mean and standard deviation are in percentage (%), with the exception of the cost metric. Best results are in bold.

Model	Accuracy	Sensitivity	Specificity	Positive Predictive Value	F_1 score	Cost
CNN	59.9±3.0	**68.2±7.2**	52.1±5.7	57.2±2.7	**62.1±3.8**	**12021±953**
BiLSTM	**62.6±3.3**	61.4±6.7	**63.8±9.0**	**62.2±4.5**	61.3±2.4	12875±963
Hybrid	60.9±4.7	63.4±5.1	58.7±7.2	59.2±5.1	61.1±4.4	12565±911

Figure 4.6; the model has a high specificity and high sensitivity for the Present class, but a low sensitivity and very high standard deviation for the Unknown class. A possible explanation for this is the low sample size of the Unknown class (that only represents 7.2% of the data), which could make its instances too varied for the model to learn in a way that can be generalized for new examples. Another possible reason is the fact that the model could be able to reliably identify the presence or absence of murmurs in these signals, since the Unknown label is only an indicator of the recordings' inferior signal quality.

All the tested models had poor results in the outcome prediction task, with a low specificity and sensitivity. Considering the large weighting of the challenge cost metric towards sensitivity, the best performing model was the CNN, and a confusion matrix of the best performing fold is presented in Figure 4.6; it is clear that the model is highly overfit to the positive class, and as such loses its usefulness as a pre-screening algorithm. This low performance could be due to the fact that the outcome labels result from an overall assessment of the patient's condition based on multiple examinations. Since the diagnosis was not done based only on auscultation data, it is possible that some of the abnormalities simply cannot be identified just from heart sound recordings, meaning that the PCG signals do not contain all the information that is necessary for the neural networks to learn.

To the test the impact of the Unknown recordings on the overall performance of the models, this evaluation was repeated using only the two main classes in the murmur detection task (Absent and Present), with the results presented in Table 4.4. There is a slight increase in the specificity, accuracy, and positive predictive value, but not in the sensitivity, and the LSTM no longer outperforms the other neural network architectures. The hybrid neural network shows the best balance between sensitivity and specificity but, when compared to the multiclass problem that includes the low quality signals, the increase in performance isn't significant.

Given these results, the chosen model was the one based on the BiLSTM architecture, and further

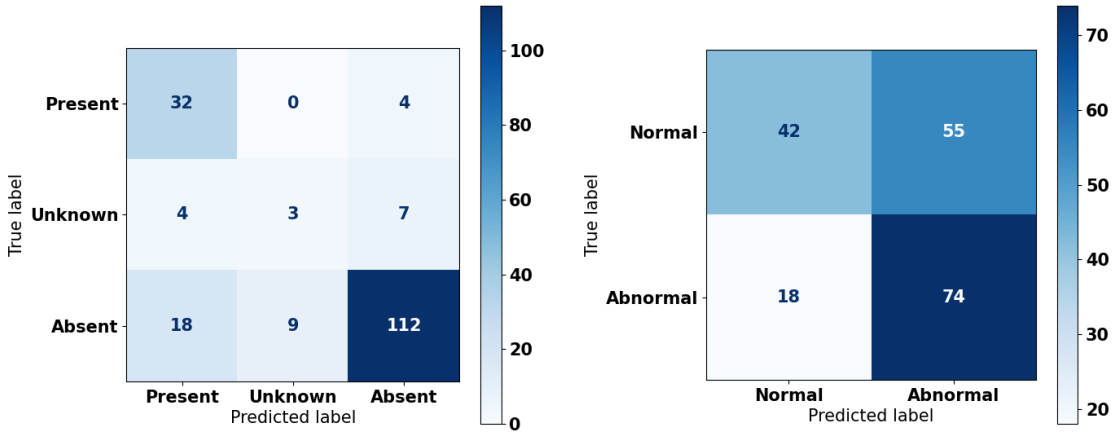


Figure 4.6: Confusion matrices of neural networks with the best F_1 score for the murmur detection (left) and outcome prediction (right) tasks.

Table 4.4: Performance metrics from binary PCG murmur classification obtained for each neural network architecture. The mean and standard deviation are in percentage (%). Best results are in bold.

Model	Accuracy	Sensitivity	Specificity	Positive Predictive Value	F_1 score
CNN	**84.7±4.1**	79.2±3.9	**85.8±6.2**	**61.9±12.8**	**68.5±5.7**
BiLSTM	81.9±1.7	81.0±2.9	84.9±4.3	55.6±5.7	65.2±5.0
Hybrid	82.8±4.3	**82.1±4.9**	83.0±5.7	56.8±8.1	66.7±5.4

evaluations were done to tune some of the hyperparameters, namely the number of hidden layers and the number of units. The use of separate models for each auscultation location was also tested.

Units and layers

The local performance of the models was evaluated for different numbers of stacked hidden layers and different number of units. In the murmur detection task, there is a significant increase in the performance of the model when the number of BiLSTM layers is increased from 1 to 2 (Table 4.5). However, when the number of layers increases from 2 to 3, the performance decreases. Even though deeper models should be able to learn more complex patterns in the signals, the small sample size can make the deeper network overfit to the training set and be unable to generalize to new data. In the outcome prediction task, the performance and scores of the model are not as correlated to the number of stacked BiLSTM layers (Table 4.6). The best challenge score and highest sensitivity and F_1 score are obtained with a single layer, but the highest specificity is obtained with two layers.

The number of units is related to the dimension of the hidden state so, the larger this number, the better the network should learn the more complex patterns of the inputs. The results were similar to the ones of the number of hidden layers: in the murmur detection task, the performance increases when the number of units goes from 32 to 64, and then decreases when the number of units goes from 64 to 128 (Table 4.7). The exception was the sensitivity for the Unknown class, that had a higher average for the

Table 4.5: Performance metrics from multiclass PCG murmur classification obtained for BiLSTMs with different numbers of hidden layers. The positive predictive value and the F_1 score refer to the Present class. The mean and standard deviation are in percentage (%). Best results are in bold.

Layers	Accuracy	Sensitivity (Present)	Sensitivity (Unknown)	Specificity	Positive Predictive Value	F_1 score	Weighted Accuracy
1	73.8±4.1	77.6±4.0	24.7±15.1	77.6±5.2	54.9±9.7	63.6±6.0	60.0±4.8
2	**77.1±1.9**	**82.7±3.7**	**31.2±11.9**	**80.1±3.1**	**58.4±4.3**	**68.3±3.0**	**65.25±4.3**
3	74.1±2.7	77.6±7.2	25.2±11.7	78.4±5.0	54.4±7.1	63.7±6.1	60.2±7.4

Table 4.6: Performance metrics from PCG clinical outcome classification with BiLSTMs with different numbers of hidden layers. The mean and standard deviation are in percentage (%), with the exception of the cost metric. Best results are in bold.

Layers	Accuracy	Sensitivity	Specificity	Positive Predictive Value	F_1 score	Cost
1	62.1±2.7	**71.1±6.2**	53.4±9.3	59.3±3.0	**64.5±2.3**	**11532±635**
2	**62.6±3.3**	61.4±6.7	**63.8±9.0**	**62.2±4.5**	61.3±2.4	12875±963
3	61.4±4.2	67.6±5.3	56.2±11.0	59.6±3.1	62.8±4.1	12033±1092

Table 4.7: Performance metrics from multiclass PCG murmur classification obtained for BiLSTMs with different dimensions for the hidden state. The positive predictive value and the F_1 score refer to the Present class. The mean and standard deviation are in percentage (%). Best results are in bold.

Nr of units	Accuracy	Sensitivity (Present)	Sensitivity (Unknown)	Specificity	Positive Predictive Value	F_1 score	Weighted Accuracy
32	72.4±4.0	79.3±5.1	23.8±11.4	74.5±6.9	51.2±6.9	62.0±5.5	60.4±4.0
64	**77.1±1.9**	**82.7±3.7**	31.2±11.9	**80.1±3.1**	**58.4±4.3**	**68.3±3.0**	**65.25±4.3**
128	74.1±2.7	82.2±7.7	**32.1±12.4**	76.1±4.4	54.7±4.3	65.3±1.7	64.9±4.9

Table 4.8: Performance metrics from PCG clinical outcome classification with BiLSTMs with different dimensions for the hidden state. The mean and standard deviation are in percentage (%), with the exception of the cost metric. Best results are in bold.

Nr of units	Accuracy	Sensitivity	Specificity	Positive Predictive Value	F_1 score	Cost
32	62.5±1.5	**68.7±5.7**	57.0±4.2	59.2±3.6	**63.4±2.7**	**12020±1166**
64	**62.6±3.3**	61.4±6.7	**63.8±9.0**	**62.2±4.5**	61.3±2.4	12875±963
128	61.3±2.5	66.6±3.1	56.2±3.0	58.8±4.0	62.4±3.3	12237±660

network with 128 units (although it still has a very high standard deviation). In the outcome prediction task, the best challenge score, highest sensitivity, and highest F_1 score are obtained with the simplest model, with 32 units, but the best specificity and positive predictive value are obtained with 64 units (Table 4.8).

Auscultation locations

The intensity and shape of the murmurs will vary when the stethoscope is placed over the different auscultation locations in the same patient. For this reason, training separate models for murmur detection with each one specialized in a recording location could lead to a more accurate diagnosis.

Based on the previous tests on the number of hidden layers and units, the chosen BiLSTM architecture was the one with two stacked layers and 64 units. Four different models were trained with recordings from each of the recording locations: aortic valve, pulmonary valve, tricuspid valve, and mitral valve. To obtain the final patient labels, the reasoning was the same as in the case with a single model: a prediction is obtained for each recording (using the model that corresponds to its location), and then the final labels and scores are generated by selecting the label of the recording with the highest probability for the Present class. The recordings that were labeled as being collected from a different and unspecified recording location were not used to train any of the models, and were evaluated using the model trained with recordings from the aortic valve.

The results are presented in Table 4.9. Using separate models for each auscultation location lead to a slight increase in the sensitivity for the Present class, but the sensitivity for the Unknown class and the specificity significantly decreased.

Table 4.9: Performance metrics from multiclass PCG murmur classification with single model and separate models for each auscultation location. The positive predictive value and the F_1 score refer to the Present class. Both the mean and standard deviation are in percentage (%). Best results are in bold.

Model	Accuracy	Sensitivity (Present)	Sensitivity (Unknown)	Specificity	Positive Predictive Value	F_1 score	Weighted Accuracy
Single model	**77.1±1.9**	82.7±3.7	**31.2±11.9**	**80.1±3.1**	**58.4±4.3**	**68.3±3.0**	**65.25±4.3**
Separate models	73.9±1.7	**84.9±6.2**	13.5±9.0	76.9±2.5	49.8±3.8	62.6±3.2	60.2±3.0

Taking all these evaluations into consideration, for scoring and ranking on the challenge the chosen model was the one based on the BiLSTM architecture with 64 units and two hidden layers, with a single model trained for all the recording locations. The scores and ranks for each task are presented in Table 4.10. The scores of the murmur detection task in the official hidden validation and test sets were superior to the scores obtained with cross-validation on the public data, which suggests that the model did not overfit to the training set. On the other hand, in the outcome prediction task, the model still had poor results, with a worse performance on the hidden test set.

Table 4.10: Challenge weighted accuracy metric for the murmur detection task and cost metric scores for the clinical outcome prediction task, with ranking on the hidden test set. 5-fold cross validation was used on the public training set, with repeated scoring on the hidden validation set and one-time scoring on the hidden test set.

	Training	Validation	Test	Ranking
Murmur Detection	0.652±0.043	0.751	0.757	6/40
Outcome Prediction	12875±963	11222	13815	25/39

The developed approach ranked highly on the murmur detection task, with the model based on BiLSTM networks and temporal and morphological features demonstrating promising performance in the detection of murmurs (with a high specificity and high sensitivity for the murmur class). However, the low prevalence of murmurs in the tested population lead to a positive predictive value of 58.4% on the training data, meaning that more than 40% of the patients that would be referred to a clinician by the algorithm would be false alarms. To further improve these results, possible steps include trying to supplement the envelope features with the provided demographic data and murmur characteristics, or adjust the decision thresholds to boost the score.

4.2 Heart Abnormality Detection based on the ECG and the PCG

The goal of the PhysioNet/CinC challenge 2016 was to develop an algorithm to automatically classify PCG recordings as normal or abnormal, to identify whether the subjects needed to be referred for expert diagnosis. It assembled eight databases from different sources, of which six were publicly available (training sets "A" through "F"). Of these, only training set "A", which corresponds to the MIT Heart Sounds Database (MITHSDB), contains simultaneous ECG data.

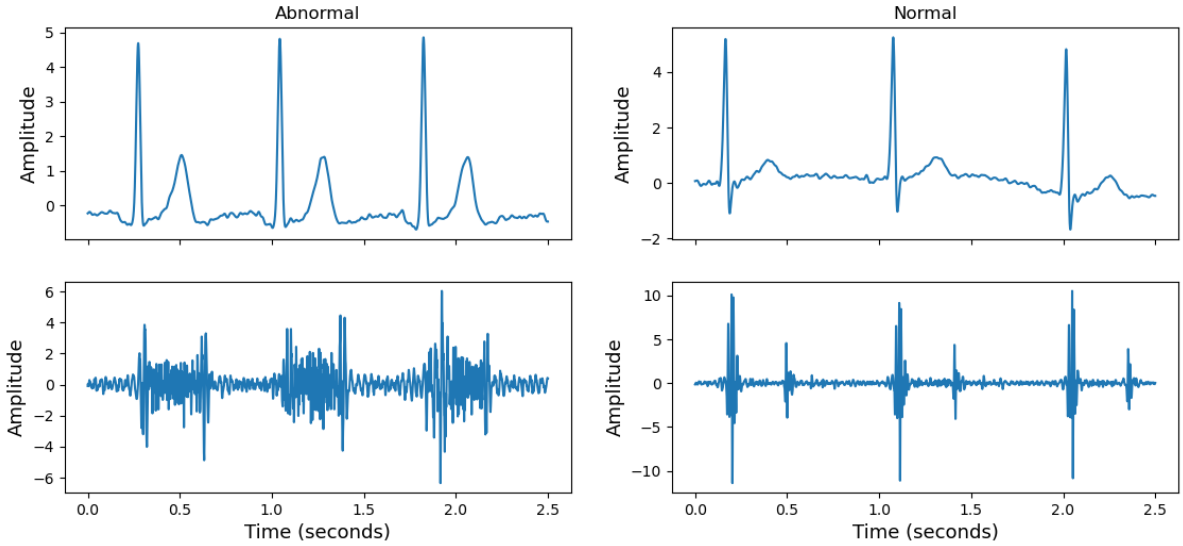


Figure 4.7: Examples of simultaneous ECG and PCG recordings from Abnormal (left) and Normal (right) classes from the MITHSDB. The signals were filtered and normalized to have zero mean and unit variance.

The MITHSDB includes 405 samples with simultaneous heart sounds and ECG recordings (Figure 4.7), acquired with a Welch Allyn Meditron electronic stethoscope with a frequency response from 20 Hz to 20 kHz [26]. The acquisitions were performed in an uncontrolled environment during in-home visits or at the hospital, meaning several PCG recordings are contaminated by noise, which includes stethoscope motion, breathing, talking, and intestinal sounds. The recordings were obtained from 121 subjects at nine different positions and orientations, meaning that each patient contributed with more than one recording. The recordings are divided into five different categories: 116 recordings are from normal controls; 132 recordings are from patients with mitral valve prolapse; 117 recordings from patients with

benign murmurs; 17 recordings from patients with aortic disease; and 23 recordings from patients with other miscellaneous conditions [26]. The dataset is therefore unbalanced when considering a binary classification: about 32.6% of the data are considered normal, while 71.4% are considered abnormal.

Given the low sample size in the MITHSDB, the approach to develop an algorithm for heart abnormality detection based on the multi-modal signals was to first apply the method described on the previous section to the entire PhysioNet/CinC challenge 2016 database (to determine the best performing neural network for PCG abnormality detection), and then to evaluate whether adding the ECG information improves the classification performance in the MITHSDB.

4.2.1 PCG Abnormality Detection

Similarly to the task in Section 4.1, the training set was split using stratified 5-fold cross validation, with the same class distribution maintained across the different folds. The pre-processing pipeline was the same, and the same methods were also used to deal with the class imbalance. On the entire Physionet/CinC challenge 2016 dataset only approximately 20.5% of the data belongs to the Abnormal class, and as such, in each fold and prior to feature extraction, the segments of recordings belonging to this class were extracted with 75% overlap. The Unsure class was not considered. A weighing was also applied to the loss function (binary cross-entropy), where each instance of class Abnormal is worth two instances of the class Normal. The network hyperparameters are the same as before (Table 4.1).

Since there is no information about which signals belong to each patient, the classification was done on a per recording basis. The results are presented in Table 4.11, and the top performance metrics are in line with the methods in the state-of-the-art applied to the same dataset. For PCG abnormality classification, the best performing models are the CNN and hybrid architectures, contrary to the results obtained for murmur detection. It can also be seen that the improvement in performance obtained from adding the parallel BiLSTM block was not significant, and that the CNN has the best sensitivity. For these reasons, the chosen model for evaluation on the MITHSDB was the CNN.

Table 4.11: Performance metrics from PCG classification obtained for each neural network architecture. Both the mean and standard deviation are in percentage (%). Best results are in bold.

Model	Accuracy	Sensitivity	Specificity	Positive Predictive Value	F_1 score
CNN	88.5±1.0	**92.8±1.5**	87.3±1.3	65.5±2.3	76.8±1.7
BiLSTM	79.7±4.3	90.8±3.2	76.8±5.8	50.9±5.0	65.0±4.3
Hybrid	**88.7±0.9**	92.3±2.3	**87.8±0.5**	**66.2±1.5**	**77.1±1.8**

4.2.2 Multi-modal signal Classification

As discussed in Section 3.1, RNNs are able to capture the temporal dependencies in the raw ECG signals without any feature extraction steps. The chosen approach was therefore to combine the CNN for the PCG classification with a parallel BiLSTM for ECG classification, in order to verify whether the

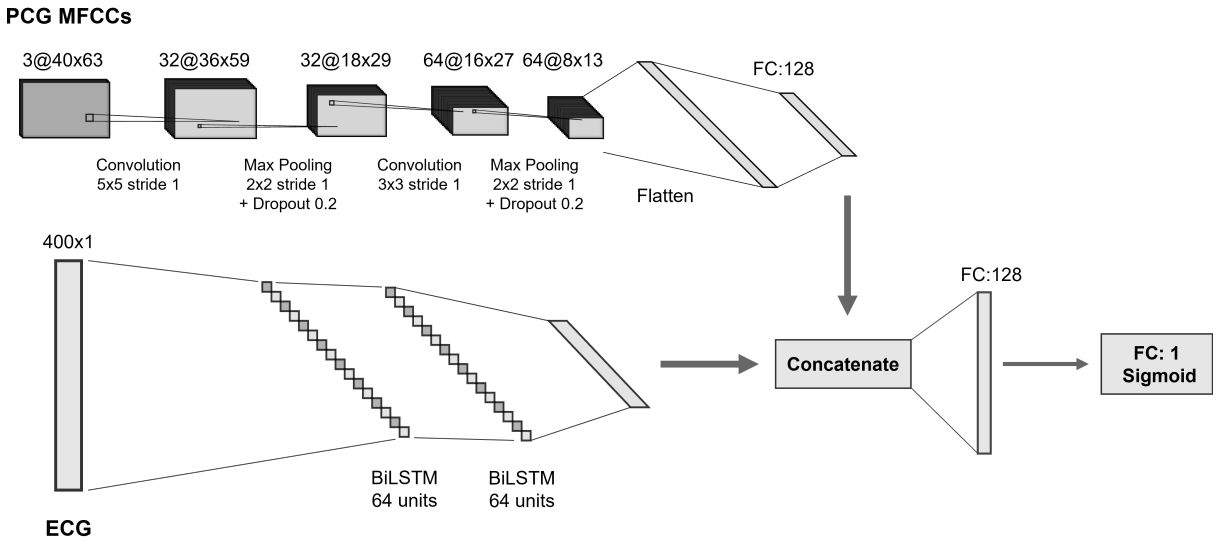


Figure 4.8: Hybrid CNN and LSTM architecture used for multi-modal ECG and PCG classification.

Table 4.12: Selected parameters of the neural networks used for multi-modal ECG and PCG classification.

Network Parameters	
Optimizer	Adam
Learning rate	10^{-3}
Batch size	32
Max. number of epochs	300

combination of the two signals enhances the neural network's classification performance.

Each ECG signal was filtered using a 100th-order band-pass Finite Impulse Response (FIR) filter with cutoff frequencies of 0.5 Hz and 50 Hz and then downsampled to a sampling frequency of 100 Hz. This sampling frequency was chosen to decrease the number of samples to be processed by the LSTM, while taking into consideration that it would still be sufficient for an accurate time domain analysis [98].

Some of the ECG signals were inverted in amplitude, and this was corrected by applying a simple method based on the median of the R-peak amplitudes: the R-peaks were detected using the Hamilton segmentation algorithm [99], and then the inverted signals were identified by checking whether they had a lower median R-peak amplitude than their inverted versions [100]. The signals were then normalized to have zero mean and unit variance.

A diagram of the used neural network is presented in Figure 4.8. It was trained with MFCC features of PCG segments at the input of the CNN block, and with the filtered and normalized ECG segments at the input of the LSTM block. The segment length was equal to 4 seconds, and these were extracted with 75% overlap for the minority class (which in this case was the Normal class). The parameters used to train the neural network are summarized in Table 4.12. Due to the lower sample size in this dataset, it was necessary to adjust the optimizer and the batch size.

In Table 4.13 the performance metrics for uni- and multi-modal signal classification are presented: the PCG uni-modal classification was done using only the CNN block; the ECG uni-modal classification was done using only the LSTM block; and the multi-modal classification was done using the two parallel

blocks. By analyzing these metrics and the confusion matrices represented in Figure 4.9, it is quite clear that the performance using the multi-modal signals is superior to that of the single modalities, showcasing the potential of using the combined signals to detect cardiac abnormalities. The results are in line with the ones of the only other deep learning method in the state-of-the-art [7]. The model has a high sensitivity for abnormality detection and a high positive predictive value, above 90%, that is influenced by the fact that there are more abnormal than normal samples on the dataset. Consequently, even though the specificity is below 80%, less than 10% of the samples that are classified as positive are false alarms.

Table 4.13: Performance metrics from multi-modal signal classification obtained for different signal combinations. Both the mean and standard deviation are in percentage (%). Best results are in bold.

Modalities	Accuracy	Sensitivity	Specificity	Positive Predictive Value	F_1 score
Only PCG	72.8±2.5	77.4±3.9	61.5±7.5	83.3±2.2	80.2±2.1
Only ECG	84.7±1.5	88.2±2.5	76.2±8.6	90.2±3.3	89.1±0.9
Simultaneous ECG + PCG	**86.7±4.0**	**89.9±3.9**	**78.7±9.3**	**91.3±3.7**	**90.5±2.9**

4.3 Discussion

From the previous experiments, several conclusions regarding the use of deep learning methods for automatic CVD screening from PCG and ECG data can be drawn.

First, it was observed that BiLSTM neural networks show promising performance in the detection of murmurs by capturing the morphological and temporal variability in PCG signals, with an average sensibility of 82.7% and average specificity of 80.1%. This incentivizes further development, but it is important to note that heart murmurs on their own aren't the best predictors of CVD, particularly in pediatric populations. Even in older populations, most of the adults with systolic murmurs have normal echocardiograms, meaning that the murmur is functional or innocent [22].

The best performing model for PCG abnormality detection in the Physionet/CinC challenge dataset was the CNN with MFCC features. Combining this architecture with the BiLSTM block to include the temporal features did not significantly increase the model's performance. These differences in the best performing models are most likely due to the properties of the datasets since, as was mentioned, the signals can have very different characteristics depending on the population and the environment in which they were acquired.

Regarding the classification based on synchronous PCG and ECG data, the results demonstrate that the multi-modal approach can improve the accuracy of automatic CVD screening, reaching an average F_1 score of 90.5%. However, unlike the methods applied to uni-modal ECG and PCG classification, the proposed approach and other deep learning based methods still don't outperform traditional machine learning methods with handcrafted features. This is likely due to the small amount of available data, since deep learning approaches are highly affected by limited datasets. Due to this constraint, the

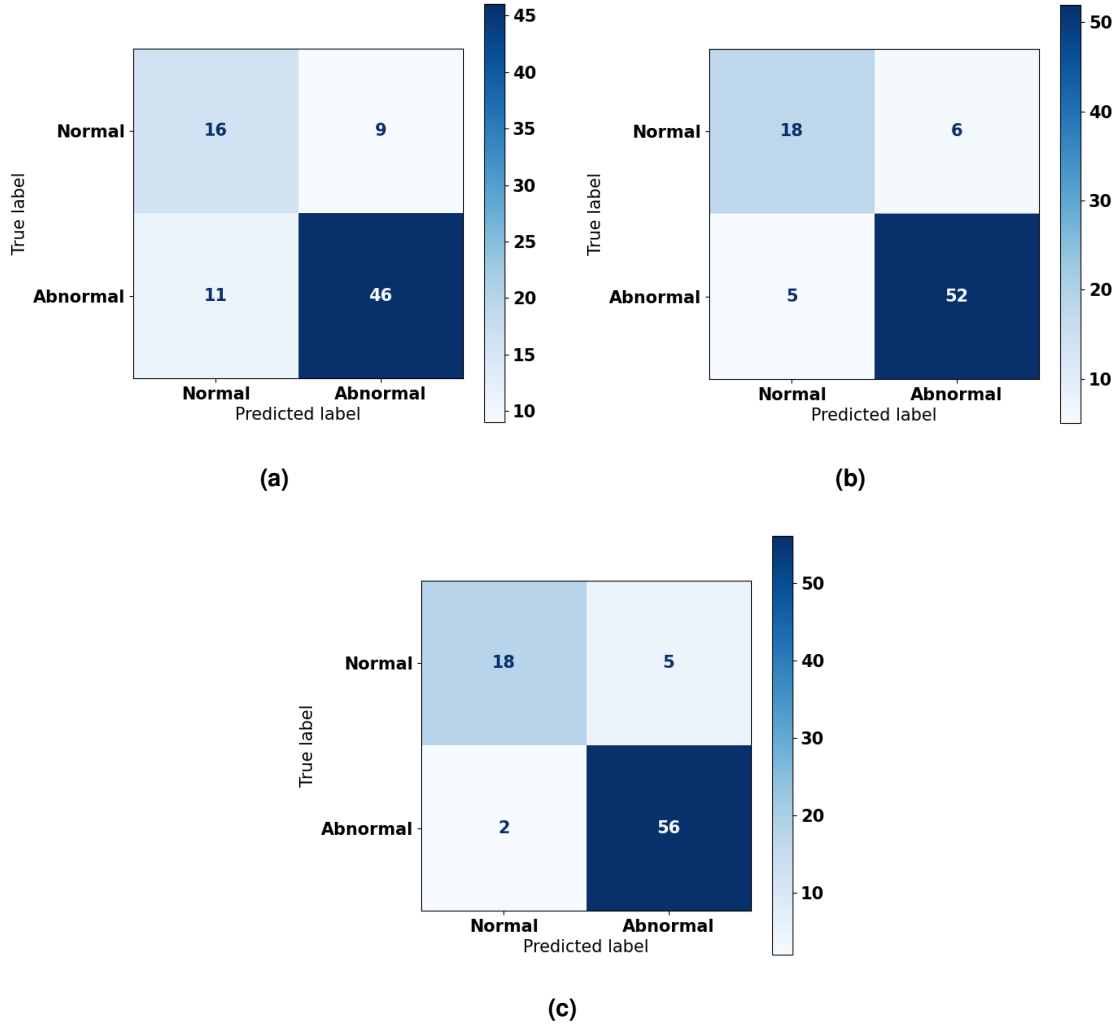


Figure 4.9: Confusion matrices of uni- and multi-modal classification. a) Uni-modal PCG. b) Uni-modal ECG. c) Multi-modal (ECG + PCG)

results should be interpreted in a tentative way since, as was mentioned, there is not only a low sample size, but many samples come from the same patient and are treated independently, which could lead to overfitting and an overestimation of the models' performance. Nonetheless, this analysis showed the potential of combining the two signals and provides a good reference for future applications and development.

Larger and more representative datasets are necessary to train and implement deep learning classifiers that can fully take advantage of the complementary information in the two modalities. However, devices that are capable of simultaneous PCG and ECG acquisition are still not widely used due to multiple limiting factors. In order to build more robust classification methods, it is first necessary to develop hardware architectures that can record the synchronous PCG and ECG and overcome the current usability constraints. This motivated the development of a novel hardware prototype based on a low-cost and open-source stethoscope, which is described in Chapter 5.

Chapter 5

A Novel Electronic Stethoscope for ECG and PCG Acquisition

In this chapter, a novel device for simultaneous electrocardiography and phonocardiography during auscultation is proposed. A summary of the chosen approach is presented (Section 5.1), along with the details of the device's form factor (Section 5.2) and instrumentation (Section 5.3). The experiments that were performed to evaluate the design, mainly related with the validation of the sensors and of the acquired biosignal data, are also described. They include an evaluation of the frequency response of the stethoscope (Section 5.4), an analysis of the concurrence between the multi-modal data (Section 5.5), a characterization of the acquired ECG leads (Section 5.6), and their comparison with Lead I (Section 5.7) . Section 5.8 then presents a discussion of the results.

5.1 Proposed Approach

A core aspect when combining ECG and PCG for disease diagnostic is the temporal synchronization of both signals. Previous work has found restrictions associated with retrofitting external sensors to existing digital stethoscopes [14, 101]. To gain more control over the construction and sensor integration, this work builds upon the Glia stethoscope to devise an end-to-end design of a novel form factor, without significantly changing the standard form factor. The Glia stethoscope[1] is an open-source 3D-printed acoustic stethoscope that is research-validated. It operates as well as a Littmann Cardiology III (considered a gold-standard device [102]), but at a fraction of the cost: approximately 20 USD, in contrast with the hundreds of dollars of the Littmann stethoscopes [102].

A model of the original Glia stethoscope is presented in Figure 5.1. It is composed of six different 3D printed parts: the stethoscope head (A), the ear tubes (D), the Y-piece which connects the head to the ear tubes (C), a spring (B), and a ring to attach the diaphragm to the head. Other hardware components include: a 40 cm silicone 12 mm outer diameter (OD), 8 mm inner diameter (ID) tube (connecting the stethoscope head and the Y-piece), and two 9 cm silicone 6 mm OD, 4 mm ID tubes (connecting the Y-piece to the ear tubes), which can be purchased off-the-shelf; a diaphragm that can be cut from standard PVC sheet material; and standard ear buds [103].

The chosen approach was to adapt this design to include phonocardiography and electrocardiography sensing capabilities, while keeping it open-source, with a low-cost, and ensuring that the acoustic

[1] https://github.com/GliaX/Stethoscope

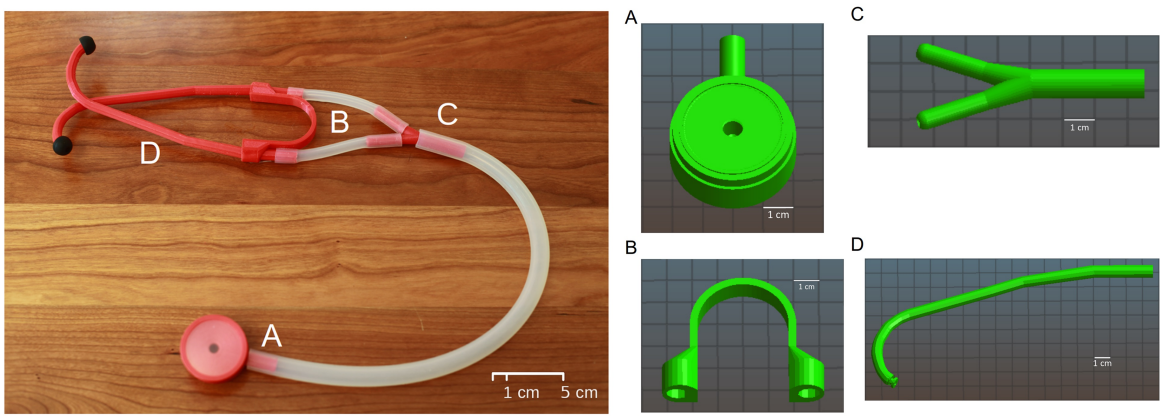


Figure 5.1: Glia stethoscope and digital models of 3D-printable parts. Image from [103].

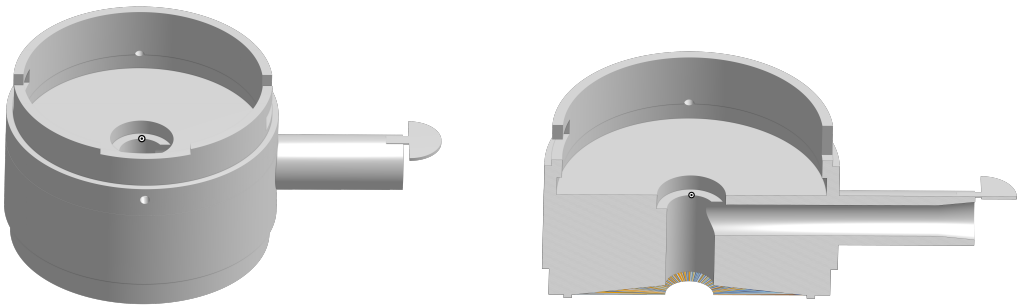
properties of the model aren't significantly altered. The form factor of the device was also considered, so it could be used similarly to standard auscultation stethoscopes. As such, the stethoscope head was adapted to include PCG and ECG sensors, with 3D-printed polymer based dry electrodes attached to the ring.

5.2 Form Factor

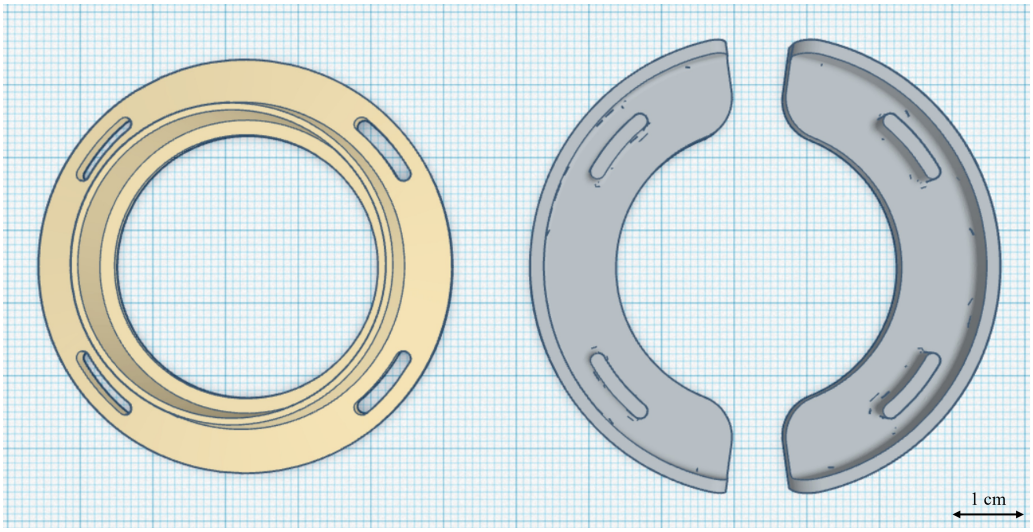
The aim was to preserve the standard analog auscultation feature of the stethoscope, to avoid creating barriers in the interaction between the end-user (the physician) and the device [101]. As such, to integrate the PCG sensor, the stethoscope head was modified by adding a small hole in the back (Figure 5.2(a)), with the same diameter as an electret microphone. With this approach, the sound will still travel through the silicone tubes and ear pieces to be listened to by the doctor during the auscultation procedure, while simultaneously being recorded. To verify that this alteration didn't significantly alter the acoustic properties of the stethoscope, the acoustic transfer of the design was compared with the one of the original stethoscope (described in Section 5.4).

To allow ECG sensing capabilities, two 3D-printed dry electrodes with a half-moon shape were added to the side of the stethoscope head, attached to an altered ring (Figure 5.2(b)). This design had several advantages, namely: 1) the fact that there was no need to alter the diaphragm of the original stethoscope, which is essential to ensure the quality and intensity of the sound reaching the earpiece and the sensor microphone; 2) a higher inter-electrode distance, which leads to a higher signal magnitude; 3) a higher electrode conductance due to the increased surface area granted by the half-moon shape, which leads to higher Signal-to-Noise Ratio (SNR) and reduced motion artifacts [104]; and 4) better ergonomics, since by fitting the electrodes along the ring the shape of the standard stethoscope isn't being modified.

The modified stethoscope parts (ring and stethoscope head) were printed using PolyLactic Acid (PLA) filaments. For the ECG dry electrodes the electrical conductivity of the material also had to considered, reason for which the Protopasta Conductive PLA (ProtoPlant, Inc) material was chosen,



(a)



(b)

Figure 5.2: Digital models of modified stethoscope parts. a) Modified stethoscope head with hole for the electret microphone on the top. b) Modified ring and half-moon dry ECG electrodes.

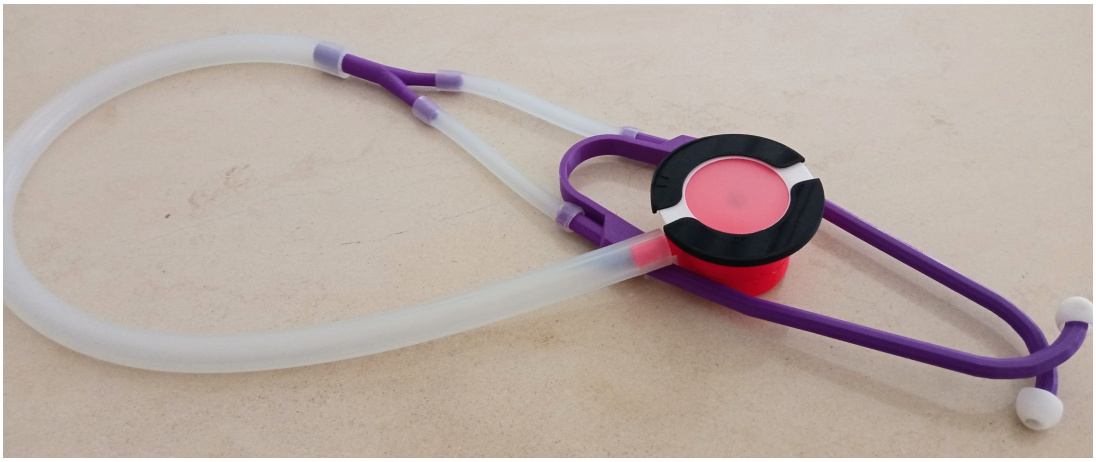


Figure 5.3: Photo of the developed device.

as it combines PLA with carbon black. It can be used on most PLA-compatible printers and has a resistivity of 30 Ω cm along the Y-axis and 115 Ω cm along the Z-axis. The electrical interface between the electrodes and the ECG sensor was done by fusing each wire to the material with heat, which has been tested and validated in [105]. The final device is presented in Figure 5.3.

5.3 Instrumentation

For data acquisition and Bluetooth wireless streaming, a novel acquisition system named ScientISST SENSE[2] was used. It is based on the ESP32 microcontroller architecture and allows multi-channel sampling with high time (up to 16kHz) and amplitude (12-bit) resolution. It was coupled with two sensors, one for each signal modality:

- A ScientISST ECG sensor [106], which has already been validated in previous studies [107]. It has a gain of 1100 and applies a bandpass filter with a high-pass (f_L) cut-off frequency of 0.5 Hz, and a low-pass (f_H) cut-off frequency of 40 Hz (the typical bandwidth for the analysis of the signal). It was used with a virtual ground configuration to measure a single-lead ECG.

- A novel PCG sensor based on an audio amplification circuit that uses an electret microphone with a frequency response between 20 Hz and 16000 Hz, 2.2 kΩ impedance, 60 Hz SNR and -40 dB sensitivity; and the LM4861 amplifier, with a gain of 40, a high-pass (f_L) cut-off frequency of approximately 21 Hz, and a low-pass (f_H) cut-off frequency of approximately 408 Hz (which were chosen based on the frequency content of the heart sounds). More details are available in Appendix A.

The SENSE Web App[3], which allows simultaneous visualization and recording, was used to record the data on a base station. Both signal modalities were acquired with a sampling frequency of 2000 Hz. Python 3.8 and the scikit-learn [108] (version 1.0.1) and BioSPPy [109] (version 1.0.0) libraries were used for the data analysis.

A total of 19 multi-modal signals were collected from 18 volunteers, 50% of which were female. One person has a previous diagnosis of arrhythmia, and in terms of age distribution 3 participants belonged to the age bracket of 18-24 years, 1 to 25-34 years, 8 to 35-44 years, 1 to 45-54 years, 3 to 55-64 years, and 2 were over 65 years old (Figure 5.4). The average height was 168.2$\pm$10.1 cm and the average weight was 66.2$\pm$12.7 kg. The evaluations were performed with the subjects seated and at rest. Initially the relevant demographic and clinical history data were collected, and after that a simultaneous phonocardiogram and electrocardiogram were recorded using the proposed device. Written consent was obtained for the participants, and all data was treated anonymously.

[2] https://www.scientisst.com/sense
[3] https://sense.scientisst.com

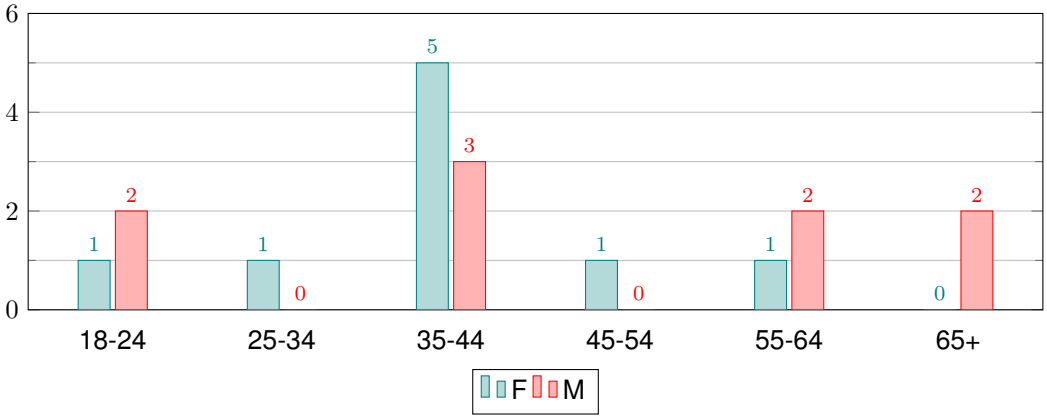


Figure 5.4: Distribution of subjects by sex and age bracket.

5.4 Acoustic Transfer

There isn't currently a standard method to compare the acoustic properties of different stethoscopes, with most evaluations being based on subjective opinions and agreement between users. More objective methods typically use a phantom to simulate the vibrations of the chest, and have the advantage of being easier to replicate for independent validations [103].

To evaluate the frequency response of the proposed device, the same experimental setup that was used to compare the frequency response of the Glia stethoscope with the one of the Littmann Cardiology III was adopted (described in [103]). A latex balloon was filled with 2 liters of water and used as a phantom, where both the proposed electronic stethoscope with the altered head and the original Glia stethoscope were applied. The phantom excitations were done by an external speaker (Adam Audio T5V nearfield monitor) which was in contact with the balloon, and white noise with frequencies between 0 and 5000 Hz (generated with Audacity® 3.1.3) was played for 15 seconds.

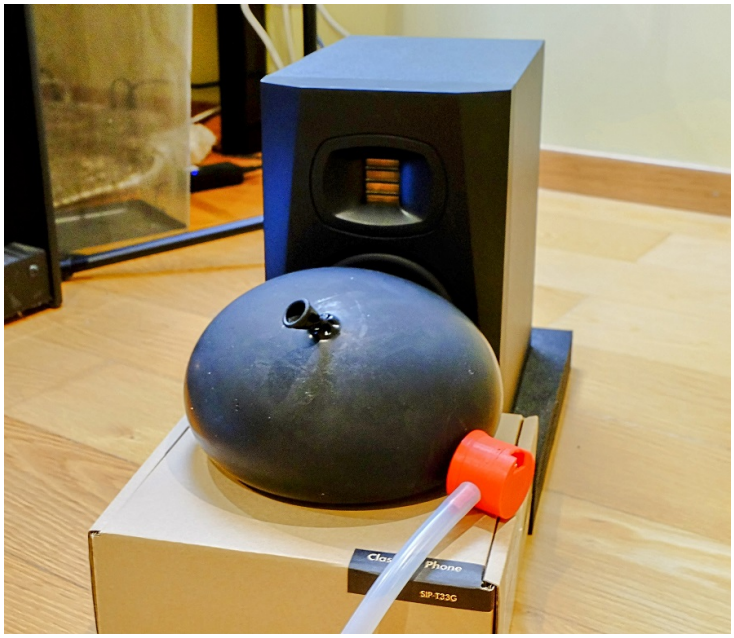
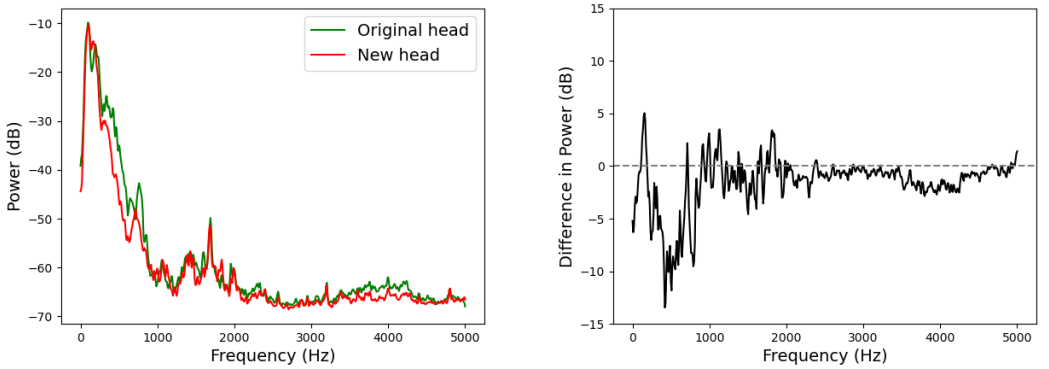
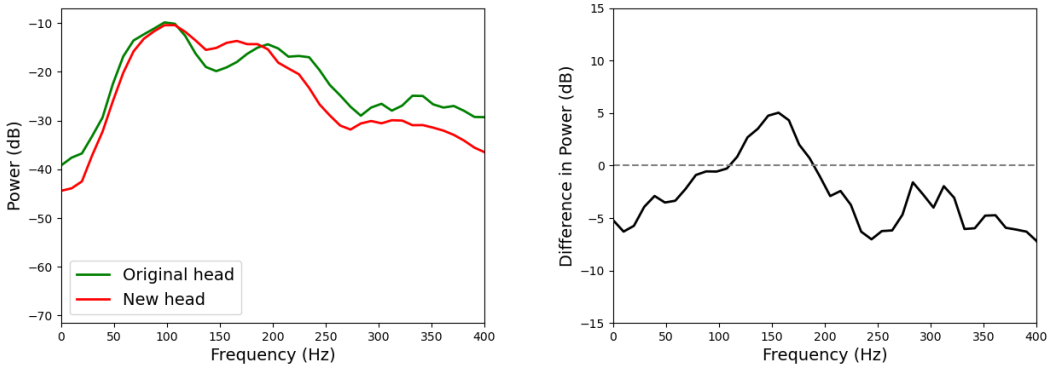


Figure 5.5: Experimental setup to evaluate the frequency response of the proposed device, with the Adam Audio T5V nearfield monitor and latex balloon filled with 2 liters of water in contact with the stethoscope head.



(a)



(b)

Figure 5.6: Stethoscope output frequency responses of the Glia stethoscope with the original head and the developed device with the altered head, for the full frequency spectrum (a) and for the frequencies of interest in PCG signals (b). The dB difference in power corresponds to the new device minus the original.

The output of the stethoscopes was recorded by a microphone placed at the end of the silicone tubes that connect the stethoscope head to the earpieces, with a sampling frequency of 10000 Hz. The setup is presented in Figure 5.5.

This experiment was repeated three times, and the average frequency responses are presented in Figure 5.6. The magnitude spectra were calculated by applying the Fast Fourier Transform (FFT) to segments of the signal with a Blackman-Harris window of size 1024, and then averaging the segments together. The proposed device performed similarly to the Glia stethoscope. In the low and mid-frequency ranges (the frequencies of interest for the analysis of PCG signals) the devices exhibit similar behavior, emphasizing the lower frequency regions between 50 and 200 Hz, with a slight roll-off in the higher frequencies.

5.5 Heart Rate and Auscultation Points

An example of unfiltered ECG and PCG signals acquired simultaneously with the device can be seen in Figure 5.7, and it is possible to identify the main fiducial points and structures and verify their alignment: the S1 sound occurs immediately after the R-peak, and the S2 sound immediately after the end of the T-wave. The PCG was segmented based on ECG fiducials and the heart sound states duration averages derived in [90].

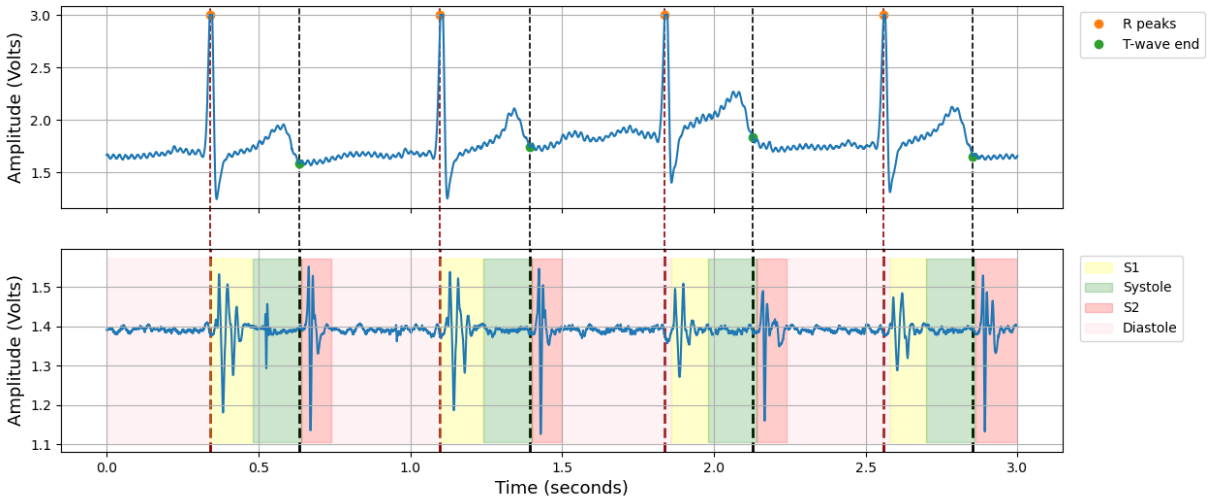


Figure 5.7: Unfiltered ECG and PCG signals acquired with the device. The PCG is segmented into the fundamental heart sound states, which are aligned with the R-peak and end T-wave positions of the ECG.

Before further analysis, both signals were filtered using a FIR filter with order equal to half of the sampling frequency, with cutoff frequencies of 0.5 and 40 Hz for the ECG and of 25 and 400 Hz for the PCG.

To evaluate the concurrence between the multimodal signals in all the subjects, the average heart rates obtained from each signal were compared. For the ECG, the average number of beats per minute was calculated by counting the number of R peaks (using the segmentation algorithm in [99]) in the whole signal duration, and for the PCG the average heart cycle duration was estimated using the method described in [90], by analyzing the autocorrelation function of the homomorphic envelope of the signal. The autocorrelation function compares a signal with a time-delayed version of itself, allowing the identification of periodic structures. As can be seen in Figure 5.8, the duration of the heart cycle can be estimated from the autocorrelation function by finding the highest peak between 500 and 1500 ms (assuming a normal heart rate that is situated between 40 bpm and 120 bpm). The homomorphic envelope is used instead of the original signal because it is smoother and less contaminated by noise [90].

In Table 5.1 and Figure 5.9 the heart rates calculated for both signals are shown. On the left plot of the figure, it can be verified that there is a high agreement between the measurements, although there are a few outliers (marked with black). They correspond to subjects that had an ECG signal with regions highly corrupted by artifacts, where the segmentation algorithm erroneously identified several R-peaks. After removing these sections and recalculating the average heart rates for the resulting segments,

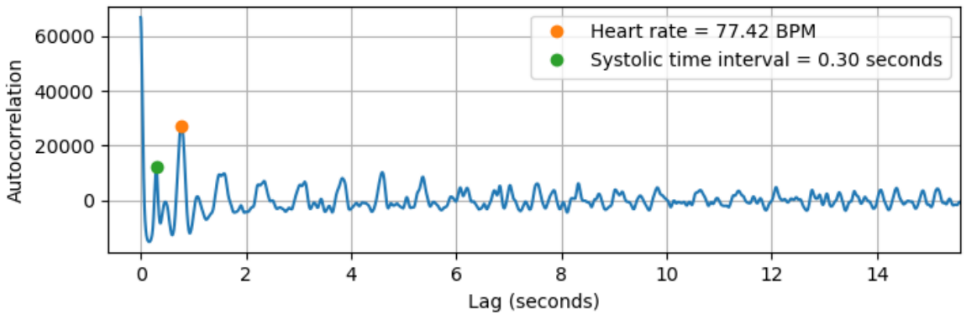


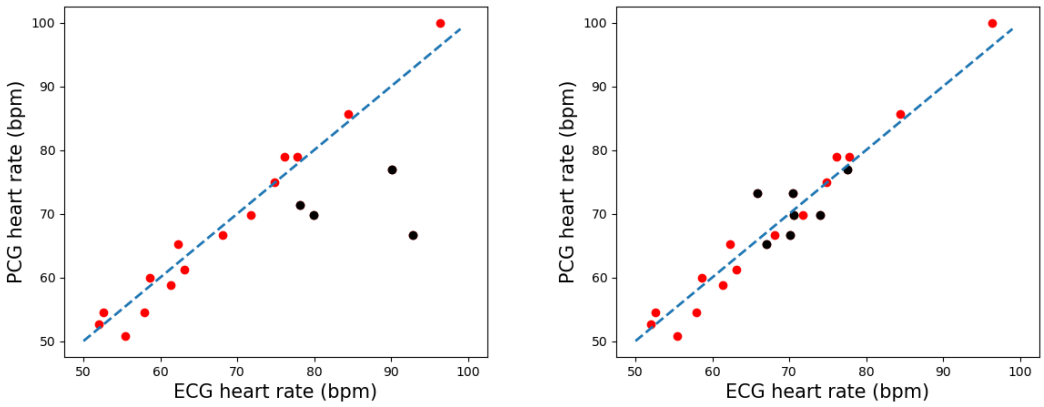
Figure 5.8: Example of estimation of heart rate durations in the PCG from the homomorphic envelope autocorrelation function.

Table 5.1: Average heart rates calculated from ECG and PCG and average absolute differences in beats per minute (bpm) of all signals before and after outlier correction.

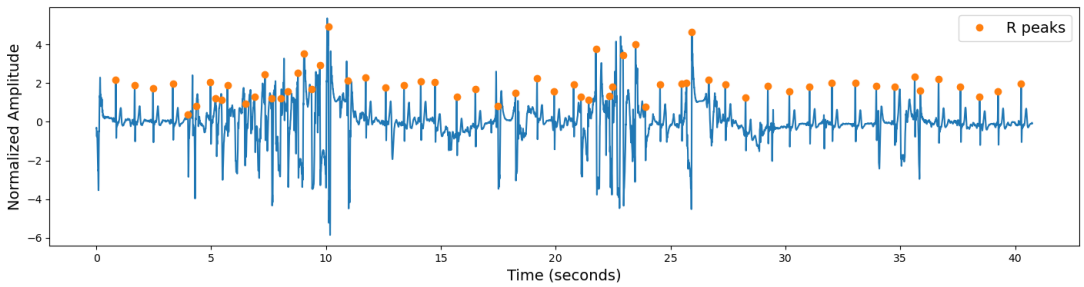
	ECG Heart Rate	PCG Heart Rate	Absolute Heart Rate Difference
Before outlier correction	71.23 ± 13.35	68.30 ± 12.18	4.64 ± 5.99
After outlier correction	68.54 ± 10.40	68.53 ± 11.42	2.41 ± 1.59

the ECG heart rates become much closer to the heart rates calculated for the PCG, and the absolute difference between the heart rates significantly decreases (Table 5.1).

As the stethoscope head is placed on different auscultation points, the heart's electrical axis will be viewed from different perspectives and, as such, the morphology of the recorded ECG waves will be different. This is clear in Figure 5.10, where alterations can also be seen in the PCG signal. S1 and S2 are present in all the signals, as expected, but depending on the area of the chest and its closeness to each heart valve, the stethoscope will amplify the cardiac sounds differently. By rotating the stethoscope head on the same auscultation point the effect on the ECG will be similar, i.e, since the electrode's position is changing, the signal will record the heart's electrical activity from a different perspective. Both signals have good quality, and the main fiducial points can be clearly identified.



(a)



(b)

Figure 5.9: a) Scatter plots comparing ECG and PCG average heart rates before (left) and after (right) artifact removal. The dashed line represents an equal heart rate in both signals. The outlier signals that were corrected are shown in black. b) Example of outlier ECG signal with R-peaks detected by the Hamilton segmentation algorithm.

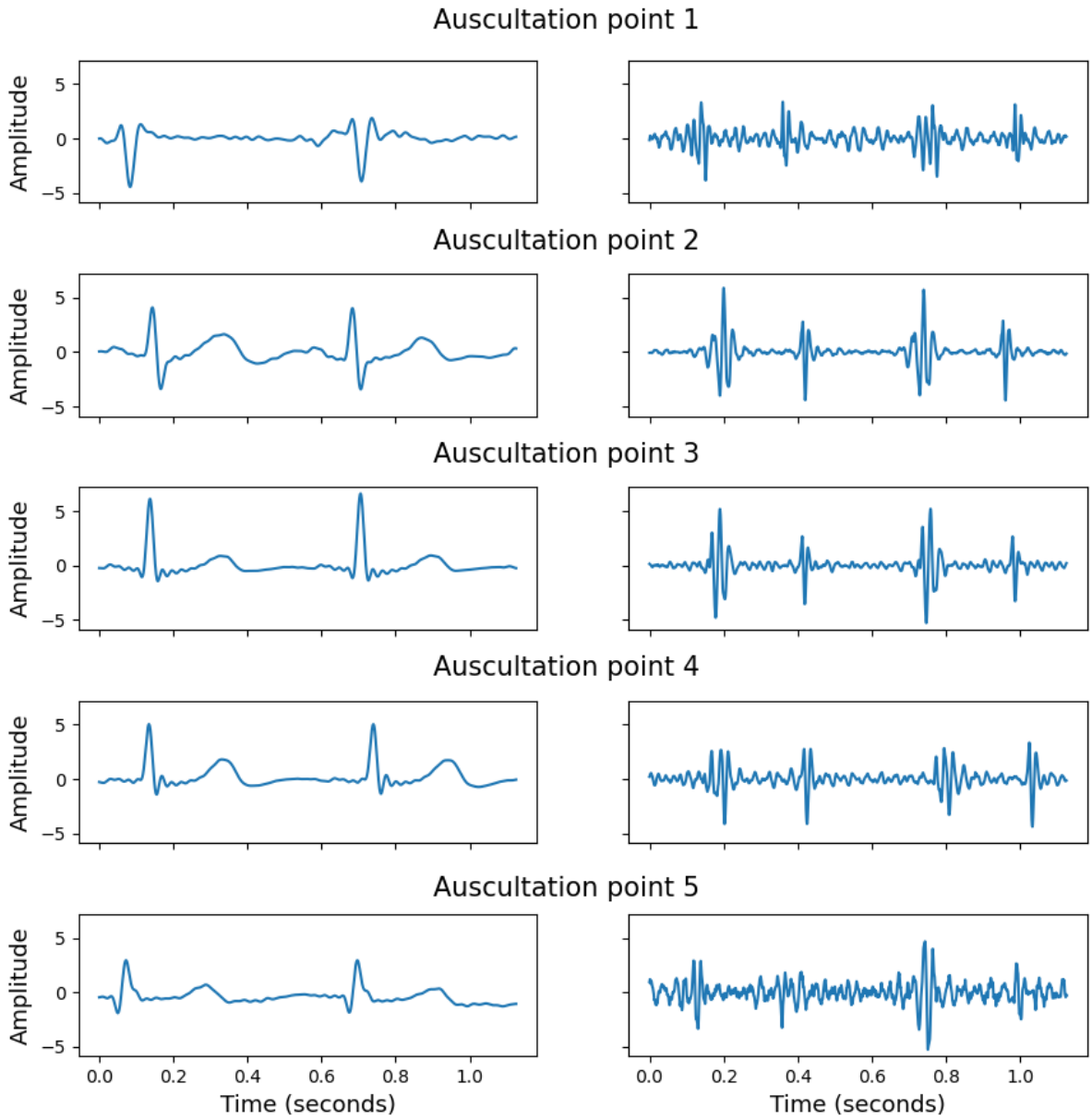


Figure 5.10: Example of normalized ECG and PCG signals acquired with the stethoscope head at the different auscultation points.

5.6 ECG Leads Characterization

The Einthoven limb leads (Leads I, II, and III) and their corresponding definitions are represented in Figure 5.11(a). This lead system is based on the assumption that the cardiac sources are represented by a dipole $\bar{p}$ located at the center of the equilateral triangle. Thus, the voltages measured by the three limb leads are proportional to the projections of the electric heart vector on the sides of the lead vector triangle [13]. Considering this, we have:

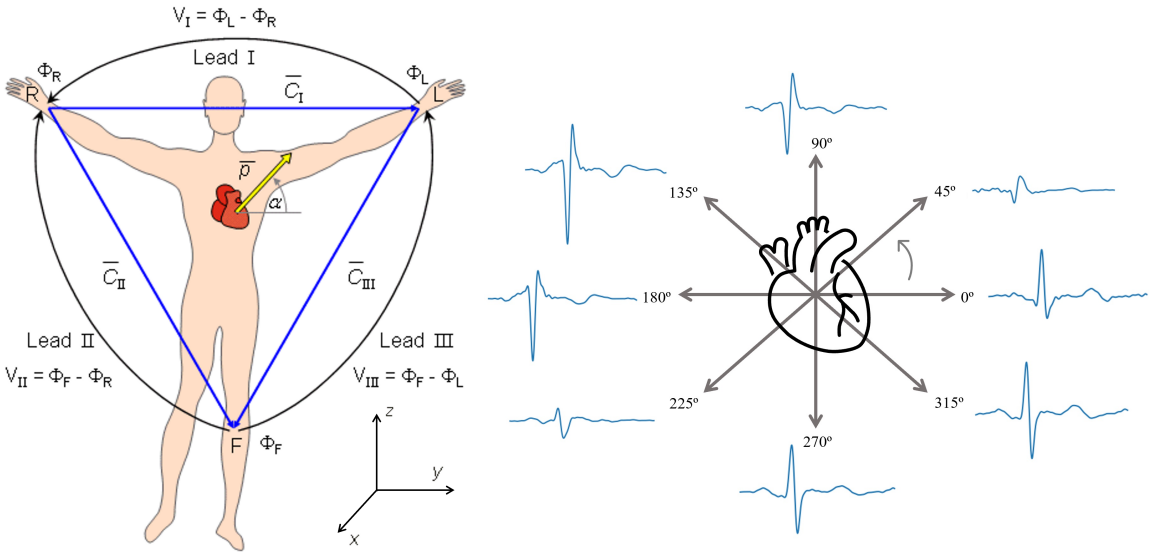
$$V_I = p\cos(\alpha) = p_y \tag{5.1}$$

$$V_{II} = \frac{p}{2}\cos(\alpha) - \frac{\sqrt{3}}{2}\sin\alpha = 0.5p_y - 0.87p_z \tag{5.2}$$

$$V_{III} = -\frac{\sqrt{p}}{2}\cos(\alpha) - \frac{\sqrt{3}}{2}p\sin(\alpha) = -0.5p_y - 0.87p_z \tag{5.3}$$

and we can verify that Kirchhoff 's voltage law is satisfied:

$$V_I + V_{III} = \frac{p}{2}\cos(\alpha) - \frac{\sqrt{3}}{2}\sin\alpha = V_{II} \tag{5.4}$$



(a) **(b)**

Figure 5.11: a) Einthoven limb leads and Einthoven triangle. Φ_L and Φ_R are, respectively, the potentials at the left and right arms, and Φ_F is the potential at the right leg [13]. b) ECG waveforms of leads measured at different angles of rotation of the stethoscope head, relative to the frontal plane.

The same assumption can be made for the ECG leads measured with the stethoscope on the chest, since they are in the frontal plane. To verify that the ECG leads obtained at different rotations also follow Kirchhoff 's law, eight signals were collected by rotating the stethoscope head counterclockwise in increments of 45º (Figure 5.11(b)). This data collection was performed always on the same auscultation

point, on the left third intercostal space. As expected, the signals measured with an angle difference of 180º have the same morphology, only inverted. Using the notation from Figure 5.11 we have, for example:

$$V_{0^\circ} = p\cos(\alpha) = p_y \tag{5.5}$$

$$V_{90^\circ} = p\sin(\alpha) = p_z \tag{5.6}$$

$$V_{45^\circ} = \frac{\sqrt{2}}{2}p\cos(\alpha) + \frac{\sqrt{2}}{2}p\sin(\alpha) = \frac{\sqrt{2}}{2}(V_0^\circ + V_{90^\circ}) \tag{5.7}$$

To compare the heartbeat waveform morphology, the ECG signals were segmented into individual beats by clipping the signals around the detected R-peaks (at -200 and 400 ms). Then, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was applied to detect the outlier beat waveforms (Figure 5.12), which were excluded from the analysis. DBSCAN is a clustering algorithm that groups together points that are close to each other (points with many nearby neighbors) according to a pre-defined distance metric, and marks points that are isolated in low-density regions as outliers. The cosine distance metric was used, due to its robustness to data scaling, with a maximum distance between two samples equal to 0.06 and a minimum number of samples per cluster equal to 2.

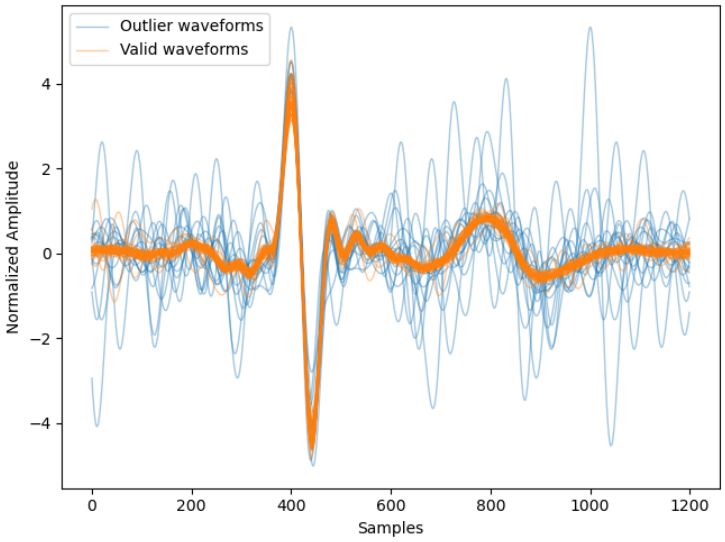


Figure 5.12: Example of use of DBSCAN to detect outlier waveforms in an ECG recording.

The Mean Squared Error (MSE) and Pearson's Correlation Coefficient (PCC) were calculated according to Equations 5.8 and 5.9 between the heartbeat waveforms of the experimental leads at different rotation angles and the equivalent waveforms obtained through Kirchhoff's law (from the adjacent ECG leads, at $\pm$ 45º).

Table 5.2: Pearson's Correlation Coefficient and Mean Squared Error between heartbeat waveforms of experimental leads at different rotation angles and the equivalent waveforms obtained through Kirchhoff 's law (from the adjacent ECG leads, at $\pm$ 45º)

Stethoscope angle	Pearson's Correlation Coefficient	Mean Squared Error
0º	0.948 $\pm$ 0.021	2.81E-3 $\pm$ 1.19E-3
45º	0.424 $\pm$ 0.210	3.99E-3 $\pm$ 1.01E-3
90º	0.974 $\pm$ 0.012	3.31E-3 $\pm$ 1.11E-3
135º	0.991 $\pm$ 0.005	1.49E-3 $\pm$ 7.99E-4
180º	0.967 $\pm$ 0.015	3.80E-3 $\pm$ 2.22E-3
225º	0.243 $\pm$ 0.264	4.55E-3 $\pm$ 1.81E-3
270º	0.985 $\pm$ 0.007	1.35E-3 $\pm$ 5.71E-4
315º	0.992 $\pm$ 0.005	1.28E-3 $\pm$ 7.27E-4

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \tag{5.8}$$

$$\text{PCC} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2}} \tag{5.9}$$

Pearson's Correlation Coefficient measures the linear correlation between the two waveforms and is invariant to data scaling, that is, invariant to the amplitude of the signals. The result is always a value between -1 and 1, and the further it is from 0 the higher the correlation. The results are presented in Table 5.2, and are in accordance with what can be observed in Figure 5.13. All the leads have a very high correlation and a low MSE, except for the ones obtained at rotation angles of 45º and 225º, which have low correlation coefficients. The most likely explanation for this is the fact that the peaks of the adjacent ECG leads at $\pm$ 45º have opposite signals. When combining the waves, the signal amplitudes cancel each other, which makes the resulting morphology more sensitive to small angle changes. As expected, these signals also have a much lower amplitude than the other leads.

This shows that the ECG leads measured with the stethoscope follow Kirchhoff's law, similarly to the Einthoven leads, meaning that to evaluate all the frontal plane components of the heart's electrical activity it is only necessary to perform two acquisitions in the same auscultation point with the stethoscope head at different angles. As a result, the design was further enhanced by adding two extra dry ECG electrodes perpendicular to the current ones. This allows the recording of two ECG leads at the same time, to obtain a full view of the frontal plane components with just one acquisition.

A preliminary evaluation to validate this modified design was done by splitting each half-moon electrode part into two separate electrodes, for a total of four electrodes on the stethoscope head. The positive electrodes were placed at the left and bottom sides (Figure 5.14). Four acquisitions were performed by rotating the stethoscope head counterclockwise at the angles of 0º, 45º, 180º, and 225º. Since there are two perpendicular ECG leads, these measurements correspond to a total of eight signals that span all the angles of rotation represented in Figure 5.11(b).

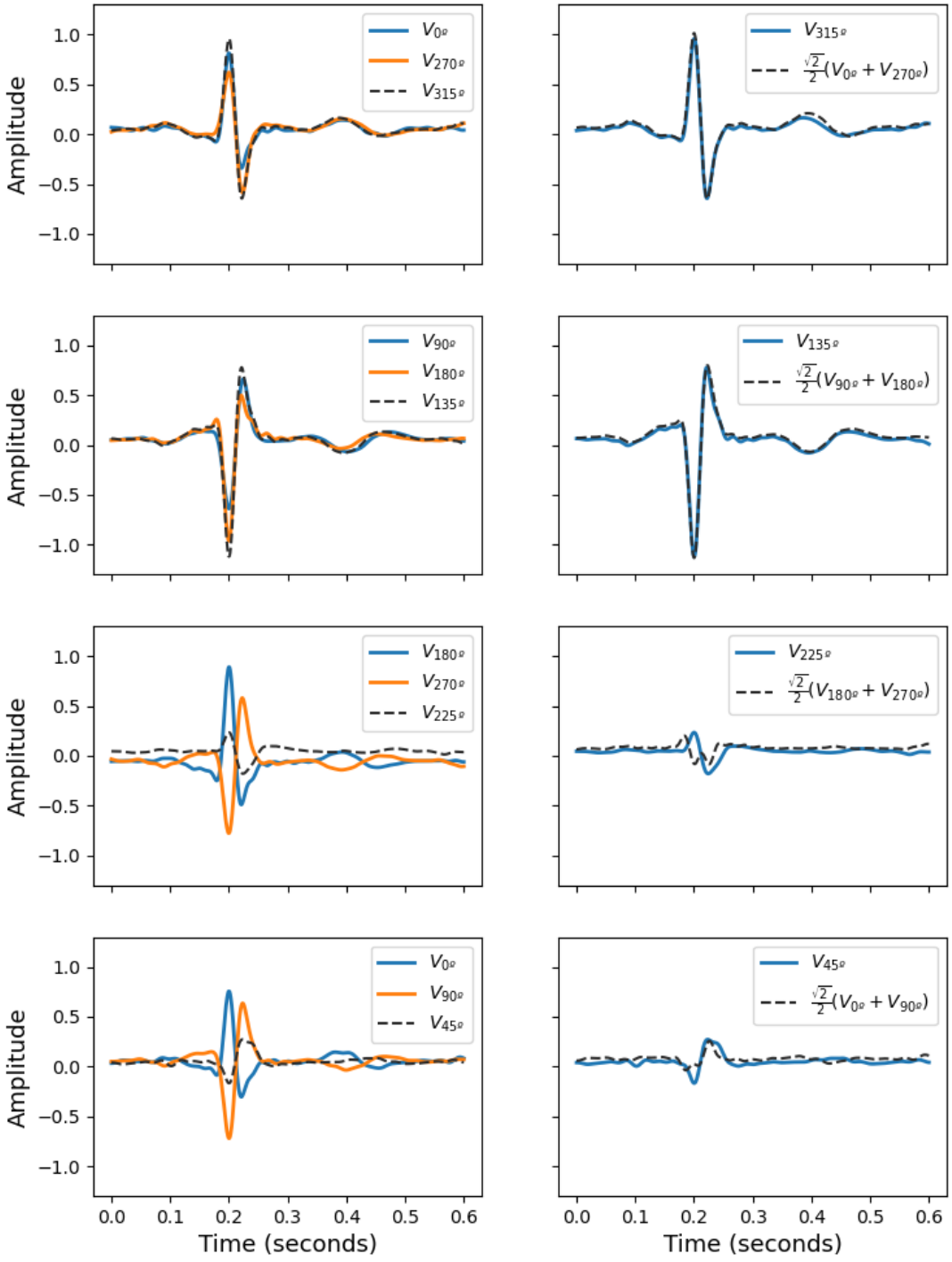


Figure 5.13: Example of heartbeat waveforms of experimental leads measured at different rotation angles and the equivalent waveforms obtained through Kirchhoff's law (from the adjacent ECG leads, at $\pm$ 45º).

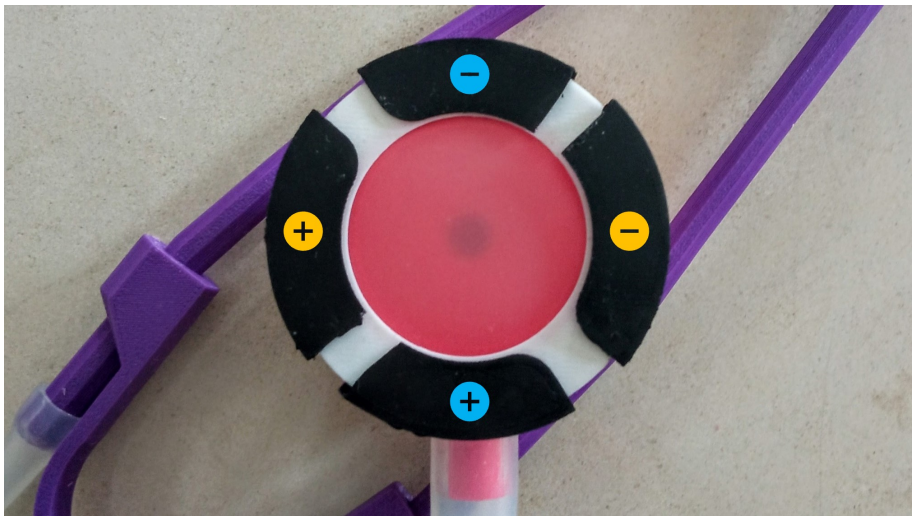


Figure 5.14: Disposition and polarity of two pairs of ECG dry electrodes on the stethoscope head, for the acquisition of two ECG leads. The leads with positive polarity are on the bottom and left sides.

Table 5.3: Pearson's Correlation Coefficient and Mean Squared Error between heartbeat waveforms of experimental ECGs at different rotation angles and the equivalent waveforms obtained through Kirchhoff's law (from the adjacent two lead ECG pair).

Stethoscope angle	Pearson's Correlation Coefficient	Mean Squared Error
0º + 270º = 315º	0.975 $\pm$ 0.014	3.93E-3 $\pm$ 1.79E-3
45º + 315º = 0º	0.959 $\pm$ 0.011	8.32E-3 $\pm$ 3.61E-3
180º + 90º = 135º	0.994 $\pm$ 0.002	4.18E-3 $\pm$ 2.91E-4
225º + 135º = 180º	0.967 $\pm$ 0.004	5.51E-3 $\pm$ 4.51E-4

Considering the positioning and polarity of the ECG electrodes on the stethoscope head, the acquired lead pairs are: 0º and 270º; 45º and 315º; 180º and 90º; 225º and 135º. The results of Table 5.3 and Figure 5.15 show that it is possible to compute additional leads from the two ECG leads measured perpendicularly. The calculated waveforms have a high correlation and low MSE when compared with the experimental measurements, although the results were slightly worse than the ones obtained for the single ECG leads (Table 5.2 and Figure 5.13). This could be because the electrode parts used in this initial study aren't completely symmetrical. By designing new electrodes with a symmetrical shape, the correlation between the acquired and computed signals should increase.

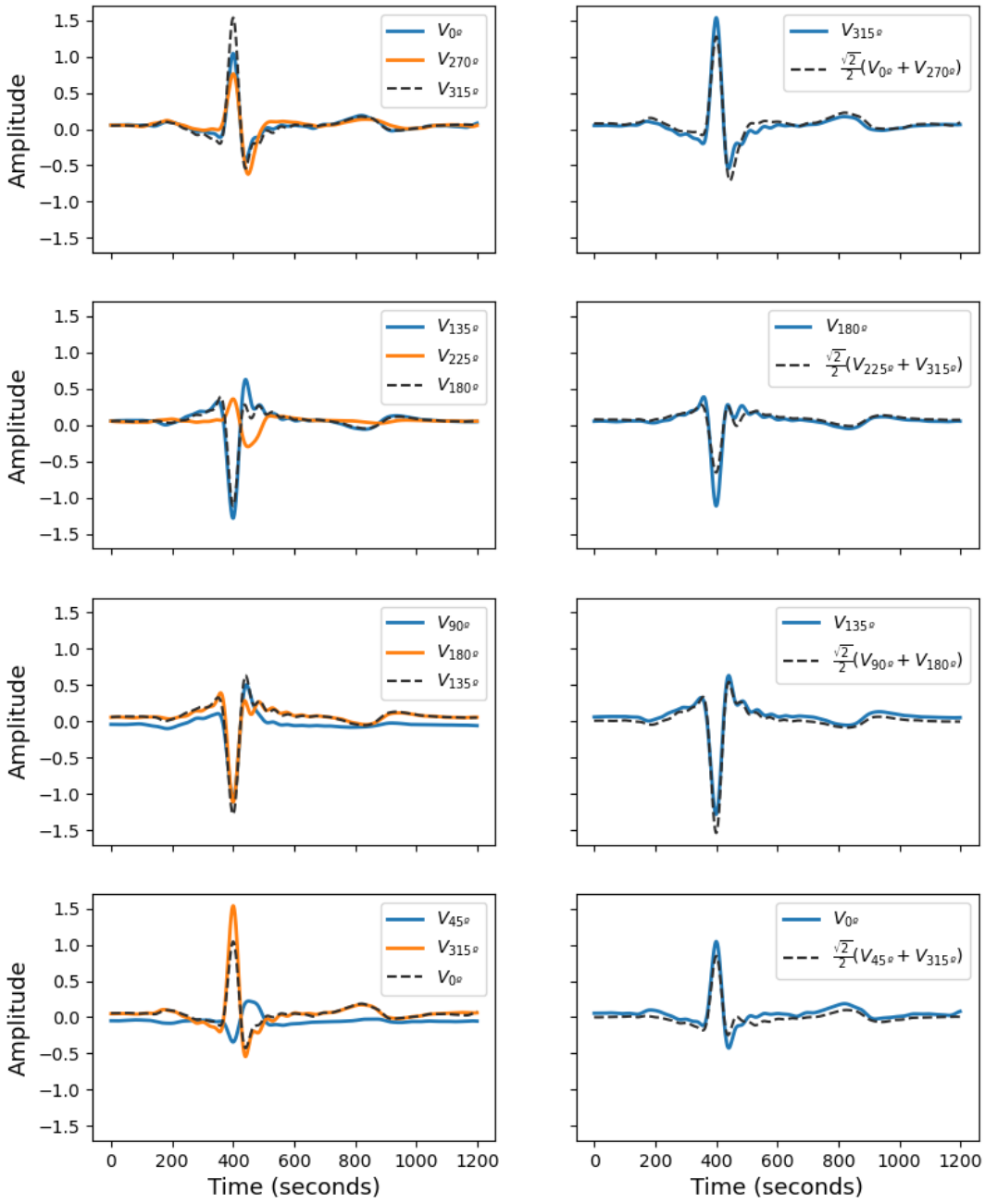


Figure 5.15: Example of heartbeat waveforms of experimental leads measured at different rotation angles and the equivalent waveforms obtained through Kirchhoff's law (from the adjacent two lead ECG pair).

5.7 Comparison of ECG leads with Lead I

The beat waveforms of the ECG leads obtained at different rotations were also compared with the ones of the standard Lead I, which was acquired simultaneously. This comparison was done using the same method as in Section 5.6: the ECGs were segmented into individual beats and, after outlier removal, the MSE and PCC were calculated for each pair of leads. All the rotation angles had a high correlation with the Lead I channel, as presented in Table 5.4.

Table 5.4: Pearson's Correlation Coefficient and Mean Squared Error between heartbeat waveforms of reference Lead I channel and experimental leads at different rotation angles.

Stethoscope angle	Pearson's Correlation Coefficient	Mean Squared Error
0º	0.876 $\pm$ 0.026	0.166 $\pm$ 0.049
45º	-0.854 $\pm$ 0.020	2.609 $\pm$ 0.409
90º	**-0.922 $\pm$ 0.016**	2.464 $\pm$ 0.262
135º	-0.887 $\pm$ 0.015	2.471 $\pm$ 0.209
180º	-0.875 $\pm$ 0.015	2.463 $\pm$ 0.262
225º	0.812 $\pm$ 0.020	0.228 $\pm$ 0.055
270º	**0.928 $\pm$ 0.007**	0.093 $\pm$ 0.014
315º	0.870 $\pm$ 0.014	0.186 $\pm$ 0.030

The MSE has a higher variation due to the polarity of the waveforms; leads with the same polarity as Lead I have a low MSE, while the ones measured with a 180º angle difference have a high MSE (but similar correlation coefficients). Surprisingly, the leads with the highest correlation were the ones obtained in the vertical axis, at 270º and 90º, instead of the ones obtained in the same direction as Lead I, in the horizontal axis. This can also be verified in Figure 5.16. A possible explanation is that the lead vectors created by the electrodes of the device (which are placed at a short distance from each other on the chest) don't span the same volume as the electrodes of Lead I (which are placed on the right and left arms). The projections of the electric heart vector on the lead vector will therefore be different, altering the morphology of the measured ECG signals. Nonetheless, this high correlation suggests that the ECG leads measured with the proposed device are clinically meaningful.

5.8 Discussion

These results show that the PCG and ECG are adequately acquired. Both signals have a different morphology depending on the point of auscultation, as expected. By using the proposed device on a standard auscultation procedure (through the 5 auscultation points), it would be possible to record 5 ECG leads that view the electrical activity of the heart from different perspectives and 5 PCG signals that will reflect the function of the different heart valves. This provides a high amount of additional information without altering the clinical routine.

The electrodes are fixed to the stethoscope head, reason for which the recorded single-lead ECG will

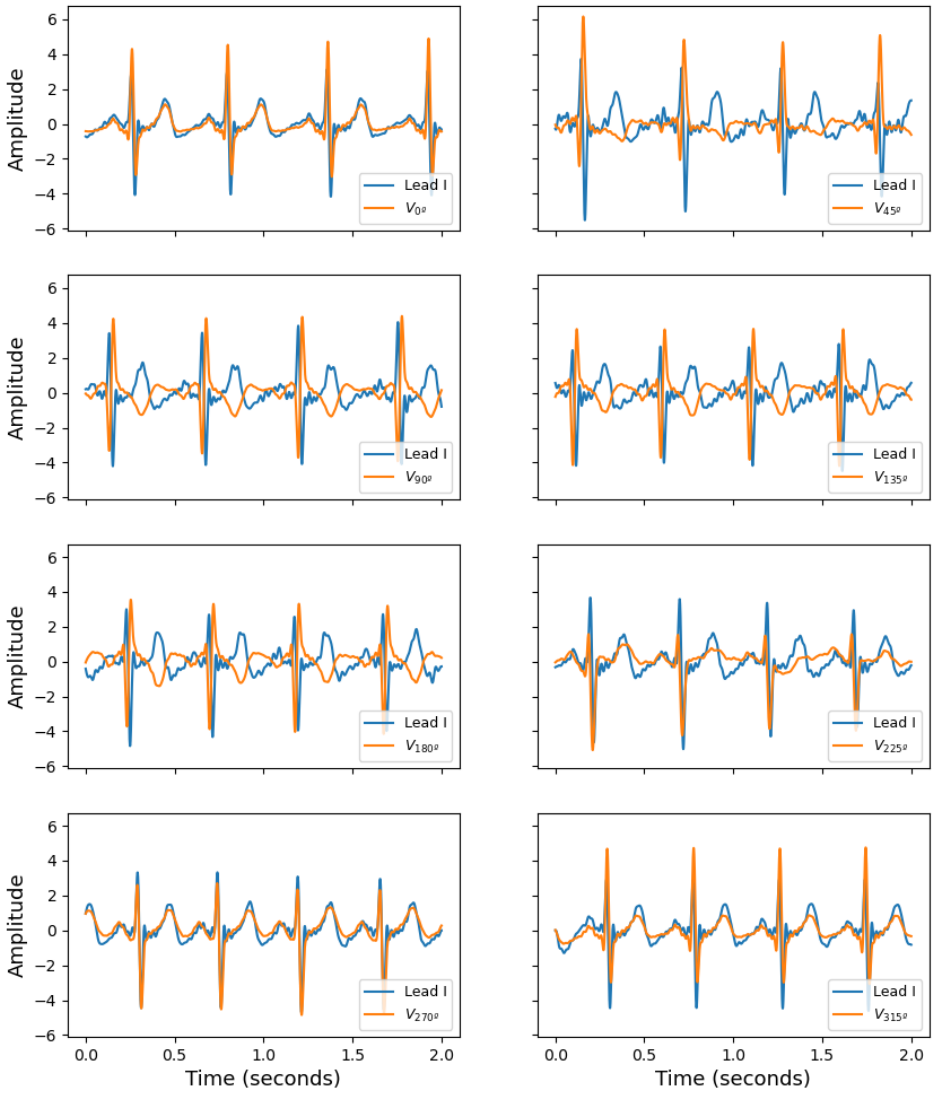


Figure 5.16: Example 2 seconds segments of ECGs acquired with experimental leads at different angles and acquired with the standard Lead I electrode positioning.

reflect a different projection of the electrical axis of the heart depending on the rotation of the device, and present a different morphology. As a result, an important contribution of this work is the experimental evaluation of the relationships between the leads obtained at different angles. They all have a high correlation with Lead I (with the highest correlation for the leads acquired along the vertical axis, at 270º and 90º), suggesting that they all contain meaningful information for CVD screening.

It was also found that the ECG leads measured with the stethoscope follow Kirchhoff's law, similarly to the Einthoven leads, meaning that to evaluate all the frontal plane components of the heart's electrical activity it is only necessary to perform two acquisitions in the same auscultation point with the stethoscope head at different angles. A preliminary evaluation was done by recording two ECG leads at the same time with the original electrodes split in two, and it was found that it is possible to calculate additional ECG leads in the frontal plane from just one acquisition. This motivates the creation of a new design for the electrodes, to have four identical parts along the ring in the stethoscope head.

Chapter 6

Conclusion

In this work, a detailed and in-depth analysis of the advantages of using simultaneous electrocardiography and phonocardiography for CVD screening was presented. This chapter outlines the main outcomes (Section 6.1) and introduces suggestions and possibilities for future work (Section 6.2).

6.1 Main Outcomes

CVDs incur a great burden in health systems and are responsible for premature deaths and a big decrease in quality of life, especially in low- and middle-income countries. Cardiac auscultation is the most common and cost-effective screening tool for CVDs. However, it is still mostly done with acoustic stethoscopes, which, despite their simplicity, have the disadvantage of needing to be operated by trained health professionals with enough clinical experience to master this skill. Electronic stethoscopes and automatic decision systems for disease screening could significantly speed up treatment and referrals, improving the patient's outcome [4].

Combining ECG and PCG acquisition in the same device allows the simultaneous inspection of the electrical and mechanical condition of the heart, further improving the accuracy of the heart disease diagnosis. The approach developed in this work enables this by incorporating ECG and PCG sensors in a validated 3D-printed acoustic stethoscope without significantly altering its form factor or acoustic transfer properties, which facilitates the integration of the new device in current clinical practice.

The implementation of deep learning architectures for classification of signals acquired during auscultation showed promising results in different tasks, including murmur detection from heart sounds and abnormality detection from multi-modal PCG and ECG data. In the latter, the developed hybrid neural network based on parallel CNN and RNN blocks had a performance at the level of other state-of-the-art methods, with an average F_1 score of 90.5%.

Overall, this evaluation demonstrated promising results for further work. The low-cost, portability, and simplicity of the proposed device, allied with computer-aided and deep learning techniques for automatic diagnosis, could significantly facilitate the early screening of CVDs. We believe this system would have the biggest impact in primary care facilities as part of the auscultation routine, in telemedicine, or in areas with difficult access to an integrated health system, where an automatic screening tool that isn't dependent on trained professionals could significantly reduce the mortality and morbidity associated with both congenital and acquired cardiopathies. Furthermore, the device is open-source, which makes it more accessible and easily upgradeable. The ability to collaborate, share the technology, and save the acquired signals for further analysis and model development could also make the device an important

tool in both research and education.

6.2 Future Work

The work developed for this thesis has established a basis for further development. Regarding the developed prototype, the next step should be a full, comprehensive evaluation of the device in a real-world setting, with both healthy subjects and patients with diagnosed cardiovascular pathologies. It is especially important to do this evaluation in comparison with standard medical tools for cardiac monitoring (such as the 12-lead ECG), in order to validate the clinical usefulness of the new device. An application has already been made to the health ethics committee of Centro Hospitalar Vila Nova de Gaia/Espinho (CHVNG/E) to evaluate the device and sensors in a hospital environment with cardiac patients.

Future work could also include improvements to the hardware, for example, by combining the ECG and PCG sensors in a single printed circuit board, or by including additional modalities, such as Functional Near-Infrared Spectroscopy (fNIRS) or Photoplethysmography (PPG), for an even more comprehensive study of the cardiac function.

This can also be articulated with the development of new datasets for the training of deep learning algorithms for automatic CVD screening based on the multi-modal data. With a larger dataset and more detailed annotations, it is expected that the performance of the current deep learning models will significantly increase, and also that they will be able to generalize better for new examples.

More improvements could be done by combining the current models with attention layers, which would not only improve the models' performance (since the neural networks would be able to better extract the most relevant features in the signals and better estimate their contribution for the final classification), but also enhance the interpretability and explainability of the results, which is highly relevant in healthcare applications.

Bibliography

[1] WHO. Cardiovascular diseases (CVD). Accessed: 22 Aug, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] M. E. Chowdhury, A. Khandakar, K. Alzoubi, S. Mansoor, A. M. Tahir, M. B. I. Reaz, and N. A. Emadi, "Real-time smart-digital stethoscope system for heart diseases monitoring," *Sensors*, vol. 19, no. 12, 2019. [Online]. Available: https://doi.org/10.3390/s19122781

[3] S. Leng, R. S. Tan, K. T. C. Chai, C. Wang, D. Ghista, and L. Zhong, "The electronic stethoscope," *BioMedical Engineering Online*, vol. 14, no. 66, 2015. [Online]. Available: https://doi.org/10.1186/s12938-015-0056-y

[4] J. Oliveira, F. Renna, P. D. Costa, M. Nogueira, C. Oliveira, C. Ferreira, A. Jorge, S. Mattos, T. Hatem, T. Tavares, A. Elola, A. B. Rad, R. Sameni, G. D. Clifford, and M. T. Coimbra, "The CirCor DigiScope Dataset: From Murmur Detection to Murmur Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 6, pp. 2524–2535, 2022. [Online]. Available: https://doi.org/10.1109/JBHI.2021.3137048

[5] F. Renna, J. Oliveira, and M. T. Coimbra, "Deep Convolutional Neural Networks for Heart Sound Segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2435–2445, 2019. [Online]. Available: https://doi.org/10.1109/JBHI.2019.2894222

[6] W. Phanphaisarn, A. Roeksabutr, P. Wardkein, J. Koseeyaporn, and P. Yupapin, "Heart detection and diagnosis based on ECG and EPCG relationships," *Medical Devices: Evidence and Research*, vol. 4, no. 1, pp. 133–144, 2011. [Online]. Available: https://doi.org/10.2147/MDER.S23324

[7] R. Hettiarachchi, U. Haputhanthri, K. Herath, H. Kariyawasam, S. Munasinghe, K. Wickramasinghe, D. Samarasinghe, A. Silva, and C. U. S. Edussooriya, "A Novel Transfer Learning-Based Approach for Screening Pre-Existing Heart Diseases Using Synchronized ECG Signals and Heart Sounds," in *Proc. of the Int'l Symp. on Circuits and Systems (ISCAS)*, 2021, pp. 1–5. [Online]. Available: https://doi.org/10.1109/ISCAS51556.2021.9401093

[8] F. Chakir, A. Jilbab, C. Nacir, and A. Hammouch, "Recognition of cardiac abnormalities from synchronized ECG and PCG signals," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 673–677, 2020. [Online]. Available: https://doi.org/10.1007/s13246-020-00875-2

[9] S. M. Monteiro, A. Fred, and H. P. da Silva, “Detection of Heart Sound Murmurs and Clinical Outcome with Bidirectional Long Short-Term Memory Networks,” in *2022 Computing in Cardiology Conference (CinC)*. Preprint, 2022.

[10] A. L. N. Fred, C. Rodrigues, and H. P. Silva, “Electrophysiology of the Heart and the Electrocardiogram: Visual Depictions,” Institute of Telecommunications, Tech. Rep., 2020. [Online]. Available: https://www.researchgate.net/publication/346910888

[11] R. M. Rangayyan, *Biomedical Signal Analysis*, 2nd ed. IEEE Press Series in Biomedical Engineering, 2015.

[12] P. Podrid, R. Malhotra, R. Kakkar, P. A. Noseworthy, H. J. Wellens, and R. Desanctis, *Podrid's Real-World A Master's Approach to the Art and Practice of Clinical ECG Interpretation. Volume 1 - The Basics*. Cardiotext Publishing, 2012.

[13] J. Malmivuo and R. Plonsey, *Bioelectromagnetism Principles and Applications of Bioelectric and Biomagnetic Fields*. New York Oxford, Oxford University Press, 1995.

[14] M. Martins, P. Gomes, C. Oliveira, M. Coimbra, and H. P. Silva, “Design and Evaluation of a Diaphragm for Electrocardiography in Electronic Stethoscopes,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 391–398, 2020. [Online]. Available: https://doi.org/10.1109/TBME.2019.2913913

[15] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, “Multi-class Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network,” *Information Fusion*, vol. 53, pp. 174–182, 2020. [Online]. Available: https://doi.org/10.1016/j.inffus.2019.06.024

[16] M. P. Witvliet, E. P. Karregat, J. C. Himmelreich, J. S. de Jong, W. A. Lucassen, and R. E. Harskamp, “Usefulness, pitfalls and interpretation of handheld single-lead electrocardiograms,” *Journal of Electrocardiology*, vol. 66, pp. 33–37, 2021. [Online]. Available: https://doi.org/10.1016/j.jelectrocard.2021.02.011

[17] E. P. Karregat, J. C. Himmelreich, W. A. Lucassen, W. B. Busschers, H. C. van Weert, and R. E. Harskamp, “Evaluation of general practitioners' single-lead electrocardiogram interpretation skills: a case-vignette study,” *Family Practice*, vol. 38, pp. 70–75, 3 2021. [Online]. Available: https://doi.org/10.1093/fampra/cmaa076

[18] J. C. Himmelreich, E. P. Karregat, W. A. Lucassen, H. C. van Weert, J. R. de Groot, M. L. Handoko, R. Nijveldt, and R. E. Harskamp, “Diagnostic Accuracy of a Smartphone-Operated, Single-Lead Electrocardiography Device for Detection of Rhythm and Conduction Abnormalities in Primary Care,” *Annals of Family Medicine*, vol. 17, no. 5, pp. 403–411, 2019. [Online]. Available: https://doi.org/10.1370/afm.2438

[19] L. Jin and J. Dong, “Normal Versus Abnormal ECG Classification by the Aid of Deep Learning,” in *Artificial Intelligence*, M. A. Aceves-Fernandez, Ed. Rijeka: IntechOpen, 2018, ch. 15. [Online]. Available: https://doi.org/10.5772/intechopen.75546

[20] V. N. Varghees and K. I. Ramachandran, “A novel heart sound activity detection framework for automated heart sound analysis,” *Biomedical Signal Processing and Control*, vol. 13, pp. 174–188, 2014. [Online]. Available: https://doi.org/10.1016/j.bspc.2014.05.002

[21] M. Waqar, S. Inam, M. A. ur Rehman, M. Ishaq, M. Afzal, N. Tariq, F. Amin, and Q. ul Ain, “Arduino Based Cost-Effective Design and Development of a Digital Stethoscope,” *Proc. of the Int’l Conf. on Emerging Technologies (ICET)*, pp. 1–6, 2019. [Online]. Available: https://doi.org/10.1109/ICET48972.2019.8994674

[22] S. R. McGee, *Evidence-Based Physical Diagnosis*, 4th ed. Philadelphia: Elsevier, 2018.

[23] S. Debal, L. H. Cherif, and F. Meziani, “Graphical Analysis of the Pathologies of the Cardiac Sounds,” *Annals of Cardiovascular Diseases*, vol. 4, no. 1, 2020.

[24] M. C. Kusko and K. Maselli, *Introduction to Cardiac Auscultation*. London: Springer London, 2015, pp. 3–14. [Online]. Available: https://doi.org/10.1007/978-1-4471-6738-9_1

[25] S. Ismail, I. Siddiqi, and U. Akram, “Localization and classification of heart beats in phonocardiography signals —a comprehensive review,” *Eurasip Journal on Advances in Signal Processing*, vol. 2018, no. 1, 2018. [Online]. Available: https://doi.org/10.1186/s13634-018-0545-9

[26] C. Liu, D. Springer, Q. Li, B. Moody, R. A. Juan, F. J. Chorro, F. Castells, J. M. Roig, I. Silva, A. E. Johnson, Z. Syed, S. E. Schmidt, C. D. Papadaniil, L. Hadjileontiadis, H. Naseri, A. Moukadem, A. Dieterlen, C. Brandt, H. Tang, M. Samieinasab, M. R. Samieinasab, R. Sameni, R. G. Mark, and G. D. Clifford, “An open access database for the evaluation of heart sound algorithms,” *Physiological Measurement*, vol. 37, no. 12, pp. 2181–2213, 2016. [Online]. Available: https://doi.org/10.1088/0967-3334/37/12/2181

[27] D. B. Springer, L. Tarassenko, and G. D. Clifford, “Logistic regression-HSMM-based heart sound segmentation,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 822–832, 2016. [Online]. Available: https://doi.org/10.1109/TBME.2015.2475278

[28] F. Safara, S. Doraisamy, A. Azman, A. Jantan, and A. R. A. Ramaiah, “Multi-level basis selection of wavelet packet decomposition tree for heart sound classification,” *Computers in Biology and Medicine*, vol. 43, no. 10, pp. 1407–1414, 2013. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2013.06.016

[29] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction Second Edition*, 2nd ed. Springer Series in Statistics, 2001. [Online]. Available: http://www.springer.com/series/692

[30] D. G. Stork, R. O. Duda, and P. E. Hart, *Pattern Classification*, 2nd ed. John Wiley and Sons, 2000.

[31] D. Graupe, *Principles of artificial neural networks*, 2nd ed., ser. Advanced series on circuits and systems. World Scientific, 2007, vol. 6.

[32] Y. Bengio, A. Courville, and I. J. Goodfellow, *Deep learning*, ser. Adaptive Computation and Machine Learning series. The MIT Press, 2016.

[33] W. Chen, Q. Sun, X. Chen, G. Xie, H. Wu, and C. Xu, "Deep Learning Methods for Heart Sounds Classification: A Systematic Review," *Entropy*, vol. 23, no. 6, 2021. [Online]. Available: https://doi.org/10.3390/e23060667

[34] C. Olah, "Understanding LSTM Networks," 2015, Accessed: 31 Jan, 2021. [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[35] S. A. Hicks, I. Strümke, V. Thambawita, M. Hammou, M. A. Riegler, P. Halvorsen, and S. Parasa, "On evaluation metrics for medical applications of artificial intelligence," *Scientific Reports*, vol. 12, 12 2022. [Online]. Available: https://doi.org/10.1038/s41598-022-09954-8

[36] S. Adhikari, S. L. Normand, J. Bloom, D. Shahian, and S. Rose, "Revisiting performance metrics for prediction with rare outcomes," *Statistical Methods in Medical Research*, vol. 30, pp. 2352–2366, 10 2021. [Online]. Available: https://doi.org/10.1177/09622802211038754

[37] G. D. Clifford, C. Liu, B. Moody, L. H. Lehman, I. Silva, Q. Li, A. E. Johnson, and R. G. Mark, "AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017," in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1–4. [Online]. Available: https://doi.org/10.22489/CinC.2017.065-469

[38] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," *Computers in Biology and Medicine*, vol. 122, p. 103801, 2020. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2020.103801

[39] R. Tripathy, M. R. Paternina, J. G. Arrieta, A. Zamora-Méndez, and G. R. Naik, "Automated detection of congestive heart failure from electrocardiogram signal using Stockwell transform and hybrid classification scheme," *Computer Methods and Programs in Biomedicine*, vol. 173, pp. 53–65, 2019. [Online]. Available: https://doi.org/10.1016/j.cmpb.2019.03.008

[40] C. Han, Y. Song, H.-S. Lim, Y. Tae, J.-H. Jang, B. T. Lee, Y. Lee, W. Bae, and D. Yoon, "Automated Detection of Acute Myocardial Infarction Using Asynchronous Electrocardiogram Signals—Preview of Implementing Artificial Intelligence With Multichannel Electrocardiographs Obtained From Smartwatches: Retrospective Study," *Journal of Medical Internet Research*, vol. 23, no. 9, p. e31129, Sep 2021. [Online]. Available: https://doi.org/10.2196/31129

[41] J. Huang, B. Chen, B. Yao, and W. He, "ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network," *IEEE Access*, vol. 7, pp. 92 871–92 880, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2928017

[42] F. Li, Y. Xu, Z. Chen, and Z. Liu, "Automated Heartbeat Classification Using 3-D Inputs Based on Convolutional Neural Network With Multi-Fields of View," *IEEE Access*, vol. 7, pp. 76 295–76 304, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2921991

[43] D. Li, J. Zhang, Q. Zhang, and X. Wei, "Classification of ECG signals based on 1D convolution neural network," in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2017, pp. 1–6. [Online]. Available: https://doi.org/10.1109/HealthCom.2017.8210784

[44] C.-H. Hsieh, Y.-S. Li, B.-J. Hwang, and C.-H. Hsiao, "Detection of Atrial Fibrillation Using 1D Convolutional Neural Network," *Sensors*, vol. 20, no. 7, 2020. [Online]. Available: https://doi.org/10.3390/s20072136

[45] S. Saadatnejad, M. Oveisi, and M. Hashemi, "LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 515–523, 2020. [Online]. Available: https://doi.org/10.1109/JBHI.2019.2911367

[46] A. Lyon, A. Mincholé, J. P. Martínez, P. Laguna, and B. Rodriguez, "Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances," *Journal of the Royal Society Interface*, vol. 15, Jan 2018. [Online]. Available: https://doi.org/10.1098/rsif.2017.0821

[47] R. Li, X. Zhang, H. Dai, B. Zhou, and Z. Wang, "Interpretability Analysis of Heartbeat Classification Based on Heartbeat Activity's Global Sequence Features and BiLSTM-Attention Neural Network," *IEEE Access*, vol. 7, pp. 109 870–109 883, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2933473

[48] T. Teijeiro, C. A. García, D. Castro, and P. Félix, "Arrhythmia classification from the abductive interpretation of short single-lead ECG records," in *2017 Computing in Cardiology Conference (CinC)*, vol. 44. IEEE, 2017, pp. 1–4. [Online]. Available: https://doi.org/10.22489/CinC.2017.166-054

[49] R. He, Y. Liu, K. Wang, N. Zhao, Y. Yuan, Q. Li, and H. Zhang, "Automatic Cardiac Arrhythmia Classification Using Combination of Deep Residual Network and Bidirectional LSTM," *IEEE Access*, vol. 7, pp. 102 119–102 135, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2931500

[50] S. Hong, C. Xiao, T. Ma, H. Li, and J. Sun, "MINA: Multilevel Knowledge-Guided Attention for Modeling Electrocardiography Signals," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint

Conferences on Artificial Intelligence Organization, 7 2019, pp. 5888–5894. [Online]. Available: https://doi.org/10.24963/ijcai.2019/816

[51] L. Guo, G. Sim, and B. Matuszewski, "Inter-patient ECG classification with convolutional and recurrent neural networks," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 3, pp. 868–879, 2019. [Online]. Available: https://doi.org/10.1016/j.bbe.2019.06.001

[52] O. Yildirim, U. B. Baloglu, R.-S. Tan, E. J. Ciaccio, and U. R. Acharya, "A new approach for arrhythmia classification using deep coded features and LSTM networks," *Computer Methods and Programs in Biomedicine*, vol. 176, pp. 121–133, 2019. [Online]. Available: https://doi.org/10.1016/j.cmpb.2019.05.004

[53] P. Liu, X. Sun, Y. Han, Z. He, W. Zhang, and C. Wu, "Arrhythmia classification of LSTM autoencoder based on time series anomaly detection," *Biomedical Signal Processing and Control*, vol. 71, p. 103228, 2022. [Online]. Available: https://doi.org/10.1016/j.bspc.2021.103228

[54] L. Shan, Y. Li, H. Jiang, P. Zhou, J. Niu, R. Liu, Y. Wei, J. Peng, H. Yu, X. Sha, and S. Chang, "Abnormal ECG detection based on an adversarial autoencoder," *Frontiers in Physiology*, vol. 13, 2022. [Online]. Available: https://doi.org/10.3389/fphys.2022.961724

[55] Z. Zhou, X. Zhai, and C. Tin, "Fully automatic electrocardiogram classification system based on generative adversarial network with auxiliary classifier," *Expert Systems with Applications*, vol. 174, p. 114809, 2021. [Online]. Available: https://doi.org/10.1016/j.eswa.2021.114809

[56] P. Wang, B. Hou, S. Shao, and R. Yan, "ECG Arrhythmias Detection Using Auxiliary Classifier Generative Adversarial Network and Residual Network," *IEEE Access*, vol. 7, pp. 100 910–100 922, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2930882

[57] T. Golany, G. Lavee, S. Tejman Yarden, and K. Radinsky, "Improving ECG Classification Using Generative Adversarial Networks," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 08, pp. 13 280–13 285, Apr. 2020. [Online]. Available: https://doi.org/10.1609/aaai.v34i08.7037

[58] G. D. Clifford, C. Liu, B. Moody, J. Millet, S. Schmidt, Q. Li, I. Silva, and R. G. Mark, "Recent advances in heart sound analysis," *Physiological Measurement*, vol. 38, no. 8, pp. E10–E25, 2017. [Online]. Available: https://doi.org/10.1088/1361-6579/aa7ec8

[59] M. Xu, L.-Y. Duan, J. Cai, L.-T. Chia, C. Xu, and Q. Tian, "HMM-Based Audio Keyword Generation," in *Advances in Multimedia Information Processing - PCM 2004*. Berlin, Heidelberg: Springer, 2005, pp. 566–574. [Online]. Available: https://doi.org/10.1007/978-3-540-30543-9_71

[60] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, "Classifying Heart Sound Recordings using Deep Convolutional Neural Networks and Mel-Frequency Cepstral Coefficients," in *2016 Computing in Cardiology Conference (CinC)*, 2016, pp. 813–816.

[61] T. Nilanon, J. Yao, J. Hao, S. Purushotham, and Y. Liu, “Normal/Abnormal Heart Sound Recordings Classification Using Convolutional Neural Network,” in *2016 Computing in Cardiology Conference (CinC)*, 2016, pp. 585–588.

[62] V. Maknickas and A. Maknickas, “Recognition of normal-abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients,” *Physiological Measurement*, vol. 38, no. 8, 2017. [Online]. Available: https://doi.org/10.1088/1361-6579/aa7841

[63] F. Demir, A. Şengür, V. Bajaj, and K. Polat, “Towards the classification of heart sounds based on convolutional deep neural network,” *Health Information Science and Systems*, vol. 7, no. 19, 2019. [Online]. Available: https://doi.org/10.1007/s13755-019-0078-0

[64] Y. Xu, B. Xiao, X. Bi, W. Li, J. Zhang, and X. Ma, “Pay More Attention With Fewer Parameters: A Novel 1-D Convolutional Neural Network for Heart Sounds Classification,” in *2018 Computing in Cardiology Conference (CinC)*, vol. 45, 2018, pp. 1–4. [Online]. Available: https://doi.org/10.22489/CinC.2018.072

[65] F. Li, M. Liu, Y. Zhao, L. Kong, L. Dong, X. Liu, and M. Hui, “Feature extraction and classification of heart sound using 1D convolutional neural networks,” *Eurasip Journal on Advances in Signal Processing*, vol. 2019, 12 2019. [Online]. Available: https://doi.org/10.1186/s13634-019-0651-3

[66] S. Latif, M. Usman, R. Rana, and J. Qadir, “Phonocardiographic Sensing Using Deep Learning for Abnormal Heartbeat Detection,” *IEEE Sensors Journal*, vol. 18, no. 22, pp. 9393–9400, 2018. [Online]. Available: https://doi.org/10.1109/JSEN.2018.2870759

[67] T.-c. I. Yang and H. Hsieh, “Classification of Acoustic Physiological Signals Based on Deep Learning Neural Networks with Augmented Features,” in *2016 Computing in Cardiology Conference (CinC)*, 2016, pp. 569–572.

[68] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B.-W. On, “Heartbeat Sound Signal Classification Using Deep Learning,” *Sensors*, vol. 19, no. 21, 2019. [Online]. Available: https://doi.org/10.3390/s19214819

[69] M. Alkhodari and L. Fraiwan, “Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings,” *Computer Methods and Programs in Biomedicine*, vol. 200, 3 2021. [Online]. Available: https://doi.org/10.1016/j.cmpb.2021.105940

[70] Yaseen, G.-Y. Son, and S. Kwon, “Classification of heart sound signal using multiple features,” *Applied Sciences*, vol. 8, no. 12, 2018. [Online]. Available: https://www.mdpi.com/2076-3417/8/12/2344

[71] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, “Heart sound classification based on improved MFCC features and convolutional recurrent neural networks,” *Neural Networks*, vol. 130, pp. 22–32, 10 2020.

[72] F. Noman, C.-M. Ting, S.-H. Salleh, and H. Ombao, “Short-segment Heart Sound Classification Using an Ensemble of Deep Convolutional Neural Networks,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 1318–1322.

[73] S. A. Singh, S. A. Singh, N. D. Devi, and S. Majumder, “Heart abnormality classification using PCG and ECG recordings,” *Computacion y Sistemas*, vol. 25, no. 2, pp. 381–391, 2021. [Online]. Available: https://doi.org/10.13053/CyS-25-2-3447

[74] J. Li, L. Ke, Q. Du, X. Chen, and X. Ding, “Multi-modal cardiac function signals classification algorithm based on improved D-S evidence theory,” *Biomedical Signal Processing and Control*, vol. 71, 1 2022.

[75] A. Das, “Knowledge representation,” in *Encyclopedia of Information Systems*, H. Bidgoli, Ed. New York: Elsevier, 2003, pp. 33–41.

[76] P. Bentley, G. Nordehn, M. Coimbra, S. Mannor, and R. Getz, “The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results,” accessed: 10 Feb, 2022. [Online]. Available: https://www.peterjbentley.com/heartchallenge/

[77] E. Messner, M. Zöhrer, and F. Pernkopf, “Heart sound segmentation - An event detection approach using deep recurrent neural networks,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1964–1974, 2018. [Online]. Available: https://doi.org/10.1109/TBME.2018.2843258

[78] E. Bondareva, T. Xia, J. Han, and C. Mascolo, “Towards Uncertainty-Aware Murmur Detection in Heart Sounds via Tandem Learning,” in *2022 Computing in Cardiology Conference (CinC)*. Preprint, 2022.

[79] D. Springer, “Logistic regression-HSMM-based heart sound segmentation (version 1.0),” *PhysioNet*, 2019. [Online]. Available: https://doi.org/10.13026/vnt9-kf93

[80] J. H. Oliveira, F. Renna, P. Costa, D. Nogueira, C. Oliveira, C. Ferreira, A. Jorge, S. Mattos, T. Hatem, T. Tavares, A. Elola, A. Rad, R. Sameni, G. D. Clifford, and M. T. Coimbra, “The CirCor DigiScope Phonocardiogram Dataset (version 1.0.3),” *PhysioNet*, 2022, https://doi.org/10.13026/tshs-mw03.

[81] A. Kazemnejad, P. Gordany, and R. Sameni, “An Open-Access Simultaneous Electrocardiogram and Phonocardiogram Database,” *bioRxiv*, 2021. [Online]. Available: https://doi.org/10.1101/2021.05.17.444563

[82] C. Landgraf, P. Goolkasian, and T. Crouch, “Wireless cardiac sensor,” U.S. Patent US20 180 256 061A1, 2017. [Online]. Available: https://patents.google.com/patent/US20180256061A1

[83] R. Kappor, “Mobile front-end system for comprehensive cardiac diagnosis,” U.S. Patent US9 492 138B2, 2016. [Online]. Available: https://patents.google.com/patent/US9492138B2

[84] C. Life. Coala Heart Monitor. Accessed: 12 Sep, 2022. [Online]. Available: https://www.coalalife.com/en/clinician/our-solution/coala-heart-monitor/

[85] M. Klum, M. Urban, T. Tigges, A. G. Pielmus, A. Feldheiser, T. Schmitt, and R. Orglmeister, "Wearable Cardiorespiratory Monitoring Employing a Multimodal Digital Patch Stethoscope: Estimation of ECG, PEP, LVET and Respiration Using a 55 mm Single-Lead ECG and Phonocardiogram," *Sensors (Switzerland)*, vol. 20, 4 2020. [Online]. Available: https://doi.org/10.3390/s20072033

[86] Y.-J. Lin, C.-W. Chuang, C.-Y. Yen, S.-H. Huang, P.-W. Huang, J.-Y. Chen, and S.-Y. Lee, "An Intelligent Stethoscope with ECG and Heart Sound Synchronous Display," in *Proc. of the IEEE Int'l Symp. on Circuits and Systems (ISCAS)*, 2019, pp. 1–4. [Online]. Available: https://doi.org/10.1109/ISCAS.2019.8702481

[87] E. Health. Eko DUO ECG + Digital Stethoscope. Accessed: 12 Sep, 2022. [Online]. Available: https://shop.ekohealth.com/products/duo-ecg-digital-stethoscope

[88] M. Engineering. CardioSleeve – ECG & Heart Sound Stethoscope Add-On. Accessed: 12 Sep, 2022. [Online]. Available: https://www.2mel.nl/projects/rijuven-cardiosleeve-ecg-heart-sound-stethoscope/

[89] M. A. Reyna, Y. Kiarashi, A. Elola, J. Oliveira, F. Renna, A. Gu, E. A. Perez-Alday, N. Sadr, A. Sharma, S. Mattos, M. T. Coimbra, R. Sameni, A. B. Rad, and G. D. Clifford, "Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022," *medRxiv*, 2022. [Online]. Available: https://doi.org/10.1101/2022.08.11.22278688

[90] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden Markov model," *Physiological Measurement*, vol. 31, no. 4, pp. 513–529, 2010. [Online]. Available: https://doi.org/10.1088/0967-3334/31/4/004

[91] D. O'Shaughnessy, *Speech Communications: Human and Machine*, 2nd ed. Wiley-IEEE Press, 1999.

[92] McFee, Brian, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in Python." in *Python in Science Conferences*, vol. 14, 2015, pp. 18–25.

[93] K. Sreenivasa and R. Manjunath, *Speech Recognition Using Articulatory and Excitation Source Features*. Springer, 2017, ch. Appendix A - MFCC Features.

[94] Q. ul Ain Mubarak, M. U. Akram, A. Shaukat, F. Hussain, S. G. Khawaja, and W. H. Butt, "Analysis of PCG signals using quality assessment and homomorphic filters for localization and classification of heart sounds," *Computer Methods and Programs in Biomedicine*, vol. 164, pp. 143–157, 2018. [Online]. Available: https://doi.org/10.1016/j.cmpb.2018.07.006

[95] Y. Chen, Y. Sun, J. Lv, B. Jia, and X. Huang, “End-to-end heart sound segmentation using deep convolutional recurrent network,” *Complex & Intelligent Systems*, vol. 7, no. 4, pp. 2103–2117, 2021. [Online]. Available: https://doi.org/10.1007/s40747-021-00325-w

[96] J. J. G. Ortiz, C. P. Phoo, and J. Wiens, “Heart sound classification based on temporal alignment techniques,” in *2016 Computing in Cardiology Conference (CinC)*, 2016, pp. 589–592.

[97] E. Alpaydin, V. Cheplygina, M. Loog, and D. M. Tax, “Single- vs. multiple-instance classification,” *Pattern Recognition*, vol. 48, no. 9, pp. 2831–2838, 9 2015. [Online]. Available: https://doi.org/10.1016/j.patcog.2015.04.006

[98] O. Kwon, J. Jeong, H. B. Kim, I. H. Kwon, S. Y. Park, J. E. Kim, and Y. Choi, “Electrocardiogram Sampling Frequency Range Acceptable for Heart Rate Variability Analysis,” *Healthcare Informatics Research*, vol. 24, no. 3, pp. 198–206, 2018. [Online]. Available: https://doi.org/10.4258/hir.2018.24.3.198

[99] P. Hamilton, “Open Source ECG Analysis Software Documentation,” E.P.Limited, 2002.

[100] R. Silva, “An Artificial Intelligence Approach for Atrial Fibrillation Detection in Single-lead Invisible ECG,” M.S. thesis, Nov 2021, Instituto Superior Técnico, Universidade de Lisboa.

[101] R. Baptista, H. Silva, and M. Rocha, “Design and development of a digital stethoscope encapsulation for simultaneous acquisition of phonocardiography and electrocardiography signals: the SmartHeart case study,” *Journal of Medical Engineering and Technology*, vol. 44, no. 4, pp. 153–161, 2020. [Online]. Available: https://doi.org/10.1080/03091902.2020.1757770

[102] Glia. Stethoscope. Accessed: 14 Sep, 2022. [Online]. Available: https://glia.org/pages/stethoscope

[103] A. Pavlosky, J. Glauche, S. Chambers, M. Al-Alawi, K. Yanev, and T. Loubani, “Validation of an effective, low cost, Free/open access 3D-printed stethoscope,” *PLOS ONE*, vol. 13, no. 3, pp. 1–10, 03 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0193087

[104] E.-R. Symeonidou, A. D. Nordin, W. D. Hairston, and D. P. Ferris, “Effects of Cable Sway, Electrode Surface Area, and Electrode Mass on Electroencephalography Signal Quality during Motion,” *Sensors*, vol. 18, no. 4, 2018. [Online]. Available: https://doi.org/10.3390/s18041073

[105] A. dos Santos Silva, H. Almeida, H. P. da Silva, and A. Oliveira, “Design and evaluation of a novel approach to invisible electrocardiography (ECG) in sanitary facilities using polymeric electrodes,” *Scientific Reports*, vol. 11, 12 2021. [Online]. Available: https://doi.org/10.1038/s41598-021-85697-2

[106] H. P. da Silva, J. Guerreiro, A. Lourenço, A. Fred, and R. Martins, “BITalino: A Novel Hardware Framework for Physiological Computing,” in *PhyCS 2014 - Proceedings of the International Conference on Physiological Computing Systems*, Jan 2014.

[107] D. Batista, H. P. da Silva, A. Fred, C. Moreira, M. Reis, and H. A. Ferreira, “Benchmarking of the BITalino biomedical toolkit against an established gold standard,” *Healthcare Technology Letters*, vol. 6, no. 2, pp. 32–36, 2019. [Online]. Available: https://doi.org/10.1049/htl.2018.5037

[108] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[109] C. Carreiras, A. P. Alves, A. Lourenço, F. Canento, H. Silva, A. Fred *et al.* (2015–) BioSPPy: Biosignal processing in Python. Accessed: 22 Oct, 2022. [Online]. Available: https://github.com/scientisst/BioSPPy/

[110] Texas Instruments. LM4861 1.1W Audio Power Amplifier with Shutdown Mode Datasheet. Accessed: 26 Oct, 2022. [Online]. Available: https://www.ti.com/lit/ds/symlink/lm4861.pdf

Chapter A

Appendix - PCG sensor

The schematic of the ScientISST PCG sensor is represented in Figure A.1. The LM4861 is an audio power amplifier capable of maintaining a constant offset equal to half of the supply voltage ($V_{cc} = 3.3\ V$).

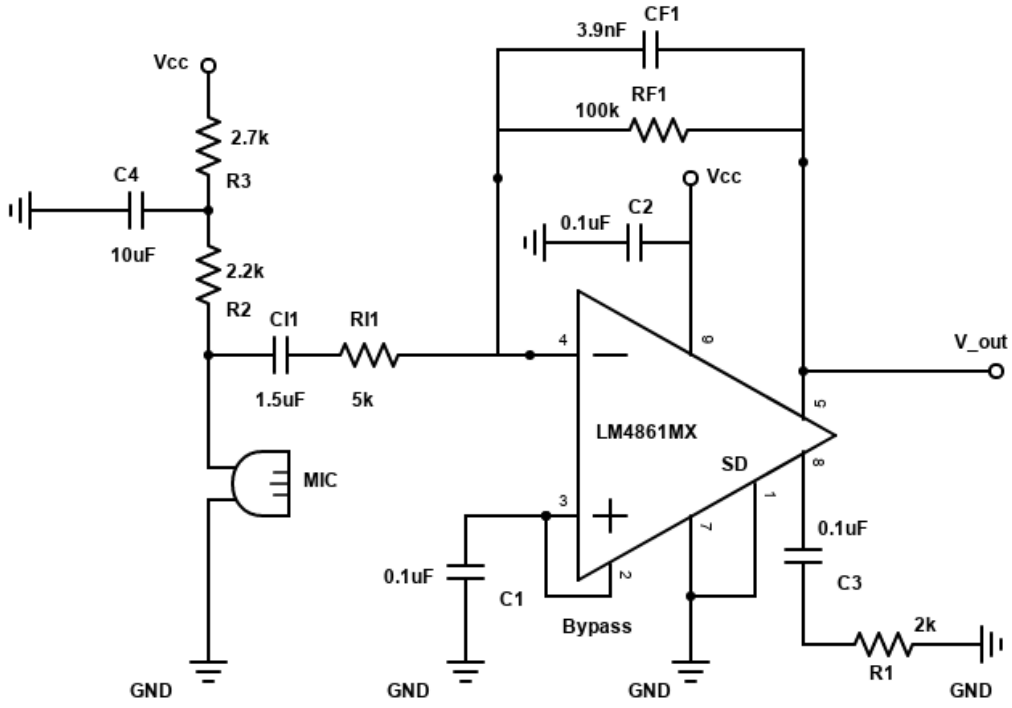


Figure A.1: ScientISST PCG sensor schematic.

The electret microphone has an impedance of 2.2 kΩ (Z_{mic}), 60 dB SNR and -40 dB sensitivity. The microphone bias circuit guarantees that the microphone is working under the conditions specified by the manufacturers (maximum current consumption of 0.5 mA, and operating voltage under 10 V), as described below:

$$V_{mic} = \frac{Z_{mic}}{Z_{mic} + R_2 + R_3} V_{cc} = 1.02\ V \tag{A.1}$$

$$I_{mic} = \frac{1}{Z_{mic} + R_2 + R_3} V_{cc} = 464.79\ \mu A \tag{A.2}$$

Additionally, in order to deal with voltage fluctuations from the power source reaching the microphone, a first order low-pass RC filter was implemented with a cut-off frequency approximately given by:

$$f_{H_{power\ source}} = \frac{1}{2\pi R_3 C_4} = 5.89\ Hz \tag{A.3}$$

The LM4861 has two operational amplifiers internally: the first amplifier's gain is externally configurable by selecting the ratio between R_{F1} and R_{I1}, while the second amplifier is fixed in a unity-gain with an inverting configuration [110]. The circuit provides an amplification of 40 at constant voltage:

$$G = 2\frac{R_{F1}}{R_{I1}} = 40 \tag{A.4}$$

and performs a bandpass filtering of the signal with a high-pass (f_L) cut-off frequency of approximately 21 Hz, and a low-pass (f_H) cut-off frequency of approximately 408 Hz:

$$f_L = \frac{1}{2\pi C_{I1} R_{I1}} = 21.22\ Hz \tag{A.5}$$

$$f_H = \frac{1}{2\pi C_{F1} R_{F1}} = 408.09\ Hz \tag{A.6}$$