Bachelor thesis

Bio-inspired Optimization Applied to Synthetic ECG Models for Generating Cardiac Arrhythmias

by **Rafael Monteiro Laranjeira**

advised by
Prof. Ph.D. Thiago Damasceno Cordeiro
Prof. Ph.D. Erick de Andrade Barboza

FEDERAL UNIVERSITY OF ALAGOAS
Computing Institute

BIO-INSPIRED OPTIMIZATION APPLIED TO SYNTHETIC ECG MODELS FOR GENERATING CARDIAC ARRHYTHMIAS

Bachelor Thesis submitted to the Computing Institute from Federal University of Alagoas as a partial requirement to obtain the bachelor degree in Computer Engineering.

Rafael Monteiro Laranjeira

Advisor: Prof. Ph.D. Thiago Damasceno Cordeiro
Advisor: Prof. Ph.D. Erick de Andrade Barboza

Examining Board:

Álvaro Alvares de Carvalho César Sobrinho   Ph.D., UFAPE
Eduardo Moraes De Miranda Vasconcellos   M.Sc., UFAL

Maceió, Alagoas
February 03, 2023

UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL

Instituto de Computação – IC

Campus A. C. Simões - Av. Lourival de Melo Mota, BL 12
Tabuleiro do Martins, Maceió/AL - CEP: 57.072-970
Telefone: (082) 3214-1401

Trabalho de Conclusão de Curso – TCC

Formulário de Avaliação

Curso: Engenharia de Computação

Nome do Aluno

| R | A | F | A | E | L | | M | O | N | T | E | I | R | O | | | | |
| L | A | R | A | N | J | E | I | R | A | | | | | | | | | |

Nº de Matrícula

| 1 | 7 | 1 | 1 | 1 | 4 | 5 | 2 | | | | | - | |

Título do TCC (Tema)

Bio-inspired Optimization Applied to Synthetic ECG Models for Generating Cardiac Arrhythmias

Banca Examinadora

Thiago Damasceno Cordeiro
Nome do Orientador

Documento assinado digitalmente
gov.br THIAGO DAMASCENO CORDEIRO
Data: 07/02/2023 13:48:43-0300
Verifique em https://verificador.iti.br

Assinatura

Álvaro Alvares de Carvalho César Sobrinho
Nome do Professor

Documento assinado digitalmente
gov.br ALVARO ALVARES DE CARVALHO CESAR SO
Data: 07/02/2023 11:22:40-0300
Verifique em https://verificador.iti.br

Assinatura

Eduardo Moraes de Miranda Vasconcellos
Nome do Professor

Documento assinado digitalmente
gov.br EDUARDO MORAES DE MIRANDA VASCONCI
Data: 07/02/2023 11:12:45-0300
Verifique em https://verificador.iti.br

Assinatura

Data da Defesa	Nota Obtida
03/ 02/ 2023	10 (DEZ)

Observações
_____
_____
_____
_____

Coordenador do Curso
De Acordo

Documento assinado digitalmente
gov.br LEANDRO DIAS DA SILVA
Data: 07/02/2023 16:00:35-0300
Verifique em https://verificador.iti.br

Assinatura

Acknowledgments

I would like to express my deepest gratitude to Prof. Thiago Cordeiro for all the support, friendship, patience, instruction, motivation, inspiration and invaluable advice during my academic experience. Special thanks to Prof. Erick Barboza, who generously provided help and guidance. I am also very grateful to Prof. Álvaro Sobrinho for his time and presence in the chair.

I could not have undertaken this journey without my parents and family's unwavering support, who believed in me even when I did not.

I would like to recognize my classmates, whom I had the pleasure of studying with, who endured several challenging moments by my side and helped me during this journey. I would like to express my most profound appreciation to Gustavo Miranda, who helped me discover computer engineering and for his amazing friendship and support. I am also extremely grateful to Gustavo Cabral for being an incredibly encouraging friend and also sharing the academic journey at UFAL. Special thanks to Eduardo Miranda for being in the chair, aiding me in this quest, being the first contact I had with this degree, and with whom I had the pleasure of working.

Words cannot express my appreciation to Giovana for being a wonderful partner and present throughout all my accomplishments and hardships, never ceasing to embolden me.

I wish to extend my special thanks to Álvaro, Roger, Sofia, Vangasse, Valério, João Arthur, Augusto, Roberto, Kevin, Hugo Tallys, Hugo Rodrigues, Andressa, Lucas Raggi, Laryssa and Maria Júlia for the continuous support, the fun times and for teaching me.

Don't ever, for any reason, do anything, to anyone, for any reason, ever, no matter what, no matter where, or who, or who you are with, or where you are going, or where you've been, ever, for any reason whatsoever.

Michael Scott

Resumo

O eletrocardiograma (ECG) é um procedimento essencial para a detecção de diversos problemas cardíacos. Com o avanço do desenvolvimento dos algoritmos de aprendizado profundo, torna-se cada vez mais interessante a construção de classificadores de doenças cardiovasculares a partir do sinal de ECG e muitos desses algoritmos apresentam um alto desempenho. Entretanto, o principal desafio ainda persiste: as bases de dados contendo sinais de ECG são caras e muitas vezes com pouca variedade e anotações de especialistas da área. Esse trabalho introduz uma nova metodologia para gerar sinais de ECG sintéticos utilizando um modelo matemático e um algoritmo bio-inspirado para estimar os parâmetros deste mesmo modelo. O principal objetivo é expandir o modelo matemático original para que ele seja capaz de reproduzir as mais diversas arritmias cardíacas utilizando sinais reais como referência. Os parâmetros do modelo são definidos para cada onda e são obtidos por um processo de otimização que consiste em minimizar a diferença entre sinal sintético gerado e o sinal real. Os resultados mostram que a metodologia proposta é capaz de estimar os parâmetros do modelo sintético utilizando funções gaussianas para cada onda e consegue se adaptar para diversas outras doenças cardiovasculares.

Palavras-chave: *Eletrocardiograma*; *Evolução diferencial*; *Otimização*; *Modelo de ECG Sintético*; *Arritmias Cardíacas*; *PhysioNet*; *Algoritmos Evolucionários*; *Algoritmos Bio-inspirados*; *Doenças cardiovasculares*.

Abstract

The Electrocardiogram (ECG) is an essential and straightforward procedure used to detect cardiac abnormalities. With the development of deep learning algorithms, the focus of many works is on the design of automatic ECG disorders detection algorithms, and many of them have achieved very high accuracy. The main challenge is still the same: clinical ECG datasets are expensive, and there are no sufficient expert annotations. This work introduces a new methodology for generating synthetic ECG signals using a mathematical model and a bio-inspired algorithm to estimate the model's parameters. The objective is to expand the original model to reproduce different cardiac arrhythmias using ECG records as a reference. The model's parameters are different for each selected ECG wave and are obtained by minimizing the difference between the ECG recordings and the synthetic ones. The results show that the proposed methodology is able to estimate the synthetic ECG model parameters using Gaussian functions for each ECG wave and is highly adaptable to different cardiac diseases.

Keywords: *Electrocardiogram*; *Differential Evolution*; *Optimization*; *Synthetic ECG Model*; *Cardiac Arrhythmias*; *PhysioNet*; *Evolutionary Algorithms*; *Bio-inspired algorithms*; *Cardiovascular diseases*.

List of Figures

List of Tables

List of Acronyms

AF Atrial Fibrillation

A-V Atrioventricular Node

CDESSA Chaotic Differential Evolution Salp Swarm Algorithm

CPU Central Processing Unit

CVD Cardiovascular Diseases

DE Differential Evolution

EA Evolutionary Algorithm

ECG Electrocardiogram

GPU Graphical Processing Unit

IPCA Independent Principal Component Analysis

LBBB Left Bundle Branch Block

LHS Latin Hypercube Sampling

MAE Mean Absolute Error

MSA Moth Search Algorithm

MSE Mean Squared Error

MSDE Moth Search Differential Evolution

PAC Premature Atrial Contraction

PCA Principal Component Analysis

PPG Photoplethysmogram

PSO Particle Swarm Optimization

RBBB Right Bundle Branch Block

RMSE Root Mean Squared Error

SAN Sinoatrial Node

SSA Salp Swarm Algorithm

VM Virtual Machine

Contents

Chapter 1

Introduction

Cardiovascular Diseases (CVDs) are the world-leading cause of death. In 2019, there were more than 523.2 million cases, with 18.6 million resulting in death. According to the American Heart Association, the prevalence of CVDs in adults ≥ 20 years old is 49.2% overall (126.9 million in 2018) and increases with age in both males and females (Alonso et al., 2021).

The early identification of these conditions is a crucial step towards bettering the prognosis. This can usually be achieved by doing an Electrocardiogram (ECG) exam. Although the ECG is a low-cost and non-invasive test that can diagnose various CVDs, it has some problems and limitations. For instance, one possible problem is that the resulting signal may have artifacts generated during the exam by sources such as muscular movement or power-line noise that may hinder the CVD diagnosis. The access to quality ECG databases is not a guarantee, which is a possible limitation.

Due to this fact, many researchers may fall back on using mathematical models to produce synthetic ECG signals that can represent a wide variety of ECG signals for several research interests and several modeling techniques can be found in the literature. McSharry (McSharry et al., 2003) developed a dynamical model for generating synthetic ECG signals that can be made noise-free and can be adjusted using a set of parameters to achieve different signal formats. Adopting the mathematical foundation provided in (McSharry et al., 2003), TiamKapen (Kapen et al., 2019) was able to obtain synthetic ECGs from eight cardiac arrhythmias with results being very close to the real signals. Still, the results were based only on a visual analysis, and no methodology was used to find the optimal model parameters.

Many papers have been using optimization techniques to find the set of best-fit parameters to a specific problem, as the work developed by Awal (Awal et al., 2021). The work provides an ECG model that can accurately represent the signal with the possibility of being adjusted to multiple cardiac dysrhythmias based on the sum of two Gaussians. This process can show significant results, but fitting more than one Gaussian function in a deterministic way has an accuracy and localization problem. To address the accuracy issue and find the proposed model's optimal parameters, (Awal et al., 2021) introduced two hybrid optimization methods: ApproxiGlo and ApproxiMul. The presented model and optimization methods are simple and capable

of reproducing ECG signals of varying duration and conditions, but they are not free of limitations. Some discussed limitations are manual extraction of the P, Q, R, S, and T waves, a limited number of ECG beats, and a lack of cardiac dysrhythmias classifiers using the proposed model.

The Differential Evolution (DE) algorithm has been used to solve several parameter estimation and optimization problems. One of the uses of DE is presented by (Zhang et al., 2022). A new framework is proposed to improve the Salp Swarm Algorithm (SSA) performance when handling complex optimization problems where the algorithm shows its limitations. The salp swarm algorithm has low search accuracy, a slow convergence speed, and high susceptibility to getting stuck on a local optimum. A chaotic initialization SSA with differential evolution (CDESSA) is suggested to enhance the convergence and accuracy of the original SSA algorithm. Multiple tests were executed to evaluate the proposed framework's performance, including several real engineering optimization problems. When compared to state-of-the-art algorithms, CDESSA displayed superior results. Elaziz (Abd Elaziz et al., 2019) proposes a new method for solving the cloud task scheduling problem, which intends to minimize the makespan required to schedule several tasks on different Virtual Machines (VMs). This method is established on the advance of the Moth Search Algorithm (MSA) using the DE algorithm and is presented as the MSDE algorithm. The DE is used as a local search method to enhance the MSA's exploitation ability. Compared to other heuristic and meta-heuristic algorithms, when applied to the task scheduling problem, the MSDE outperforms the other algorithms. The main drawback of the alternative method is the time complexity, which can be improved by supplying a good initial population.

Important applications of DE can also be found in the ECG area, as in the work of (Dolinský et al., 2018). Dolinský introduces a new mathematical model for generating synthetic ECG signals based on the geometrical properties of an actual ECG signal. The model can generate several custom-made ECG signals for diseases like cardiac arrhythmias using a trigonometric function or a Gaussian monopulse. The model can also take into consideration the noise generated by the equipment and the respiratory frequency. To obtain the optimal model parameters, the DE algorithm was utilized. The results show that the model is suitable for modeling different diseases.

This work proposes a new methodology for obtaining McSharry's model parameters using the DE algorithm and real ECG signals. This study will describe the DE algorithm and its importance in the proposed method. To conclude, the results will be presented, discussing their implications, addressing applications, and discussing possible future works.

1.1 Objectives

The main objective of this study is to introduce a new methodology to generate synthetic ECGs that are similar to real ones with the ability to replicate the most varied conditions by using the DE algorithm to find the optimal parameters.

1.2   Specific Objectives

For the main goal, the following specific objectives will be contemplated.

1. Discuss the importance of ECGs on the diagnosis of multiple cardiac disorders.

2. Extend the base McSharry synthetic ECG model to generate unhealthy ECG signals.

3. Use the differential evolution algorithm to find the optimal parameters of the extended model.

4. Simulate Premature Ventricular Contraction, Right Bundle Branch Block, Left Bundle Branch Block and Atrial Fibrillation using the parameters found with DE.

1.3   Applications

Several applications can be made using synthetic ECGs. For instance, noisy signals can be generated for further use in testing different signal-processing techniques. Different signals can be generated and used to enrich ECG datasets, improving lead diversity and the number of unique diseases. These enhanced datasets can be used to train deep neural networks for ECG classification or wave identification tasks. An educational tool can also be developed to help medical students understand the ECG and also train more experienced doctors in clinical ECG interpretation.

1.4   Structure

In the first chapter, the thesis will introduce the topic of study. In the second chapter, the background behind the methodology discussed in this thesis will be the main focus. That includes the definition of a few cardiac problems, synthetic ECG signals and the general idea of the differential evolution algorithm. For the third chapter, the complete methodology will be the topic of discussion. There, the steps taken to obtain the results will be talked through. The fourth chapter will show the results. Finally, the last chapter is where the main considerations will be displayed.

Chapter 2

Background

This chapter explores a background regarding the main concepts involving cardiac impulses, ECGs, synthetic ECG models, cardiac disorders and the differential evolution algorithm.

2.1 The Electrocardiogram Signal

The Electrocardiogram signal is obtained by placing electrodes on opposite sides of the heart in order to record the electrical potentials generated by the cardiac impulse. A small fraction of the electric current that propagates throughout the nearby tissues travels to the body's surface, which can be detected by the electrodes placed on the skin beforehand. It is a non-invasive painless test that can identify heart problems and keep track of the heart's health state. Some of the issues that can be identified are evidence and detection of myocardial infarction (Sridhar et al., 2021) or morphological ECG abnormalities (Ribeiro et al., 2020).

2.1.1 Generation of the Electric Impulse

The electric impulse is generated by the process of self-excitation of the sinus nodal fibers. The extracellular fluid on the exterior of the nodal fiber has a high sodium ion concentration and a modest number of open sodium channels. These concentrations cause positive sodium ions to leak to the interior of the sinus nodal fibers, meaning that the stream of positive ions implies a steady positive increase in the resting membrane electric potential. After reaching the potential of about -40 millivolts, the sodium-calcium channels become active, inducing the action potential. Because the ends of the sinus nodal fibers are directly connected to the adjacent atrial muscle fibers, the action potential beginning at the sinus node will advance through the atria, ultimately reaching the atrioventricular node. After a delay produced by the A-V node and the adjacent conductive fibers, the transmission carries on to the ventricles. Some specialized fibers called Purkinje fibers transmit action potentials at a much faster speed than the ventricular muscle and the A-V nodal fibers. These special fibers transmit the potential passing through the A-V bundle and into the ventricles. After reaching the terminations of the Purkinje fibers, the

transmission is maintained by the ventricular muscle mass at a slower velocity. Finally, the impulse goes through the ventricular muscle, eventually reaching the epicardial surfaces (Hall and Hall, 2020). Some structure involved on the generation of the eletric impulse can be seen in Figure 2.1.
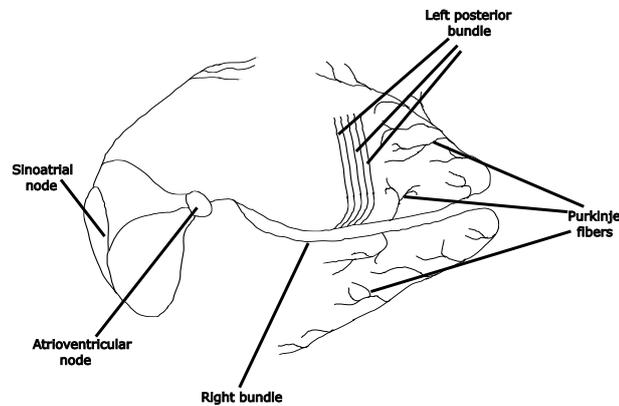
Figure 2.1: Illustrative heart showing main structures for the rhythmical excitation of the heart including sinus node, Purkinje system, A-V node, atrial internodal pathways, and ventricular bundle branches.

2.1.2 ECG Waves, Segments and Intervals

The normal electrocardiogram usually consists of five separate waves: a P wave, a Q wave, a R wave, a S wave and a T wave. The Q, R and S waves regularly compose the QRS complex. The P wave results from potentials generated by the depolarized atria preceding the atrial contraction. Like the P wave, the QRS complex is caused by depolarization. However, in this case, it is the depolarization of the ventricles before the ventricular contraction. Lastly, the T wave results from the potentials produced by the recovery of the ventricles after the depolarization. In conclusion, the ECG is comprised of depolarization waves (P, Q, R, and S waves) and polarization waves (T wave) (Hall and Hall, 2020).

Sometimes it may be interesting to refer to specific moments of the electrocardiogram using intervals. These moments can be seen in Figure 2.2 There are a few key intervals that are commonly used:

1. P-Q interval: It is the time starting from the beginning of the electrical excitation of the atria until the beginning of the electrical excitation of the ventricles. Usually lasts 0.16 second.

2. P-R interval: It encompasses the same time span of the P-Q interval, but it is utilized when the Q wave is missing.

3. Q-T interval: It is the time starting from the beginning of the Q wave until the end of the T wave. Sometimes the Q wave may be absent, so in these situations the time starts at the beginning of the R wave. It usually lasts 0.35 second.

4. R-R interval: It is the time between two consecutive peaks from the R wave.

There are also specific segments that can be used to refer to specific moments, they represent the length between two specific points:

1. P-R segment: This segment starts at the end of the P wave and ends at the start of the Q wave.

2. S-T segment: This segment starts at the end of the S wave and ends at the start of the T wave.

3. T-P segment: This segment starts at the end of the T wave and ends at the start of the P wave from the next cycle.
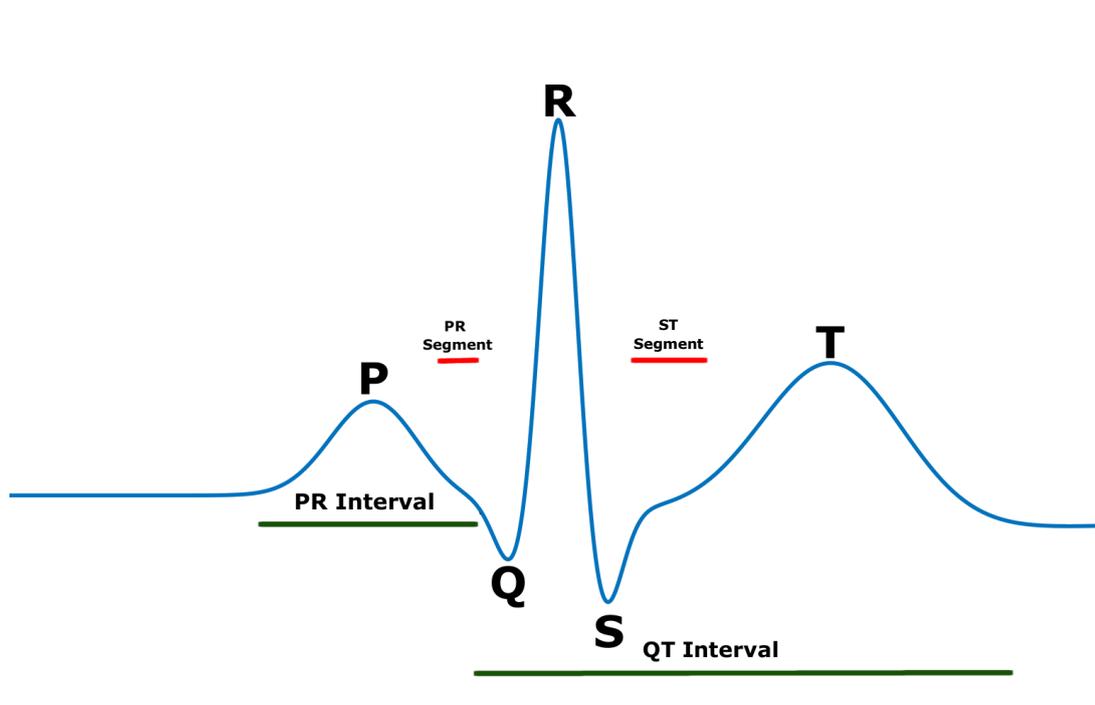
Figure 2.2: ECG excerpt representation showing waves, intervals and segments.

2.1.3   Lead Configuration and Systems

Leads are the tracing of the potential difference between two points. Some lead systems can be utilized to achieve the task of recording the electrocardiogram. The 12-lead system is the most common system and the electrodes placement can be seen in Figure 2.4. Three of these configurations are (Hall and Hall, 2020):

1. Standard bipolar leads. This system is recorded from two electrodes located on opposite sides of the heart, usually on the limbs. This configuration has three leads: Lead I, Lead II, and Lead III. The placement of the leads form a triangle called Einthoven's triangle.

Moreover, Einthoven's Law affirms that if any two of the three leads are known, the third can be figured out by summing the known leads, taking into account their signs.

- Lead I: For this lead, the negative terminal of the electrocardiograph is connected to the right arm and the positive terminal is connected to the left arm.

- Lead II: In this lead, the negative terminal of the electrocardiograph is connected to the right arm and the positive terminal is connected to the left leg.

- Lead III: Here the negative terminal of the electrocardiograph is connected to the left arm and the positive terminal is connected to the left leg.

2. Precordial leads. Electrocardiograms can also be recorded with one electrode placed directly over the heart at six possible points. The electrode placed at any of the six possible points is connected to the positive terminal of the electrocardiograph. The negative terminal will be connected simultaneously through equivalent electrical resistances to the left arm, left leg, and right arm.

3. Augmented unipolar limb leads. In this system of leads, two limbs are connected to the negative terminal of the electrocardiograph, while the third limb is connected to the positive terminal. The lead's name will depend on which limb is connected to the positive terminal.

- aVR lead: When the positive terminal is connected to the right arm.

- aVL lead: when the positive terminal is connected to the left arm.

- aVF lead: when the positive terminal is connected to the left leg.

Every lead can be seen as a different perspective of the human heart, which means that some leads may be better than other in diagnosing diseases that have a specific effect on the human body. These perspectives can be grouped in 2 electrical planes: horizontal and vertical planes. The precordial leads are situated in the horizontal plane and the standard bipolar leads and augmented unipolar limb leads are situated in the vertical plane. The Figure 2.3 illustrates the electrical planes mentioned above.
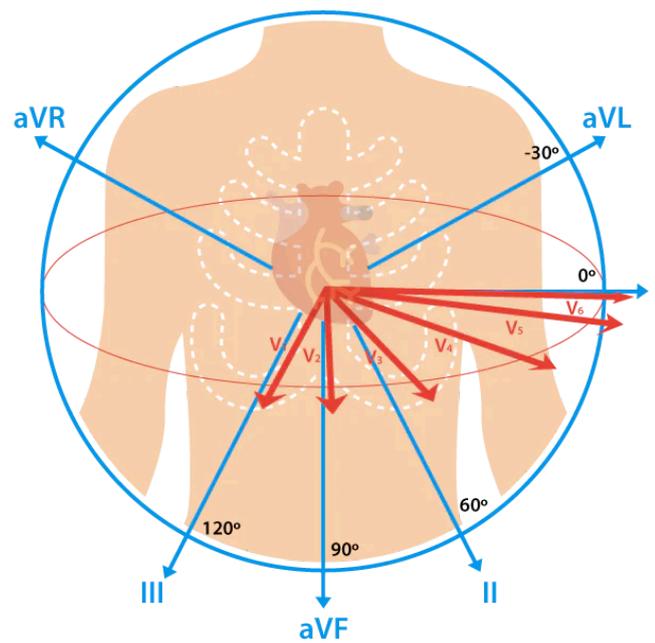
Figure 2.3: Illustration of the horizontal and vertical electrical planes. From Cables and Sensors.
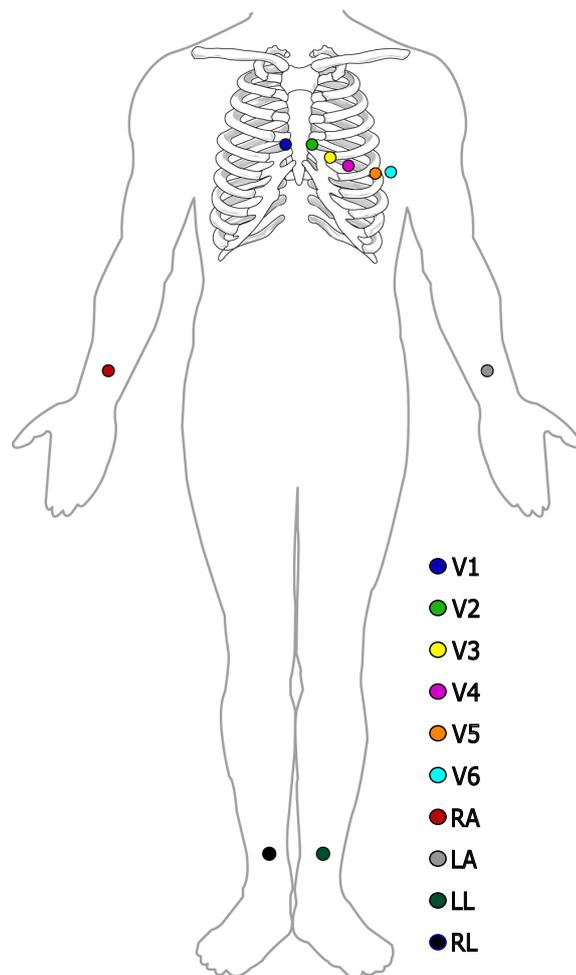
Figure 2.4: Figure displaying the 12-lead system. Adapted from (OpenStax College, 2013).

2.2   Cardiac Disorders

Cardiovascular diseases are conditions that can affect the structures and/or the function of the heart, the main affected structures being the heart and the blood vessels. There are several types of heart disease and different causes. Some diseases are congenital, which means that the person was born with it and others may be caused by genetics, other diseases, lifestyle choices, and other influences. As discussed earlier, heart disease is the leading cause of death worldwide. However, the risk of acquiring some cardiovascular diseases can be lowered with certain choices or at least have its symptoms mitigated with adequate treatment. Below are the principal cardiovascular diseases and a brief description of each one[1] [2]:

- Arrhythmias: Condition where the heart gets out of rhythm.

- Aorta Disease and Marfan Syndrome: Conditions that cause the widening or tear of the aorta.

- Cardiomyopathies: Diseases related to the heart muscle.

- Congenital Heart Disease: A problem that occurs in one or multiple parts of the heart before birth.

- Coronary Artery Disease: Also known as atherosclerosis, it happens when plaque is accumulated and hardens the arteries.

- Deep Vein Thrombosis and Pulmonary Embolism: Condition where blood clots form in the deep veins and may travel through the bloodstream and reach the lungs, causing the interruption of the blood flow.

- Heart Failure: Condition where the heart fails to pump as strongly as it should.

- Heart Valve Disease: Diseases associated with the valves located at the exit of the four heart chambers.

- Rheumatic Heart Disease: Condition that happens when an inflammatory disease damages the heart valves.

- Stroke: Occurs when the blood flow to the brain is blocked or slowed.

In this work, the emphasis will be on heart rhythm disorders. They are one of the most common diseases that have defining characteristics that can be identified on an ECG. There are a few types of arrhythmias and some of them will be discussed below.

---

[1]https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed on September 22th, 2021.

[2]https://www.webmd.com/heart-disease/guide/diseases-cardiovascular. Accessed on September 22th,2021

2.2.1 Heart Blocks

Heart blocks are disorders that affect the heart rhythm due to an obstruction in the heart's electrical conduction system. Sometimes there are no symptoms, and the severity of the problem is related to the distance from the Sinoatrial Node (SAN). One category of heart blocks is the bundle branch blocks. Right bundle branch block (RBBB) and left bundle branch block (LBBB), as can be seen in Figures 2.5 and 2.6, occur when the physiologic electrical conduction system of the heart, specifically in the His-Purkinje system, is altered or interrupted, resulting in a widened QRS and electrocardiographic vector changes (Ten Tusscher and Panfilov, 2008). The widened QRS can be identified in Figures 2.5 and 2.6. An enlarged S wave can also be spotted in Figure 2.6.
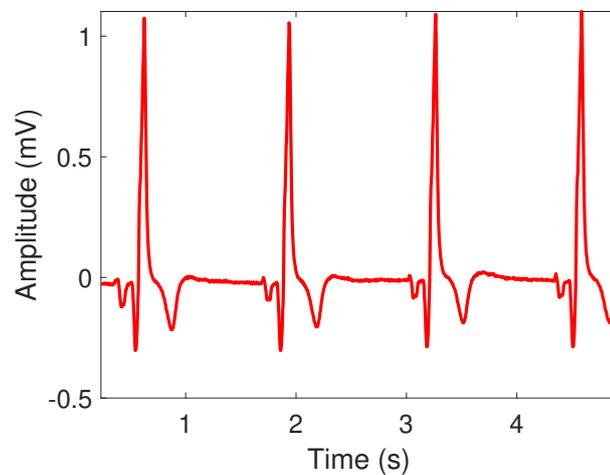
Figure 2.5: Lead V1 displaying multiple cycles with RBBB extracted from patient I16 from St Petersburg INCART 12-lead Arrhythmia Database.
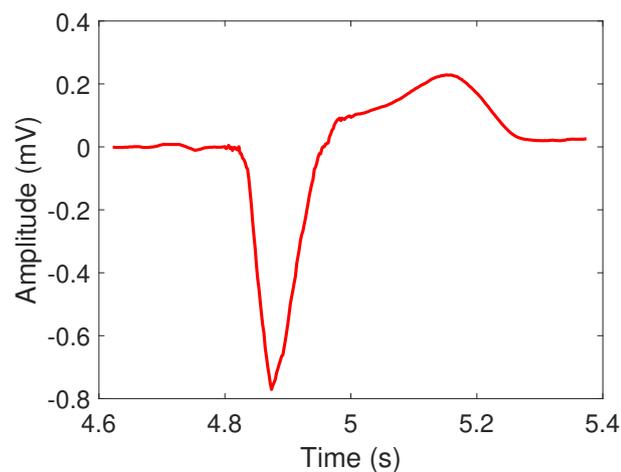
Figure 2.6: Lead V1 displaying a LBBB extracted from patient 24 from Lobachevsky University Electrocardiography Database.

2.2.2 Premature Contractions

Premature contractions occur when the heart's upper or lower chambers generate an extra heartbeat. If the premature beat starts in the atria, triggered by the atrial myocardium, it's called premature atrial contraction (PAC). A PAC can be seen in Figure 2.7. A premature ventricular contraction (PVC) is a disorder that happens when the heartbeat initiates by the Purkinje fibers rather than the SA node. Since a PVC occurs before a regular heartbeat, there is a pause before the next regular heartbeat (Gianni et al., 2018). A PVC can be seen in Figures 2.8 and 2.9. Note that a repolarisation abnormality can be identified in Figure 2.9, as the ST segment and the T wave have opposite directions compared to the QRS complex.
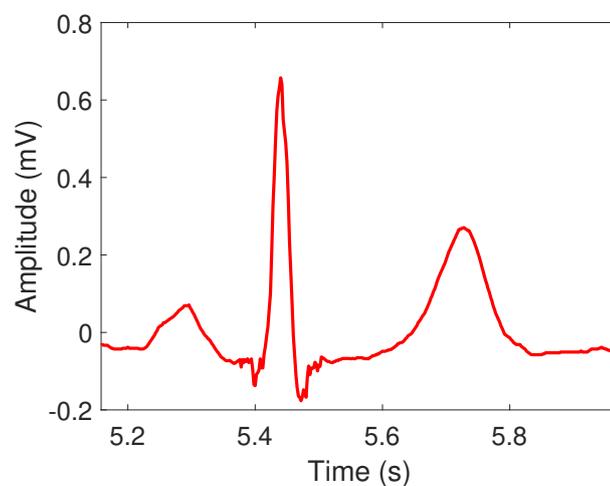
Figure 2.7: Lead II displaying a PAC extracted from patient 66 from Lobachevsky University Electrocardiography Database.
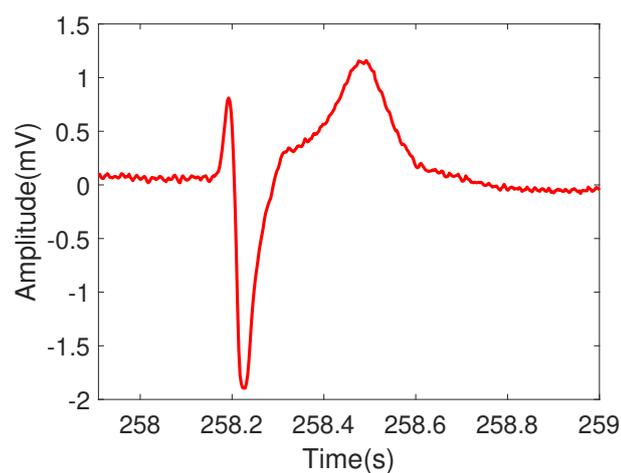
Figure 2.8: Lead MLII displaying a PVC extracted from patient 114 from MIT-BIH Arrhythmia database.
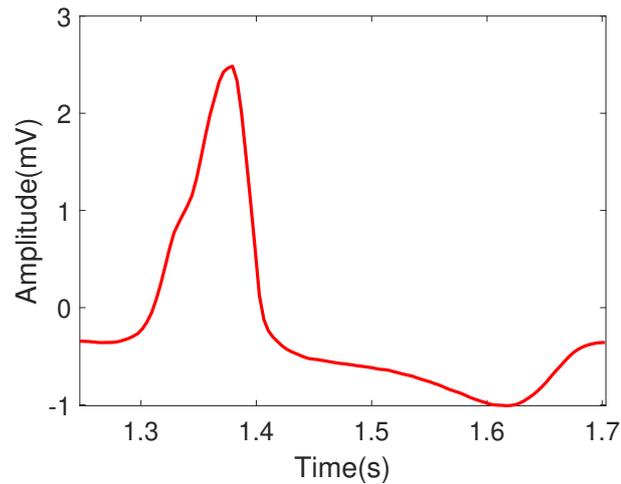
Figure 2.9: Lead II displaying a PVC extracted from patient I13 from St Petersburg INCART 12-lead Arrhythmia Database.

2.2.3 Fibrillations

When there are chaotic and irregular electrical signals in the heart muscle, an atrial or ventricular fibrillation occurs. If the irregular heartbeat happens on the heart's upper chambers, it will be called atrial fibrillation. If it happens on the heart's lower chambers, it will be called ventricular fibrillation. This type of heart rhythm irregularity will cause different symptoms depending where the irregular heartbeat occurs. An atrial fibrillation (AF) can be seen in Figure 2.10. Note the absence of the P wave and the presence of fibrillatory waves (f-waves).
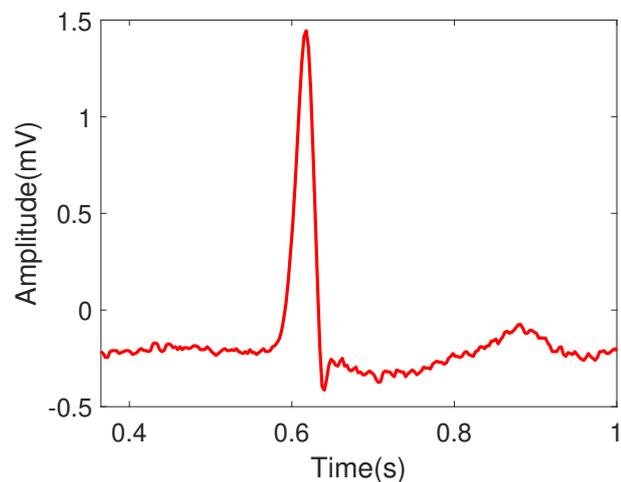
Figure 2.10: Lead MLII showing atrial fibrillation extracted from patient 221 from MIT-BIH Arrhythmia Database.

2.3 ECG Databases

Many biomedical research databases are available for free on the Internet, most of them containing digital recordings of multiple physiologic signals, like ECGs. The databases can be used for a variety of studies and analyses. They include many cardiac disorders, contain several different leads, and have different recording conditions and other valuable data. One popular biomedical database source that includes ECGs is the PhysioNet (Goldberger et al., 2000) community resource. Most databases available on PhysioNet have open access with few or no restrictions associated, making them a very favored choice. Other databases with limited or exclusive access can also be found on several papers.

Various ECG database applications can originate from artificial intelligence, machine learning/deep learning, diagnosis, medical equipment certification, education, and other areas. For machine learning/deep learning, many classifiers are using a myriad of models to classify certain diseases, normal and healthy ECGs, and other vital classifications.

The work introduced by (Ribeiro et al., 2020) uses a deep neural network for the automatic classification of 12 lead ECGs. The model was trained using over 2 million labeled exams that the Telehealth Network of Minas Gerais analyzed. The dataset was collected from 811 counties in the state of Minas Gerais, Brazil. The network showed excellent performance using the dataset.One of the limitations of the classifier is the lack of statistical significance to assert that the model performs better than medical students. This limitation is possibly caused by the presence of class imbalance, where small mistakes can noticeably affect the scores. When dealing with small datasets that contain even less data diversity, the drop in performance can be considerably more significant.

(Weimann and Conrad, 2021) used transfer learning for the classification of atrial fibrillation using a database provided by the PhysioNet Computing in Cardiology Challenge 2017 (Clifford et al., 2017). The Convolutional Neural Network (CNN) utilized was pre-trained using the Icentia11k dataset (Tan et al., 2019). This dataset was the biggest public ECG dataset at the time, containing 11 thousand patients and 2 billion labeled beats. The network then went through some fine-tuning using the atrial fibrillation data. This process resulted in an improvement of up to 6.57% in comparison to CNNs that were not pre-trained with the dataset, highlighting the importance of a good dataset.

Although the Physionet databank allows for numerous analyses and provides a good number of data, it has limitations. First, many databases contain only a small number of leads, sometimes only one lead. An analysis involving multiple leads may be required for some diseases to obtain a more accurate diagnosis. There are also problems related to the diversity of the data. Some diseases appear much more frequently than others and are recorded under many different conditions. When dealing with classification tasks, for instance, having much fewer samples of certain diseases can lead to class imbalance and impact the model's performance. A combination of different cardiac disorders within the same ECG can be uncommon. Therefore these

specific scenarios will also lead to imbalanced data.

2.4 Synthetic ECG Modeling

A synthetic ECG is generated using mathematical equations that simulate the functioning of a real heart. The model produces a trajectory in a 3D state-space, where each rotation represents a RR-interval. The dynamic equations of motion presented in (McSharry et al., 2003) are defined as:

$$\dot{x} = \alpha x - \omega y \tag{2.1}$$

$$\dot{y} = \alpha y - \omega x \tag{2.2}$$

$$\dot{z} = - \sum_{i \in \{P,Q,R,S,T\}} a_i \Delta\theta \exp\left(-\frac{\Delta\theta_i}{2b_i^2}\right) - (z - z_0) \tag{2.3}$$

where $\alpha = 1 - \sqrt{x^2 + y^2}$, $\theta = \text{atan2}(y, x)$, $\Delta\theta = (\theta - \theta_i)$, ω is the angular velocity of the trajectory as it moves around the limit cycle, and z_0 is the baseline wander. Essential points are described by events that occur in the z direction, like the P, Q, R, S, and T waves. The baseline wander is a signal artifact that can originate from multiple noise sources and may hinder the interpretation or clinical diagnosis. (McSharry et al., 2003) defines the baseline wander as:

$$z_0(t) = A sin(2\pi f_2 t) \tag{2.4}$$

where $A = 0.15 mV$ is the amplitude and $f_2 = 0.25$ is the respiratory frequency. All the equations were numerically integrated using a fourth-order Runge–Kutta method with a time step equivalent to $\frac{1}{f_s}$, where f_s is the sampling frequency. The ECG curve obtained by using the parameters provided by (McSharry et al., 2003) to generate a normal ECG that can be seen in Figure 2.11.

2.5 Evolutionary Algorithms

Evolutionary algorithms (EAs) are general bio-inspired optimization methods that use concepts like populations, replication, variation, and selection. Bartz-Beielstein defined that EAs consist of three principal components: search operators (recombination and mutation), an imposed control flow, and adequate mapping of variables to potential solution candidates (also known as genotype-phenotype mapping) (Bartz-Beielstein et al., 2014). Sometimes, EAs can be described as algorithms that systematically seek the best solution contained in the problem space. The strategies used to guide the search are known as metaheuristics. These algorithms
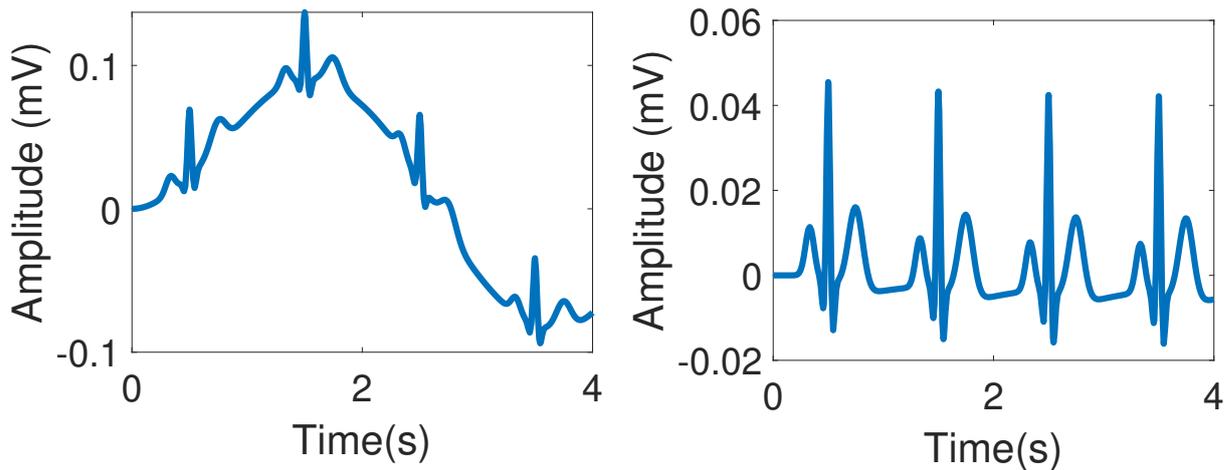
Figure 2.11: Synthetic ECG with baseline wander and without.

can solve complex problems that are difficult to solve using traditional methods like gradient-based solutions. Due to the increased availability of computation, evolutionary algorithms will probably be utilized in multiple applications in the future. Many aspects of these optimization methods draw inspiration from the Theory of Evolution (Darwin, 1859) or mimic natural evolution in some way. A definition that is generally accepted for EAs is:

> "Evolutionary algorithms: collective term for all variants of (probabilistic) optimization and approximation algorithms that are inspired by Darwinian evolution. Optimal states are approximated by successive improvements based on the variation-selection-paradigm. Thereby, the variation operators produce genetic diversity and the selection directs the evolutionary search", (Beyer et al., 2002).

2.6 Differential Evolution

The main focus of this work is the Differential Evolution (DE) algorithm that belongs to the class of evolutionary algorithms, more specifically bio-inspired algorithms. It's an evolutionary computation algorithm, proposed by Storn and Price (Storn and Price, 1997), that uses a heuristic method for the global optimization of functions that can have properties such as nonlinearity.

There are a few requirements for a minimization technique to be considered practical and the DE algorithm is designed to fulfill all of these requirements. They are listed as follow:

- Potential to deal with non-differentiable, nonlinear and different types of cost functions, such as multimodal functions;

- Ability to be parallelized in order to make use of multiple cores/threads to speed up computation;

- A small selection of variables that control the optimization process;

- Great convergence capabilities. The algorithm can provide consistent results in various independent trials.

The DE algorithm uses a parallel direct search method to generate variations of the parameter vectors, which can be seen as the main advantage of this method, i.e., it can significantly reduce the time to find an optimal solution. With the advancement in computing power in recent years, DE has become a compelling option for finding an optimal solution for these problems.

The methodology to select the best vector, i.e., the mutation strategy, can vary according to the objective. Several mutation strategy approaches have been discussed and a nomenclature is frequently used in the literature:

$$DE/x/y/z,$$

where x is the vector to be mutated, y is the number of difference vectors utilized and z is the crossover strategy selected. For instance, the mutation strategy shown in Fig 2.12 is the $DE/best/1/bin$. The vector to be mutated is the best vector of the population (the vector that produces the lowest fitness in this case), the number of the difference vectors is 1, and the crossover strategy is the binomial crossover, where the mutated components are randomly selected.

The function to measure the quality of the solution can be one of the common ones to evaluate estimators, like Mean Squared Error (MSE), Mean Absolute Error (MAE) or Root Mean Square Error (RMSE).

The DE algorithm only needs three parameters: population size (NP), an integer representing the number of individuals in a population; differential weight (F), a real value usually varying in the range $[0, 2]$, representing the scaling factor; and crossover rate (CR), a real value usually in the range $[0, 1]$, representing the probability of recombination. These are the values determined by (Storn and Price, 1997).

These three parameters must be carefully selected to ensure convergence to obtain the best solution. Furthermore, choosing which differential mutation will generate the new vectors is essential, as some modifications may produce better results and are more well-suited for specific problems.

The DE cycle has three main steps: mutation, crossover and selection. Let S_1 be the target vector selected from the population's i_{th} position. For the mutation step, each population member has a mutant vector created using Eq. 2.5:

$$v_i = X_{best} + F(x_{r2} - x_{r3}) \tag{2.5}$$

where v_i is the mutant vector, r_2 and r_3 are random integers $\in \{1, 2, 3, ..., \text{NP}\}$ and i is the i_{th} population index. For the crossover step, every individual is combined with its respective mutant vector to create a trial vector u_i, using the binomial crossover strategy and the resultant

vector is defined by Eq. 2.6:

$$u_i = \begin{cases} v_i[k] \text{ if } rand[0,1] < \text{CR} \\ \text{else } S_1[k] \end{cases} \tag{2.6}$$

where k is the index representing the dimension of each parameter of the population and $rand$ is a random number between 0 and 1.

Lastly, in the selection step, a comparison is made between the trial vector u_i and the target vector S_1. If the fitness of the trial vector is less than the target vector, the next generation member at the ith position becomes the trial vector. The flowchart in Figure 2.12 can be summarized below.

First, the population is generated using one population generation method. For this work, the population was generated using the Latin Hypercube Sampling method. Every individual in the population is then evaluated with the chosen fitness function. After the evaluation the mutant vectors are created using Eq. 2.5, and then the mutant vector is recombined with their parents to generate a trial vector using Eq. 2.6. Ultimately, the target and the trial vector are compared and the fitter ones are added to a new population. These steps are repeated until the termination condition is met.

Initial population

Choose target vector S_1

Random choice of two
population members
x_{r2}, x_{r3}

Build weighted difference vector
$F(x_{r2} - x_{r3})$

Add the populations
best vector x_{best}:
$v = x_{best} + F(x_{r2} - x_{r3})$

Do crossover with target
vector S_1 to get trial vector u

$u < S_1$

Yes   No

Replace S_1 with u Keep S_1

Create new population
generation

No   Enough
generations?   Yes   Stop
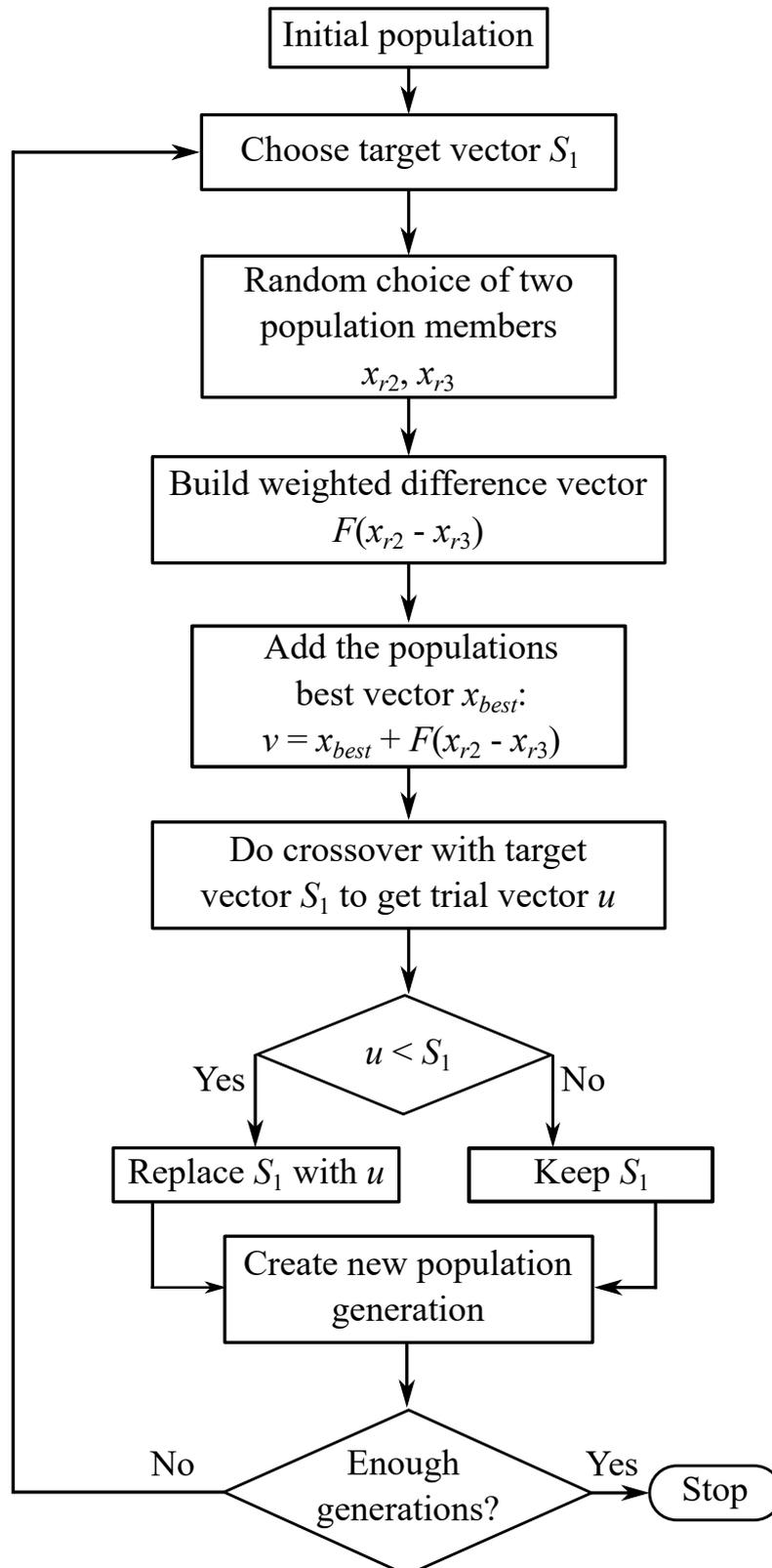
Figure 2.12: Flowchart displaying the process of the differential evolution algorithm.

2.7   Final Considerations

This chapter discussed about many theoretical aspects of ECGs and a few appointed diseases. Synthetic ECG models were introduced, and the importance of finding the optimal parameters to simulate different heart characteristics and conditions was considered. Many other diseases could also be noteworthy, but the scope of this thesis was limited to PVC, RBBB, LBBB and AF.

Evolutionary algorithms were briefly introduced, with the emphasis being on the Differential Evolution Algorithm. This algorithm was adopted mainly because of its simplicity and great convergence properties, especially when paired with parallel computing. The original DE algorithm was chosen among its multiple variants because it showed satisfactory performance for this problem. As discussed before, the parameters were also mostly taken from the original paper. Multiple studies have tried to find the best values for these parameters (Gämperle et al., 2002)(Ronkkonen et al., 2005)(Liu, 2002). However, they will not be taken into consideration for this study.

In the next chapter, the methodology will be established and analyzed at each step of the process.

Chapter 3

Methodology

The methodology for estimating the ECG model parameters can be divided into five steps: signal selection, signal extraction, signal filtering, wave selection and parameter estimation with DE.

Figure 3.1: Flowchart showing the methodology steps.

3.1   Signal Selection

In the first step, the ECG signals are selected from the databases. The chosen databases were the St Petersburg INCART 12-lead Arrhythmia Database, the MIT-BIH Arrhythmia Database (Moody and Mark, 2001) and the Lobachevsky University Electro-cardiography Database (Kalyakulina et al., 2020) from PhysioNet (Goldberger et al., 2000). The St Petersburg INCART 12-lead Arrhythmia Database contains 75 annotated recordings extracted from 32 Holter records. Each record has a duration of 30 minutes and is composed of the 12 typical leads, all of these leads being sampled at 257 Hz. The MIT-BIH Arrhythmia Database contains 48 thirty minutes excerpts obtained from 47 patients. Each record contains two leads and all of these leads are sampled at 360 Hz. The Lobachevsky University Electrocardiography Database consists of 200 recordings, each having 10 seconds of duration and having all the 12 standard leads. The signals were sampled at 500 Hz.

Like previously stated, the emphasis of this work is on cardiac arrhythmias, so specific diseases from this category were selected. After choosing the patient with the desired signal from the selected databases, we proceed to the next step.

3.2 Signal Extraction

After the data selection, the ECG signal will be extracted from the chosen lead. To do so, the functions present in the WFDB Toolbox for MATLAB® were used to specify the part of the signal that was selected (Silva and Moody, 2014). The selection was made based on a visual analysis of the signal, as shown in Figure 3.2.
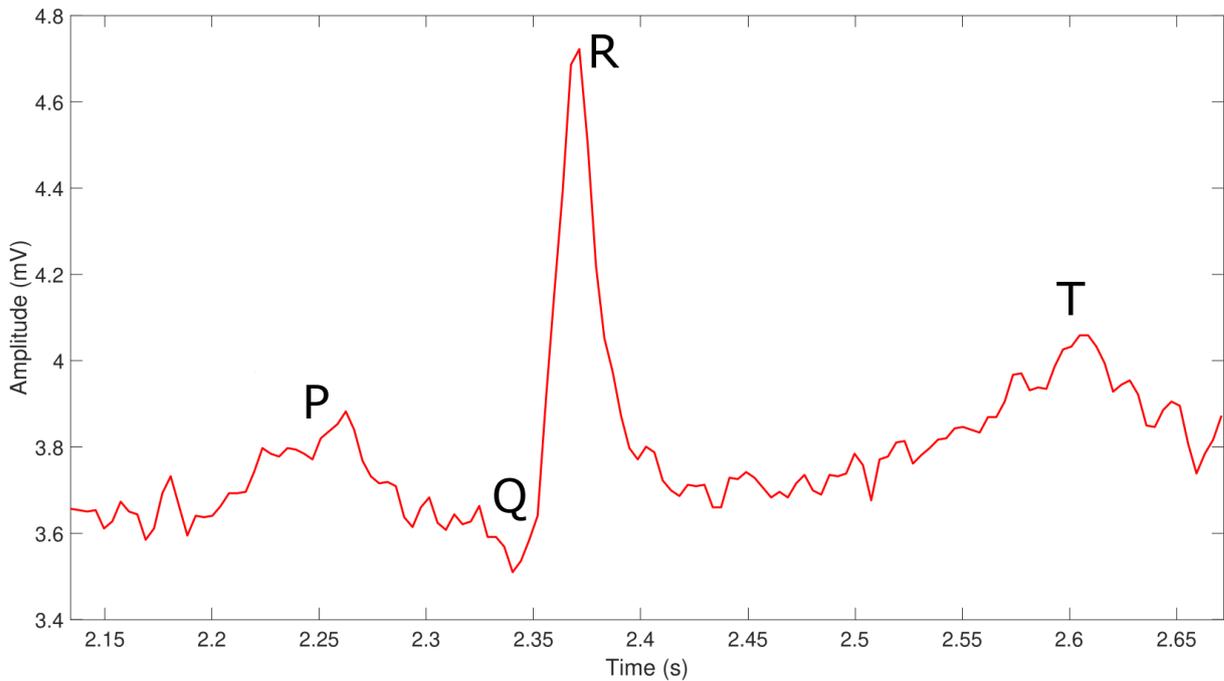
Figure 3.2: Extracted cycle of a healthy ECG from patient I01 and lead I from the St Petersburg INCART 12-lead Arrhythmia Database. It should be noted that the S-wave is absent in this lead.

The functions available on the toolbox cut the signal in the desired segment, returning the sampling time and the signal amplitude for each sample in the excerpt. The database provides the sampling frequency for each signal.

3.3 Signal Filtering

The noisy signal extracted in the previous step went through a filtering process in this step. The chosen filter was the Butterworth filter. The Butterworth filter is a signal processing filter that aims to have a frequency response as flat as possible in the passband, invented by Stephen Butterworth (Butterworth et al., 1930). This type of filter was chosen because of its properties and effectiveness on filtering ECG signals. The process can be visualized on Figure 3.3.
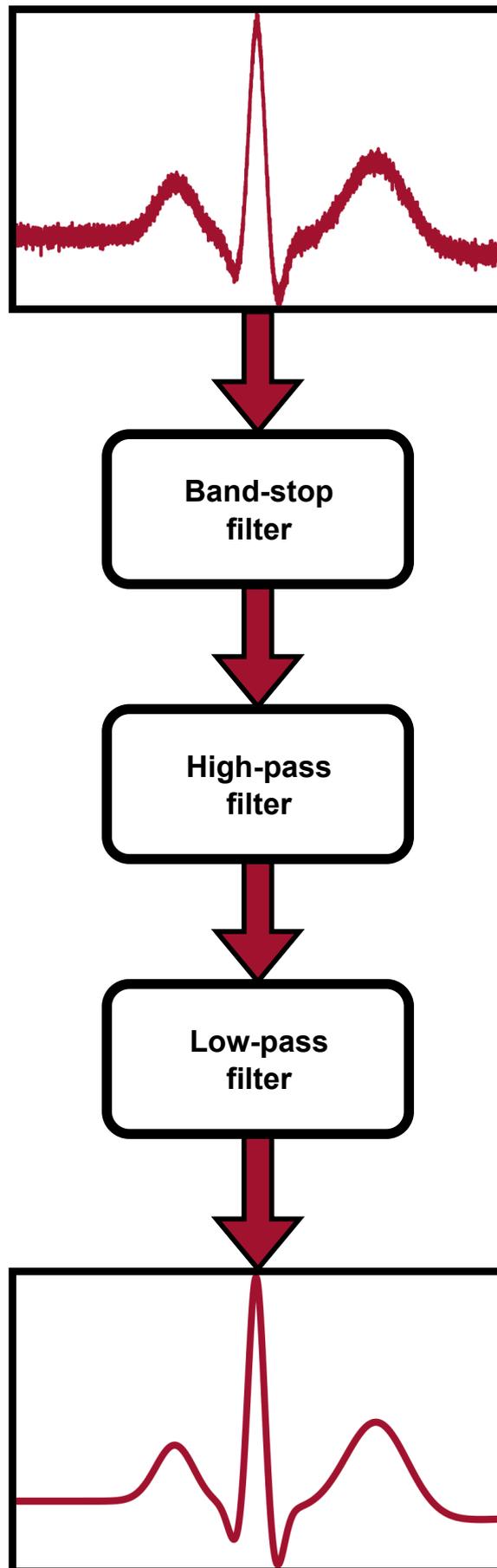
Figure 3.3: Flowchart of the filtering step.

3.3.1 Band-stop Filter

The band-stop filter was designed to attenuate the power-line interference on the signal. The filter used was a 4th-order filter and the selected cutoff frequencies were 49.5 and 50.5 Hz. The Bode Diagram of the filter can be seen in the Figure 3.4. The resultant signal after the application of the band-stop filter is exhibited in Figure 3.5.
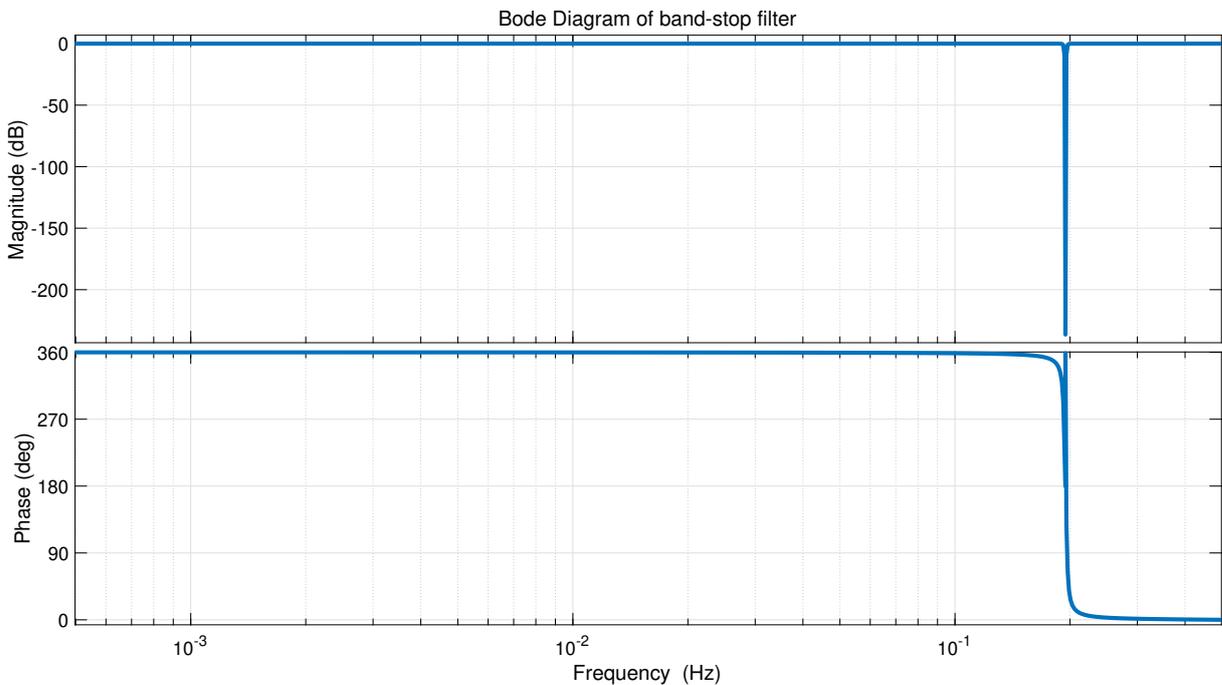
Figure 3.4: Bode Diagram of the band-stop filter.
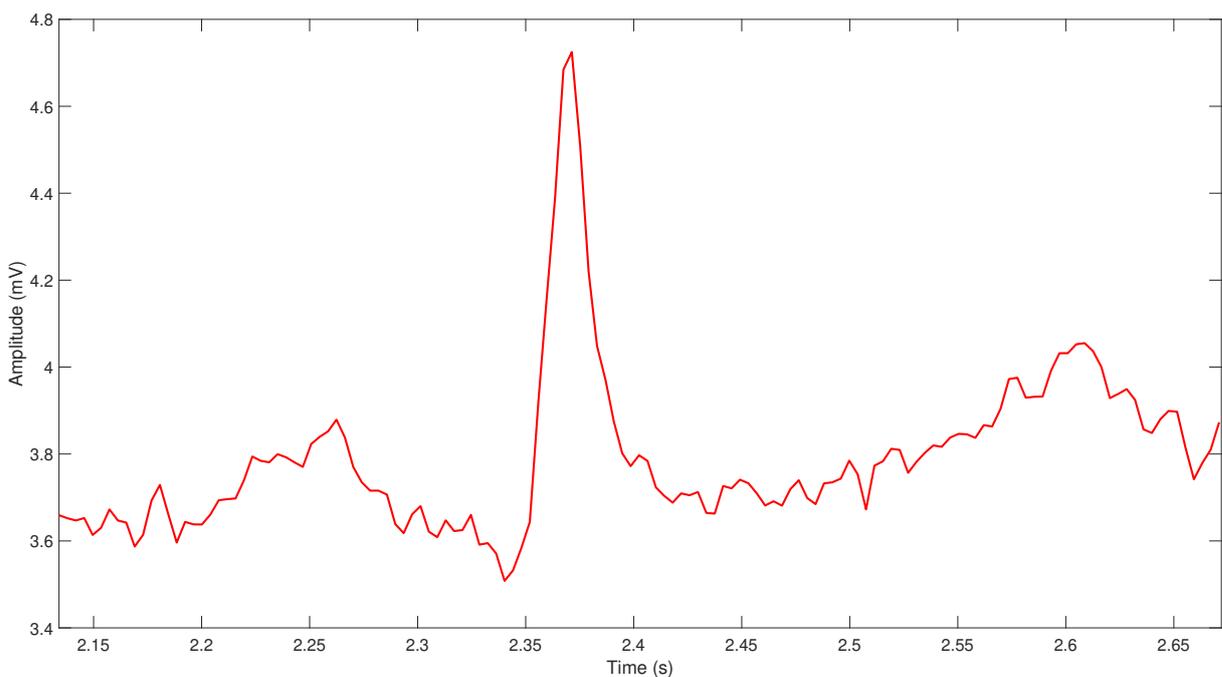
Figure 3.5: Extracted cycle of a healthy ECG from patient I01 and lead I from The St Petersburg INCART 12-lead Arrhythmia Database (Figure 3.2) after applying the band-stop filter.

3.3.2   High-pass Filter

The high-pass filter was designed to attenuate the DC noise interference on the signal and the baseline wander. The filter used was a 2th-order filter and the selected cutoff frequency was 0.5 Hz. The Bode Diagram of the filter can be seen in the Figure 3.6. The resultant signal after the application of the high-pass filter is exhibited in Figure 3.7.
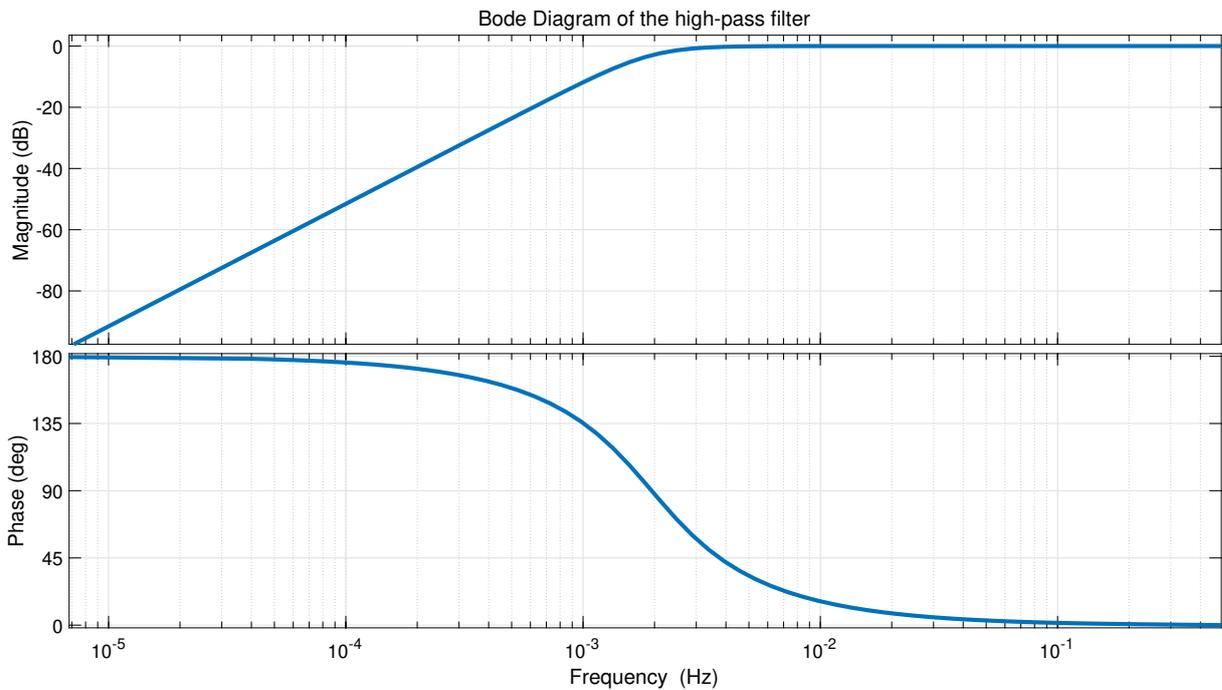
Figure 3.6: Bode Diagram of the high-pass filter.
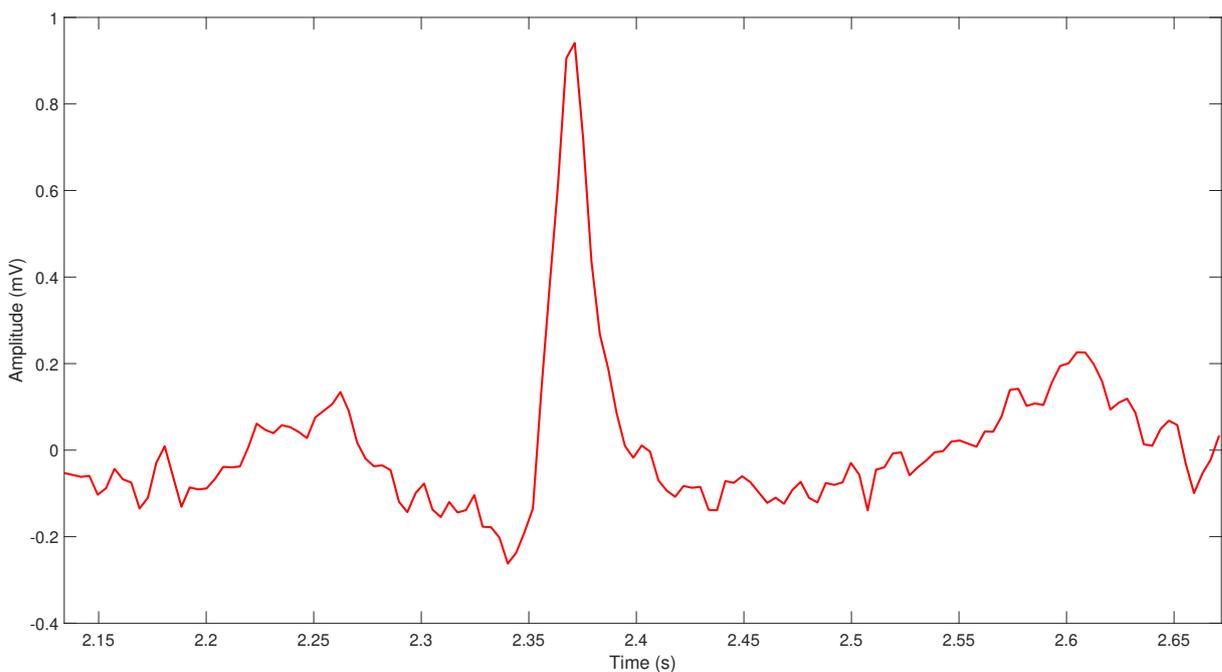
Figure 3.7: Extracted cycle of a healthy ECG from patient I01 and lead I from The St Petersburg INCART 12-lead Arrhythmia Database (Figure 3.2) after applying the high-pass filter.

3.3.3 Low-pass Filter

The low-pass filter was designed to attenuate the high-frequency noise present on the signal. The filter used was a 2th-order filter and the selected cutoff frequency was 40 Hz. The Bode Diagram of the filter can be seen in the Figure 3.8. The resultant signal after the application of the low-pass filter is exhibited in Figure 3.9.
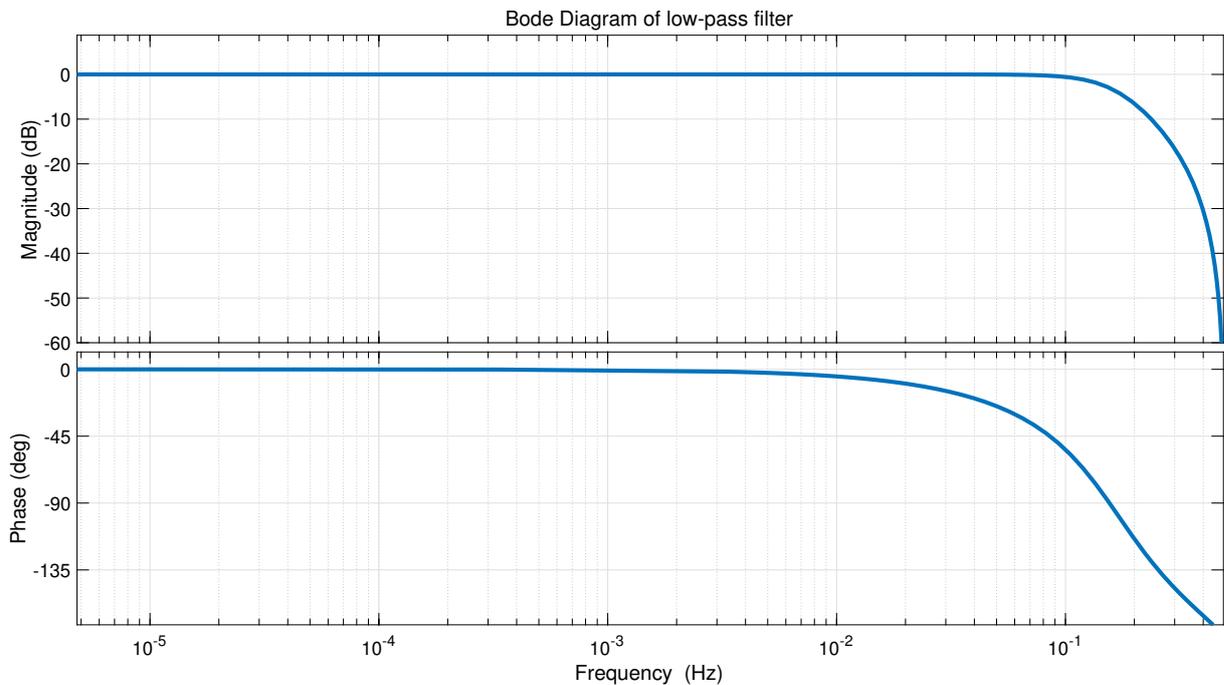
Figure 3.8: Bode Diagram of the low-pass filter.
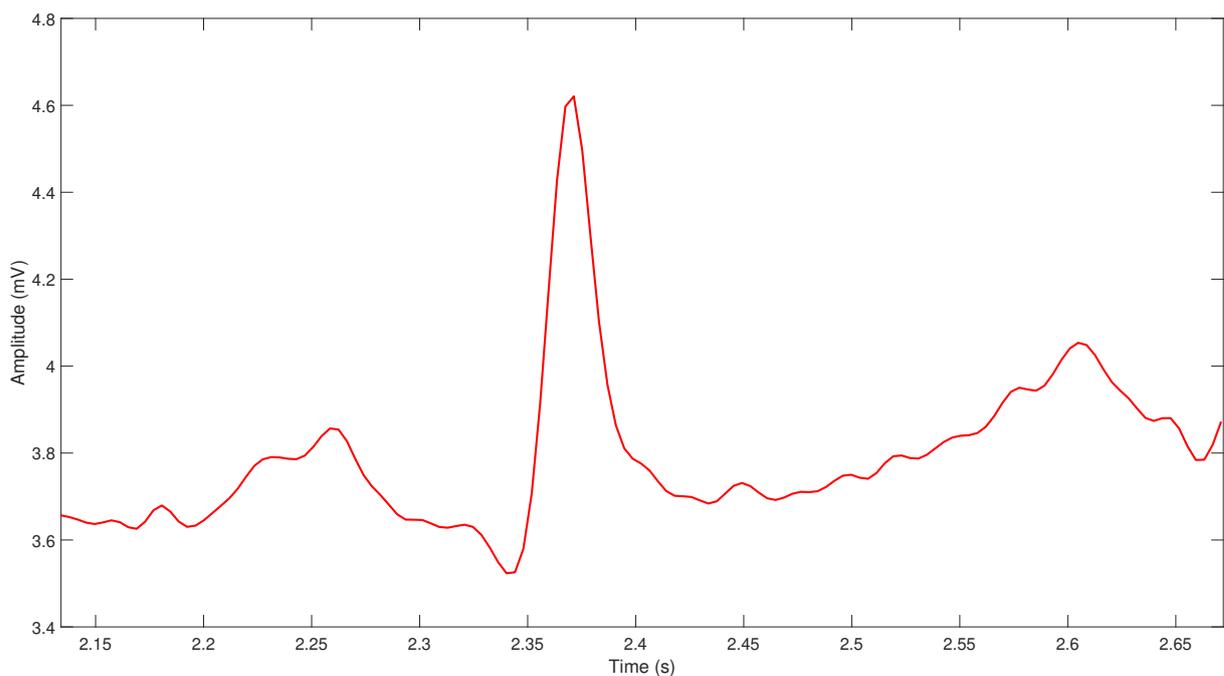
Figure 3.9: Extracted cycle of a healthy ECG from patient I01 and lead I from The St Petersburg INCART 12-lead Arrhythmia Database (Figure 3.2) after applying the low-pass filter.

The final result after applying all the three filters is shown in Figure 3.10. The effect of the three filters can be compared in Figure 3.11.
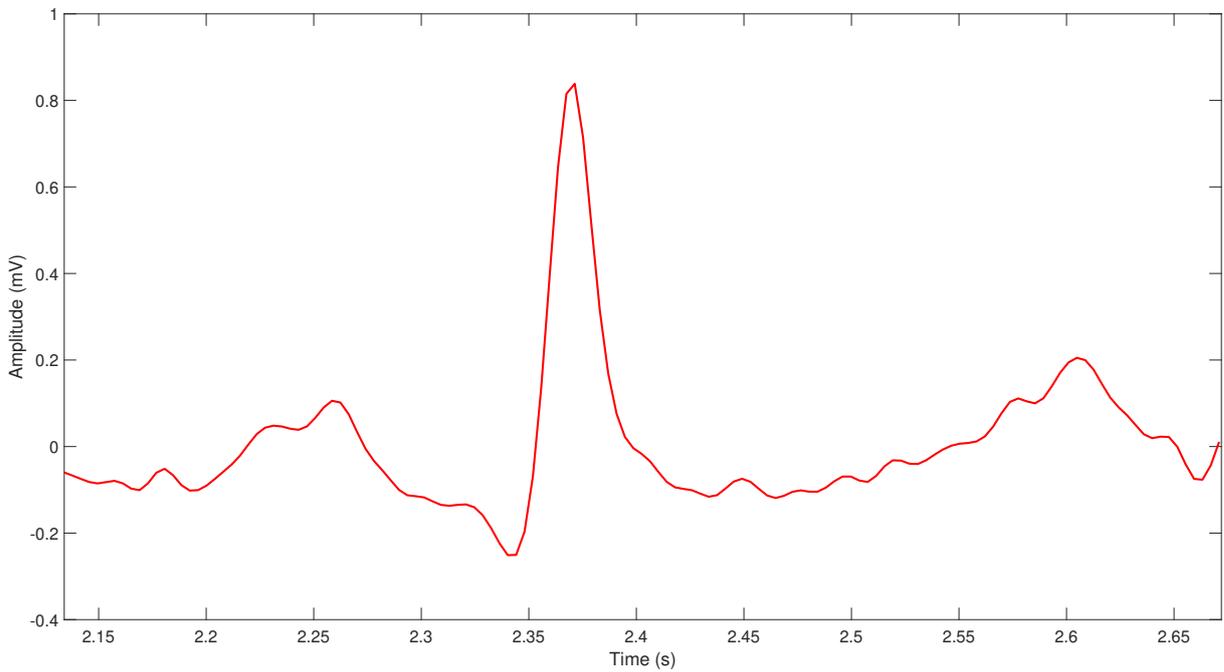
Figure 3.10: Extracted cycle of a healthy ECG from patient I01 and lead I from The St Petersburg INCART 12-lead Arrhythmia Database (Figure 3.2) after applying all the filters.
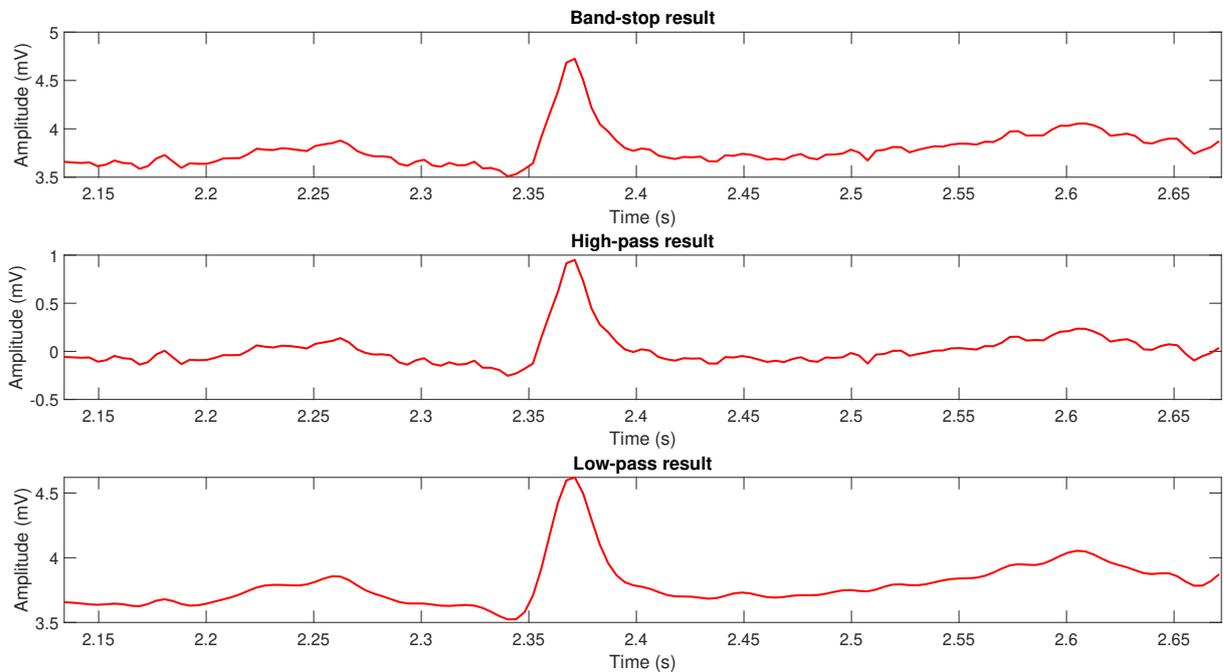
Figure 3.11: Comparison of each filter effect on this methodology step.

3.4   Wave Selection

The filtered signal was separated into waves. The number of waves varies according to the cardiac problem and the chosen lead. The separation methodology was based on a visual analysis. Some ECGs do not have all the five waves mentioned earlier. As shown in Figure 3.2, the S wave is absent. The signal was divided into several individual or combined waves, such as the qRs complex. For example, the Figure 3.10 was divided into its four identified waves, as seen in the Figures 3.12, 3.13, 3.14 and 3.15. In contrast, the Figure 2.8 was divided into the qRs complex, as exhibited in Figure 3.16 and also the T wave, displayed on 3.17.
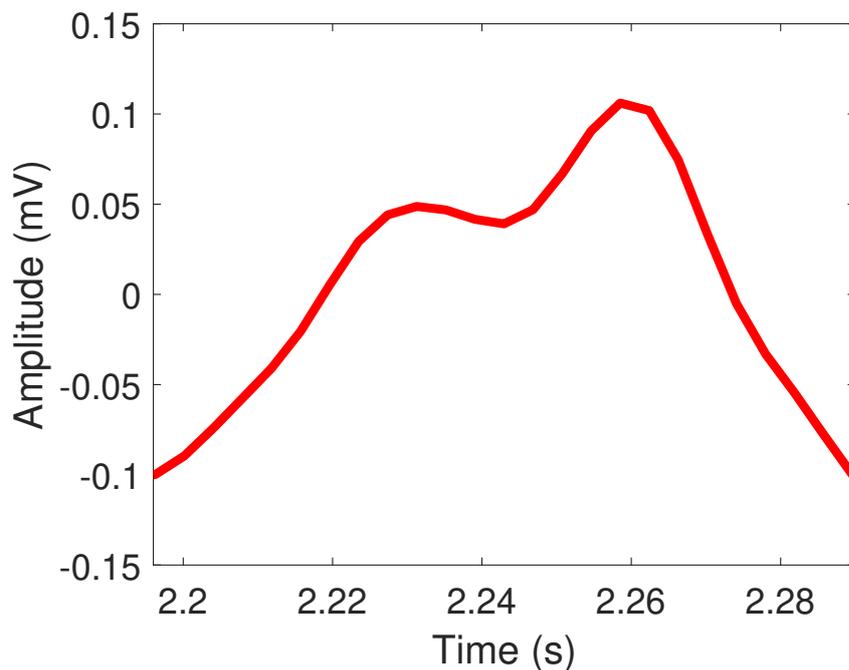
Figure 3.12: P wave from the extracted cycle of a healthy ECG from patient I01 and lead I from The St Petersburg INCART 12-lead Arrhythmia Database (Figure 3.2).
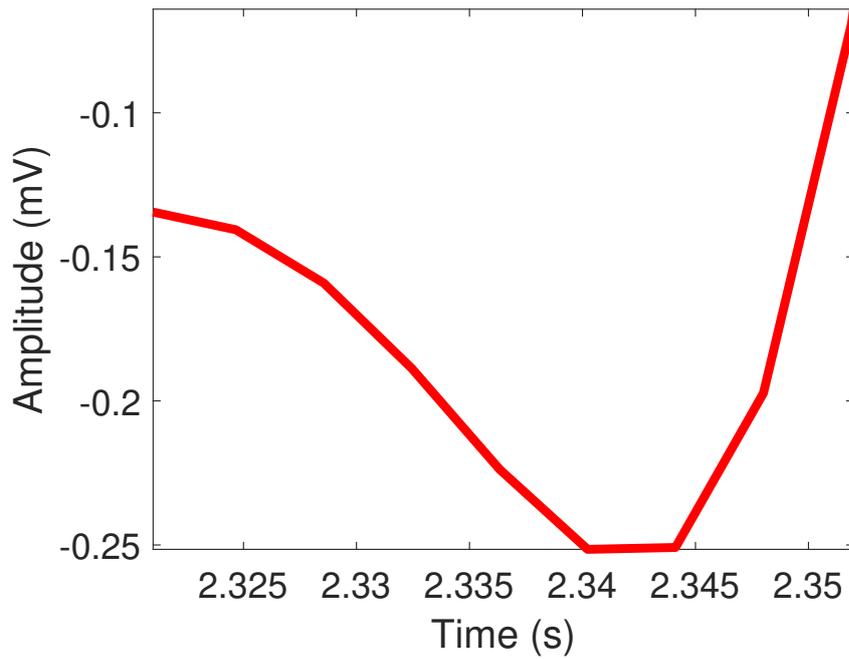
Figure 3.13: Q wave from the extracted cycle of a healthy ECG from patient I01 and lead I from The St Petersburg INCART 12-lead Arrhythmia Database (Figure 3.2).
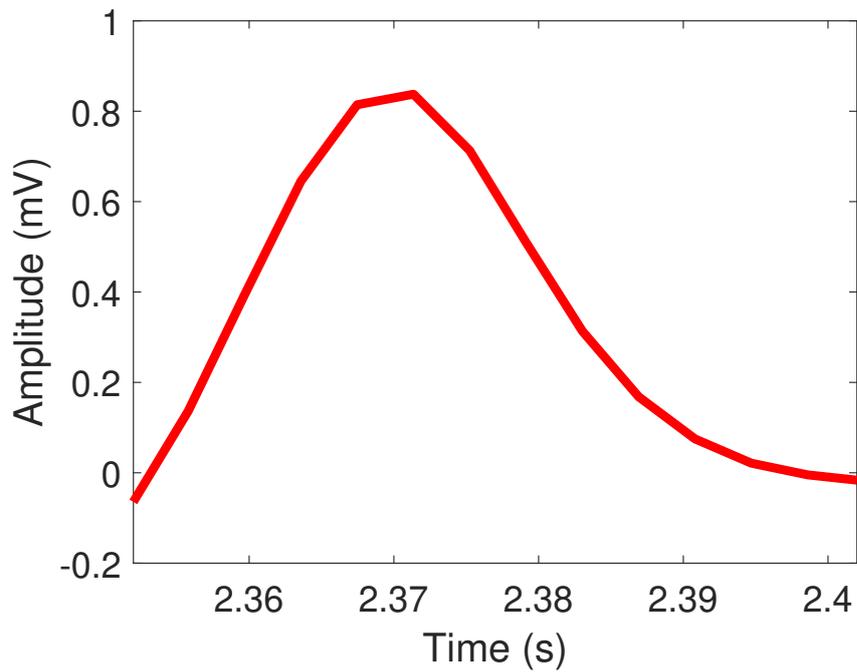
Figure 3.14: R wave from the extracted cycle of a healthy ECG from patient I01 and lead I from The St Petersburg INCART 12-lead Arrhythmia Database (Figure 3.2).
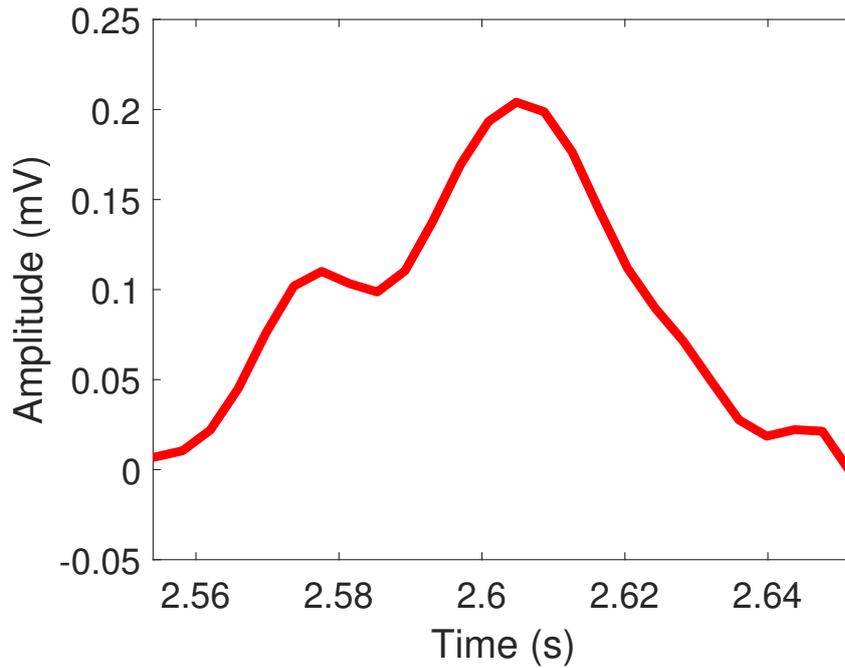
Figure 3.15: T wave from the extracted cycle of a healthy ECG from patient I01 and lead I from The St Petersburg INCART 12-lead Arrhythmia Database (Figure 3.2).
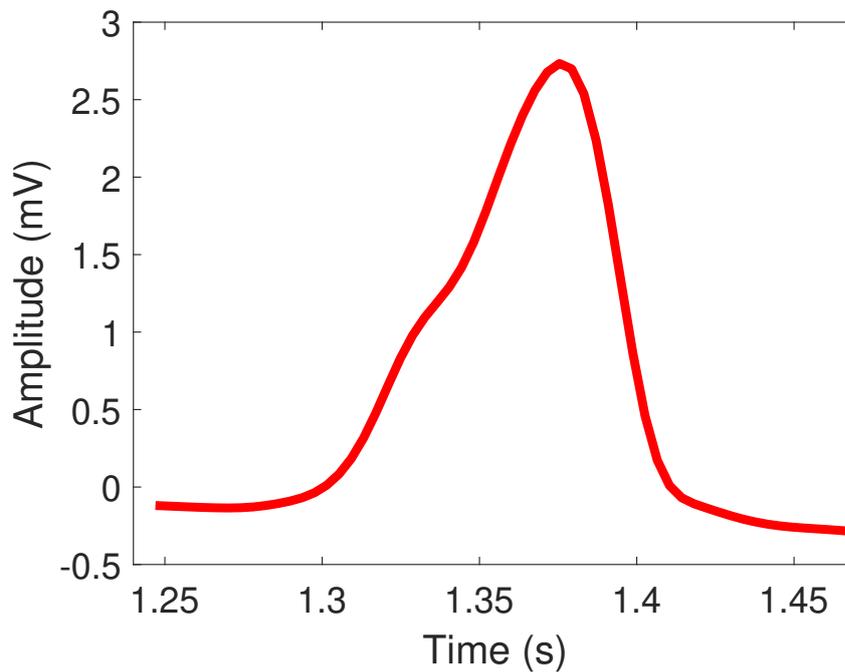
Figure 3.16: qRs complex obtained from the Lead II displaying a PVC extracted from patient I13 from St Petersburg INCART 12-lead Arrhythmia Database. (Figure 2.9).
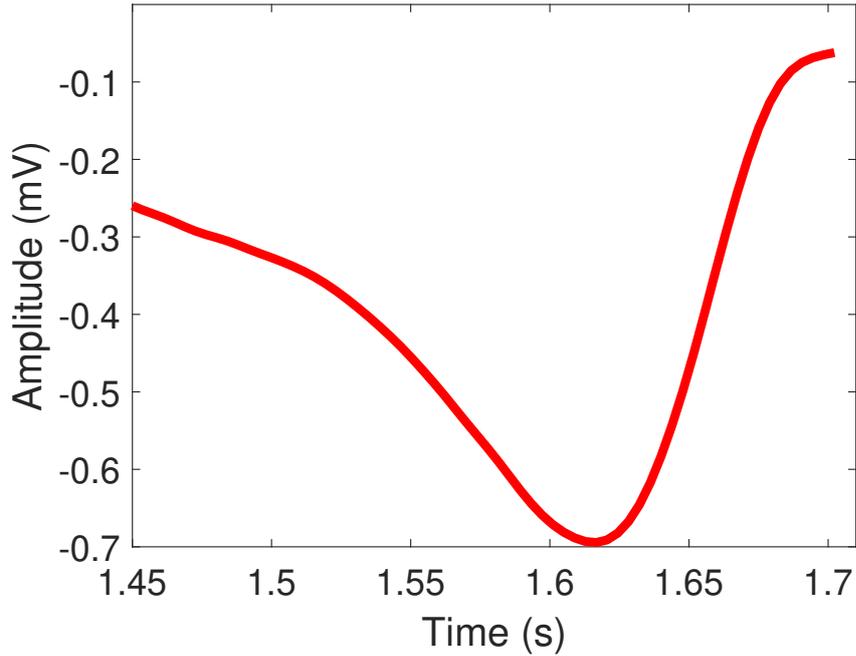
Figure 3.17: T wave obtained from the Lead II displaying a PVC extracted from patient I13 from St Petersburg INCART 12-lead Arrhythmia Database. (Figure 2.9).

3.5 Parameter Estimation

For the final part, the signal is ready to be used for estimating the parameters of the mathematical model. The differential evolution implementation from the SciPy API (Jones et al., 2001–) was utilized with the following parameters and options:

- Each individual of the population was composed of 3 parameters (a, b, θ) for each wave;

- Bounds were defined for each parameter and they varied for different waves and different signals;

  – The bounds of the parameters b and θ were not changed, they were defined as $b \in [-30, 30]$ and $\theta \in [-\pi, \pi]$;

  – The initial guess of the bounds was done by studying the Gaussian function parameters and analysing each wave. After a suitable initial guess was made, the interval was reduced to optimize searching time.

- The cost function used was the RMSE between the extracted signal and the synthetic signal generated by the model with the parameters obtained from each iteration, defined by:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{3.1}$$

- The population size (NP) was 50 individuals;

- The mutation strategy used was the 'best-1-bin' strategy;

- The maximum number of iterations was 200.

Chapter 4

Results

All the results were obtained using a laptop with an Intel® Core™ i5-11400 CPU, 32GB of 3200 MHz RAM and a NVIDIA Geforce RTX 3060 GPU.

4.1 Premature Ventricular Contraction

Following the proposed methodology, the signal I13, lead II, from the St Petersburg IN-CART 12-lead Arrhythmia Database, was selected. The signal is referred to a 39-year-old woman acquired with a sampling frequency of 257 samples per second. A second-order Butterworth filter was designed for the filtering process, as shown before. The chosen PVC cycle can be seen in Figure 4.1. The beginning and the endpoint of the signal were defined by visual analysis.
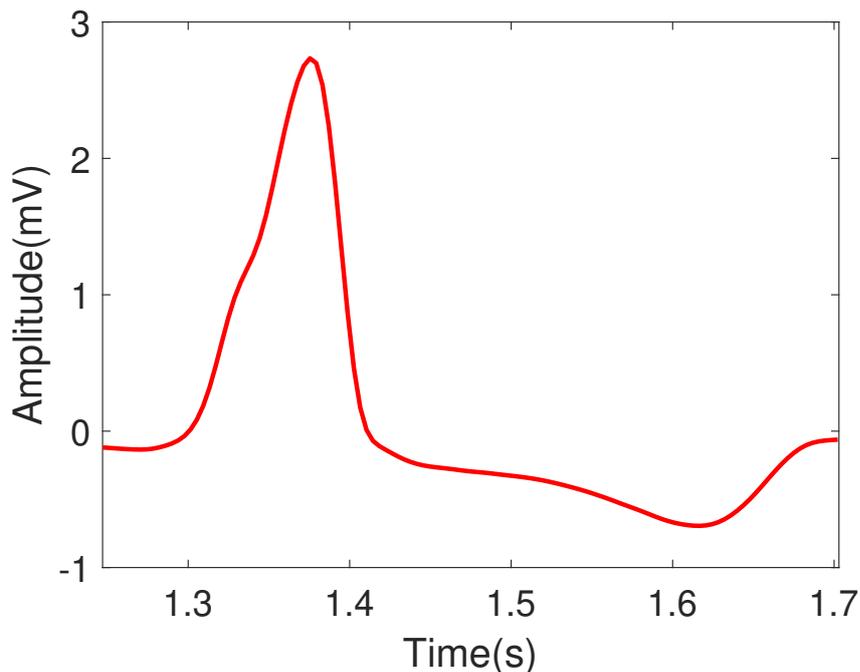
Figure 4.1: Filtered PVC signal used as reference.

The DE algorithm was executed for each ECG wave, modifying only the bounds for the a parameter of each wave. The values of the parameters b and θ were not changed for this analysis, they were defined as $b \in [-30, 30]$ and $\theta \in [-\pi, \pi]$. A few tests were expected to analyze the visual effect and its corresponding fitness value. The attempt to reduce the final interval was to help the algorithm converge faster, checking if any parameter was at the maximum or minimum value of the defined interval. The bounds that produced the best fit and the model parameters are shown in Table 4.1.
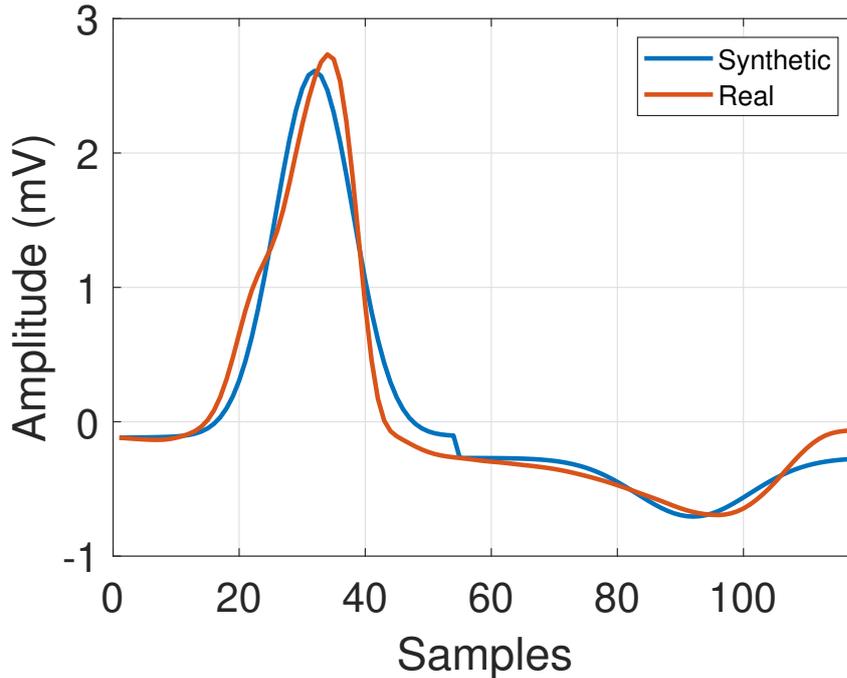
Figure 4.2: PVC resultant signal after going through every step of the methodology and the original filtered signal.

Table 4.1: Results from PVC and the best parameters for the model.

Wave	Bounds	Fitness	Time(s)	a	b	θ
qRs	[9000,10000]	0.030094	22.32	9989.230	-0.0151	-3.063
T	[-4000,-3000]	0.053662	21.86	-3583.960	0.0220	-3.048

4.2 Right Bundle Branch Block

The signal I16 from the INCART Arrhythmia Database, V1 lead, was selected using the same methodology. The patient is a 64-year-old that was diagnosed with a transient ischemic attack. In this ECG signal, 1520 cycles were labeled as RBBB. One of these cycles was selected (Fig. 4.3), and the same filters used in the PVC signal were applied.
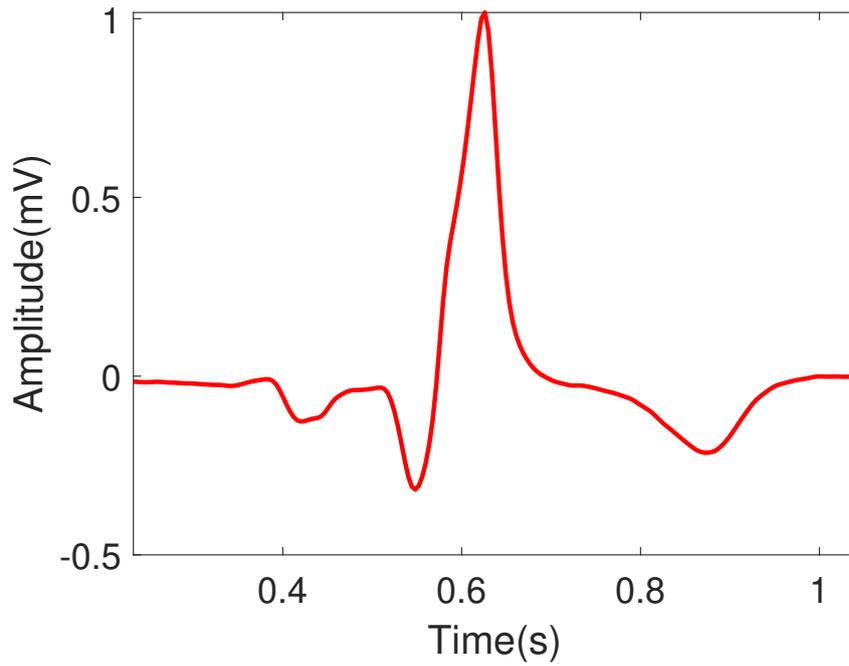
Figure 4.3: Filtered RBBB signal used as reference.

For the RBBB signal, only the bounds for the parameter a were modified. A few tests were run to find the best interval, analyzing the visual result and its corresponding fitness value. The bounds that produced the best fit and the model parameters are shown in Table 4.2.
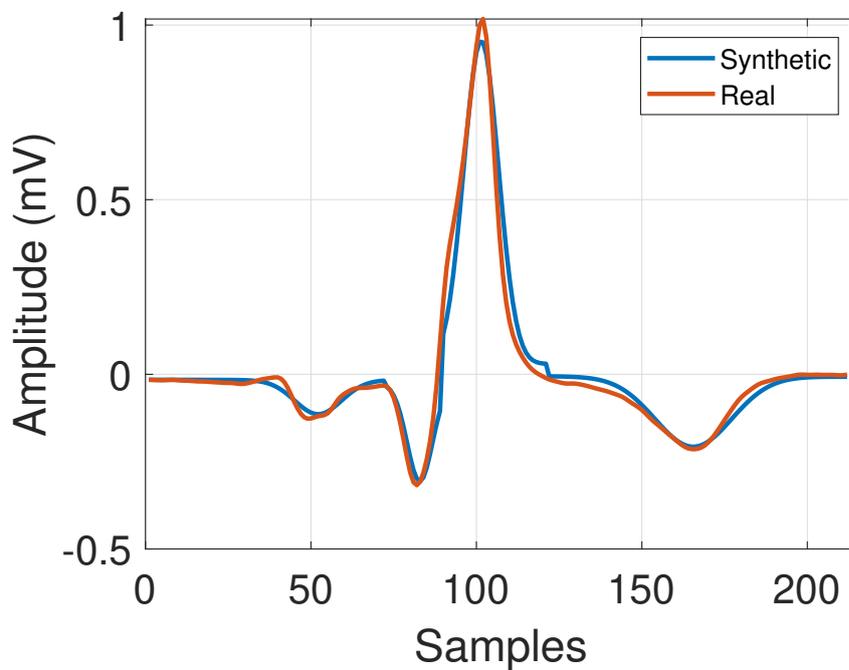
Figure 4.4: RBBB resultant signal after going through every step of the methodology and the original filtered signal.

Table 4.2: Results from RBBB and the best parameters for the model.

Wave	Bounds	Fitness	Time(s)	a	b	θ
P	[-7000,-6000]	0.039292	23.18	-6999.999	0.017	-3.013
Q	[-30000,-21677]	0.030174	10.80	-27547.246	0.009	-3.114
R	[13000,15000]	0.022525	14.83	13688.664	-0.012	-3.111
T	[-3000,-1000]	0.02226	34.51	-2953.654	-0.028	-3.032

4.3 Left Bundle Branch Block

The signal 24 from the Lobachevsky University Electrocardiography Database (Kalyakulina et al., 2020), V1 lead, was also selected using the same steps as the other signals. The patient is a 64-year-old that was diagnosed with a complete left bundle branch block. One cycle with LBBB (Fig. 4.5) was chosen and the same filter used on the other diseases was applied to the cycle. Both the S and T waves were identified on the extracted cycle.
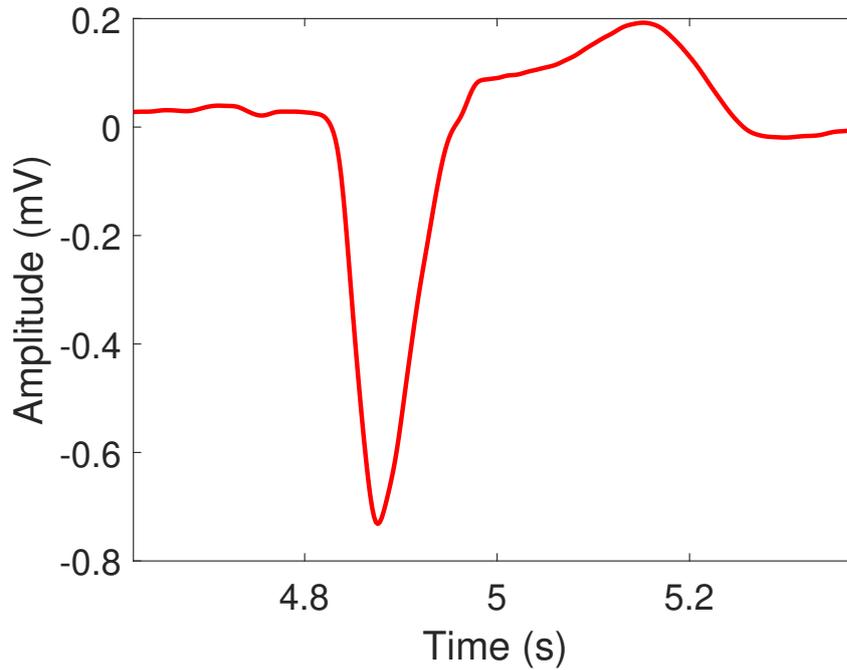
Figure 4.5: Filtered LBBB signal used as reference.

The a parameter was the only one that had its bounds modified. After running some experiments with the bounds, the final interval was determined and the best outcome was saved. The results can be seen in Table 4.3:
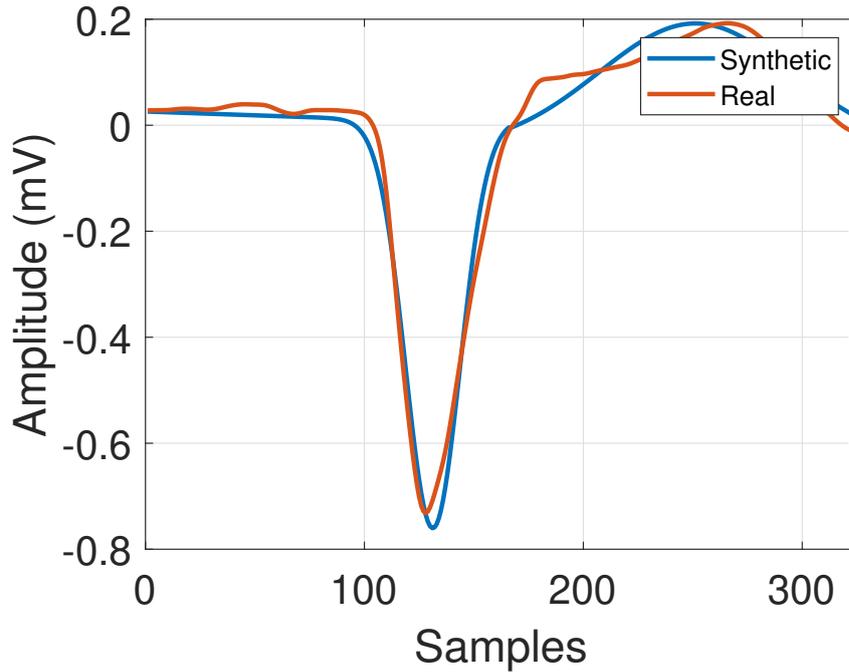
Figure 4.6: LBBB resultant signal after going through every step of the methodology and the original filtered signal.

Table 4.3: Results from LBBB and the best parameters for the model.

Wave	Bounds	Fitness	Time(s)	a	b	θ
S	[-3000,-1000]	0.016598	56.72	-2705.490	0.030	-2.819
T	[100,300]	0.052943	49.88	250.000	0.107	-2.932

4.4 Atrial Fibrillation

The last signal analyzed was the signal 221 from the MIT-BIH Arrhythmia Database (Moody and Mark, 2001). The patient is an 83-year-old that had 12 episodes of atrial fibrillation. One of these episodes was picked and had all the filters from the proposed methodology applied. This episode can be seen on Figure 4.7.
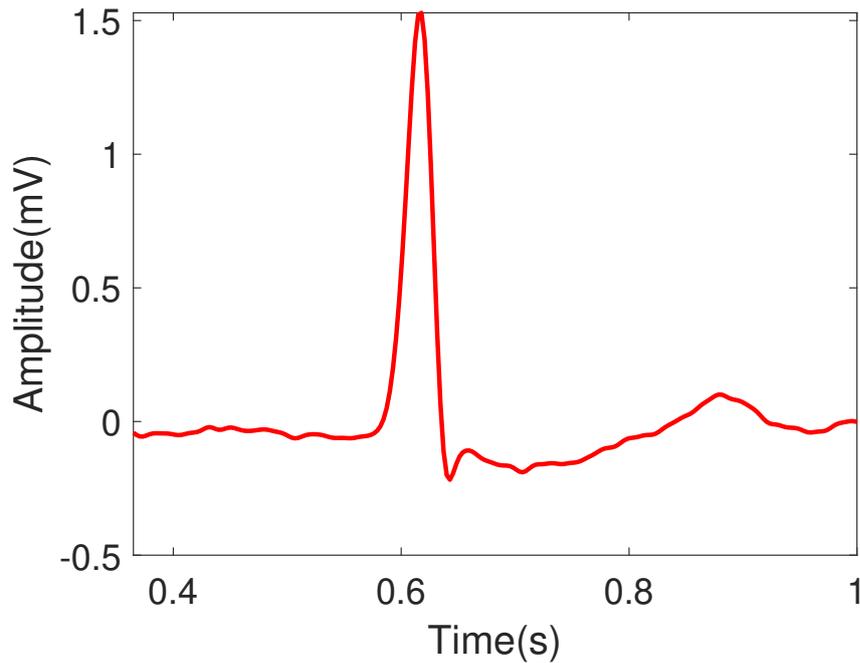
Figure 4.7: Filtered AF signal used as reference.

Here it was also observed that the disease required more than one gaussian per wave to reflect the real signal characteristics. The signal was split in 3 waves, R, S and T waves. The T wave required one more gaussian to fully represent and guarantee the signal's continuity.

Like all other signals, the only parameter considered for the bound analysis was the a parameter. A range of values was experimented and the interval that resulted in the best outcome was saved. The results can be seen in Table 4.4:
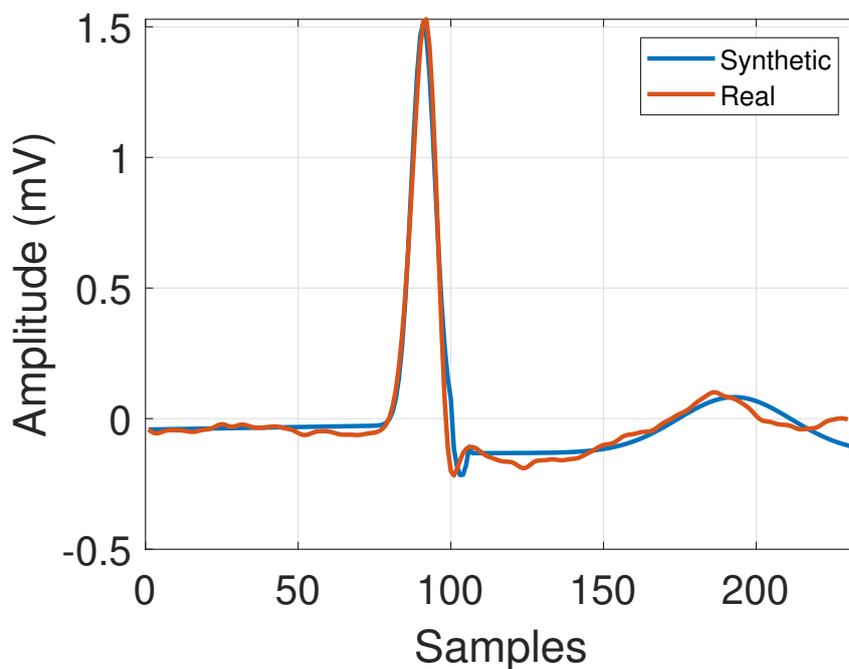
Figure 4.8: AF resultant signal after going through every step of the methodology and the original filtered signal.

Table 4.4: Results from AF and the best parameters for the model.

Wave	Bounds	Fitness	Time(s)	a	b	θ
R	[48000,50000]	0.012115	35.33	48486.325	-0.0068	-2.981
S	[-295000,-265000]	0.063315	7.41	-265000.038	1.883	-3.135
T	[-20000,-15000],[1000,3000]	0.043738	90.11	-10220.592, 1726.812	-0.1237, 0.0327	1.349, -2.988

Chapter 5

Conclusion and Future Works

In this work, the synthetic ECG model developed in (McSharry et al., 2003) was used. The objective was to obtain the parameters that best fit the model to real ECGs with some cardiac arrhythmia. There are several different models in the literature using two or more gaussians for each wave, which can lead to better results. For specific ECG signals, it was shown that multiple gaussians can boost the estimation accuracy. Moreover, some cardiac issues can benefit significantly from the increment in the number of gaussians. The methodology showed in this work can be further expanded to optimize different models, such as the synthetic photoplethysmogram (PPG) model (Tang et al., 2020).

An enhanced methodology for the individual wave selection can also be an essential improvement to the process described in this work. A new method to divide the signal into multiple waves that does not require human interaction could reallocate the time spent manually separating the waves to other essential parts of the procedure. Since the estimation was done per wave, they were merged to display the final result correctly. This fact can lead to discontinuities between the waves and impact the final result. A better way to join the waves and smooth the transitions can boost the final results.

To achieve this, we utilized a method of parameter estimation using an evolutionary computation algorithm: Differential Evolution. Given the ease of use, the excellent convergence properties, and the potential to be used in parallel, this algorithm brought satisfactory results in a relatively short time. The results show that the proposed methodology is able to maintain the properties of the original signal using only one gaussian per ECG wave for the most part and is highly adaptable to different cardiac diseases. Moreover, the model can replicate ECG recordings at their original frequency, and other noises can be added. It is also possible to work with different leads, if available.

Although good results with the original DE algorithm have been achieved, there is room for further improvement in several aspects. For instance, a few variants of the DE algorithm could be investigated, such as CODE, JADE, SADE, and JDE (Georgioudakis and Plevris, 2020). These variants could promote better convergence properties and other benefits.

Furthermore, another area to be investigated is the utilization of GPUs to improve the al-

gorithm's performance. The possibility of using parallel computing alongside the incredible power of the GPUs can lead to a significant performance increase. A computer with a better CPU for parallel applications can also speed up convergence.

In addition, there are several possible applications where this work could be used. Firstly, the estimation of real ECG signals can be a helpful tool to aid doctors in diagnosing various heart problems. The generation of different ECG signals can also be used for education purposes, e.g., teaching medical students using software that replicates several ECGs under other conditions. Another possibility is to enrich ECGs datasets used for classification techniques. With relation to the DE algorithm, it can be tested with different synthetic ECG models, as for Awal's model (Awal et al., 2021).

Bibliography

M. Abd Elaziz, S. Xiong, K. Jayasena, and L. Li. Task scheduling in cloud computing based on hybrid moth search algorithm and differential evolution. *Knowledge-Based Systems*, 169: 39–52, 2019.

A. Alonso, F. H. J. Aparicio, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, F. A. P. Carson, F. A. M. Chamberlain, B. M. Kissela, F. K. L. Knutson, C. D. Lee, et al. Heart disease and stroke statistics—2021 update. *Circulation*, 2021(143), 2021.

M. Awal, S. S. Mostafa, M. Ahmad, M. A. Alahe, M. A. Rashid, A. Z. Kouzani, M. Mahmud, et al. Design and optimization of ecg modeling for generating different cardiac dysrhythmias. *Sensors*, 21(5):1638, 2021.

T. Bartz-Beielstein, J. Branke, J. Mehnen, and O. Mersmann. Evolutionary algorithms. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(3):178–195, Apr. 2014. doi: 10.1002/widm.1124. URL `https://doi.org/10.1002/widm.1124`.

H.-G. Beyer, T. B. To, B. Sendhoff, H. Pohlheim, W. Jakob, and E. Brucherseifer. Glossary, Feb 2002. URL `https://homepages.fhv.at/hgb/ea-glossary/ea-terms-engl.html`.

S. Butterworth et al. On the theory of filter amplifiers. *Wireless Engineer*, 7(6):536–541, 1930.

Cables and Sensors. 12 lead ECG placement guide. `https://www.cablesandsensors.com/pages/12-lead-ecg-placement-guide-wit` Accessed: 2023-1-29.

G. D. Clifford, C. Liu, B. Moody, H. L. Li-wei, I. Silva, Q. Li, A. 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)*, pages 1–4. IEEE, 2017.

C. Darwin. *On the Origin of Species by Means of Natural Selection*. Murray, London, 1859. or the Preservation of Favored Races in the Struggle for Life.

P. Dolinský, I. Andras, L. Michaeli, and D. Grimaldi. Model for generating simple synthetic ecg signals. *Acta Electrotechnica et Informatica*, 18(3):3–8, 2018.

R. Gämperle, S. D. Müller, and P. Koumoutsakos. A parameter study for differential evolution. *Advances in intelligent systems, fuzzy systems, evolutionary computation*, 10(10):293–298, 2002.

M. Georgioudakis and V. Plevris. A comparative study of differential evolution variants in constrained structural optimization. *Frontiers in Built Environment*, 6:102, 2020.

C. Gianni, J. D. Burkhardt, C. Trivedi, S. Mohanty, and A. Natale. The role of the purkinje network in premature ventricular complex-triggered ventricular fibrillation. *Journal of Interventional Cardiac Electrophysiology*, 52(3):375–383, 2018.

A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101 (23):e215–e220, 2000.

J. E. Hall and M. E. Hall. *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences, 2020.

E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. URL `http://www.scipy.org/`.

A. Kalyakulina, I. Yusipov, V. Moskalenko, A. Nikolskiy, K. Kosonogov, N. Zolotykh, and M. Ivanchenko. Lobachevsky university electrocardiography database. *Type: Dataset. Available online: https://physionet. org/content/ludb/1.0. 0/(accessed on 10 July 2021)*, 2020.

P. T. Kapen, S. U. K. Kouam, and G. Tchuen. A comparative study between normal electrocardiogram signal and those of some cardiac arrhythmias based on mcsharry mathematical model. *Australasian physical & eng. sciences in medicine*, 42(2), 2019.

J. Liu. On setting the control parameter of the differential evolution method. In *Proceedings of the 8th international conference on soft computing (MENDEL 2002)*, pages 11–18, 2002.

P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Trans. on Bio. Engineering*, 50(3):289–294, 2003.

G. B. Moody and R. G. Mark. The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001.

OpenStax College. Illustration from anatomy physiology, connexions web site, 2013. URL `http://cnx.org/content/col11496/1.6/`. This work is licensed under the Creative Commons Attribution 3.0 International License. To view a copy of this license, visit `http://creativecommons.org/licenses/by/3.0/`.

A. H. Ribeiro, M. H. Ribeiro, G. M. Paixão, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. Ferreira, C. R. Andersson, P. W. Macfarlane, W. Meira Jr, et al. Automatic diagnosis of the 12-lead ecg using a deep neural network. *Nature communications*, 11(1):1–9, 2020.

J. Ronkkonen, S. Kukkonen, and K. V. Price. Real-parameter optimization with differential evolution. In *2005 IEEE congress on evolutionary computation*, volume 1, pages 506–513. IEEE, 2005.

I. Silva and G. B. Moody. An open-source toolbox for analysing and processing physionet databases in matlab and octave. *Journal of open research software*, 2(1), 2014.

C. Sridhar, O. S. Lih, V. Jahmunah, J. E. Koh, E. J. Ciaccio, T. R. San, N. Arunkumar, S. Kadry, and U. Rajendra Acharya. Accurate detection of myocardial infarction using non linear features with ecg signals. *Journal of Ambient Intelligence and Humanized Computing*, 12(3): 3227–3244, 2021.

R. Storn and K. Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

S. Tan, G. Androz, A. Chamseddine, P. Fecteau, A. Courville, Y. Bengio, and J. P. Cohen. Icentia11k: An unsupervised representation learning dataset for arrhythmia subtype discovery. *arXiv preprint arXiv:1910.09570*, 2019.

Q. Tang, Z. Chen, R. Ward, and M. Elgendi. Synthetic photoplethysmogram generation using two gaussian functions. *Scientific Reports*, 10(1):1–10, 2020.

K. Ten Tusscher and A. V. Panfilov. Modelling of the ventricular conduction system. *Progress in biophysics and molecular biology*, 96(1-3):152–170, 2008.

K. Weimann and T. O. Conrad. Transfer learning for ecg classification. *Scientific reports*, 11 (1):1–12, 2021.

H. Zhang, T. Liu, X. Ye, A. A. Heidari, G. Liang, H. Chen, and Z. Pan. Differential evolution-assisted salp swarm algorithm with chaotic structure for real-world problems. *Engineering with Computers*, pages 1–35, 2022.