People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University of M'hamed BOUGARA Boumerdes

Institute of Electrical and Electronic Engineering

Department of Electronics

Final Year Project Report Presented in Partial Fulfillment of the Requirements of the Degree of

MASTER

In Computer Engineering

Title:

ECG Identification and Monitoring System

Presented By:

AMMARKHODJA Dalia

HAMADOUCHE Chahrazed

Supervisor and Co-supervisors:

Pr. A.DAAMOUCHE

Pr. A.AMIRA

Dr. R.NAMANE

Spring 2021

Dedication

To the two persons who have meant and continue to mean so much to me. Although they are no longer part of this world, their memory keeps me moving forward, to the memory of my father and grandmother.

To my lovely mother, the strongest woman I have ever met, the woman who taught me how to be autonomous and strong, words can never be enough to express how grateful I am for her, without her I would have never become the person I am today.

To my grandfather, the wisest man I met, you were always my go-to person when I feel bad.

To my three sisters, thank you for supporting me through my hard times, thank you for always being here for me, thank you for your endless love.

To all my friends who believed in me and motivated me through my 5 years at IGEE, I am glad that I met you and lucky to have you.

Chahrazed

It is my genuine gratefulness and warmest regard that I dedicate this work to my loving parents for their limitless love and countless encouragement, and for being there day after day to make sure my life turned out this way. I can never thank them adequately for all what they have done and always do for me.

To my brother for protecting me when needed and teaching me how to be a better person and for making every day of my life so fun and enjoyable.

To my fiance for his great support, advice and for hearing all my pointless dramas about this work over and over again.

To all my friends who have encouraged and supported me. Words cannot describe how truly grateful I am to have such amazing people in my life; without each of them I'd be nowhere near the person I am.

Dalia

Acknowledgements

Praise to Allah the Almighty and most merciful for giving us the wisdom, persistence and potency to complete this project.

We revere the patronage and moral support extended with love by our families, especially our parents whose financial and constant passionate encouragement made this undertaking a reality.

It is our genuine and deepest pleasure to submit our heartiest gratitude and appreciation to our respected supervisors Pr.Daamouche, Pr.Abbes, and Dr Namane for their guidance, support, timely advice and scientific approach that made the completion of this project possible.

We perfusely thank our colleagues, distinguished teachers and people whose assistance and contributions are greatly acknowledged.

Abstract

Technologies emerging the world have raised new threats. They made the traditional identification modalities vulnerable and easy to counterfeit. Hence, more sophisticated biometrics become indispensable to protect the integrity of the devices and the sensitive data. In the last few years there has been a growing interest in electrocardiogram (ECG) due to its uniqueness. In this work, we developed an ECG identification system. We employed this system into a real world application while solving a second problem which is the digitization of the health system. More precisely, it is an ECG remote monitoring system with ECG based identification. Two datasets were used: the MIT-BIH Arrhythmia and our own collected dataset using our implemented hardware system. The collected data was denoised using a wavelet transform based filter. After segmenting the data and extracting the necessary fiducial features, we stored the generated templates. A template matching algorithm based on the correlation factor was considered to make the matching decisions. Finally, the whole process is presented to the user through a user-friendly android application. Preliminary experimental results indicate that the system achieved 87.5% to 90.90% accuracy.

Keywords: ECG, biometric, identification, monitoring, android application, template matching.

Contents

List of Figures

List of Tables

List of Acronyms

ADC Analog to Digital Converter.

DBMS Database Management System.

DNA Deoxyribonucleic Acid.

DWT Discrete Wavelet Transform.

ECG Electrocardiogram.

EER Enhanced Entity-Relationship.

EMF Electromagnetic Field.

FIR Finite Impulse Response.

GPU Graphics Processing Unit.

IDE Integrated Development Environment.

IIR Infinite Impulse Response.

IWT Inverse Wavelet Transform.

MIT Massachusetts Institute of Technology.

PWM Pulse Width Modulation.

SDK Software Development Kit.

SQL Structured Query Language.

SSP Serial Port Protocol.

SVM Support Vector Machine.

SWT Stationnary Wavelet Transform.

UI User Interface.

UML Unified Modeling Language.

General Introduction

The practice of distinguishing humans based on intrinsic physical or behavioral traits goes back thousands of years. It all started with the fingerprints used in Babylonian business transactions in 500 BC. Then, with Chinese merchants in the fourteenth century using children's palms and footprints to distinguish them. Moving to Egyptian history, where traders were differentiated by their physical characteristics [43]. This longstanding concern of identifying people, was developed with time. In other words, the more we advance in technology, the more accurate techniques we use. However, the goal always remains the same "identifying people", also known as biometric identification.

Nowadays, more innovative biometrics are used such as face recognition, eye iris, DNA, voice recognition, and even keystroke dynamics. These identifiers are used in several applications starting with military access control, criminal or civil identification, banking, mobile commerce, and even healthcare. However, this technological advancement did not only make identifying people easier, but it also made falsifying and identity theft easier. According to the statistics, cybercrimes and identity theft has increased especially during the COVID-19 crisis [41]. There were 4.8 million identity theft and fraud reports received by the FTC in 2020 [5]. In order to avoid such threats, more sophisticated biometrics are required. One of the proposed methods is the continuous biometric authentication. In addition to its high accuracy in identifying people, it also provides a proof of vitality of the authenticated person. This might be of a huge advantage in many applications.

ECG is among the most promising techniques in continuous authentication. In the last decade it has attracted much attention from research teams. It provides an accurate identification and a vitality proof at the same time. Furthermore, this bio-signal is used by physicians to diagnose cardiovascular diseases. Therefore, it can combine both the security and healthcare applications. In other words, it can be deployed in a remote healthcare monitoring system while providing the maximum security for the confidential patient's information.

During the COVID-19 pandemic, the whole world realized the importance of the healthcare systems and remote monitoring. The innovation of such systems would economize a huge amount of money, efforts and most importantly save lives. These are the main reasons why researchers are now interested in the health system digitalization or what we call the e-health. This project comes as a proposed solution for enhancing the remote monitoring, hence the e-health system while protecting the user's privacy through using the same ECG signal used for monitoring as a biometric. In other words, this project implements an ECG biometric based system with remote monitoring. It consists of three major parts: a hardware system for collecting the ECG signals, a software part for processing the collected signal and identifying the person, and a phone application to introduce the whole system to the user. The report is organized as follows:

- Chapter 1: Comprises a background on biometric modalities in general and ECG in particular. In addition, it briefly states the literature and research done before.
- Chapter 2: Describes the methods used in data collection, denoising, segmentation, feature extraction and template matching.
- Chapter 3: Contains the detailed implementation of both the template matching and the Android application.
- Chapter 4: Entails the obtained results as well as a discussion.

Finally, the report is ended with a conclusion and future suggestions to improve and extend our system.

Chapter 1

Background and Related Work

Introduction

Biometric based systems provide a way to identify individuals from their unique characteristics. Thus, such systems are increasingly being adopted to improve security, privacy and provide different applications in various fields. Some of the most significant biometric identifiers are: the face, fingerprints, and iris. However, the technological advancement made such identifiers vulnerable to forgery and security attacks [33]. To avoid such threats, continuous biometric authentication was proposed. It does not only offer biometric authentication, but it also presents a proof of vitality of the authenticated person. Hence, ECG has gained a lot of attention as a promising biometric technique due to its uniqueness and difficulty to counterfeit. In addition to security applications, this bio-signal can reveal a lot of information about the person's physical and emotional health state. Therefore, it is also used in monitoring systems. In this chapter, we will discover more about this interesting signal as well as the research that has been done on the ECG biometric based systems.

1.1 Biometrics

1.1.1 Definition

The term came from the two Greek words, "bio" meaning life and "metrics" meaning to measure. Thus, biometrics literally means "to measure life" [26].
Biometrics are known to be the body measurements and calculations related to human characteristics [45]. They represent the most suitable means of identifying and authenticating individuals especially for security purposes.

1.1.2 Biometric identification and authentication

Biometric authentication

It is the process of comparing data of a person's characteristics to the same person's biometric "template" and determine resemblance [9]. It consists of two main steps:

- Storing the reference model
- After that, the stored model is compared to the person's biometric data to be authenticated.

Biometric Identification

It consists of determining the identity of a person by capturing an item of biometric data from this person and comparing it to the several people's biometric data stored in a database [9].

1.1.3 Biometric Identifiers

1.1.3.1 Behavioral

Behavioral Biometrics is the field of study related to the measure of uniquely identifying and measurable patterns in human activities [35]. It identifies patterns in the ways that particular bodies perform particular tasks—patterns in walking, speaking, typing, or even touchscreen and mouse behavior. These patterns are prohibitively difficult to capture and replicate, and they evolve over time [31].

1.1.3.2 Physiological

1. **Biological**: obtained by making biological analyses of: blood, DNA, saliva...etc
2. **Morphological**: face id, fingerprints, the eye's iris and ECG

As can be seen in figure 1.1, ECG has a really high efficiency. It is completely unique and differs from one human being to another, even between identical twins. In addition to its uniqueness, ECG signal provides the real-time vitality feedback (i.e: ensures that the biometric sample is being collected from an alive individual), this represents a huge advantage from a security point of view. Furthermore, ECG is often used by physicians in cardiovascular diseases diagnosis. Through the deployment of ECG-enabled biometric system the identity of the persons can be verified online during ECG monitoring or offline through their medical records.

Moreover, it can be used in mobile health monitoring systems by providing a high protection of the person's identity and condition. This private health information is strictly protected under the privacy regulation standardization of the Health Insurance Portability and Accountability Act (HIPAA) of 1996 [50].

Biometric Modalities	Univers-ality	Unique-ness	Perma-nence	Collect-ability	Perfor-mance	Accep-tance
Fingerprint	Medium	High	High	Medium	High	Medium
Retina	High	High	Medium	Low	High	Low
Hand Geometry	Medium	Medium	Medium	High	Medium	Medium
Palm print	Medium	High	High	Medium	High	Medium
Voice	Medium	Low	Low	High	Low	Medium
Face	High	Low	Medium	High	Low	High
Signature	Low	Low	Low	Medium	Low	High
DNA	High	High	High	Low	High	Low
ECG	High	High	High	Medium	High	Medium

Figure 1.1: Comparative study of different biometric modalities [37]

1.2 ECG Signal

1.2.1 Definition

ECG is a test that detects and records the strength and timing of the heart's electrical activity. This information is recorded on a graph that shows each phase of the electrical signal as it travels through the heart. It is obtained by placing electrodes on the skin which detect the small electrical changes that are a consequence of cardiac muscle depolarization followed by repolarization during each cardiac cycle (heartbeat) [47].

1.2.2 ECG Waves

Before digging into the different components of an ECG signal, we would like to introduce briefly the heart anatomy for a better understanding of its electrical activity.

The Heart

As shown in figure [1.2] the human heart has four chambers. The two upper ones are called the left and right atria whereas the lower chambers are called the left and right ventricles. A wall of muscle called the septum separates the left and right atria and the left and right ventricles.

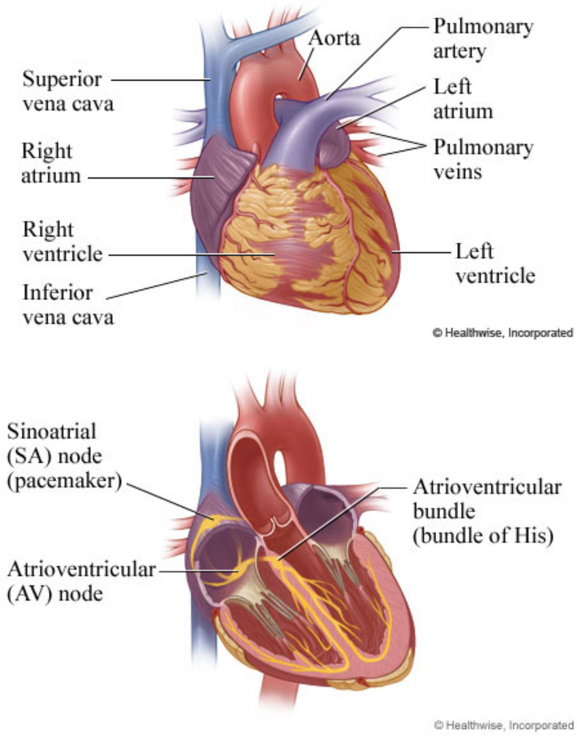


Figure 1.2: Anatomy of the human heart [28]

The Heart's Electrical Activity

The electrical signal travels through the network of conducting cell "pathways," which stimulates your upper chambers (atria) and lower chambers (ventricles) to contract. The signal is able to travel along these pathways by means of a complex reaction that allows each cell to activate one next to it, stimulating it to "pass along" the electrical signal in an orderly manner. As cell after cell rapidly transmits the electrical charge, the entire heart contracts in one coordinated motion, creating a heartbeat.

The electrical signal starts in a group of cells at the top of the heart called the sinoatrial (SA) node. The signal then travels down through the heart, triggering first the two atria and then the

two ventricles. In a healthy heart, the signal travels very quickly through the heart, allowing the chambers to contract in a smooth, orderly fashion [28].
The heartbeat happens as follows:

1. The SA node (called the pacemaker of the heart) sends out an electrical impulse
2. The upper heart chambers (atria) contract
3. The AV node sends an impulse into the ventricles
4. The lower heart chambers (ventricles) contract or pump
5. The SA node sends another signal to the atria to contract, which starts the cycle over again

This cycle of an electrical signal followed by a contraction is one heartbeat.

ECG Waves

The ECG signal consists of three main components (figure 1.3):

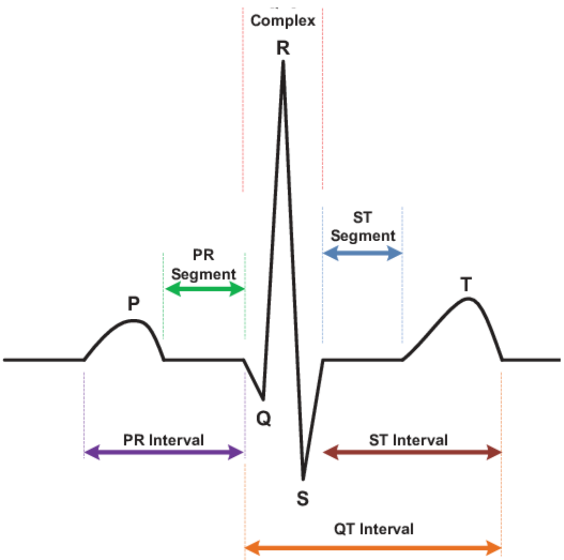


Figure 1.3: The ECG signal waves [47]

- **The P wave:** representing the depolarization of the artia
- **The QRS complex:** three closely related waves corresponding to the depolarization of the ventricles

- **The T wave:** representing the repolarization of the ventricles

We may also mention:

- **The PR interval:** time taken for electrical activity to move between atria and ventricles
- **ST segment:** isoelectric time representing the time between depolarization and repolization of the ventricles

The relation between each of the ECG waves and the excitation process of the heart is depicted in figure [1.4]

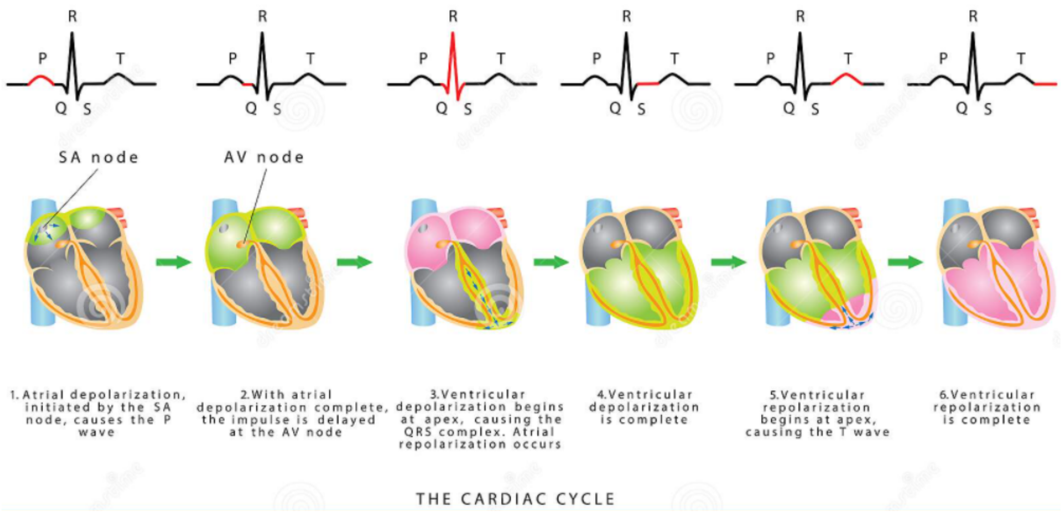


Figure 1.4: The cardiac cycle

1.3 Related Work

Recently, a lot of research has been done on the ECG biometric based systems to obtain high accuracy. However, the performance of such systems is affected by many factors: the filtering type, segmentation method, type of the extracted features, the matching algorithm and even the health status of the subjects. The particularity of each research lies in the method used in each of the mentioned factors. Here are some of the most used methods:

1.3.1 Methods Used In Filtering

Different filters were applied on ECG signals to remove intra-class variation. We may cite:

- **Bandpass filters:** most studies in ECG biometrics have employed the bandpass filter (BF) by cascading a low-pass (LP) and high-pass (HP) filters with different cutoff frequencies [18].
- **Finite impulse response FIR** it works well for the attenuation of the known frequency bands,such as the noise coming from the electrical network (50 Hz or 60 Hz), since they allow quick and easy application of the reject-band-filter. The problem with this approach is that the frequency of the noise is not always known [14].
- **Infinite Impulse Response:** IIR filters are generally chosen for applications where linear phase is not too important and memory is limited. They have been widely deployed biomedical sensor's signal processing. The main drawback of this approach is the stability of the system.
- **Wavelet transform:** many methods based on wavelet transforms have been employed to remove noise, since they preserve ECG signal properties avoiding loss of its important physiological details and are simple from a computational point of view.

1.3.2 Methods Used In Segmentation

Different techniques were used to segment continuous ECG signals we may mention:

- **Fixed length window:** it is a common method consisting of capturing one heart or one cycle of ECG by using a fixed-length window.
- **RR interval:** first, it starts with localising the R-peak using any R peak detection algorithm, the most used one is *Pan-Tompkins* technique. Then, a fixed window is considered by taking the identified R-peak as a reference to segment the ECG signal in terms of the *R-R* interval (RR) [18].
- **QRS complex detection:** This method lies essentially in detecting all the QRS complex to segment the ECG into individual beats [38].

1.3.3 Methods Used in Feature Extraction

ECG biometric feature extraction can be categorized as a *fiducial point* or *non-fiducial point*:

- **Fiducial points feature extraction:** Algorithms based on fiducial features use the characteristic local features of ECG beats such as temporal or amplitude difference between consecutive fiducial points, and peak (minimum or maximum), extracted from single ECG beat or segment. For example, the P, Q, R, S, and T peak wave, the time difference between the peaks of the Q and T waves, and the QT interval are considered as fiducial features. Fiducial methods rely on the accurate detection of the main ECG characteristic points such as P, Q, R, S, and T waves. Several subsets of these fiducial features have been used in the literature .

- **Non fiducial points feature extraction:** Non-fiducial point feature extraction holistically analyzes an ECG, typically by applying time or frequency analysis to obtain other statistical features. This method aims to extract discriminative information from the ECG waveform without having to localize fiducial points [18].

1.3.4 Methods Used In Matching

The signals are matched with reference models stored in a database using the following methods:

- **Softmax:** It is used as the last activation function of a neural network to normalize the output of a network to a probability distribution over predicted output classes (each class represents a person).
- **SVM:** support-vector machine, is a supervised learning model with associated learning algorithm that analyzes data for classification and regression analysis.
- **Euclidean distance** employed by calculating the euclidean distances between the templates, the smaller the distance the higher the probability of the match.
- **Hamming distance** proceeds by comparing the resulting hamming distance with a threshold.

Table 1.1, adapted from the comparative study [14], summarizes all the above methods.

Step	Methods	Work References
Filtering	BP	[3] [19] [20] [4]
	FIR	[18] [2]
	IIR	[17] [40] [21]
	Wavelet	[12]
Segmentation	Fixed length window	[17] [40] [21] [3]
	RR interval	[22] [27]
	QRS complex detection	[23] [39]
Feature extraction	Fiducial points	[22] [27] [36]
	Non fiducial points	[15]
Matching	Softmax	[13] [6]
	SVM	[3] [20] [22] [27] [12]
	Euclidean distance	[30] [44] [51]
	Hamming distance	[17] [16] [11]

Table 1.1: Summary of the state of the art approaches for ECG biometrics (adapted from [14])

Conclusion

In this chapter, we have seen all the necessary information needed to understand the rest of the report. We started by stating the different existing biometric identifiers. Moving to the advantages of using ECG as a biometric over the rest of the biometric modalities. We have also seen, the ECG signal, how it is generated, and its different waves. Moreover, we have stated the most relevent research that has been done in this field. In the second chapter, we will introduce our proposed ECG biometric based system.

Chapter 2

Proposed Solution

Introduction

An ECG biometric based system is characterized by four stages. It starts with the data collection phase, where data is collected from different subjects. The second stage is pre-processing the collected data in order to eliminate noise. The next one, is feature extraction by processing the data until getting the characteristics of all subjects. Finally, the obtained information is used in the identification phase. In this stage, an unknown ECG is presented to the system to perform the previous mentioned phases before a matching algorithm assigns the extracted features to a best matching subject's data stored in the database. Therefore, each of the previous stages is of a huge importance in order to achieve good performance. In this chapter we will present our proposed system stage per stage.

2.1 General Block Diagram

First, we started by the data collection which consists of two datasets: one was collected using our hardware system and the other one is the MIT_BIH Arrhythmia database. After that, the collected data was denoised in order to remove the different noises that might affect the latter steps. Then, the denoised ECG recordings were segmented into heart beats by detecting the R peaks. After the cleaning phase, the data was processed to extract the necessary features for our biometric system. Using all the obtained information, a template matching algorithm was implemented in order to identify the user. Finally, a user friendly Android application was developed to visualize the results. It offers two features: ECG monitoring and heart rate calculation. Each of the previous steps will be detailed in the following sections.

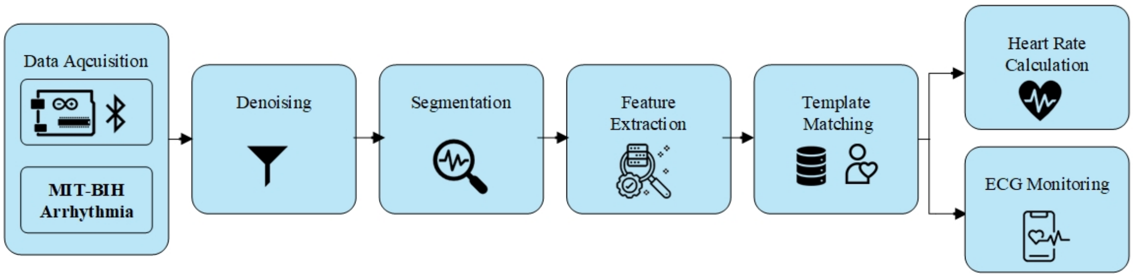


Figure 2.1: General block diagram of the proposed system

2.2 Data Acquisition

All the work present in this report was trained and tested using two databases: the MIT-BIH Arrhythmia to cover normal and abnormal ECG recordings and our own collected dataset to familiarize our programs with the sensor's data.

2.2.1 MIT BIH Arrhythmia Database

The MIT-BIH Arrhythmia Database [1] contains a set of over 4000 long-term Holter recordings (i.e: a type of ambulatory electrocardiography device). These recordings were obtained by the Beth Israel Hospital Arrhythmia Laboratory between 1975 and 1979.

The database contains 48 records taken from 25 men aged 32 to 89 years, and 22 women aged 23 to 89 years (Records 201 and 202 came from the same male subject). Each of the recordings is slightly about 30 minutes long. They are divided into two groups [29]:

- The first one contains 23 records (record 100 to 124 with a missing number). It serves as a representative sample of the variety of waveforms and artifact that an arrhythmia detector might encounter in routine clinical use
- The second one contains 25 records (record 200 to 234 with some missing numbers). It includes a variety of rare but clinically important phenomena (complex ventricular, junctional, and supraventricular arrhythmias and conduction abnormalities).

The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10-mV range. Two or more cardiologists independently annotated each record; disagreements

[1] The database can be found at: `https://physionet.org/content/mitdb/1.0.0/`

were resolved to obtain the computer-readable reference annotations for each beat (approximately 110,000 annotations in all) included with the database [29].

2.2.2 The Collected Data

The second dataset, contains eight recordings collected using our hardware system. These records, were taken from 6 women and 2 men aged 19 to 92 years. Each of the recordings is approximately about 10 minutes long.

Figure 2.2 represents the general schematic of our circuit used to collect the previous data. It consists of the following components:

- **A microcontroller: Arduino Uno** It is a microcontroller board based on the ATMEGA328P. It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz ceramic resonator (CSTCE16M0V53-R0), a USB connection, a power jack, an ICSP header and a reset button. In other words, it contains everything needed to support the microcontroller
- **The ECG Sensor: AD8232** [2] is an integrated signal conditioning block for ECG and other biopotential measurement applications. It is designed to extract, amplify, and filter small biopotential signals in the presence of noisy conditions, such as those created by motion or remote electrode placement. This design allows for an ultra-low power analog-to-digital converter (ADC) or an embedded microcontroller to acquire the output signal easily.
- **The Bluetooth Module: HC_05 (ZS_040)** [3] is an easy to use Bluetooth SPP (Serial Port Protocol) module, designed for transparent wireless seria connection setup. Its makes an easy way to interface with controller or PC. HC-05 Bluetooth module provides switching mode between master and slave mode.

We start by placing the electrodes on the chest. After that, the microcontroller starts reading the recorded values from the AD8232. These data, are sent via bluetooth to an edge device (computer in our case) in order to process them. The described steps are illustrated in figure 2.3 and through the flowchart 2.4 which represents both the setup and loop functions of the arduino program.

[2] The datasheet can be found at: `https://www.analog.com/media/en/technical-documentation/data-sheets/ad8232.pdf`

[3] The HC_05 user manual can be found at: `https://www.gme.cz/data/attachments/dsh.772-148.1.pdf`

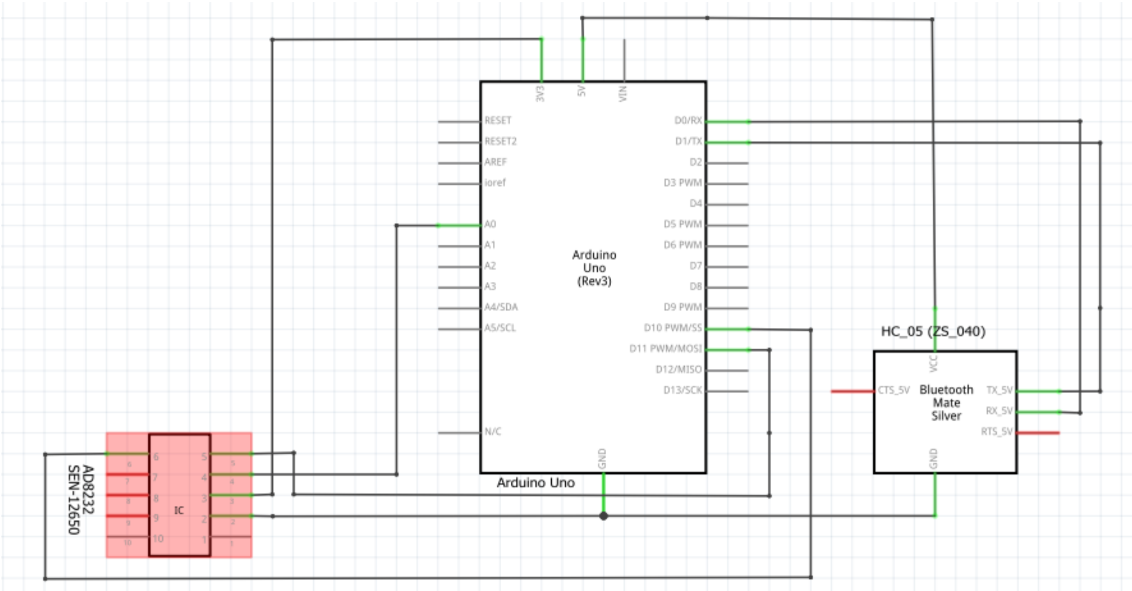


Figure 2.2: General schematic of the implemented system

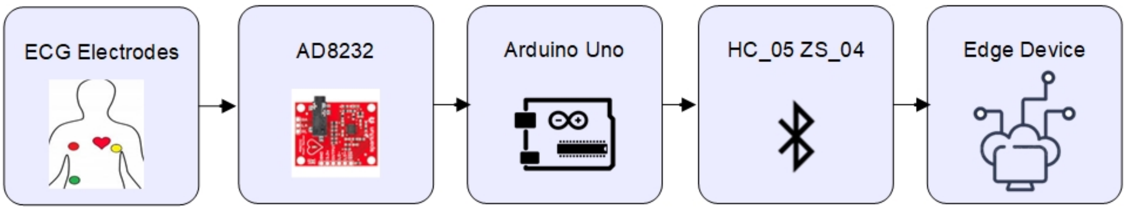


Figure 2.3: General block diagram of the implemented circuit

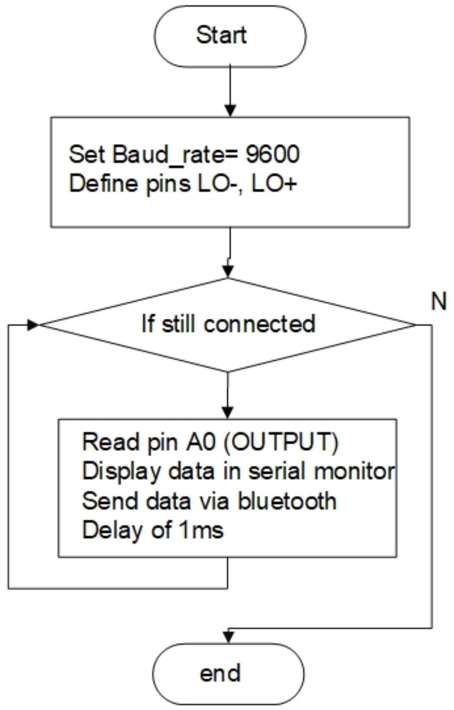


Figure 2.4: Arduino program flowchart

2.3 Denoising

Denoising is a required step when it comes to ECG signals. It removes noise which interfere with the ECG components and affect the quality of the analysis. There exist various types of noises, but the most significant ones are [10]:

- **Baseline wandering (BW):** It is a slow varying artifact that considerably affects the ST segment of the ECG signal. The BW noise spectrum is localized in low frequencies, usually less than 0.5HZ but it can be higher due to stress and nervousness. It is mainly due to patient respiration, more precisely, the spatial change caused by lung movement. It can be also generated from motion and change in the electrode-skin impedance.

- **Powerline Interference (PLI):** The PLI component and its harmonics are known to be the most disturbing noise sources in bio-potential recordings. It makes the boundary regions of P-waves and T-waves unidentifiable, its frequency spectrum is quite narrow and centered around 50HZ /60HZ (depending on the country). This noise comes from several sources, among which:

- The power line electromagnetic interference
- Electromagnetic field (EMF) by the machinery which is placed nearby
- Electrical equipment such as air conditioner, elevators and X-ray units draw heavy power line current, which induce 50 Hz / 60HZ signals in the input circuits of the ECG machine

- **Muscle artifacts:** Known as Electromyography (EMG) noise, it occurs at the time of muscle activity during an ECG recording especially in stress or nervousness tests. It consists of a maximum frequency of 10 KHz
- **Patient–Electrode Motion Artifacts:** The movement of the electrodes away from the contact area on the skin leads to variations in the impedance between the electrode and skin causing potential variations in the ECG.
  We may also state other noise sources such as the instrumentation noise produced by the equipment composing the recording section (probes, cables, analog to digital converter, etc.). Obviously, these types of interference could be significantly reduced by a careful choice of high-quality devices.

There are many techniques that can be used to denoise the ECG signal, but much attention has been placed recently on the wavelet transform as it is much more effective than the traditional filtering methods. The Hard and soft threshold method are the most commonly used wavelet threshold denoising techniques [34].
The classical Discrete Wavelet Transform DWT suffers from a major drawback: it is not a time-invariant transform, which is a valuable property for signal analysis [42]. This means that DWT brings out a transform shift. In contrast, the stationary wavelet transformation (SWT) is proposed on the foundation of orthogonal wavelet transformation, which brings the properties of shift and scale invariance.

2.3.1 The Wavelet Transform

Wavelets are finite support duration waveforms that have an average of zero. The wavelet analysis decomposes signals using an orthonormal family of basis functions. This analysis is suited for transient, time-varying signals, thus well suited for the ECG signals in our case. The wavelet transform, is computed by convolving the signal under processing and the wavelet function [36].

There are several wavelet families that can be used for denoising and filtering purposes, among which we may mention: Daubechies, biorthogonal, Haar, and Symlet.
The mother wavelet is chosen from the previous families and it must satisfy the admissibility condition illustrated by the equation 2.1. It implies that the Fourier transform of Ψ(t) vanishes at the zero frequency. In other words, wavelets must have a bandpass-like spectrum. The property also means that the average value of the wavelet in the time domain must be zero.

$$C_{\Psi} = 2\pi \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega \tag{2.1}$$

Where $\Psi(\omega)$ is the Fourier transform of Ψ(t).
The wavelet transform of a signal x(t) is then defined by:

$$W(b, a) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{a}} \Psi^{*}(\frac{t-b}{a}) dt \tag{2.2}$$

Where a is the scale parameter, b is the shift parameter, and Ψ*(t) is the complex conjugate of the mother wavelet Ψ(t).
If we take the discrete value of the scale and the shift parameter respectively: $a = 2^j$, $b = 2^j$k. The discrete wavelet transforms (DWT) of x(t) can be defined as:

$$W(k, j) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{2^j}} \Psi^{*}(2^{-j}t - k) dt \tag{2.3}$$

The DWT is implemented by combining high-pass filters H_j and low-pass filters L_j as shown in figure 2.5.

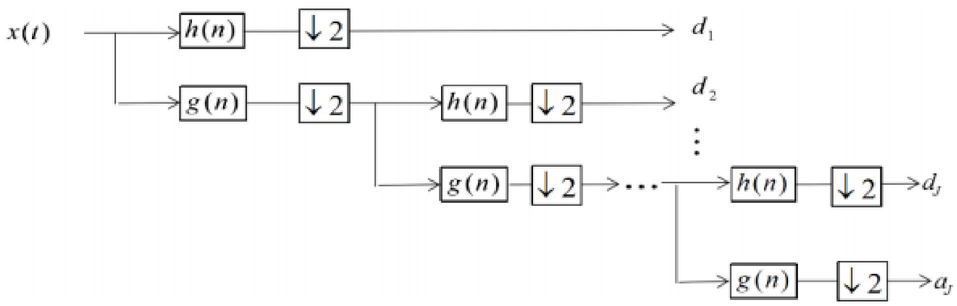


Figure 2.5: Block diagram of multi-level DWT of x(t) [49]

Figure 2.5 shows a decomposition of wavelet transform implementation of a signal given by $x(t)$.

It consists of a high pass filter $h(n)$ and a low pass filter $g(n)$ which is basically a mirror version of $h(n)$. The down sampled outputs help in providing the detail coefficients d_n and the approximation a_n.

As explained earlier, the DWT is not a time-invariant transform; whereas the SWT makes such a transform possible by up-sampling the filter coefficients of the high-pass filters and the low-pass filters.

The wavelet coefficients are given by the sequences d_j and the scaling coefficient is given by the sequence a_j . Where j represents the order of SWT. This process is illustrated in figure 2.6 where filters in each level are up-sampled versions of $h(n)$ and $g(n)$.

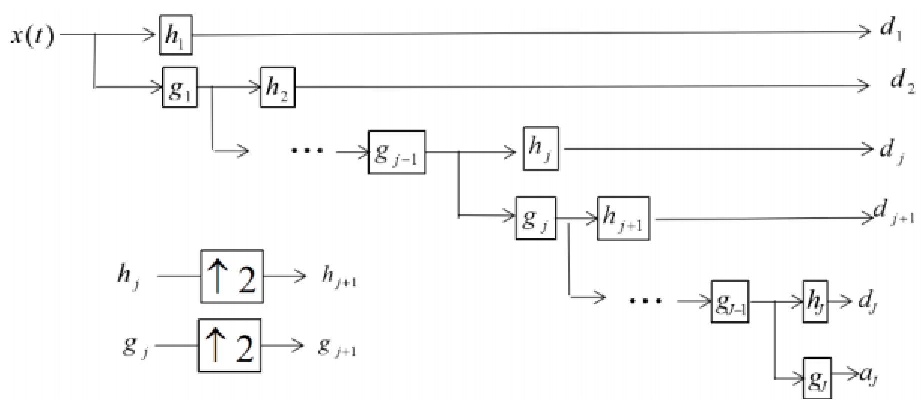


Figure 2.6: Block diagram of multi-level SWT of x(t) [49]

2.3.1.0.1 Threshold Function

Hard and soft threshold functions are the most commonly used approaches; they have their own characteristics and limitations. Hard threshold function hurts the signal continuity, which brings out the oscillation and poor smoothness of the reconstructed signal. Whereas, soft thresholding is a method implemented by first setting to zero coefficients whose absolute values are lower than the threshold λ then shrinking the nonzero coefficients toward zero.

2.3.2 The Denoising Procedure

The SWT based method provides effective signal denoising with minimum computational complexity. It consists of the following steps:

1. Input the signal to be denoised
2. Select a wavelet and determine the decomposition level J for denoising
3. Take the J level discrete stationary wavelet transform (SWT) of the signal, referred to as detail components of each level and the Jth approximation component
4. Select a threshold for each detail component and apply the proposed threshold function to the detail components
5. Take the inverse stationary wavelet transform (ISWT)

The denoising specifications applied on the MIT-BIH database are:

- Mother Wavelet: Daubechies 1 [4]
- Level of Decomposition: N=4
- Threshold Function: Soft
- Selected Threshold Value: $\lambda = 7.574$

As an example, figures 2.7 and 2.8 depict the denoising process of the MIT_BIH's record 100.

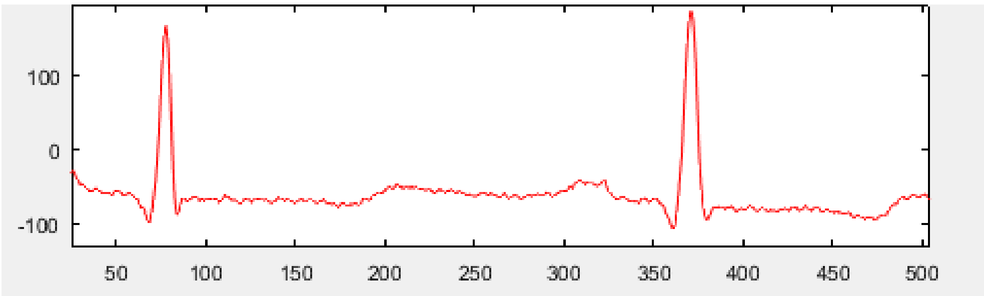


Figure 2.7: Record 100 MIT_BIH before denoising

[4]Daubechies wavelets extend the Haar wavelets by using longer filters, that produce smoother scaling functions and wavelets.

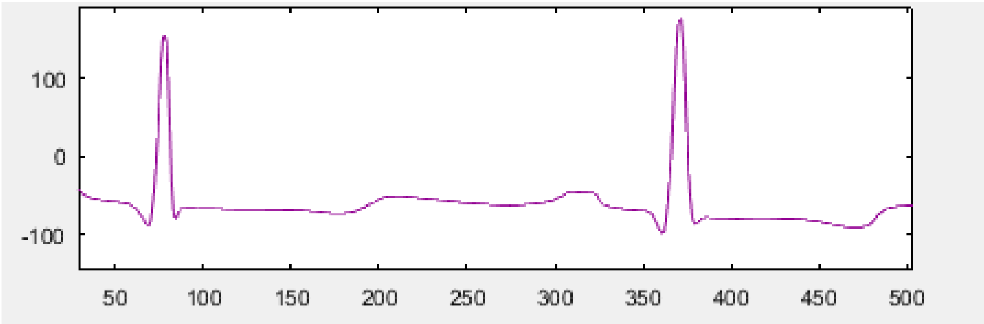


Figure 2.8: Record 100 MIT_BIH after denoising

The denoising specifications for the collected data

Since the collected recordings were 10min long, a signal extension was necessary to satisfy the compatibility between the level of decomposition and the length of the signal. For this purpose, the zero padding technique was used; it refers to adding zeros to the end of a time domain signal to increase its length.

2.4 Segmentation

The main objective of the segmentation process is to divide the recordings into individual beats, by detecting the R peak locations. This, would allow us to further process the ECG signals and extract the necessary features. We have used two different methods, one for each database:

2.4.1 MIT_BIH Arrhythmia

A DWT based algorithm [5] was implemented using the Symlet4 wavelet with N=4. As can be seen in figure 2.9, the Symlet4 was chosen due to its resemblance with the QRS complex.
The aim of this algorithm is to preserve R peaks while eliminating the low frequency bands (i.e., p-wave and T-wave frequencies). This implies that a band pass filtering action is needed.
First, we applied the undecimated wavelet transform which separates the signal components into different frequency bands. After that, the approximation coefficient a_4 was eliminated because it carries all the low frequency details of the ECG signal. Similarly, the detail coefficients d_1 and

[5]This detection can be done using a normal frequency selective bandpass filter, but the wavelet transform gives much better results and higher accuracy since the QRS complex is not always sharp.

d_2 were not considered. After considering only d3 and d4 to achieve the band pass filtering, we constructed a signal using the IWT where the R peaks are well presented. Finally, a peak detection technique was used by taking the maximum peak in a window. As an output, the obtained R peak locations were stored in an array for each record. Figure 2.10 shows an example of the detected R peaks on record 100.

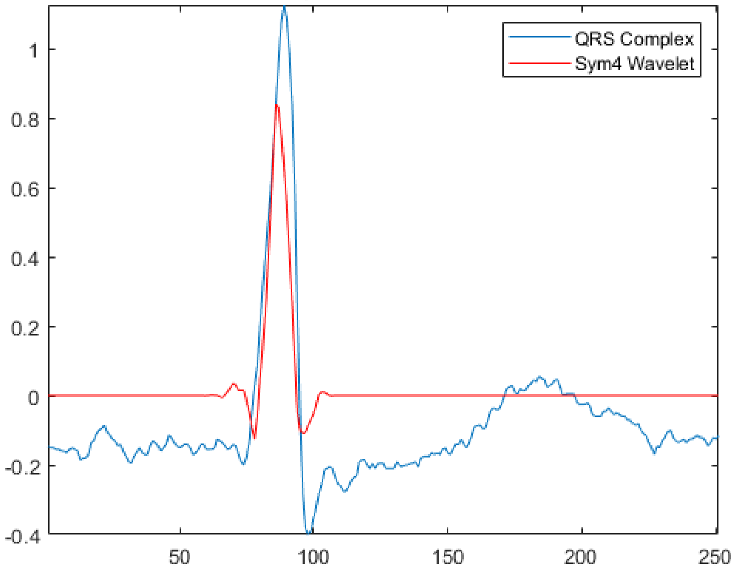


Figure 2.9: Symlet 4 wavelet vs. the QRS complex

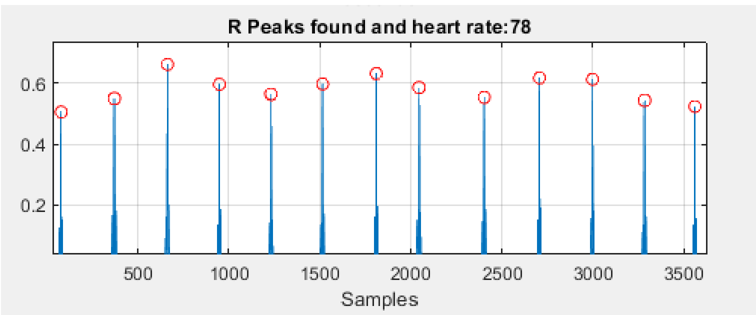


Figure 2.10: Detected R peaks MIT BIH Arrhythmia record 100

2.4.2 The Collected Data

For the denoised collected ECG signals, we utilized manual segmentation using a sliding window with fixed conditions. First, we started by setting the window size (135 samples). The idea is about sliding this window and detecting the peak each time. However, some restrictions were added to avoid some types of errors:

- Comparing the indexes of the currently detected peak and the previous one, to avoid detecting the same R peak twice
- Comparing the amplitude of the detected peak with a certain threshold to avoid detecting a noise as an R peak

Figure 2.11 shows the detected R peaks for record 308.

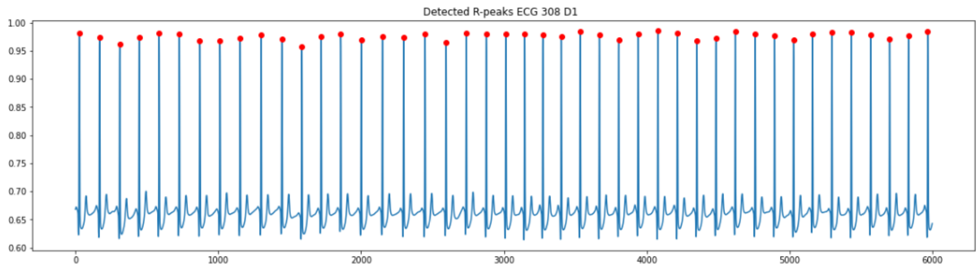


Figure 2.11: Detected R peaks record 308 (collected data)

2.5 Feature Extraction

After denoising and segmentation, the data is now ready to be processed to extract the needed features for our algorithm. Mainly three features were generated to form what we call a *template* for each record. Each template contains the following information:

1. **R peaks magnitudes:** nine successive R peak magnitudes
2. **RR distances:** the interval between two successive R waves. Nine RR intervals were extracted from each record in our datasets. They consist of $RR_{previous}$ and $RR_{following}$:
   - $RR_{previous}$: $RR[i-1]$ the interval between the i^{th} R wave and the previous R wave.
   - $RR_{following}$: $RR[i+1]$ the interval between the i^{th} R wave and the next R wave.

3. **An average beat of that record:** we took fifteen R peak to calculate the morphological points by summing each of the identical points of the 15 beats, then averaging them. The process is explained in flowchart 2.12. This results in samples representing a typical beat of that person.

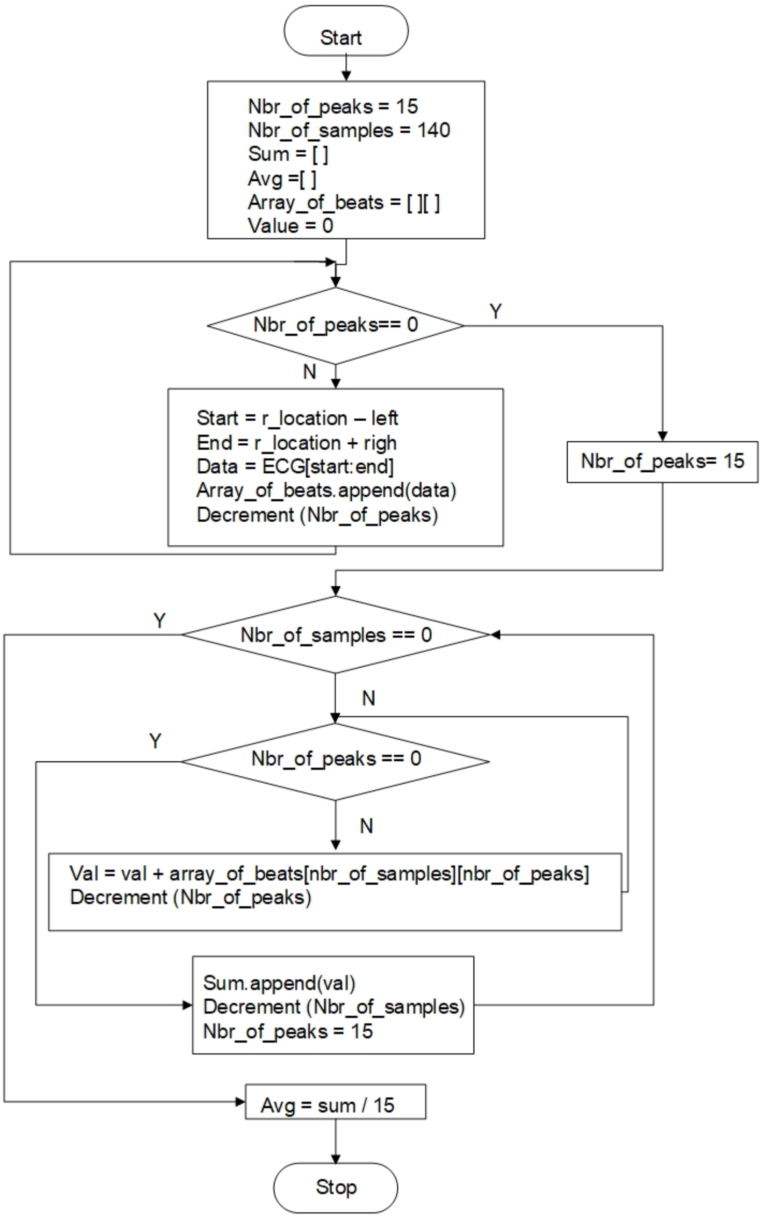


Figure 2.12: Average beat calculation flowchart

2.6 Template Matching

Each of the previously generated templates were stored in a database in order to train and test our matching algorithm. The main idea is that each time we want to identify a person, we generate a template and compare it with the already stored ones in the database. If it achieves a certain threshold, a possible match is considered. This is called *"template matching"* and will be seen in details in the next chapter.

Conclusion

In this chapter we have seen the overall structure of our system, starting by the used databases. Moving to cleaning the data by eliminating noise, processing it to extract the needed features and generating *templates* for each record. Finally, matching ECGs in order to properly identify people. In the coming chapter, we will be seeing the operation of our matching algorithm. In addition to the implementation of an Android application to introduce the whole system to the user.

Chapter 3

Implementation

Introduction

In the previous chapter, we have seen the most important stages of any ECG biometric based system. These stages can be further divided into two main categories: the preprocessing and template matching stages. The preprocessing phase prepares the data from raw data to perfectly defined templates ready to use, by passing through denoising, segmenting, and feature extraction. Whereas, the template matching phase takes an input signal, generates a template of that signal using the same previous process, and finally compares the generated template with the ones stored in the database. In this chapter, we will see the detailed implementation of this template matching, how it is done, and how can we deduce whether there is a match or not. Moreover, obtaining a result and identifying the persons will not be enough. In order to employ this ECG identification feature in a useful situation, we thought of adding an android application. This user friendly application, will complete our software/ hardware system. It will serve as a guide to our hardware system and as a visualization tool of the results. Besides, it will serve as a link between the doctor and her/his patients.

3.1 Tools and Technologies

In this section each of the tools and technologies used to realize our project will be presented briefly.

3.1.1 Development Tools

- **Google Colaboratory:** Colaboratory, or "Colab" for short, is a product from Google Research. It allows writing and executing arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs [7].

- **Processing:** Processing is a free graphical library and integrated development environment (IDE) built for the electronic arts, new media art, and visual design communities. The Processing Environment includes a text editor, a compiler, and a display window. The Processing IDE is similar to Arduino in terms of structure. It has setup functions and draw functions like an Arduino has a setup and loop function. It can establish connection with the Arduino through serial communication. This way, data can be exchanged between the Arduino and the Processing IDE [32].
- **Visual Studio Code:** Visual Studio Code is a free coding editor.It supports several programming languages, including Python, Java, C++, JavaScript, and more [25].
- **Flutter:** Flutter is an open-source UI framework created by Google and released in May 2017. It is used to develop cross platform applications for Android, iOS, Linux, Mac, Windows, Google Fuchsia, and the web from a single codebase. Flutter consists of two important parts [8]:
  - *An SDK (Software Development Kit):* A collection of tools to help developing applications. This includes tools to compile code into native machine code (code for iOS and Android).
  - *A Framework (UI Library based on widgets):* A collection of reusable UI elements (buttons, text inputs, sliders, and other elements) that can be personalized if needed.
- **MySQL Workbench:** MySQL Workbench is a visual database design tool that integrates SQL development, administration, database design, creation and maintenance into a single integrated development environment for the MySQL database system.

3.1.2 Programming Languages

- **Python:** Python is an interpreted high-level general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Its language constructs as well as its object-oriented approach aim to help writing clear, logical code for small and large-scale projects [48].
- **Matlab:** An abbreviation of *"matrix laboratory"*, *MATLAB* is a proprietary multi-paradigm programming language and numeric computing environment developed by MathWorks. It allows matrix manipulations, plotting of functions and data, implementation of algorithms,

creation of user interfaces, and interfacing with programs written in other languages [24].

- **Dart:** Dart is a programming language optimized for applications on multiple platforms. It is developed by Google and is used to create mobile, desktop, server and web applications. It is an object-oriented, class-based, memory-recovery language with C-like syntax. Dart can be compiled as native code or as JavaScript [46].

3.2 Template Matching

Before digging into the implementation of the template matching stage described in section 2.6, we would like to introduce a statistical measure *"The Correlation Factor"* since it represents the basis of our comparison.

3.2.1 Correlation Factor

Correlation coefficients are used to measure how strong a relationship is between two variables. The values range between -1.0 and 1.0. A correlation of -1.0 shows a perfect negative correlation, while a correlation of 1.0 shows a perfect positive correlation. A correlation of 0.0 shows no linear relationship between the movement of the two variables.
There are several types of correlation coefficients, but the most common one is the *Pearson correlation* (Γ). It measures the strength and direction of the linear relationship between two variables. The formula of this coefficient is given by equation 3.1.

$$\Gamma_{xy} = \frac{\sum_{n=1}^{N}(x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_{n=1}^{N}(x_n - \bar{x})^2 \sum_{n=1}^{N}(y_n - \bar{y})^2}} \tag{3.1}$$

Where Γ_{xy} is the correlation coefficient between the input template y_n and the stored templates x_n, n denotes the number of samples, and N is the length of the template.

3.2.2 Template Matching Algorithm

Template matching represents the last step of the identification process. We start by introducing an ECG signal to our template generator (the signal will be denoised, segmented into individual beats; then, the features seen earlier ,in section 2.5, will be extracted to generate a template). After that, the generated template will be compared with the templates stored in our database based on the correlation coefficient.
If the highest correlation coefficient is greater than a certain threshold ($threshold = 0.8465$), a

possible match between the template and the person to be identified is considered. In other words, the person associated with the input signal with the highest correlation coefficient is selected as a match if and only if it exceeds the specified threshold. The whole process is illustrated in figure 3.1.

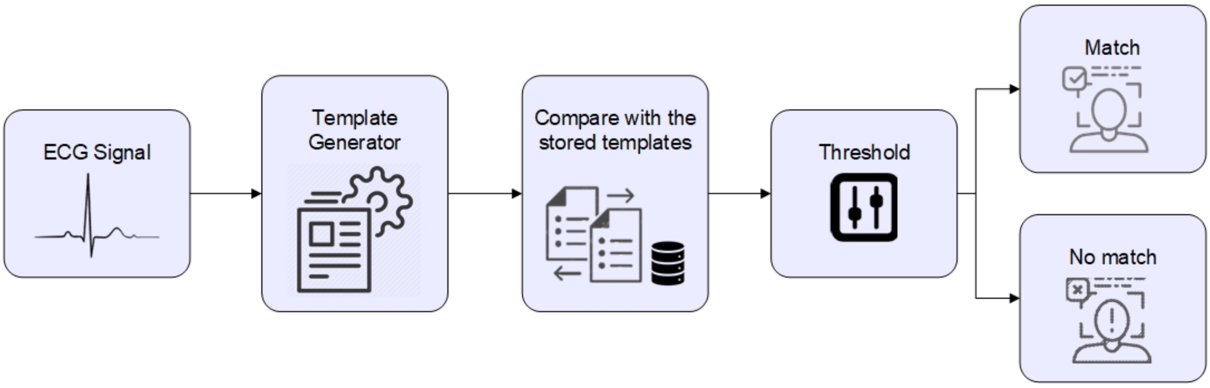


Figure 3.1: Template Matching Process

3.3 Android Application Implementation

Our android application was implemented using Visual Studio Code, Flutter, Dart, and Firebase. In this section, we will see the overall design of our application in addition to its different possible scenarios.

3.3.1 Design Requirements

To implement an application, a mandatory step is to start with having a clear idea about the external entities that will interact with the system, the possible interactions, and their order of occurrence. For this purpose, *Unified Modeling Language* ”UML” is used by software engineers to model the system before starting the actual implementation.

3.3.1.1 Unified Modeling Language

UML is a standardized modeling language consisting of an integrated set of diagrams. It is a very important part of developing object oriented software and the software development process in general. The UML uses mostly graphical notations to express the design of the projects. It makes use of elements and forms linked together to form diagrams which can be classified into two types [1]:

- **Structural Diagrams:** Capture static structure of the system. Structural Diagrams include:
  - Component diagram
  - Object diagram
  - Deployment diagram.
- **Behavior Diagrams:** Capture dynamic aspects of the system. Behavior diagrams include:
  - Use case diagram
  - State diagram
  - Activity diagram
  - Interaction diagram.

Figure 3.2 illustrates the hierarchy of diagrams in UML2.2.

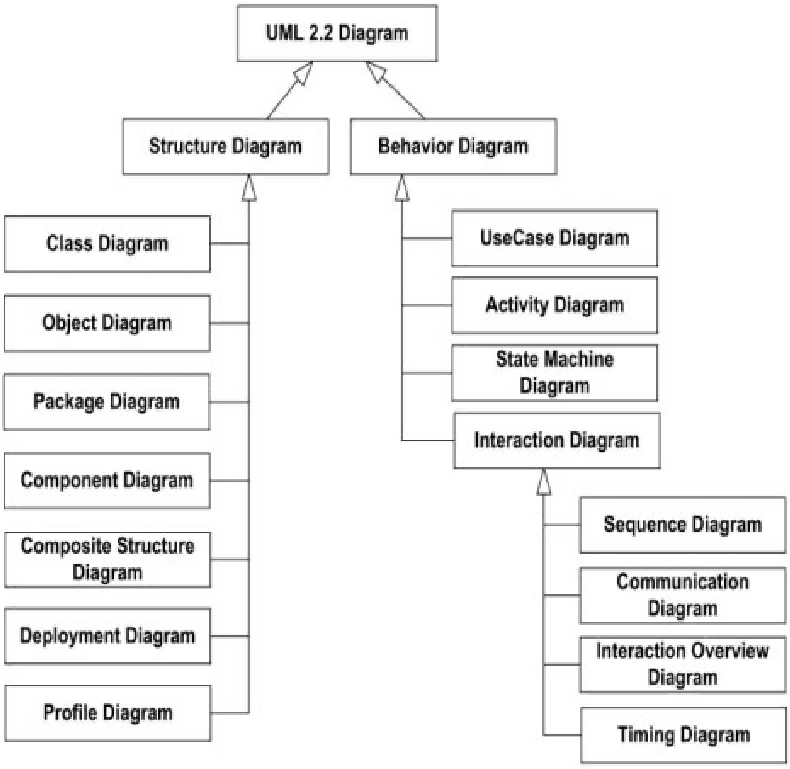


Figure 3.2: UML2.2 diagrams structure

3.3.2 Modeling

3.3.2.1 Use Case Diagram

Definition

A use case diagram is a graphical depiction of an actor's possible interactions with a system. It shows the several use cases and types of users the system has.

Use Case Diagram Components

Any use case diagram has three components which are:

- **Use cases:** a set of actions, services and functions that the system needs to perform. They are represented using labeled circles.
- **Actors:** individuals involved with the system defined according to their roles. An actor can be a human or another external system.

- **Relationships:** we used four types of relationships:
  - **Association:** it is the relationship between an actor and a use case. It indicates that an actor can use a certain functionality of the system.
  - **Include:** it is a relationship between two use cases. It signifies that the use case on the side to which the arrow points is included in the use case on the other side of the arrow. This means that for one functionality that the system provides, another functionality of the system is accessed.
  - **Extend:** it indicates alternative options under a certain use case. In other words, it extends the base case and adds more functionality to the system.
  - **Generalization of a use case:** it refers to the relationship which can exist between two use cases. It shows that one use case (child) inherits the structure, behavior and relationships of another use case (parent)

In our application, we have two actors:**The Patient** and **The doctor**. Table 3.1 resumes all the use cases for each actor depending on his role. Moreover, figure 3.3 illustrates the use case diagram of our application.

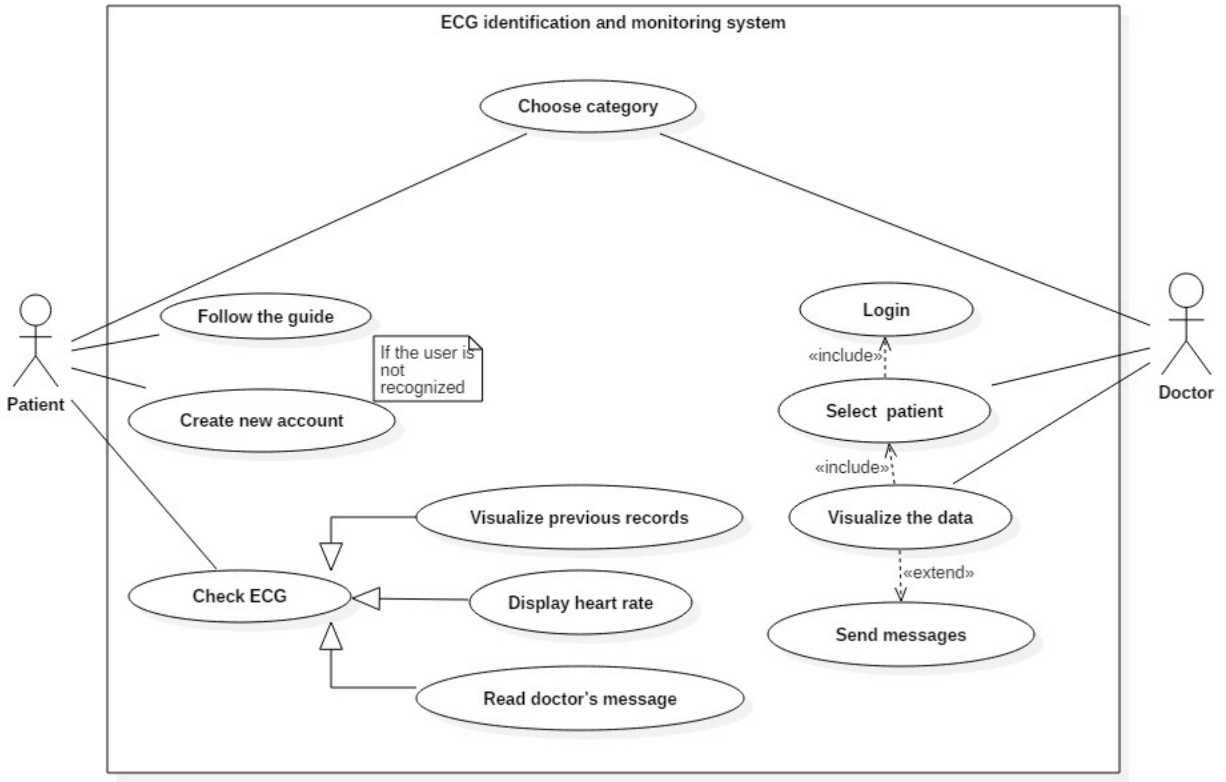


Figure 3.3: Use case diagram of our application

3.3.2.2 Sequence Diagram

Definition

Sequence diagrams in UML show how objects interact with each other and the order of the occurrence of these interactions. Each sequence diagram shows the interactions for a particular scenario. The processes are represented vertically whereas interactions are represented as arrows.

Used Sequence Diagram Components

- **Actor:** shows entities that interact with the system.
- **Life line:** represents the passage of time as it extends downward. This dashed vertical line shows the sequential events that occur to an object during the charted process. Life lines

Actor	**Use cases**
Patient	- Choose category - Follow the guide - Create new account (if the user is not recognized) - Check ECG • Visualize previous records • Display heart rate • Read doctor's messages
Doctor	- Choose category - Login - Select patient - Visualize the recordings - Send messages

Table 3.1: Actors use cases summary

may begin with a labeled rectangle shape or an actor symbol.

- **Activation box:** represents the time an object needs for an object to complete a task.
- **Synchronous message:** represented by a solid line with a solid arrowhead. This symbol is used when a sender must wait for a response to a message before it continues. The diagram should show both the call and the reply.
- **Asynchronous message:** represented by a solid line with a lined arrowhead. Asynchronous messages do not require a response before the sender continues. Only the call should be included in the diagram.
- **Return message:** represented by a dashed line with a lined arrowhead, these messages are replies to calls.
- **Self message:** used when a call is sent from a life line to itself.
- **Alternative fragment operator:** divides fragments into groups and defines conditions for each group. Only the one whose conditions are satisfied will be executed.

- **Reference fragment:** refers to an interaction defined on another diagram. The frame will cover all the life lines involved in the interaction.

Sequence Diagrams Of Our Application

Our system consists of seven sequence diagrams. Each of them, will be followed by its textual description.

1. **Sequence diagram 01: Doctor Login**: represented by table 3.2 and figure 3.4.
2. **Sequence diagram 02: Data visualization**: represented by table 3.3 and figure 3.5.
3. **Sequence diagram 03: Patient Identification**: represented by table 3.4 and figure 3.6.
4. **Sequence diagram 04 : Sign in**: represented by table 3.5 and figure 3.7.
5. **Sequence diagram 05 : Check ECG: See previous recordings**: represented by table 3.6 and figure 3.8.
6. **Sequence diagram 06: Check ECG: Heart rate**: represented by table 3.7 and figure 3.9.
7. **Sequence diagram 07: Check ECG: read doctor's message:** represented by table 3.8 and figure 3.10.

Use case name	Doctor Login
Actor	Doctor
Scenario	- The user launches the application - The system displays the main layout - The user selects the doctor category - The system displays the login layout - The user enters ID and password - The system needs to verify the entered ID and password
Alternative	- If ID and password correct:the system will display list of patients - Else: the system displays an error message

Table 3.2: Scenario 1: Doctor Login textual description

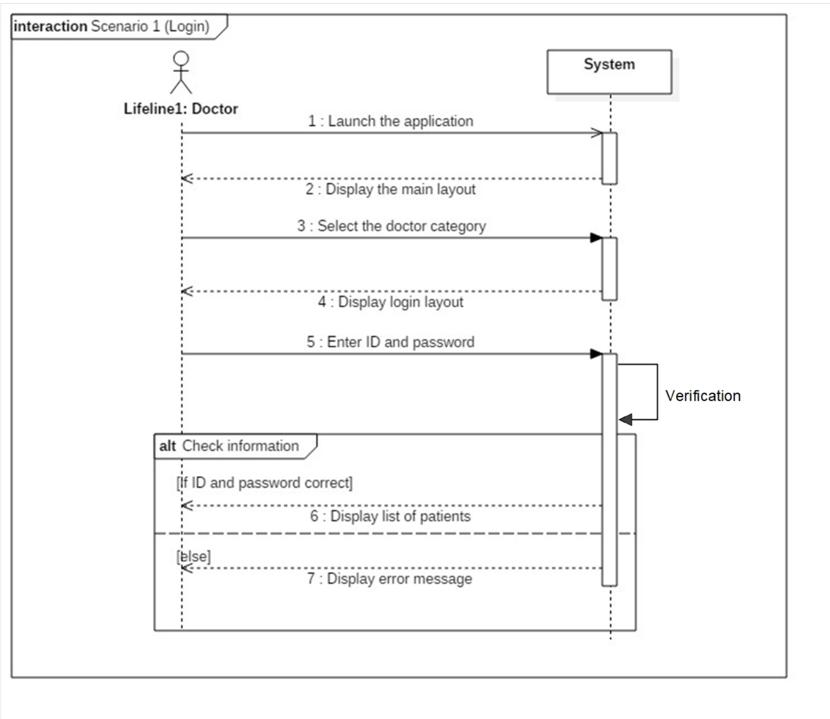


Figure 3.4: Scenario 1: Doctor Login

Use case name	Data Visualization
Actor	Doctor
Precondition	Scenario 1: Doctor Login
Scenario	- The doctor selects one patient from the list of patients - The system displays the list of recordings of the selected patient - The doctor selects a recording - The system visualizes the ECG signal
Optional	- The doctor will have the choice either to write a message to his patient or not.

Table 3.3: Scenario 2: Data visualization textual description

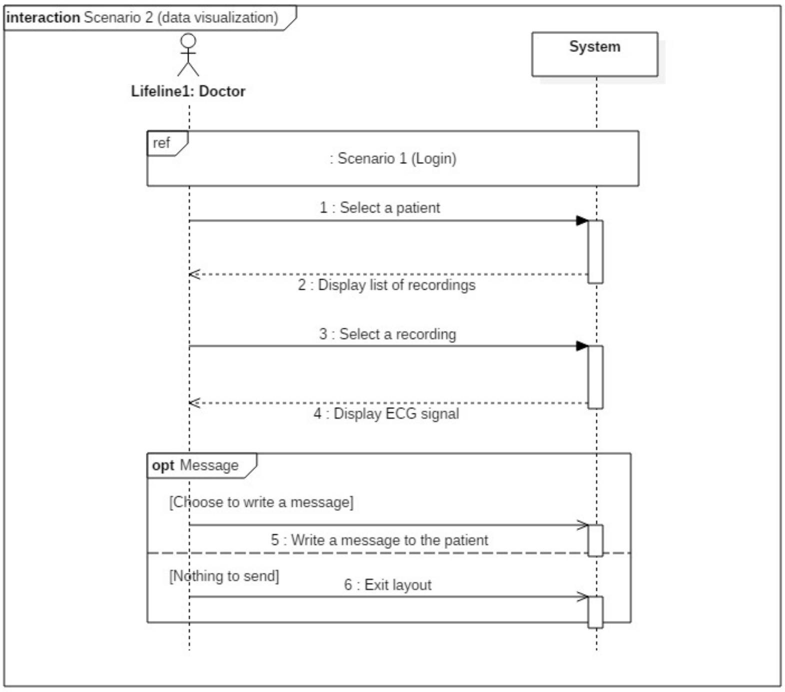


Figure 3.5: Scenario 2: Data visualization

Use case name	Patient Identification
Actor	Patient
Scenario	- The user launches the application - The system displays the main layout - The user selects the patient category - The system displays the guide layout which illustrates the instructions for placing the electrodes - The user clicks the next button when ready - The system displays a waiting layout while processing in the background the ECG of the patient
Alternative	- If the ECG is matched, hence the person is identified, the system displays the check ECG layout - Else, if the person is not identified, the system displays a sign up layout

Table 3.4: Scenario 3: Patient Identification textual description

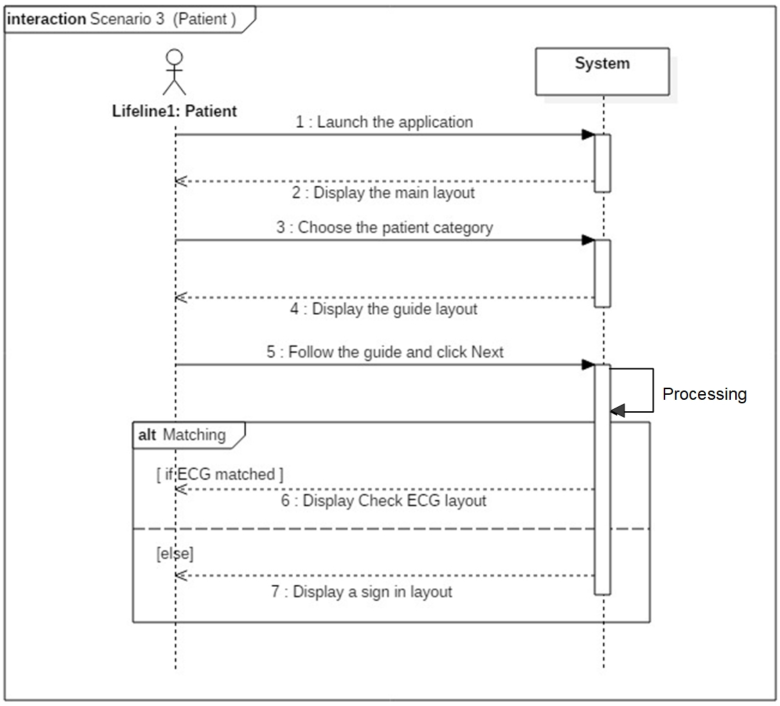


Figure 3.6: Scenario 3: Patient

Use case name	Patient Sign up
Actor	Patient
Precondition	Scenario 3: Patient Identification
Scenario	- The user fills in the registration form (the input fields) - The system sends a query to the server to insert the patient information - The server executes the query and checks the validity of the entered data and sends a confirmation message to the system - The system displays the Check ECG layout

Table 3.5: Scenario 4: Patient sign up textual description

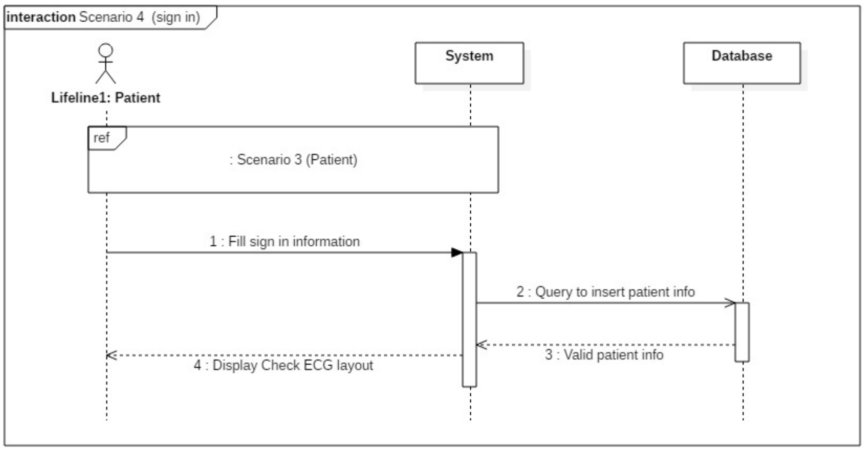


Figure 3.7: Scenario 4: Patient Sign up

Use case name	See previous recordings
Actor	Patient
Precondition	Scenario 3: Patient Identification
Scenario	- The user clicks the "See previous recordings" button - The system displays the list of the saved recordings - The user chooses one recording from the displayed list - The system visualized the selected recording

Table 3.6: Scenario 5: "See previous recordings" textual description

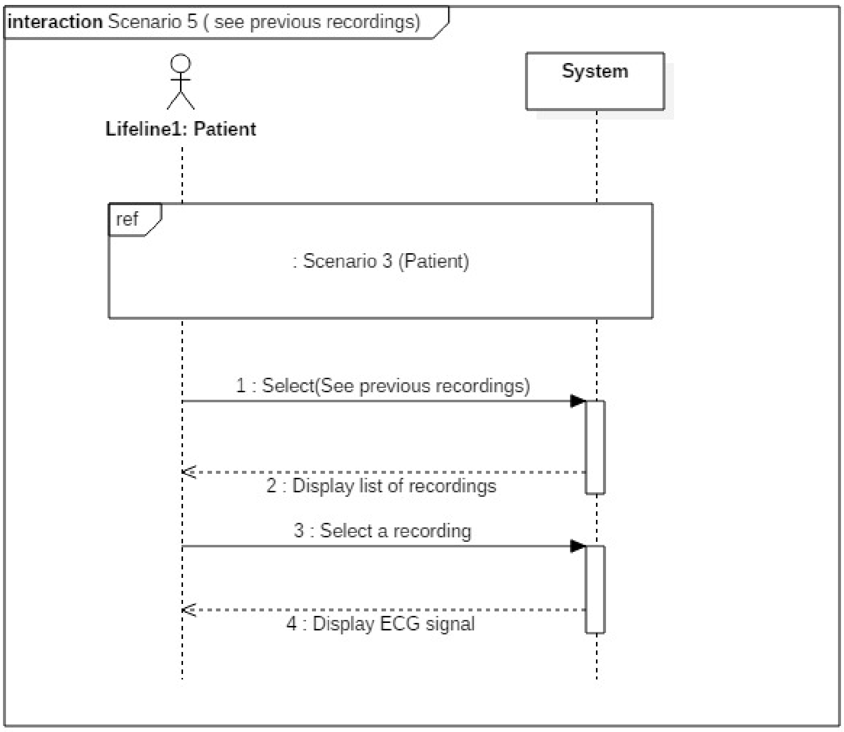


Figure 3.8: Scenario 5: See previous recordings

Use case name	See previous recordings
Actor	Patient
Precondition	Scenario 3: Patient Identification
Scenario	- The user clicks the "Heart rate" button - The system displays the calculated heart rate of the patient

Table 3.7: Scenario 6: "Heart Rate" textual description

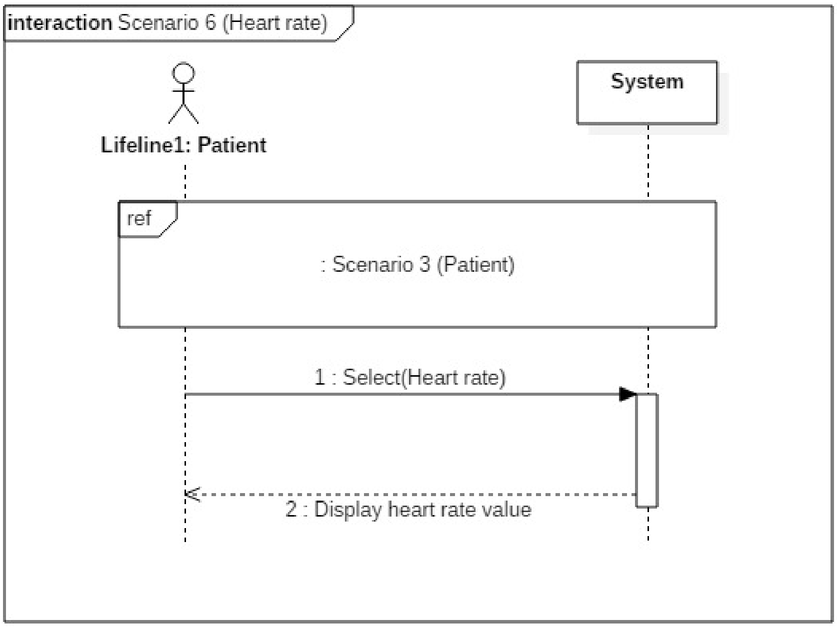


Figure 3.9: Scenario 6: Heart rate

Use case name	See previous recordings
Actor	Patient
Precondition	Scenario 3: Patient Identification
Scenario	- The user clicks on the "Read doctor's messages" button - The system displays the doctors messages

Table 3.8: Scenario 7: "Check doctor's messages" textual description

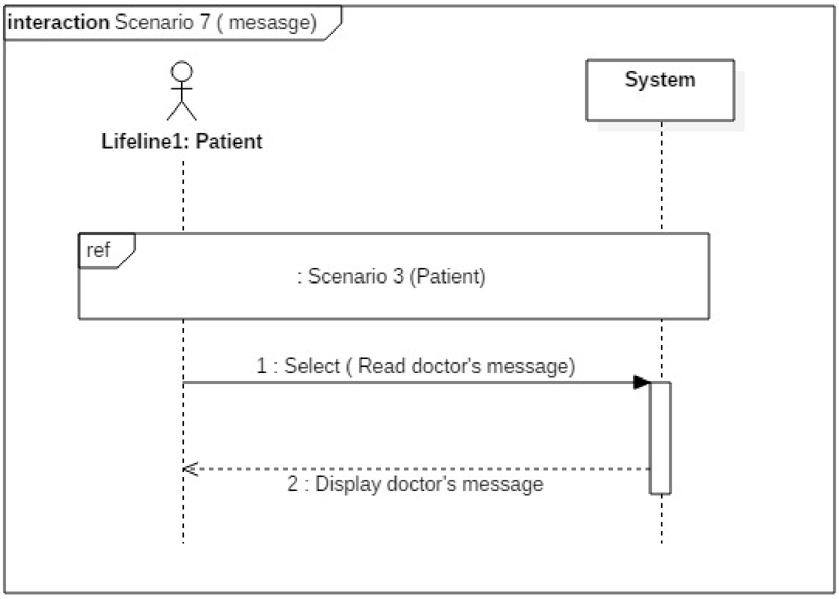


Figure 3.10: Scenario 7: Check doctor's messages

3.3.3 Database

Definition

A database is a collection of related data or information usually stored and accessed using a computer system. A DBMS is used for defining, creating, maintaining, and controlling access to the database[1].

Class Diagram

Before creating the database, designing the appropriate UML class diagram is a crucial step. It presents an overview of the used data.

As figure 3.11 illustrates, the following rules were taken into consideration:

- A patient can have only one doctor but a doctor can have several patients.
- A patient can have many records and a record belongs to only one patient.
- A patient can send several messages but a message is sent to a single patient.

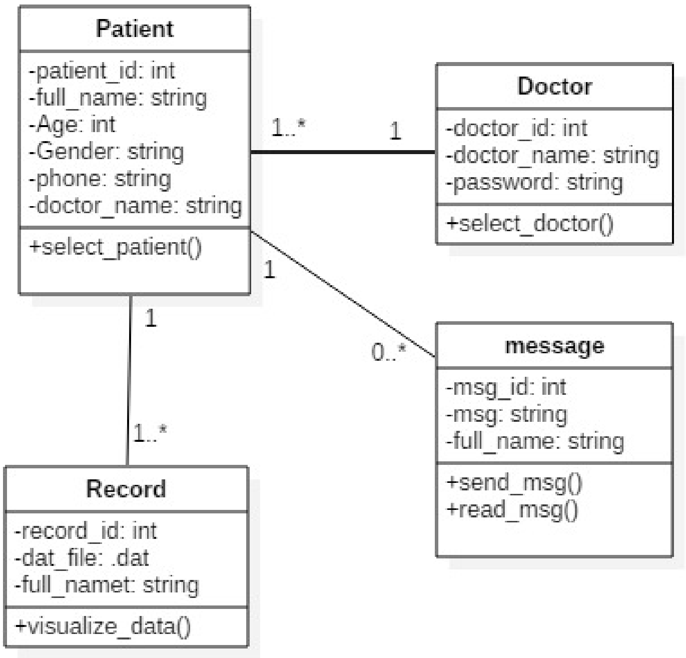


Figure 3.11: Class Diagram

After setting the management rules and realizing the class diagram, the next step is to transform this diagram into an equivalent relational model as follows:

- First, we create a table for each class whose fields are attributes of the class diagram.
- After that, we choose a primary key for each table.

- Then, we assign foreign keys.
- Finally, we set the class diagram associations as database relationships the obtained EER diagram is represented in figure 3.12 .

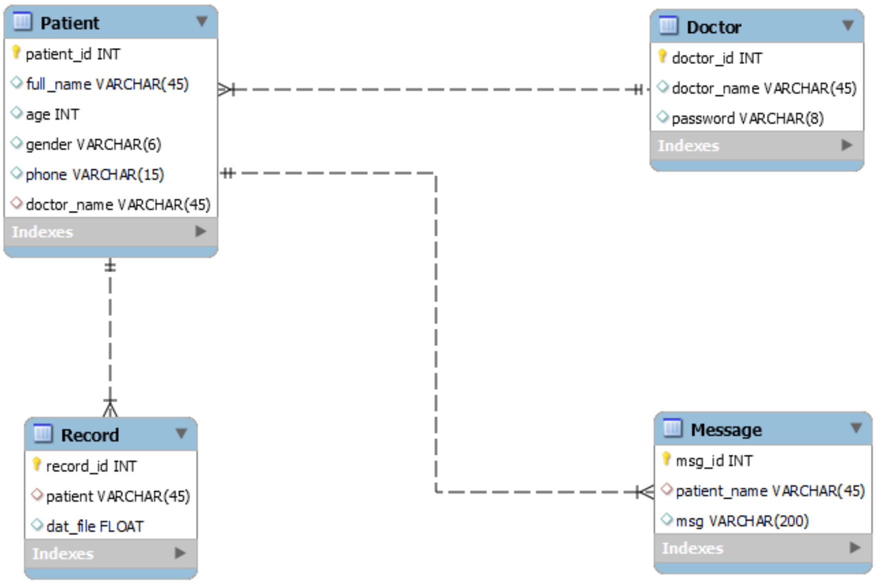
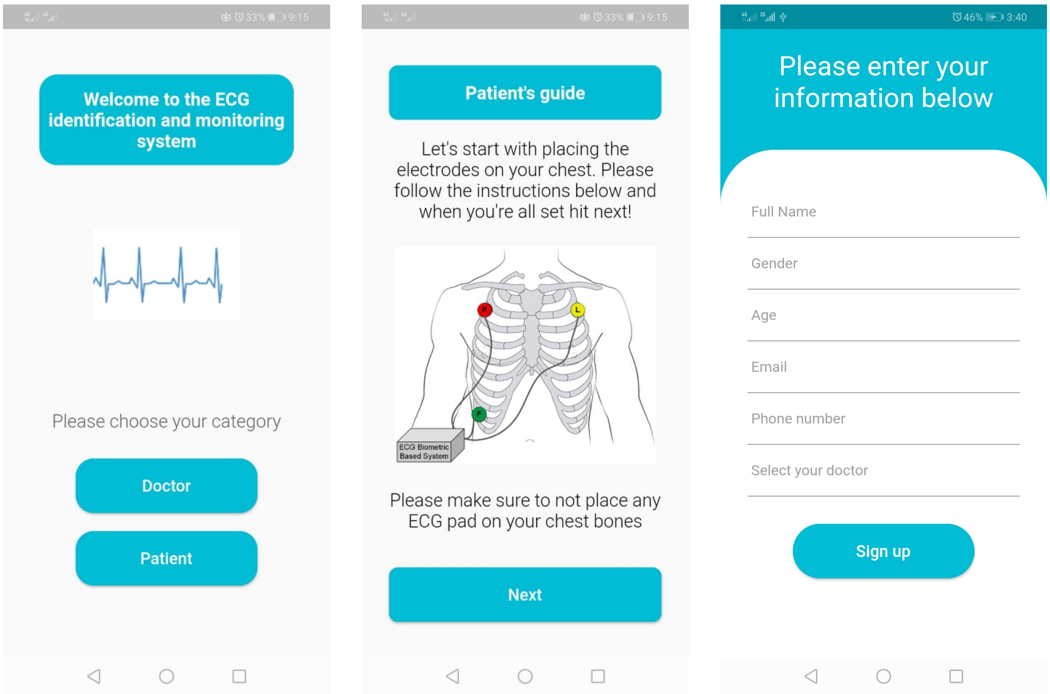


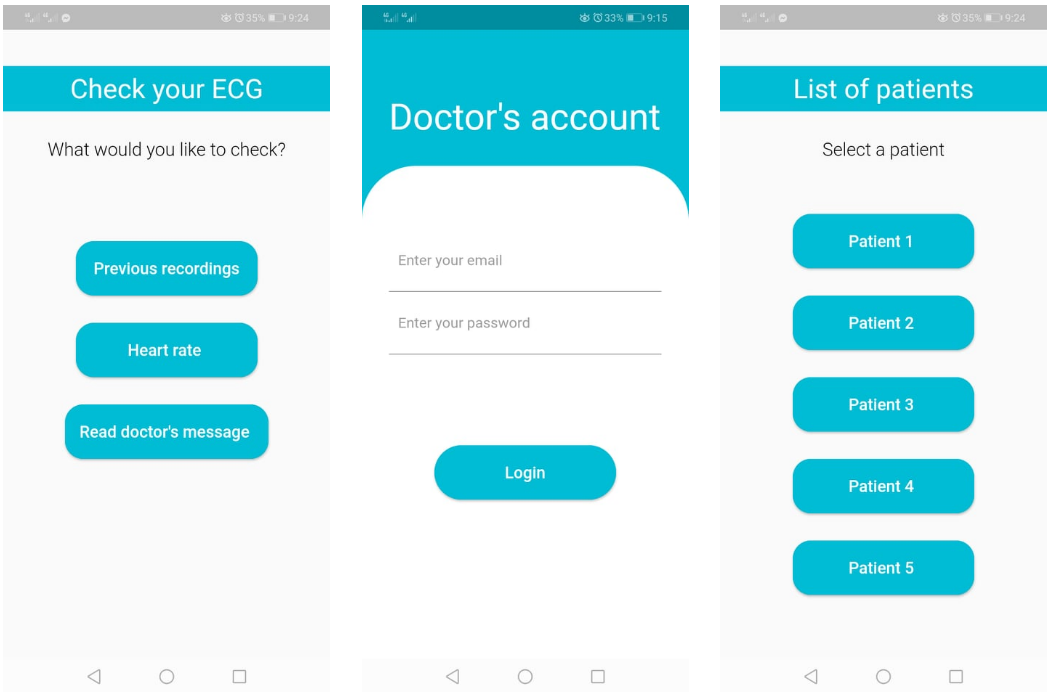
Figure 3.12: Enhanced Entity Relationship Diagram

3.3.4 User Interfaces

Figure 3.13 shows the design of our user interface through some layouts screenshots.



(a) The main layout (b) The guide layout (c) The patient sign up layout



(d) The check ECG layout (e) The doctor login layout (f) The list of patients layout

Figure 3.13: Some user interfaces from the android application

Conclusion

In this chapter we have covered, the last step in the ECG identification process, which is the Template Matching. In addition, we have seen both the skeleton and the design of our user friendly application through UML diagrams and the different layouts. In the coming chapter, we will present and discuss all the obtained results at each of the previously discussed stages.

Chapter 4

Results and Discussion

Introduction

In this chapter, we will outline the various methods we used to evaluate our models. Following that, we will discuss the results achieved through the aforementioned techniques.

4.1 Evaluation Metrics

In equations 4.1, 4.2, 4.3, and 4.4:

- TP: is true positives
- FP: is false positives
- TN: is true negatives
- FN: is false negatives

4.1.1 Sensitivity

Sensitivity is the metric that evaluates a model's ability to predict true positives of each available category. Also called *"Recall"*, it is the true positive rate, i.e., the ratio of correctly predicted positive to the total number of actually positive observations. It is given by equation 4.1.

$$sensitivity = \frac{TP}{TP + FN} \tag{4.1}$$

4.1.2 Positive Predictive Value

The Positive Predictive Value (PPV) or precision is the proportion of correct positive predictions for the total number of positive observations. It is given by equation 4.2.

$$PPV = \frac{TP}{TP + FP} \tag{4.2}$$

4.1.3 False Rejection Rate

The FRR (false rejection rate), is the probability of cases for which a biometric system fallaciously denies access to an authorized person. It happens when a biometric system, solution or application fails to match the biometric input with a stored template, fallaciously returning a no-match and denying access to an authorized person. It is given by equation 4.3.

$$FRR = \frac{FN}{TP + FN} \tag{4.3}$$

4.1.4 Accuracy

Accuracy represents the proportion of the total number of predictions that were correct. In other words, it is the ratio of the correct predicted observations made by the model over all-kinds of predictions made. It is given by equation 4.4.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4.4}$$

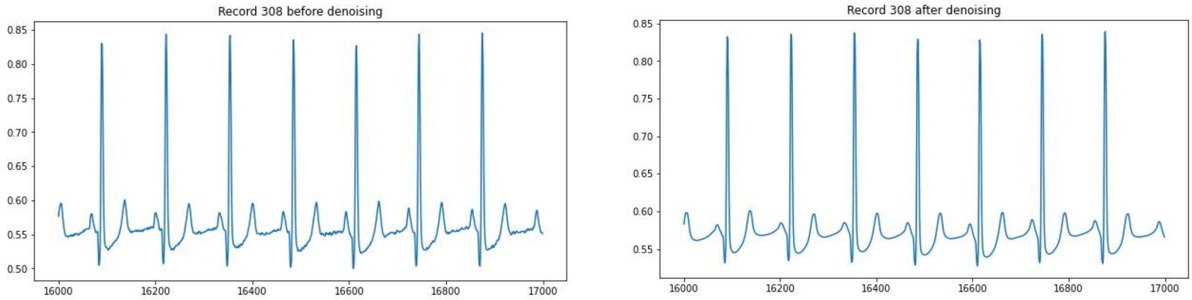
4.2 Results

In this section we will show the different results obtained in each of the previously discussed stages, as well as the overall performance of our system.

4.2.1 Denoising

By applying the wavelet transform explained in section 2.3 we could obtain good results by eliminating the apparent noise without losing the main characteristics of our ECG signals. Figure 4.1 shows the effect of applying the filter on a record from each of the collected data. [1]

[1]There is no evaluation metric for the denoising in our case, since the amount of original noise in unknown.



(a) Record 308 before denoising

(b) Record 308 after denoising

Figure 4.1: Record 308 before and after denoising

4.2.2 Segmentation

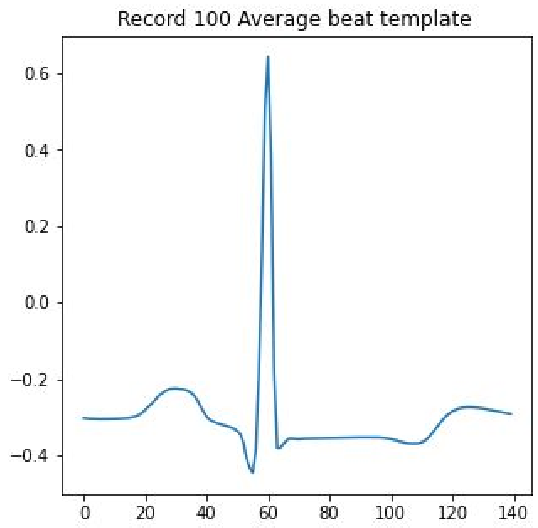
As explained earlier in section 2.4 two different approaches were used on our datasets in order to segment the ECG signals into beats by identifying the R-peaks. Table 4.1 summarizes the obtained results on both the collected data and the MIT BIH Arrhythmia.

Dataset	Sensitivity	Positive Predictivity
Collected data	99.76%	99.46%
MIT BIH Arrhythmia	100%	100%

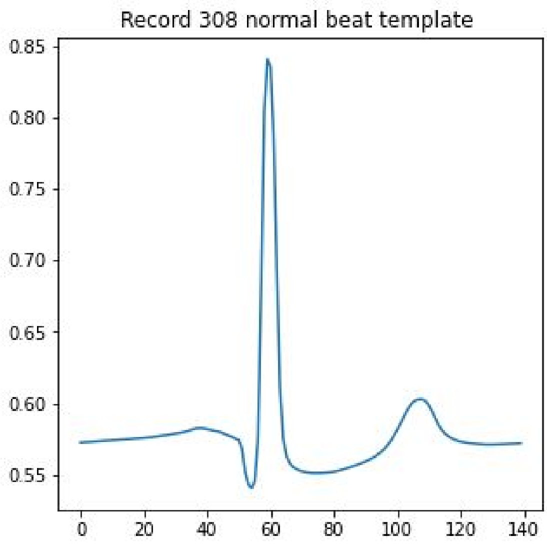
Table 4.1: Detected R peaks on both the collected data and MIT BIH Arrhythmia datasets

4.2.3 Feature Extraction

As mentioned before we extracted three main features in order to generate the templates for each record: the $Rmagnitude$, $RR_{distances}$, and the average beat. Figure 4.2 shows an average beat for record 100 (from the MIT BIH Arrhthmia database) and record 308 from the collected data.



(a) Record 100 normal beat template (MIT BIH Arrhythmia)



(b) Record 308 Normal beat template (Collected data)

Figure 4.2: Normalized beat templates

4.2.4 Template Matching

In order to measure the efficiency and test the performance of our system, we calculated the different evaluation metrics presented in section 4.1 The results of the calculations are summarized in tables 4.2 and 4.3. [2]

Database	number of subjects	number of correctly identified subjects	number of misidentified subjects
MIT BIH Arrhythmia	44	40	4
Collected data	8	7	1

Table 4.2: Results of identification on MIT BIH and collected data

[2] The MIT BIH Arrhythmia database contains four paced beat recordings, they were excluded

Database	Accuracy	FRR
MIT BIH Arrhythmia	90.90%	0.047
Collected data	87.5%	0

Table 4.3: Results of identification on MIT BIH and collected data (calculations)

4.3 Discussion

In this work we proposed an ECG based monitoring system. We suggested different methods for each of the major phases of any identification system.

We started by cleaning the ECG signals. As can be seen in the results section, the wavelet transform was a good choice for the denoising stage. It represents an effective method for eliminating the noise from the ECG signals while preserving the main properties and important information carried by the signal. Unfotunately, the mathematical evaluation of the applied filter was not possible because we ignore the amount of noise that affects our signals.

For the segmentation phase, we could notice that the results were quite high on the two used datasets despite using two different methods (ranging from 99.46% to 100%). However, the wavelet based R peak detection was slightly more accurate by not missing any R peak, this represents a huge advantage in avoiding the errors while generating the templates and even calculating the heart rate. The only drawback of this method is the execution time which is longer than the conditioned window method.

Concerning the most important step of the identification process, the template matching, we could achieve acceptable results. It is known that the MIT-BIH Arrhythmia database subjects contain a lot of pathologies which makes identifying such signals quite difficult due to the different changes in the cardiac rhythm and even the shape of the heart beats. Despite this difficulty, we managed to correctly identify 40 subjects out of 44 recordings. We excluded 4 recordings from the original database (records 102, 104, 107, and 217) because they belong to subjects with paced beats. This led us to achieve an accuracy of 90.90%. For the collected data, it was a much smaller database because we collected it by ourselves. Nevertheless, we could gather a good quality data appropriate for a study. The used sensor was able to output a stable ECG signal compared to its cost. We were able to identify 7 subjects out of 8 with an accuracy of 87.5%. Due to the size of the dataset we cannot really judge the real performance of the system over the collected data. Maybe a larger dataset would give better results by more familiarizing the system with this type of data.

In overall, the performance of our system in each of the previous stages was good. The originality of our work lies in proposing a complete system with hardware, software, and even a phone application. Moreover, we could employ a trending research topic ,which is using ECG as a biometric, in a real world application to solve one of the major problems in Algeria as well as the world "the digitization of the health system". Despite the good performance, there is still a room for future improvement.

Conclusion

In this chapter, we described the metrics used for the evaluation of our proposed system. First, we showed the results of the denoising step using the wavelet transform on both datasets. Then, we compared the performance of the different approaches used in the segmentation phase. After that, we gave an example of the generated average beats by taking one record from each dataset. Finally, we calculated the described evaluation metrics in order to measure the performance of our entire system. We could notice that the results were pretty good despite the difficulty of identifying the abnormal records from the MIT BIH arrhythmia database.

Conclusion and Future Work

The proposed system aims to auto identify the users based on their ECG signals. It also acts as a bridge between the physician and his patients by offering the remote monitoring option. We can divide the implemented work into three main pillars: the hardware system that serves as a data collector, the ECG biometric system which represents the identification part of the project, and an Android application that introduces the whole project through a user friendly interface making the interaction much easier.

Generalities and theoretical background were presented in the first chapter. It highlighted the other biometric modalities and the reasons of choosing ECG over them. It also provided a brief overview of the literature and the work already implemented in this field. The second chapter, discussed the general structure of the system. Starting from the data collection through our hardware system and the used database "MIT-BIH Arrhythmia". It also described the used methods in all of the denoising, segmentation, feature extraction, and template matching stages. The third chapter, described the used tools and technologies in our implementation. Moreover, it included the detailed implementation of the correlation factor based template matching. It also presented the design requirements, modeling and user interface of the implemented phone application.

From the results discussed in the fourth chapter we can conclude that the system was successfully implemented and our initial objectives were met. The achieved results are quite promising: clean denoised ECG signals, more than 99.46% of the R peaks were correctly detected, 90.90% of the subjects from the MIT BIH Arrhythmia database were correctly identified with 4.7% of false rejection rate, and 87.5% for the collected data with 0% of false rejection rate.

Despite the promising results, further work needs to be carried out in order to improve our system.

- One of the ways we could improve our work is by testing it on larger datasets with different cardiovascular illnesses not only arrhythmia. That way, our system will be familiar with a variety of abnormalities.

- Another key of improvement is adding an automatic heart beat classifier. This would make the system more of an autonomous monitoring system without waiting for the physician's diagnosis.
- Furthermore, treating the identification process as a classification problem instead of a matching problem would give a higher accuracy. However, this approach might be computationally expensive.
- Finally, we believe as well that the used hardware plays a major role in obtaining good results, so using more sophisticated sensors or edge device would reduce the response time and give more steady results.

Bibliography

[1] BOUAMAMA Abdelbaki. Design and Implementation of a Web-based Management System. Case of the “IGEE” Scientific Council. Master’s thesis, IEEE ex INELEC, 2020.

[2] Carlos Carreiras, André Lourenço, Ana Fred, and Rui Ferreira. Ecg signals for biometric applications - are we there yet? In *2014 11th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, volume 02, pages 765–772, 2014.

[3] Hyun-Soo Choi, Byunghan Lee, and Sungroh Yoon. Biometric authentication using noisy electrocardiograms acquired by mobile sensors. *IEEE Access*, 4:1266–1273, 2016.

[4] Iulian B. Ciocoiu and Nicolae Cleju. Off-the-person ecg biometrics using convolutional neural networks. In *2019 International Symposium on Signals, Circuits and Systems (ISSCS)*, pages 1–4, 2019.

[5] Federal Trade Commission. New data shows ftc received 2.2 million fraud reports from consumers in 2020. `https://www.ftc.gov/news-events/press-releases/2021/02/new-data-shows-ftc-received-2-2-million-fraud-reports-consumers`. Accessed: June 2021.

[6] Eduardo José da Silva Luz, Gladston J. P. Moreira, Luiz S. Oliveira, William Robson Schwartz, and David Menotti. Learning deep off-the-person heart biometrics representations. *IEEE Transactions on Information Forensics and Security*, 13(5):1258–1270, 2018.

[7] Google. Colaboratory frequently asked questions. `https://research.google.com`. Accessed: June 2021.

[8] Google. Flutter. `https://flutter.dev`. Accessed: June 2021.

[9] Thales Group. Biometrics: definition, use cases and latest news. `https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/inspired/biometrics`. Accessed: may 2021.

[10] Medina Hadjem and Farid Naït-Abdesselam. An ecg t-wave anomalies detection using a lightweight classification model for wireless body sensors. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 278–283, 2015.

[11] Siddarth Hari, Foteini Agrafioti, and Dimitrios Hatzinakos. Design of a hamming-distance classifier for ecg biometrics. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3009–3012, 2013.

[12] Maryamsadat Hejazi, S. A. R. Al-Haddad, Shaiful Jahari Hashim, Ahmad Fazli Abdul Aziz, and Yashwant Prasad Singh. Non-fiducial based ecg biometric authentication using one-class support vector machine. In *2017 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pages 190–194, 2017.

[13] Pei-Lun Hong, Jyun-Ya Hsiao, Chi-Hsun Chung, Yao-Min Feng, and Shun-Chi Wu. Ecg biometric recognition: Template-free approaches based on deep learning. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2633–2636, 2019.

[14] Mohit Ingale, Renato Cordeiro, Siddartha Thentu, Younghee Park, and Nima Karimian. Ecg biometric authentication: A comparative analysis. *IEEE Access*, 8:117853–117866, 2020.

[15] Woo-Hyuk Jung and Sang-Goog Lee. Ecg identification based on non-fiducial feature extraction using window removal method. *Applied Sciences*, 7:1205, 11 2017.

[16] Emna Kalai Zaghouani, Adel Benzina, and Rabah Attia. Ecg based authentication for e-healthcare systems: Towards a secured ecg features transmission. In *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 1777–1783, 2017.

[17] Nima Karimian, Mark Tehranipoor, Damon Woodard, and Domenic Forte. Unlock your heart: Next generation biometric in resource-constrained healthcare systems and iot. *IEEE Access*, 7:49135–49149, 2019.

[18] Nima Karimian, Damon Woodard, and Domenic Forte. Ecg biometric: Spoofing and countermeasures. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(3):257–270, 2020.

[19] Hanvit Kim and Se Young Chun. Cancelable ecg biometrics using compressive sensing-generalized likelihood ratio test. *IEEE Access*, 7:9232–9242, 2019.

[20] Majid Komeili, Wael Louis, Narges Armanfard, and Dimitrios Hatzinakos. Feature selection for nonstationary data: Application to human recognition using medical biometrics. *IEEE Transactions on Cybernetics*, 48(5):1446–1459, 2018.

[21] Ruggero Donida Labati, Roberto Sassi, and Fabio Scotti. Ecg biometric recognition: Permanence analysis of qrs signals for 24 hours continuous authentication. In *2013 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 31–36, 2013.

[22] Jikui Liu, Liyan Yin, Chenguang He, Bo Wen, Xi Hong, and Ye Li. A multiscale autoregressive model-based electrocardiogram identification method. *IEEE Access*, 6:18251–18263, 2018.

[23] Vu Mai, Ibrahim Khalil, and Christopher Meli. Ecg biometric using multilayer perceptron and radial basis function neural networks. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2745–2748, 2011.

[24] Mathworks. What is matlab. `https://www.mathworks.com/discovery/what-is-matlab.html`. Accessed: June 2021.

[25] Microsoft. Learn to code with visual studio code. https://code.visualstudio.com/learn. Accessed: June 2021.

[26] Alleydog.com's online glossary. Biometrics. `https://www.alleydog.com/glossary/definition.php?term=Biometrics`. Accessed: May 2021.

[27] Joana S. Paiva, Duarte Dias, and João P. S. Cunha. Beat-id: Towards a computationally low-cost single heartbeat biometric identity check system based on electrocardiogram wave morphology. *PLOS ONE*, 12(7):1–32, 07 2017.

[28] PeaceHealth. Electrical system of the heart. `https://www.peacehealth.org/medical-topics/id/te7147abc`. Accessed: May 2021.

[29] PhysioNet. Mit-bih arrhythmia database directory. `https://archive.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm`. Accessed: May 2021.

[30] João Ribeiro Pinto and Jaime S. Cardoso. An end-to-end convolutional neural network for ecg-based biometric authentication. In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8, 2019.

[31] Plurilock. Behavioral biometrics history, applications, and technology. https://www.plurilock.com/behavioral-biometrics-guide/1-what-is-behavioral-biometrics/. Accessed: May 2021.

[32] Processing. Processing ide. https://processing.org/reference/environment/. Accessed: June 2021.

[33] Chander Kant Rubal Jain. Attacks on Biometric Systems: An Overview. *International Journal of Advances in Scientific Research*, 2015.

[34] B. Yu S. G. Chang and M. Vetterli. Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Trans. Image Process.*, vol. 9 No. 9, p. 1522-1531, 2000.

[35] XTN Cognitive Security. Behavioral biometrics for beginners: 10 questions you might have! https://xtn-lab.com/10-questions-about-behavioral-biometrics/. Accessed: May 2021.

[36] Abdelkader Sellami, Amine Zouaghi, and Abdelhamid Daamouche. Ecg as a biometric for individual's identification. pages 1–6, 10 2017.

[37] ZOUAGHI A. SELLAMI A. ECG as Biometric for Individual's Identification. Master's thesis, IEEE ex INELEC, 2017.

[38] Khairul A. Sidek, Ibrahim Khalil, and Herbert F. Jelinek. Ecg biometric with abnormal cardiac conditions in remote monitoring system. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(11):1498–1509, 2014.

[39] Khairul A. Sidek, Ibrahim Khalil, and Herbert F. Jelinek. Ecg biometric with abnormal cardiac conditions in remote monitoring system. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(11):1498–1509, 2014.

[40] Pathoumvanh Somsanouk, Surapan Airphaiboon, and Kazuhiko Hamamoto. Robustness study of ecg biometric identification in heart rate variability conditions. *IEEJ Transactions on Electrical and Electronic Engineering*, 9, 05 2014.

[41] Statista. Global increase in identity theft risk since the coronavirus outbreak as of august 2020. https://www.statista.com/statistics/1175657/increase-identity-theft-coronavirus-outbreak/. Accessed: June 2021.

[42] B. W. Silverman T. Cai. Incorporating information on neighbouring coefficients into wavelet estimation. *Sankhya, Indian J. Statist.*, vol.63(2001), No. 2, p. 127-148, 2001.

[43] Government Technology. Tracing the history of biometrics. `https://www.govtech.com/public-safety/tracing-the-history-of-biometrics.html`. Accessed: June 2021.

[44] Saeid Wahabi, Shahrzad Pouryayevali, Siddarth Hari, and Dimitrios Hatzinakos. On evaluating ecg biometric systems: Session-dependence and body posture. *IEEE Transactions on Information Forensics and Security*, 9(11):2002–2013, 2014.

[45] Wikepedia. Biometrics. `https://en.wikipedia.org/wiki/Biometrics`. Accessed: may 2021.

[46] Wikepedia. Dart programming language. `https://en.wikipedia.org/wiki/Dart_(programming_language)`. Accessed: June 2021.

[47] Wikepedia. Electrocardiography. `https://en.wikipedia.org/wiki/Electrocardiography`. Accessed: May 2021.

[48] Wikepedia. Python programming language. `https://en.wikipedia.org/wiki/Python_(programming_language)`. Accessed: June 2021.

[49] Yongfeng Dai1 Xiaoli Wang1. An Improved Denoising Method Based on Stationary Wavelet Transform. *College of Mechanical Electronic Information, China University of Geosciences, Wuhan, Hubei.*

[50] Amit Kumar Ray Yogendra Narain Singh, Sanjay Kumar Singh. Bioelectrical Signals as Emerging Biometrics: Issues and Challenges. *International Scholarly Research Notices*, vol. 2012, 2012.

[51] Badr Eddine ZEBBICHE. Implementation of a Biometric Identification System using Electrocardiogram ECG/EKG. Master's thesis, IEEE ex INELEC, 2015.