PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of higher education and scientific research
University Mohamed Khider of Biskra
Faculty of exact sciences and natural science and life
Department of computer science

Order N° : SIOD 22/M2/2019

Memory

Presented for the academic master's degree in

Computer Science

Option : Information system, optimization and decision support

---

HEART DISEASE ANALYSIS BASED ON DATA MINING, DEEP LEARNING AND WEARABLE TECHNOLOGIES

By :
KABOT ABDELGHANI

Submitted on 06/07/2019, before the jury composed of :

DJEDI Noureddine	Professor	President
ABDELLI Belkacem	MC-B	Supervisor
BELOUNNAR Saliha	MA-A	Examiner

Acknowledgments

All the thanks and praise is for god, which give me the force and the will to understand accomplish this modest study.

By the way, I would be so proud to thank deeply my sincere teacher and supervisor Dr. Belkacem ABDELLI for his patience, help and advice.

Many thanks to all my teachers who encourage me at the University of Mohamed Khider Biskra.

Abdelghani KABOT

DEDICATION

To my:

Mother and Father

Sisters : Khedidja, Asma, Ferdous, Nour elhouda and Salsabil.

And to all my family members.

Abdelghani KABOT.

Contents

Abbreviations vii

Abstract vii

1 introduction 3
- 1.1 Overview : 3
- 1.2 Why this title? 6
- 1.3 Motivation : 6
- 1.4 Organization of the Thesis : 7

2 The State of the Art 10
- 2.1 Introduction : 10
- 2.2 Wearable Health Devices Market Trends : 10
- 2.3 Automatic system and Vital signs (General Works): 12
- 2.4 Related works : 13
  - 2.4.1 ECG Denoising : 14
  - 2.4.2 ECG Heartbeats Classification : 15
  - 2.4.3 ECG Arrhythmia Classification : 16
  - 2.4.4 Heart disease prediction based on DM techniques: 16

3 Background about heart disease , Data Mining and Deep Learning Techniques 17
- 3.1 Heart : 17
  - 3.1.1 The structure of the heart: 17
  - 3.1.2 The cardiac cycle : 18
  - 3.1.3 The process of a Heartbeat : 19
- 3.2 Cardiac Background : 20
- 3.3 Heart diseases : 22
  - 3.3.1 Myocardial Infarction and heart attack : 22

3.4 Data Mining and Deep Learning : 23
3.4.1 DM Definition, Techniques and KDD process : 23
3.4.1.1 DM Definition : 23
3.4.1.2 KDD process : 23
3.4.2 DL Definition and process : 24
3.4.2.1 DL Definition : 24
3.4.2.2 DL process : 24
3.4.3 Techniques and Methods Used : 24
3.4.3.1 Supervised Learning : 24
3.4.3.2 Semi-Supervised Learning : 25
3.4.3.3 Classification : 25
3.4.3.4 K Neighbors classifier : 25
3.4.3.5 Neural Networks, CNN , AE and LSTM : 26
3.4.3.5.1 Neural Networks: 26
3.4.3.5.2 Convolutional Neural Networks (CNN) : 27
3.4.3.5.3 Autoencoders (AE) : 29
3.4.3.5.4 Long short-term memory (LSTM) : 30
3.5 Conclusion : 31

4 Conception and Realization **32**
4.1 Introduction : 32
4.2 Global Vision and Conception: 32
4.2.1 Global Conception : 33
4.3 Detailed conception and realization : 34
4.3.1 Data Selection and Acquisition : 34
4.3.1.1 Data Selection : 34
4.3.1.2 Data Acquisition : 34
4.3.1.2.1 PhysioNet website : 34
4.3.1.2.2 UCI Machine learning website : 37
4.3.1.2.3 E-Health platform : 37
4.3.2 Environment of work for the rest of the project : 42
4.3.2.1 The machine used : 42
4.3.2.2 The development environment : 42
4.3.3 Pre-processing (ECG Denoising) : 43
4.3.4 Heartbeats Classification based on 1-D CNN : 43
4.3.4.1 Dataset used and Split : 43
4.3.4.2 Model Training and validation 44
4.3.5 ECG Arrhythmia classification : 46
4.3.5.1 Detection Myocardial Infarction (MI) Based on LSTM: . . 46

4.3.5.1.1 Dataset used : Exploploration and Split : 46
4.3.5.1.2 Building the model : 46
4.3.5.2 CVDs Classification : 47
4.3.5.2.1 Dataset used : 47
4.3.5.2.2 The model : 47
4.3.5.2.3 Cleaning data : 48
4.3.5.3 Heart disease prediction based on DM : 49
4.3.5.3.1 Dataset used : Visualization and Understanding: 49
4.3.5.3.2 The model : 52
4.4 Our future vision : 52

5 Results and discussion : 54
5.1 Introduction : 54
5.2 ECG Denoising: 54
5.3 ECG Heartbeats classification : 55
5.3.1 Comparison : 55
5.4 ECG Arrhythmias Classification : 56
5.4.1 Detection Myocardial Infarction (MI) Based on LSTM : 56
5.4.2 CVDs Classification : 56
5.5 Heart disease prediction based on DM Techniques: 56

6 Conclusion and Future works : 58
6.1 Conclusion : 58
6.2 Future works Plan : 59

Appendix : 60
.1 Web Application : 60
.2 Android Application : 63

Bibliographie 66

List of Figures

1.1.1 Informatics pyramid [12] 4

2.2.1 The total revenue the growth rate in billions (Adapted from[24] 11
2.2.2 Horizontal bar graphic showing the trend of global market value of wearable computing devices, in millions, between 2017 and 2019[15]. 11
2.2.3 (A) Connected home medical monitoring devices (in millions) 2011 to 2017; (B) The world market for telehealth from 2014 divided in the main areas (CHF-congestive heart failures; COPD-chronic obstructive pulmonary disease) [15] 12
2.3.1 Healthcare wearable sensors. 13
2.3.2 Schematic overview of the four main Analytics processes (activity, prediction, anomaly detection and diagnose/decision support) in relation to different aspects of wearable sensing in wearable health devices. Filled line-Medical purposes; Traced line: Activity purposes. 13
2.4.1 Number of retrieved scientific papers related with WHDs and a specific physiological sign. The distribution is divided in two time intervals and according to the purpose: medical or activity. (BT—Body Temperature; BP—Blood Pressure; RR—Respiration Rate; BG—Blood Glucose; HR—Heart Rate; SpO2—Blood Oxygen Saturation; ECG—Electrocardiogram) [15]. . 14

3.1.1 The Human Heart 18
3.1.2 The phases of a cardiac cycle 19
3.2.1 Schematic representation of 12 leads (electrodes) placed around the heart components [45] 20
3.2.2 Normal ECG recording 21
3.3.1 Myocardial Infarction Interpretation on ECG 22
3.4.1 The KDD process 23
3.4.2 ML vs DL 24
3.4.3 KNN illustration 26

3.4.4 Neuron in Neural Network. . . . 26
3.4.5 Small Neural Network Architecture. . . . 27
3.4.6 CNN Architecture. . . . 28
3.4.7 The convolution f g. Adopted from [44] . . . 28
3.4.8 Max Pooling. Adopted from [11] . . . 29
3.4.9 AE architecture . . . 29
3.4.10 Schema of an LSTM unit . . . 30

4.2.1 Approaches of our Vision . . . 33
4.2.2 Global Conception . . . 33
4.3.1 the sample of *.dat File . . . 36
4.3.2 the sample of *.atr File . . . 36
4.3.3 the sample of *.atr File . . . 37
4.3.4 Prototype's Schematic . . . 38
4.3.5 From left to right: -The E-HP, -The Connection bridge, -Raspberry Pi3 B . 38
4.3.6 ECG sensors plugged into the e-HP (with appropriate placements) . . . 39
4.3.7 The assembled platform . . . 40
4.3.8 The Power Bank . . . 41
4.3.9 The platform with the belt . . . 42
4.3.10 Difference between the B [28] model and our model A. . . . 44
4.3.11 Python Implementation of Heartbeats classification Model. . . . 44
4.3.12 Detailed Conception of the model - by using the python code line plot_model() which is available on keras library-. . . . 45
4.3.13 Comparison of the first channel of the ECG of a Healthy control and a subject who had a Myocardial infarction. . . . 46
4.3.14 Python code of LSTM model used on Detecting MI disease. . . . 47
4.3.15 The Architecture of Stanford ML group. Adopted from[23] . . . 48
4.3.16 Different CVDs on Dataset of Stanford ML group . Adopted from[23] . . . 49
4.3.17 Dataset Describe . . . 49
4.3.18 Correlation Matrix . . . 50
4.3.19 The main factors causing more heart diseases . . . 50
4.3.20 The sex which is most effecting the heart disease . . . 51
4.3.21 The sex vs Cholesterol levels. . . . 52

5.2.1 Result of ECG Denoising based on Autoencoder . . . 54
5.3.1 Accuracy and F1 score of the Heartbeats classification. . . . 55
5.4.1 Accuracy and F1 score of the some CVDs classification. . . . 56
5.5.1 Python implementation of K Neighbors classifier . . . 57
5.5.2 K Neighbors Classifier scores for different K values. . . . 57

.1.1 Home page 60
.1.2 Register page 60
.1.3 Login page 60
.1.4 DashBoard : Statics page 61
.1.5 Doctor profile : to Edit personal infrmations 61
.1.6 Doctor Management : to manage patients 61
.1.7 Patient Management : to add a new patient 61
.1.8 Patient Diagnostic 62
.1.9 Patient Diagnostic Histories - sequenc eof heart disease classes - 62
.1.10 Heart disease class : using our deep learning model 62
.1.11 Tables of our database 62
.2.1 Home interface 63
.2.2 Login interface 63
.2.3 ECG Analysis interface - notification from doctors- 64
.2.4 Doctor's Advises - notification from doctors- 64
.2.5 The Nearest Hospital 65

List of Tables

3.2.1 ECG Features and their normal duration[27]. 21

4.3.1 Summary of mappings between beat annotations and AAMI EC57 categories. Adopted from [28] 35
4.3.2 Diagnosis classes in the PTB ECG Database. Adopted from [41] 35
4.4.1 Database illustrate our future vision 53

5.3.1 Comparison of heartbeats classification results. 55

Abbreviations

CNN **C**onvolution **N**eural **N**etworks
CVD **C**ardio**v**ascular **D**isease
DL **D**eep **L**earning
DM **D**ata **M**ining
ECG **E**lectro**c**ardio**g**ram
e-HP **e** **H**ealth **P**latform
IOT **I**internet **O**f **T**hings
LSTM **L**ong-**S**hort **T**ime **M**emories
ML **M**achine **L**earning
MI **M**yocardial **Infarction**
NN **N**eural **N**etworks
OS **O**perating **S**ystem
RL **R**einforcement **L**earning
RNN **R**ecurrent **N**eural **N**etworks
RPi3 **R**aspberry **Pi3**
WHD **W**earable **H**ealth **D**evices
WT **W**earable **T**echnologies

Abstract

Information technology has virtually altered every aspect of human life in the present era. The application of informatics in the health sector is rapidly gaining prominence and the benefits of this innovative paradigm are being realized across the globe. However, this evolution has a real challenge of how to get maximum out of the patient's data wherever already available, which can employed by high computer technologies such as data mining, deep learning and wearable health technologies, and turned into useful information and knowledge. This data can be used to develop expert systems to save expert clinicians, help in diagnosing some life-threating diseases such as heart diseases and predicting a different diseases before that happening like heart attack, all of this with less cost, processing time and improved diagnosis accuracy by decreasing the number of misdiagnosis.

In this study, we have chosen the heart disease as a case of study due to its direct influence on the human life and the high number of death caused by this type of disease. This study aims to develop a health informatics system for the prediction of heart diseases using data mining, deep learning and wearable health technologies. We worked on the ECG signal and risk factors of the heart state regarding to its capabilities to classify a different heartbeats, some heart diseases based on ECG Arrhythmia and detecting Myocardial Infarction disease as known as heart attack. Data used are gathered from three sources, the first one is the e-Health platform mounted on Raspberry Pi3, which allows to generate and record ECG signals which is a sequence time series, the second source are records provided by Physionet website and the source are patient's life data that are the main risk factors of the heart state similar as heart disease data provided by UCI Machine learning website. We propose in this memory to use CNN, LSTM and k neighbors in the training and testing steps, in order to classify heartbeat types. some CVDs and predict a heart disease. The scored results are quite acceptable, however some adjustments can be introduced to the way of collecting data from patients. Despite this, the trained model still improve high capabilities on classifying heart beat types and Arrhythmias.

Chapter 1

introduction

1.1 Overview :

Information and communication technologies have revolutionized the living way of people in the 21 st century. It would not be wrong to say that information technology has altered virtually every aspect of human lifestyle in the present era. Health informatics is the study of resources and methods for the management of health information. This area of study supports health information technology, medical practice, medical research and medical informatics. The application of informatics in the health sector has been rapidly gaining prominence and the benefits of this innovative paradigm are being realized across the globe.

The importance of using analytics in healthcare, or healthcare informatics as it is popularly known as, has gained a significant importance in the last few decades, due to the capabilities offered by discipline of health informatics as acquiring, saving, securing, analyzing and extracting meaning full insights from row health data; this improves patient's diagnoses and treatment. Health informatics also facilitates proper management, analysis and use of health-related data for more efficient healthcare delivery and service to the clients and patients. The inherent philosophy of health informatics is to transfer greater control of healthcare to the care providers and patients [3].

Other terms such as clinical informatics and health information management are also very popular and they all focus on the aspect of incorporating the power of information technology into modern health practices and medical data management. Even though there are different names for the concept, the inherent philosophy of all of them is essentially the same. Whether it is called health informatics, biomedical informatics, or clinical informatics, it represents a process through which data is analysed and utilised to generate knowledge that can be applied successfully to address clinical problems and facilitate rapid health care delivery in a highly time-sensitive manner [12]. Figure 1.1.1 below is an informatics pyramid highlighting the intricate relationship between data, information and knowledge that all the above-mentioned fields or paradigms subscribe to.

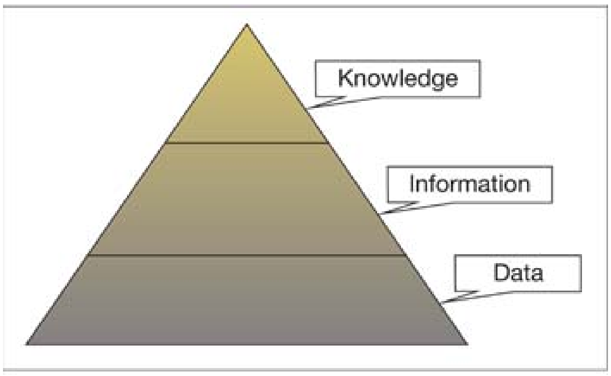


Figure 1.1.1: Informatics pyramid [12]

The Electronic Health Record (EHR) is a systematic collection of electronic health data about individual patients or populations. It is capable of being shared across healthcare providers in a certain state or throughout a country [21]. Health records may include a range of data including general medical records, patient examinations, patient treatments, medical history, allergies, immunization status, laboratory results, radiology images, and some useful information for examination. This rich information data set may help researchers in examining and diagnosing diseases using computer techniques. Using EHRs can help in reducing the cost of legacy systems, improving the quality of care, and increasing the mobility or sharing of records.

The existence of EHRs have encouraged researchers to the idea of electronic healthcare system where the components of the legacy healthcare systems come together and electronically share and transfer patient information across the public infrastructure across a country.

Today, the World, is moving fast toward electronic healthcare information systems. This movement will lead to the production of huge health-related information and data that can be a great asset in increasing overall quality of healthcare and well being of the people, if used judiciously. The aim of this current work is to investigate the aspects of utilising health data using deep learning and data mining techniques to better diagnose and predict the heart failure.

Heart disease or cardiovascular disease (CVD) is the class of diseases that involve the heart or blood vessels (arteries and veins). According to the World Health Organization (WHO), CVDs are the number one of cause of deaths occur in low- and middle-income countries and today in many countries heart disease is viewed as a "second epidemic," replacing infectious diseases as the leading cause of death.

Early diagnosis of heart diseases can help reduce the rate of mortality. One of the ways to diagnose heart diseases is by diagnosing the heart rhythms,also know arrhythmias -is a representation type of CVDs that refers to any irregular change from the normal heart rhythms- from Electrocardiogram (ECG) signals , also by Transoesophageal Echocardiography (TEE) and the health data of patient life histories that known as the risk factors of heart state.

Somtimes, person's might have symptoms and know that have Arrhythmia [1], but other times you may not have symptoms and still have arrhythmia . to do the diagnostic they do ECG test, TEE diagnostic or health life data analysis. this test is often done for a few minutes in the hospitals and the finding is basically that in a few minutes we can't really capture a person's abnormal heart rhythms and can't helping to predict the heart failure.

The analysis of ECG signals, TEE and the patient data histories which are voluminous, varied and very moving data by doctors is time consuming and this is in concomitant with the shortage of experts possessing knowledge on the analysis of it. Because the evolution of Medicine ,can't stay in the traditional doctor's method .For that needs to new technologies to improve Doctor's diagnostic performance.But How ?

Traditional computer science aided is routinely used by doctor despite multicenter studies showing these programs do not improve their diagnostic performance. but always computer science researchers try to find new techniques to live in healthy world. Thus, automated methods can solve limitations of traditional diagnostic methods and provide medical knowledge for diagnoses purposes. To solve this and many other problems in the health sector related to heart diseases diagnosis, one must come up with a way to extract hidden information from enormous data sets that are collected in the past. If we talk about new techniques we should to talk about Data Mining, Ai especially deep learning and IOT especially wearable technologies. recent developments in DM, DL and WTs open possibilities for creating new generation that can help doctor's in their diagnostics and safe life of patients make it decentralized on the hospitals from those enormous data which can be used in this data sets readings.

Data mining has recently become one of the most progressive and promising fields for the extraction and manipulation of data to produce useful information.

Machine learning allows systems the ability to automatically learn and improve from experience without programming explicitly. This is an application of artificial intelligence. Machine learning methods are often classified as supervised learning, unsupervised learning, semi-supervised learning as knows as reinforcement learning.

The Internet of Things (IoT), a rapidly expanding technology area that is shaping up to bring the next revolution in information systems and computing technologies in general. It is upon us. Sensors and embedded devices in automobiles, phones, watches, supermarkets, homes, roads and bridges, appliances and industrial and farm equipment, and wearable technology are already making new kinds of information available and changing the way

[1] is an irregular heartbeat. It occur when the electrical impulses in the heart that coordinate the heartbeat do not work properly. These make the heart beat in a way it should not, whether that be too fast, too slowly, or too erratically [20]

information is produced, consumed, and experienced. IoT obviously represents a great opportunity for advances in information analysis [40].

Internet of Things (IoT)-based medical devices -or the Internet of Things for healthcare, or the Internet of Medical Things- can get patients out of the hospital more quickly, or keep them out altogether, and save organizations money. On the other hand, interconnectivity can provide for easy data collection, asset management, OTA updates and device remote control and monitoring [1].

The connections between IoT especially wearable technologies and data storage and processing as well as data mining and machine learning in healthcare sector especially EHR are obvious and gaining attention already [4].

In order to know how DM, DL and Iot are helping to deal with this new healthcare data generation, we have selected heart disease as a sector of study,in the aim to understand the benefit of using this High new technologies on the heart data. The ECG signal is considered as a the main case of study and we focusing on it, and analysis the life data histories of patient, to combining with the both results whish is our new vision to predict heart attack.

In this work, we will classify the different heart beats and classify the different CVDs within this ECG arrhythmia in a real time ,and trying to analysis the health life histories of the patient which will help to predict the heart failure.

1.2 Why this title?

For me, As a Computer Science student which always working hard to helping people live on beautiful and healthy world. And we always cared about the most important and then the most important. for this challenge, we can't be proved this only in the health sector. For that, I'm interesting and focusing to apply the new technologies on the healthcare field.

Because saving human life is a gold investment. For this, it is no better than trying to diagnose heart diseases and early prediction of the heart failure because it is the number one cause of death for both men and women in the world.

1.3 Motivation :

because Our study have a big challenges :

- **Challenge 1 :** in a few minutes of ECG test ,doctor's can give a wrong diagnostic like can said : the person p have a Atrial Fibrillation disease in this few minutes but he is really before the test have normal heart and after the test he have also the normal heart ; for that we can't really capture a person's abnormal heart rhythms.This is really motivates why we don't diagnose in hours to improve the performance of doctor's diagnostic .

- **Challenge 2 :** amount of data capture : 1,6 million heart beats in 14 days , it's hard for cardiologist to do the tedious job of ECG interpretation and there are a very few doctors who'd be willing to go through two weeks of ECG reading for each patients, this is really motivates why we need automated interpretation.

- **Challenge 3 :** Difficult to diagnose with tow leads or single lead ECG , because Normally diagnosis with 12 leads ECG – in the hospitals is from 6 to 12 lead ECG in cause what doctors need –

- **Challenge 4:** Differences between heart rhythms are often subtle .

- **Challenge 5 :** The integration of EHR and Iot provide the Collection of data remotely -decentralized on the hospitals - and support long-distance clinical healthcare and the telemonitoring.

- **Challenge 6 :** The integration of DM, DL and Iot provide a real time heart disease analysis which being a very big challenge.

- **Challenge 6 :** We know that in the hospitals we don't have a many cardiologists, just there are a general doctors, so this project can help general doctors to do a perfect heart diagnosis.

- **Challenge 7 :** Combination with the ECG analysis and the health data life patient's histories which we suppose that can predicting heart attack with a high accuracy.

- **Challenge 8 :** The gold challenge of this study is to introduce a new sequence data set of the heart life analysis in the reinforcement learning field which containing a label sequence of a long time data of ECG signal with the diagnostic class of ECG signal , health life patient stories. the class of this data set is patient died or not. The main goal of this database is to predicting the heart attack before hours or may be days, which would be beneficial to heart diseases researchers.

1.4 Organization of the Thesis :

The study is organised into six chapters, each focusing on different features of the study work. The following is a summary of the contents of each chapter.

Chapter 1 provides an overview of the importance of the information especially the health information and the heart diseases detection and a more detailed investigation of the problems. My reasons to choose this title, the motivations which have led to this study and my future vision in this case study of heart disease.

Chapter 2 considers the state-of-the-art of the Wearable Technologies and the use of DM and DL on WHDs. Also, there is section about the related works in this field and a section part which proposed in the field of heart diseases,data mining and deep learning. In addition to other classification methods proposed for heart diseases diagnosis and the real time analysis.

Chapter 3 describes the available information on Heart Diseases from clinical view and a backgrounds about DM , DL and their process and techniques. It is divided into separate sections which has more sub-sections. In the section 3.1, there is details about the background study in this field. Section 3.1 is divided into 3 subsections where subsection 3.1.1 contains details about the structure of the heart with the diagram of a human heart, subsection 3.1.2 contains information about the cardiac cycle, that is how blood flows inside a human heart and the changes that take place in the cardiac muscle during this time, with a diagram of the phases of the cardiac system and subsection 2.1.3 contains details about the process of a heartbeat with another diagram to explain it better. In the Section 3.2, there is details about the cardiac background that includes description about the ECG, how the 12 leads are connected to the body to do ECG and also the duration table for each section in the ECG PQRST graph. In the Section 3.3, there is details about the heart diseases particularly: Myocardial Infraction in the section 3.3.1 . In the Section 3.4 describes the main parameters that can predict the heart attack. In the section 3.5, it gives a look at the DM, DL process and explains some techniques used by the both.

Chapter 4 describes our global and detailed vision of our system and try to give and explain the different steps allowing the realization of our project . It is also devided into separate sections. In the section 4.1, there is a background about this chapter. In the section 4.2 contains details about the global conception of our study. In the section 4.3 which has more sub-sections,in the section 4.3.1, explains how we collect and acquire the data from patient with the mobility aspect and learning data sets that gathered from PhysioNet and UCI Machine learning web sites. In the section 4.3.2, is about the work environment. In the section 4.3.3 contains the conception and realization of the ECG signal pre-processing (denoising) using deep Autonecoders. In the section 4.3.4, contains the detailed conception model and the realization of ECG heart beats classification using CNN . In the section 4.3.5, gives the detailed conception model of Myocardial Infraction using LSTM, some CVDs Classification based on two-leads ECG arrhythmias, Heart disease prediction and the realization steps of each phase. In the section 4.4., try to to give the vision model of the heart attack prediction system and how to realize it .

In chapter 5, there is information about the results we found and discussion on them. We have also compared our result with some works that did in the recent few years to get a better overview.

Chapter 6 which is the last chapter contains details about the conclusion and future works that can be done under this topic.

At the end, an appendix that contains some captures about our website and android application. Finally all the references that we used are given.

Chapter 2

The State of the Art

2.1 Introduction :

Today, the world is witnessing an unbelievable development in the field of mobile technology, especially the wearable sensors. Several domains have taken great advantages of those wearable sensors especially the Healthcare industry.

Furthermore, the great forces of DM and DL in different domains especially medicine which have great chance in the recent years and achieve a comparable accuracy with the expert of medicine in neurology, radiology and cardiology.

Nowadays, the combination of DM and DL Algorithms with the Wearable technologies (WT) is really big challenge which allowing the follow, the record, the process and the mine of the patient vital sign remotely, continuous ambulatory monitoring of human vital signs during daily life (during work, at home, during sport activities, etc.) or in a clinical environment, with the advantage of minimizing discomfort and interference with normal human activities[14], long time record and in real time. These improvements are leading to increase the decentralized of the diagnostics on hospitals which will help to predict different diseases before being sick it for hours, months or even years, safe human life and help doctors on epidemiology to know the really causes and the risk factors of the disease.

In this chapter, we will see the trends of the WHDs Markets and try to examine some relevant works in the domain of the wearable sensor of vital signs. As our spot is the heart disease as case of study in this work, we will outline some of relevant works in the field of heart disease analytics, and try to spotlight the more common DM and DL tasks, approaches and algorithms used to diagnose the different CVDs.

2.2 Wearable Health Devices Market Trends :

Wearable devices market value is in constant grow and this year it is estimated to reach a value of approximately $12 billion. It is a market that is in constant growth, if we think that in 2010 the market was only $6.3 million, it is possible to understand that in these recent years it has increased substantially (around 200 hundred percent). According to IDTechEx, in terms of global revenue, the following five year's trend is to increase at a higher rate as it can be seen in figure 2.2.1 [15] which that Horizontal bar graphic showing the total revenue in billions ($) (left axis) from 2015 to 2017, and estimated until 2026. The blue line shows the revenue growth rate in billions ($) (right axis).

Figure 2.2.1: The total revenue the growth rate in billions (Adapted from[24]

Wearable devices market can be categorized according to the classification shown in Figure 2.2.2. According to this study made by ABI Research , from 2017 to 2019 the use of wearable devices in healthcare will constantly increase. This fact is a good indicator for WHDs companies that have the aim to develop products for healthcare applications.

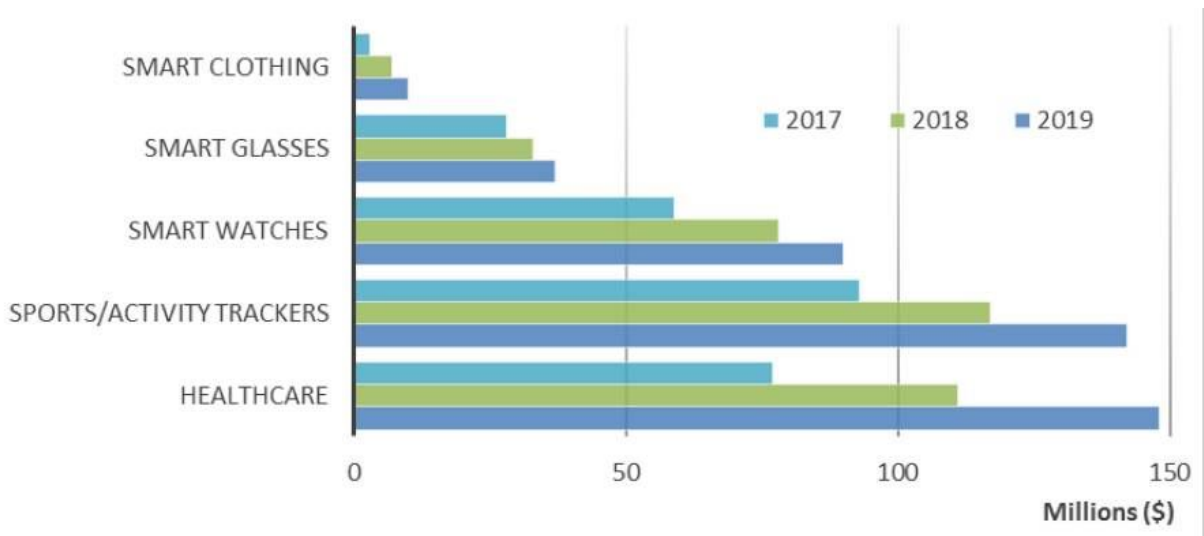


Figure 2.2.2: Horizontal bar graphic showing the trend of global market value of wearable computing devices, in millions, between 2017 and 2019[15].

With the technology and internet of things (IoT) revolution, the healthcare wearable devices segment is increasing and with this the telehealth sector is also rapidly changing. In 2014 it was predicted that this year the revenues of telehealth devices and services reach \$4.5 billion, which is almost the double of the 2017 value of \$2.8 billion. The number of home healthcare monitoring devices connected to a data center also has a growing trend considering the past years (2.2.3 A) according to a study made by Berg Insight. In this study it is possible to understand the evolution of this trend in the sector of home medical monitoring devices: diabetes care devices; blood pressure monitors; multi-parameter patient monitoring; apnea and sleep monitors; holter monitors; and heart rate meters. Although most of these device cannot have wearable features, it is possible to conclude that this is a continuous growing market segment, creating a market opportunity for the growth of WHDs in home healthcare for following years.

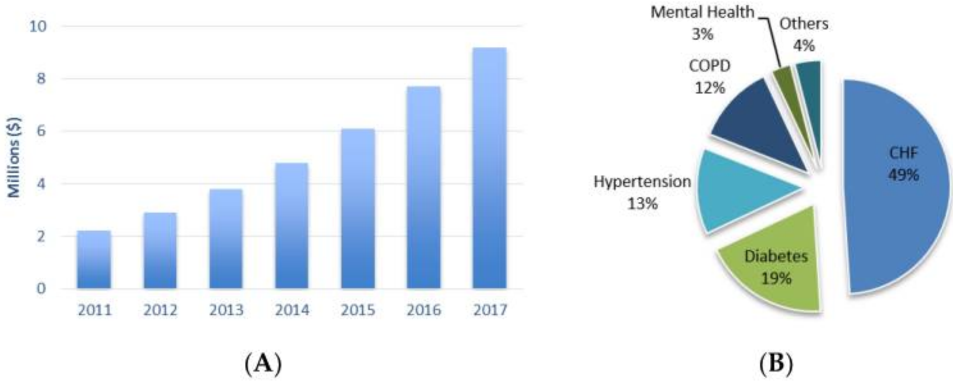


Figure 2.2.3: (A) Connected home medical monitoring devices (in millions) 2011 to 2017; (B) The world market for telehealth from 2014 divided in the main areas (CHF-congestive heart failures; COPD-chronic obstructive pulmonary disease) [15]

Healthcare ambulatory monitoring segment, according to a study made by IHS Inc. (London, UK) can be divided in several areas according to the diseases or the type of monitoring, resulting in five main areas: (2.2.3 B): congestive heart failures (CHF); chronic obstructive pulmonary disease (COPD); diabetes; hypertension; and mental Health. This study conclusion is that mobile telehealth solutions are going to become the standard in remote patient monitoring, leading to a larger market for the WHDs [15].

2.3 Automatic system and Vital signs (General Works):

The human body has multiple different physiological signs (figure 2.3.1) that can be measured: from electrical signs to biochemical, human biosignals are possible to be extracted and be used to better understand the bodily health status and reaction to external factors.

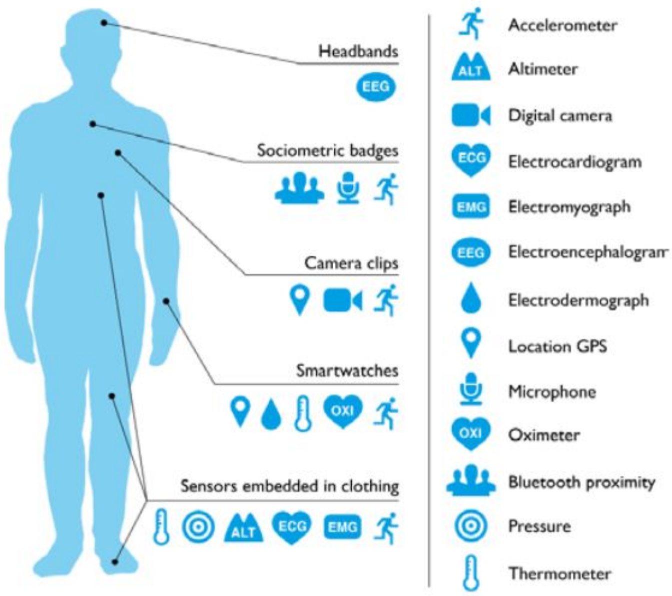


Figure 2.3.1: Healthcare wearable sensors.

Nowadays, Automatic system and wearable technologies let us to classify WHDs according to three aspects (figure 2.3.2): scenario of use (home/remote or clinical environment); the type of monitoring (offline or online); and the type of user (healthy or patient)

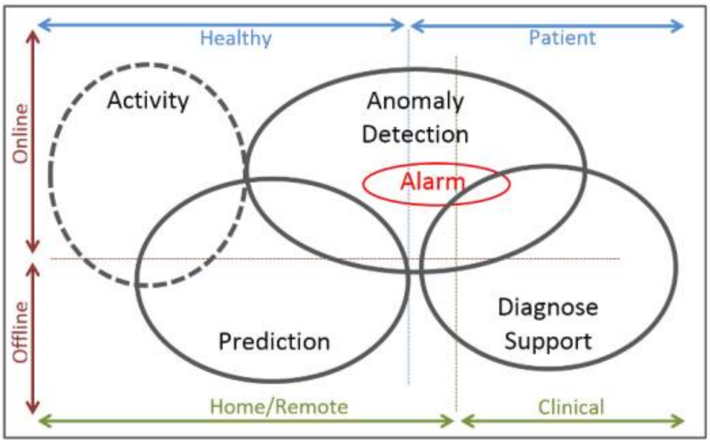
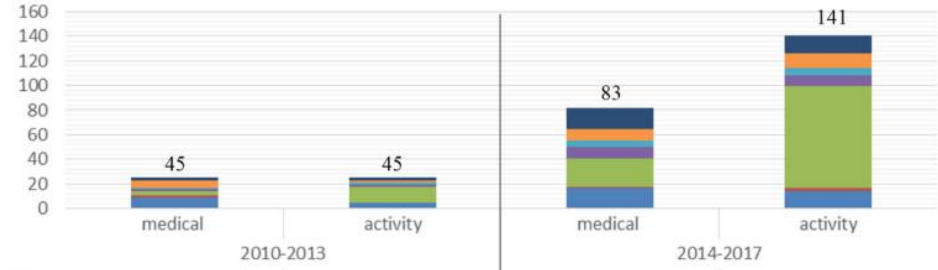


Figure 2.3.2: Schematic overview of the four main Analytics processes (activity, prediction, anomaly detection and diagnose/decision support) in relation to different aspects of wearable sensing in wearable health devices. Filled line- Medical purposes; Traced line: Activity purposes.

2.4 Related works :

We live in an era where computation is being moved from vast centralized servers to PCs and cloud. For a long time researchers have been working on identifying and predicting different diseases using machine learning and wearable technologies. Exploration supervised learning, unsupervised learning and reinforcement learning, As we see in figure 2.4.1 that has been the most works are on the main tests of heart disease which are HR and ECG.



	2010-2013 medical	2010-2013 activity	2014-2017 medical	2014-2017 activity
BT	2	2	17	15
BP	6	2	10	12
RR	1	2	5	6
BG	2	1	9	8
HR	3	13	23	83
SpO2	2	0	1	3
ECG	9	5	17	14

Figure 2.4.1: Number of retrieved scientific papers related with WHDs and a specific physiological sign. The distribution is divided in two time intervals and according to the purpose: medical or activity. (BT—Body Temperature; BP—Blood Pressure; RR—Respiration Rate; BG—Blood Glucose; HR—Heart Rate; SpO2—Blood Oxygen Saturation; ECG—Electrocardiogram) [15].

Here some related works ECG in the field of Signal denoising, Heartbeats classification, ECG arrhythmia classification and heart disease prediction based on risk factors :

2.4.1 ECG Denoising :

Due to nature of the human body and the way of how ECG is measured, the signal is exposed to several types of noise from various sources like muscular activities, power-line, skin stretching and electrode motion, movement of heart due to respiration, etc. ECG filtering is aimed to eliminate the unwanted signals from an ECG record, without hampering the clinical information contained within the signal itself[48].Denoising or filtering ECG data is a known problem with relatively long history; a number of techniques therefore exists. Common approaches are listed in [46].They include finite and infinite impulse response filters, wavelet-based methods, filtered reside technique and empirical mode decomposition[7]. For a long past years,most researchrs on ECG filtering that based on Fast Fourier transformation (FFT) and Wavelet transformation which have an excellent accuracy but need a long time proccessing.

In recent years, one can observe an appearance of novel approaches to signal filtering, utilizing machine learning methods and neural networks. For example, Moein [35] investigated multi-layer perceptron networks for ECG noise removal. For training, a relatively small dataset of 100 signal samples was used. The expected out-puts were produced by denoising input signals using Kalman filter. The network was able to achieve error rate less than 0.5 for all of them. However, it is worth noting that due to the nature of the dataset, the network learned in fact to simulate the Kalman filtering. Therefore, by training the network this way, one cannot achieve better performance than the Kalman filtering itself.

An other approach to neural network based noise reduction is described in [38].It utilizes both neural networks and wavelet transform, in a form of Wavelet Neural Networks (WNN). Such networks are a special kind of three layer feed forward neural networks, employing a set of wavelets as activation functions. The network training is a two phase process; first, using 400 iterations of specialized algorithm – Adaptive Diversity Learning Particle Swarm Optimization (ALDPSO) that performs global search in the population of 20 candidate networks. The second phase are 1600 iterations of gradient descent of the best performing network from previous phase. The training and test data are real signals from PhysioBank database [7].

In the last few years, the combination of CNN with Deep Auto-encoder which this last have a luck of use on denoising on different signals such us audio, biomedical signals, etc .

2.4.2 ECG Heartbeats Classification :

Needless to say, classification of heartbeats is a challenging problem. This is due to near chaotic behaviors observed in heart abnormalities. ECG signals of heartbeats are characterized by features known as the P-wave, the QRS complex, and the T-wave. Typical features in classification include signal samples from the primitive features (P,QRS, T) and mathematical transformations thereof such as Fourier and Wavelets[30]. Automatic high-accuracy methods for R-peak extraction have existed at least since the mid 1980's [37]. Current algorithms for R-peak extraction tend to use wavelet transformations to compute features from the raw ECG followed by finely-tuned threshold based classifiers [32, 33]. Because accurate estimates of heart rate and heart rate variability can be extracted from R-peak features, feature-engineered algorithms are often used for coarse-grained heart rhythm classification, including detecting tachycardias (fast heart rate), bradycardias (slow heart rate), and irregular rhythms.However, such features alone are not sufficient to distinguish between most heart arrhythmias since features based on the atrial activity of the heart as well as other features pertaining to the QRS morphology are needed.[39] Before deep learning became applicable, much work has been done to automate the extraction of other features from the ECG. For example, Q. Zhao and L. Zhang applied in [49] two different feature extraction methods to obtain the feature vector of ECG data. The wavelet transform is used to extract the coefficients of the transform as the features of each ECG segment, then support vector machine (SVM) is performed with Gaussian kernel to classify different ECG heart rhythm. Artificial neural networks have also been used for the task of beat detection [34]. While these models have achieved high-accuracy for some beat types, they are not yet sufficient for high-accuracy heart arrhythmia classification and segmentation[39].

Researchers in the recent few years have focused their efforts on using on deep learning models for ECG features extraction. In [47], researchers develop an algorithm for automatic ECG features extraction based on Deep Belief Network (DBN). Along with th development of DL models especially CNN that give the chance to process the ECG data with a maximum accuracy. For example, Acharya et al [5] applied a CNN model with data augmentation, also Mohammad kachuee et al [28] applied a deep residual CNN based on a transferable representation.

2.4.3 ECG Arrhythmia Classification :

The power of DM and ML give the chance to classify a lot of diseases such us heart disease especially arrhythmia supervised classification that is the most decisive on CVDs analysis. For example, researchers [22] have proposed using a combination of SVM and DBN, in which DBN is used for feature learning and SVM is then applied for the classification tasks using the learned features. also Recent works Kiranyaz et al., 2016 [29]; Al Rahhal et al., 2016 [6]; Rajpurakar P et al.,2017 [39]; Awni Y et al., 2019 [23] started applying DNN models on ECG signals for arrhythmia classification and achieved good performance. The both works of Stanford researchers in [23, 39] which that have a gold challenge and motivate me to working in this project because, Rajpurkar P , Awni Y et al. have shown that this approach achieved cardiologist-level performance using a dataset that's 500 times larger than others of its kinds which is really motivate, they use a private dataset which contain 300 million ECGs records. in their research develop a model from deep residual network in [25], where a 34 layer deep CNN is applied directly, without adopting any complex preprocessing and feature engineering step.

2.4.4 Heart disease prediction based on DM techniques:

In the recent years this part doesn't have a luck studies by compaines or researcher teams, the most of work are making by developers using SVM, Decision Tree, Random forest and K Neighbor Classifiers.

Chapter 3

Background about heart disease , Data Mining and Deep Learning Techniques

Data Mining and Machine learning are the study of making computers act using the knowledge gathered from historical data without being expressly modified. This chapter contain details about the background study, structure of heart and its cardiac cycle, how a heartbeat occurs, cardiac background, how to read an ECG graph and background about heart diseases and heart attack. Furthermore it also has information about DM and DL .

3.1 Heart :

3.1.1 The structure of the heart:

The heart is a hollow muscular organ that lies in the middle of the chest cavity and is surrounded by a double membrane known as the pericardium. This protects the heart and facilitates its pumping action. The heart is responsible for controlling the circulatory system in mammals and other animals. The heart is divided into four chambers internally. Two Atria (singular : atrium) and two ventricles. The upper chambers are the atria (left atrium and right atrium) and the lower chambers are the ventricles (left ventricle and right ventricle) [45]. Muscular walls (septum) divide two sides of the heart. Each atrium is connected to its ventricle by a opening that is guarded by a valve (the bicuspid valve on the left and the tricuspid valve on the right). The bicuspid valve which is also known as mitral valve is made up of two flaps and the tricuspid valve has three flaps. These valves are also known as atrioventricular valves and they prevent backflow of blood into the atria. All the four chambers have a smooth membranous lining internally known as the endocardium. The walls of the heart are made up of cardiac muscles which is unlike any other muscle in our bodies. It never gets fatigue like skeletal muscle. The cardiac muscle fibres contract and then relax again about 70 times a minute on average [26].

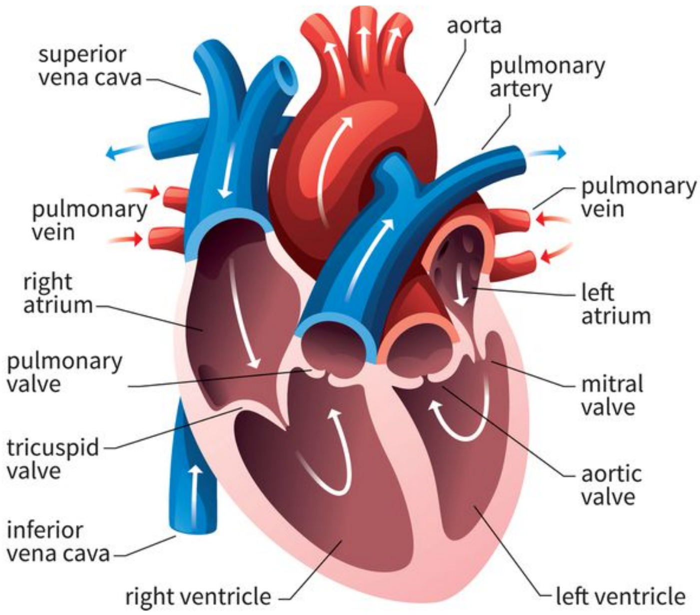


Figure 3.1.1: The Human Heart

The heart is attached to four blood vessels which are Aorta, Vena cava, Pulmonary Artery and Pulmonary vein. The vena cava is connected to the right atrium and pulmonary vein is connected to the left atrium. Similarly, the aorta is connected to the left ventricle and pulmonary artery is connected to the right ventricle. The ventricles are separated from the aorta and pulmonary artery by means of semilunar valves. These valves are also responsible for preventing backflow of blood into the heart. The atrial walls are thinner than the wall of the ventricles since ventricles need to withstand more blood pressure.The vena cava supplies de-oxygenated blood from the body, which then flows into the right atrium then the right ventricle. This gets pumped through the pulmonary artery to the lungs where it gets oxygenated, before returning to the heart via the pulmonary vein. This flows through the left atrium into the left ventricle, and then gets pumped to the body via the aorta. It finally returns to the heart through the vena cava, and the process repeats[31]. Figure 3.1.1 shows a Human Heart.

3.1.2 The cardiac cycle :

The human heart is a pump and it pumps blood around the body at different speeds and at different pressures according to the need of the body. Blood is moved through the heart by a series of contractions and relaxations of the muscle in the walls of the four chambers. The contraction of the heart is called 'systole' and the relaxation is called 'diastole'. The contraction and relaxation together constitute the heartbeat. Cardiac cycle is formed by the cyclical repetition of contraction and relaxations of these muscles and consists of three main stages which are atrial systole, ventricular systole and diastole. The cardiac cycle is explained briefly below

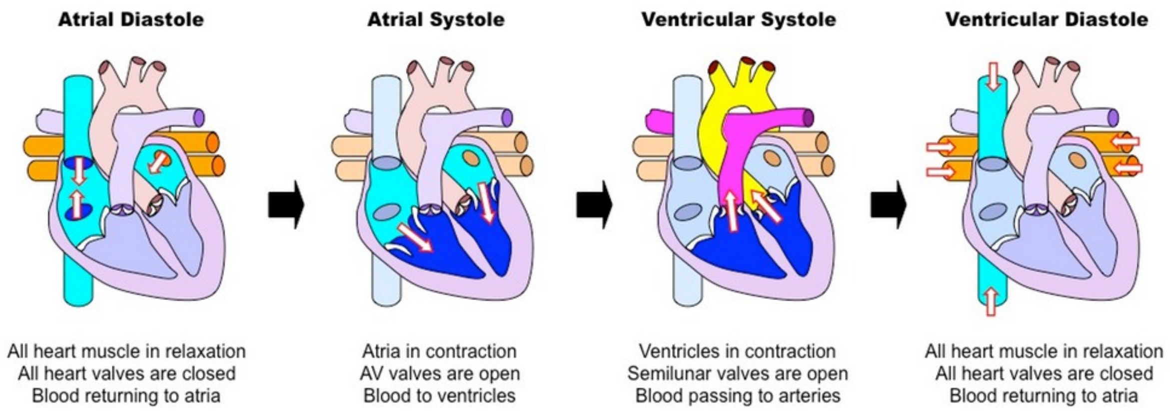


Figure 3.1.2: The phases of a cardiac cycle

As blood enters the atria, it can not pass into the ventricles as the atrioventricular valves are closed and during atrial systole, the walls of the atria contracts and this raises the blood pressure in the atria which forces the atrioventricular valves to open, then the blood then passes to the ventricles through these valves. During ventricular systole, the ventricles contract when they are full and this raises the blood pressure in the ventricles which closes the atrioventricular valves and thus preventing any backflow of blood to the atria, therefore, The pressure continues to increase as the ventricle contracts and this forces open the semilunar valves and blood travels to the two arteries. The pulmonary artery carries blood to the lungs and aorta has several branches that carries blood to all other parts of the body. As blood leaves the ventricles, the semilunar valves are closed due to the increase of pressure in the pulmonary artery and aorta. This happens during the diastole. The cycle then begins again as the atria start to fill with blood. Figure 3.1.2 is an image of the phases of the cardiac cycle [26].

3.1.3 The process of a Heartbeat :

. The pumping of the heart, or the heartbeat, is caused by alternating contractions and relaxations of the myocardium. These contractions are stimulated by electrical impulses from a natural pacemaker, the sinoatrial, or S-A, node located in the muscle of the right atrium. An impulse from the S-A node causes the two atria to contract, forcing blood into the ventricles. Contraction of the ventricles is controlled by impulses from the atrioventricular, or A-V, node located at the junction of the two atria. Then we find special conduction tissue - his purkinje network send the impulse to the rest of two ventricules Following contraction, the ventricles relax, and pressure within them falls. Blood again flows into the atria, and an impulse from the S-A starts the cycle over again. This process is called the cardiac cycle. The period of relaxation is called diastole. The period of contraction is called systole. Diastole is the longer of the two phases so that the heart can rest between contractions [8, 40].

3.2 Cardiac Background :

Number of people suffering from heart disease is increasing day by day mostly because of the unhealthy lifestyle led by people. ECG provides us with series of sinus rhythm which defines the condition of heart. An Electrocardiogram (ECG) is a procedure that shows the electrical activity of the human heart over a period of time. It consists of several sensors placed around the body connected to a monitor. The electrical signals are recorded by the device, by attaching electrodes to the outer surface of the skin. Then the electrodes detect the small electrical changes on the skin that arise from the heart muscle's electrophysiological pattern of depolarizing and repolarizing during each heartbeat [45]

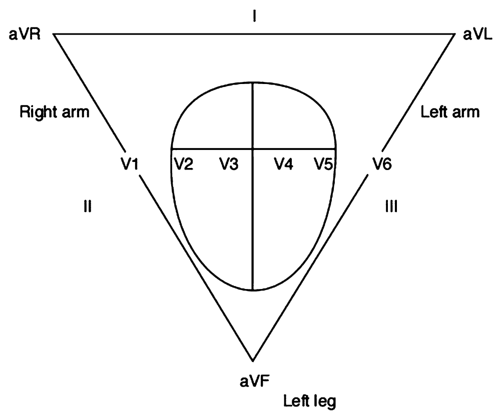


Figure 3.2.1: Schematic representation of 12 leads (electrodes) placed around the heart components [45]

The electrical signals travel through the electrodes to the graph device, which records them as characteristic waves. Different waves reflect the activity of different areas of the heart which generate the respective flowing electrical currents .The electrocardiogram comprises of 12 leads around the heart (figure 3.2.1. For scientific purposes, the heart is at the focal point of a triangle . The cathode situations are assigned as follows:The three appendage drives: lead I joins the privilege and left arms, lead II associates the correct arm and left leg and lead III joins the left arm and left leg. The three enlarged leads: aVL is situated confronting the heart from the correct arm VL from the left arm and aVF from the left foot and these cathodes are set in a frontal plane. The precordial leads (V1–V6): these are put on the front of the thorax and record even driving forces [18].

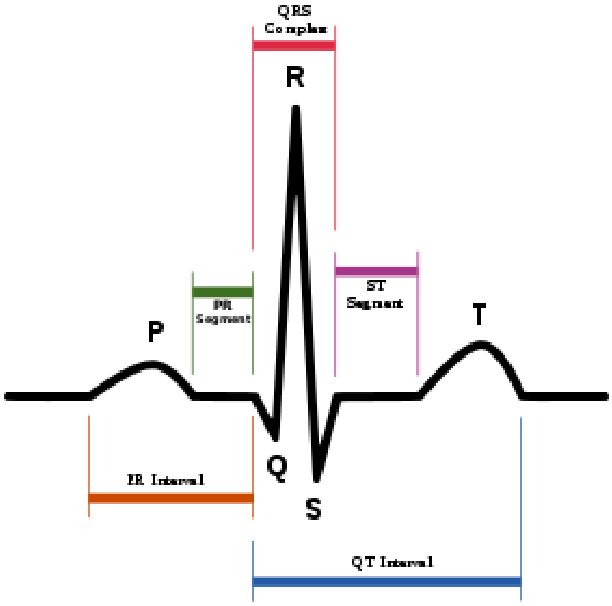


Figure 3.2.2: Normal ECG recording

An ECG cycle consists of 5 waves which are PQRST (fig 5). The P wave corresponds to the atrial depolarization and the pumping of blood from the atrium to the ventricle. The QRS complex usually consists of Q, R and S waves which corresponds to the depolarization of blood and the pumping of blood out from the ventricle to the body and lung. Both the P wave and the QRS complex are depolarization waves. Finally the T wave corresponds to the repolarization of the ventricle and the recovery of the ventricle for next cycle [27].

Feature	Description	Duration
RR	interval between R wave and the next R wave	0.6-1.2 s
P	first short upward movement of the ECG tracing	80ms
PR	measured from the beginning of the P wave to the beginning of the QRS complex	120-200 ms
QRS	normally begins with a downward deflection Q, a larger upwards deflection R and ends with a downward S wave	80-120ms
PR	connects the P wave and the QRS complex	50-120ms
J-point	The point at which the QRS complex finishes and the ST segment begins is called J-point	Not applicable
ST	connects the QRS complex and the T wave	80-120ms
T	normally a modest upward waveform	160ms
ST	measured from the J point to the end of the T wave	320ms
QT	measured from the beginning of the QRS complex to the end of the T wave	420ms

Table 3.2.1: ECG Features and their normal duration[27].

One ECG signal consists of several ECG beats and each ECG beat contains P wave, QRS complex, and T wave. Each peak (P, Q, R, S and T), intervals (PR, RR, QRS, ST, and QT) and segments (PR and ST) of ECG signals have their normal amplitude or duration values [27]. These peaks, intervals, and segments are called ECG features. All these features for one ECG cardiac cycle are described in Table 1.

3.3 Heart diseases :

3.3.1 Myocardial Infarction and heart attack :

A myocardial infarction (MI), commonly known as a heart attack, occurs when a portion of the heart is deprived of oxygen due to blockage of a coronary artery. Coronary arteries supply the heart muscle (myocardium) with oxygenated blood. Without oxygen, muscle cells served by the blocked artery begin to die (infarct). Myocardial infarction can be characterized from a sum of different perspectives related to clinical, electrocardiographic (ECG), biochemical and pathologic characteristics. It is accepted that the term myocardial infarction reflects death of cardiac myocytes caused by prolonged ischemia. Changes in ECG when MI detected are ST rise and ST depression and T-wave changes shown in figure 3.3.1. New ST rise at the J-point in two bordering leads with the cut-off focuses: 0.2 mV in men or 0.15 mV in ladies in leads V2– V3 and additionally 0.1 mV in different leads and new flat or down-slanting [36]

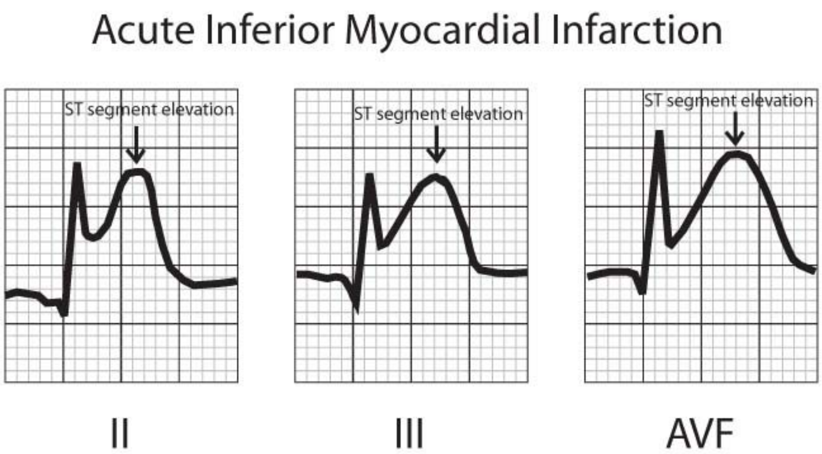


Figure 3.3.1: Myocardial Infarction Interpretation on ECG

3.4 Data Mining and Deep Learning :

3.4.1 DM Definition, Techniques and KDD process :

3.4.1.1 DM Definition :

Data mining,in computer science is the process of discovering interesting and useful patterns and relationships in large volumes of data. The field combines tools from statistics and artificial intelligence with database management to analyze large digital collections, known as data sets [10].

3.4.1.2 KDD process :

In 1996, in the proceedings of the first international conference on KDD, "UM. Fayyad " gave one of the best known definitions of Knowledge Discovery from Data: "The non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data." [19]

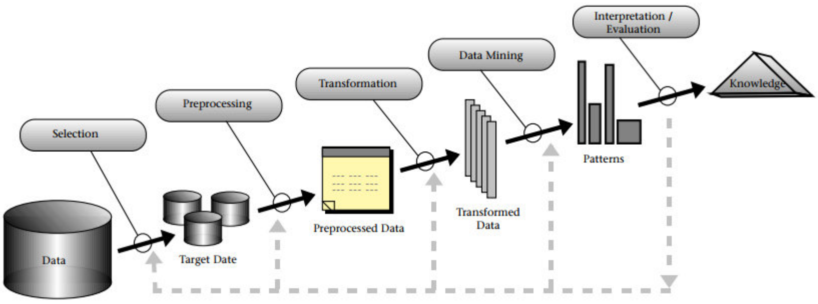


Figure 3.4.1: The KDD process

- **Selection:** ata Selection is the process where data relevant to the analysis task is retrieved from the database.

- **Preprocessing:** Due to the huge size, the multiple heterogeneous sources, and various forms, data is highly susceptible to noisy, missing, and inconsistent. Low -quality data will lead to low quality mining results. In this step, we try to preprocess data to improve its efficiency, and ease the future mining process.

- **Transformation:** In this step, data will be transformed or consolidated into forms appropriate for mining, by performing summary or aggregation operations.

- **Data Mining:** is the main part of the KDD process, here we use effective techniques and tools to mine efficient knowledge .

- **Interpretation:** Once data is mind we get some patterns, not all patterns or results are tolerable, so we have to evaluate the usefulness of results. Otherwise, patterns are presented in the format, which is understood by the user, and precious knowledge can be discovered.[16]

3.4.2 DL Definition and process :

3.4.2.1 DL Definition :

Deep learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in artificial intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network.

3.4.2.2 DL process :

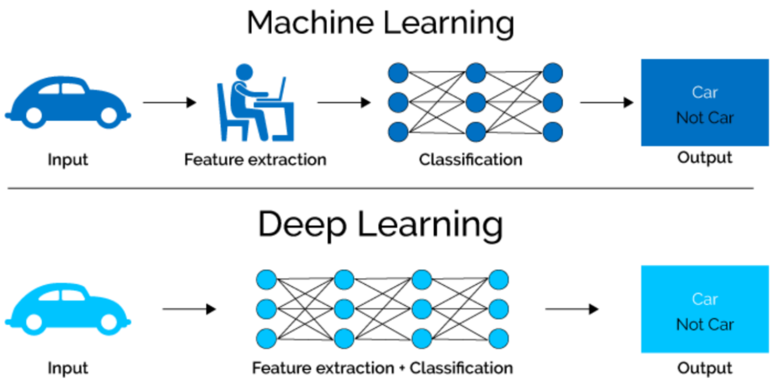


Figure 3.4.2: ML vs DL

As we see in the figure 3.4.3, in the DL we don't need to do the features extraction as like us in the DM and ML.

3.4.3 Techniques and Methods Used :

3.4.3.1 Supervised Learning :

The majority of practical learning uses supervised learning. Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output. Y = f(X) In Supervised Learning, algorithms learn from labeled data. After understanding the data, the algorithm determines which label should be given to new data based on pattern and associating the patterns to the unlabeled new data.

3.4.3.2 Semi-Supervised Learning :

I believe that this area will be the future of high technologies, So what is it? Semi-Supervised Learning or Reinforcement learning (RL) is where you have a large amount of input data (X) and only some of the data is labeled (Y) are called semi-supervised learning problems.

3.4.3.3 Classification :

is one of the most widely used techniques in DM and ML . Is the processing of finding a set of models (or functions) which describe and distinguish data classes or concepts, for the purposes of being able to use the model to predict the class of objects whose class label is unknown. The derived model is based on the analysis of a set of training data. Many algorithms are used in the classification techniques we site for example: Zero R, One R, Decision Tree, Support Vector Machine, Artificial Neural Networks, Convolutional Neural Network and Recurrent Neural Network,LSTM, etc.

In the next we'll give a look about the algorithms that we used in this study.

3.4.3.4 K Neighbors classifier :

K-Nearest Neighbours is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection.

Algorithm Let m be the number of training data samples. Let p be an unknown point.

1-Store the training samples in an array of data points arr[]. This means each element of this array represents a tuple (x, y).

2-for i=0 to m: Calculate Euclidean distance d(arr[i], p).

3- Make set S of K smallest distances obtained. Each of these distances correspond to an already classified data point.

4- Return the majority label among S.

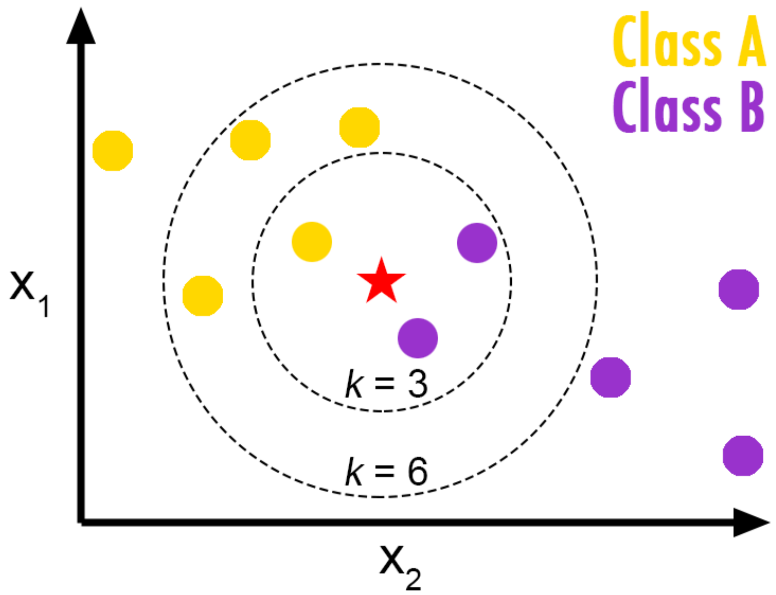


Figure 3.4.3: KNN illustration

3.4.3.5 Neural Networks, CNN , AE and LSTM :

3.4.3.5.1 Neural Networks:

What is a Neuron? neural networks were inspired by the neural architecture of a human brain, and like in a human brain the basic building block is called a Neuron. Its functionality is similar to a human neuron, i.e. it takes in some inputs and fires an output. In purely mathematical terms, a neuron in the machine learning world is a placeholder for a mathematical function, and its only job is to provide an output by applying the function on the inputs provided.

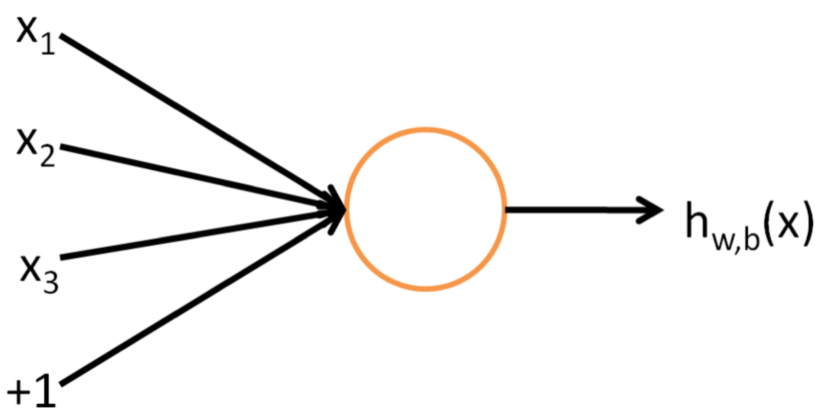


Figure 3.4.4: Neuron in Neural Network.

The function used in a neuron is generally termed as an activation function. The major activation functions tried to date, step, sigmoid, tanh, and ReLU. Researchers have generally found the last function which is ReLU has be performing better than sigmoid or tanh functions.

understanding a Neural Network : Before understanding a NN, it's imperative to understand what is a layer in a NN. A layer is nothing but a collection of neurons which take in an input and provide an output. Inputs to each of these neurons are processed through the activation functions assigned to the neurons. For example, here in figure 3.4.6 is a small neural network.

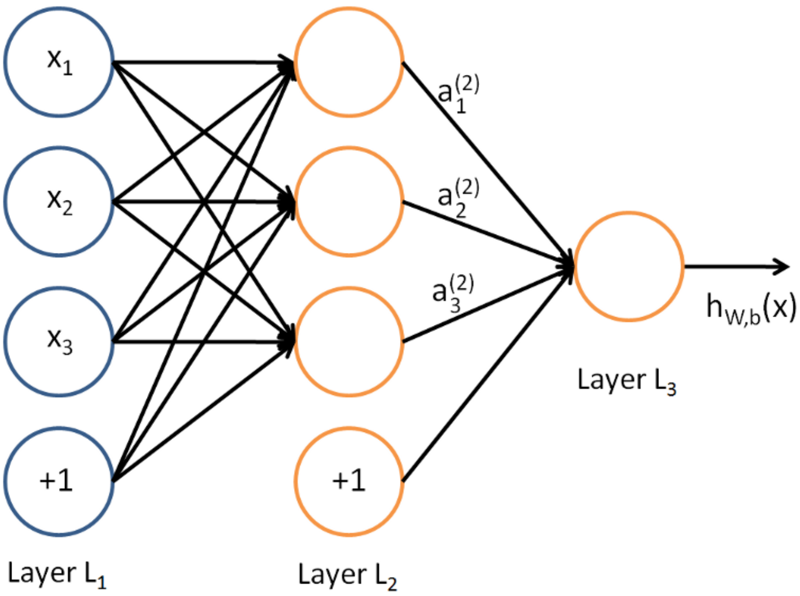


Figure 3.4.5: Small Neural Network Architecture.

In figure 3.4.6, the leftmost layer of the network is called the input layer, and the rightmost layer the output layer (which, in this example, has only one node). The middle layer of nodes is called the hidden layer because its values are not observed in the training set or calling Black Box.
For a neural network to make accurate predictions each of these neurons learn certain weights at every layer. The algorithm through which they learn the weights is called back propagation [42].

3.4.3.5.2 Convolutional Neural Networks (CNN) :

CNN is one of the variants of neural networks used heavily in the field of Computer Vision. It derives its name from the type of hidden layers it consists of. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers, and normalization layers. Here it simply means that instead of using the normal activation functions defined above, convolution and pooling functions are used as activation functions.

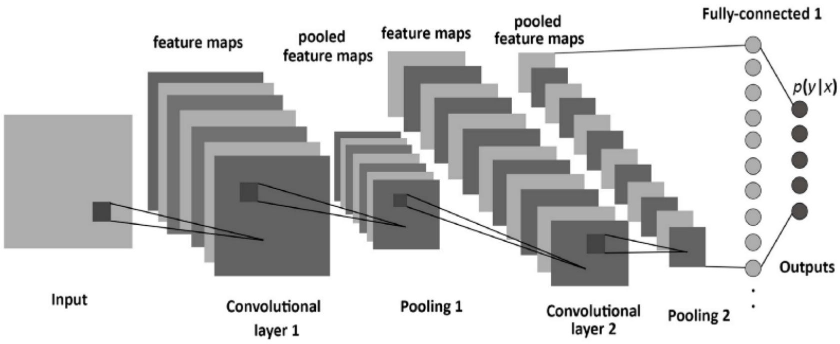


Figure 3.4.6: CNN Architecture.

To understand it in detail one needs to understand what convolution and pooling are. Both of these concepts are borrowed from the field of Computer Vision and are defined below.

Convolution: Convolution operates on two signals (in 1D) or two images (in 2D): you can think of one as the "input" signal (or image), and the other (called the kernel) as a "filter" on the input image, producing an output image (so convolution takes two images as input and produces a third as output).
In layman terms it takes in an input signal and applies a filter over it, essentially multiplies the input signal with the kernel to get the modified signal [44].
Mathematically, a convolution of two functions f and g is defined as in figure 3.4.7

$$(f * g)(i) = \sum_{j=1}^{m} g(j) \cdot f(i - j + m/2)$$

Figure 3.4.7: The convolution f g. Adopted from [44]

Pooling: is a sample-based discretization process. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned[11].
There are two main types of pooling commonly known as max and min pooling. As the name suggests max pooling is based on picking up the maximum value from the selected region and min pooling is based on picking up the minimum value from the selected region. Here the figure below shows max pooling with a 2x2 filter and stride = 2 [11].

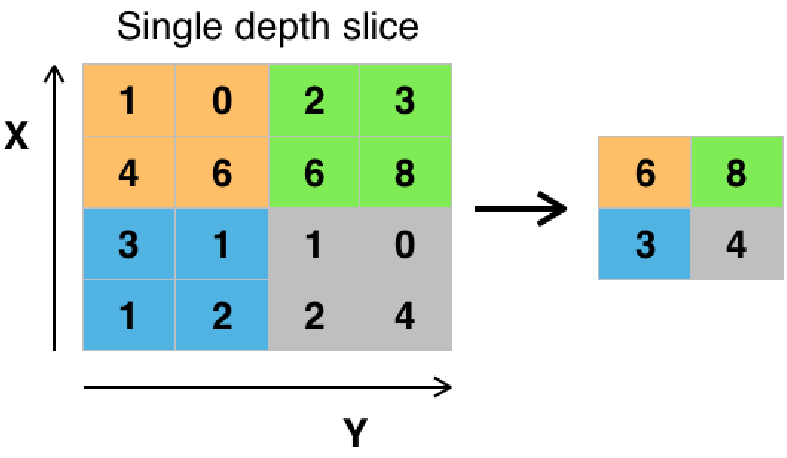


Figure 3.4.8: Max Pooling. Adopted from [11]

3.4.3.5.3 Autoencoders (AE) :

AE are a specific type of feedforward neural networks where the input is the same as the output. They compress the input into a lower-dimensional code and then reconstruct the output from this representation. The code is a compact "summary" or "compression" of the input, also called the latent-space representation.

Figure 3.4.9 below shows that an autoencoder consists of 3 components: encoder, code and decoder. The encoder compresses the input and produces the code, the decoder then reconstructs the input only using this code.

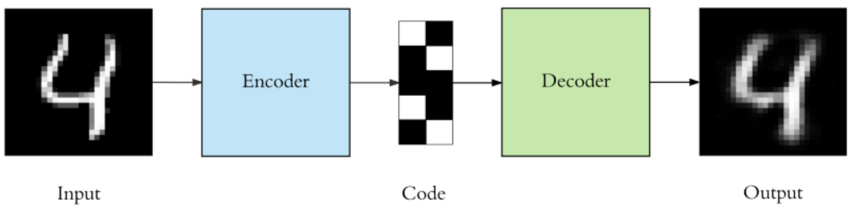


Figure 3.4.9: AE architecture

First the input passes through the encoder, which is a fully-connected ANN, to produce the code. The decoder, which has the similar ANN structure, then produces the output only using the code. The goal is to get an output identical with the input. Note that the decoder architecture is the mirror image of the encoder. This is not a requirement but it's typically the case. The only requirement is the dimensionality of the input and output needs to be the same. Anything in the middle can be played with [13].

Types of Autoencoders :

1. Denoising autoencoder.

2. Sparse autoencoder.

3. Variational autoencoder (VAE).

4. Contractive autoencoder (CAE).

3.4.3.5.4 Long short-term memory (LSTM) :

LSTM units allow to learn very long sequences. The main characteristic of an LSTM is the presence of three gates:

- update gate

- forget gate

- output gate

Below is a schema of an LSTM unit. The three gates are shown by the presence of the three sigmas:

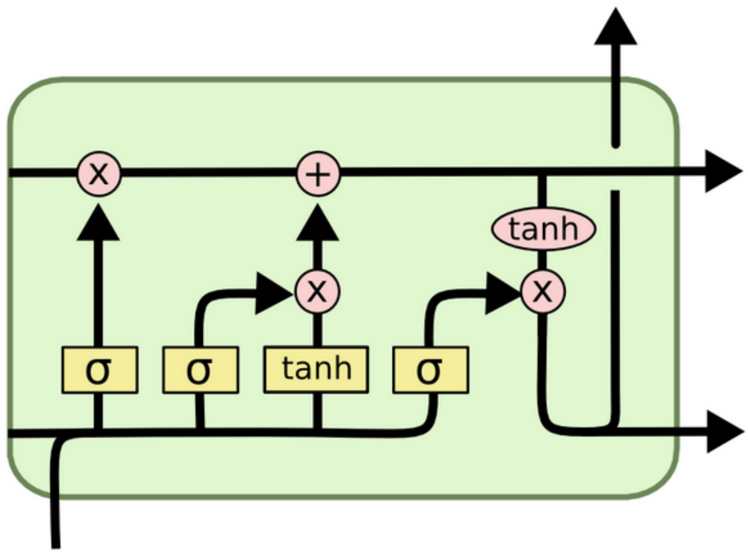


Figure 3.4.10: Schema of an LSTM unit

From the figure above, notice that there is a line from left to right at the top of the LSTM unit. This line represents how earlier information can be passed on to further steps in the network, and this is why an LSTM unit is so good at memorizing long sequences. Consequently, this allows a model to capture longer range dependencies.

This might seem like there is not a lot to know about LSTM, but understand that this is simply a unit within a recurrent neural network. The green box above simply represents one unit of an RNN. Then, multiple units are connected, resulting in a full network [9].

3.5 Conclusion :

In this chapter, we have tried to cover Heart,ECG signal and heart diseases, and covering up to DM, DL, their processes and define their techniques and algorithms that we used in the study.

The next chapter is assigned to the conception and the realization parts.

Chapter 4

Conception and Realization

4.1 Introduction :

As we have seen in previous chapters, our century is witnessing an unbelievable development in the field that combining high great technologies which are DM, DL and wearable technologies on the sector of healthcare, also as we have remarked that the most researches in the recent years on WHDs are focusing on heart disease based on ECG signal and life style healthcare histories.

In this chapter, and as our selection is the CVD as a case of study, we will give our conceptual vision and how can we realize our remote automatic system, my new proposition that can help to predict the heart attack, which is also based on the DM and DL processes. Firstly, before we'll give our conception, we will give a some details about data sets that we are used on this study, then we'll present our conception and our schematic's prototype, with the main process steps, and then in next step, we will give the conception and implementation details of each step which are ECG denoising, heartbeats and Arrhythmia classification, heart disease prediction that based on the risk factors. Finally we'll try to give our view of my new proposition that can help to predict a heart attack before that happen.

4.2 Global Vision and Conception:

Our future vision is to develop an automatic clinical system that combining the four approaches that shows in figure 4.2.1. This is my new proposition that can help to predict a heart attack by create a sequence of Algerian dataset relating to heart disease.

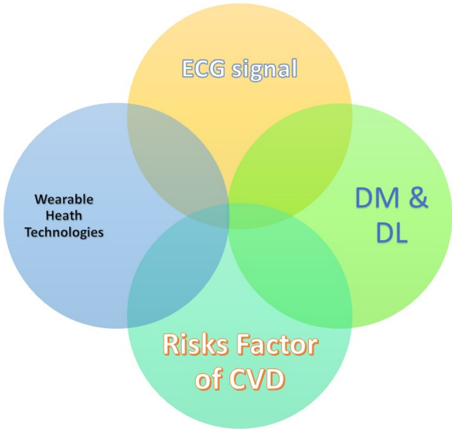


Figure 4.2.1: Approaches of our Vision

In the next, we'll describe our global conception based on KDD and DL processes.

4.2.1 Global Conception :

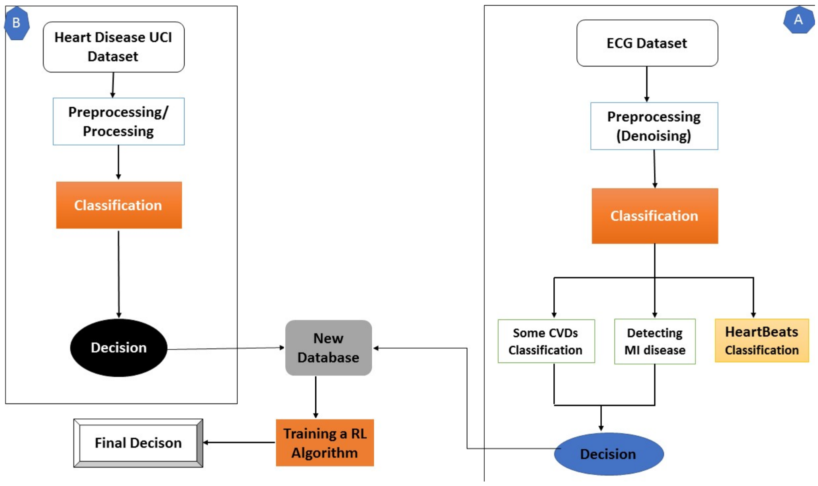


Figure 4.2.2: Global Conception

Our decision making process is designed to be with great similarity to the KDD and DL processes. The first step is the selection of ECG data (A) and Risk factor data (B) that similar to what is in Heart Disease UCI as treated data. After we target data field, ECG signal and Risk factor data files are generated, collected and assembled to form the database source files. However, the ECG signal present in our database, was exposed to several types of noise during the generation and record step, the preprocessing step is dedicated to the noise elimination (or minimization) . After we have our clean data, we proceed to the classification step, where we classify heartbeats and some Arrhythmias and save the decision-making of heart disease class on the database with the date of the decision. In the other side (B), where the risk factors data file is present in our database, after that, when the decision is making save it in the database with the new risk factors data. Then, try to make a sequence dataset which contain the two model decisions of the both sides which are ecg signal (A) and the risk factors (B). and when the patient die, we save it in the database with the date of death as a class of our new database. The resulted data will form the new database, which be divided into two parts, the first one used as training data, while the second one will be used to validate our classifier. If the obtained results are acceptable this means that our model is approved and can be used for the help in decision-making, otherwise we should return back to perform the necessary modifications and adjustments in the steps which precede the training, this adjustment operation is known as the tuning process.

In the next part, we will try to explain the previous steps with more details, how we realized it and about data that we are used.

4.3 Detailed conception and realization :

4.3.1 Data Selection and Acquisition :

4.3.1.1 Data Selection :

As we have seen before that our vision is divided in two parts (figure 4.2.2 which are the most representative of the heart state. The ECG is the widely used in the monitoring of the heart disease, and the risk factors that can help doctors to know the heart state of patient. For this, we selected as the studied data sources.

4.3.1.2 Data Acquisition :

We have chosen to get our data from three sources, the downloaded databases from PhysioNet website [2] and UCI ML website [43]. Also, the generated data from the our conceived e-health platform.

4.3.1.2.1 PhysioNet website :

In this study, we used the MIT-BIH dataset to classify different heartbeats and some CVDs , PTB datasets to detect Myocardial Infraction disease, Computing in the Cardiology Challenge 2017 (CINC17) dataset to detect Atrial Fibrillation disease, Here, in the next there are some details about this datasets are we used and the download scripts.

A- MIT-BIH dataset : The MIT-BIH dataset includes the ECG signals for 48 half-hour recording of different subjects, collected from 47 patients between 1975 and 1979. The ECG recording at the sampling rate of 360Hz. Each record contains two ECG leads; ECG lead II and lead V1. Usually, the lead II is used to detect heartbeats and different CVDs. This database is recommended by the American association of medical instrumentation (AAMI), since it includes the five essential arrhythmia groups as described in Table 4.3.1.

Category	Annotations
N	- Normal - Left/Right bundle branch block (L/R BBB) - Atrial escape - Nodal escape
S	- Atrial permature -Aberrant atrial permature - Nodal permature - Supra-ventricular permature
V	- Permature ventricular contraction - Ventricular escape
F	- Fusion of ventricular and normal
Q	- Paced - Fusion of paced and normal - Unclissifiable

Table 4.3.1: Summary of mappings between beat annotations and AAMI EC57 categories. Adopted from [28]

B- PTB dataset : The PTB Diagnostics dataset consists of ECG records from 290 subjects: 148 diagnosed as MI , 52 healthy control, and the rest are diagnosed with 7 different disease as described in Table 4.3.2.

diagnosis	Patients
Myocardial Infraction	148
Cardiomyopathy/Heart failure	18
Bundle banch block	15
Dysrythmia	14
Myocardial hypertrophy	7
Valvular heart disease	6
Myocarditis	4
Miscellaneous	4
Healthy Control	52

Table 4.3.2: Diagnosis classes in the PTB ECG Database. Adopted from [41]

C- Computing in the Cardiology Challenge 2017 Dataset (CINC17): The 2017 PhysioNet/CinC Challenge aims to encourage the development of algorithms to classify, from a single short ECG lead recording (between 30 s and 60 s in length), whether the recording shows normal sinus rhythm, atrial fibrillation (AF), an alternative rhythm, or is too noisy to be classified.

The three precedent datasets have a four types of files :

- ***.dat or *.mat file:** containing rows of digitized real values,

Elapsed time (seconds)	ECG1 (mV)	ECG2 (mV)
0.000	0.160	-0.290
0.008	0.180	-0.300
0.016	0.170	-0.320
0.023	0.180	-0.290
0.031	0.180	-0.310
0.039	0.170	-0.330
0.047	0.170	-0.330
0.055	0.130	-0.360
0.062	0.130	-0.390
0.070	0.100	-0.390
0.078	0.060	-0.350
0.086	0.040	-0.310
0.094	0.010	-0.270
0.102	-0.010	-0.200
0.109	-0.020	-0.150
0.117	-0.020	-0.150
0.125	-0.020	-0.160
0.133	-0.030	-0.120

Figure 4.3.1: the sample of *.dat File

- ***.atr file:** containing the annotation of all beats in the *.dat file,

0:00.594	76	N
0:01.414	181	N
0:02.219	284	N
0:03.000	384	N
0:03.773	483	N
0:04.578	586	N
0:05.391	690	N
0:06.195	793	N
0:06.992	895	N
0:07.781	996	N
0:08.586	1099	N
0:09.391	1202	N
0:10.180	1303	N
0:11.000	1408	N
0:11.813	1512	N
0:12.625	1616	N
0:13.422	1718	N
0:14.227	1821	N
0:15.031	1924	N
0:15.820	2025	N

Figure 4.3.2: the sample of *.atr File

- ***.hea file:** containing complement information about the patient and the ECG signal.

```
14046 2 128 10828800
14046.dat 212 0 12 0 32 -13462 0 ECG1
14046.dat 212 0 12 0 -58 -28804 0 ECG2
#Age: 46  Sex: M
```

Figure 4.3.3: the sample of *.atr File

4.3.1.2.2 UCI Machine learning website :

also we used the heart disease dataset that available on UCI Machine learning Repository. This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Here the 14 used :

1. 3 (age)
2. 4 (sex)]
3. 9 (cp)
4. 10 (trestbps)
5. 12 (chol)
6. 16 (fbs)
7. 19 (restecg)
8. 32 (thalach
9. 38 (exang)
10. 40 (oldpeak)
11. 41 (slope)
12. 44 (ca)
13. 51 (thal)
14. 58 (num) (the predicted attribute)

4.3.1.2.3 E-Health platform :

The e-Health Sensor Shield (figure 4.3.5) allows Arduino and Raspberry Pi users to perform biometric and medical applications where body monitoring is needed by using several sensors. This information can be used to monitor in real time the state of a patient or to get sensitive data in order to be subsequently analyzed for medical diagnosis. Biometric information gathered can be wirelessly sent using any of the 6 available connectivity options such as: Wi-Fi, 3G, GPRS, Bluetooth, 802.15.4 and ZigBee depending on the application. In this project, we use Wi-fi and 3G as a connectivity options and Electrocardiogram sensor (ECG).

The output of e-Health platform (e-HP) is a file containing rows of real values, representing electrical activity of the heart.

In the next sections, we'll describe how to realize the data acquisition for our prototype's schematic that we see it in the figure 4.3.4.

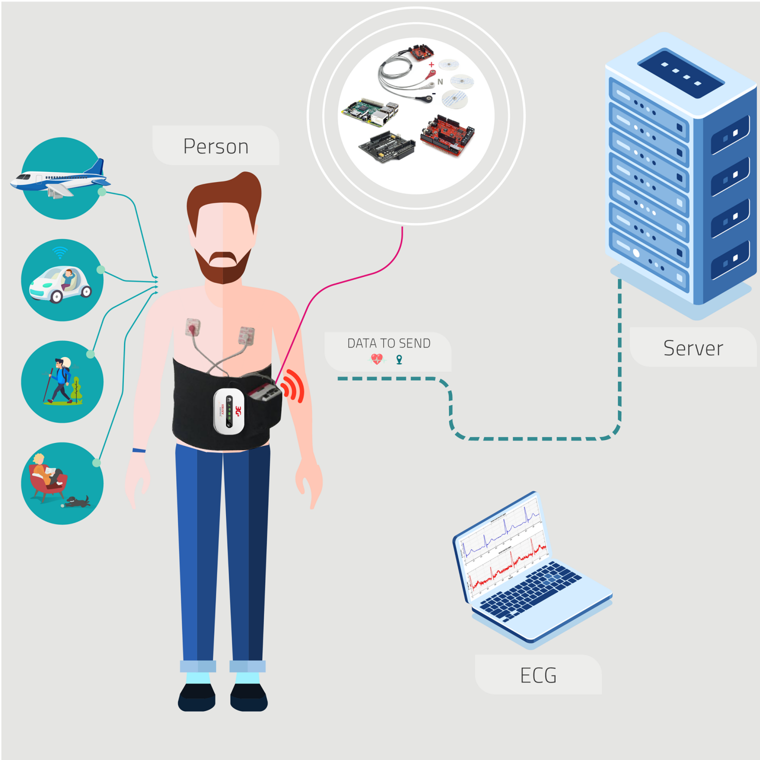


Figure 4.3.4: Prototype's Schematic

The e-HP sensor platform can be used with an Arduino (by default) or Raspberry kit. In our work, we are using the V2.00 of the platform with Raspberry Pi3 B kit. To plug the e-HP into the Raspberry kit we need to use the Raspberry Pi to Arduino shields connection bridge, which allow to use the e-HP with the Raspberry Pi kit.

Figure 4.3.5: From left to right: -The E-HP, -The Connection bridge, -Raspberry Pi3 B

To make the entire system operational we followed the next steps:

-a. Preparing the Raspberry Pi3 (RPi3): The RPi3 is a of small single-board computer developed in the United Kingdom by the Raspberry Pi Foundation.
To operate the RPi3, we have used Raspbian O.S, which is a version of Debian Linux O.S dedicated to the Raspberry Pi kit, and is recommended by Raspberry Pi Foundation.

b. Connecting the Raspberry Pi to Arduino shields: The next step is to configure the RPi3 to be ready to communicate with the Raspberry Pi to Arduino shields, for that simply execute the next command in the command prompt of the Raspbian O.S:

```
wget http://www.cooking-hacks.com/media/cooking/images/documentation/ras
pberry_arduino_shield/raspberrypi2.zip && unzip raspberrypi2.zip && cd c
ooking/arduPi && chmod +x install_arduPi && ./install_arduPi && rm insta
ll_arduPi && cd ../..
```

c. Installing the e-HP library: The e-Health library includes high level functions for an easy manage of the board. The library can be downloaded from [44]. The e-Health library for Raspberry Pi requires the ArduPi library and both libraries should be in the same path.

d. Pluging the ECG sensors and assemeble the system : Now our platform is ready to acquire the e-HP. The next step is plugging the ECG sensors into e-HP and then we assemble our system.

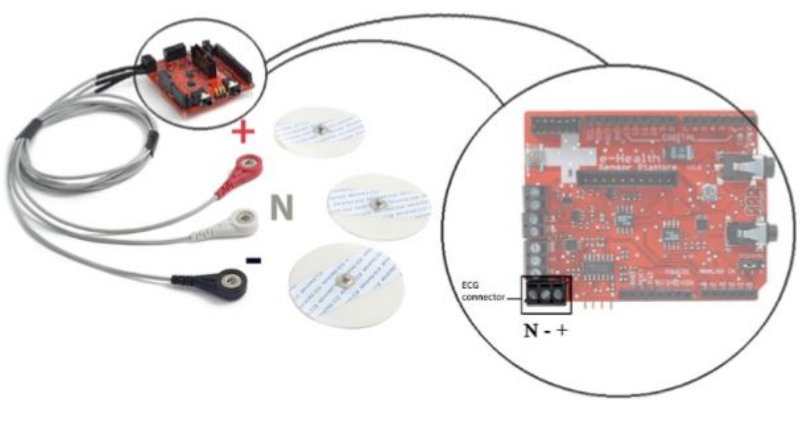
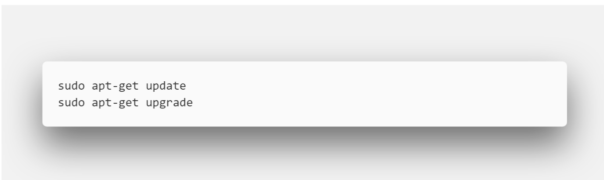


Figure 4.3.6: ECG sensors plugged into the e-HP (with appropriate placements)

Figure 4.3.7: The assembled platform

e. Enabling the interface with the e-HP : In order to interfacing the RPi3 with e-HP. At the end of the configuration steps, update the O.S with the latest patches using those commands:



```
sudo apt-get update
sudo apt-get upgrade
```

Reboot the RPi3, and then compile the arduPi library.

Now our system is ready for the acquisition step.

f. Data acquisition with the platform : Once our system is ready, we proceed to the acquisition step. For that, we have to create the necessary program; the original proposed version is on C++ available on the cooking-[17].

Where ECG.cpp is the name of file to be compiled, and ECG the resulting file which can be executed by: Sudo ./ECGToFile

h. Platform mobility realization : To realize the mobility aspect of the platform we proceeded to two operations:

- The first one is realizing the auto-lunching of acquisition program in Raspbian O.S by modifiying the /etc/rc.local file, by adding the path and the command to lunch the acquisition program before the exit 0 line, like following:
  Cd /home/pi/myProjectsFolder

Sudo ./ECGToFile
Exit 0

- Using a power-bank to power the platform and make it independent from the outlet. However, the autonomy the platform depend on the power-bank capacity, according several forums discussion a 50.000 mAh power-bank can power a Raspberry Pi3 for 24 hours. Unfortunately we was unable to acquire a such like power-bank, we contend with a SONY cycle Energy 10.000 mAh power-bank.

Figure 4.3.8: The Power Bank

h. Remotely and real time recording : to assure the remote monitoring of ECG we use the Wi-Fi key of Djezzy which able to share the Internet to our platform by capture it using the Wi-Fi on Raspberry Pi. This step offer to sending ECG signal record in real time and everywhere.

After the precedent steps, we'll obtain our prototype as see it in the figure 4.3.9

Figure 4.3.9: The platform with the belt

In the next parts, we'll describe the different Deep learning models and the detail's implementation of the next steps that we presented in the detailed conception. But, before this, we should to present our machine and the framework that we used on the implementation step.

4.3.2 Environment of work for the rest of the project :

4.3.2.1 The machine used :

For the remaining tasks of our project, we used the "Kali linux 2018.4" as O.S, which is mounted on HP compatible machine dotted with the following characteristics:

- **CPU :** intel pentium , 2.57GHz,

- **RAM :** 4Gb.

4.3.2.2 The development environment :

Google Colab or "the Colaboratory" is a free cloud service hosted by Google to encourage Machine Learning and Artificial Intelligence research, where often the barrier to learning and success is the requirement of tremendous computational power.

- **Benefits of Colab:** Besides being easy to use (which I'll describe later), the Colab is fairly flexible in its configuration and does much of the heavy lifting for students or experts working on DM and Ai.

- Python 2.7 and Python 3.6 support,
- Free GPU acceleration,
- Pre-installed libraries: All major Python libraries like Keras, TensorFlow, PyTorch, and OpenCV, Scikit-learn, Matplotlib among many others are pre-installed and ready to be imported,
- Built on top of Jupyter Notebook Collaboration feature (works with a team just like Google Docs): Google Colab allows developers to use and share Jupyter notebook among each other without having to download, install, or run anything other than a browser,
- Supports bash commands,
- Google Colab notebooks are stored on the drive,
- etc.

Here we give some basics on how use google colab :

- **Creating Folder on Google Drive :** Since Colab is working on your own Google Drive, we first need to specify the folder we'll work. I created a folder named "app" on my Google Drive. Of course, you can use a different name or choose the default Colab Notebooks folder instead of app folder.

- **Creating New Colab Notebook :** Create a new notebook via Right click,go to More, then click to Colaboratory .

  Rename notebook by means of clicking the file name.

4.3.3 Pre-processing (ECG Denoising) :

Random noise exists in wearable sensor based measurements whose frequency spreads from -p to p . Frequency filtering or smoothing is very likely damage critical information. Remove noise from heartbeat using an trained autoencoder could be a promising direction to go as heartbeat is repeatable pattern. The autoencoder is trained using noisy heartbeat as input and corresponding ground truth heartbeat as output. Noise is modeled as uniform distribution.

In this work, we considered this step as we use the trained model, the entry is ecg file with noise , and the result is cleaned ecg signal.

4.3.4 Heartbeats Classification based on 1-D CNN :

4.3.4.1 Dataset used and Split :

The original datasets used are the MIT-BIH Arrhythmia Dataset that were preprocessed by [28] based on the methodology described in III.A of the paper (see figure ..). The assembled dataset will be split into two parts, the first one called Training data (about

80% of the dataset) will be used to learn the model chosen, and the second part (about 20% of the data set), will be used as test set.

4.3.4.2 Model Training and validation

the model that we are chosen is similar to [28]that we use a neural network based on 1D convolutions but without the residual blocks. Here the figure 4.3.10 shows the difference between the both models, and the figure 4.3.11 shows the python implementation of the model.

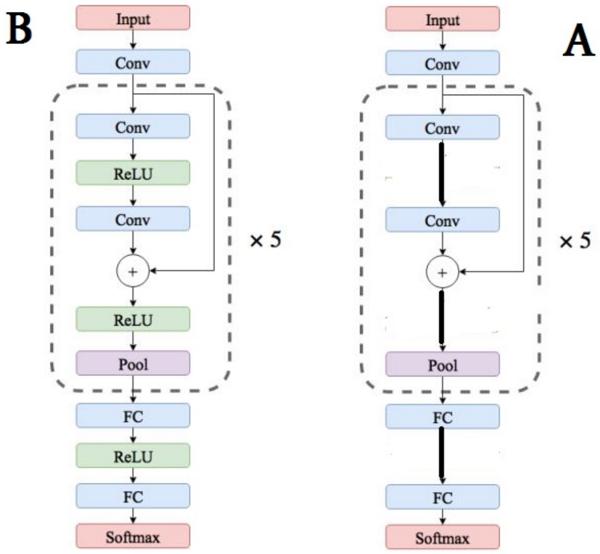


Figure 4.3.10: Difference between the B [28] model and our model A.

```python
def HB_model():
    nclass = 5
    inp = Input(shape=(187, 1))
    img_1 = Convolution1D(16, kernel_size=5, activation=activations.relu, padding="valid")(inp)
    img_1 = Convolution1D(16, kernel_size=5, activation=activations.relu, padding="valid")(img_1)
    img_1 = MaxPool1D(pool_size=2)(img_1)
    img_1 = Dropout(rate=0.1)(img_1)
    img_1 = Convolution1D(32, kernel_size=3, activation=activations.relu, padding="valid")(img_1)
    img_1 = Convolution1D(32, kernel_size=3, activation=activations.relu, padding="valid")(img_1)
    img_1 = MaxPool1D(pool_size=2)(img_1)
    img_1 = Dropout(rate=0.1)(img_1)
    img_1 = Convolution1D(32, kernel_size=3, activation=activations.relu, padding="valid")(img_1)
    img_1 = Convolution1D(32, kernel_size=3, activation=activations.relu, padding="valid")(img_1)
    img_1 = MaxPool1D(pool_size=2)(img_1)
    img_1 = Dropout(rate=0.1)(img_1)
    img_1 = Convolution1D(256, kernel_size=3, activation=activations.relu, padding="valid")(img_1)
    img_1 = Convolution1D(256, kernel_size=3, activation=activations.relu, padding="valid")(img_1)
    img_1 = GlobalMaxPool1D()(img_1)
    img_1 = Dropout(rate=0.2)(img_1)

    dense_1 = Dense(64, activation=activations.relu, name="dense_1")(img_1)
    dense_1 = Dense(64, activation=activations.relu, name="dense_2")(dense_1)
    dense_1 = Dense(nclass, activation=activations.softmax, name="dense_3_mitbih")(dense_1)

    model = models.Model(inputs=inp, outputs=dense_1)
    opt = optimizers.Adam(0.001)

    model.compile(optimizer=opt, loss=losses.sparse_categorical_crossentropy, metrics=['acc'])
    model.summary()
    return model
```

Figure 4.3.11: Python Implementation of Heartbeats classification Model.

The figure below shows the detailed conception of the model.

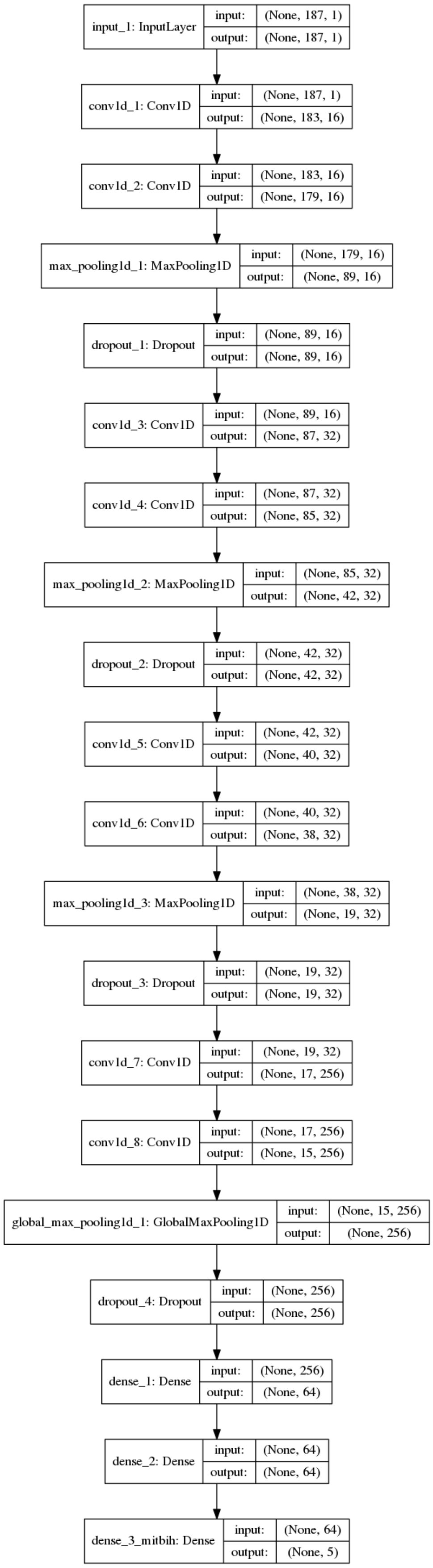


Figure 4.3.12: Detailed Conception of the model - by using the python code line plot_model() which is available on keras library-.

4.3.5 ECG Arrhythmia classification :

4.3.5.1 Detection Myocardial Infarction (MI) Based on LSTM:

As we have seen in the chapter 3, the importance of detecting MI which is more commonly known as a heart attack. and as we know the power of the use RNN especially LSTM on the classification tasks concerning time series data such as ECG signals. For this, we choose LSTM to detect the MI disease. In the next, we'll see the dataset used and how we prepared it to our classify model based on LSTM.

4.3.5.1.1 Dataset used : Exploploration and Split :

In this part, we use the PTB Diagnostic ECG database, that contains nine different CVDS. But for this study we only aim to discriminate between healthy control and MI.

- **Data exploration: Comparing the data** As can be seen from figure 4.3.15 , the ECG signal of the healthy control is a lot more constant and less noisy compared to that of the myocardial patient. This difference is visible over multiple channels and over multiple patients. The codeblock below shows the code which allows you to compare more patients and channels with each other. If you do, you'll notice that the ECG's of the myocardial patients generally look a lot more noisy and has less profound spikes.

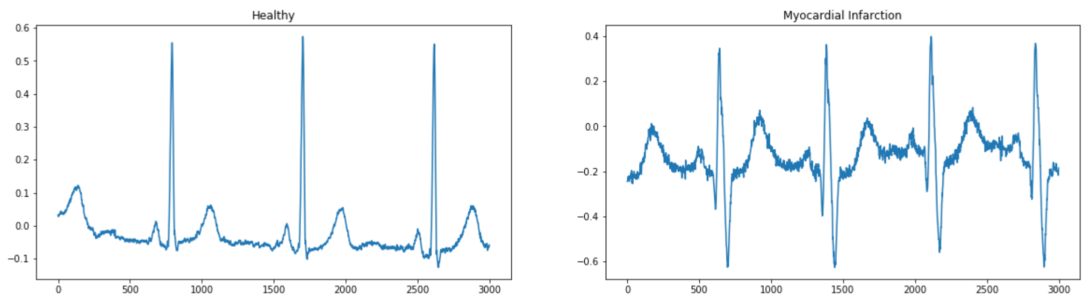


Figure 4.3.13: Comparison of the first channel of the ECG of a Healthy control and a subject who had a Myocardial infarction.

- **Data split :** The assembled dataset will be split into two parts, the first one called Training data (about 80% of the dataset) will be used to learn the model chosen, and the second part (about 20% of the data set), will be used as test set.

4.3.5.1.2 Building the model :

We will be using a simple 3 layered LSTM network with dropout after each layer. The number of LSTM nodes per layer start with 256 and are halved every next layer. The first LSTM two layers return the whole output sequence, while the last LSTM layer only returns the last step of its output sequence, thus dropping the temporal dimension. The code block below 4.3.14 shows the python implementation of the LSTM model .

```
def make_model(input_shape, output_dim, dropout=0.2):

    print("model dim: ", input_shape, output_dim)
    model = Sequential()
    model.add(CuDNNLSTM(256, input_shape=input_shape, batch_size=None, return_sequences=True))
    model.add(Dropout(dropout))
    model.add(CuDNNLSTM(128, return_sequences=True))
    model.add(Dropout(dropout))
    model.add(CuDNNLSTM(64))
    model.add(Dropout(dropout))
    model.add(Dense(output_dim, activation='softmax'))
    model.compile(loss='categorical_crossentropy', optimizer='adam')
return model
```

Figure 4.3.14: Python code of LSTM model used on Detecting MI disease.

4.3.5.2 CVDs Classification :

4.3.5.2.1 Dataset used :

As we have seen, the most works on Arrhythmias classification in the two precedent years are on CINC17 to join the PhysioNet website competition. For this, I choose to use the MIT-BIH Arrhythmia Dataset that were cleaned. The assembled dataset will be split into two parts, the first one called Training data (about 80% of the dataset) will be used to learn the model chosen, and the second part (about 20% of the data set), will be used as test set.

4.3.5.2.2 The model :

As we have present in th chapter 2 that ,the research paper in NATURE by Stanford ML group [23] is surprising. It says the F1 score (frequenc-weighted) average of the their trained model is superior to average cardiologist. This research paper really motivate me to test their model on MIT-BIH Arrhythmia Dataset.

Their model shows in figure 4.3.15 is mainly made of 3 blocks, the input plus a resnet block, and a resnet loop-block and a output block. One of the 1-D convolution layer in the loop-block reduces the half of the size of the layer per bi-loops. Thus, during 15 loops, it shrink down the input size of the signal to 1/256 times. And its filter length gets double up with zeropad every 4 loops. Then the filter grows from 1 to 32 at the input block and enlarges from 32 to 256 at the loop-block.

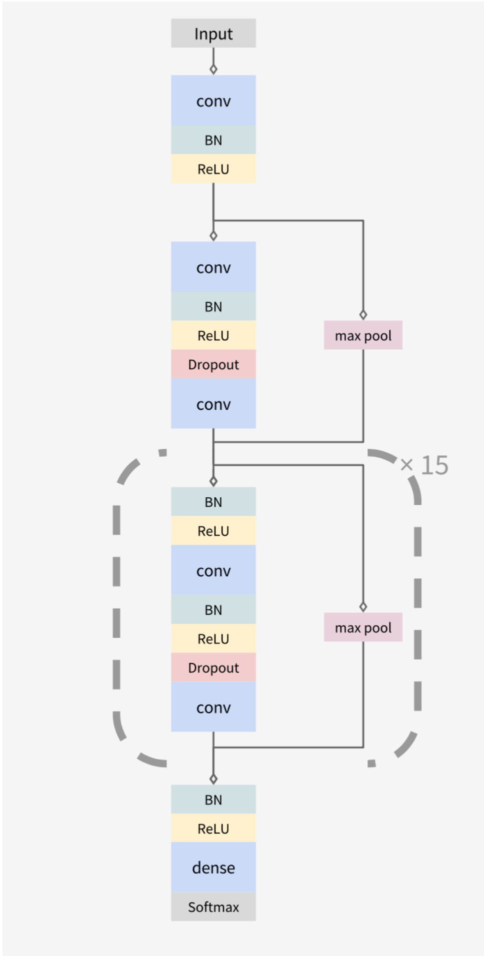


Figure 4.3.15: The Architecture of Stanford ML group. Adopted from[23]

4.3.5.2.3 Cleaning data :

When we used the same model of Stanford ML group in the first times didn't give a good accuracy, for this we thought that because the difference between the two datasets.

Let us look at the dataset of Stanford research group used in the above figure 4.3.15. It looks well-balanced. There are still majority of normal sinus signals but not so much as MIT-BIH data. Thus, we decided to remove randomly 85 percents of sinus signals in the MIT-BIH dataset.There are 5 types, MLII, V1, V2, V4, V5 in MIT-BIH dataset. we mainly trained and tested MLII and V1 features among them because other features do not have many data. for this we also reduced the number of categories from more than 15 to only 5 by getting rid of minor number of categories. These five kinds of heart beat signals are secured to have more data than others for a qualified classifier. I have also removed L, R labels. These are branch bundle block beats which were originally labeled as normal beats before.

```
Training set
          AF 9313
         AVB 3043
    BIGEMINY 3181
         EAR 4367
         IVR 2572
  JUNCTIONAL 2484
       NOISE 10751
       SINUS 33540
         SVT 8228
   TRIGEMINY 3119
          VT 5722
 WENCKEBACH 2395
Testing set
          AF 59
         AVB 48
    BIGEMINY 22
         EAR 22
         IVR 34
  JUNCTIONAL 36
       NOISE 40
       SINUS 213
         SVT 34
   TRIGEMINY 20
          VT 17
 WENCKEBACH 29
```

Figure 4.3.16: Different CVDs on Dataset of Stanford ML group . Adopted from[23]

4.3.5.3 Heart disease prediction based on DM :

4.3.5.3.1 Dataset used : Visualization and Understanding:

in this part, I have used the Heart Disease UCI dataset to help me in our future vision.

Before any analysis, I just wanted to take a look at the data. So, I used the info() method.

As you can see from the output above, there are a total of 13 features and 1 target variable. Also, there are no missing values so we don't need to take care of any null values. Next, I used describe() method.

	age	cp	trestbps	chol	fbs	restecg	thalach	oldpeak	slope	ca
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	30
mean	54.366337	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	1.039604	1.399340	0.7
std	9.082101	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	1.161075	0.616226	1.0
min	29.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.0
25%	47.500000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	1.000000	0.0
50%	55.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.800000	1.000000	0.0
75%	61.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.600000	2.000000	1.0
max	77.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	6.200000	2.000000	4.0

Figure 4.3.17: Dataset Describe

The method revealed that the range of each variable is different. The maximum value of age is 77, 25% of samples are around 61 years old or older.but for cholesterol it is 564. Thus, feature scaling must be performed on the dataset and 50% of samples have a cholesterol level up to 240.

To more understanding this data,we'll plot the correlation matrix. colorbar() shows the colorbar for the matrix.

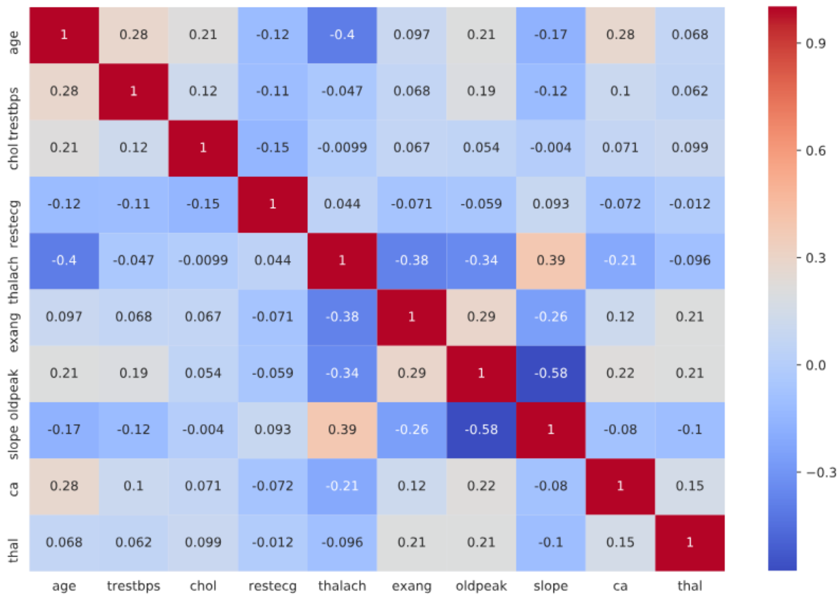


Figure 4.3.18: Correlation Matrix

It's easy to see in figure 4.3.18 that there is no single feature that has a very high correlation with our target value. Also, some of the features have a negative correlation with the target value and some have positive.

Which factors causing more heart diseases ? As we shows in the figure 4.3.19

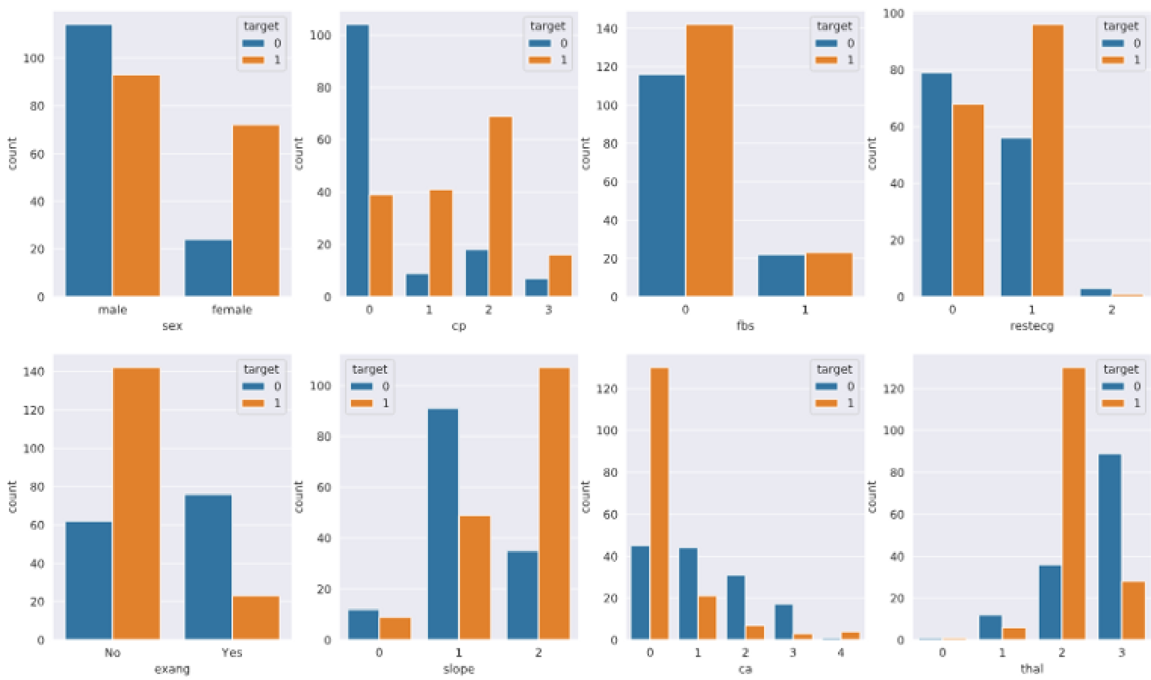


Figure 4.3.19: The main factors causing more heart diseases

It's easy to determine that :
Chest pain: the heart disease diagnosis is greater among the patients that feel any chest pain.
Restegc-Electrocardiograph results: the rate of heart desease diagnoses higher for patients with a ST-T wabe abnormality .
Slope: The ratio of patients diagnosed with heart desease is higher for slope = 2
Ca: The diagonosed ratio decreases fo ca between 1 and 3.
Thal: the diagnosed ratio is higher for thal = 2.

Let's perform a multivariate analysis, comparing the number of healthy and unhealthy people by sex. to do this we implemented the code below that shows the figure 4.3.20

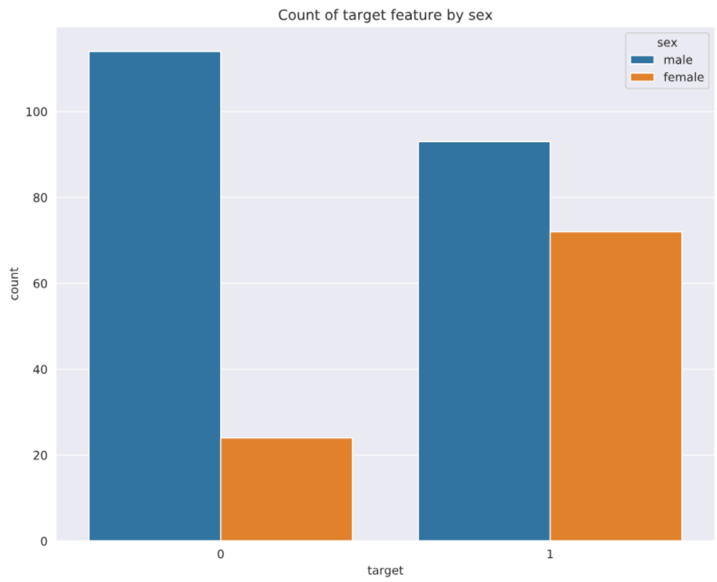


Figure 4.3.20: The sex which is most effecting the heart disease

It's easy to remark that, the amount of healthy male people is greater than the amount of unhealthy. For women, the number of unhealthy women is higher.

to see which sex is having higher cholesterol levels ? we implement the code below that the Quantitative/Categorical Plots is in figure 4.3.21 which present that female patients has higher cholesterol indices than male patients.

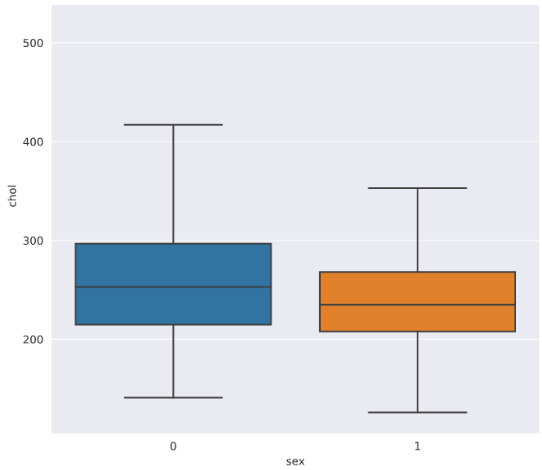


Figure 4.3.21: The sex vs Cholesterol levels.

4.3.5.3.2 The model :

Here in the next, we'll try to implement the K Neighbors Classifier based on KDD process. Before we go to the Classification task, we should to preprocess our datasets, in this step we convert some categorical variables into dummy variables and scale all the values before training the Machine Learning models. First, I'll use the get dummies method to create dummy columns for categorical variables. Then, we'll try to classify it by K neighbors and playing on parameters to got the best accuracy.

4.4 Our future vision :

Or in other words, How can we realize our new proposition which can predict the heart attack before hours, days or months. Unfortunately, we can't realize it because is a big project to building it in few months, but the idea is very clear for me. Here in the next, the steps of how can realize it in the future.

1- Create Database : to realize our vision, we need first to create a new sequence database that containing the Health Life Style of patients which are :

- The Heart disease class that we get it from the decision of our ECG Arrhythmias model and the Cardiologist experts, with the date of diagnostic and the patient location.
- The patient data similar to the Heart Disease UCI database - that we studied in the previous subsection- and the decision making by our classification model and doctors, with the date of diagnostic and the patient's location.
  Like the below :
  ((Diagnostic Date: Heart disease Class: patient's Location)* (Diagnostic Date: Some Health data : Heart disease stage : patient's Location)*)

2- Decision Class of the Database : we still set the data that we cite it and when the patient is die by the cause of a heart attack, we set in the database that he/she dies with the date of the death.

Here is a table that illustrate our database.

Data	**Class**
(..(date i :Age=25,....cholesterol :150..etc : Stage 2 : Alger)...	(date i : die)
(...(date i : MI disease :Batna).... (date N: Paced : Djanet)..())	(date I : None)
.....	....
((date1:Age=61,..etc : Stage 4 : UK).. (date N: Nornal : Setif))	(date i : die)

Table 4.4.1: Database illustrate our future vision

3- Development of the automatic system : The data base that we are created in is the reinforcement learning because some patients that we still alive. For this we need to train and test our system using the new algorithm which is Generative Adversarial Networks (GANs), RNN or try to use the measurement similarity which based on math.

4- Input and Output: the input to our system will be like what we have seen on the table 4.4.1. and the output will be like : "77.7% of humans like your health state was died after two weeks ".

In this chapter, we have presented an approach to use in the conception of our classification model. We have also discussed the several steps of this conception and the most used technics to evaluate, approve and realize the efficiency of our global vision. In the last of proposed steps, we have given our point of view for a prediction of heart attack model based on heart disease class and the main risk factors of heart state.

The next chapter is assigned to the discussion of the result obtained are explained and try to compare our results with the recent works.

Chapter 5

Results and discussion :

5.1 Introduction :

In the previous chapter, we have explained our conceptual approach to build an ECG Arrhythmias classifier based on DL models and heart disease prediction based of K Neighbors Classifier, and we have discussed the several steps, from data selection to implementation of the model validation, to achieve the decision making phase.

In this chapter, results obtained on different parts that we explained on previous chapter will discussed with some comparisons.

5.2 ECG Denoising:

As we have seen in the previous chapter that we used a DAE pre-trained model to clean the denoising on ECG signals.

- **Input :** The input of this phase is *.dat file with some noising.

- **Output :** The output will be a denoised ECG signal as we shows on figure 5.2.1

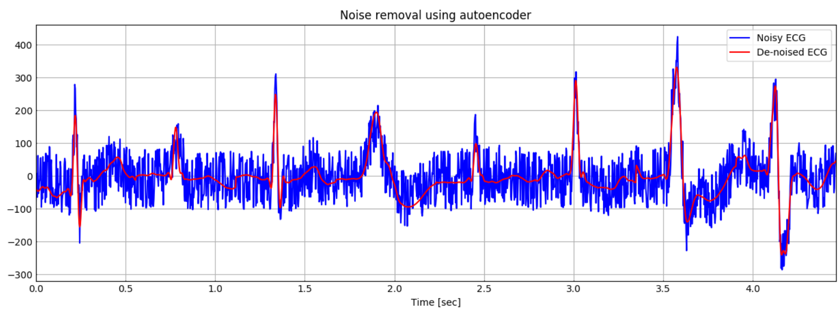


Figure 5.2.1: Result of ECG Denoising based on Autoencoder

5.3 ECG Heartbeats classification :

After Launching the training step on preprocessed MIT-BIH dataset with 1-D CNN on our machine, the setup stopped improve the Accuracy on the epoch 34, the result shows on figure 5.3.1 which are :

- **Accuracy :** 98.6

- **F1 score :** 92.1

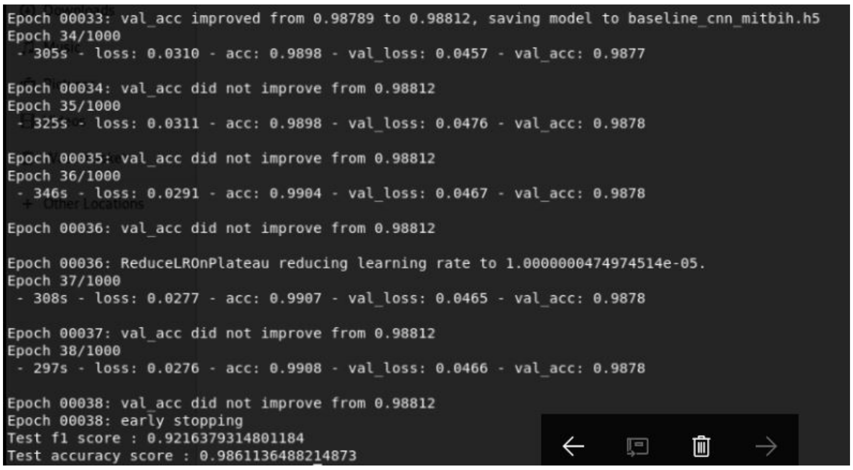


Figure 5.3.1: Accuracy and F1 score of the Heartbeats classification.

5.3.1 Comparison :

Here is a comparative table 5.3.1 shows our result with the results of other works .

	N	S	V	F	Q	Accuracy
this study	0.98	0.89	0.96	0.87	0.986	0.98
Drid, Bitam [16]	0.99	0.84	0.84	0.40	0.0	0.990

Table 5.3.1: Comparison of heartbeats classification results.

The performance of our model is comparable with the [16] that they use Random Forest and classify 15 different heartbeats, however, it noteworthy to mention that they used on the prepossessing and processing phases a black tool box to denoise ECG signal and extract the main features. The table 5.3.1 shows that our model is better then the model they [16] used on S,V,F and Q heartbeats.

- **Remark :** When we train the model on our machine used, the execution time was 4 hours, but when we trained it on google colab was a surprising time execution for me which was just 3 minutes .

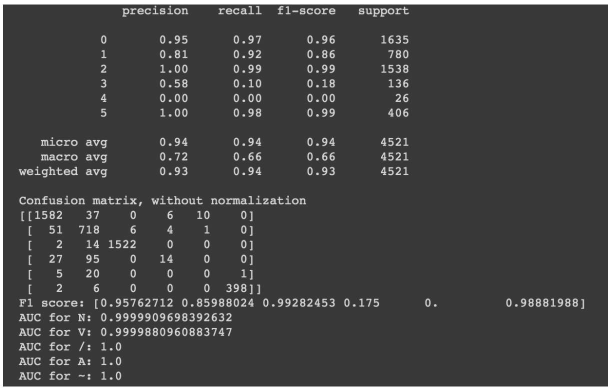
5.4 ECG Arrhythmias Classification :

5.4.1 Detection Myocardial Infarction (MI) Based on LSTM :

When we setting the random seed to 1337 in the implementation step, the results can differ a bit depending on the division of the test and training set, but you can expect an precision and recall of at least 0.90 on Myocardial infarction. For the healthy control these numbers are a bit lower, around 0.7 for both precision and recall that's because the healthy control is not strange considering there was a lot less data available of healthy subjects, also we are not tune the hyperparameters carefully such as the random seed.

5.4.2 CVDs Classification :

After Launching the training step on MIT-BIH dataset based on Stanford ML group [23, 39] model on google colab, we got very good f1-scores for normal, ventricular, paced beats and noise of validation. Figure 5.4.1 present the accuracy that we get it is 94% which is really motivate to continue developing our automatic classification system.



```
              precision    recall  f1-score   support

           0       0.95      0.97      0.96      1635
           1       0.81      0.92      0.86       780
           2       1.00      0.99      0.99      1538
           3       0.58      0.10      0.18       136
           4       0.00      0.00      0.00        26
           5       1.00      0.98      0.99       406

   micro avg       0.94      0.94      0.94      4521
   macro avg       0.72      0.66      0.66      4521
weighted avg       0.93      0.94      0.93      4521

Confusion matrix, without normalization
[[1582   37    0    6   10    0]
 [  51  718    6    4    1    0]
 [   2   14 1522    0    0    0]
 [  27   95    0   14    0    0]
 [   5   20    0    0    0    1]
 [   2    6    0    0    0  398]]
F1 score: [0.95762712 0.85988024 0.99282453 0.175      0.         0.98881988]
AUC for N: 0.9999909698392632
AUC for V: 0.9999880960883747
AUC for /: 1.0
AUC for A: 1.0
AUC for ~: 1.0
```

Figure 5.4.1: Accuracy and F1 score of the some CVDs classification.

5.5 Heart disease prediction based on DM Techniques:

in the previous chapter, we seen that we choose to use K Neighbors Classifier, when we launch the training model as in figure 5.5.1 . However, the number of neighbors varied them from 1 to 20 neighbors. the figure 5.5.2 explore the test score in each case.

```
knn_scores = []
for k in range(1,21):
    knn_classifier = KNeighborsClassifier(n_neighbors = k)
    knn_classifier.fit(X_train, y_train)
    knn_scores.append(knn_classifier.score(X_test, y_test))
```

Figure 5.5.1: Python implementation of K Neighbors classifier

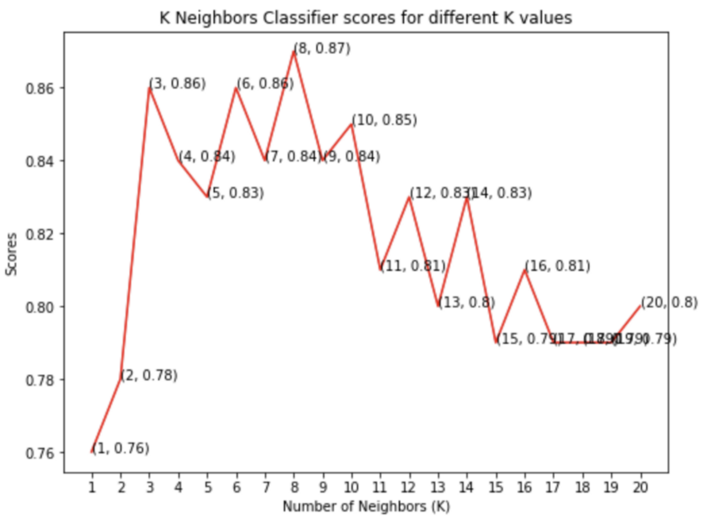


Figure 5.5.2: K Neighbors Classifier scores for different K values.

As we can see, we achieved the maximum score of 87% when the number of neighbors was chosen to be 8.

In this chapter, we have seen the gold and surprising results of our models that are better then a lot of works in the last recent years.

In the next chapter, we'll give a general conclusion with some future works that we want do it in the future .

Chapter 6

Conclusion and Future works :

This chapter gives a conclusion on our study and has information about our future work plan .

6.1 Conclusion :

DM and DL on Wearable Technologies was emerging and becoming the most talked about technology trends nowadays, influenced by the data explosion phenomenon and the great need to new techniques and tools to deal with the new generation data.

The opportunities of using DM and DL lies on how it enable to extract and deliver accurate results faster even in real time. Many fields have benefited from the opportunities offered by the Analytics on Wearable technologies.

In this work, we have seen the immense values of these opportunities when applying the DM and DL techniques in the healthcare sector, especially when focusing the deadliest diseases in our age, which are the heart disease.

In our case of study, we used the CNN model to classify the heartbeats present on recorded ECG signal, also we use LSTM model to detecting MI disease which can be define as a heart attack and use Stanford ML researcher group model to classify remotely six different CVDs on two ECG leads, also using the K neighbors classifier to predict a heart disease that based of the risks factors of heart state. Results scored showed a very accurate rate of the model trained with the data set used.

High accuracy diagnosis from ECG signal and the main factors of heart state with an automatic approach can save expert clinicians and cardiologists considerable time and decrease the number of misdiagnoses.

This improvements can lead to better on our new proposition to predict the heart attack before happening by developing a smart automatic system.

6.2 Future works Plan :

Because that our vision is bigger then build it in few months, for this as further works:

- continue to develop the smart automatic system that can help to predict the heart attack,

- i suggest to training our system based on GANs Algorithm, RNN or using the similarity measurements techniques,

- As we have been dealing with few diseases that we classified, try to classify other CVDs.

- try to augment our data and re-trained again .

- try to detecting other heart diseases on single or two lead ECG record.

- As we using a pre-trained model on ECG denoising, I'll try to develop our model based on Denoising Autoencoder (DAE).

working on echocardiography and develop an automatic classification system using Deep learning models like CNN .

- compare our results with Algerian cardiologists.

To end our modest study, we can say that we can't deny the great importance of computer science technology in realizing fruitful achievements in different domains especially that of health care.

Appendix :

.1 Web Application :

Here some captured photos of our web server application which is for Doctor's to manage our patients and see the results of our Automatic system.

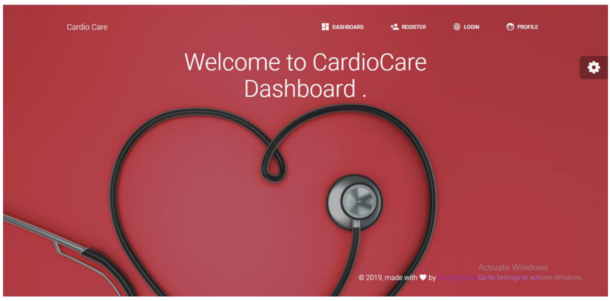


Figure .1.1: Home page

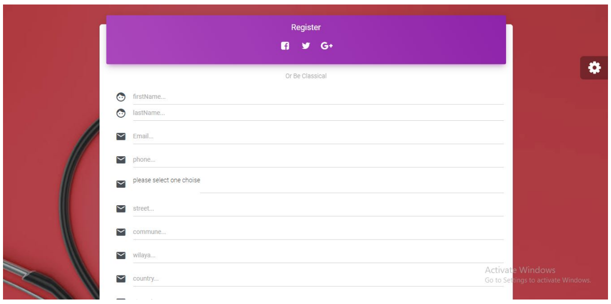


Figure .1.2: Register page

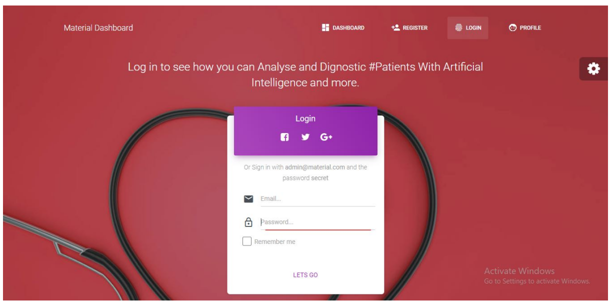


Figure .1.3: Login page

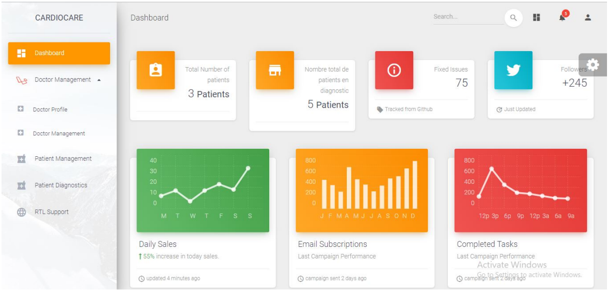


Figure .1.4: DashBoard : Statics page

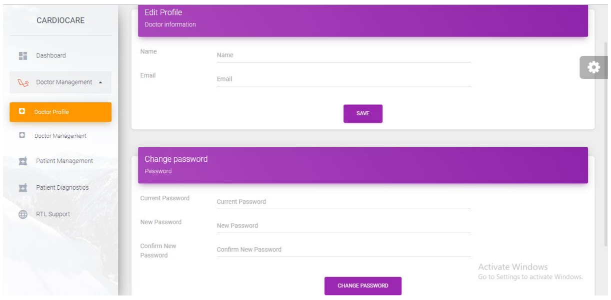


Figure .1.5: Doctor profile : to Edit personal infrmations

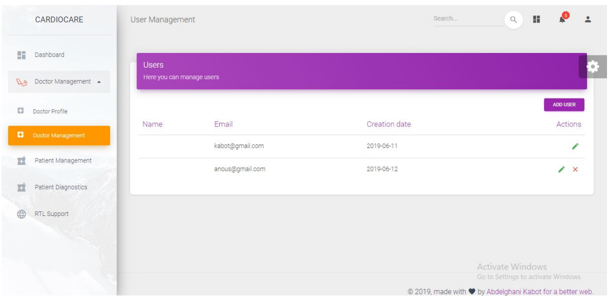


Figure .1.6: Doctor Management : to manage patients

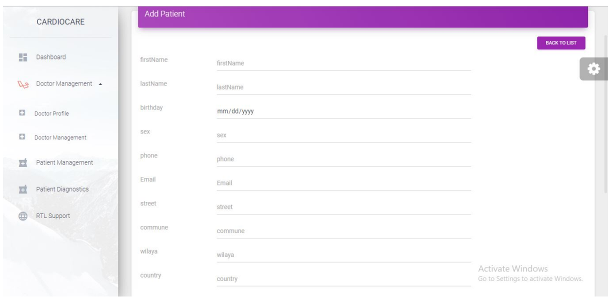


Figure .1.7: Patient Management : to add a new patient

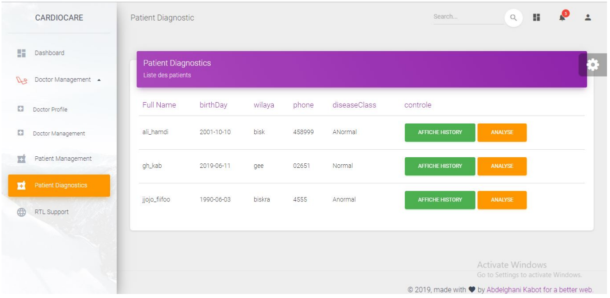


Figure .1.8: Patient Diagnostic

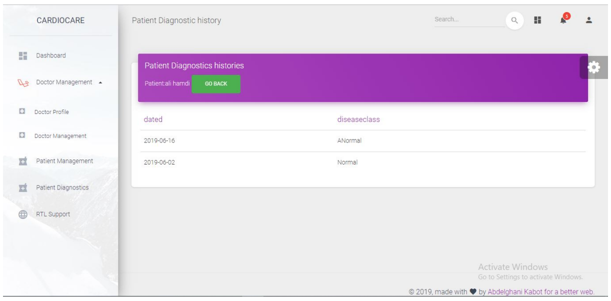


Figure .1.9: Patient Diagnostic Histories - sequenc eof heart disease classes -

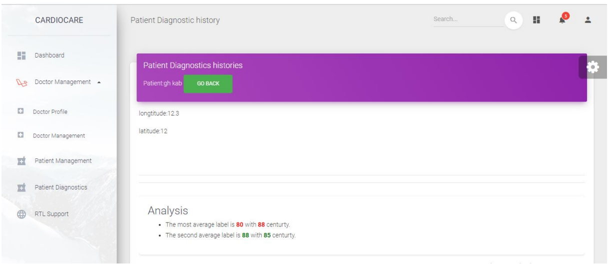


Figure .1.10: Heart disease class : using our deep learning model

Table	Action	Rows	Type	Collation	Size	Overhead
data	Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_unicode_ci	32 KiB	-
diagnostics	Browse Structure Search Insert Empty Drop	7	InnoDB	utf8mb4_unicode_ci	32 KiB	-
migrations	Browse Structure Search Insert Empty Drop	5	InnoDB	utf8mb4_unicode_ci	16 KiB	-
password_resets	Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_unicode_ci	16 KiB	-
patients	Browse Structure Search Insert Empty Drop	3	InnoDB	utf8mb4_unicode_ci	48 KiB	-
tests	Browse Structure Search Insert Empty Drop	4	InnoDB	utf8_general_ci	16 KiB	-
users	Browse Structure Search Insert Empty Drop	2	InnoDB	utf8mb4_unicode_ci	32 KiB	-
7 tables	Sum	21	InnoDB	utf8_general_ci	192 KiB	0 B

Figure .1.11: Tables of our database

We used Laravel 5.8 framework that based on php 7and mysql to implement the back-end side of our CardioCare website .

.2 Android Application :

Here some captured photos of our android application which is for patient to show the advices of him/her doctor, show the nearest hospitals on real time, the ECG diagnostic, Doctors advises and diagnostics.

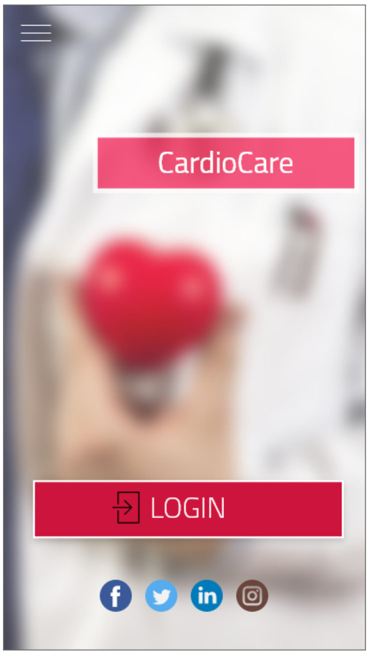


Figure .2.1: Home interface

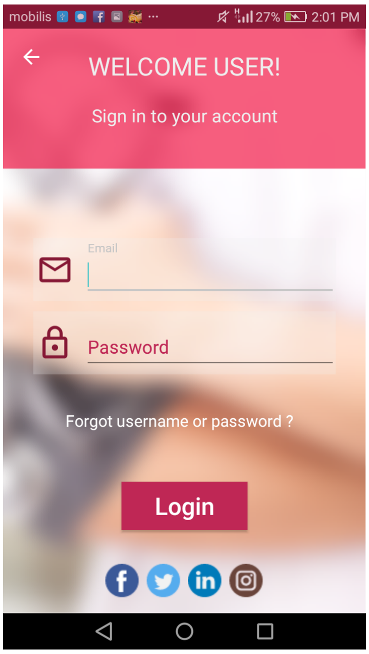


Figure .2.2: Login interface

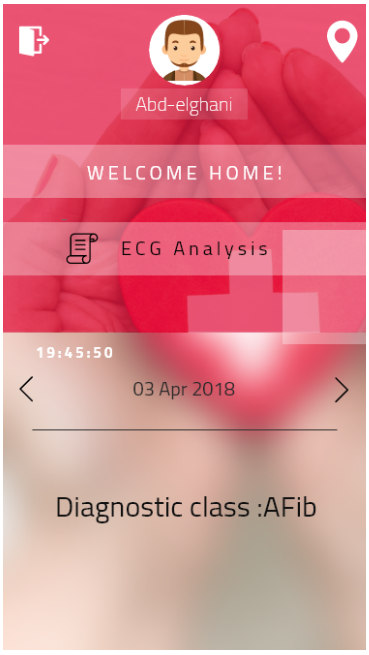


Figure .2.3: ECG Analysis interface - notification from doctors-

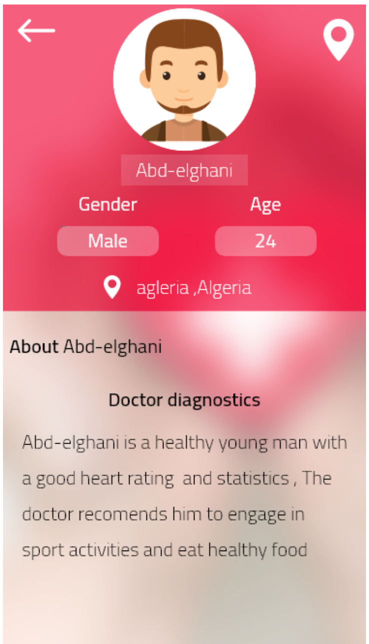


Figure .2.4: Doctor's Advises - notification from doctors-

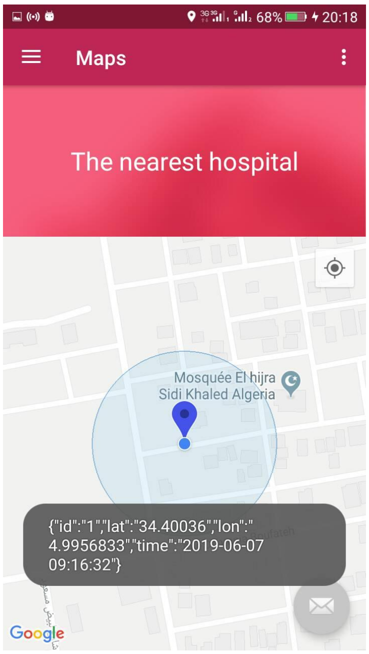


Figure .2.5: The Nearest Hospital

We used android studio and java to implement our CardioCare android application.

Bibliography

[1] URL: www.medium.com (visited on 02/12/2019).

[2] URL: https://physionet.org/.

[3] University of Toronto 2013. *What is Health Informatics.* URL: http://ihpme.utoronto.ca/academics/pp/mhi/faq/ (visited on 02/12/2019).

[4] Kayleigh Davis 2019. *Introduction to the Internet of Medical Things (IoMT).* URL: https://dzone.com/articles/internet-of-medical-things-iomt (visited on 02/15/2019).

[5] U Rajendra Acharya et al. "A deep convolutional neural network model to classify heartbeats". In: *Computers in biology and medicine* 89 (2017), pp. 389–396.

[6] Mohamad Mahmoud Al Rahhal et al. "Deep learning approach for active classification of electrocardiogram signals". In: *Information Sciences* 345 (2016), pp. 340–354.

[7] Karol Antczak. "Deep Recurrent Neural Networks for ECG Signal Denoising". In: *arXiv preprint arXiv:1807.11551* (2018).

[8] *Atrio ventricular node*. URL: https://en.wikipedia.org/wiki/Atrioventricular_node (visited on 04/15/2019).

[9] colah's blog. *Understanding LSTM Networks.* URL: https://colah.github.io/posts/2015-08-Understanding-LSTMs// (visited on 05/10/2019).

[10] Christopher Clifton. "Encyclopædia britannica: definition of data mining". In: *Retrieved on* 9.12 (2010), p. 2010.

[11] *Convolutional neural network*. URL: https://en.wikipedia.org/wiki/Convolutional_neural_network#Pooling (visited on 05/10/2019).

[12] Prudence W Dalrymple. "Data, information, knowledge: The emerging field of health informatics". In: *Bulletin of the American Society for Information Science and Technology* 37.5 (2011), pp. 41–44.

[13] Arden Dertat. *Applied Deep Learning - Part 3: Autoencoders.* URL: hhttps://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798 (visited on 05/10/2019).

[14] Marco Di Rienzo et al. "MagIC system: A new textile-based wearable device for biological signal monitoring. Applicability in daily life and clinical setting". In: *Conf Proc IEEE Eng Med Biol Soc.* Vol. 7. 2005, pp. 7167–7169.

[15] Duarte Dias and João Paulo Silva Cunha. "Wearable health devices—vital sign monitoring, systems and technologies". In: *Sensors* 18.8 (2018), p. 2414.

[16] Bitam S Drid B. "Pattern Recognition For Healthcare Analytics". In: *CS Department ,Biskra University* (2017).

[17] *e-Health Sensor Platform V2.0 for Arduino and Raspberry Pi [Biometric / Medical Applications].* URL: https://www.cooking-hacks.com/documentation/tutorials/ehealth-biometric-sensor-platform-arduino-raspberry-pi-medical (visited on 04/18/2019).

[18] *Electrical signals of the heart.* URL: https://www.ebme.co.uk/articles/clinical-engineering/electrical-signals-of-the-heart (visited on 04/15/2019).

[19] UM Fayyad. "Advances in knowledge discovery: making sense out of data". In: *IEEE Expert* 11.5 (1996), pp. 22–3.

[20] Adam Felman. *Everything you need to know about heart disease.* URL: https://www.medicalnewstoday.com/articles/237191.php (visited on 02/18/2019).

[21] Tracy D Gunter and Nicolas P Terry. "The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions". In: *Journal of medical Internet research* 7.1 (2005), e3.

[22] Li Guo, Gavin Sim, and Bogdan Matuszewski. "Inter-Patient ECG Classification with Convolutional and Recurrent Neural Networks". In: *arXiv preprint arXiv:1810.04121* (2018).

[23] Awni Y Hannun et al. "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network". In: *Nature medicine* 25.1 (2019), p. 65.

[24] J Hayward, G Chansin, and H Zervos. "Wearable Technology 2017–2027: Markets, Players, Forecasts". In: *IDTexEx Report* (2017).

[25] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016, pp. 770–778.

[26] *Heart.* URL: https://en.wikipedia.org/wiki/Heart (visited on 03/07/2019).

[27] Shweta H. Jambukia, Vipul K. Dabhi, and Harshadkumar B. Prajapati. "Classification of ECG signals using machine learning techniques: A survey". In: *2015 International Conference on Advances in Computer Engineering and Applications* (2015), pp. 714–721.

[28] Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. "Ecg heartbeat classification: A deep transferable representation". In: *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE. 2018, pp. 443–444.

[29] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. "Real-time patient-specific ECG classification by 1-D convolutional neural networks". In: *IEEE Transactions on Biomedical Engineering* 63.3 (2015), pp. 664–675.

[30] Choudur Lakshminarayan and Tony Basil. "Feature Extraction and Automated Classification of Heartbeats by Machine Learning". In: *arXiv preprint arXiv:1607.03822* (2016).

[31] Tanya Lewis. "Human heart: Anatomy, function & facts". In: *LiveScience. TechMedia Network* 7 (2015).

[32] Cuiwei Li, Chongxun Zheng, and Changfeng Tai. "Detection of ECG characteristic points using wavelet transforms". In: *IEEE Transactions on biomedical Engineering* 42.1 (1995), pp. 21–28.

[33] Juan Pablo Martínez et al. "A wavelet-based ECG delineator: evaluation on standard databases". In: *IEEE Transactions on biomedical engineering* 51.4 (2004), pp. 570–581.

[34] SL Melo, LP Caloba, and J Nadal. "Arrhythmia analysis using artificial neural network and decimated electrocardiographic data". In: *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*. IEEE. 2000, pp. 73–76.

[35] Sara Moein. "An MLP neural network for ECG noise removal based on Kalman filter". In: *Advances in Computational Biology*. Springer, 2010, pp. 109–116.

[36] *Myocardial Infarction*. URL: https://en.wikipedia.org/wiki/Myocardial_infarction (visited on 04/15/2019).

[37] Jiapu Pan and Willis J Tompkins. "A real-time QRS detection algorithm". In: *IEEE Trans. Biomed. Eng* 32.3 (1985), pp. 230–236.

[38] Suranai Poungponsri and Xiao-Hua Yu. "Electrocardiogram (ECG) signal modeling and noise reduction using wavelet neural networks". In: *2009 IEEE International Conference on Automation and Logistics*. IEEE. 2009, pp. 394–398.

[39] Pranav Rajpurkar et al. "Cardiologist-level arrhythmia detection with convolutional neural networks". In: *arXiv preprint arXiv:1707.01836* (2017).

[40] Munindar P. Singh. *Heart : anatomy*. URL: https://www.britannica.com/science/heart (visited on 03/25/2019).

[41] Nils Strodthoff and Claas Strodthoff. "Detecting and interpreting myocardial infarction using fully convolutional neural networks". In: *Physiological measurement* (2018).

[42] UFLDL Tutorial. *Multi-Layer Neural Network*. URL: http://ufldl.stanford.edu/tutorial/supervised/MultiLayerNeuralNetworks/ (visited on 05/10/2019).

[43] *UC Irvine Machine Learning Repository*. URL: https://archive.ics.uci.edu/ml/index.phpt.

[44] Cornell University. "Convolution". In: (2013). URL: https://www.cs.cornell.edu/courses/cs1114/2013sp/sections/S06_convolution.pdf.

[45] Romeo Vecht, Michael A Gatzoulis, and Nicholas Peters. *ECG diagnosis in clinical practice*. Springer Science & Business Media, 2009.

[46] Aswathy Velayudhan and Soniya Peter. "Noise Analysis and Different Denoising Techniques of ECG Signal-A Survey". In: *IOSR Journal of Electronics and Communication Engineering* 3.Iii (2015), pp. 641–644.

[47] Dan Wang and Yi Shang. "Modeling physiological data with deep belief networks". In: *International journal of information and education technology (IJIET)* 3.5 (2013), p. 505.

[48] John G Webster. "Reducing motion artifacts and interference in biopotential recording". In: *IEEE Transactions on Biomedical Engineering* 12 (1984), pp. 823–826.

[49] Qibin Zhao and Liqing Zhang. "ECG feature extraction and classification using wavelet transform and support vector machines". In: *2005 International Conference on Neural Networks and Brain*. Vol. 2. IEEE. 2005, pp. 1089–1092.