People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
University 8 May 1945 – Guelma
Faculty of Science and Technology
Department of Electronics and Telecommunications

Master Thesis

Domain: **Science and Technology**
Sector: **Electronics**
Specialty: **Instrumentation**

Automatic classification of ECG heartbeats using deep neural networks

---

Presented by:

Sadoun Mohammed Seghir

Supervised by :

Pr. Nemissi Mohamed

June 2022

ABSTRACT

Abstract

Automatic ECG classification systems are a valuable tool for assisting doctors and supervising patients. Deep neural networks have been widely used as an alternative to the existing classification systems based on distinct feature extraction and classification phases. In this study, we propose an automatic system that classifies ECG heartbeats based on 1D convolutional neural network. To evaluate the proposed model, we perform tests on the MIT-BIH arrhythmia database and we considers four classes.

Key words: Electrocardiography (ECG), Deep neural network, 1D Convolutional neural network (CNN).

ملخص

ملخص

تعد أنظمة التصنيف التلقائي لتخطيط القلب أداة قيمة لمساعدة الأطباء والإشراف على المرضى. تم استخدام الشبكات العصبية العميقة على نطاق واسع كبديل لأنظمة التصنيف الحالية القائمة على مراحل استخراج السمات المميزة ومراحل التصنيف. في هذه الدراسة ، نقترح نظامًا آليًا يصنف نبضات القلب على أساس الشبكة العصبية التلافيفية .لتقييم النموذج المقترح ، نجري اختبارات على قاعدة بيانات عدم انتظام ضربات القلب وننظر في أربع فئات.

الكلمات المفتاحية: تخطيط القلب الكهربائي، الشبكة العصبية العميقة ، الشبكة العصبية التلافيفية

RESUME

Resume

Les systèmes de automatiques de classification de l'ECG sont des outils précieux pour aider les médecins et superviser les patients. Les réseaux de neurones profonds ont été largement utilisés comme alternative aux systèmes de classification existants basés sur des phases distinctes d'extraction des caractéristiques et de classification . Dans cette étude, nous proposons un système automatique qui classifie les battements de cœur ECG sur la base d'un réseau de neurones convolutionnels 1D. Pour évaluer le modèle proposé, nous effectuons des tests sur la base de données d'arythmie MIT-BIH et nous considérons quatre classes.

Mots clés : Électrocardiographie (ECG), Les réseaux de neurones profonds, réseau de neurones convolutionnels 1D (CNN).

DEDICATION

I dedicate this work to:

To my Father may God rest his soul,

To my mother,

To my brothers and sisters,

To all my family,

To all my friends ,

To all my teachers .

To everyone I love

May they find here the expression of all my gratitude.

Mohammed seghir

ACKNOWLEDGEMENT

“ Thanks to the one God, Light of the heavens and the earth, who help and guide ”

I would like to thank and express my deep gratitude to my supervisor Pr.M.Nemissi for his availability, explanations, observations, orientations, advice and suggestions during the period of this modest work.

To all professors and to all who taught me a letter throughout my educational career.

Finally, I extend my sincere gratitude and thanks to all my loved ones and friends, who have always supported and encouraged me during the realization of this thesis.

Thank you.

Table of Contents

List of Figures

List of Tables

GENERAL INTRODUCTION

Arrhythmias, which are irregular heartbeats, may show serious problems leading to sudden cardiac death. Consequently, automatic heartbeat classification systems constitute important tools in clinical cardiology, especially in real time and long-time records. Several methods have been proposed for the automatic ECG heartbeat classification.These methods are based on two main phases: feature extraction and classification. Extracting features is generally time consuming and may increase the computational complexity of the overall process. Therefore, deep convolutional neural networks (CNN) have been used to perform both phases. In this work, we address ECG heartbeat classification using 1D-CNN. ECG records from the MIT/BIH arrhythmia database are used for the performance evaluation.

This thesis consists of three chapters

- **Chapter one**

  In this chapter we give the basics of pattern recognition in general, then we present the basics of electrocardiography, the cardiac pathology and ECG acquisition.

- **Chapter two**

  In this chapter, we first present neural networks and their common models: perceptron, adaline and multi-layer perceptron. Then, we detail the training using backpropagation algorithm. Finally, we give some advantages and disadvantages of neural networks.

- **Chapter three**

  In this chapter, we focus on the proposed 1D convolutional neural network. We first present the general 2D-CNN, then we present the used 1D-CNN. Then, we present the well-known MIT-BIH arrythmia database that we use for the evaluation. Finally, we discuss the obtained results over the MIT-BIH arrythmia database.

Finally, we conclude this thesis by a general conclusion where we resume our main contribution.

CHAPTER I

AUTOMATIC CLASSIFICATION OF ECG HEARTBEATS

I.1 Introduction

The process of classifying arrhythmias is generally very difficult for doctors because sometimes it is necessary to analyze each heartbeat of the ECG records during hours, or even days. In addition, human errors can occur during the analysis the ECG records. An important alternative is to use automatic classification systems to assist doctors. This chapter gives some basic definitions about pattern recognition systems and heartbeat classification processes.

I.2 Basics of pattern classification

I.2.1 Definition of pattern classification

Pattern recognition is the scientific discipline whose objective is the classification of objects into a number of categories or classes. Depending on the application, these objects can be pictures or signal wave forms or any sort of estimations that need to be classified [1].

I.2.2 Applications of pattern classification

Pattern recognition is an integral part of most machine intelligence systems built for decision making. This includes:

Machine vision is a range in which pattern recognition is of importance. A machine vision framework captures pictures by means of a camera and analyzes them to produce descriptions of what is imaged.

Character (letter or number) recognition is another critical zone of pattern recognition, with major suggestions in automation and information taking care of. Optical character recognition (OCR) frameworks are already commercially accessible and recognizable to all of us.

Computer-aided diagnosis is another critical application of pattern recognition, aiming at helping specialists in making symptomatic choices. The ultimate diagnosis is, of course, made by the specialist. Computer-assisted determination has been connected to and is of intrigued for a assortment of restorative data, such as X-rays, computed topographic images, ultrasound pictures, electrocardiograms (ECGs), and electroencephalograms (EEGs). The require for a computer-aided determination stems from the truth that medical data are frequently not effectively interpretable, and the translation can depend very much on the aptitude of the doctor.

Speech recognition is another zone in which a lot of investigatation and development effort has been contributed. Speech is the foremost common means by which humans communicate and trade data. In this way, the objective of building intelligent machines that recognize talked information has been a long-standing one for scientists and engineers as well as science fiction writers. Potential applications of such machines are numerous.

Data mining and knowledge discovery in databases is another key application area of design acknowledgment. Data mining is of seriously intrigued in a wide run of applications such as medication and science, advertise and money related investigation, business management, science investigation, picture and music recovery. Its notoriety stems from the reality that within the age of information and information society there's an ever increasing request for recovering data and turning it into information. Moreover, this data exists in gigantic sums of information in various forms including text, images, sound and video, put away in numerous places conveyed all over the world.

Mining for biomedical and DNA data analysis has delighted in a big growth since the mid-1990s. All DNA arrangements include four fundamental building elements; the nucleotides: adenine (A), cytosine (C), guanine (G) and thymine (T). Like the letters in our alphabet and the seven notes in music, these four nucleotides are combined to form long groupings in a bent stepping stool frame. Qualities comprise of usually, hundreds of nucleotides organized in a specific arrange. Particular gene-sequence patterns are related to specific illnesses and play a critical part in medication.

I.2.3 Process of pattern classification:

The basic questions emerging in a classification task [1] are:

- How are the features generated?
- What is the finest number l of features to utilize? This is also an important task and it concerns the feature selection stage of the classification system. In practice, a bigger than essential number of features candidates is generated, and at that point the "best" of them is adopted.
- Having received the suitable features, how can we design the classifier?
- Finally, once the classifier has been designed, how can we evaluate the performance of the designed classifier? That is, what is the classification error rate? This is the assignment of the framework assessment stage.

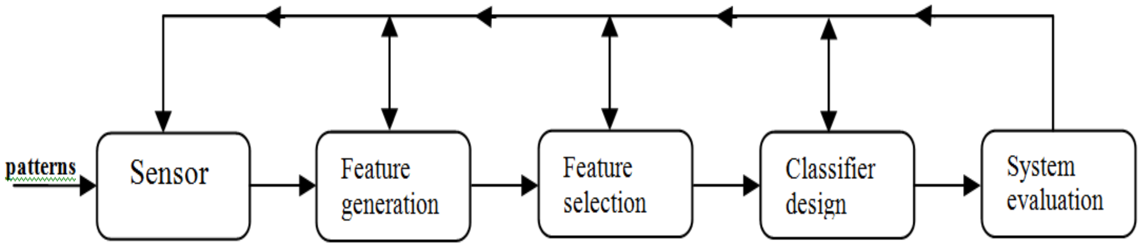

Figure I.1: The basic stages involved in the design of a classification system.

Figure I.1 represents the different phases of designing a classification system. As is clear from the feedback arrows, these stages are not independent. On the contrary, they are interrelated and, depending on the results, we may go back to modify the previous stages in order to enhance the overall performance. Furthermore, there are some approaches that combine phases, for example, the feature selection and the classifier design, in a same optimization problem [1].

I.3 Basics of ECG

I.3.1 Definition of ECG

An electrocardiogram (ECG) is a basic test that may be performed to assess the rhythm and electrical activity of the heart. Sensors placed on the skin detect the electrical impulses generated by the heart every time it beats. A machine records these signals, which a doctor examines to see whether they are normal or not. An ECG may be recommended by a heart expert (cardiologist) or any doctor who suspects heart problem. A professionally qualified healthcare expert can perform the test at a hospital, clinic, or even wearing devices. Despite the name, an ECG is not the same as an echocardiogram, which is a scan of the heart. An ECG is frequently used in conjunction with other tests to help diagnose and monitor cardiac problems. It can be used to look into signs of a cardiac disease, such as chest discomfort, palpitations (rapid heartbeats), dizziness, and shortness of breath. An ECG can aid in the detection of:

- Arrhythmias are conditions in which the heart beats too slowly, too rapidly, or irregularly.
- Coronary heart disease occurs when the blood flow to the heart is obstructed or stopped by a buildup of fatty substances.
- Heart attacks occur when the blood supply to the heart is suddenly cut off.
- Cardiomyopathy is a condition in which the heart walls thicken or expand. A series of ECGs can also be obtained over time to monitor a person who has previously been diagnosed with a heart problem or is receiving medicine that may affect the heart.

I.3.2 Waves, intervals and segments

The ECG graph is made up of a succession of waves or deflections. Each electrocardiographic deflection has been assigned an alphabet letter. As a result, a wave sequence that represents a single cardiac cycle is referred to as P Q R S T and U. (Figure I.2). The entire QRS complex is considered as a unit as it represents a ventricular depolarization. The positive deflection is always referred to as the R-wave. The negative deflection before the R-wave is the Q-wave, while the negative deflection after the R-wave is the S-wave [2].

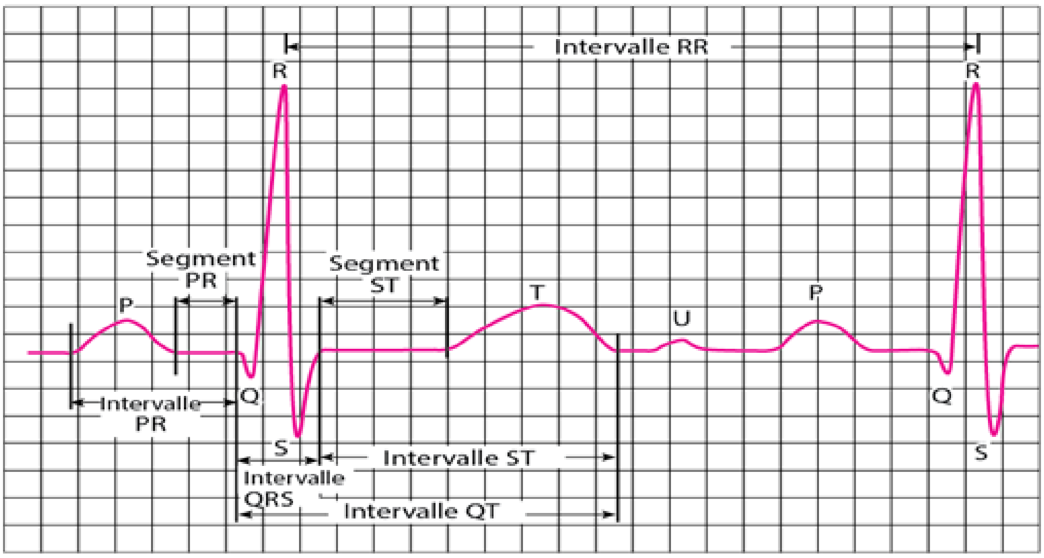

Figure I.2: Waves, segments and intervals

a. Significance of ECG Deflections

P wave : Produced by atrial depolarization.
QRS complex : Produced by ventricular depolarization.
It consists of:

Q wave : First negative deflection before R wave.
R wave : First positive deflection after Q wave.
S wave : First negative deflection after R wave.

T wave : Produced by ventricular repolarization.
U wave : Produced by Purkinje repolarization.

b. The Intervals

When analyzing an ECG diagram, the distances between certain waves are relevant in order to establish a temporal relationship between consecutive events during a cardiac cycle. Because the spacing between waves is expressed on a time axis, these spacings are referred to as ECG intervals. The following ECG intervals are clinically important [2]:
P-R Interval
Q-R Interval

c. The Segments

Two segments occurring between the waves of a single cardiac cycle are clinically important [2]:
P-R Segment
S-T Segment

I.3.3 ECG leads

During activation of the myocardium, electrical forces, or action potentials, propagate in different directions. These electrical forces can be picked up from the body surface using electrodes and recorded in the form of an electrocardiogram. A pair of electrodes, consisting of a positive and a negative electrode, form an electrocardiography lead. Each lead is oriented to record electrical forces as seen from one aspect of the heart. The position of these electrodes can be changed so that different leads are obtained.
There are twelve conventional ECG lead placements that make up the routine 12-lead ECG.
The 12 ECG leads are:
Limb leads or extremity leads, six in number.
Chest leads or precordial leads six in number.

a. The limb leads

Limb leads come from electrodes attached to the limb. Electrodes are attached to each of the three limbs, the right arm, left arm, and left leg. Right leg electrode acts as Ground electrode.

Augmented limb leads – three in number.

Standard limb leads – three in number.

Standard Limb Leads

The standard limb leads get a chart of the electrical strengths as recorded between two limbs at a time. Subsequently, the standard limb leads are too called bipolar leads. In these leads, on limb carries a positive electrode and the other limb carries a negative electrode. There are three standard limb leads :

Lead LI

Lead LII

Lead LIII

Lead	Positive electrode	Negative electrode
I	LA	RA
II	LL	RA
III	LL	LA

Table I.1: Standard Limb Leads

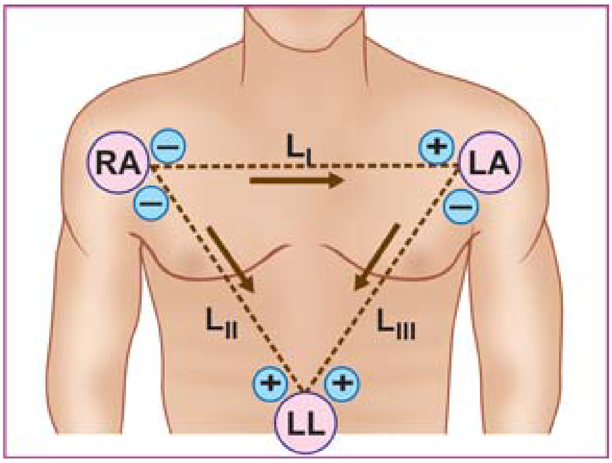

Figure I.3: The three standard limb leads: LI, LII and LIII

Augmented Limb Leads

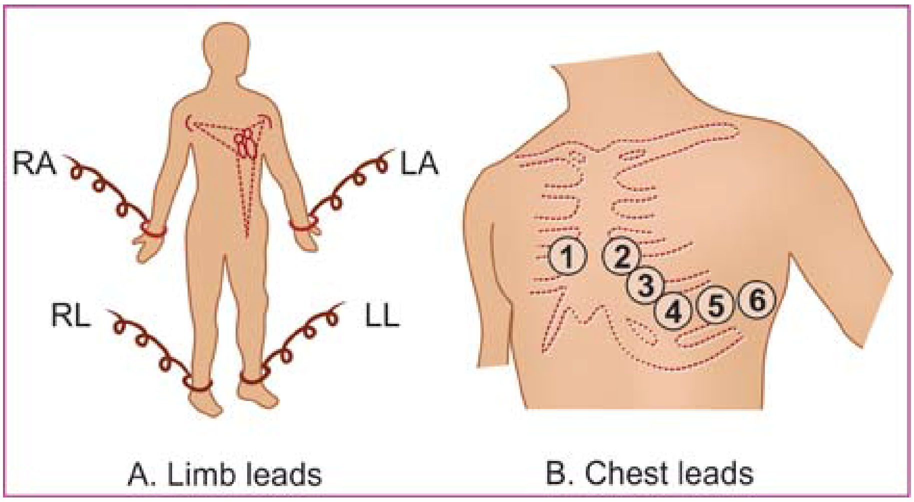

Figure I.4: Electrode placement for ECG recording

The augmented limb leads provide a graph of electrical forces from one limb at a time. As a result, augmented limb leads are often referred to as unipolar leads. One limb of these leads has a positive electrode, whereas the center terminal represents the negative pole, which is at zero potential.There are three augmented limb leads:

Lead aVR (Right arm)
Lead aVL (Left arm)
Lead aVF (Foot left).

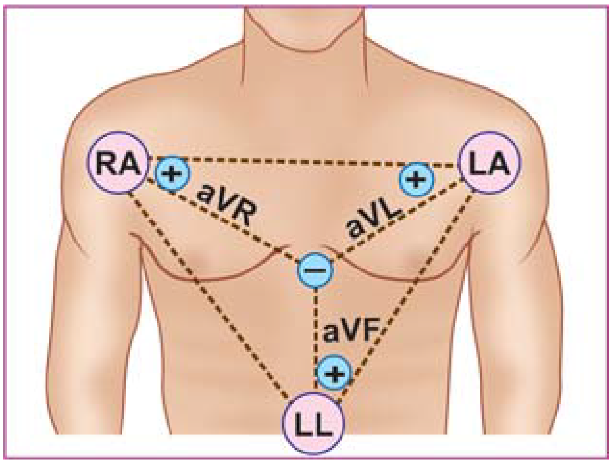

Figure I.5: Electrode placement for ECG recording

b.The chest leads

Electrodes are implanted on the precordium in certain places to get the chest leads. On the left side of the chest, an electrode can be put in six different positions, each position representing one lead . As a result, there are six chest leads namely:
Lead V1: Just to the right of the sternal border, over the fourth intercostal space.

Lead V2: Just to the left of the sternal border, over the fourth intercostal space.
V3: Over a point in the middle of V2 and V4 (see V4 below).
Lead V4: In the midclavicular line, over the fifth intercostal gap.
Lead V5 is located at the same level as lead V4 and crosses the anterior axillary line.
Lead V6 is located at the same level as leads V4 and V5 and crosses the midaxillary line.

I.4 Cardiac pathology

This paragraph very briefly describes the different cardiac pathologies and especially those that may be detected using a Holter record. Our objective in this section is not to analyse precisely the origins of these diseases and their consequences on cardiac function or to describe the treatments that these pathologies require, but simply to relate some abnormal observations of the ECG trace to the most common pathologies. Among the most common diseases, those that affect the heart rhythm and are called cardiac arrhythmias. But before talking about arrhythmias, it is interesting to know the characteristics of the normal rhythm also called sinus rhythm [3].

I.4.1 Sinus rhythm

It is the "normal" rhythm of the heart that corresponds to a physiological activation of the atria, then the ventricles, from the sinus node. The sinus rhythm is characterized by a regular heart rhythm, normally between 60 and 80/ minute in adults at rest. It is characterized on the ECG by a succession of P-waves, followed by ventriculogram (ventricular activity (QRS complex and T-wave)). As shown in Figure I.6

Figure I.6: Normal sinus rhythm

I.4.2 Cardiac arrhythmias

For a normal subject, every heartbeat pushing blood into the arteries (the systolic phase) is triggered by an electrical impulse. This excitation is generated by a group of cardiac cells of a different nature, called the sinus node, located in the right atrium. Arrhythmia occurs when electrical excitation occurs elsewhere than in the sinus node, in the atrioventricular node or ventricles, for example, or when the electrical pulse no longer follows normal pathways.

I.4.3 Types of common arrhythmias

At rest, the heart normally beats between 60 and 80 beats per minute (this is called the pulse or heartbeat). An acceleration (tachycardia), a slowdown (bradycardia) or a change in heart rhythm (irregular rhythm) are the forms of arrhythmia.

The arrhythmia is sometimes accompanied by palpitations. Palpitations are not a heart rhythm disorder, but simply the feeling of the heartbeat. This most often occurs when the pulse is beating too fast or irregularly. In addition, there may be an arrhythmia without the patient feeling palpitations (so it is not known that the heart beats abnormally). There are various forms of arrhythmia, the main ones being:

Extrasystole:

Extrasystole is a premature heartbeat or excess that is felt as an erratic or missing heartbeat, as if the heart were missing a turn. This is the most common arrhythmia. Many people have extrasystoles that they don't even perceive. If these are not accompanied by other symptoms, which is the majority of cases, this abnormality is benign and can occur in a healthy heart. Sometimes the extrasystole is accompanied by a brief dizziness, but it is not serious.

Atrial flutter (atrial):

In atrial flutter, abnormal depolarization constantly travels a loop path in the right atrium (usually rising along the inter-atrial septum, descending on the outer wall of the right atrium, the left atrium being depolarized secondarily). The rotation frequency is 300 beats/min. Depolarization will therefore occur at the entrance of the atrioventricular node 300 times per minute and will cross the junction to the ventricles only once in two or three, or less. The frequency of the ventricles will therefore be a submultiple of 300 bat/min. This continuous activity of the atria is registered on the ECG by very characteristic ear waves called F waves: they have a serrated appearance in D2, D3 and aVF (see figure I.7) Note that, in this type of pathology, there is a major risk for the patient's health, due to the possibility of embolism generation, due to turbulence in the blood flow in the atria.

Atrial fibrillation (atrial fibrillation):

Atrial fibrillation is a supraventricular arrhythmia without any organization. Depolarization is divided into a multitude of fronts of different directions and amplitudes, performing an electrical activity at the level of the atria totally disordered.

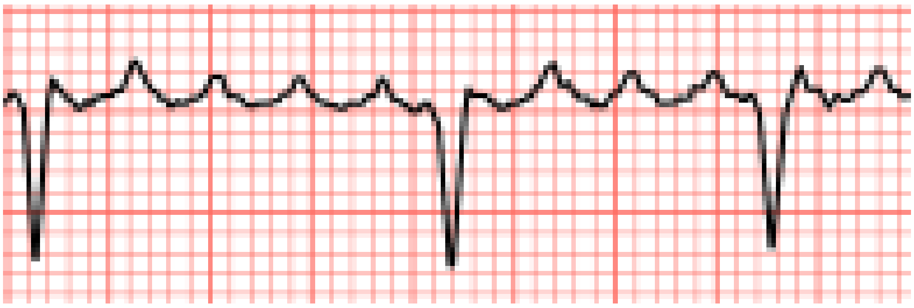

Figure I.7: Atrial flutter (atrial)

This activity, most often sustained, leaves the atrial myocardium no electrical rest. It is translated on the ECG by the disappearance of organized ear waves in favor of a continuous activity resembling a kind of irregular sinusoid. The frequency of depolarization is variable from one point to another within the atria, but in any case most often very high. Multiple depolarization fronts are thus presented at the ventricular auriculo node which fulfils its role of filter by letting pass randomly only a few fronts, the frequency of ventricles becoming completely irregular, generally between 90 and 140/min depending on the permeability state of the node.

The absence of atrial systoles is not serious in itself, but it is nevertheless responsible for a significant decrease in cardiac efficiency at two levels: on the one hand, because the heart does not benefit from the atrial systole which provides part of the blood filling of the ventricles, and, on the other hand, because the average rhythm is generally greater than 100 bat/min and can reach 200 bat/min. The major risk associated with this pathology is that of flutter, that is to say, the possible sending, in traffic, of emboles formed at the level of the atria (especially harmful if it is the left atrium, since the left ventricle sends blood into the general circulation, and in priority to the heart and brain). This risk is relatively low when an AF is permanent, while it is increased during episodes of FA on a sinus background (paroxysmal FA), especially when moving from one rhythm to another.

Supraventricular tachycardia (atrial or junctional):

Atrial tachycardia can originate from an ectopic focal spot, a stimulation loop (flutter) or a pathway that short-circuits the AV pathway, known as the accessory pathway, with re-entry through the AV node.

In the case of an ectopic focus, it is a group of cells located in the atria, which depolarize spontaneously and faster than the sinus, thus taking its place. The depolarization of the atria is not of sinus origin, the propagation of the nerve impulse differs from that of the sinus, and an unusual P-wave is observed. The typical frequency of discharge of this type of focal spot is between 120 and 200, in the absence of AV conduction problems, the ventricles are trained at the same rate. Regular discharge from a localized ectopic focal spot in the AV node can also be the cause of tachycardia, called junctional tachycardia (nodal), in which case the frequency of beats can reach 250 bat/min. Unlike atrial tachycardia, no P-wave precedes QRS complexes as there is no atrial activity before the beat (there may be retrograde atrial activity with the trace on the ECG embedded in the QRS complexes). The main risk of this type of pathology is the lack of efficiency of the ventricles which, forced to contract very frequently, do not have the time to fill themselves properly with blood: the body's supply of oxygen can be altered.

ventricular fibrillation:

It's the physiological equivalent of atrial fibrillation, but it matches the ventricles. The ventricles then discharge completely out of sync, and there is no more cardiac systole. The ECG registers irregular, anarchic, fast ventricular activity, taking an oscillatory appearance. (see figure I.8). Ventricular fibrillation is therefore a particularly serious arrhythmia,

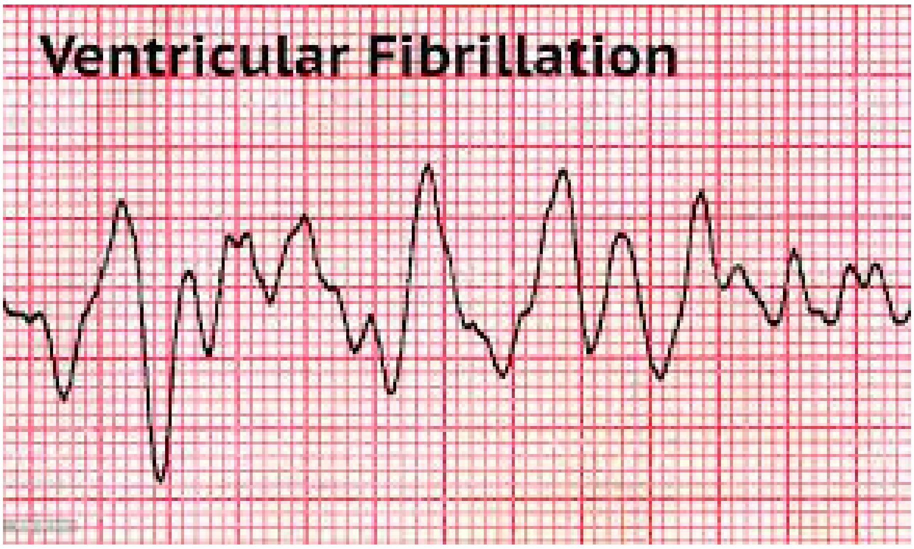

Figure I.8: ventricular fibrillation

since it is a threat of sudden death, in fact the heart does not at all perform its pumping work, the blood no longer circulates, which leads to asphyxiation of all the tissues of the body, including the myocardium itself. Without an immediate intervention (defibrillation), likely to resynchronize the depolarization of the myocardial cells and thus restart the cardiac movement, death ensues. People who present such risks can now benefit from the implantation of a defibrillator: placed at the level of the chest, like a pacemaker, it is equipped with a probe that can detect the rhythmic anomaly and lead the device to deliver a strong electric discharge

ventricular tachycardia:

Ventricular tachycardia originates from one or more ventricular ectopic focal(s) (which depolarize in turn). The beats are thus in the form of very close ventricular extrasystoles (Figure I.9). This type of rhythm is dangerous because of its possible evolution into

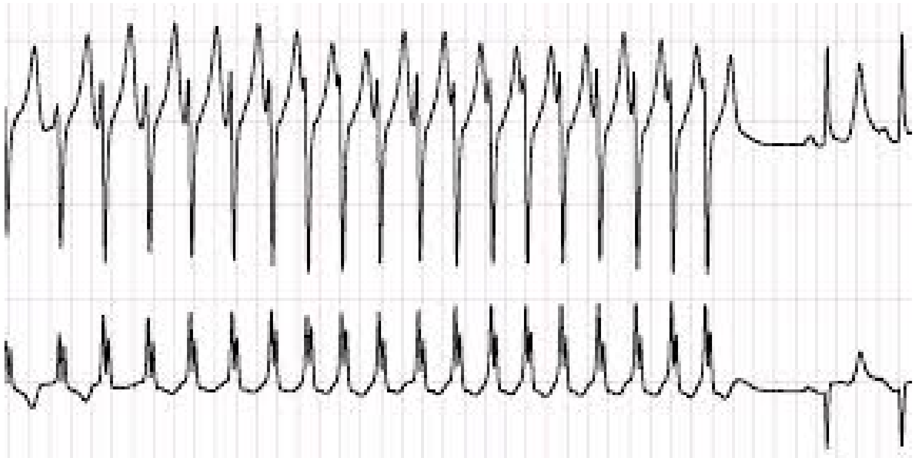

Figure I.9: ventricular tachycardia

ventricular fibrillation, which leads to the death of the patient if it is not treated with a defibrillator within a few minutes of its appearance

I.4.4 Blocks

Cardiac blocks are caused by a rupture of the electrical impulse conduction in the heart cells. These ruptures can be: Slowing: lengthening of the journey time, Intermittent: the blocking of the conduction is done randomly Complete: no conduction. At the level of the rapid conduction pathways of nerve impulses in the heart, there are essentially three types of blocks classified according to their locations : the sino-auricular block (SA) located between the sinus and the atria, the atrioventricular block (AV) between the atria and ventricles, and the branch block (right and/or left) at the level of transmission to each of the two ventricles; it is mainly the first two (SA and AV blocks) that are likely to cause rhythm disturbances.

Sinoatrial block

In the case of the SA block, the electrical impulse from the sinus is not transmitted to the atria, so the atrial and ventricular muscles do not contract. On the ECG, we observe the absence of a beat where there should be one regularly. In addition, if the block is well installed, it is possible that several pulses of the sinus are not transmitted, in such a case, we often observe the relay by one or more pacemakers of reserve (ectopic foci).

Atrioventricular block

In the case of an AV block, the influx propagates correctly at the level of the atria but is not transmitted to the ventricles: the contraction of the atria is not followed by the contraction of the ventricles. This type of block is usually characterized by its degree of severity:

An AV block of order I corresponds to the lengthening of the PR distance, that is, the conduction between the atria and the ventricles is difficult, slow, but not absent. Each P-wave is followed by a QRS complex; the rhythm is regular.

An AV-Block II blocks certain pulses from the atria, which then do not give birth to ventricular contractions (Figure I.10); the ECG trace occasionally presents isolated P-waves, not followed by QRS complexes.

An AV block of order III corresponds to the total absence of conduction between the atria and the ventricles. The activities of these two parts of the heart, usually correlated, are, in this case, totally independent. The sinus continues to ensure the regularity of the atrial beats, but the absence of transmission of the atrial influx to the ventricles leads an ectopic focus, at the junction node (AV) or ventricular level, to take control of the ventricles, which, depending on the case, can introduce a regular rhythm or not, but certainly slower than the sinus rhythm.

Figure I.10: Bloc AV d'ordre II

The irregularities of the rhythm are therefore mainly observable during the blocks AV I and AVII, and especially during an intermediate situation where we find a so-called Wenckebach period, which corresponds to a periodic lengthening of the PR distance resulting in a P wave not driven, followed by a P wave driven, etc. This pathology thus introduces a regular irregularity of the rhythm.

Branch blocks

Anatomically, the atrioventricular conduction, after having been filtered by the Tawara node, borrows the trunk of the His beam which is divided into 2 branches: a right branch to the right ventricle and a left branch to the left ventricle it-even dividing into 2 branches is the anterior and posterior branch. These branches end in multiple branches called the Purkinje Network.

A simple slowdown or complete blockage of conduction in one of these branches is called a branch block. This conduction anomaly therefore causes a delay in depolarization of the ventricle, or enlargement of the QRS. We can distinguish classically:

The right branch blocks

The interruption of the seat conduction at the right branch while the transmission of the influx is normally done through the left branch, the right ventricle is thus activated in a delayed and abnormal way.

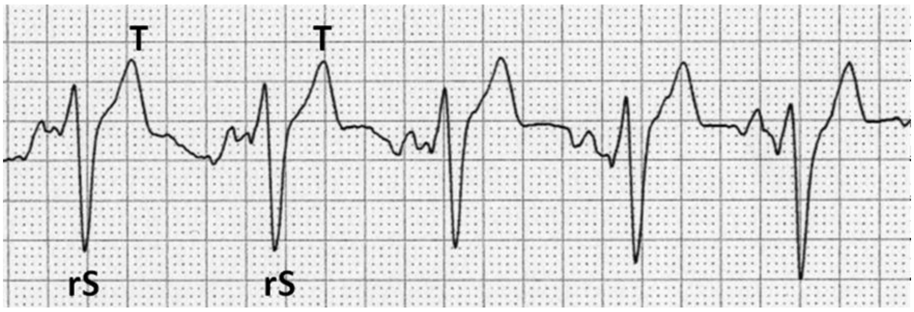

Figure I.11: Right Branch Block

The left Branch Block

The conduction disorder relates here to the left branch while the right branch is healthy, it is the left ventricle that presents a delay of activation, this being done from the right ventricle that is normally activated.

Figure I.12: left Branch Block

I.5 ECG acquisition

In this section we describe an example of an ECG acquisition system, “K and H products” [4]. As described above, the right foot is always used as a point of reference to the mass. From the test points on the right arm, left arm and left foot, six ECG lead signals, including leads I, II, III, aV_R ,aV_L and aV_F can be performed.

The circuit is single channel with the possibility of selecting multiple leads. In general, the frequency is between 0.1 and 100 Hz and the maximum amplitude is 1 mV in a normal ECG signal. In addition, to avoid electrical shock caused by a leaking power supply or test instrument, the isolation circuit must be considered for ECG detection. Figure I.6

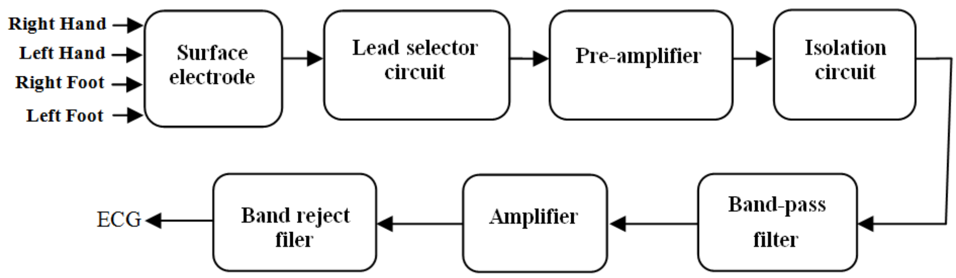

Figure I.13: Block diagram of ECG measurement circuit

shows the block diagram of the ECG measurement circuit. In ECG measurement, surface electrodes on four ends are used to capture very low and time-varying potentials. The conductor selection circuit contains a voltage tracking circuit to adapt the impedance between the electrode and the skin, this design can increase the measurement sensitivity.

I.5.1 Preamplifier Circuit

Figure I.7 shows the preamp circuit consisting of an instrumentation amplifier. If $Z_{11} = Z_{12}$, $Z_{13} = Z_{14}$ and $Z_{15} = Z_{16}$, then the voltage gain can be determined by:

$$Av = \frac{Z_{15}}{Z_{13}}\left(1 + \frac{2Z_{11}}{Z_{10}}\right) \tag{I.1}$$

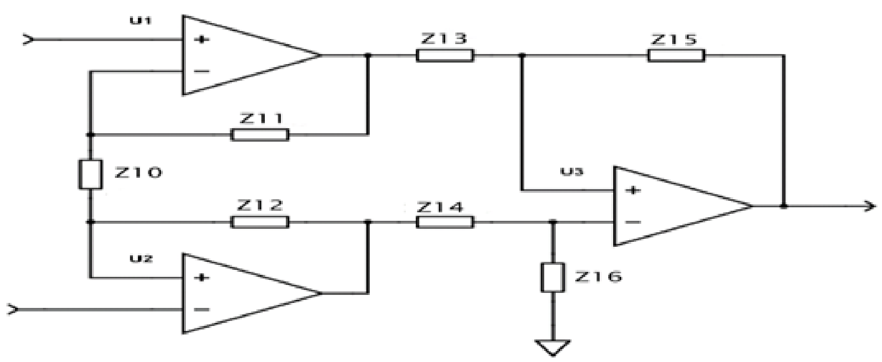

Figure I.14: Preamplifier Circuit

I.5.2 Band-Pass Filter Circuit

In circuit design, the operational amplifier is used to create a 2nd order active high pass filter as shown in Figure I.9 (a). The filter cut-off frequency (f_L) is set to 0.1 or 1 Hz and can be calculated using Z_{17}, Z_{18}, Z_{19} and Z_{20} as shown in equation

$$f_L = \frac{1}{2\pi\sqrt{Z_{17}Z_{18}Z_{19}Z_{20}}} \tag{I.2}$$

And the bandwidth gain is explained in the following equation:

$$\frac{(Z_{21} + Z_{22})}{Z_{21}} = 1.56 \tag{I.3}$$

As shown in Figure I.9 (b), the active second-order low-pass filter has a cut-off frequency (fH) of 100 Hz and can be calculated using Z_{25}, Z_{26}, Z_{27} and Z_{28} as shown in this equation:

$$f_H = \frac{1}{2\pi\sqrt{Z_{25}Z_{26}Z_{27}Z_{28}}} \tag{I.4}$$

So the bandwidth gain is as follows:

$$\frac{(Z_{29} + Z_{30})}{Z_{29}} = 1.56 \tag{I.5}$$

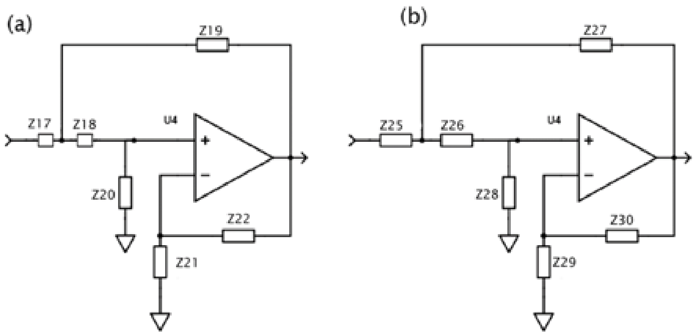

Figure I.15: Band-Pass Filter Circuit

I.5.3 Isolation Circuit

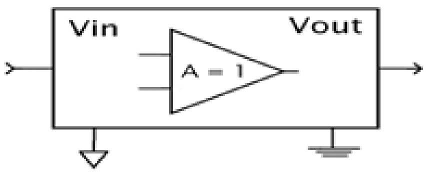

Figure I.16: Isolation Circuit

As shown in Figure I.8, signal isolation is achieved by an optical approach.

I.5.4 Amplifier Circuit

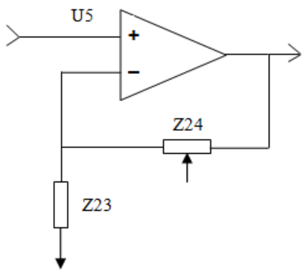

Figure I.17: Amplifier Circuit

Figure I.10 shows a non inverter amplifier. In the amplifier, Z_{24} is used for gain adjustment as shown below:

$$Av = \frac{Z_{13} + Z_{24}}{23} \tag{I.6}$$

I.5.5 Band-Reject Filter Circuit

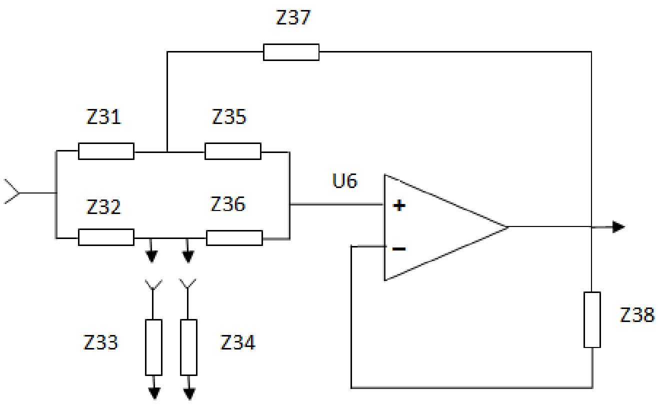

Figure I.18: Band-Reject Filter Circuit

Figure I.11 shows a dual T-band filter consisting of RC networks, including the operational amplifier Z_{31}, Z_{32}, Z_{33} (or Z_{34}), Z_{35}, Z_{36} and Z_{37}. If Z_{31}= Z_{35}, $Z_{32} = Z_{36}$, $Z_{33} = 12Z_{31}$ (or $Z_{34} = 12Z_{31}$) and $Z_{37} = 2Z_{32}$. The center frequency can be calculated using the following equation:

$$f = \frac{1}{2\pi Z_{31} Z_{32}} \tag{I.7}$$

I.6 Systems of ECG heartbeat Classification

I.6.1 Feature extraction

The effectiveness of the heartbeat classification of the arrhythmia using the ECG signal relies heavily on the feature extraction step. Any information taken from the heartbeat that is utilized to distinguish the kind of heartbeat might be termed a feature. The features can be retrieved in various ways directly from the morphology of the ECG signal in the time domain and/or frequency domain, as well as from the cardiac rhythm.

The most common feature discovered in the literature is determined from cardiac rhythm (or heartbeat interval), also known as RR interval. The RR interval is the time between two beats. it is the distance between the R peak of one beat and the R peak of another, which might be its precursor or successor.

Other features based on heartbeat intervals including other distances between the fiducial points of a heartbeat, have been reported in the literature.

The features that have produced the best accuracies in literature to date include features taken from the domain of time/frequency, as well as aspects of the RR interval. [5].

I.6.2 Classification

After extracting and selecting features, they can be used to design a classification model. Techniques from machine learning domain can then be used to design a system for arrhythmia heartbeat classification. Among the most common models used as classifiers [5]: Support vector machines (SVM), artificial neural networks (ANN), and linear discriminant analysis (LDA). They have been successfully used for this purpose.

I.7 Conclusion

In this chapter, we focused on systems of automatic ECG heartbeat classification. These systems include two important phases: feature extraction and classification. Many approaches have proposed for these two steps. In this work, we aim to introduce a system for ECG heartbeat classification using deep neural networks. The next chapter constitutes an introduction to artificial neural networks.

CHAPTER II

NEURAL NETWORKS

II.1 Introduction

The discipline of neural networks models the human mind. The normal human mind comprises of almost 10^{11} neurons of different kinds, with every neuron associating with up to huge number of synapses. In that capacity, neural network models are additionally called connectionist models. Mental capacities, including language, theoretical thinking, also, learning and memory, address the most perplexing mind activities to characterize as far as brain systems. During the 1940s, McCulloch and Pitts observed that a neuron can be demonstrated as a basic treshold gadget to perform rationale work. In 1949, Hebb proposed the Hebbian rule to portray how learning influences the synaptics between two neurons. Rosenblatt proposed the perceptron model, and Widrow and Hoff proposed the adaline (versatile direct component) model, prepared with a most un-mean squares (LMS) technique [6].

In 1969, Minsky and Papert demonstrated numerically that the perceptron can't be utilized for complex rationale work. This significantly melted away the interest in the field of neural networks.

The most unmistakable milestone in neural network research is the backpropagation(BP) learning algorithm proposed for the multilayer perceptron (MLP) model in 1986 by Rumelhart, Hinton, and Williams . The neuron, or nerve cell, is the basic physical and utilitarian unit of the sensory system including the mind. A neuron has four parts: the dendrites, the soma (cell body), the axon, and the synapse. A soma contains a cell nucleus. Dendrites branch into a ragged organization around the cell to get input from other neurons, though the axon loosens up for a significant distance, ordinarily a centimeter also to the extent that a meter in outrageous cases. The axon is a result channel to other neurons; it branches into strands and substrands to interface with the dendrites and cell groups of different neurons. The interfacing intersection is known as a neurotransmitter. Each cortical neuron gets $10^4 - 10^5$ synaptic associations, with most information sources coming from far off neurons.

II.2 Biological neuron

The nervous system is thought to be over 1000 billions of interconnected neurons. Although neurons are not all identical, their form and some characteristics make it possible to divide them into a few large classes. Indeed, it is as important to know, that the not all neurons have similar behavior in function their position in the brain.

The cell body:

It contains the nucleus of the neuron as well as the machine necessary for the synthesis of enzymes. This body spherical or pyramidal cell form also contains the other molecules essential to the life of the cell. Its size is few microns in diameter.

Dendrites:

These are thin tubular extensions that branch around and form a kind of vast tree. The signal sent to the neuron are captured by the dendrites. Their size is a few tens of microns in length.

The axon:

It is along the axon that the signals leave the neuron. Unlike dendrites that branch around the neuron, the axon is longer and branches at its end or it connects dendrites from other neurons. Its size can vary between a few millimeters to several meters.

Synapse:

A synapse is a junction between two neurons and usually between the axon of one neuron and a dendrite of another neuron (but there are also axo-axonal synapses for example)

II.3 Artificial neuron

II.3.1 Architecture

A neuron is an essential handling unit in a neural network. A node processes all fan-in from different nodes and creates a result as indicated by a transfer function called the activation function. The activation function addresses a direct or nonlinear mapping

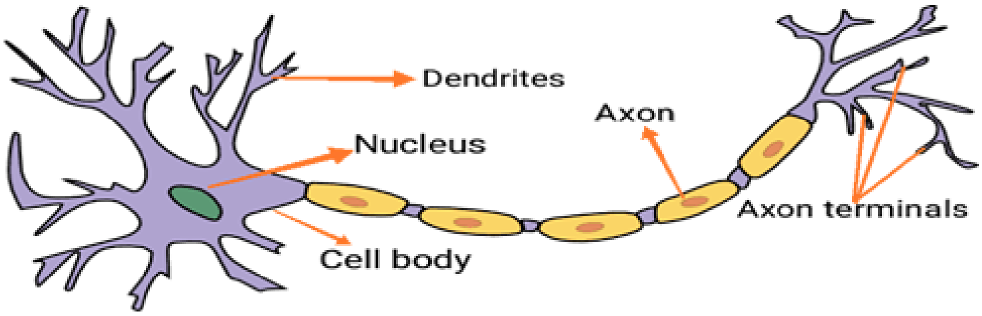

Figure II.1: Biological neuron

from the contribution to the result and is indicated by $\phi(\bullet)$. The variable synapses are demonstrated by weights. The McCulloch-Pitts neuron model , which utilizes the sigmoid activation function, was enlivened organically. Figure II.2 outlines the straight forward McCulloch-Pitts neuron model [7]. The result of the neuron is given by:

$$net = \sum_{i=1}^{J1} w_i x_i - \theta = w^T x - \theta \tag{II.1}$$

$$y = \phi(net) \tag{II.2}$$

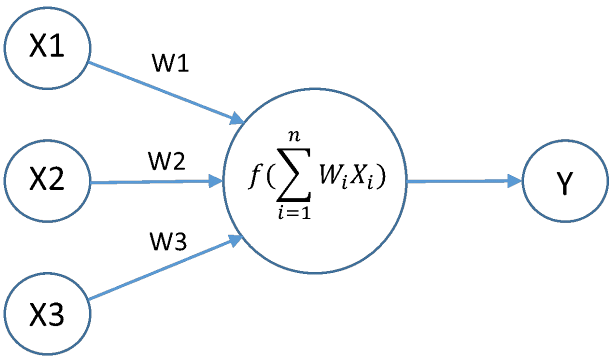

Figure II.2: Artificial Neuron

Where x_i is the ith input,w_i is the connection weight from the ith input,$w = (w_1, \ldots, w_{J1})^T$ $x = (x_1, \ldots, x_{J1})^T T$, is an edge or inclination, and J_1 is the quantity of information sources. The activation function $\phi(\bullet)$ is generally some continuous or discontinuous function, mapping the genuine numbers into the interval (-1, 1) or (0, 1).

II.3.2 Activation functions

Activation functions are functions utilized in neural networks to computes the weighted entirety of input and biases, of which is utilized to choose in case a neuron can be terminated or not. It controls the displayed information through a few gradient handling usually gradient plummet and afterwards create an output for the neural organize, that contains the parameters within the information. These AFs are regularly referred to as a transfer function in some literature.

The most commonly used activation functions.

a. Softmax

The softmax function is a function that turns a vector of K real values into a vector of K real values that sum to 1. The input values can be positive, negative, zero, or greater than one, but the softmax transforms them into values between 0 and 1, so that they can be interpreted as probabilities. If one of the inputs is small or negative, the softmax turns it into a small probability, and if an input is large, then it turns it into a large probability, but it will always remain between 0 and 1.

$$\sigma(z)_i = \frac{e^{zi}}{\sum_{j=1}^{k} e^{zj}} \tag{II.3}$$

b. ReLU (Rectified Linear Unit)

The rectified linear unit (ReLU) activation work was proposed by Nair and Hinton 2010, and ever since, has been the foremost broadly utilized activation function for deep learning applications with state-of-the-art comes about to date [8]. The ReLU could be a speedier learning AF , which has demonstrated to be the foremost effective and broadly utilized work . It offers the way better execution and generalization in profound learning compared to the Sigmoid and tanh actuation capacities . The ReLU represents a nearly linear function and therefore preserves the properties of straight models that made them simple to optimize, with gradient-descent strategies [9]. The ReLU activation function performs a limit operation to each input component where values less than zero are set to zero in this way the ReLU is given by

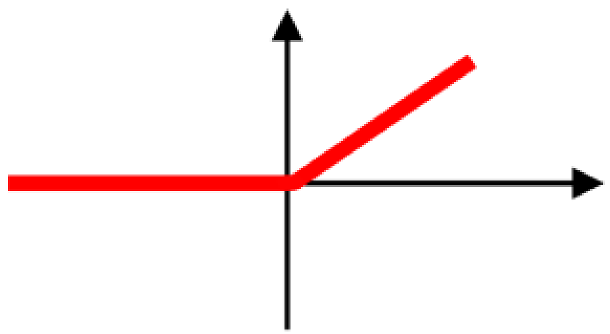

Figure II.3: ReLU (Rectified Linear Unit)

$$f(x) = \begin{cases} 0, & \text{si } x < 0 \\ x, & \text{si } x \geq 0 \end{cases} \tag{II.4}$$

c. Sigmoid

The Sigmoid AF is some of the time alluded to as the logistic function or squashing function in some writing . The Sigmoid function inquire about results have delivered three variations of the sigmoid AF, which are utilized in DL applications. The Sigmoid could be a non-linear AF used mostly in feedforward neural networks. It may be a bounded differentiable genuine function, characterized for real input values, with positive derivatives all over and a few degree of smoothness . The Sigmoid function is given by the relationship

$$f(x) = \frac{1}{1 + e^{-x}} \tag{II.5}$$

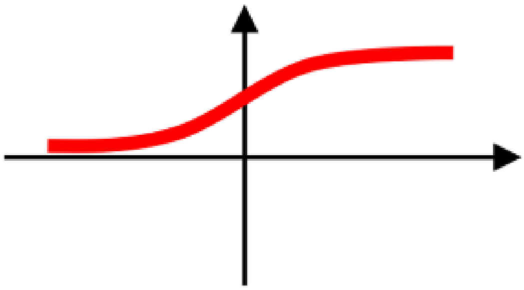

Figure II.4: Sigmoid activition function

d. ArcTangente (tanh(x))

The hyperbolic tangent function is another type of AF used in DL and it has some variants used in DL applications. The hyperbolic tangent function known as tanh function, is a smoother zero-centred function whose range lies between -1 to 1, thus the output of the tanh function is given by

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{II.6}$$

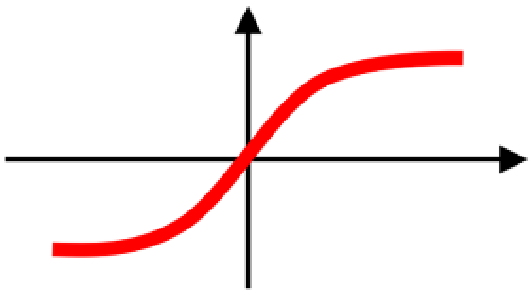

Figure II.5: ArcTangente activition function (tanh)

II.4 Perceptron

II.4.1 Architecture

The perceptron is a binary formal neuron. This means that its only output is either 0, or 1, corresponding to two predictable classes. All its inputs are connected to its exit. The action of a neuron on its inputs (x_i) corresponds to an aggregation function, which in the case of the perceptron is a weighted sum. To calculate it, parameters are associated with each input (i), which we call: weight (w_i). The perceptron aggregation function is expressed as follows: $\sum_{i=1}^{n} w_i x_i$

Its result is called the aggregation value, noted z. Today, the perceptron is a very simplistic architecture. Its limit is its capacity to solve only linearly separable problems.

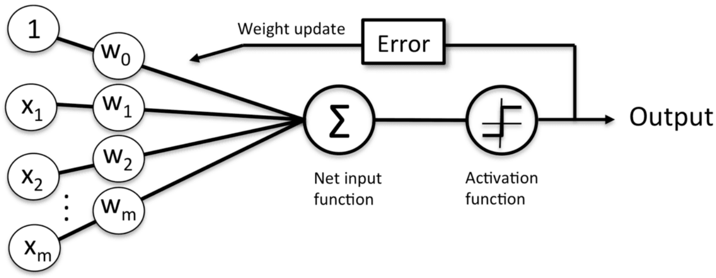

Figure II.6: Perceptron

II.4.2 Training

The perceptron is trained through steps as follows :

Step 1: Initialize the weights W_1,W_2...W_n and threshold to small random values.
Step 2: Present new input X_1,X_2,..X_n and desired output d_k .
Step 3: Calculate the actual output based on the following formula:

$$y_k = f_h(\sum_{i=1}^{n}(X_i W_i) - \theta_k) \tag{II.7}$$

Step 4: Adapt the weights according to the following equation:

$$W_i(new) = W_i(\text{ old }) + \eta\,(d_k - y_k)\,x_i, 0 \leq i \leq N \tag{II.8}$$

Where η is a positive gain fraction less than 1 and k_d is the desired output. Note that the weights remain the same if the network makes the correct decision.
Step 5: Repeat the procedures in steps $2-4$ until the classification task is completed.

II.5 Adaline

II.5.1 Architecture

Adaptive Linear Neuron is a network of single-layer artificial neurons. It was developed by Professor Bernard Widrow and one of his students, Ted Hoff, of Stanford University in 1960. Adaline relies on the formal neuron of McCulloch and Pitts. It consists of a synaptic weight, a bias (a constant that is added to the input) and a summation function.

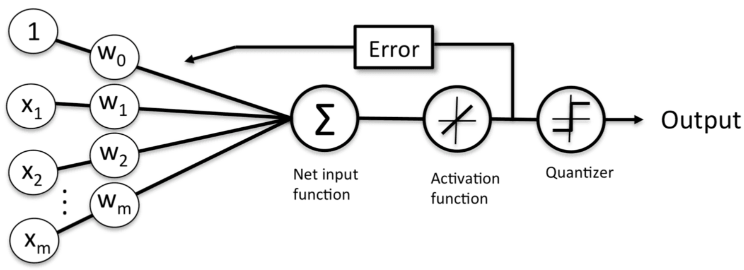

Figure II.7: Adaline

II.5.2 Training (Delta Rule)

The delta rule updates the weights between the connections so as to minimize the difference between the net input to the output unit and the target value. We write the weight update in each iteration as: $w_j{=}w_j{+}w_j$
Where

$$\Delta w_j = \eta\left(\text{target}^{(i)} - output^{(i)}\right) x_j^{(i)} \tag{II.9}$$

- Step 1- initialize the weights, bias and learning rate to start the training.

- Step 2- Continue step 3-8 when the stopping condition is not true.
- Step 3- Continue step 4-6 for every bipolar training pair s: t.
- Step 4- Activate each input unit as follows

$$X_i = S_i(\text{ where } i = 1 \text{ to } n) \tag{II.10}$$

- Step 5- Obtain the net input with the following relation

$$y_{in} = b + \sum_{i}^{n} x_i w_i \tag{II.11}$$

- Step 6- Apply the following activation function to obtain the final output

$$f(y_{in}) = \begin{cases} 1 \text{ if } & y_{in} \geq 0 \\ -1 \text{ if } & y_{in} < 0 \end{cases} \tag{II.12}$$

- Step 7- Adjust the weight and bias as follows -Case 1. If y t then

$$w_i(\text{ new }) = w_i(\text{ old }) + \eta (t - y_{in}) x_i \tag{II.13}$$

$$b(\text{ new }) = b(\text{ old }) + \eta (t - y_{in}) \tag{II.14}$$

-Case 2. if y=t then

$$w_new = w_old \tag{II.15}$$

$$b(new) = b(old) \tag{II.16}$$

Where y is the actual output and t is the desired/target output.
$(t - y_{in})$ is the computed error.

- Step 8 – Test for the stopping condition, which will happen when there is no change in weight or the highest weight change occurred during training is smaller than the specified tolerance.

II.6 Multi-layer Perceptron (MLP)

II.6.1 Architecture

The MultiLayer Perceptron (MLP) principle [Rumelhart et al. 1985], is to organize in collaboration several simple perceptrons so as to use the output of a neuron as input to one or more others. Several neurons can be present side by side, we speak of neuronal layer and they can be connected to several neurons of a next layer. A multilayer perceptron is necessarily equipped with an input neural layer and an outlet layer. It can also be supplemented by one or more intermediate layers, called hidden layers.

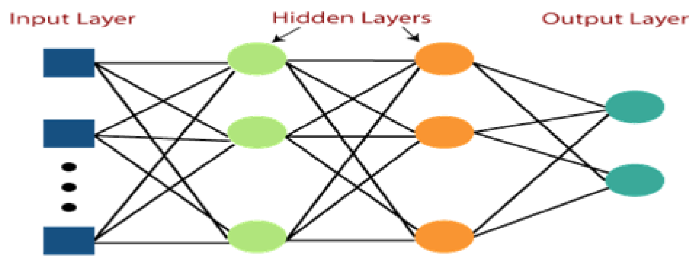

Figure II.8: Multi-layer Perceptron (MLP)

II.6.2 Training using Backpropagation

(BP, Backpropagation) [10], is one of the most common methods and most commonly used for learning neural networks. BP consists of minimize the distance between the calculated output $Z^{(q)}$ and the desired output $T^{(q)}$ corresponding to each learning example $X^{(q)}$. Backpropagation of errors is basically just gradient descent. Mathematically speaking, backpropagation is:

$$w_{updated} = w_{old} - \eta \nabla E \tag{II.17}$$

Quadratic error is often used as being the cost function of BP. For a set of Q learning examples, the error total quadratic is given by:

$$E = \sum_{q=1}^{Q} \sum_{j=1}^{J} \left(t_j^{(q)} - z_j^{(q)} \right)^2 \tag{II.18}$$

II.6.3 Equation and Algorithm

The BP algorithm consists of four main steps. After there the random initialization of all the network weights, the forward propagation algorithm to calculate the outputs, then back to calculate the necessary corrections and finally update the weights.
The updates of the output weights, $u_{mj}^{(r+1)}$ and hidden weights, $w_{nm}^{(r+1)}$ at iteration (r + 1), are given by:

$$u_{mj}^{(r+1)} = u_{mj}^{(r)} - \eta \frac{\partial E^{(r)}}{\partial u_{mj}} \tag{II.19}$$

$$w_{nm}^{(r+1)} = w_{nm}^{(r)} - \eta \frac{\partial E^{(r)}}{\partial w_{nm}} \tag{II.20}$$

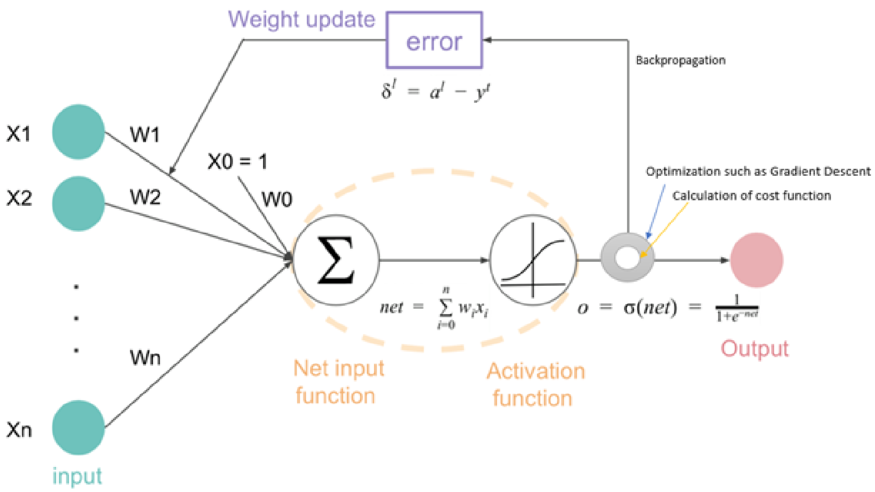

Figure II.9: Process of Backpropagation

Updating these weights in the case of a unipolar sigmoid, with a value between 0 and “1”, are given as follows:

$$S_j = \sum_{(m=1.M)} u_{mj} y_m \tag{II.21}$$

$$\partial E/\partial u_{mj} = (\partial E/\partial s_j)\,(\partial s_j/\partial u_{mj}) \tag{II.22}$$

$$= [(\partial E/\partial z_j)\,(\partial z_j/\partial s_j)\,(\partial s_j/\partial u_{mj})] \tag{II.23}$$

$$= \left[\left[(\partial/\partial z_j)\left(\sum\nolimits_{(p-1j)} (t_p - z_p)^2\right)(\partial/\partial s_j)\, g\,(s_j)\right][\partial s_j/\partial u_{mj}]\right] \tag{II.24}$$

$$= -2\,(t_j - z_j)\, g'\,(s_j)\, y_m \tag{II.25}$$

But

$$g'\,(s_j) = (d/ds_j)\, g\,(s_j) = (d/ds_j)\,[1 + \exp\,(-s_j + b)]^{-1} \tag{II.26}$$

$$= [z_j]^2\,[1/z_j - 1] = [z_j]^2\,[(1 - z_j)\,/z_j] = z_j\,(1 - z_j) \tag{II.27}$$

$$= z_j\,(1 - z_j) \tag{II.28}$$

So

$$\partial E/\partial u_{mj} = -2\,(t_j - z_j)\, z_j\,(1 - z_j)\, y_m \tag{II.29}$$

The weight $[w_n m]$ is derived using:

$$r_m = \sum_{(n=1.N)} w_{nmj} x_n \tag{II.30}$$

Afterward

$$\partial E/\partial w_{nm} = (\partial E/\partial r_m)\,(\partial r_m/\partial w_{nm}) \tag{II.31}$$

$$= \left[(\partial/\partial y_m) \left(\sum (j = 1.J) \, (t_j - z_j)^2 \right) (\partial/r_m) \, y_m \right] [\partial r_m/\partial w_{nm}] \tag{II.32}$$

$$= \left\{ \sum (j = 1.J)(-2) \, (t_j - z_j) \, [z_j \, (1 - Z_j)] \, [u_{mj}] \right\} [y_m \, (1 - y_m)] \, [x_n] \tag{II.33}$$

Then

$$w_{nm}(r+1) = w_{nm}(r) + \eta \left\{ = \left\{ \sum (j = 1.J) \, (t_j - z_j) \, [z_j \, (1 - z_j)] \, [u_{mj}] \right\} [y_m \, (1 - y_m)] \, [x_n] \right\} \tag{II.34}$$

To recap, the learning equations of the unipolar update are:

$$u_{mj} \leftarrow u_{mj} + \eta_1 \left(t_j^{(q)} - z_j^{(q)} \right) z_j^{(q)} \left(1 - z_j^{(q)} \right) y_m^{(q)} \tag{II.35}$$

and

$$w_{nm} \leftarrow w_{nm}^{(r)} + \eta_2 \left\{ \sum (j = 1.J) \left(t_j^{(a)} - z_j^{(a)} \right) \left[z_j^{(q)} \left(1 - z_j^{(q)} \right) \right] [u_{mj}] \right\} \left[y_m^{(a)} \left(1 - y_m^{(q)} \right) \right] \left[x_n^{(q)} \right] \tag{II.36}$$

The BP algorithm is given by:

Step 1:/ Read the learning data

Set network parameters: M, N, J/

Step 2: /Set learning parameters $\eta_1.\eta_2$/

Step 3: /Generate initial weights

Step 4: /adjust all weights using the steepest descent method/

for r = 1 to I do

for q = 1 Q do

Calculation of network outputs (eq II.19 and II.20);

for m =1 to M do

for j = 1 to J do

Update of u_{mj} according to equation (II.35)

for n = 1 to N do

w_{nm} update according to equation (II.36)

II.7 Advantages and inconvenient of neural networks

In the following we will present the advantages and disadvantages of NNs.

II.7.1 Advantages

Implementation of parallelism.

High learning capacity.

Resistance to noise and unreliable data.

High generalization capacity.

Finding solutions to non-linear and complex problems.

II.7.2 Disadvantages

Black boxes: weights are not interpretable .

Lack of initialisation methods .

Problem of parameters setting, like training step.

II.8 Applications of neural networks

Neural networks have been successefully applied in various domain, such as:

Robotics and Sensors.

Modeling of non-linear dynamic processes.

The process control.

Pattern recognition and classification.

Signal processing. Approximation and prediction of functions and others. . . .

II.9 Conclusion

This chapter constitutes an introduction to neural networks. We presented the architecture and training of some common neural models, such as: perceptron , adaline and MLP. We focused on the training using back propagation as it is the most used training algorithm for deep neural networks. We also mentioned some advantages and inconvenient of neural networks .

CHAPTER III

HEARTBEAT CLASSIFICATION USING 1D CNN

III.1 Introduction

In conventional pattern recognition systems, the feature extractor was created by experts. This takes considerable effort. Instead, deep neural networks, like CNN, incorporates the feature extractor into the training process. The aim of this work is to use a 1D-CNN to classify ECG heartbeats. This chapter first introduces general 2D-CNN, then it presents the used 1D-CNN. Finally, it discusses the obtained results over the well-known MIT-BIH arrythmia database.

III.2 Architecture of CNN

CNN is more than just a deep neural network with several hidden layers. It is a deep network that mimics how the brain's visual cortex analyzes and recognizes pictures [11]. CNN differs from the other neural networks in both idea and functioning. The multiclass classification neural network is integrated in the CNN's output layers.

Regardless of the recognition method, utilizing the original data, for example images, directly in the recognition systems produces poor results; the images need to be modified to extract features. Before CNN, the feature extractor was created by experts. As a result, it took a substantial amount of money and effort while not producing good results. Machine Learning is independent of these feature extractors. This process is shown in Figure III.1

[11]. Rather than creating it manually, CNN incorporates the feature extractor into the

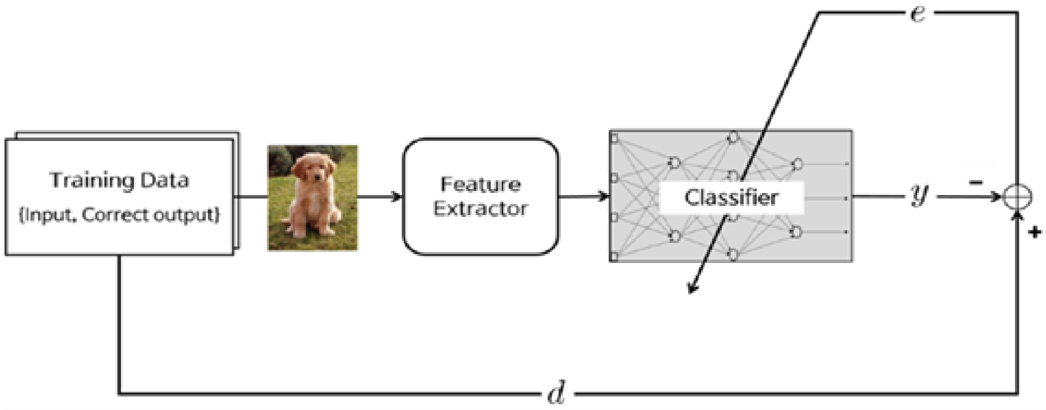

Figure III.1: Feature extractors used to be independent of Machine Learning

training process. CNN's feature extractor is performed using a special type of neural networks, the weights of which are established throughout the training process. CNN's main feature and benefit is that it convertes the manual feature extraction design into an automated procedure. Figure III.2 demonstrates the CNN training idea.

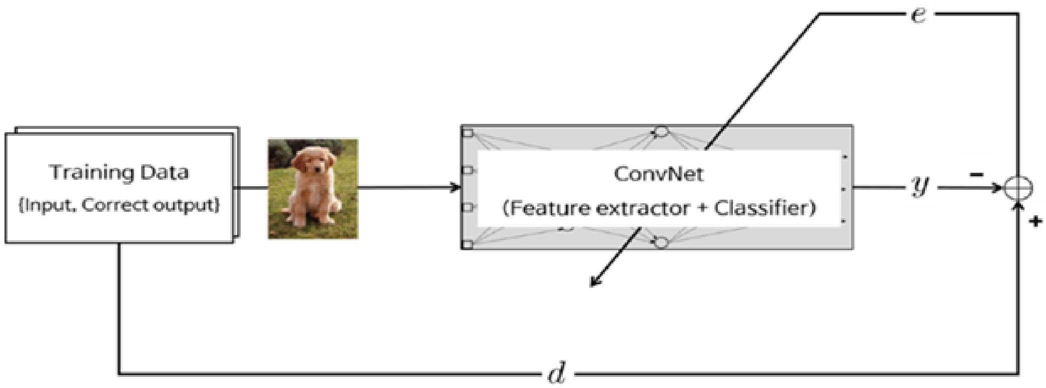

Figure III.2: CNN's feature extractor is composed of special kinds of neural networks

CNN produces superior performance when its feature extraction neural network is deeper (has more layers), at the expense of training challenges. CNN is made up of two neural networks: one that extracts features from the input data and another that classifies the features. CNN's typical architecture is shown in Figure III.3 [12].

The images are the inputs of the feature extraction network, which provide features. These feature signals are sent into the classification neural network. The classification neural network then acts on the image's characteristics to get the result. The feature extraction neural network is made up of stacked convolution layers and pooling layers. The convolution layer, as the name suggests, uses the convolution technique to transform the picture. It can be compared to a set of digital filters. The pooling layer merges neighboring pixels

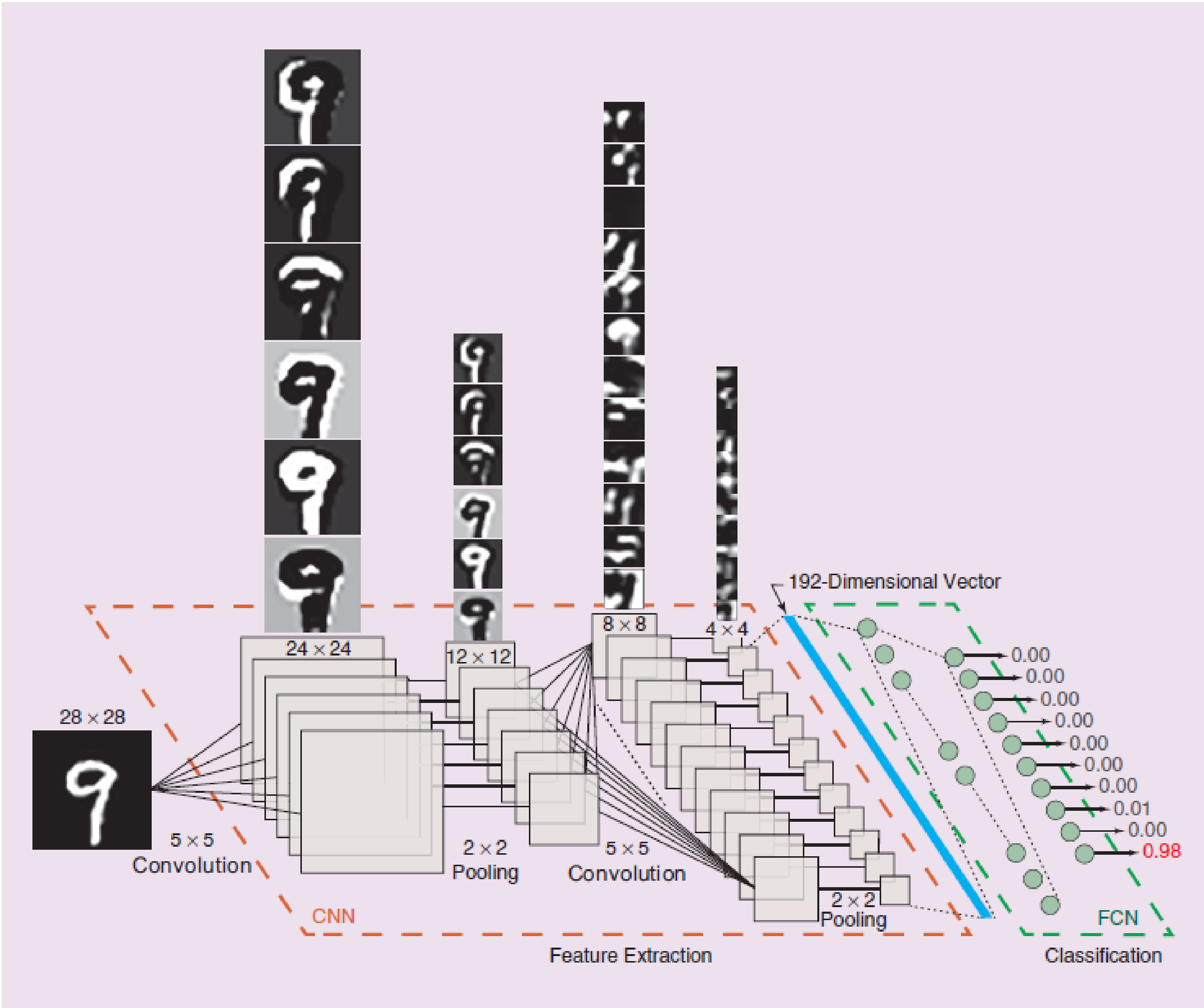

Figure III.3: A CNN trained to extract features that are then used by an FCN to classify handwritten numerals. The input image shown is from the National Institute of Standards and Technology database. (A formatted version of this database is available for experimental work at yann.lecun.com/exdb/mnist.)

into a single pixel. As a result, the pooling layer diminishes the image's dimension. Since the images are the CNN's major concern, the operations of the convolution and pooling layers are conceived in a two-dimensional plane. This is the main distinction between CNN and the other neural networks.

III.2.1 Convolution layer

The convolution layer produces new pictures known as feature maps. The feature map emphasizes the original image's distinguishing qualities. In comparison to the other neural network layers, the convolution layer acts in a fundamentally distinct manner. This layer is not based on connection weights and a weighted sum. Instead, it includes filters that modify images. These filters are referred as convolution filters. The feature map is produced by passing the images through the convolution filters.

Figure III.4 shows the convolution layer process, where the circled * mark represents the convolution operation and the mark φ represents the activation function. The convolution filters are represented by the square grayscale icons that appear between these operators.

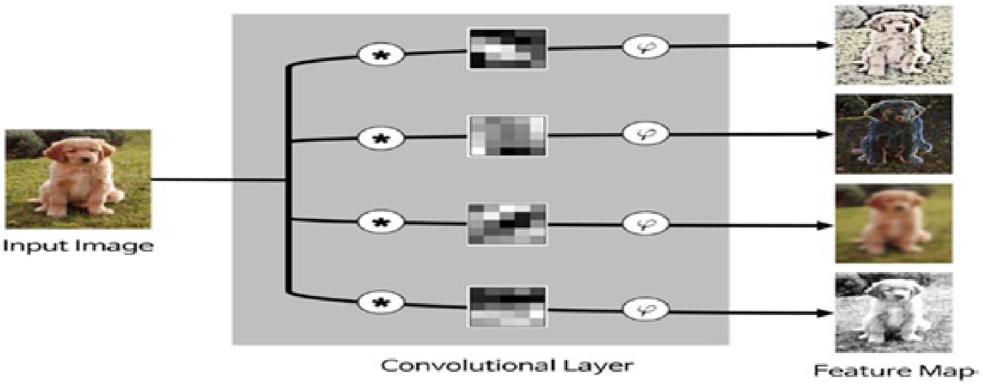

Figure III.4: The convolution layer process

III.2.2 Pooling layer

The pooling layer compresses the image by combining nearby pixels in a specific area of the image into a single representative value. Pooling is a common technique that many other image processing techniques have already used.

To perform the actions in the pooling layer, we must first figure out how to choose the pooling pixels from the picture and how to set the representative value. The surrounding pixels are normally chosen from the square matrix, and the amount of pixels merged varies depending on the task. The representative value is often set as the mean or maximum of the pixels chosen.

III.2.3 Fully connected layer

Fully connected layers in neural networks are those in which all of the inputs from one layer are connected to every activation unit in the next layer. The final few layers of most popular machine learning models are full connected layers that assemble the data retrieved by previous layers to generate the final output. In the classification problems, the number of the neurons in the output layer is equal to the number of classes. Figure III.5 illustrate an example of two fully connected layers. Softmax activation function are usually used in the output layer.

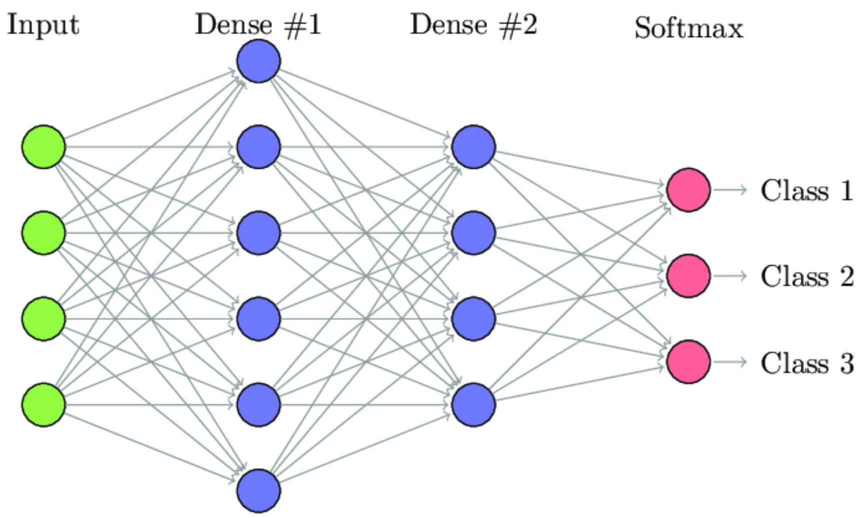

Figure III.5: A Neural Network with Fully Connected Layers

III.3 1D convolutional neural networks

Deep CNNs are often designed to be used only on 2D data such as photos and movies. This is why they are frequently referred to as "2D CNNs." As an alternative, 1D Convolutional Neural Networks (1D CNNs) have recently been constructed as a modified form of 2D CNNs . According to [13], 1D CNNs are beneficial and hence superior to their 2D counterparts in dealing with 1D signals for the following reasons:

- There is a significant difference in terms of computational complexities of 1D and 2D convolutions
- As a general observation, most 1D CNN applications have employed compact (with 1–2 hidden CNN layers) designs, but practically all 2D CNN applications have used "deep" architectures.
- Deep 2D CNN training typically necessitates the use of specialized hardware. For training small 1D CNNs with few hidden layers, however, any CPU implementation on a typical computer is doable and quite quick.
- Compact 1D CNNs are well-suited for real-time and low-cost applications, particularly on mobile or handheld devices, because to their minimal processing needs .

  Compact 1D CNNs have proven higher performance in recent research on applications with little labeled data and large signal fluctuations gathered from various sources (i.e., patient ECG). In 1D CNNs, two separate layer types are suggested, as shown in III.6 , which include 1D convolutions, activation functions, and subsampling (pooling), and 2 fully-connected (dense) layers that are similar to the layers of a standard Multi-layer Perceptron (MLP), and hence referred to as "MLP-layers." The following hyperparameters define the configuration of a 1D-CNN:

1. Number of hidden CNN and MLP layers/neurons (in the sample 1D CNN shown in III.6, there are 3 and 2 hidden CNN and MLP layers, respectively).
2. Filter (kernel) size in each CNN layer (in the sample 1D CNN shown in III.6, filter size is 41 in all hidden CNN layers).
3. Subsampling factor in each CNN layer (in the sample 1D CNN shown in III.6, subsampling factor is 4).
4. The choice of pooling and activation functions.

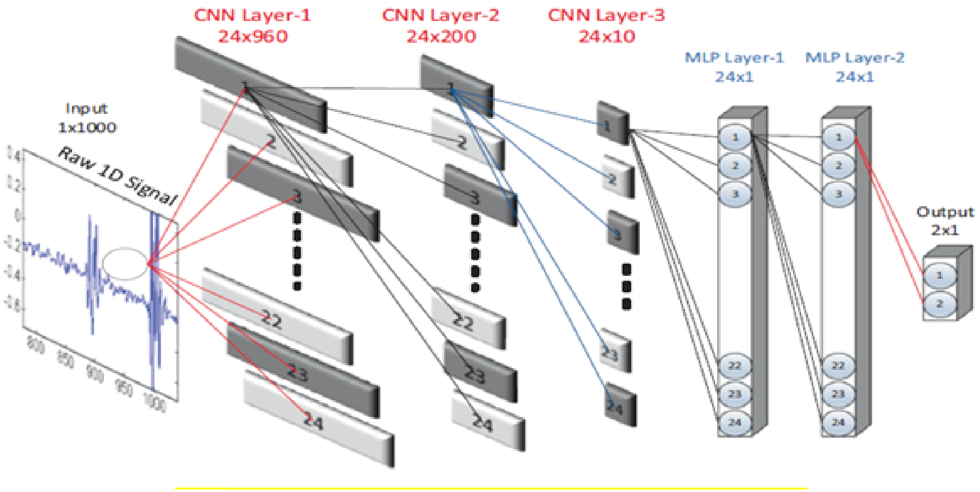

Figure III.6: A sample 1D CNN configuration with 3 CNN and 2 MLP layers

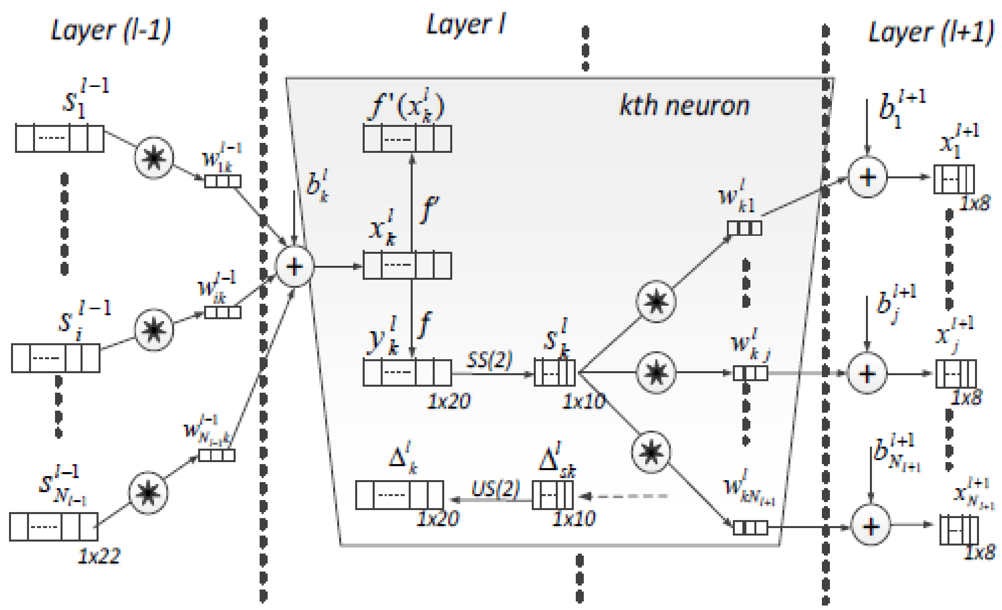

Figure III.7: The adaptive 1D CNN implementation

Figure III.7 illustrate an example of 1d convolution layer [14] . This figure focus on a neuron in the lth layer connected with neurons from the previous layer (l-1) and neurons from the next layer (l+1). In this figure SS and US denote sub-sampling and up-sampling respectively. The weights w_{ij} represent the kernels. The 1D forward propagation (FP) can be expressed as

$$x_k^l = b_k^l + \sum_{i=1}^{N_{l-1}} \text{conv}\,1D\left(w_{ik}^{l-1}, s_i^{l-1}\right) \tag{III.1}$$

Where x_k^l is the input, b_k^l is the bias of the k^{th} neuron at layer l, and s_i^{l-1} is the output of the i^{th} neuron at layer l-1. w_{ik}^{l-1} is the kernel from the the i_{th} neuron at layer l-1 to the k^{th} neuron at layer l. With such a design it is aimed that the number of hidden CNN layers can be set to any number.

III.4 MIT-HIT Database

III.4.1 Presentation

The MIT-BIH arrhythmia database [15] includes 48 half-hour extracts of two-channel ECG recordings from 47 patients studied by the BIH arrhythmia laboratory between 1975 and 1979. Twenty-three recordings were chosen at random from 4,000 outpatient ECG records per hour obtained from a diverse group of inpatients (roughly 60%) and outpatients (about 40%) at Boston's Beth Israel Hospital. The remaining 25 recordings in the same collection were chosen to represent less common but clinically important arrhythmias that are often underrepresented in small samples.

The recordings were made at 360 samples per second and per channel, with a resolution of 11 bits over a 10 mV range. Each record was annotated separately by two or more cardiologists; discrepancies were handled in order to provide reference, computer-readable data [16]. In this work we considered only six patients and four classes, namely: normal, left branch block, right branch block, Premature ventricular contractions. These classes are the most present in this database.

III.4.2 Examples of beats from different classes

Figure III.8 illustrates some example of beats from the MIT-BIH arrhythmia database. We present the four classes considered in this study, namely: normal, left branch block, right branch block, Premature ventricular contractions. We note that beats from the same classes may be different for each patient.

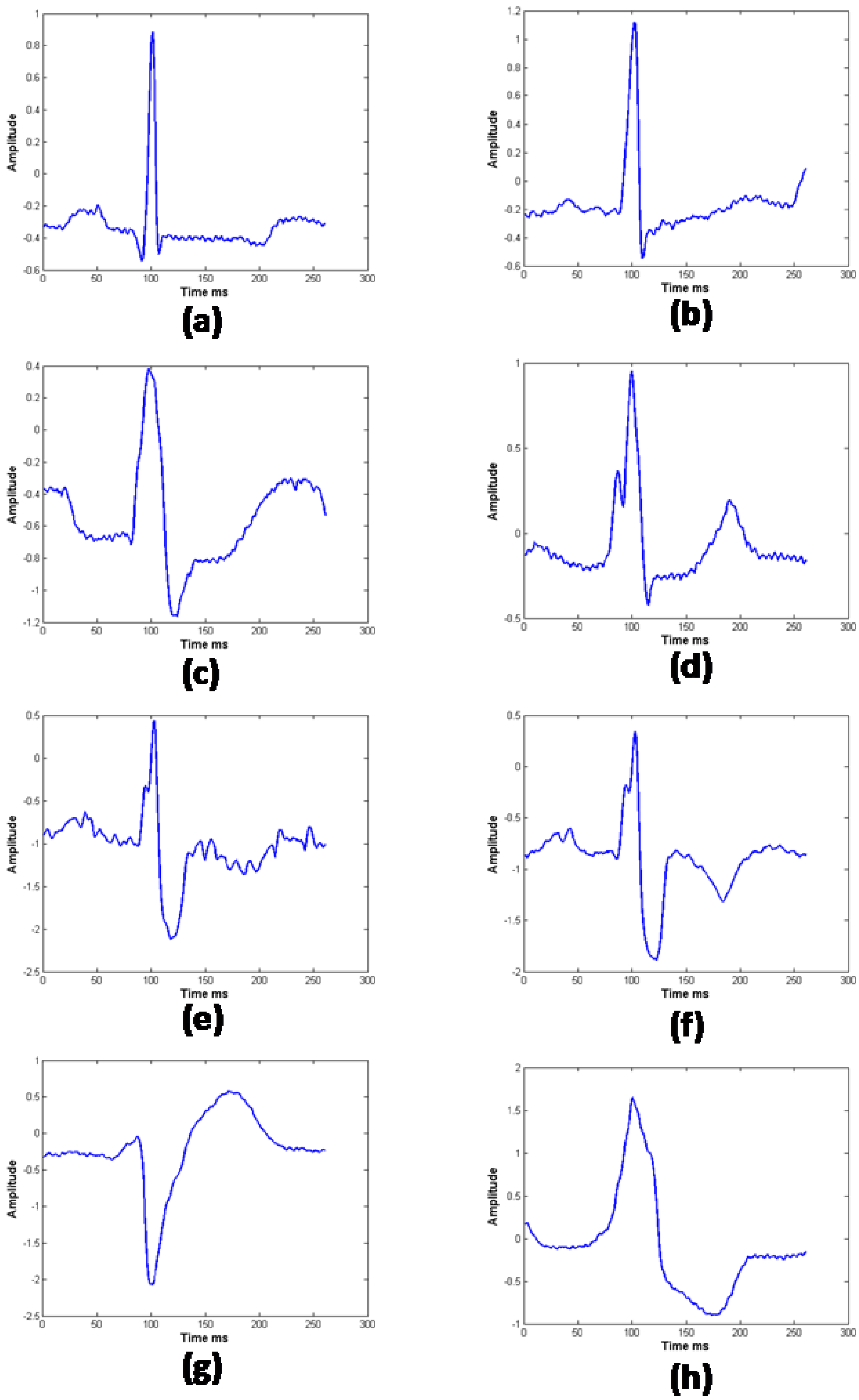

Figure III.8: Some example of beats from different patients
a,b : Normal beats from patient 100 and patient 200
c,d : Left branch block beats from patient 109 and patient 111
e,f : Right branch block beats from patient 118
g,h : Premature ventricular contractions from patient 200 and patient 208

III.5 Results

The obtained results are represented in table III.I. To evaluate the generalisation performance of the proposed classifier, we stratified three partition training/test data: 90/10, 70/30 and 50/50. Obviously, 50/50 is the most challenging as the test data are equal to the training data. For each partition, we adopted tow architectures: 3-3-45-4-4-4 and 5-5-75-8-8-4. The numbers in these architectures are described as follows:

- The first number: the number of neurons in the first convolution layer.
- The second number: the number of neurons in the second convolution layer.
- The third number: the number of features obtained after convolution.
- The fourth number: the number of neurons in the first hidden layer of the MLP.
- The fifth number: the number of neurons in the second hidden layer of the MLP.
- The sixth number: the number of neurons in the output layer and it represents the number of classes.

In each architecture, we used two sizes of the filters in the first and second convolution layers: 3-3 and 5-5.

From the obtained results, we note the following points:The partition 90/10 is the simplest, the proposed model provided almost 100% on the test data. 70/30 and 50/50 are more difficult. The proposed model provided good results with large architecture. We note also that better results were obtained with the filters with size 5. In general, the results are promising in all architectures.

Parameters			Results		
Ratio Training / test data (%)	Architecture: Number of neurons in each layer	Size of filters in CNN layers	Training error	Training rate	Test rate
90 / 10	3-3-45-4-4-4	3-3	2.1370e-005	100	100
	3-3-45-4-4-4	5-5	0.0011	99.9214	99.2908
	5-5-75-8-8-4	3-3	1.7821e-005	100	100
	5-5-75-8-8-4	5-5	4.0663e-006	100	100
70 / 30	3-3-45-4-4-4	3-3	0.0144	99.1919	99.1919
	3-3-45-4-4-4	5-5	0.0043	99.7980	99.2925
	5-5-75-8-8-4	3-3	3.5503e-005	100	100
	5-5-75-8-8-4	5-5	1.2990e-005	100	99.7642
50 / 50	3-3-45-4-4-4	3-3	0.0038	99.8586	99.2928
	3-3-45-4-4-4	5-5	5.7654e-005	100	99.2928
	5-5-75-8-8-4	3-3	2.7875e-005	100	99.5757
	5-5-75-8-8-4	5-5	2.8029e-004	100	99.7171

Table III.1: The obtained results over MIT-BIH database.

Figure III.9 illustrates an example of the evolution of the error (mean sum squared) during the training process. In this case, the system converge after about 25 iterations.

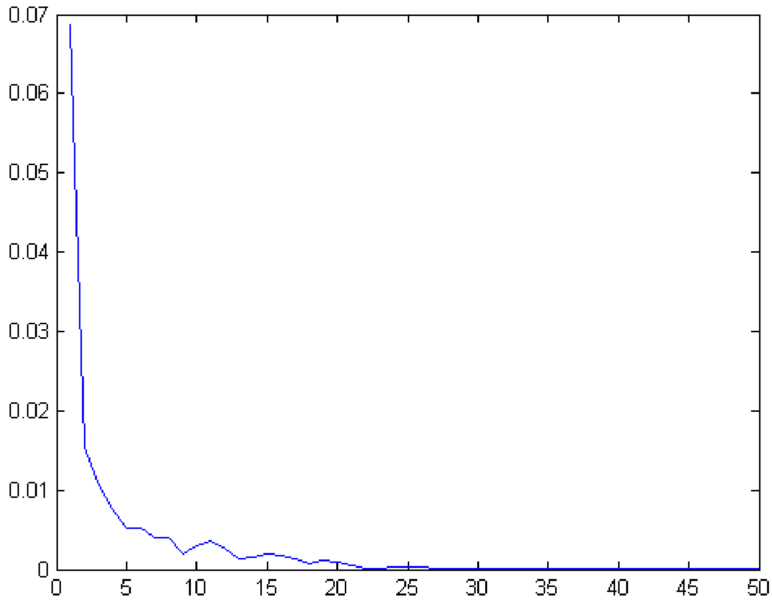

Figure III.9: Example Mean Sum Squared Error during the training

III.6 Conclusion

In this chapter, we first presented the CNN in general and then we presented the used 1D-CNN. To evaluate the proposed model, we performed tests on MIT-BIH database. We performed tests with different architectures and sizes of the filters. In general, the results are promising in all architectures. The classification using the 1D-CNN has the following advantages:

- It can be directly applied to raw ECG without pre-treatment
- It avoids the problem of time-consuming feature extraction and selection phases.

However, CNN has the following disadvantages:

- It need a lot of parameters to be set: the number of layers , the number of neurons, the number and the size of filters,

GENERAL CONCLUSION

In this work, we addressed the problem of classifying ECG heartbeats. In conventional pattern recognition systems, the feature extraction and classification were performed in separate stages. This takes considerable effort and needs expert intervention. Deep neural networks, like CNN, constitute an important alternative as they incorporate the feature extractor into the training process. In this work, we used a 1D-CNN to classify ECG heartbeats. To evaluate the proposed model, we performed tests on MIT-BIH database. We performed tests with different architectures and sizes of the filters. In general, the results were promising with all architectures. We noted that the classification using the 1D-CNN has the following advantages:

- It can be directly applied to raw ECG without pre-treatment.
- It avoids the problem of time-consuming feature extraction and selection phases.

However, CNN needs a lot of parameters to be set. This includes:

- Number of CNN layers and neurons.
- The number of fully connected layers and neurons.
- The number and size of filters (kernels).
- The subsampling factors.
- The type pooling and activation functions.

REFERENCES

[1] K. Koutroumbas and S. Theodoridis, *Pattern recognition.* Academic Press, 2008.

[2] J. Hampton and J. Hampton, *The ECG made easy e-book.* Elsevier Health Sciences, 2019.

[3] R. Benali, "Analyse du signal ecg par réseau adaptif d'ondelettes en vue de la reconnaissance de pathologies cardiaques," *Doctorat en Sciences, Université Abou Bekr Belkaid*, 2013.

[4] https://www.kandh.com.tw/kl-730-biomedical-measurement-training-system-kl-730.html.

[5] E. J. d. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "Ecg-based heartbeat classification for arrhythmia detection: A survey," *Computer methods and programs in biomedicine*, vol. 127, pp. 144–164, 2016.

[6] K.-L. Du and M. N. Swamy, *Neural networks and statistical learning.* Springer Science & Business Media, 2013.

[7] S. Haykin, *Neural networks and learning machines, 3/E.* Pearson Education India, 2009.

[8] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Icml*, 2010.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning.* MIT press, 2016.

[10] C. G. Looney, *Pattern recognition using neural networks: theory and algorithms for engineers and scientists.* Oxford University Press, Inc., 1997.

[11] P. Kim, *Matlab deep learning: With machine learning, neural networks and artificial intelligence*, 2017.

[12] R. C. Gonzalez, "Deep convolutional neural networks [lecture notes]," *IEEE Signal Processing Magazine*, vol. 35, no. 6, pp. 79–87, 2018.

[13] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1d convolutional neural networks and applications: A survey," *Mechanical systems and signal processing*, vol. 151, p. 107398, 2021.

[14] S. Kiranyaz, T. Ince, R. Hamila, and M. Gabbouj, "Convolutional neural networks for patient-specific ecg classification," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 2608–2611.

[15] https://physionet.org/physiobank/database/mitdb/.

[16] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.