UNIVERSITÀ POLITECNICA DELLE MARCHE

FACOLTÀ DI INGEGNERIA

_____

Master's Degree in Biomedical Engineering

ELECTROCARDIOGRAM CLASSIFICATION BY COMBINING WAVELET TRANSFORMS AND CONVOLUTIONAL NEURAL NETWORKS

Supervisor

Prof. Laura Burattini

Author

Enrico Emaldi

Assistant Supervisor

Dr. Agnese Sbrollini

A.Y. 2020/2021

Abstract

The electrocardiogram (ECG) is a powerful tool for diagnosis and screening of many heart-related disorders which are of primary concern and thus need to be correctly diagnosed as early as possible. Its characteristic signal holds abundant information, primarily regarding electrical functioning but also related to functional and structural elements of the heart, some of which are evident just by visual inspection, but others are hidden, even to an expert eye. In this work, the wavelet transform was used to access the frequency domain and highlight potentially crucial features. It is well suited for non-stationary signals with dynamic components like the ECG while keeping at the same time a good temporal resolution, differently from the more common Fourier transform for example. Once obtained the transformed ECG representations, called scalograms, an artificial intelligence algorithm was developed for the automated classification of pathological signals. In particular, a convolutional neural network (CNN), well suited for images like scalograms, used supervised deep learning to classify data without explicitly selecting features. The data were retrieved from the publicly available datasets provided by PhysioNet, containing almost one-hundred-thousand labelled real-world ECG, collected through four different contents. Due to the extremely unbalanced nature of the classes representing a realistic distribution of heart-related disorders, the model was trained only on a limited selection of four of them, while the multi-label problem of having more than one disorder associated with some of the signals was simplified to be only multi-class. The end-to-end architecture was built to input one scalogram for each lead as an image channel in a hyperspectral fashion; doing so, it was also possible to design four more models with a reduced number of leads, respectively with six, four, three, and two leads. These models were cross-validated and achieved similar performances in terms of accuracy (86% average), precision (83% average), and sensitivity (82% average) across the different classes and models. These results show the validity of the approach for ECG classification, in addition to a certain redundancy in the different lead's information with the five models almost superimposable. Overall, wavelet transform provides a useful representation of the ECG, which can be combined with CNN to effectively diagnose different disorders, even when a reduced number of leads is used.

CONTENTS

INTRODUCTION

The electrocardiogram (ECG) is often the first clinical report that comes to mind when speaking about biomedical data. The reason why is easy to explain from a medical point of view. Cardiovascular diseases (CVD) are listed by the World Health Organization (WHO) as the main cause of death worldwide, primarily in the form of heart diseases or strokes [1]. It follows that a versatile tool such as the ECG, which can be used to diagnose multiple cardiomyopathies or their precursors, is crucial for medicine and research to invest in.

From a clinical standpoint, the ECG offers a lot of benefits too. It requires relatively cheap instrumentation, which is also easy to use and maintain. It can be performed under many different circumstances to monitor the health status of the patient, from intensive care units and ambulances to physical training and exercise. It comes also in the form of wearable sensors, ranging from Holter recorders to smartwatches. But most importantly, the ECG signal can be interpreted just by visual inspection.

The fact that a physician can read an ECG at first glance should not be underestimated, but at the same, it draws some intrinsic limitations. First, one recording can last for few minutes, but also multiple hours, up to twenty-four or forty-eight hours, and sometimes is recorded in continuous, like in intensive care units. In addition, the standard 12-lead ECG signal is made of 12 time-series. If a physician wanted to analyze it just by visual inspection, not it would be extremely time-consuming but also prone to error due to the expertise of the examiner. It is also possible that some features are hidden in the tracing, not visible in the time domain. On the other hand, the information contained in a 12-lead ECG may be redundant; there could be a minimal representation to describe concisely yet exhaustively the heart state.

In this context, the work of bioengineers plays a key role: using an interdisciplinary approach, merging tools from electrophysiology, signal processing, and computer science, perhaps it is possible to elaborate a solution. This thesis will present an innovative method to automatically classify the ECG signal, consisting of a convolutional neural network that inputs an adjustable number of leads using their representation as scalograms obtain using the wavelet transform.

1. BIOMEDICAL OVERVIEW OF THE HEART

A comprehensive review of the anatomical and physiological characteristics of the heart is beyond the scope of this thesis, but in order to properly understand some peculiar attributes of the following work, some unavoidable elements must be addressed first. In particular, the essential anatomical structure of the heart and the proper names of its component, the physiological elements which characterize the cardiac cycle, and the most significant and common cardiomyopathies and disorders.

1.1. ANATOMY OF THE HEART

The heart is the central part of the circulatory system: with its rhythmic contractions, it pumps the blood inside the blood vessels.

The heart is a hollow muscular organ, sited in the middle of the mediastinum in the thoracic cavity, with a slight offset to the left side, fully surrounded by a double-membraned sac called the pericardium.

It is shaped like a cone, slightly flattened from front to back, with its base positioned upwards and tapering down to the apex. Its average length is 12 cm, width is 10,5 cm, thickness is 5,5 cm, and it weighs about 250 g [2].

The heart presents four internal cavities: two superior chambers, which are the right and left atria, and two inferior chambers, which are the right and left ventricles (Figure 1). The conjunction of the right atrium and right ventricle is often referred to as the right heart, while the conjunction of the left atrium and left ventricle represent the left heart. Each atrium opens into its relative ventricle via the corresponding atrioventricular (AV) valve.

The two "hearts" do not communicate with the down cavities on the left in a healthy adult because separated by the mainly muscular but also fibrous cardiac septum. The latter can be divided into the interatrial septum which separates the two atria, the AV septum which separates the right atrium and the right ventricle, and the interventricular septum which separates the two ventricles. During the foetal period, the right atrium and the left one can communicate via the foramen ovale, an orifice in

the interatrial septum. After birth, in some cases later during infancy, the foramen closes spontaneously, and the interatrial septum definitively separates the two atria and their respective circulations [2].

The cardiac skeleton consists of dense connective tissue, forming the AV septum on which the four heart valves are attached. In addition to its structural function, it also provides an important boundary for the heart's electrical conduction system thanks to its insulating property.

The tricuspid valve is located between the right atrium and the right ventricle. It has three cusps that link to chordae tendineae and three papillary muscles that are termed by their relative positions: anterior, posterior, and septal. Between the left atrium and the left ventricle is the mitral valve. Because it has two cusps, an anterior and a posterior cusp, it is also known as the bicuspid valve. These cusps are also connected to two papillary muscles that overhang from the ventricular wall through chordae tendineae [2].
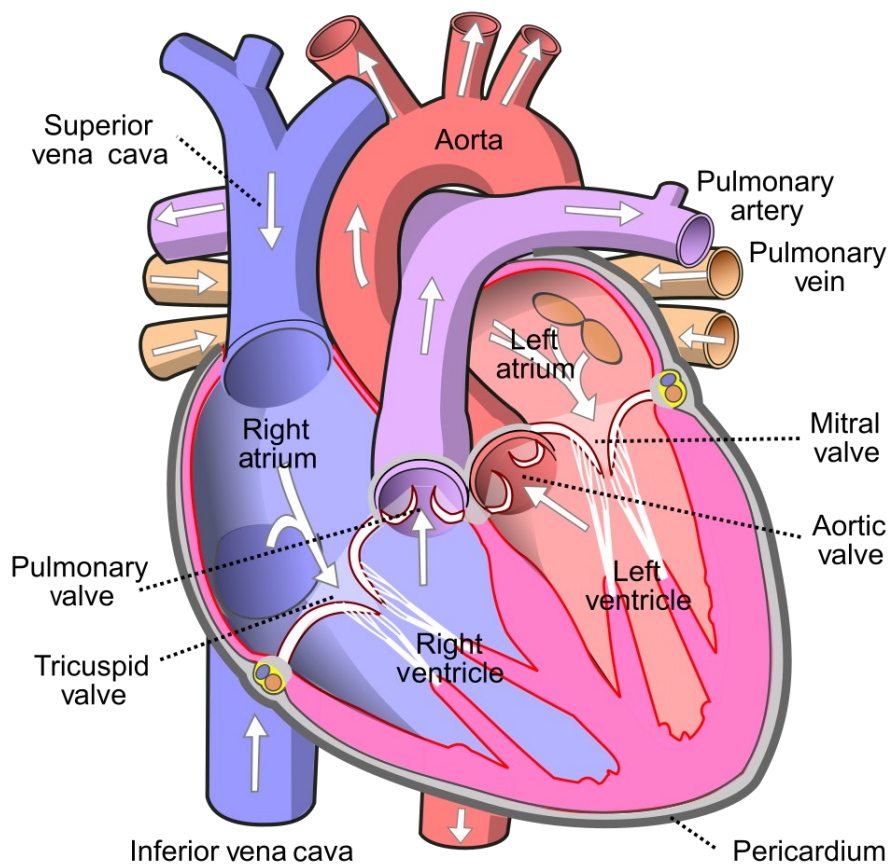

FIGURE 1 SCHEMATICAL REPRESENTATION OF THE HEART'S ANATOMY

Chordae tendineae are cartilaginous connections that connect the papillary muscles to the heart's walls and valves. When the valves close, these muscles keep them from falling too far back. The papillary muscles relax during the relaxation phase of the cardiac cycle, whereas the chordae tendineae are slightly tense. The papillary muscles contract in tandem with the heart chambers. This puts pressure on the chordae tendineae, which helps to keep the AV valves' cusps in place and keep them from blowing back into the atria [2].

Each ventricle has two extra semilunar valves at the outflow. At the base of the pulmonary artery is the pulmonary valve. There are three cusps on this, none of which are linked to any papillary muscles. When the ventricle relaxes, blood from the artery flows back into the ventricle, filling the pocket-like valve and pressing on the cusps, which close to seal the valve. The semilunar aortic valve is located near the bottom of the aorta and is not connected to the papillary muscles. This also contains three cusps that close in response to the pressure of blood returning from the aorta [2].

The heart wall is made up of three layers. The endocardium is the innermost one. It is a basic squamous epithelial lining that covers the heart chambers and valves. It is connected to the myocardium by a thin layer of connective tissue that is continuous with the endothelium of the heart's veins and arteries. The myocardium, or cardiac muscle, is the middle layer and consists of an involuntary striated muscle tissue surrounded by a collagen framework; it presents a complex swirling pattern of muscle cells that maximize its pumping function. The outermost layer is called the epicardium, which can be considered the myocardium's serosa, and constitutes the inner interface with the pericardium [2].

The heart tissues are supplied with oxygen and nutrients by the coronary circulation, which includes arteries, veins, and lymphatic vessels.

The vagus nerve and nerves from the sympathetic trunk send nerve impulses to the heart. The heart rate and the force of cardiac contraction are influenced by these nerves, but not controlled by them. The medulla oblongata contains two paired cardiovascular centres that send signals along these neurons. The parasympathetic nervous system's vagus nerve lowers heart rate, whereas the sympathetic nervous system's sympathetic trunk nerves raise it. The cardiac plexus is a network of nerves that runs along the top of the heart [2].

1.2. PHYSIOLOGY OF THE CARDIAC CYCLE

Human blood circulation is defined as double and complete, that is, the two halves of the heart function autonomously, like a double heart, and the mixing of venous and arterial blood (commixtio sanguinis) never takes place.

The heart muscle works like a pushing and aspiring pump: it "sucks" into its cavities the venous blood and then sends it to the lungs through the pulmonary artery in the system of the small circulation. From the lungs, the pulmonary veins bring oxygen-laden blood back to the atria and then to the ventricles which, through the aorta, distribute it throughout the body, up to the capillary networks in the system of great circulation [3].

The heart, contracting, pushes blood into the arteries; in fact, all blood vessels in which blood flows with a centrifugal direction with respect to the heart are called arteries. The arteries branch successively, giving rise to branches of gradually smaller calibre and with an increasingly thin wall; finally, they end by resolving themselves in very slender blood vessels, called capillary vessels. From these originate the veins, which flow together into veins of even greater calibre and finally put their head to the heart; in fact, all blood vessels in which the current of the blood has a centripetal direction with respect to the heart are called veins [3].

The coronary circulation must be considered separately: from the epicardial arteries originates an extensive network of intramural arteries, arterioles, and capillaries. Given the high oxygen demand of the myocardium, the capillary network amounts to about 15% of the total heart mass and this facilitates the exchange of nutrients to the myocytes and from them the emission of waste products, which are collected in a network of intramural venules and transported by them in large epicardial veins of capacitance. Most of the venous blood of the left ventricle drains into the coronary sinus and from there into the right atrium. The rest of the drainage takes place through the Thebesian veins and the anterior cardiac vein, which drain into the right heart chambers [4].

A heart rate measured in beats per minute (bpm) between 60 and 100 beats is considered normal; a frequency of less than 60 bpm is called bradycardia, while one above 100 bpm is called tachycardia. In the newborn, the frequency reaches 120 bpm, while in the fetus it is even higher [3].

During sleep, the heart pumps 5 liters of blood in a minute, while during moderate physical activity the amount doubles. For heavy activity or vigorous athletic activity, up to 20 liters per minute are reached. At rest, blood pressure is considered normal if it does not exceed the value of 130/80 mmHg, although lower values are equally considered in the norm and in general, these values depend on a

wide range of parameters such as sex, age or build and it is therefore difficult to establish universally effective ranges [3].

The cardiac cycle includes the two essential phases in which the activity of the heart takes place: diastole and systole (Figure 2). During diastole, the whole heart is relaxed, allowing blood to flow into the four cavities. It flows from the vena cava into the right atrium and from the pulmonary veins in the left atrium. The heart valves are simultaneously open and allow the passage of blood from the atria to the ventricles. Diastole lasts about 0,4 s, enough time to allow the ventricles to fill up almost completely [3].

The systole begins with a contraction of the atria, lasting about 0,1 s, which determines the complete filling of the ventricles. Then the ventricles contract for about 0.3 s; their contraction closes the AV valves and opens the semilunar valves so that oxygen-poor blood can be pushed towards the lungs, while oxygen-rich blood heads towards the whole body, through the aorta [3]. During the contraction, the myocardium acts as a unique giant fiber (syncytium). This means that the heart can be interpreted with the law "all-or-nothing", in its entirety and not for a single cell as it would happen in skeletal muscles. The amount of blood expelled from the heart into the arterial circulation with each beat is called systolic (or cardiac) output.
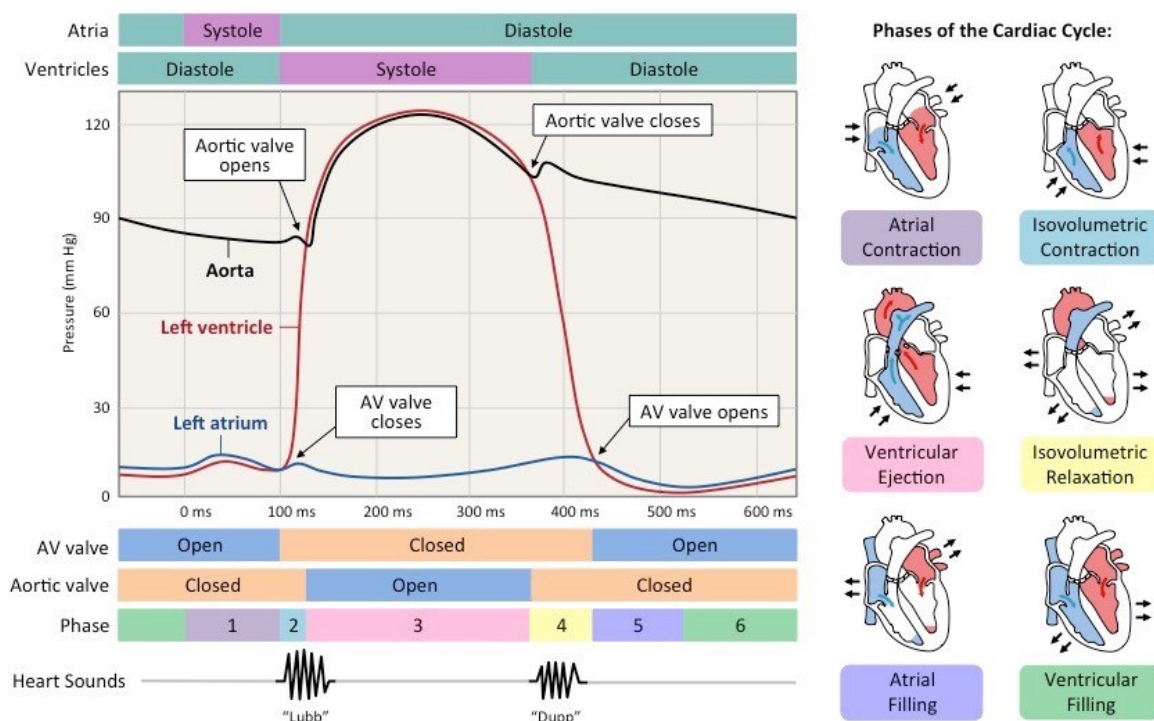

FIGURE 2 THE CARDIAC CYCLE FEATURING (TOP-DOWN): ON-SET OF THE ATRIA AND OF THE VENTRICLES; PRESSURE WAVES OF THE AORTA (BLACK), LEFT VENTRICLE (RED), LEFT ATRIUM (BLUE); ON-SET OF THE AV VALVE AND OF THE AORTIC VALVE; CYCLE PHASES (COLOR-CODED WITH THE DRAWINGS ON THE RIGHT); HEART SOUNDS

1.3. ELECTROPHYSIOLOGY

The heart is composed of two kinds of cells that allow cardiac contraction: the mechanical ones (myocytes) and those belonging to the conduction system of the heart (pacemaker).

The plasmatic membrane of myocytes, which divides the intracellular and extracellular spaces, is semipermeable to ions and due to their difference in concentration allows the existence of a voltage difference of -90 mV, with the negative pole inside the cell, called resting potential. The potassium ion K^+ is more abundant on the inside of the cell and tends to spread outside following the diffusion gradient, while the opposite holds for the sodium ion Na^+ which tends to spread out toward the inside of the cell. Dynamic equilibrium is reached when the force due to the diffusion gradient becomes equal to one due to the electrochemical gradient, limiting any further diffusion of ions.

The voltage difference of these cells from the resting potential can vary if a current travels through the membrane; if such stimulating current is enough to reach the threshold of -65 mV the action potential takes place. It has a duration of 1-2 ms and an amplitude of 105-120 mV, generating a positive spike of around 20 mV, which is higher than those of skeletal muscle cells because it must maximize the efficiency of the cardiac pump.

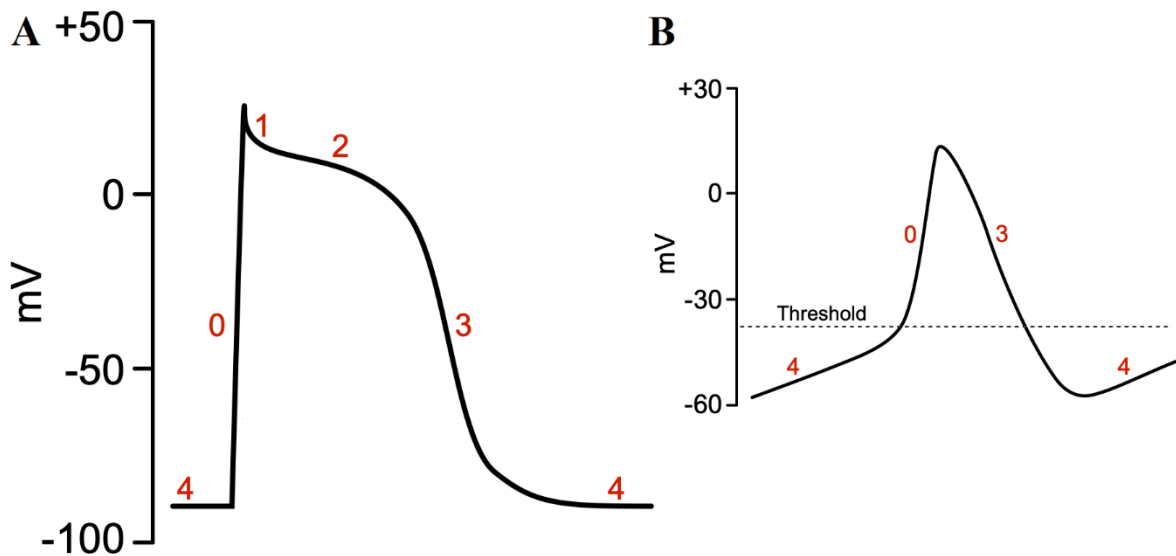The action potential of a myocyte follows these steps (Figure 3 , A):

FIGURE 3 (A) ACTION POTENTIAL OF A CARDIAC MYOCYTE; (B) ACTION POTENTIAL OF A PACEMAKER CELL

- <u>Phase 0</u>: Rapid depolarization. When a cell is reached by an electrical impulse of such entity to reach the threshold, a sudden opening, of the Na^+ channels, takes place with a quick influx of said ion inside the cell. It follows an inversion of electrical polarity: the inside becomes positive, deactivating the Na^+ channels
- <u>Phase 1</u>: Early repolarization. The K^+ channels open and the ion starts exiting from the cell, with a consequent initial repolarization
- <u>Phase 2</u>: Plateau. During this phase, the outflux, of K^+ ions, continues, accompanied by the simultaneous opening of calcium ions Ca^{++} channels and its related influx. With both of them being two positive ions, the net charge exchange does not change
- <u>Phase 3</u>: Actual repolarization. While the outflux of K^+ continues, the Ca^{++} channels close. This allows the complete repolarization of the cell. Nevertheless, the ionic conditions are inverted: the inside of the cell contains a lot of Na^+ entered during phase 0, while K^+ is concentrated outside.
- <u>Phase 4</u>: Resting potential. The Na^+-K^+ pump, an active exchange protein channel, uses adenosine triphosphate (ATP) to expel the Na^+ ions in exchange for K^+ ions. At the end of this phase, the cell will be able to be stimulated once more by an electrical impulse and to give rise to a new action potential.

If an electrical stimulus reaches the myocytes during the action potential, it will not arise a new one. The cell is in its absolute refractory period from the beginning of phase 0 to the first half of phase 3 and cannot be excited. The cell is in its relative refractory period after the second half of phase 3 and can be excited precociously if the stimulus is more intense than the critical threshold.

Pacemaker cells exhibit different behavior (Figure 3 , B). They are also called calcium-dependent because they have not got Na^+ channels, their depolarization, and successive repolarization depend exclusively upon Ca^{++} influx and outflux. Pacemaker cells present an intrinsic automatism and spontaneous depolarization, while myocytes are sodium-dependent and must be stimulated by electrical stimuli. Their phase 4 is unstable, thus do not present a well-defined resting potential. They are located in the sinus node, in the AV node, in the bundle branches, and in the Purkinje fibres, with the largest spontaneous rate in the sinus node dictating the rhythm of the heart. These cells undergo the effect of the autonomic control system which regulates the pacemaking rate: sympathetic influences increase the rate of depolarization and shift the depolarization threshold downward, while vagal influences have the opposite effect.

The stimulus generated in the pacemaker cells of the sinus node reaches almost simultaneously the atrial muscle fibres; the electrical signal then propagates to the AV node in about 20 ms. When the

signal is transmitted from the atria to the ventricles, it is possible to highlight a delay in conduction, essential to obtain a sufficient filling of the ventricles, thus maximizing the efficiency of the cardiac pump. The AV node introduces a 90 ms delay before the impulse can reach the His fascicle, followed by another delay of 40 ms due to the transit through the fibrous part of the AV septum, characterized by slow conduction velocity. Then, the impulse propagates through the Purkinje fibers which from the AV node reach the ventricles. These fibers are characterized by a fast conduction velocity in their proximal extensions which allows almost instantaneous conduction to the ventricles; this velocity decreases toward their distal extremities, adding up more delay. Overall, the time required by the ventricles to properly contract is about 60 ms.

2. THE ELECTROCARDIOGRAPHIC SIGNAL

Given the previous anatomical and physiological premise, it follows an overview of the heart electrical conduction system and its behavior which gives rise to the ECG. Then, the main characteristics of this peculiar signal are also explored. The theoretical support was provided, if not stated otherwise, by the following books [5][6]. At the end of this chapter, some prevalent heart-related disorders are presented too.

2.1. ELECTROCARDIOGRAM

The human body's surface develops small electric currents at the nervous and muscular levels. This phenomenon coincides with the ionic variations inside and outside of myocytes relative to the opposite directions of Na^+ and K^+ fluxes. These small ionic currents are extremely difficult to measure at a cellular level but can be evaluated via the integration of biopotentials, in the heart's case using an electro-cardiographer: the change in electrical potential produce the deflection of a writing system, thus allowing the conversion of ionic currents in electric currents at the level of the electrodes and the electrical energy in mechanical energy. Modern systems do not present any longer a mechanical apparatus, using electronics to digitalize the information. The electrical energy however must be properly amplified in order to obtain a readable signal. The recording of an electro-cardiographer is the ECG [7].

During each and every beat, waves of depolarization follow each other in the manner that was discussed previously. This very model of ordered depolarizations gives rise to the ECG. The deflections recorded at the body's surface level represent the sum of all the depolarizations and repolarizations of the entirety of the myocardial cells active at the same time.

The timing of depolarization and successive repolarization of the cardiac cells give rise to the ECG characteristic shape, which by the way is completely different from what could be measured at the cellular level [7]. The shape of ECG waves is determined by the electrical state present at the electrodes positioned in different regions of the body. Four positive waves (P, R, T) and two negatives

(Q, S) constitute the ECG. Their sign is expressed with respect to the isoelectric line, which is the ECG baseline, considering a positive shift upward, negative one downward. The isoelectric is essential for the interpretation of the ECG, especially of the ST segment, which is considered normal only if it does not differ from it of 0,1 mV, either upward or downward.

Here it follows a more in-depth yet practical presentation of the ECG waves and segments (Figure 4) [8]:

- <u>P wave</u>: correspond to the depolarization of the atria and originates in the sinus node. When the electric impulse leaves the sinus node produces the depolarization of the myocytes nearby which contracts and then propagates in a radial way. Its duration ranges from 60 to 120 ms, its amplitude is less or equal to 0.25 mV.

- <u>PR segment</u> (sometimes, PQ segment): A-V conduction time. Normal measurements belong to the interval between 120 ms up to 200 ms; if the PR segment is too short (<120 ms), high-frequency arrhythmias can happen; on the other hand, if the PR segment is too long (>200 ms), it can be due A-V block, leading too low heart rates.

- <u>QRS complex</u>: constituted of three waves that follow each other, it corresponds to the ventricles' depolarization. The Q wave is normally negative, of small entity, and corresponds to the depolarization of the interventricular septum; the R wave presents a big, positive spike, corresponding to the depolarization of the apical part of the ventricles; the S wave is a negative wave, normally of small entity, corresponding to the depolarization of the basal and posterior regions of the left ventricle. The QRS complex is considered normal if its duration ranges from 60 to 90 ms. This interval is comprised also of the repolarization of the atria which is not visible because completely masked by the more prominent ventricular depolarization.

- <u>ST segment</u>: represents the interval during which all the ventricular cells are depolarized and there are no significant electrical effects until the beginning of the repolarization (usually the ST segment is isoelectric).

- <u>T wave</u>: is the first repolarization wave of the ventricles. It is not always identifiable because it can be of very small amplitude.

- <u>QT segment</u>: represents the electric systole, meaning the summation of ventricular depolarization and repolarization. Its duration varies depending on heart rate, usually between 350 and 440 ms.

- <u>U wave</u>: worth mentioning even if rarely visible in an ECG because of its small amplitude; it is normally due to papillary muscles repolarization.
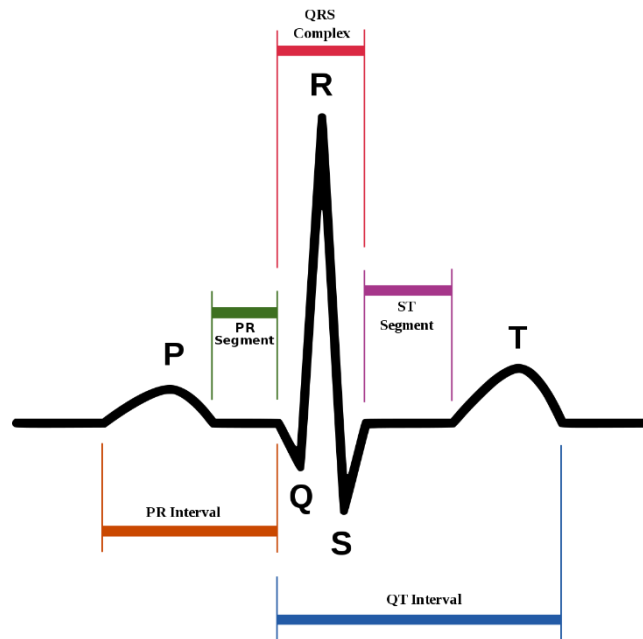
- FIGURE 4 ECG SIGNAL WITH ANNOTATED WAVES AND SEGMENTS

2.2. LEAD SYSTEMS AND HEART VECTOR

From an electrical standpoint, the heart can be seen as an electric dipole, meaning a system of two charges of opposite signs placed close to the other. One myocardial fibre in resting condition, isolated and measured placing two electrodes near its extremities does not produce any voltage, because both extremities are at their resting potential (positive outside of the cell, negative inside), and there is not any current flow.

When the fibre is excited, the action potential starts propagating and the polarity of the fibre is suddenly inverted (negative outside of the cell, positive inside) due to the opening of the ionic channels on the cell membrane. One of the two electrodes is invested first by the action potential and will be negative, while the other, being still at rest, will be positive. In this case, the dipole is said to be moving [7].

Until now only a single cardiac fibre was considered, but, during the propagation of the action potential, each fibre acts as a dipole and thus its electrical activity can be represented using an electric vector (or dipole vector) in three dimensions, characterized by direction and magnitude.

The ECG does not record the electrical activity for single fibre, but the activity of the myocardium as a whole. Nevertheless, the cardiac mass acts as a syncytium, which is the equivalent of a single giant

fibre, both in a mechanical and electrical sense; then it is possible to consider the whole myocardium as a single dipole, characterized by one vector that is the heart vector. At every moment numerous fibers are active simultaneously, each one with its own electrical activity and dipole vector, making the heart vector the summation of all these contributions and thus representing the electrical activity of the heart as a whole.

For the recording of the ECG, it is necessary to apply at least two electrodes on the body surface to measure the voltage variations. These electrode couples are called leads and can be either bipolar or unipolar: the first ones are characterized by a couple of exploring electrodes, while the second ones use a single exploring electrode, and the electric potential difference is recorded with respect to one null-potential indifferent electrode.

Although different lead-systems exist, the standard ECG uses 10 electrodes, 4 peripheral (wrists and ankles) and 6 precordial electrodes, in order to record 12 leads: 6 from the limbs, of which 3 are bipolar and 3 unipolar, which measure the activity on the frontal plane (bipolar ones are called I, II, and III, while unipolar ones are called aVR, aVL, aVF), and 6 precordial and unipolar ones, which measure the electrical activity on the horizontal plane (they are called V1, V2, V3, V4, V5, V6)[7].

To obtain the bipolar leads (also called limb leads), one must build an equilateral triangle (Einthoven triangle) starting from the electrodes positioned on the right wrist (right arm, RA), on the left wrist (left arm, LA), and on the left ankle (left leg, LL), with the heart being virtually the centre of such triangle. The bipolar leads obtained as electrode couples are the following [9]:

- I: connects RA and LA, RA as the negative pole
- II: connects RA and LL, RA as the negative pole
- III: connects LA and LL, LA as the negative pole

With respect to these three leads, it is possible to estimate the projection of the dipole vector on the frontal plane, dividing the 360° area into three sections of 120° each. This procedure, however, is not enough to record in an adequate manner all the changes in the heart vector and hence it is necessary to select more leads that allow a better approximation of the events; in this regard, the Goldberger leads were introduced.

The Goldberger leads aVR, aVL, and aVF, also called "augmented" because they are amplified in order to be compared with the bipolar signals, are unipolar. The exploring electrode coincides with one vertex of the Einthoven triangle and is considered positive, while the negative polo is constituted by the Goldberger central terminal. The latter is the average of the remaining limbs' electrodes; connecting the extremities of each lead with two equal electrical resistances and using the junction

between them as a reference, with respect to the electrode placed on the opposite vertex of the triangle, it is possible to record alongside three more directrices, corresponding precisely with the bisectrices of the triangle itself. The three leads are obtained in this way:

- aVF: central terminal between RA and LA with respect to LL electrode

$$aVF = LL - \frac{1}{2}(RA + LA) \tag{1}$$

- aVR: central terminal between LA and LL with respect to RA electrode

$$aVR = RA - \frac{1}{2}(LA + LL) \tag{2}$$

- aVF: central terminal between RA and LL with respect to LA electrode

$$aVL = LA - \frac{1}{2}(RA + LL) \tag{3}$$

The conjunction of limb leads, and augmented leads put the basis of the hexaxial reference system, which is used to calculate the electrical axes of the heart in the frontal plane. In this manner it is possible to map the cardiac activity using 6 leads, dividing the frontal plane into regions of 60° each, with an accurate, still insufficient, level of detail [9].

In order to improve the definition, more electrodes placed in the immediate surrounding area of the heart are implied. They can identify and localize in a much more precise manner, for instance, lesions that could be missed using only the first six leads, and to analyze the heart depolarization vector on the transversal plane, which is perpendicular to the frontal plane seen previously.

In this case, another reference electrode is used, called Wilson terminal, obtained as the mean potential of the Einthoven voltages, used as the negative pole, and six exploring electrodes place in the precordial position as the positive poles. The position of the six electrodes, that gives the name "precordial" to this set of leads, is shown below [9]:

- V1: in the fourth intercostal space (between ribs 4 and 5) just to the right of the sternum (breastbone)
- V2: in the fourth intercostal space (between ribs 4 and 5) just to the left of the sternum
- V3: in between leads V2 and V4
- V4: in the fifth intercostal space (between ribs 5 and 6) in the mid-clavicular line
- V5: horizontally even with V4, in the left anterior axillary line
- V6: horizontally even with V4 and V5 in the mid-axillary line

The 12-lead system allows a complete and extensive description of the electrical activity of the heart (Figure 5)[9].

Other configurations are possible, and there is a general lack of standardization among them. One very common alternative lead system is the Mason-Likar: it is obtained from the standard 12-lead system but places the limbs electrodes respectively outwardly on the right and left shoulders and just below the umbilicus instead. This system is widely spread in emergency medicine (i.e., ambulances) and exercise or prolonged dynamic monitoring (e.g., Holter recorders) because it reduces the space occupied by wires and hardware. Although valid, this system lacks standardization and normalization and often is compared with standard 12-lead ECGs, although different positions of the electrodes describe different heart vectors, and their records differ during some moments of the activation [10].
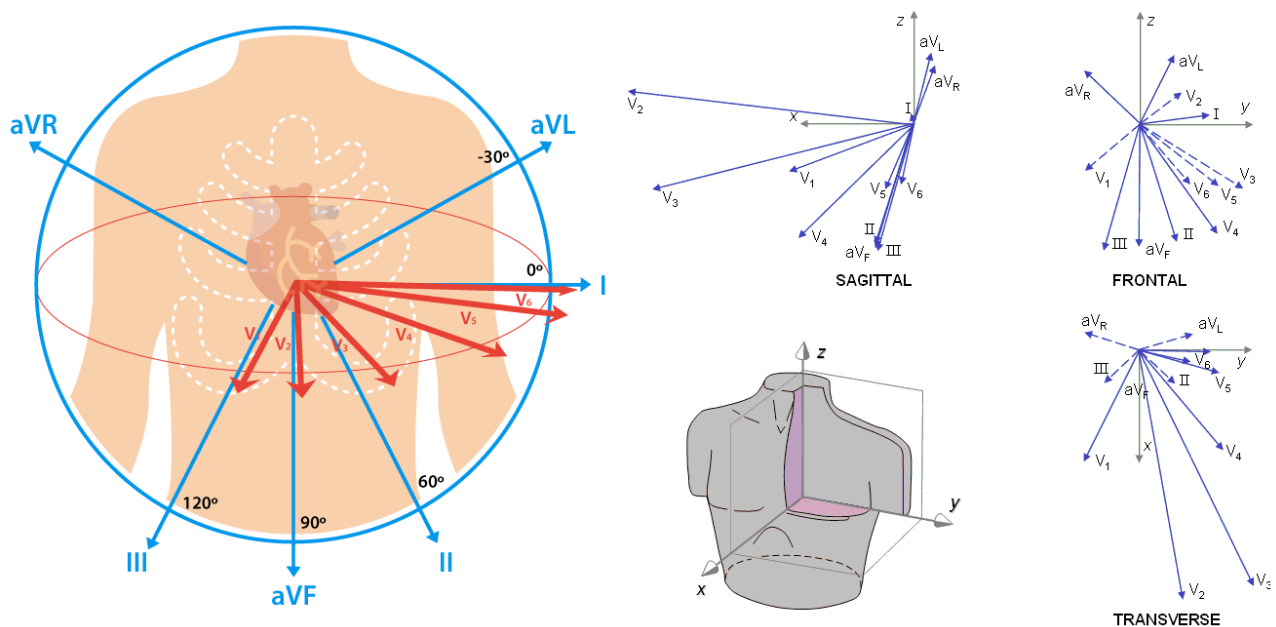
FIGURE 5 (LEFT) SCHEMATIC REPRESENTATION OF THE 12-LEAD VECTORS WITH RESPECT TO THE ELECTRODES' POSITION; (RIGHT) ACTUAL REPRESENTATION OF THE TRUE LEAD VECTORS IN 3D

2.3. PREVALENT ECG-DIAGNOSED DISORDERS

The main applications of the ECG to cardiological diagnosis are identifying the electric axis of the heart, monitoring its rate, highlighting irregularities (arrhythmias) as well as disorders in the activation sequence; it is also possible to use empirical criteria to diagnoses structural abnormalities, such as an increase in wall thickness or chambers size, myocardial ischemia, and infarction. These aspects will be discussed briefly. The theoretical support was provided mainly by the following book [5].

Recalling previously discussed topics, the electric axis of the heart is the direction of the heart vector, resulting from the average direction of all the dipole vectors. The normal range lies between +30° and -110° in the frontal plane and +30° and -30° in the transversal plane. The direction can be computed using the 12-lead ECG numerically. Deviation in the axis to the right denotes an increased electric activity in the right ventricle caused by an increase in right ventricular mass; the same holds for deviation to the left, which involves the left ventricle activity and mass. This disorder can be due to congenital heart diseases or pulmonary hypertension for right deviation and ischemic heart or conduction defect for left deviation.

The rhythm of a healthy heart is characterized by a normal sinus rhythm, meaning that the cardiac activation is triggered by the sinus node. It shows the normal P-QRS-T deflections and has a frequency that ranges from 60 to 100 bpm (Figure 6).

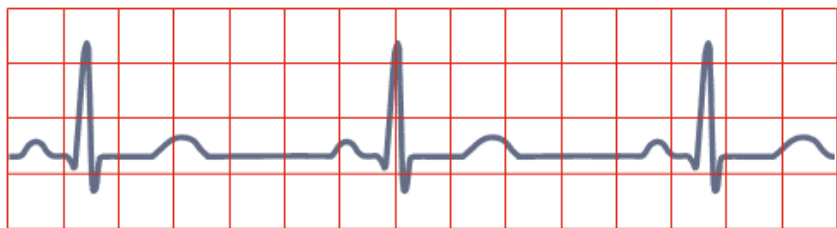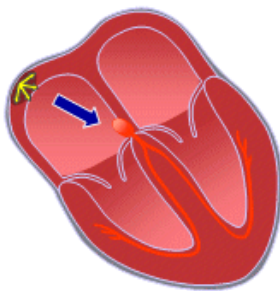FIGURE 6 NORMAL SINUS RHYTHM [5]

Irregularities in the heart rhythm originating in the atria or AV junction are called supraventricular arrhythmias. If the rate is slowed at the level of the sinus node below 60 bpm but the rhythm is regular, sinus bradycardia is diagnosed; it may be caused by an increase in vagal tone (Figure 7). The opposite effect is sinus tachycardia, occurring when the rate is greater than 100 bpm, usually in response to physical or psychical stress, but it may also result from congestive heart failure (Figure 8).

The sinus rhythm can also present significant discontinuities in the PP or RR interval, generally known as sinus arrhythmias. This situation is common to all age groups and is not considered a medical condition in young people.

The origin of the atrial contraction can also be delocalized, leading to a non-sinus atrial rhythm. If its position changes with time, this condition is called a wandering pacemaker. The P-waves will change in polarity, reaching the opposite sign when close to the AV node.

FIGURE 7 SINUS BRADYCARDIA [5]

FIGURE 8 SINUS TACHYCARDIA [5]

Paroxysmal atrial tachycardia takes place when the P-waves result from a re-reentrant activation front in the atrial, usually through the AV node in a circular fashion. One example is atrial fluttering: the heart rate is so elevated that the isoelectric interval between the end of T and the beginning of P disappears (Figure 9).
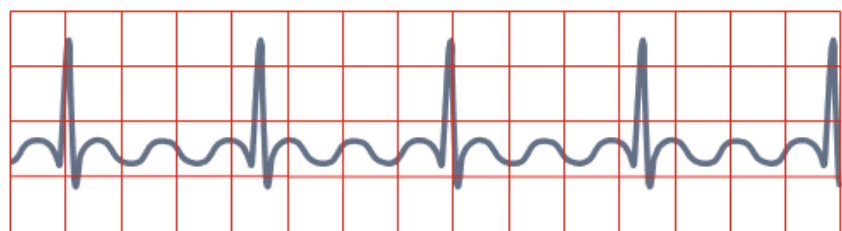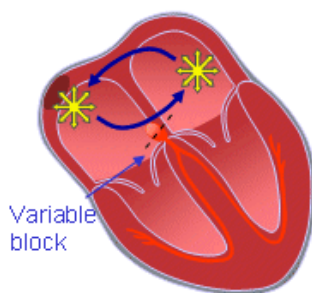
If the activation in the atria are fully chaotic and generates irregular fluctuations in the baseline, atrial fibrillation may occur, leading to a disjointed action of the atria and the ventricles. It can be seen as a temporary loss of coordination between the atria and the ventricles, as it is recovered automatically; it is fairly common and is not considered a medical condition unless it is a periodic event.

It is possible to have activation at the ventricles' level, leading to ventricular arrhythmias. These are considered the most dangerous because activations not originating in the AV does not propagate radially toward the outer walls and result in longer conduction times which can block the whole system.

Premature ventricular contractions occur abnormally early and can originate in the atria or the AV node (supraventricular) as well as in the ventricle. If the ventricles are continuously activated at the ventricular level (ectopic focus), the rhythm is called idioventricular. It may be due to slower conduction, such as in ischemic ventricular muscle, which leads to circular re-entry. In this case, the ventricle is activated at a high rate causing ventricular tachycardia (Figure 10).

FIGURE 9 ATRIAL FLUTTER [5]

VENTRICULAR TACHYCARDIA
Impulses originate at ventricular pacemaker
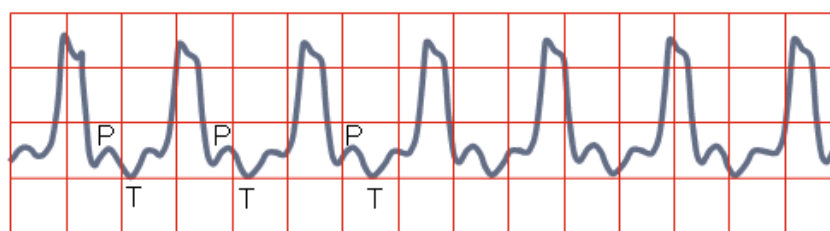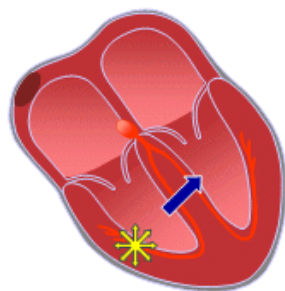
Wide ventricular complexes. Rate > 120/min

FIGURE 10 VENTRICULAR TACHYCARDIA [5]

In a similar fashion to atrial fibrillation, if activation in the ventricles are fully chaotic and generates irregular fluctuations in the baseline, ventricular fibrillation occurs. It is life-threatening; it does not resolve by itself and leads to death if left untreated within few minutes. During ventricular fibrillation, the heart loses its pumping function and cardiac output drops to zero. Cardiopulmonary resuscitation and defibrillation are needed to treat this condition.

As mentioned before, the ECG can diagnose conduction disorders too. They are referred to as blocks, with the most common being the AV block and bundle branch block (BBB).

The first one can occur in three degrees of severity: the first degree denotes a partial block at the AV level, the second degree causes sudden dropouts of QRS-complex, and the third degree presents a total block which leads to complete asynchrony between the P-wave and the QRS-complex.

BBB instead denotes a conduction defect in the right (right BBB, RBBB) or in the left (left BBB, LBBB). This means that the activation must of the ventricle, either the right or the left, must await initiation by the opposite side. This activation proceeds solely on a cellular basis, without the involvement of the conduction system, resulting in much slower and irregular conduction.

Anatomical abnormalities, such as right or left atrial hypertrophy and right or left ventricular hypertrophy, are likely to present conduction abnormalities. These are usually finer in the 12-leads representation, especially in the atrial cases, but can be almost immediately spotted constructing the heart vector. One example can be seen in Figure 11.
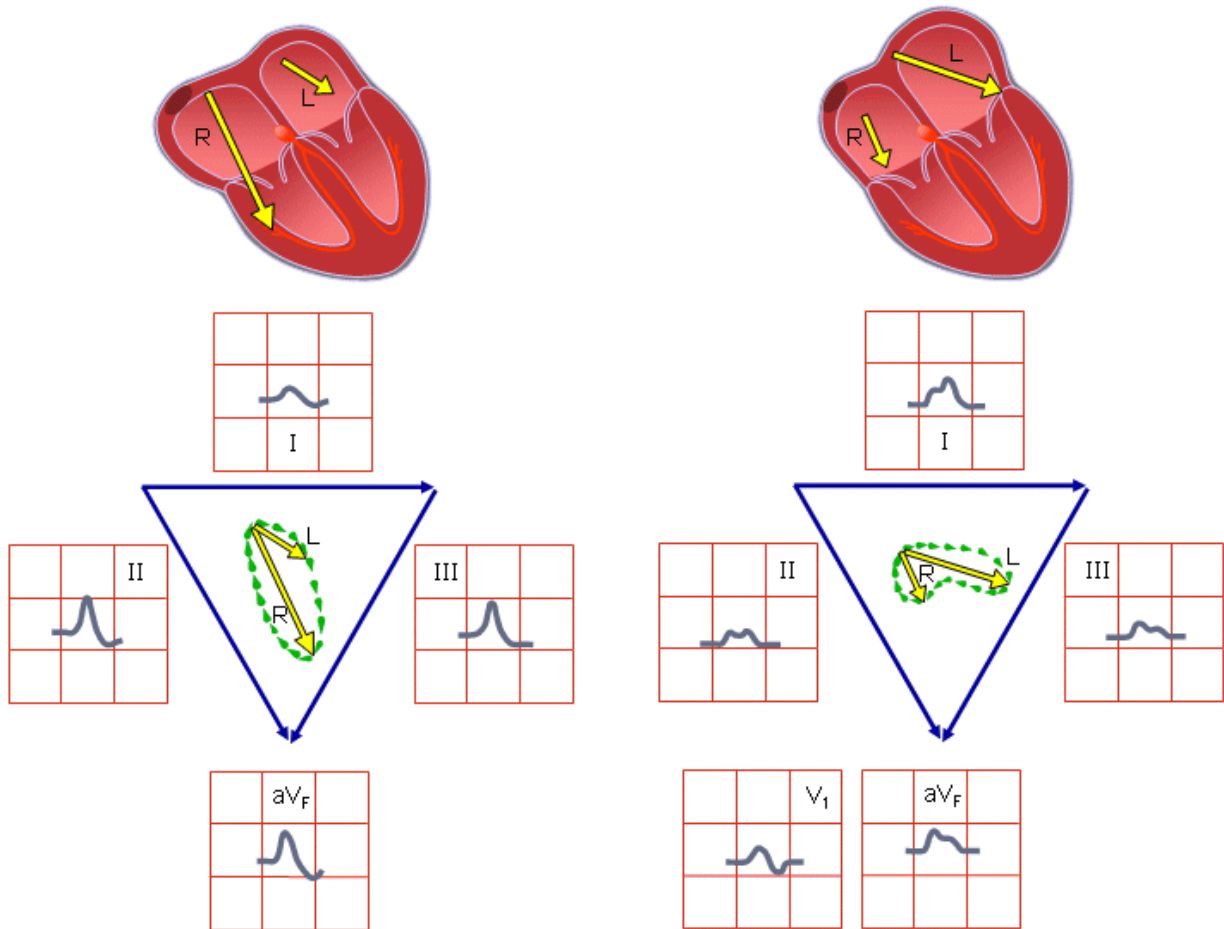
FIGURE 11 ATRIAL HYPERTROPHY [5]

3. ELEMENTS OF SIGNAL PROCESSING

The ECG is a critical signal for diagnosing and monitoring CVDs. But even if its standard interpretation relies primarily on direct visual assessment, there is a lot of critical information hidden underneath its surface [11]. The information content of the ECG is not limited to its tempo-spatial characteristics, with some of its features appreciable only in the frequency domain. This chapter will present a brief and practical introduction to signal processing in the frequency domain. The information will be spot on in order to have the correct number of elements to understand what is up to come but will not focus too much on analyzing tools proper of Mathematics.

3.1. THE FOURIER TRANSFORM

The most common approach to represent time-amplitude signals like the ECG in the frequency domain in order to extract new features is the Fourier Transform (FT) [12]. According to Fourier, a periodic function in time can be decomposed in a series of basic sinusoids, being sines and cosines (Euler's notation: $e^{j\theta} = \cos\theta + j\sin\theta$), each of them defined by their frequency, called harmonics.

$$s(t) = \sum_n \hat{s}(n)e^{j2\pi nf_0 t}$$

(4)

If the function is not periodic, the sum becomes an integral. The coefficients $\hat{s}$ are the weights of each harmonic, representing how similar the singular component is to the original signal, and are called FT. The FT of a periodic signal can be obtained integrating over its period:

$$\hat{s}(n) = \int_0^T s(t)e^{-j2\pi nf_0 t}dt$$

(5)

The FT of a non-periodic signal can be computed by changing the limits of the integral:

$$\hat{s}(f) = \int_{-\infty}^{+\infty} s(t)e^{-j2\pi ft}dt \tag{6}$$

The FT is no longer time-dependent and is calculated in the frequency domain as the convolution of the signal and a generic sinusoid; its value is the largest when they correlate the most (Figure 12).

In practice the signal has a very well-defined time of acquisition, meaning the integral never runs between -∞ and +∞, and is also not continuous in time, but it is sampled. This results in a set of a finite quantity of frequency called Discrete Fourier Transform (DFT):

$$\hat{s}_k = \sum_{n=0}^{N-1} s(n)e^{-j2\pi\frac{k}{N}n} \tag{7}$$

It is implemented in software with algorithms called Fast Fourier Transform (FFT) which reduces the number of operations. It is important to state that every Fourier transformation is invertible.

Representing the magnitude of the FT is possible to evaluate the distribution of energy in the signal across different frequencies, nevertheless losing information about the sequence of the events, since the FT does localize them in frequencies and not in time. A possible solution to this is implementing what is called the Short-Time Fourier Transform (STFT), consisting in having the FT of a temporal portion of the signal and then slide this window in time [13]:

$$\hat{s}_w(f, t_0) = \int_{-\infty}^{+\infty} s(t)w(t - t_0)e^{-j2\pi ft}dt \tag{8}$$
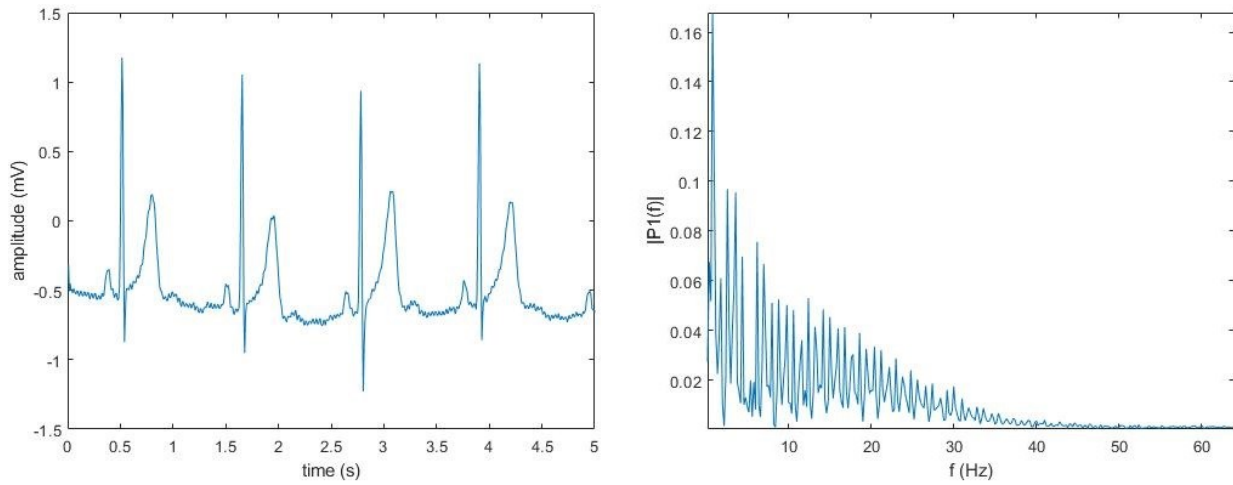
FIGURE 12: LEFT) ECG IN THE TIME DOMAIN; RIGHT) THE AMPLITUDE OF THE FFT IN THE FREQUENCY DOMAIN

3.2. THE WAVELET TRANSFORM

An alternative approach is the Wavelet Transformation (WT). WT decomposes the signal into a set of basic functions called wavelets (literally little waves), obtained from a proto-wavelet named mother wavelet by dilatation/contraction and by shifts.

The minimum requirement imposed on a function to qualify as a mother wavelet is that it satisfies the admissibility condition reported below, where the constant C_ψ is called the admissibility constant [14]:

$$0 < C_\psi := \int\limits_{-\infty}^{+\infty} \frac{|\Psi(\omega)|}{|\omega|} d\omega < \infty \tag{9}$$

In order to provide a useful time-frequency localization, the wavelet must be a well-localized function in time and in the frequency domain. One mild decay condition for $\psi(t)$ is square integrability. If the degree of decay is sufficient, the admissibility condition is equivalent to the requirement $\Psi(0) =$

- 22 -

$\int_{-\infty}^{+\infty} \psi(\omega)dt = 0$. Practically, the function must oscillate around the t-axis like a small wave, thus the name wavelet (Figure 13).

After choosing a mother wavelet, one can obtain a family of daughter wavelets by scaling and translating the function ψ [15]:

$$\psi_{\tau,s}(t) := \frac{1}{\sqrt{|s|}} \psi\left(\frac{1-\tau}{s}\right), \qquad s, \tau \in \mathbb{R}, \qquad s \neq 0 \tag{10}$$

where s is the scaling/dilation factor, acting on the width of the wavelet, and τ is the translation parameter. Changing the first causes the wavelet to stretch if $|s| > 1$ or to compress if $|s| < 1$, while the second causes shifting the wavelet position in time.

Given a time series $x(t)$, its continuous wavelet transform (CWT) can be expressed as a function of two variables (s, τ) given the mother wavelet ψ [15]:

$$W_{x;\psi}(\tau, s) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{|s|}} \psi * \left(\frac{t-\tau}{s}\right) dt \tag{11}$$
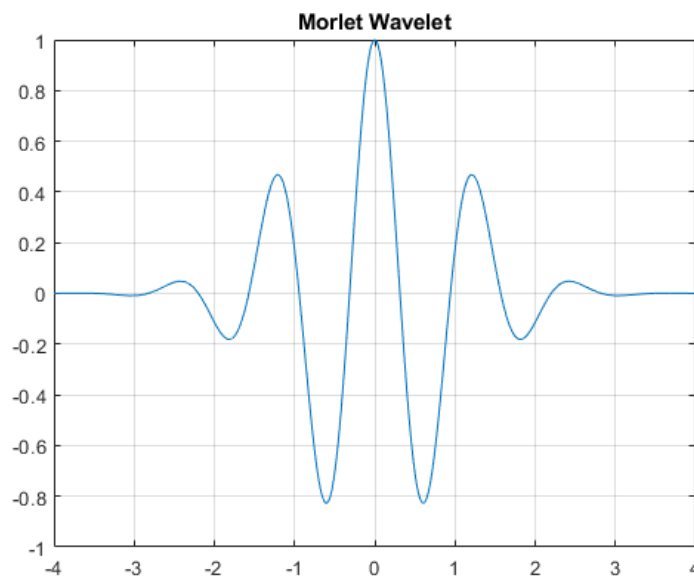

FIGURE 13 AN EXAMPLE OF WAVELET

The wavelet's position in the time domain is provided by τ, whereas its frequency domain position is given by s. By mapping the original series into a function of τ and s, the WT provides information on both time and frequency. The key distinctions between wavelet and Fourier transform are that the FT has no time localization parameter and instead uses cosine and sine functions instead of a wavelet function.

The admissibility condition from Equation (9) guarantees that the energy of the original function is preserved after the WT, thus ensuring the invertibility of the CWT. In fact, many reconstruction formulas are available because the WT is highly redundant.

Analogously to the FT case, the WT power spectrum (WPS), called scalogram or periodogram, can be defined both locally and globally when averaged over time (GWPS)[15]. (Note: the notation $W_{x;\psi}$ will be exchanged with W_x for brevity)

$$(WPS)_x(\tau, s) = |W_x(\tau, s)|^2 \tag{12}$$

$$(GWPS)_x(s) = \int\limits_{-\infty}^{+\infty} |W_x(\tau, s)|^2 d\tau \tag{13}$$

The mother wavelet can be complex-valued, and thus the corresponding WT will be complex, allowing the separate analysis in terms of real $\Re\{W_x(\tau, s)\}$ and imaginary $\Im\{W_x(\tau, s)\}$ part or alternatively of amplitude $|W_x(\tau, s)|$ and phase $\phi_x(\tau, s)$. The two representations are linked by the following relationship on the phase (phase-angle)[15]:

$$\phi_x(\tau, s) = Arctan \left(\frac{\Re\{W_x(\tau, s)\}}{\Im\{W_x(\tau, s)\}} \right)^2 \tag{14}$$

Real-valued functions have undefined phases; thus, it is necessary to use complex wavelets to separate amplitude and phase information. If it is the case, one should choose a so-called analytical wavelet, a particular mother that meets the following condition: $\Psi(\omega) = 0$ for $\omega < 0$. When ψ is analytical and $x(t)$ is real, it possible to recover $x(t)$ from the WT computed only for the positive value of s, by using the Morlet formula, often implied in numerical applications [16][17]:

$$x(t) = 2\Re \left[\frac{1}{K_{psi}} \int_0^{+\infty} W_x(\tau, s) \frac{ds}{s^{3/2}} \right] , K_{\backslash pis} := \int_0^{+\infty} \frac{\Psi^+(\omega)}{\omega} d\omega \tag{15}$$

Following the Heisenberg uncertainty principle from quantum mechanics and applying it to this particular context, there is a trade-off between the precision in frequency and in time [15]. Considering the probability density function $\frac{|\psi(t)|^2}{||\psi||^2}$, its mean $\mu_{t;\psi}$ gives a measure of the localization of the wavelet in time (centre in time). (Note: the notation $\mu_{t;\psi}$ will be exchanged with μ_t for brevity)

$$\mu_t = \frac{1}{||\psi||^2} \int_{-\infty}^{+\infty} t|\psi(t)|^2 dt \tag{16}$$

In the same way, the standard deviation $\sigma_{t;\psi}$ can give a measure of dispersion (radius in time). (Note: the notation $\sigma_{t;\psi}$ will be exchanged with σ_t for brevity)

$$\sigma_t = \frac{1}{||\psi||} \left[\int_{-\infty}^{+\infty} (t - \mu_t)^2 |\psi(t)|^2 dt \right]^{\frac{1}{2}} \tag{17}$$

The centre in frequency and radius in frequency can be defined by changing the probability density function $\frac{|\Psi(\omega)|^2}{||\Psi||^2}$

$$\mu_\omega = \frac{1}{||\Psi||^2} \int_{-\infty}^{+\infty} \omega|\Psi(\omega)|^2 d\omega \tag{18}$$

$$\sigma_\omega = \frac{1}{||\Psi||} \left[\int_{-\infty}^{+\infty} (\omega - \mu_\omega)^2 |\Psi(\omega)|^2 d\omega \right]^{\frac{1}{2}} \tag{19}$$

The interval $[\mu_t - \sigma_t, \mu_t + \sigma_t]$ is the set where $\psi(t)$ has its most significant values; the same interval can be definite in frequency as such $[\mu_\omega - \sigma_\omega, \mu_\omega + \sigma_\omega]$.

Back to the uncertainty principle, it is possible to define the Heisenberg window for the function ψ around the point (μ_t, μ_ω) on the (t, ω) plane:

$$H_\psi := [\mu_t - \sigma_t, \mu_t + \sigma_t] \times [\mu_\omega - \sigma_\omega, \mu_\omega + \sigma_\omega] \tag{20}$$

The uncertainty is given by the radius product and establishes the lower bound of the uncertainty principle

$$\sigma_t \sigma_\omega \geq \frac{1}{2} \tag{21}$$

Since the daughter wavelet $\psi_{\tau,s}$ can be obtained from ψ by simple translation (τ) and scaling (s), the centre and radius in time of the wavelet are given by respectively:

$$\mu_{t;\psi_{\tau,s}} = \tau + s\mu_t, \qquad \sigma_{t;\psi_{\tau,s}} = s\sigma_t \tag{22}$$

The WT $W_x(\tau, s)$ provides temporal information on $x(t)$ around the instant $t(\tau) = \tau$, with precision $s\sigma_t$, and frequency information about $X(\omega)$ around the frequency

$$\omega(s) = \frac{\mu_\omega}{s} \tag{23}$$

With resolution $\frac{\sigma_\omega}{s}$.

The WT provides a time-scale representation of the analysed function, not strictly a time-frequency representation. It is possible to convert scales into pseudo frequency using the formula (23) when considering μ_ω given by (23).

The most used wavelet family across different applications are the analytical Morlet wavelets (Figure 13)[18]. They are a one-parameter family of functions:

$$\psi_{\omega_0}(t) = K e_o^{i\omega t} e^{-\frac{t^2}{2}} \qquad (24)$$

The above functions are not proper wavelets because they do not satisfy the admissibility condition but, assuming $K = \pi^{-1/4}$ (in this it has unit energy) for values $\omega_0 \geq 5$, the values of $\Psi_{\omega_0}(\omega)$ for $\omega \leq 0$ are so small that it can be considered numerically an analytical wavelet [19].

The Morlet wavelet is so popular because of four properties: it is an analytical wavelet (only for numerical purposes); it has peak frequency, energy frequency, and central instantaneous frequency equal to ω_0 (facilitating the conversion from scales to frequencies); it has optimal joint time-frequency concentration (the Heisenberg box area reaches it is lower bound with this wavelet); it has a time radius and frequency radius equal to $\frac{1}{\sqrt{2}}$ (achieving an excellent compromise between time and frequency accuracy).

3.3. WAVELETS FOR BIOMEDICAL SIGNALS

WT remains nowadays much more obscure and unutilized than FT, but its applications could be certainly wider than its current niche. One counterintuitive example may be biomedical signals, in particular, but not limited to the ECG.

First, FT assumes the signal to be stationary, which means the frequencies present in the signal are not time-dependent; FT offers only an approximation for non-stationary, which is worse when the signal exhibits more dynamic components. The ECG, like other biological signals, is only assumed to be stationary, but instead, it is not. WT, on the other hand, is not limited by that, but instead can be specifically designed to work with dynamical systems.

Another crucial aspect is that the FT is not localized in time, because the sine waves range from -∞ to +∞, and hence have no resolution in the time domain, only in the frequency one. Signals which exhibit different temporal dynamics can contain the same set of frequencies, thus losing information

when transforming the signal. This problem could be solved using the STFT like mentioned before, but this could lead to the uncertainty principle, the theoretical limits of the FT. Reducing the window's size allows to gather more information in the time-domain at the cost of losing information in the frequency one; the vice versa also holds (Figure 14).

Wavelets, instead, are localized in time, thus intrinsically the WT has both time and frequency resolution. This means that with the WT is possible not only which frequencies are present in a signal, but also when they occur. Using the scale factor, it is possible to obtain high resolution in the frequency domain for small frequency (at the cost of low resolution in the time domain), but at the same time high resolution in the time domain for high frequency (at the cost of low resolution in the frequency domain)(Figure 14).

Another crucial aspect of WT is its representation. Using the convolution operator, as mentioned before, it is possible to multiply a portion of the ECG signal in the time domain for the wavelet, whereupon shifting the latter until the end of the original signal, and after that repeat the process for each scaled wavelet. This will end in a bunch of time-amplitude signals, one for each scale used; thus, considering the signals' amplitude as "latitude" using a colour scale, the resulting representation will be a 2D colour plot with axes scale and time, which is the scalogram (Figure 15).
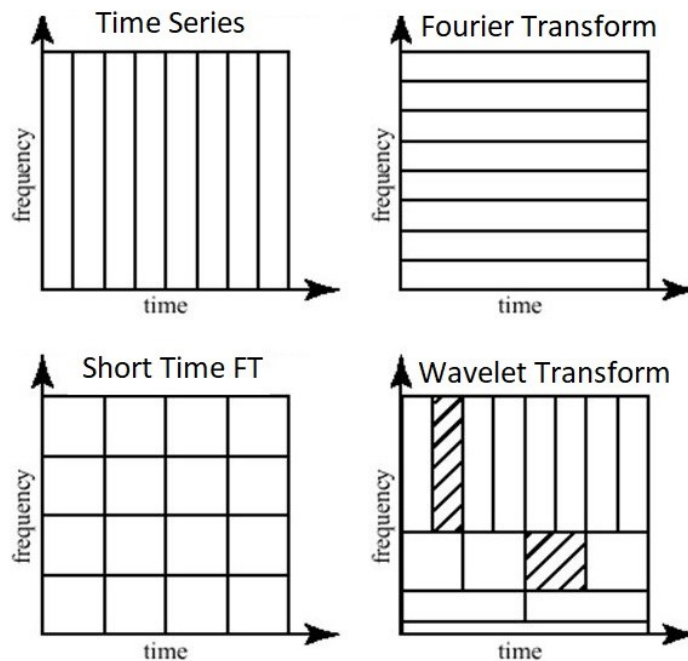

FIGURE 14 GRAPHICAL REPRESENTATION OF THE DIFFERENT RESOLUTIONS OF VARIOUS TRANSFORMS
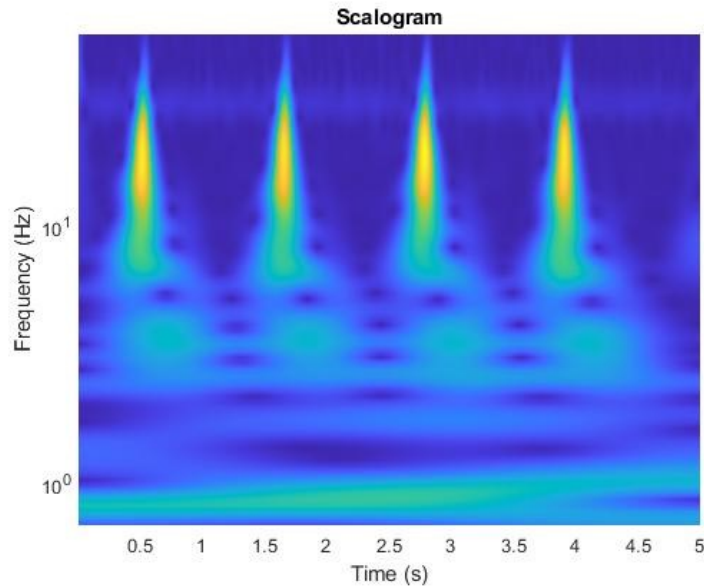
The last worth-mentioning characteristics of the WT for biomedical signals are still related to their scalograms representation. In addition, to give to provide information about the frequency domain, localize it in time, and understand the dynamical behaviour of a system, scalograms can be used to differentiate types of signals produced by a system from each other. Transferring this idea to the ECG, it is possible to distinguish between normal and pathological tracings, then classify them just by looking at their scalograms. Of course, even if such representations are richer in information content, a physician would more confidently evaluate a standard time-amplitude ECG, or even its FT power-spectrum rather than the WT scalograms. But the same cannot be said for computer vision algorithms, which can rather easily take as input 2D RGB images than a time series.

4. ARTIFICIAL INTELLIGENCE

Artificial Intelligence (AI) is a vast and daily overgrowing field in computer science. Its many shapes and forms allow application and research practically everywhere. In this chapter, only a subset of AI will be presented. The following will try to be as exhaustive as possible, focusing at the same time on those tools and practical aspects that will come in handy later on.

4.1. MACHINE LEARNING AND DEEP LEARNING

In machine learning (ML), one develops and studies methods that give computers the capability to solve problems by learning from experiences [20]. This is achieved using mathematical models which can be "trained" to produce relatable outputs using the input data. Once optimized, the model must be able to generalize what it has learned on new, unseen data, and still obtain good results.

Thus, it is obvious, ML relies on data, possibly loads of data. For this reason, these models are called data-driven and sometimes using the black-box parallelism, both to shift the focus on their inputs/outputs and because of their general lack of interpretability.

Datasets used for ML are usually split into two subsets. The training subset is required in order for the model to learn, which in general is done via the minimization of an optimization function. The validation subset, instead, is used when the model is already trained and is used to prevent overfitting and determine the optimal design of the model.

One second dataset of smaller entities is commonly implied too. The testing dataset serves to evaluate how well the model generalizes what is learned on new data.

There are several kinds of ML, depending on how the input data are categorized. The most common is supervised learning, where the data are labelled, for example with a pathological status, by an expert. In semi-supervised learning, similarly, the majority of the data presents labels, but they're mixed with unlabelled or weakly labelled (e.g., annotations in forms of scribbles) [21]. Self-supervised learning instead uses self-generated labels, without human intervention. Unsupervised

learning, on the other hand, does not present any labels and is used to explore hidden structures in the data.

ML models are trained to perform their tasks using manually selected features extracted from the raw data. This kind of approach proved to be very effective, as long as one is able to select the best-representing set of features. But there is a subset of ML models that do not require selecting any features, but instead, learn by themselves which are the most relevant ones just from the raw data. That is called deep learning (DL) [20].

The most common application of DL is artificial neural networks (ANN), and in particular convolutional neural networks (CNN), which will be discussed later. These models started outperforming any other AI image analysis algorithm due to the high computational power of graphical processing units (GPU) and the availability of larger datasets. Notably, in 2012 a DL model won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) halving the second-best error rate on the image classification task [22]. The quick progress and wide range of applications of DL attracted investments and attention over the years, making the entire field of ML the hot trend of studies worldwide [23]. Medical imaging techniques that produce huge amounts of data, at a pace higher than the traditional analysis methods can process, are, on paper, an easy match for DL models [21].

4.2. ARTIFICIAL NEURAL NETWORKS

ANNs are hierarchical structures, consisting of a set of stacked layers, each composed of functional units referred to as neurons.

The first layer is the input layer, where the data enter the network. Then, it follows at least one hidden layer, where the data is transformed as it flows through the network. Lastly, there is the output layer, where the network makes its prediction. The result is then compared with the expected value, the initial label of the data, via an objective function. Then the ANN adjusts its internal parameters and starts again. Going through this cycle of trial and error, during which the objective function, is what for a machine means to learn.

Now more in detail, it will be presented the functioning of a basic ANN, in order to have a grasp of what is really happening inside one of them. The example of a feedforward neural network (FNN) or multilayer perceptron is useful to visualize the process (Figure 16).

The FNN is a parametrized mathematical function $y = f(x; \theta)$, where an input x, which is fed through a series of nonlinear transformations is mapped to an output y:

$$y = f(x; \theta) = (f_n \circ \dots \circ f_1)$$

(25)

Each transformation is a layer of the network, composed of a simple linear transformation of the previous component's output and a non-linearity σ_k; the latter is generally a sigmoid function or, even more commonly, a rectified-linear-unit (ReLU) function (Figure 17), while the matrix θ_k contains the so-called weights of the k-layer [24]:

$$f_k = \sigma_k(\theta_k^T f_{k-1}), \qquad k = 1, \dots, n$$

(26)

Each layer then computes the weighted sum of all the outputs from every neuron in the previous layer, followed by the activation function. The result is called layer activation. Those are then fed into the next layer of FNN and so on until the output layer is reached.
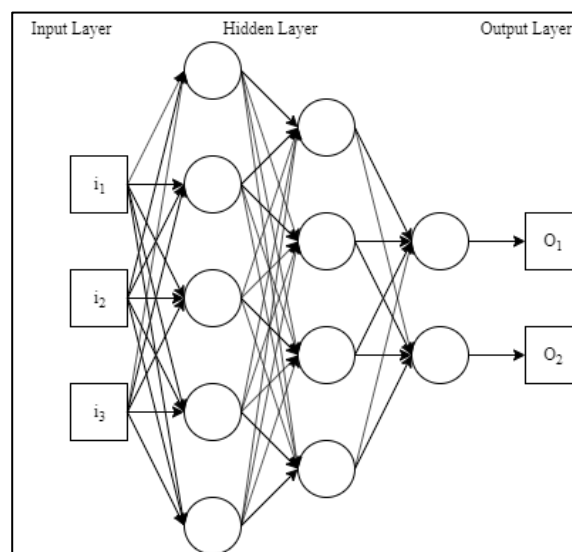

FIGURE 16 MULTI-LAYER PERCEPTRON (FNN)
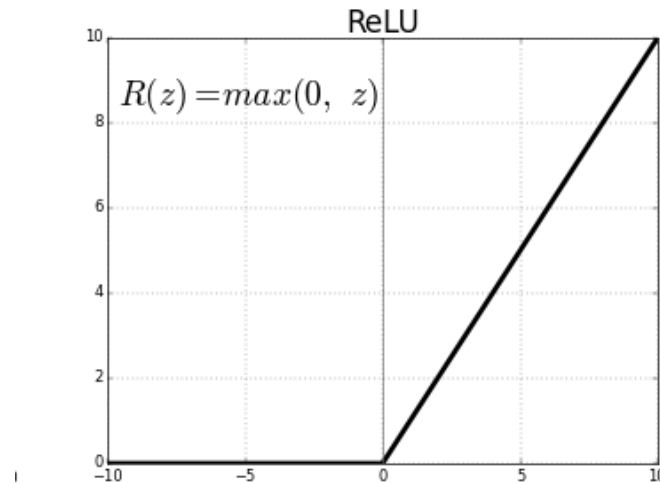
$$R(z) = max(0, \ z)$$

FIGURE 17 ReLU ACTIVATION FUNCTION

During training, the FNN uses the input data to perform predictions at the output layer, trying to match the original labels. The model must learn to use the intermediate representations to form complex hierarchical representations of data to achieve a correct prediction. The earlier features are very general, usually referring to the general shape of the data, but the deeper the network goes, the finer details it is able to extract. It is worth noting that a deep network is not always synonymous with good, because, in addition to becoming more computationally expensive, overfitting may occur.

In order to do so, its weights are adjusted during the second path, also called back-propagation, in which an objective function, called cost or loss function, is minimized. Normally implied loss functions depend on the performed task: for classification purposes, it is possible to use mean squared error as a basic option, or more involved ones such as categorical cross-entropy [24]. The derivative of the objective function is computed, and the weights are updated backwards by means of an optimization algorithm called gradient descent. Typical choices are the classical stochastic gradient descent (SGD), the root-mean-square propagation (RMSProp), and the Adam optimizer [25].

4.3. CONVOLUTIONAL NEURAL NETWORKS

In the matter of image analysis, even if it is possible to use simple FNN, it would be extremely inefficient. This is due to the fully connected structure of the hidden layers, where each neuron is connected to all the neurons in the following layer. Considering a model which keeps intact the structure of the image results in a much better performance.

(CNN) are the most common DL networks for image analysis and classification because are able to preserve the spatial relationships in the data using fewer connections.

The input layer of CNN is arranged in a grid-like shape; the feedforward path preserves this relationship, with each layer operating on a small region of the previous one.

The core element of every CNN is the presence of at least one convolutional layer. In order to understand the extreme importance of this peculiar element, it can be useful to approach the discussion using practical examples.

Supposing that the input of the CNN is an image with dimensions of 32x32 pixels and depth equal to 3, which is the number of colour channels for an RGB image). In order to connect this input to a single neuron in FNN, it would require a number of weights equal to input size multiplied by 1, so 32x32x3. It may be thought though that more than one neuron might be required in order to perform meaningful classification, and that a suitable number of neurons in the first hidden layer could be equal to the number of pixels of the input image, raising the number of parameters up to 32x32x3 by 32x32.

It is obvious that this approach is not efficient. Passing just from the input layer to the first hidden layer requires training 3145728 parameters [26].

One alternative approach could be to look for local regions in the input and connecting those regions, instead of a single pixel, to the next layer. The hidden neurons in this way only receive inputs from the associated part of the previous layer, for example from 5x5 neurons. Considering like before the first input layer with 32x32, the number of connections will be 5x5x3 by 32x32, which results in 76800. Even if the reduction in computational terms is important, the number of parameters is still elevated [27].

An additional assumption is to keep the local connection weights fixed for the entire neurons of the next layer. This will connect the neighbour neurons in the next layer with exactly the same weight to the local region of the previous layer. Numerically, this means reducing the number of parameters to 5x5x3=75 [28].

This approach, which mathematically corresponds to the convolution operation, can, as illustrated, significantly reduce the number of parameters, but also can detect features in the image regardless of their positions in the image. This is because conceptually this method is equivalent to sliding a window of fixed dimensions (e.g., 5x5x3) in the input layer and mapping the output to the corresponding place (Figure 18). In this way, it is possible to design multi-dimensional matrices called

tensors able to detect edges and shapes in the image. For this reason, a convolutional operation can be seen as a filter [29].

In a brand-new CNN, these tensors are not initialized, and their value will change during training in order to complete their given task. It is also possible to stack multiple convolutional layers in the same CNN; in this way, one can extract features that are the more particular and hidden the deeper the network goes [30].

The size of a convolutional layer also referred to as kernel, is not the only parameter that must be initialized when designing this kind of layer. The stride controls the overlap between the filter and the layer's neurons. Controlling the stride allows controlling the size of the input too. Assuming the input dimension equal to $N \times N$ and the filter size of $F \times F$, the output size $O \times O$ will be given by:

$$O = 1 + \frac{N - F}{S} \tag{27}$$

Where S is the stride value. If, for example, N is equal to 7, considering a unitary stride, the filter will move one node every time, and O will be equal to 5; but if instead, the stride is equal to 2, then O will be 3 (Figure 19). This means that the output size will be inevitably reduced when performing a convolution [30].
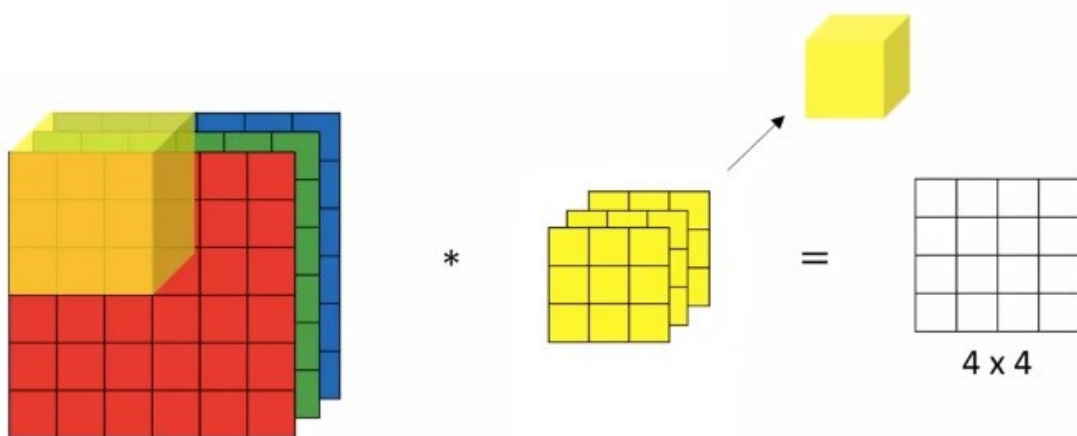

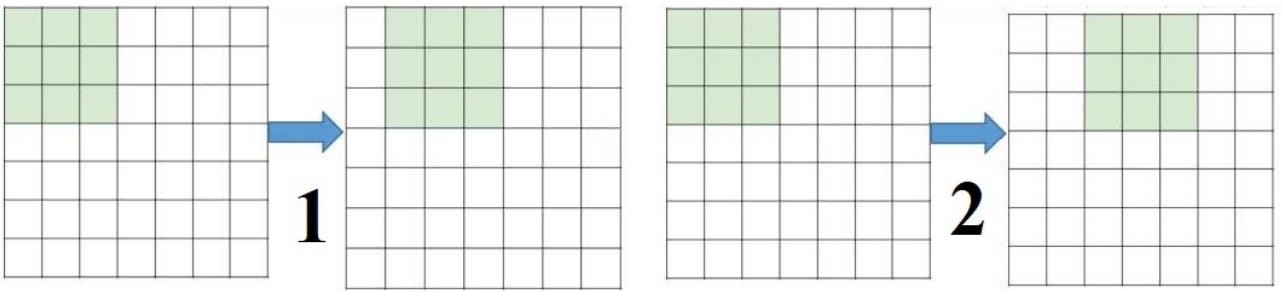FIGURE 18 CONVOLUTION OF A 6X6X3 (RGB) INPUT FOR A 3X3X3 FILTER

FIGURE 19 EXAMPLES OF DIFFERENT STRIDE VALUES

It is possible to solve this issue as well as the loss of information at the borders of the input, which are interested in the smallest number of convolutions, acting on the padding parameter. Similar to other applications, zero-padding increases the size of the input adding zeros at the outer borders (Figure 20). With respect to Equation (27), also consider the contribution P for each size increment achieved using padding:

$$O = 1 + \frac{N + 2P - F}{S} \tag{28}$$

Padding is particularly useful with deep CNN, which features many convolutional layers and would lose every spatial information by reducing too much the input size of each layer [24].

Nevertheless, the weight sharing of the convolutional layer bring translational invariance to the model, which heavily favours feature extraction regardless of spatial properties; if the spatial content is somewhat important for the given task, it is better to avoid using shared weight as a general concept, with some exceptions [31].

The illustrated concepts are fully valid for multi-resolution and multi-dimensional applications, such as hyper-resolution and 3D images [30].
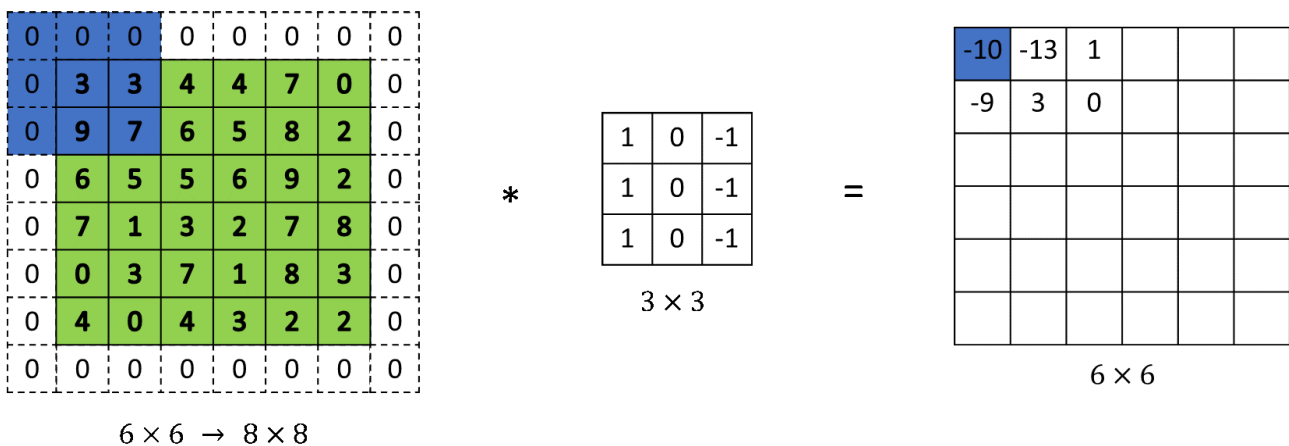

FIGURE 20 2D CONVOLUTION (BLUE) WITH AN INPUT (GREEN) CONSIDERING PADDING = 1 (DOTTED)

Convolutional layers also use non-linearities as their action layers, with the most common being ReLU.

Other recurring elements of a CNN are explained below [20][24]:

- Pooling layers: are a type of non-linear down-samplings. Pooling can be implemented using a variety of non-linear functions, the most common of which being max-pooling (also average pooling). It divides the input image into a series of rectangles and outputs the maximum for each sub-region (Figure 21). These layers reduce both the dimension and number of parameters, allowing to control overfitting; for this reason, such layers are often included between successive convolutional layers. They contribute to local translation invariance, but not to global one unless they perform a global pooling operation (kernel equal to the input size). It is important to control the down-sampling effect of pooling layers, or either avoid them altogether [24].

- Dropout regularization layers: randomly remove neurons during training, obtaining slightly different networks for each training batch; in this way, the weights are tuned based on the optimization of more than one possible network. This technique is an averaging procedure based on the stochastic sampling of neural networks. The average of different models ensembled together tends to perform better. This relatively simple idea proved very useful to prevent overfitting and generally to boost the CNNs' performances.

- Batch normalization layers are usually placed after activation ones and produce normalized feature maps, subtracting the mean and diving by the standard deviation. These layers force the CNN to periodically reset its activation to zero mean and unit standard deviation, which grant regularization, speed up the network, and reduce the dependency to close-to-correct parameter initialization.
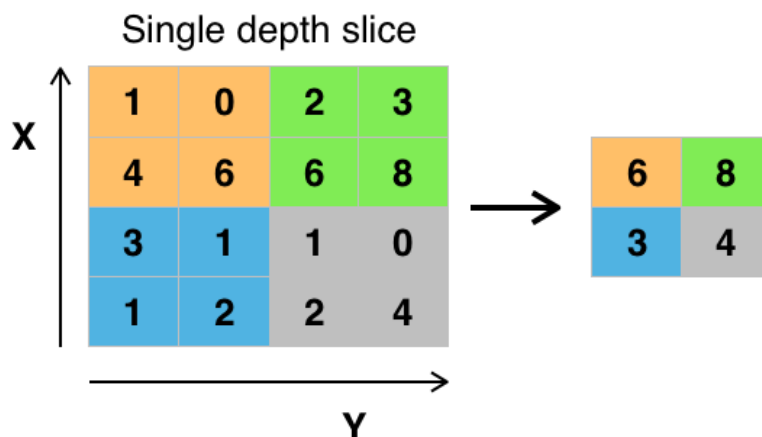

FIGURE 21 2D MAX POOLING WITH A 2X2 FILTER AND STRIDE EQUAL TO 2

- Fully connected layers are the type of layer discussed before when talking about FNN. Their relevance is restricted to prediction, thus are frequently referred to as the classifier part of the network. These layers can also have their own activation function, like a sigmoid, but often they use a SoftMax function, which normalizes its inputs into a probability distribution, with each component in the interval between 0 and 1 and adding up to 1; in this way, the larger input component will correspond to the function output, which is the prediction of the network. The main drawback of fully connected layers is their large number of parameters which burdens the CNN performance. One way to reduce the computational expense is to reduce the number of nodes and connections, for example by using the dropout technique.

The presented elements are sufficient to understand the majority of CNN architecture used for classification purposes (AlexNet [22], GoogLeNet [32], ResNet [33], VGG [34]), which can be described as a combination of the previous elements with small variations.

Other CNN architectures are possible and can be designed ad hoc in order to perform a certain task. Just to bring up a few examples: fully CNN (FCN) don't present fully connected layers to avoid diving each image into multiple patches and train one CNN to predict each central pixel for every patch and instead perform end-to-end pixel-wise segmentation using an encoder-decoder structure which can input images of any size and output of the same size (usually achieved via upsampling operations) [31]; recurrent CNN (RCN) on their behalf are used for sequences of images because they have "memory" of the past and can use it for making future decisions (often used for video classification) [35].

For many applications, there is a scarcity of training data, but in order to avoid overfitting, CNNs often require a substantial amount of them. Training the network on a larger data set from a related domain is a common practice. After the network parameters have converged, transfer learning is used to fine-tune the network weights using in-domain data. Transfer learning means applying information, strategies, and skills learned in a previous context to a new, different yet similar, context. Furthermore, convolutional network topologies can be successfully applied to issues with small training sets using this technique.

5. EXPERIMENTAL SETTINGS

With a solid theoretical background, it is possible to introduce the problem of the ECG classification with an adjustable number of leads. The following chapter will introduce the experimental setting as well as the material and methods to solve the anticipated problem.

5.1. BACKGROUND

For the 2021 edition of the international conference Computing in Cardiology (CinC), PhysioNet proposed its annual challenge going by the name "*Will two do*?". The goal was to identify clinical diagnoses from the following lead set of ECG recordings:

1. Twelve-leads (I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, V6)
2. Six-leads (I, II, III, aVR, aVL, aVF)
3. Four-leads (I, II, III, V2)
4. Three-leads (I, II, V2)
5. Two leads (I and II)

Being able to perform classification using a reduced-lead system can be useful in some contexts, such as in emergency medicine or when one or more leads are affected by noise. Evidence in the literature shows that subsets of the standard 12-lead systems can be used for capturing useful pieces of information, comparable to the full 12-lead ECG even [36][37][38].

This challenge was a direct follow-up to the CinC 2020 Challenge going by the name "*Classification of 12-lead ECGs*". Since the results on both the available and the undisclosed (hidden) datasets, which are partially shared with this year's, were recently published, it is possible to draw some expectations [42].

Of the seventy teams' codebases that successfully ran on the test data, forty-one qualified for the final rankings. The most common algorithmic approach was based on DL and CNNs, but the majority of

entries (70%) used standard clinical or hand-crafted features with ML classifiers (support vector machines, gradient boosting, random forests, and shallowed neural networks) [42].

The performances of each team's final algorithm were evaluated on the CinC validation set, the hidden CPSC set, the hidden G12EC set, the hidden undisclosed set, and the test set. The results show that on average, the challenge scores dropped 47% from the hidden CPSC set to the hidden G12EC set and another 57% from the hidden G12EC set to the hidden undisclosed set, as well as an average drop of 50% from the validation score set to the test set [42].

This under-performance could be due to models being over-trained on the CPSC available data while the hidden CPSC constituted only a small fraction of the validation and testing set. The poorer scores and ranks show the importance of including multiple sources of data in order to properly generalize the algorithms. Some developed algorithms were adapted from different applications.

The class imbalance between the datasets was, and still is, significant, but is representative of the real-world problem of interpreting 12-lead ECGs in a clinical environment.

In this regard, most models performed at their best on the CPSC dataset because it had fewer and more balanced diagnoses while at the same time being the least representative.

The challenge discouraged a priori usage of information on distributions since the algorithms are likely to be used in a variety of unknown populations.

For a comprehensive description of the performances of the undisclosed models with respect to the proposed scoring function, please refer to the official review paper of CinC 2020 [42].

5.2. MATERIALS AND METHODS

5.2.1. DATA

The data provided by the challenge staff were collected from six different sources from four countries in three continents around the world. The grand total was over 100000 12-lead ECG, but the publicly available ones were 84800, while the others were retained as private validation and testing data.

The available sources with their characteristics are the following:

1. China Physiological Signal Challenge in 2018 (CPSC 2018), held during the 7th International Conference on Biomedical Engineering and Biotechnology in Nanjing, China [39]. It contains two databases, the CPSC Database and the CPSC-Extra Database, for a total of 10330 public ECG recordings with a duration between 6 and 144 s and a sampling frequency of 500 Hz

2. St Petersburg INCART 12-lead Arrhythmia Database [40]. It contains 74 ECG recordings with a duration of 30 minutes and a sampling frequency of 257 Hz

3. Physikalisch-Technische Bundesanstalt (PTB)[41]. It contains two databases, the PTB and the PTB-XL, for a total of 22353 ECG recordings with a duration between 10 and 120 s and a sampling frequency of either 500 or 1000 Hz

4. Georgia South-eastern United States database (G12EC)[42]. It contains 10334 public ECG recordings with a duration between 5 and 10 s and a sampling frequency of 500 Hz

5. Chapman University, Shaoxing People's Hospital (Chapman-Shaoxing), and Ningbo First Hospital (Ningbo) database [42][43]. It contains 45152 ECG recordings with a duration of 10 s and a sampling frequency of 500 Hz

All the data were formatted using the standard WFDB. Each ECG recording is made of a coupling of a binary MATLAB v4 file containing the actual signal data and a WFDB header of plain text for the recording attributes, the patient specifics, and diagnosis. The latter corresponds to the label, and it is expressed as the SNOMED-CT code of the corresponding pathology. Each recording can be associated with more than one label. One example is shown in Figure 22.


```
1    A0001 12 500 7500 12-May-2020 12:33:59
2    A0001.mat 16+24 1000/mV 16 0 28 -1716 0 I
3    A0001.mat 16+24 1000/mV 16 0 7 2029 0 II
4    A0001.mat 16+24 1000/mV 16 0 -21 3745 0 III
5    A0001.mat 16+24 1000/mV 16 0 -17 3680 0 aVR
6    A0001.mat 16+24 1000/mV 16 0 24 -2664 0 aVL
7    A0001.mat 16+24 1000/mV 16 0 -7 -1499 0 aVF
8    A0001.mat 16+24 1000/mV 16 0 -290 390 0 V1
9    A0001.mat 16+24 1000/mV 16 0 -204 157 0 V2
10   A0001.mat 16+24 1000/mV 16 0 -96 -2555 0 V3
11   A0001.mat 16+24 1000/mV 16 0 -112 49 0 V4
12   A0001.mat 16+24 1000/mV 16 0 -596 -321 0 V5
13   A0001.mat 16+24 1000/mV 16 0 -16 -3112 0 V6
14   #Age: 74
15   #Sex: Male
16   #Dx: 59118001
17   #Rx: Unknown
18   #Hx: Unknown
19   #Sx: Unknown
```

FIGURE 22 HEADER OF "A0001.MAT" FILE, CPSC DATABASE

The recording number is A0001, and the file name is A0001.mat. The recording has 12 leads, recorded at a 500 Hz sampling frequency, and contains 7500 samples. The following 12 lines of the file show that each lead was written at 16 bits with an offset of 24 bits, the floating-point number (analogic-to-digital converter (ADC) units per physical unit) is 1000/mV, the resolution of the analogic-to-digital converter (ADC) used to digitize the signal is 16 bits, and the baseline value corresponding to 0 physical units is 0. The last three values are to the first value of the signal (-1716, etc.), the checksum (0, etc.), and the lead name (I, etc.). The final 6 lines give information about the patient: 74-year-old male with a diagnosis (Dx) of 59118001, SNOMED-CT code for RBBB. The medical prescription (Rx), history (Hx), and symptom or surgery (Sx) are unknown

The dataset is loaded on MATLAB. This allows a first inspection of the data and their labels. The number of classes of this classification problem corresponds to the number of diagnosed pathologies present in the dataset. If this number is greater than two, which would result in a binary classification problem where each recording would either belong to one class or the other, the problem is said to be multi-class.

The dataset is of course expected to have more than two classes, so a second evaluation is made for the labels associated with the data. If one recording presents more than one class (i.e., more than one pathology is diagnosed), the classification is called multi-label.

The initial 84800 files were arranged in seven folders. After loading, they present a total number of 126 unique classes. Refer to Table 1 to see the contribution of each dataset

Dataset name	Number of files	Number of classes	Unique classes
Chapman	10247	54	5
CPSC2018	6877	9	0
GA	10344	67	9
Ningbo	34905	80	19
PTB	516	17	5
PTBXL	21837	50	3
StPetersburg	74	37	6

TABLE 1 PRELIMINARY SCROLL OF THE NOT-YET MERGED DATASETS

During the preliminary labels scouting, 40036 files (47,2%) presented a single label. The maximum number of concurrent labels was 12, found in 1 file.

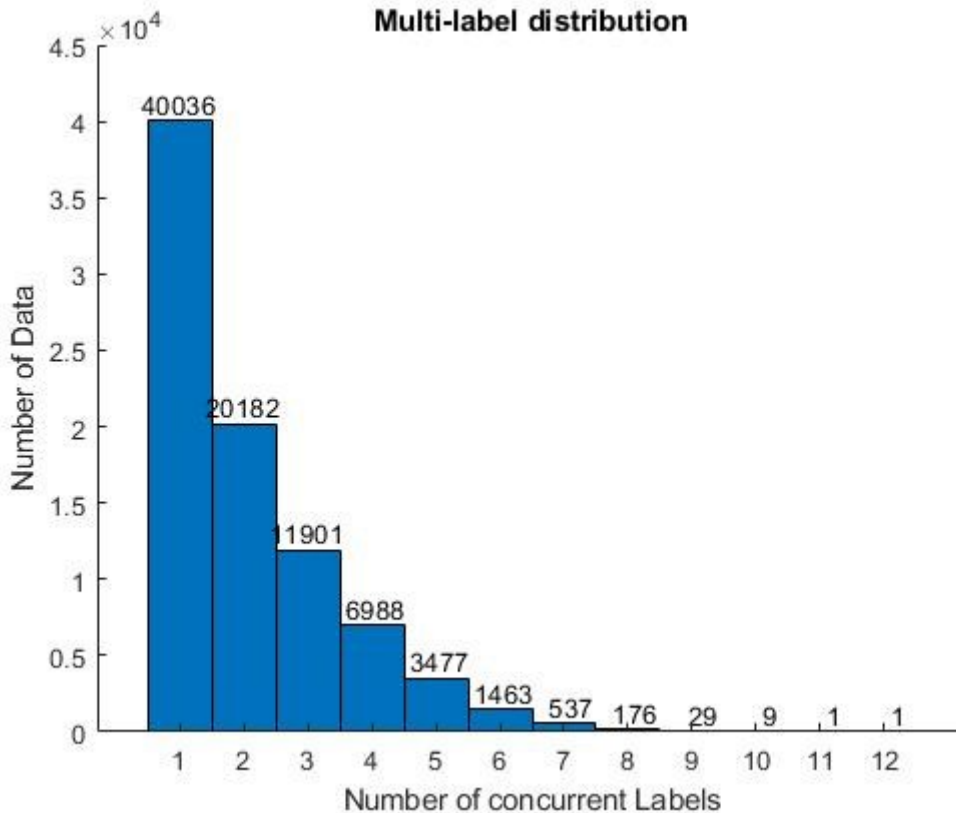In Figure 23 it is shown the distribution of data in function of the number of labels.

FIGURE 23 NUMBER OF DATA AS A FUNCTION OF THEIR NUMBER OF LABELS

The most represented class was '426783006' (ECG: sinus rhythm) with 28967 files (34,2%). The least represented classes were '251168009' (Supraventricular bigeminy), '413444003' (Acute myocardial ischemia), and '74615001' (Tachycardia-bradycardia) with 1 file each. The mean number of elements for each class was 1395, with a standard deviation of 3555.

The statistical distribution of labels shows the number of signals which present more than one class. In order to simplify the problem and avoid solving a multi-class multi-label classification, if the data present more than one label in their one-hot encoded vector, only the one corresponding with the most popular class will be kept, and the others will be discarded. This procedure will automatically prune the low represented class while at the same time increase the diversity of the remaining class, which is the most common pathology diagnosed for that signal.

After the re-labelling, the total number of unique classes dropped to 87. The 39 lost classes were not present in single-label data or statistically predominant in multi-label ones.

The most represented class was still '426783006' (ECG: sinus rhythm) with 28967 files (34,2%). The least represent classes were 14 and had only 1 element each. The mean number of elements for each class was 975, with a standard deviation of 3828.

The statical visualization of the distribution of classes, which, as a reminder, are pathologies, shows if they are balanced or not. In case of great classes' unbalance, the training process will be biased toward the most predominant classes and will likely ignore the others, leading to an overfitting model; this phenomenon is referred to as data leakage and can be solved either by augmenting the under-represented classes or pruning the over-represented ones [24]. Both instances are inconvenient: in the first case, data augmentation will invalidate the real distribution of the data, and in the second case, pruning will lower diversity among the sample in addition to reduce the training pool.

In Table 2 are reported the non-insignificant classes, meaning the feature at least more than 100 elements.

SNOMED-CT code	Disorder/Finding	Number of elements	Presence (%)
426783006	Sinus Rhythm	28967	34,2%
426177001	Sinus Bradycardia	18514	21,2%
427084000	Sinus Tachycardia	7593	9,0%
164934002	T Wave Abnormal	6621	7,8%
164890007	Atrial Flutter	6145	7,2%
164889003	Atrial Fibrillation	2892	3,4%
59118001	RBBB	1836	2,2%
427393009	Sinus Arrhythmia	1660	2,0%
39732003	Left Axis Deviation	1342	1,6%
164865005	Myocardial Infarction	984	1,2%
270492004	1st Degree AV Block	978	1,2%
429622005	ST Depression	860	1,0%
284470004	Premature Atrial Contraction	748	0,9%
164873001	Left Ventricle Hypertrophy	694	0,8%
428750005	Nonspecific ST-T Abnormality	689	0,8%
164884008	Ventricular Ectopics	664	0,8%
10370003	Rhythm From Artificial Pacing	584	0,7%
426761007	Supraventricular Tachycardia	416	0,5%
55827005	Left Ventricular Hypertrophy	388	0,5%
164909002	LBBB	245	0,3%
111975006	Prolonged QT Interval	233	0,3%
59931005	Inverted T Wave	227	0,3%
164931005	ST Elevation	198	0,2%
55930002	ST Segment Changes	186	0,2%
67741000119109	Left Atrial Enlargement	144	0,2%
106068003	Atrial Rhythm	111	0,1%

TABLE 2 THE 26 MOST REPRESENTED CLASSES AMONG THE DATASETS AND THEIR NUMBER OF ELEMENTS

The combination of available datasets proved to be extremely unbalanced among the different classes and not servable for DL purposes without intervention. The simplification of considering a multi-class single-label problem was necessary to lower the overall complexity but it is worth considering almost the majority of the data presented a single class, anyway, as shown in Figure 23. Due to computational power limitations, the number of classes was also reduced to four. In this way, though it was possible to consider balanced classes, essential for unbiased learning.

The four considered classes were: Left Ventricle Hypertrophy ('164873001'), Ventricular Ectopics ('164884008'), Premature Atrial Contraction ('284470004'), and Nonspecific ST-T Abnormality ('428750005'). The total number of elements was 2546.

5.2.2. PREPROCESSING

Once the data are correctly re-labelled, it is possible to start the pre-processing. Using MATLAB Wavelet Toolbox features, it is possible to design a CWT filter-bank just by setting the signal sampling frequency and length. This will save a lot of computational power during the generation of scalograms, because, by precomputing the filter-bank once, it is then possible to apply it to all the signals with the corresponding sampling frequency. Since the dataset comprehends data sampled at different frequencies, it is necessary to precompute one filter bank for each sampling frequency. The signals are also windowed with respect to the shortest in the duration of their group with the same sample frequency.

The mother-wavelet used in designing the CWT filter-bank is the Morlet (Gabor) or bump wavelet, with symmetry parameter γ equal to 3 and time-bandwidth product P^2 equal to 60, as by default [44].

Applying the precomputed CWT filter-bank to the signal, it is possible to obtain one scalogram for each one of the 12 leads. Those are saved offline in ".tiff" format as a stack 12 of greyscale images of 128x128 pixels with "jpeg" compression (Figure 25). In this way, it is possible to efficiently organize the new dataset in folders named as the corresponding labels, where each file actually contains the CWT of all the 12 leads of the original signal as their scalograms.
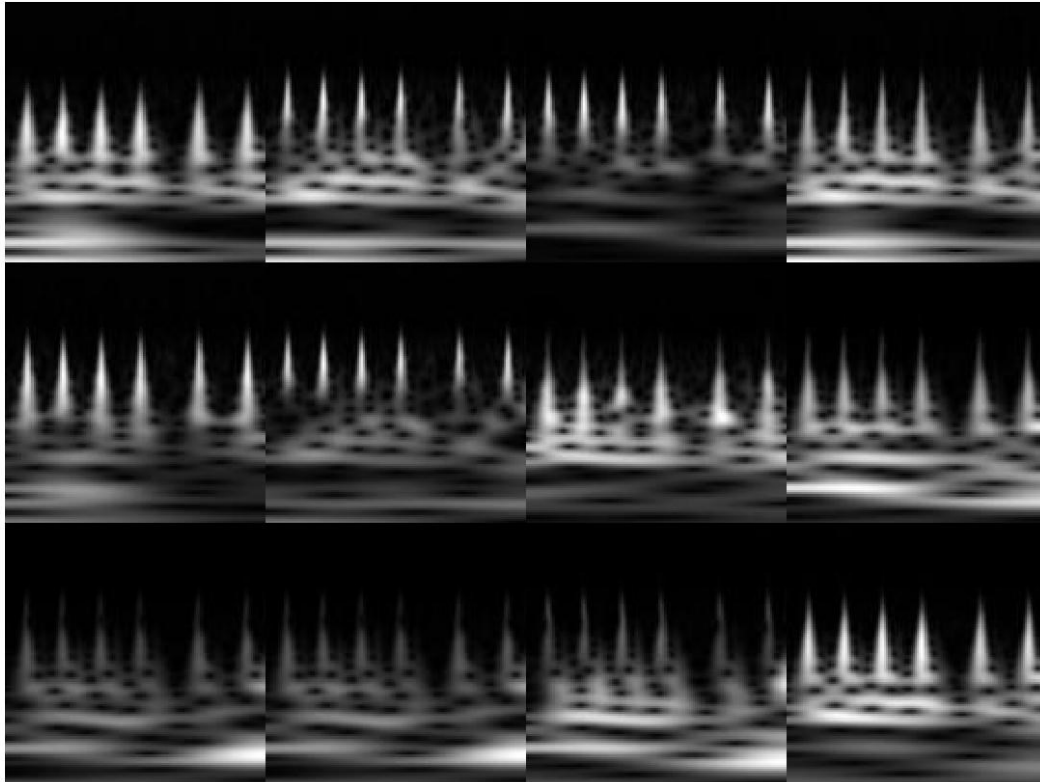
5.2.3. MODEL ARCHITECTURE

The architecture of the CNN and the successive training and cross-validation were performed using the Keras and TensorFlow API in Python.

The CNN designed for this task must be accepted as input the information contained in 12 scalograms. Usually, architectures trained for image classification have one channel (greyscale) or three channels (RGB) inputs. It is possible to consider each lead's spectrogram as one channel of a hyperspectral image, thus designing an architecture that accepts images of size 128x128 pixels x12 channels. This will not impact the function of the CNN, affecting only the number of filters, which will be four times more compared to RGB images.

The end-to-end CNN is quite simple and is shown in Figure 25.

1. First Convolutional Layer with 32 filters, kernel size of 3x3, unitary stride, and no padding
2. ReLU Activation

3. Max Pooling Layer with a pool size of 2x2 and stride equal to 2.

4. Second Convolutional Layer with 64 filters, kernel size of 3x3, unitary stride, and no padding

5. ReLU Activation

6. Second Max Pooling Layer with a pool size of 2x2 and stride equal to 1

7. Drop-out layer with a 50% probability

8. Fully connected layer with 1000 neurons in the output

9. ReLU Activation

10. Drop-out layer with a 70% probability

11. Fully connected layer with output neurons equal to the number of classes

12. SoftMax activation
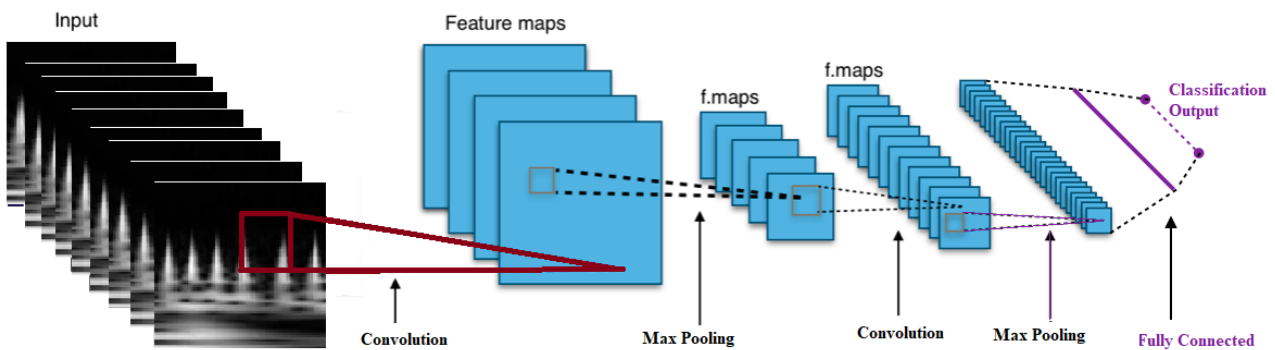
FIGURE 25 END-TO-END CNN ARCHITECTURE

5.2.4. EXPERIMENTS

The 12-lead model was successfully built and set to run its first instance with a batch size of 12 elements for 50 epochs.

The dataset was randomly split into training and validation subsets with a ratio of 4 to 1.

The 12-lead model was trained and validated during 10 splits of stratified shuffle split validation.

Once done with the 12-lead DL model, it is possible to start training a new CNN with a different lead-set. This time though, it is possible to import the initial parameters from the previously trained model instead of beginning from nothing. This practice is referred to as fine-tuning in the process of transfer learning [24].

In this way, it would be possible to train the 6-lead, the 4-lead, the 3-lead, and the 2-lead models using the "knowledge" obtained from previous training just by changing the first layer.

Nevertheless, the input dimension can't be directly changed without nullifying the effect of previous training. This was solved in a manner reminiscent of how CNN trained to classify RGB images can be used for greyscale images: the single-channel is repeated three times. In this case, the original input had 12 channels, one for each lead, and all the subsets present a number of channels that can be repeated for two, three, four or six times in order to match the correct size.

In this way, it was possible to train four different models reducing the effective input size each time in succession.

This cascade fashion is bound exclusively to training. When all the models have been trained, to evaluate a signal for testing purposes, one needs just to select the model with the correct number of leads and does not require to run them all.

The four models were then validated using stratified shuffle splitting, a combination of shuffle split and stratified K-fold, which is more suitable for classification problems with unbalanced numbers of classes. The folds are made by preserving the percentage of samples for each class.

One last model was trained from scratch using a 2-channel input for the 2-lead system. The batch size, number of epochs, training-validation ratio, and number of splits remained the same also for this run.

5.2.5. STATISTICS

Precision, sensitivity, and F_1 score were calculated for each class.

The precision (or positive predictive value) defines the ability of the classifier not to label as positive a sample that is negative and is calculated as:

$$Precision = \frac{number\ of\ true\ positives}{number\ of\ true\ positive + number\ of\ false\ potisitives} \tag{29}$$

The sensitivity (or recall in ML, or true positive rate) defines the ability of the classifier to find all the positive samples and is calculated as:

$$Sensitivity = \frac{number\ of\ true\ positives}{number\ of\ true\ positive + number\ of\ false\ negatives} \qquad (30)$$

F-scores are a measure of a test's accuracy; F_1 score, in particular, is the harmonic mean of precision and sensitivity:

$$F_1 = \frac{2}{precision^{-1} + sensitivity^{-1}} \qquad (31)$$

The specificity defines how many negative selected elements are truly negative and is calculated as:

$$Specificity = \frac{number\ of\ true\ negative}{number\ of\ true\ negatives + number\ of\ false\ positives} \qquad (32)$$

The Receiver Operating Characteristic (ROC) curve, which shows the diagnostic ability of the model as its discrimination threshold is varied, gives an estimate of the sensitivity and specificity of the model, as well as the Area Under the Curve (AUC); the latter expresses the probability that the model ranks a randomly chosen positive higher than a randomly chosen negative. Considering true positive rate (TPR, sensitivity) as a function of false-positive rate (FPR, 1-specificity), AUC is the following integral:

$$AUC = \int_{x=0}^{1} TPR\big(FPR^{-1}(x)\big)dx \qquad (33)$$

The AUC can be used to compare more reliably their performances. It may be useful to remember that the coherence of this statistic, although questioned due to its noisy characteristic, has been vindicated as a measure of aggregated classification performance in terms of a uniform rate distribution [45].

5.3. RESULTS

The last iteration of the 12-lead model reached an accuracy of 78.43% with a loss of 0.48. The average accuracy, calculated over the 10 splits, was (74.84±3,79)% while the average loss was 0.55±0.04.

Precision, Sensitivity and F_1 score for each class are reported in Table 3.

Class	Precision	Sensitivity	F_1 score	Support
1. **Left Ventricle Hypertrophy**	0.83	0.76	0.79	139
2. **Ventricular Ectopics**	0.74	0.90	0.81	124
3. **Premature Atrial Contraction**	0.84	0.61	0.71	109
4. **Nonspecific ST-T Abnormality**	0.76	0.85	0.80	138

TABLE 3 PRECISION, SENSITIVITY AND F_1 FOR THE 12-LEAD MODEL. SUPPORT REFERS TO THE TOTAL NUMBER OF CORRECT TARGET VALUES (GROUND TRUTH)

The confusion matrix for the 6-lead model is reported in Figure 26. The mean ROC curve and AUC are reported in Figure 27 instead.
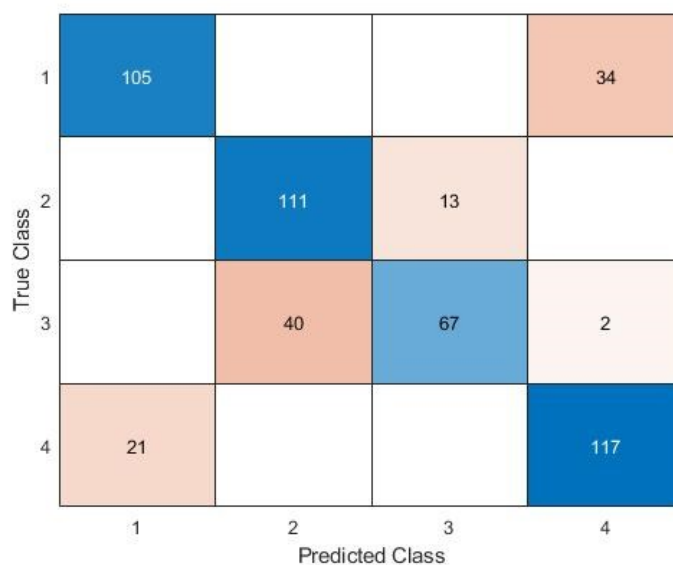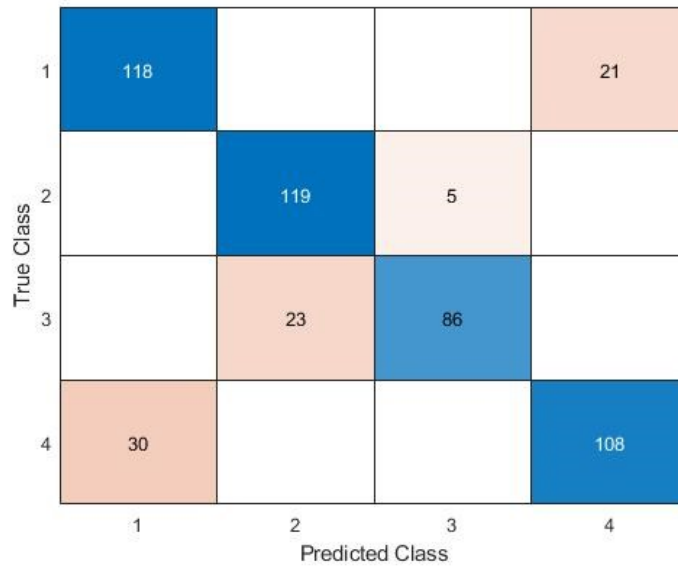

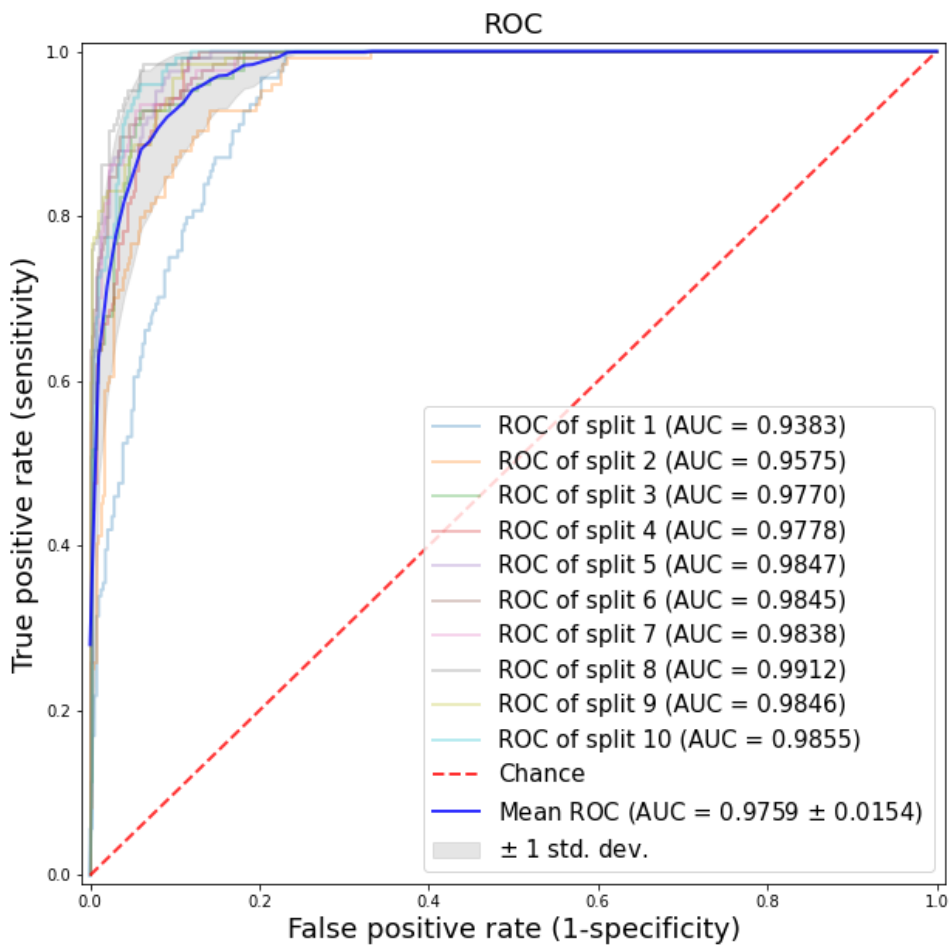FIGURE 26 CONFUSION MATRIX OF THE 12-LEAD MODEL
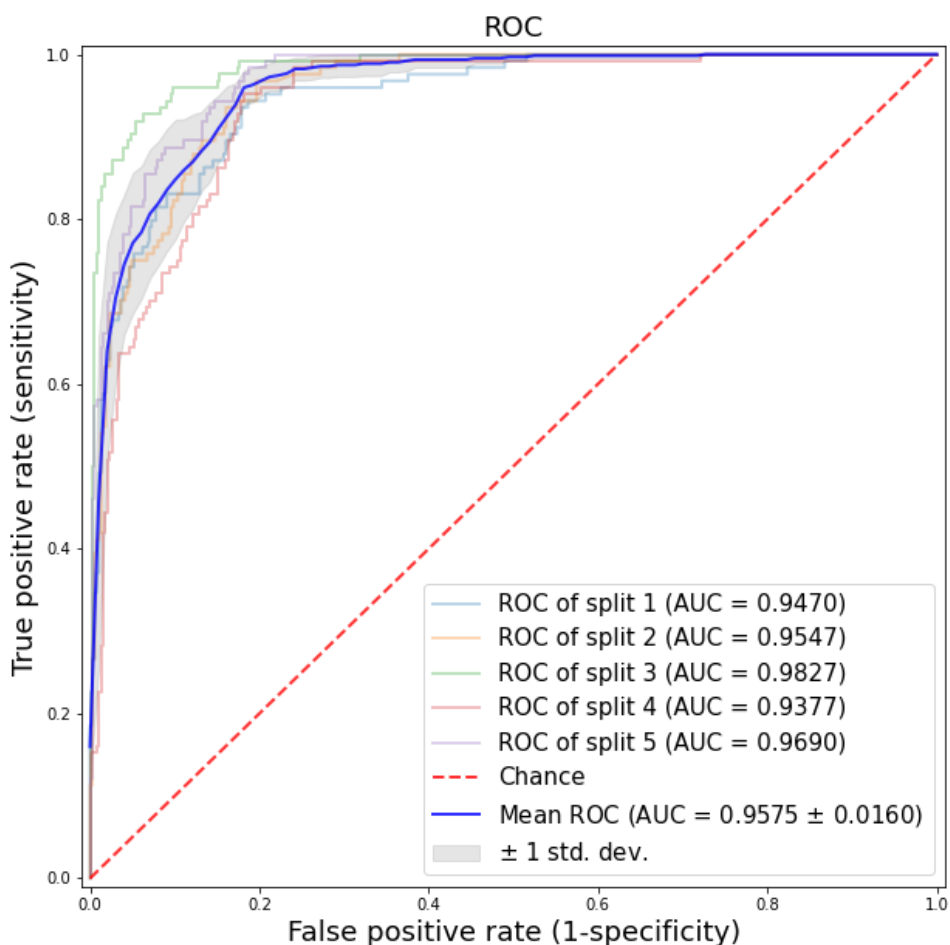
FIGURE 27 MEAN ROC CURVE AND AUC OF THE 12-LEAD MODEL

The last iteration of the 6-lead model reached an accuracy of 84.50% with a loss of 0.41. The average accuracy was (78.92±5.09)% while the average loss was 0.45±0.08.

Precision, Sensitivity and F_1 score for each class are reported in Table 4.

Class	Precision	Sensitivity	F_1 score	Support
1. **Left Ventricle Hypertrophy**	0.80	0.85	0.82	139
2. **Ventricular Ectopics**	0.84	0.96	0.89	124
3. **Premature Atrial Contraction**	0.95	0.79	0.86	109
4. **Nonspecific ST-T Abnormality**	0.84	0.78	0.81	138

TABLE 4 PRECISION, SENSITIVITY AND F_1 FOR THE 6-LEAD MODEL. SUPPORT REFERS TO THE TOTAL NUMBER OF CORRECT TARGET VALUES (GROUND TRUTH)

The confusion matrix for the 6-lead model is reported in Figure 28. The mean ROC curve and AUC are reported in Figure 29 instead.
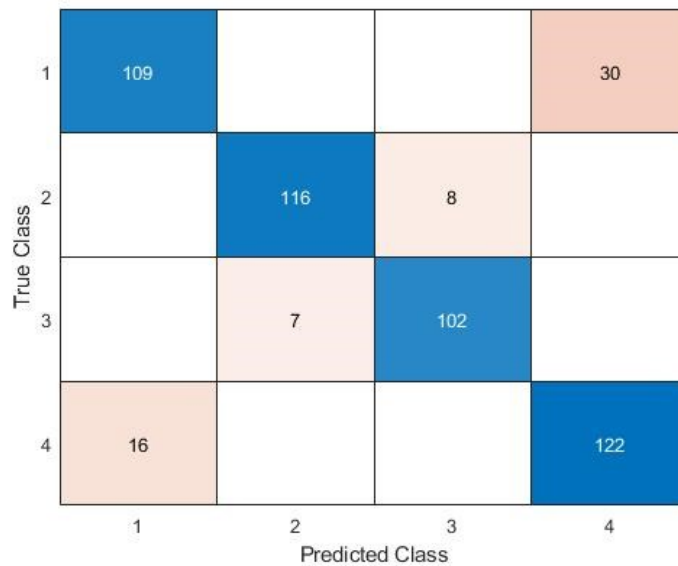

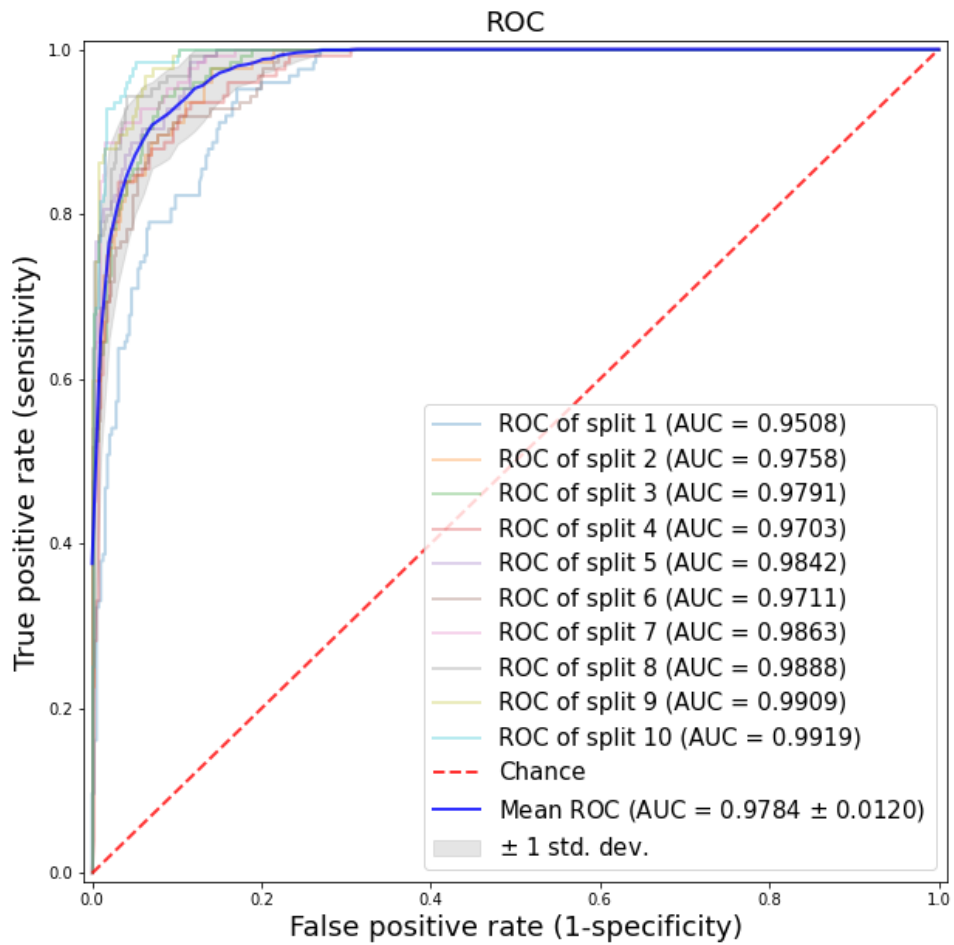FIGURE 28 CONFUSION MATRIX OF THE 6-LEAD MODEL

FIGURE 29 MEAN ROC CURVE AND AUC OF THE 6-LEAD MODEL

The last iteration of the 4-lead model reached an accuracy of 88.03% with a loss of 0.33. The average accuracy was (79.47±5.20)% while the average loss was 0.44±0.08.

Precision, Sensitivity and F_1 score for each class are reported in Table 5.

Class	Precision	Sensitivity	F₁ score	Support
1. **Left Ventricle Hypertrophy**	0.87	0.78	0.83	139
2. **Ventricular Ectopics**	0.94	0.94	0.94	124
3. **Premature Atrial Contraction**	0.93	0.94	0.93	109
4. **Nonspecific ST-T Abnormality**	0.80	0.88	0.84	138

TABLE 5 PRECISION, SENSITIVITY AND F_1 FOR THE 4-LEAD MODEL. SUPPORT REFERS TO THE TOTAL NUMBER OF CORRECT TARGET VALUES (GROUND TRUTH)

The confusion matrix for the 4-lead model is reported in Figure 30. The mean ROC curve and AUC are reported in Figure 31 instead.
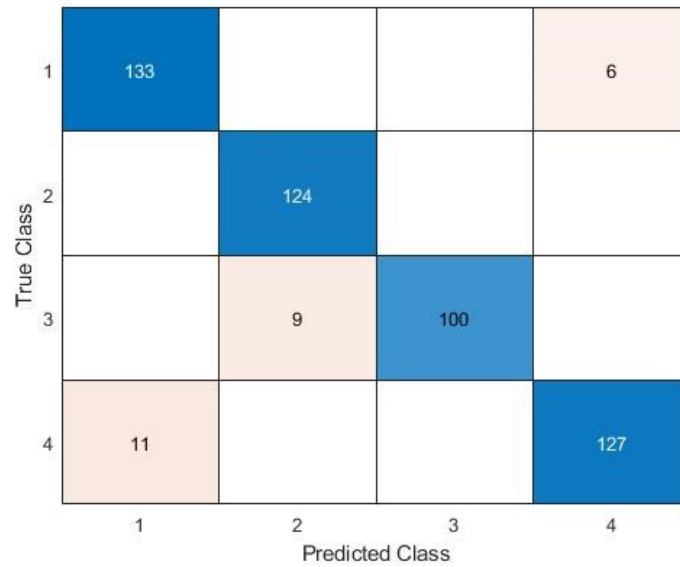

FIGURE 30 CONFUSION MATRIX OF THE 4-LEAD MODEL

The last iteration of the 3-lead model reached an accuracy of 94.90% with a loss of 0.16. The average accuracy was (84.57±9.37)% while the average loss was 0.34±0.15.

Precision, Sensitivity and F_1 score for each class are reported in Table 6.

Class	Precision	Sensitivity	F_1 score	Support
1. **Left Ventricle Hypertrophy**	0.92	0.96	0.94	139
2. **Ventricular Ectopics**	0.93	1.00	0.96	124
3. **Premature Atrial Contraction**	1.00	0.92	0.96	109
4. **Nonspecific ST-T Abnormality**	0.95	0.92	0.94	138

TABLE 6 PRECISION, SENSITIVITY AND F_1 FOR THE 3-LEAD MODEL. SUPPORT REFERS TO THE TOTAL NUMBER OF CORRECT TARGET VALUES (GROUND TRUTH)

The confusion matrix for the 3-lead model is reported in Figure 32. The mean ROC curve and AUC are reported in Figure 33 instead.

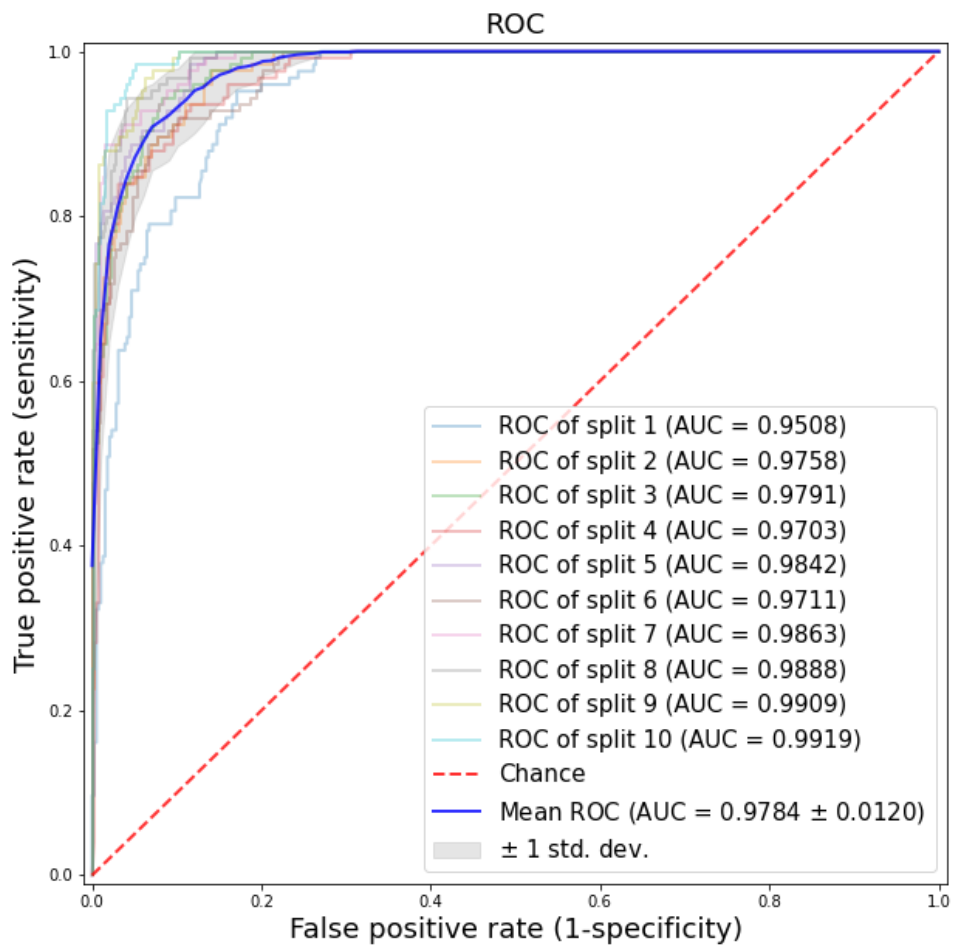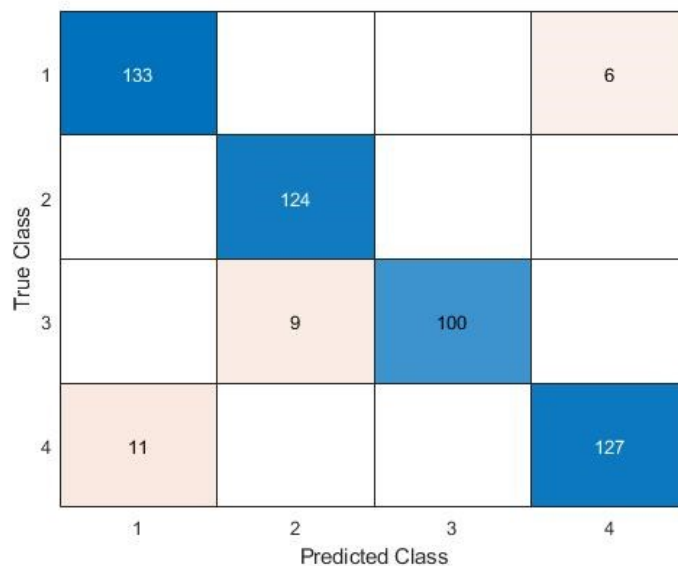**FIGURE 32 CONFUSION MATRIX OF THE 3-LEAD MODEL**

FIGURE 33 MEAN ROC CURVE AND AUC OF THE 3-LEAD MODEL

The last iteration of the 2-lead model reached an accuracy of 87.25% with a loss of 0.33. The average accuracy was (75.01±8.30)% while the average loss was 0.49±0.11.

Precision, Sensitivity and F_1 score for each class are reported in Table 7.

Class	Precision	Sensitivity	F₁ score	Support
1. **Left Ventricle Hypertrophy**	0.81	0.95	0.87	139
2. **Ventricular Ectopics**	0.83	0.98	0.90	124
3. **Premature Atrial Contraction**	0.98	0.77	0.86	109
4. **Nonspecific ST-T Abnormality**	0.94	0.78	0.85	138

TABLE 7 PRECISION, SENSITIVITY AND F₁ FOR THE 2-LEAD MODEL. SUPPORT REFERS TO THE TOTAL NUMBER OF CORRECT TARGET VALUES (GROUND TRUTH)

The confusion matrix for the 2-lead model is reported in Figure 34. The mean ROC curve and AUC are reported in Figure 35 instead.

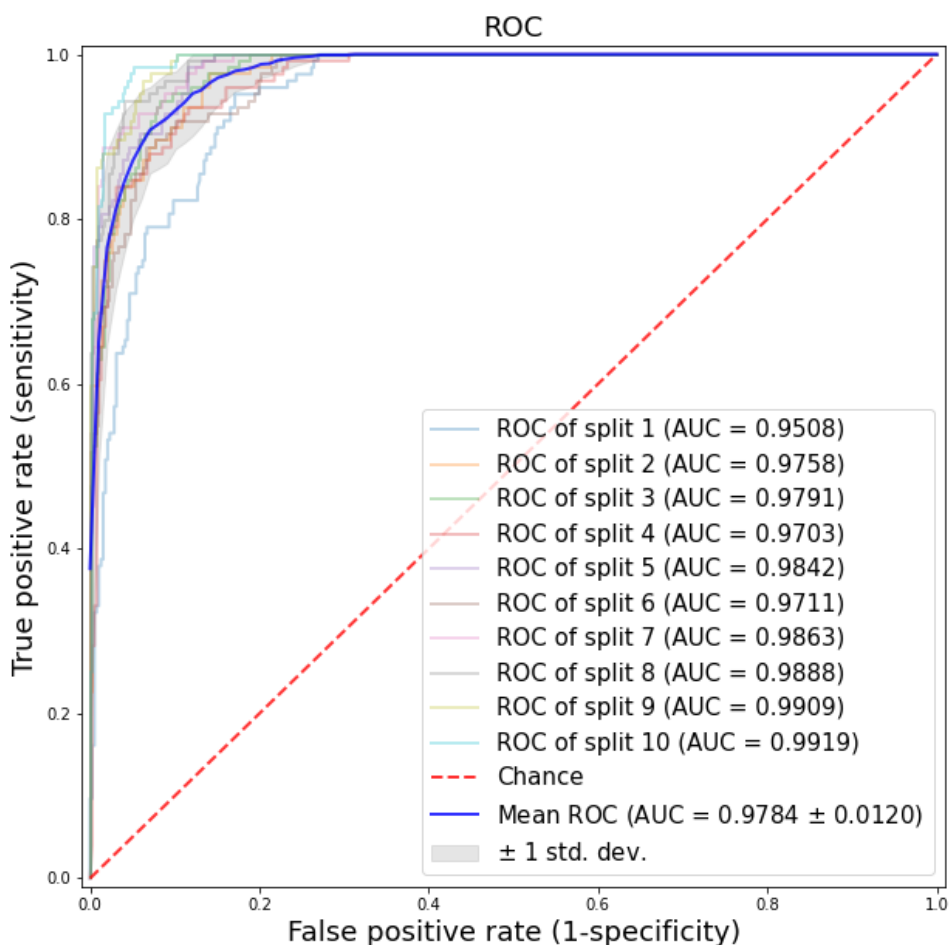FIGURE 34 CONFUSION MATRIX OF THE 2-LEAD MODEL

FIGURE 35 MEAN ROC CURVE AND AUC OF THE 2-LEAD MODEL

The last iteration of the 2-lead model with 2 channels reached an accuracy of 92.94% with a loss of 0.21. The average accuracy was (80.61±6.60)% while the average loss was 0.42±0.11.

Precision, Sensitivity and F_1 score for each class are reported in Table 8.

Class	Precision	Sensitivity	F_1 score	Support
1. **Left Ventricle Hypertrophy**	0.81	0.95	0.87	139
2. **Ventricular Ectopics**	0.83	0.98	0.90	124
3. **Premature Atrial Contraction**	0.98	0.77	0.86	109
4. **Nonspecific ST-T Abnormality**	0.94	0.78	0.85	138

TABLE 8 PRECISION, SENSITIVITY AND F_1 FOR THE 2-LEAD "FROM SCRATCH" MODEL. SUPPORT REFERS TO THE TOTAL NUMBER OF CORRECT TARGET VALUES (GROUND TRUTH)

The confusion matrix for the 2-lead "from scratch" model is reported in Figure 36. The mean ROC curve and AUC are reported in Figure 37 instead.
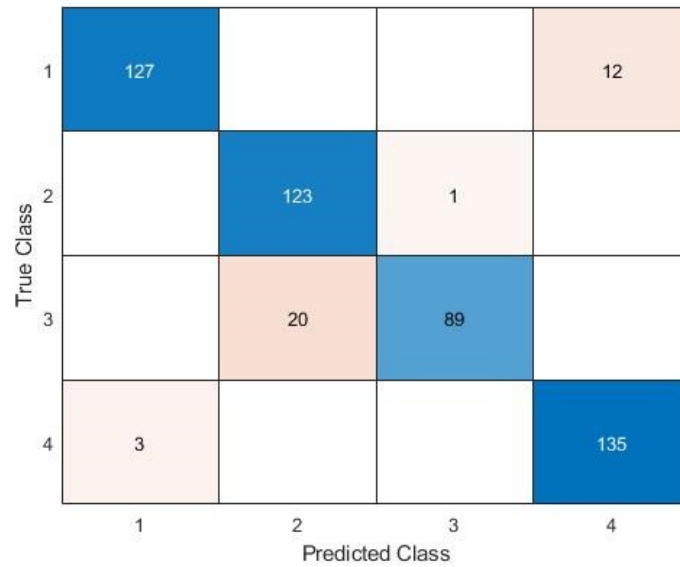

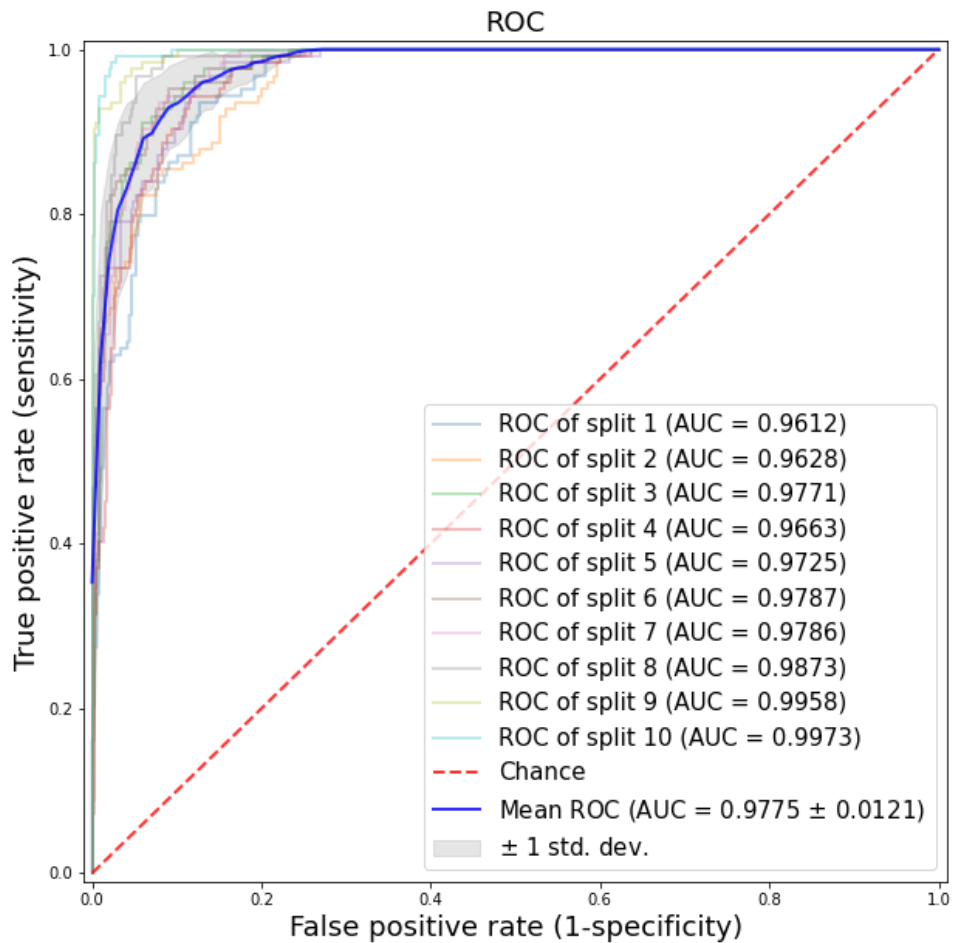**FIGURE 36 CONFUSION MATRIX OF THE 2-LEAD "FROM SCRATCH" MODEL**

FIGURE 37 MEAN ROC CURVE AND AUC OF THE 2-LEAD "FROM SCRATCH" MODEL

5.4. DISCUSSION

The 12-lead model was the first to be trained and achieved up to 78% of accuracy. This result by itself is sufficient to show the raw power of this approach. Despite the model being trained from scratch, with no optimization on the hyper-parameters of the network, such as convolution stride or dropout chance, the model was capable to perform meaning classification at the first attempt. The ROC space shows a well-shaped curve leaning toward the perfect classification (no false negatives, no false positives), hence has good sensitivity and specificity.

The 6-lead, 4-lead, 3-lead, and 2-lead models were trained in succession starting from the pre-initialized weights of the 12-lead model. As shown in the result section, the accuracy of the model increased by almost 6%, up to 84% for the 6-lead model, then slightly up to 88% for the 4-lead model, reaching 94% for the 3-lead model, before dropping down to 87% for the 2-lead model. The average accuracy did not vary significantly among the four models with a reduced number of leads. The AUC remained constant around 97%, meaning the models have an equal aggregated classification performance

The boost in accuracy, as well as in sensitivity and specificity, that reached its best in the 3-lead model, is partially due to the overfitting of the model. For this reason, a fifth model was built from scratch, featuring an input with only 2 channels to classify 2-lead ECG. This model achieved very good results, superimposable to those obtained with the fine-tuning approach.

The advantage of using transfer learning lays in having pre-initialized CNN weights which can be very convenient to boost training especially for deep networks with a lot of parameters; if the model is particularly light-weighted, it can be convenient to act on the input shape instead, reducing the number of channels, and thus saving computational power upstream.

The good results obtained in the models working with a reduced number of leads may indicate that some of the information contained in the ECG are redundant, and performing classification with less inputs is possible, at least for certain pathologies.

CONCLUSION

In this thesis, a novel approach for ECG classification using an adjustable number of leads was presented. Thanks to the properties of the CWT, it was possible to obtain the scalograms of the signal, which is an adequate input for a simple end-to-end CNN. The proposed architecture features common DL elements and is particularly light-weighted in terms of computational requirements. It was different from canonical implementation in the number of channels provided at the input layer, which was equal to the maximum number of available leads, twelve. This kind of input, although not very popular in other fields of DL application, is perfectly supported by the mathematical tools of CNN and is often referred as to hyperspectral input. Overall, the conjunction of WT and CNN was a success for the classification of the ECG, and it was even possible to use it with a reduced number of leads. This approach presents the same drawbacks of any other DL model, such as the high computational cost, the generally low interpretability and generalizability, in addition to a burdensome preprocessing, but provides great benefits such as high accuracy, sensitivity, and specificity, as well as being easy to use and tune.

The future holds great promises for the advancement of alike models. For instance, increasing the available computational power could be possible to classify more classes. Due to the very nature of the data, the classes will always be imbalanced because some heart-related disorders are much more common than others but there exist tools to solve this problem artificially, like data augmentation. In this way, though the dataset would no longer represent the real-world distribution of such disorders. Another improvement could be solving the full multi-label. There are currently two possible solutions: problem transformation and adapted algorithms. The first one still does not address the full multi-label solutions but instead simplifies it in multiple single-label problems, solves them individually and merges their solutions; some practical examples are binary relevance, classifier chains, and label powersets. The adapted algorithms on the other hand are designed starting from existing models explicitly to solve the multi-label problem; one example is the multi-label version of the k-nearest neighbours (kNN) algorithm. Last but not least, the design of CNN can be addressed, either acting on the hyper-parameters on the depth of the network.

Once trained, the proposed model can perform ECG classification in a very short time and, assuming there is internet access, the computation can be performed via cloud computing, making the number of applications of this technique countless.

BIBLIOGRAPHY

[1] W. Geneva., "Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019. Geneva, World Health Organisation.," *World Heal. Organ.*, 2020.

[2] J. G. Betts *et al.*, *Anatomy & Physiology -OpenStax*. 2013.

[3] D. U. Silverthorn, "Human Physiology: An Integrated Approach, Global Edition," *Pearson*, 2015.

[4] E. Falk and V. Fuster, "Hurst The Heart," *Hurst's The Heart, 14e*. 2017.

[5] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. 2012.

[6] R. E. Klabunde, *CARDIOVASCULAR PHYSIOLOGY CONCEPTS - Second edition*. 2012.

[7] R. Phillips and M. Feeney, "The Cardiac Rhythms," *Dimens. Crit. Care Nurs.*, 1991, doi: 10.1097/00003465-199109000-00007.

[8] S. Scheidt, "Basic electrocardiography: leads, axes, arrhythmias.," *Clin. Symp.*, 1983.

[9] J. E. Hall and M. E. Hall, "Guyton and Hall: Textbook of Medical Physiology 14th Edition," *Elsevier*, 2021.

[10] M. Papouchado, P. R. Walker, M. A. James, and L. M. Clarke, "Fundamental differences between the standard 12-lead electrocardiograph and the modified (mason-likar) exercise lead system," *Eur. Heart J.*, 1987, doi: 10.1093/eurheartj/8.7.725.

[11] T. Reichlin *et al.*, "Advanced ECG in 2016: is there more than just a tracing?," *Swiss medical weekly*. 2016, doi: 10.4414/smw.2016.14303.

[12] H. Doose and W. K. Baier, "Theta rhythms in the EEG: A genetic trait in childhood epilepsy," *Brain and Development*. 1988, doi: 10.1016/S0387-7604(88)80091-3.

[13] S. J. Schiff, A. Aldroubi, M. Unser, and S. Sato, "Fast wavelet transformation of EEG," *Electroencephalogr. Clin. Neurophysiol.*, 1994, doi: 10.1016/0013-4694(94)90165-1.

[14] I. Daubechies and C. Heil, "Ten Lectures on Wavelets," *Comput. Phys.*, 1992, doi:

10.1063/1.4823127.

[15]   L. Aguiar-Conraria and M. J. Soares, "The continuous wavelet transform: Moving beyond uni- and bivariate analysis," *J. Econ. Surv.*, 2014, doi: 10.1111/joes.12012.

[16]   M. Farge, "Wavelet Transforms and their Applications to Turbulence," *Annu. Rev. Fluid Mech.*, 1992, doi: 10.1146/annurev.fl.24.010192.002143.

[17]   D. S. Stoffer and M. Holschneider, "Wavelets: An Analysis Tool.," *J. Am. Stat. Assoc.*, 1997, doi: 10.2307/2965590.

[18]   P. Goupillaud, A. Grossmann, and J. Morlet, "Cycle-octave and related transforms in seismic signal analysis," *Geoexploration*, 1984, doi: 10.1016/0016-7142(84)90025-5.

[19]   E. Foufoula-Georgiou and P. Kumar, "Wavelet Analysis in Geophysics: An Introduction," in *Wavelet Analysis and Its Applications*, 1994.

[20]   A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift fur Medizinische Physik*. 2019, doi: 10.1016/j.zemedi.2018.11.002.

[21]   C. Chen *et al.*, "Deep Learning for Cardiac Image Segmentation: A Review," *Front. Cardiovasc. Med.*, 2020, doi: 10.3389/fcvm.2020.00025.

[22]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, 2017, doi: 10.1145/3065386.

[23]   P. H. Burke, Brian, "Top Strategic Technology Trends for 2021- Gartner," *Gartner*, 2020.

[24]   D. Learning, "Deep Learning - Goodfellow," *Nature*, 2016.

[25]   D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," 2015.

[26]   Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, 2016, doi: 10.1016/j.neucom.2015.09.116.

[27]   O. Abdel-Hamid, L. Deng, and D. Yu, "Exploring convolutional neural network structures and optimization techniques for speech recognition," 2013, doi: 10.21437/interspeech.2013-744.

[28]   S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," 2018, doi: 10.1109/ICEngTechnol.2017.8308186.

[29]   H. Krizhevsky, A., Sutskever, I., "'Imagenet classification with deep convolutional neural network', in Advances in Neural Information Processing Systems, p. 1097-1105.," *Elsevier Ltd*, 2012.

[30]   N. Milosevic, *Introduction to Convolutional Neural Networks*. 2020.

[31]   E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, doi: 10.1109/TPAMI.2016.2572683.

[32]   C. Szegedy *et al.*, "Going deeper with convolutions," 2015, doi: 10.1109/CVPR.2015.7298594.

[33]   K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016, doi: 10.1109/CVPR.2016.90.

[34]   K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[35]   R. P. K. Poudel, P. Lamata, and G. Montana, "Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation," 2017, doi: 10.1007/978-3-319-52280-7_8.

[36]   B. J. Drew, M. M. Pelter, D. E. Brodnick, A. V. Yadav, D. Dempel, and M. G. Adams, "Comparison of a new reduced lead set ECG with the standard ECG for diagnosing cardiac arrhythmias and myocardial ischemia," *J. Electrocardiol.*, 2002, doi: 10.1054/jelc.2002.37150.

[37]   M. Green, M. Ohlsson, J. Lundager Forberg, J. Björk, L. Edenbrandt, and U. Ekelund, "Best leads in the standard electrocardiogram for the emergency detection of acute coronary syndrome," *J. Electrocardiol.*, 2007, doi: 10.1016/j.jelectrocard.2006.12.011.

[38]   H. R. Aldrich *et al.*, "Identification of the optimal electrocardiographic leads for detecting acute epicardial injury in acute myocardial infarction," *Am. J. Cardiol.*, 1987, doi: 10.1016/S0002-9149(87)80062-0.

[39]   F. Liu *et al.*, "An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection," *J. Med. Imaging Heal. Informatics*, 2018, doi: 10.1166/jmihi.2018.2442.

[40]   A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," *Circulation*, 2000, doi:

10.1161/01.cir.101.23.e215.

[41]   R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet," *Biomed. Tech.*, 1995, doi: 10.1515/bmte.1995.40.s1.317.

[42]   E. A. Perez Alday *et al.*, "Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology Challenge 2020," *Physiol. Meas.*, 2020, doi: 10.1088/1361-6579/abc960.

[43]   J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, and C. Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," *Sci. Data*, 2020, doi: 10.1038/s41597-020-0386-x.

[44]   Z. Qibin and Z. Liqing, "ECG feature extraction and classification using wavelet transform and support vector machines," 2005, doi: 10.1109/icnnb.2005.1614807.

[45]   P. Flach, J. Hernández-Orallo, and C. Ferri, "A coherent interpretation of AUC as a measure of aggregated classification performance," 2011.

ACKNOWLEDGEMENTS