Deep learning and unsupervised machine learning for the quantification and interpretation of electrocardiographic signals

Guillermo Jiménez Pérez

TESI DOCTORAL UPF / 2021

Director of the thesis

Prof. Oscar Camara
Department of Information and Communication Technologies

Directors

Oscar Camara	Full Professor Universitat Pompeu Fabra Barcelona, Spain

Review Committee:

Blanca Rodríguez	Full Professor University of Oxford Oxford, United Kingdom
Pablo Laguna	Full Professor Universidad de Zaragoza Zaragoza, Spain
Gemma Piella	Full Professor Universitat Pompeu Fabra Barcelona, Spain

This work was carried out in the *Sensing in Physiology and Biomedicine (PhySense)* research group, at the Departament de Tecnologies de la Informació i les Comunicacions at the Universitat Pompeu Fabra, Barcelona, Spain. This thesis was supported by a scholarship from the Secretariat for Universities and Research of the Ministry of Business and Knowledge of the Government of Catalonia (2017 FI_B 01008) and a scholarship of the Department of Information and Communication Technologies at Universitat Pompeu Fabra (DTIC-UPF).

Resumen

Las señales electrocardiográficas, ya sea adquiridas en la piel del paciente (electrocardiogamas de superficie, ECG) o de forma invasiva mediante cateterismo (electrocardiogramas intracavitarios, iECG) ayudan a explorar la condición y función cardíacas del paciente, dada su capacidad para representar la actividad eléctrica del corazón. Sin embargo, la interpretación de las señales de ECG e iECG es una tarea difícil que requiere años de experiencia, con criterios diagnósticos complejos para personal clínico no especialista, que en muchos casos deben ser interpretados durante situaciones de gran estrés o carga de trabajo como en la unidad de cuidados intensivos, o durante procedimientos de ablación por radiofrecuencia (ARF) donde el cardiólogo tiene que interpretar cientos o miles de señales individuales. Desde el punto de vista computacional, el desarrollo de herramientas de alto rendimiento mediante técnicas de análisis basadas en datos adolece de la falta de bases de datos anotadas a gran escala y de la naturaleza de "caja negra" que están asociados con los algoritmos considerados estado del arte en la actualidad. Esta tesis trata sobre el entrenamiento de algoritmos de aprendizaje automático que ayuden al personal clínico en la interpretación automática de ECG e iECG. Esta tesis tiene cuatro contribuciones principales. En primer lugar, se ha desarrollado una herramienta de delineación del ECG para la predicción de los inicios y finales de las principales ondas cardíacas (ondas P, QRS y T) en registros compuestos de cualquier configuración de derivaciones. En segundo lugar, se ha desarrollado un algoritmo de generación de datos sintéticos que es capaz de paliar el impacto del reducido tamaño de las bases de datos existentes para el desarrollo de algoritmos de delineación. En tercer lugar, la metodología de análisis de datos de ECG se aplicó a datos similares, en registros electrocardiográficos intracavitarios, con el mismo objetivo de marcar inicios y finales de activaciones locales y de campo lejano para facilitar la localización de sitios de ablación adecuados en procedimientos de ARF. Para este propósito, el algoritmo de delineación del ECG de superficie desarrollado previamente fue empleado para preprocesar los datos y marcar la detección del complejo QRS. En cuarto y último lugar, el algoritmo de delineación de ECG de superficie fue empleado, junto con un algoritmo de reducción de dimensionalidad, Multiple Kernel Learning, para agregar la información del ECG de 12 derivaciones y lograr la identificación de marcadores que permitan la estratificación del riesgo de muerte súbita cardíaca en pacientes con cardiomiopatía hipertrófica.

Palabras clave: aprendizaje profundo, aprendizaje automático no supervisado, cuantificación, interpretación, electrocardiograma, electrocardiograma intracavitario.

Abstract

Electrocardiographic signals, either acquired on the patient's skin (surface electrocardiogam, ECG) or invasively through catheterization (intracavitary electrocardiogram, iECG) offer a rich insight into the patient's cardiac condition and function given their ability to represent the electrical activity of the heart. However, the interpretation of ECG and iECG signals is a complex task that requires years of experience, difficulting the correct diagnosis for non-specialists, during stress-related situations such as in the intensive care unit, or in radiofrequency ablation (RFA) procedures where the physician has to interpret hundreds or thousands of individual signals. From the computational point of view, the development of high-performing pipelines from data analysis suffer from lack of large-scale annotated databases and from the "black-box" nature of state-of-the-art analysis approaches. This thesis attempts at developing machine learning-based algorithms that aid physicians in the task of automatic ECG and iECG interpretation. The contributions of this thesis are fourfold. Firstly, an ECG delineation tool has been developed for the markup of the onsets and offsets of the main cardiac waves (P, QRS and T waves) in recordings comprising any configuration of leads. Secondly, a novel synthetic data augmentation algorithm has been developed for palliating the impact of small-scale datasets in the development of robust delineation algorithms. Thirdly, this methodology was applied to similar data, intracavitary electrocardiographic recordings, with the objective of marking the onsets and offsets of events for facilitating the localization of suitable ablation sites. For this purpose, the ECG delineation algorithm previously developed was employed to pre-process the data and mark the QRS detection fiducials. Finally, the ECG delineation approach was employed alongside a dimensionality reduction algorithm, Multiple Kernel Learning, for aggregating the information of 12-lead ECGs with the objective of developing a pipeline for risk stratification of sudden cardiac death in patients with hypertrophic cardiomyopathy.

Keywords: Deep learning, unsupervised machine learning, quantification, interpretation, electrocardiogram, intracavitary electrocardiogram.

I will be brief, as it is my belief that dedications should mostly be private. I have been lucky to have extraordinary people aiding me during the hardships and joys of pursuing a doctoral degree. First and foremost, I want to thank my wife, Loli, my family and my friends, to whom I owe everything I am and without whom this document would never exist. Secondly, to my supervisor, Oscar Camara, who has always helped me in any way he could. Thirdly, to the amazing team at the Pompeu Fabra, comprising researchers, developers and administrative staff, who have been incredibly nice and fun, and with whom I have shared beautiful moments. And finally to you, the reader, without whom this work would not valid.

Contents

List of Figures

List of Tables

List of Acronyms

AF	atrial fibrillation
AI	artificial intelligence
ANN	artificial neural network
AP	accessory pathway
AS	amplifier saturation
AUC	area under the curve
AV	atrioventricular
AVRT	atrioventricular reentrant tachycardia
AWGN	additive white Gaussian noise
BN	batch normalization
BW	baseline wander
CA	cardiac arrhythmia
CNN	convolutional neural network
CS	coronary sinus
CSkel	cardiac skeleton
CT	computed tomography
CVD	cardiovascular diseases

DA	data augmentation
DEEP	decrement-evoked potential
DL	deep learning
DR	dimensionality reduction
DSP	digital signal processing
EAM	3D electroanatomical mapping
ECA	efficient channel attention
ECG	electrocardiogram
ECS	electrical conduction system of the heart
EGM	electrogram
EP	electrophysiology
FCN	fully-convolutional network
FF	far field
FN	false negatives
FNN	feedforward neural network
FP	false positives
FRLP	fractionated late potential
FRP	fractionated potential
FT	Fourier transform
GAN	generative adversarial network
GMM	Gaussian mixture model
GT	ground truth
HCM	hypertrophic cardiomyopathy
HF	heart failure
HFRP	highly fractionated potential
HMM	hidden Markov models
HSC	hidden slow conduction
HT	Hermite transform
ICD	implantable cardioverter defibrillator
iECG	intracavitary electrocardiogram
ILP	isolated late potential
k-NN	k-nearest neighbors

kM	k-means
LA	left atrium
LAD	left axis deviation
LAT	local activation time
LAVA	local abnormal ventricular activities
LBB	left bundle branch
LDA	linear discriminant analysis
LF	local field
LGE	late gadolinium enhancement
LP	late potential
LSTM	long short-term memory
LV	left ventricle
LVH	left ventricular hypertrophy
LVOT	left ventricular outflow tract
LVOTO	left ventricular outflow tract obstruction
M	mean
MI	myocardial infarction
MKL	multiple kernel learning
MKR	multiscale kernel regression
ML	machine learning
MR	mitral regurgitation
MRI	magnetic resonance imaging
NSR	normal sinus rhythm
OTVA	outflow tract ventricular arrhythmia
PCA	principal component analysis
PN	powerline noise
PS	pacemaker spikes
PT	phasor transform
PVs	pulmonary veins
RA	right atrium
RBB	right bundle branch
ReLU	rectified linear unit

RFA	radio frequency ablation
RNN	recurrent neural network
RPS	random periodic spikes
RV	right ventricle
SA	sinoatrial
SCD	sudden cardiac death
SD	standard deviation
SDr	spatial dropout
SoO	site of origin
SVM	support vector machine
TP	true positives
TWI	T wave inversion
U-Net	U-Net
UB	unidirectional block
VA	ventricular arrhythmia
VAE	variational autoencoders
VF	ventricular fibrillation
VT	ventricular tachycardia
VUS	variant of uncertain significance
W-Net	W-Net
WPW	Wolff-Parkinson-White
WT	wavelet transform

CHAPTER

1 Introduction

1.1. Cardiac anatomy and function

1.1.1. Cardiac anatomy and function

The heart is an organ whose purpose is to pump blood through the blood vessels to the rest of the body. The heart is is placed in the middle mediastinum of the thoracic cavity, and is protected and fixed within the thoracic chamber through the pericardium, a fluid-filled fibro-serous structure that encloses, protects and lubricates the heart [1]. The heart is usually 12 centimeters in length and 8 centimeters in width, and its apex is oriented towards the left side of the chest, although it is affected by genetic factors [1]. The heart is structurally organized in four main chambers and five main vessels. The cardiac chambers are responsible of holding and pumping blood, and comprise the left ventricle (LV) and the right ventricle (RV), two cone-shaped structures, as well as the left atrium (LA) and the right atrium (RA). The main cardiac vessels are responsible of distributing blood to different regions in the body, and comprise the pulmonary artery, the pulmonary veins, the aorta and the superior and inferior vena cava. Figure 1.1 depicts the distribution of the cardiac chambers and main cardiac vessels within the heart.

All cardiac chambers are physically isolated from each other. The ventricles are separated from the atria by the cardiac skeleton, a structure of dense connective tissue that surrounds the mitral and tricuspid valves and extends to the origins or

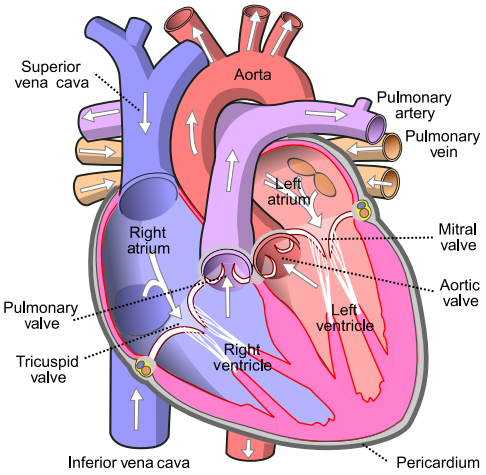


Figure 1.1: Diagram of the normal cardiac anatomy, depicting the main cardiac chambers, cardiac vessels and valves. Source: https://commons.wikimedia.org/.

the aorta and pulmonary trunk [2, 3]. The cardiac skeleton provides structural support and electrical isolation due to their fibrous composition, allowing atria and ventricles to contract at different times. The right heart (comprising the RV and RA), on its behalf, is separated from the left heart (composed of the LV and LA) by the septum. The septum is mainly composed of cardiac muscle but becomes fibrous closest to the cardiac skeleton and in its atrial portion, reminiscent of fetal development (ductus ovale). Surrounding the junction between the atria and the ventricles lies the coronary sulcus, a structure in which the main trunks of the coronary arteries and veins run [3, 4]. Coronary circulation is responsible of irrigating the myocardium itself, with the main cardiac arteries feeding oxygenated blood to the cardiac muscle and the cardiac veins returning deoxygenated blood into the RA. Coronary circulation returns to the heart through the coronary sinus (CS), a structure of importance in electrophysiology procedures as the atrial electrical activity can be explored with a catheter placed in it.

Functionally, the objective of the organ is to pump blood; it is performed through a four-stage process called the cardiac cycle (Figure 1.2). Phases 1 and 2 correspond to the relaxation of the heart (diastole) and phases 3 and 4 correspond to myocardial contraction (systole). During phase 1 (isovolumetric relaxation), the cardiac muscle relaxes after the previous pump: the pulmonary, aortic, mitral and tricuspid valves close. During phase 2 (ventricular filling), the ventricles are filled with blood coming from two main sources: the lungs, in the case of the LV, and the rest of the body for the RV. Firstly, the mitral and tricuspid valves open and the ven-

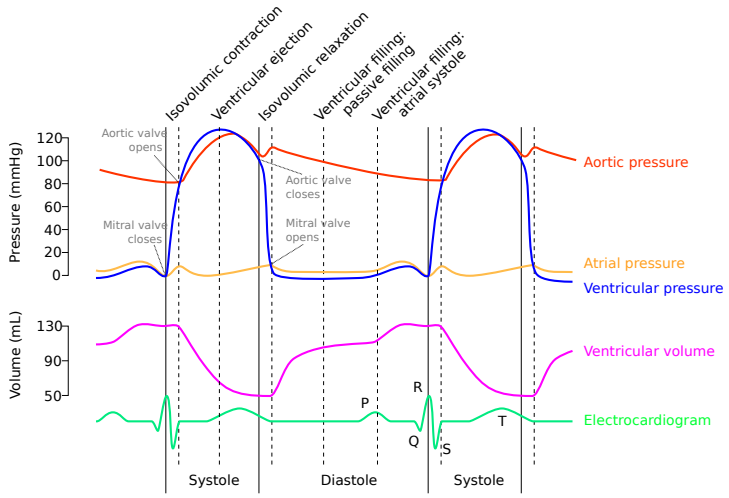


Figure 1.2: Wiggers diagram of aortic, (left) atrial and ventricular pressure related to the ventricular volume and the phases of the electrocardiogram. The diagram indicates the different phases of the cardiac cycle (isovolumetric contraction, ventricular ejection, isovolumetric relaxation, ventricular filling) and its relationship to the electrocardiogram's waves: P wave (atrial depolarization), QRS wave (ventricular depolarization) and T wave (ventricular repolarization). Adapted from `https://commons.wikimedia.org/`.

tricles are passively filled during ventricular and atrial relax relaxation, increasing their volume to accommodate incoming blood; secondly, the LA and RA contract to completely fill the ventricles. During phase 3 (isovolumetric contraction), the ventricles contract, building up pressure. The increased pressure impedes blood to regurgitate through the mitral and tricuspid valves, but the pressure differential between the ventricles and the outflow vessels (aorta and pulmonary artery) is not high enough for the aortic and pulmonary valves to open. Finally, during phase 4 (ventricular ejection), the aortic and pulmonary valves open and blood is ejected from the ventricles to the rest of the body, gradually reducing the pressure differential [3, 5]. The relationship of the different phases of the cardiac cycle with pressure, volume and electrocardiographic data (Wiggers diagram) has been depicted in Figure 1.2.

1.1.2. Cardiac histology

Histologically, the heart is composed of cardiac muscle, specialised conductive tissue, blood vessels and connective tissue. The cardiac muscle, on its behalf, is composed of cardiomyocytes, which are specialised muscle cells that are able to

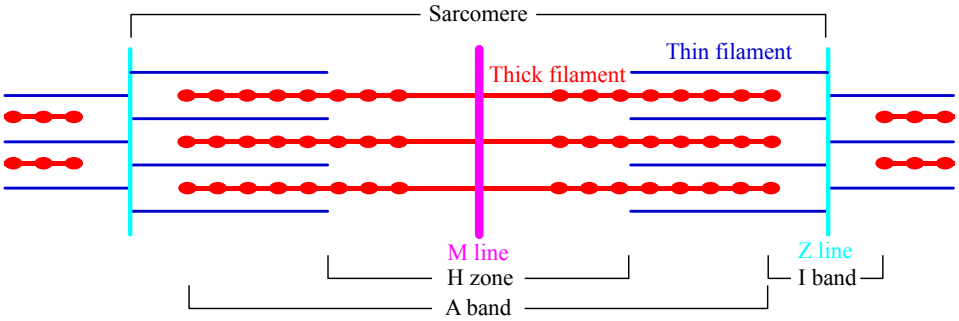


Figure 1.3: Diagram of a myofibril, containing many sarcomeres. When contracting, the space left between the thick filament and the Z line (I band) is reduced, causing contraction. Adapted from https://commons.wikimedia.org/.

contract when electrically stimulated.

Cardiomyocytes contain myofibrils, which are the cell's contractile elements. Myofibrils are large proteic structures internally arranged in sarcomeres, which are the smallest contractile unit in cardiac muscle. Sarcomeres are proteic structures composed of thick actin filaments surrounded by thin filaments (composed of troponin, actin and tropomyosin), producing contraction when the concentration of calcium ions within the cell increases, causing the filaments to attract each other (Figure 1.3). Contraction is performed anisotropically, i.e., the force is larger on the longitudinal direction of the fibers as compared to the transversal direction. Contractile force, in turn, reduces the internal volume of the heart, which causes a rise in blood pressure in the heart, allowing blood to flow to a region of lower pressure (e.g., the LV builds pressure, causing blood to open the aortic valve and flowing to the aorta) [6]. Disruptions in the normal functioning of the sarcomeres, which occurs in some genetic mutations such as in hypertrophic cardiomyopathy (HCM), can cause the myocardium to contract inefficiently and trigger cardiac remodelling, leading to overdeveloped left ventricular myocardium (left ventricular hypertrophy), and can have fatal consequences, as explored in chapter 5.

Structurally, the components are assembled in a three-layered structure, consisting in the endocardium, the myocardium and the epicardium. The endocardium is the innermost layer, in contact with the blood pool, and is composed of smooth muscle cells and connective tissue. It lines the atria, the ventricles and heart valves. The myocardium is the middle layer and is mainly composed of cardiomyocytes, capillaries and connective tissue. It is responsible for myocardial contraction. Finally, the epicardium is the outermost layer of the heart, is composed connective and adipose tissue and holds blood and lymphatic vessels. It is in contact with the pericardial fluid, which help avoid damaging the myocardium in blunt trauma [6].

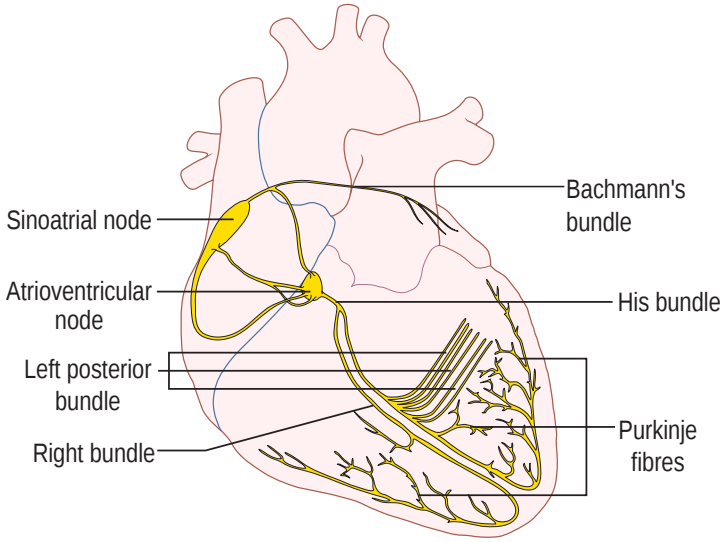


Figure 1.4: Diagram of the electrical conduction system of the heart. The electrical impulse is generated in the sinoatrial node and transmitted to the left atrium through the Bachmann's bundle and to the ventricles through the atrioventricular node (where it is delayed), the His bundle and the left/right bundle branches, spreading throughout the myocardium thanks to the Purkinje fibers. Source: https://commons.wikimedia.org/.

1.1.3. Cardiac electrophysiology

As has been outlined in the last section, mechanical activation of the myocardium is initiated by electrical activations. The exposure of troponin proteins to calcium ions trigger a chain reaction forcing actin and myosin to bind and, thus, to contract [7]. Myocardial contraction, however, must be performed in a precise and timely manner, as the phases of the cardiac cycle must not overlap; e.g., a simultaneous atrial and ventricular contraction is undesirable. The structure controlling the passing of electrical impulses in the myocardium is called the electrical conduction system of the heart (ECS), and has two main functions: impulse communication, for quickly connecting distant areas, and heartbeat initiation, for myocardial contraction to be produced without external stimuli. The ECS consists in a network of specialized cardiac cells that present increased conduction velocities and excitability, as well as mechanisms for autonomously generating stimuli (automaticity). Figure 1.4 represents the anatomical pathways that compose the ECS.

Cardiac electrical impulses occur due to changes in voltage across the membrane of myocardial cells, which produce the release of ionic currents between the in-

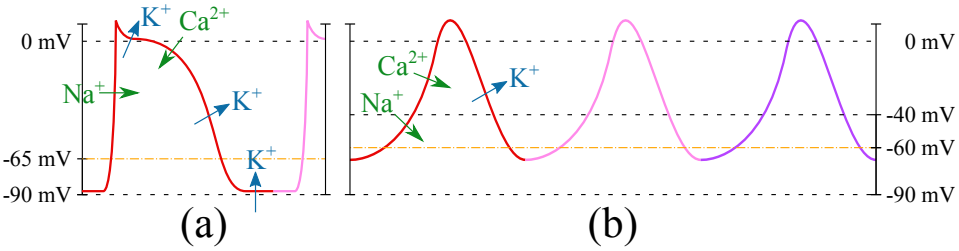


Figure 1.5: Action potential in a normal cardiomyocyte (a) and a pacemaker cell (b), with the main ionic interchange currents displayed in different parts of the cardiac cycle. Represented are several action potentials with different colors. The orange, dash-dotted line represents the threshold potential that triggers the massive displacement of ions within the cell.

tracellular and extracellular medium via each cell's ion channels. A full activation cycle of a cardiomyocyte is coined an "action potential", and is depicted in Figure 1.5. The main ionic currents involved in an action potential depend on the opening and closing of sodium (Na^+), calcium (Ca^{2+}) and potassium (K^+) ion channels, which are voltage-gated input/output channels of charged atoms present in the cell membrane. Once the cell's membrane potential reaches a certain threshold (e.g., -65 mV for non-specialized cardiomyocytes), dedicated Ca^{2+} ion channels open, binding to the troponin in the sarcomeres and triggering contraction. The electrical impulse, however, must be communicated to neighbouring cells for contraction to spread throughout the myocardium, which is performed through "gap junctions" that allow ions to pass through connecting the cytoplasm of two cells.

The electrical conduction system of the heart

The electrical impulse is generated in the sinoatrial (SA) node, a group of cells embedded in the wall of the RA. The SA node is connected to the LA through a band of specialized cells called the Bachmann's bundle, which quickly transfer the electrical impulse from the RA to the LA, allowing for synchronous contraction. The impulse is also transmitted to a structure named the atrioventricular (AV) node, located at the septal junction between the atria and the ventricles, whose function is to delay the signal before it traverses to the ventricles, permitting ventricular filling. The electrical isolation of the cardiac skeleton prevents atrial depolarization to trigger ventricular contraction, so the electrical impulse is forced to traverse the AV node to further continue the contraction. The AV node also possesses a degree of automaticity, being able to initiate the heartbeat in case of dysfunction of the SA node. The AV node further passes the impulse via the bundle of His (which branches into the right bundle branch, RBB, and the left bundle branch,

LBB), a pathway of cells embedded in the interventricular septum. The LBB, on its behalf, stems into the left posterior and anterior fascicular bundles, which communicate the impulse to the heart's walls. The RBB and LBB ramify into multiple Purkinje fibers, which are responsible of disseminating the impulse to the rest of the unspecialized cardiomyocytes, causing fast mechanical contraction [8, 9]. Figure 1.4 depicts the described pathways alongside an overview of the anatomy for reference.

Automaticity and pacemaker cells

It is important to remind that cardiomyocytes need external stimuli to be activated and, thus, contract. The external stimuli are produced by a specialized kind of cardiomyocytes called pacemaker cells, which contain fewer myofibrils as compared normal cardiomyocytes but can autonomously depolarize. Pacemaker cells present specific Na^+ channels that slowly positivize the cell's intracellular medium by introducing Na^+ ions in the cell, which causes a slow rise in the cell's voltage and periodically cause the intracellular medium to reach threshold potential, allowing the transmission of the electrical impulse to neighbouring cells through gap junctions.

Pacemaker cells, thus, are able to automatically and autonomously initiate electrical impulses. However, the rate at which pacemaker cells depolarize vary among different regions. As an example, the SA node outpaces any other region in the heart by producing 60 to 100 stimuli per minute (during rest), thus becoming the heart's natural pacemaker. Other regions, however, produce stimuli at slower rates: the AV node produces 40 to 60 stimuli per second, whereas other cells (e.g., Purkinje fibers, whose action potential is most similar to non-pacemaker cells but present automaticity due to specialized Na^+ ion channels) can fire at up to 20 to 40 stimuli per second [7–10]. Non-SA-mediated rhythms are called "escape" rhythms, whose function is enabling the heart to have alternative sources of autonomous cardiac contraction should the SA node malfunction, but escape rhythms can cause abnormal alterations in automaticity patterns that can lead to cardiovascular diseases (CVD) and are usually not fast enough to provide sufficient cardiac output.

Decremental properties of cardiac tissue and unidirectional blocks

An important characteristic of cardiac tissue is its decremental properties, which causes electrical impulses to be transmitted slower at faster firing rates. Not all cardiac tissue shows decremental properties: non-specialized cardiomyocytes do not exhibit decremental behaviour (if healthy), whereas some pacemaker cells do,

with the canonical example of the AV node. The elucidation of whether some tissue displays decremental properties is an important diagnostic tool during radio frequency ablation procedures and in electrophysiology exploration in general, as damaged tissue exhibits decremental properties whereas some structures naturally conduct impulses decrementally and, thus, informs about the mechanism of action of many types of arrhythmias [8, 9]. E.g., the decremental property of the AV node allows to limit ventricular activity under the circumstance of enhanced atrial activity, such as during atrial fibrillation (AF) episodes. In chapter 4, an algorithm is trained with the objective of measuring delays between signals for assessing physiological decremental properties in the AV node.

Another important characteristic are unidirectional blocks. Unidirectional blocks occur when a small amount of cardiomyocytes (source) attempt at propagating an electrical impulse to a large amount of cells (destination). Given that ionic currents that transmit stimuli to neighbouring cells are partly dissipated to the extracellular medium, if a cell has to stimulate a too large amount of myocardial mass, the concentration of ions might be insufficient to make them reach to their threshold potential, terminating the wavefront at that point. Unidirectional blocks are direction dependent: a cell might not be able to depolarize a large mass of cells in a given direction but, if the direction of the stimulus is reversed, that large myocardial mass could depolarize said cell, giving continuity to the electrical current.

1.1.4. General mechanisms of cardiac arrhythmias

Two main mechanisms are responsible for arrhythmia generation: alterations in impulse generation or in impulse conduction. Although these mechanisms cover a wide array of possible conditions, this introduction focuses on changes in automaticity and in reentry, which are the most frequent proarrhythmic mechanisms.

Changes in automaticity occur when cells outside the normal location for impulse generation (the SA node) have a faster firing rate than usual, competing with the normal wavefront propagation. Increased automaticity could happen in several locations, giving rise to different types of arrhythmias: atrial fibrillation might be caused by spontaneous depolarization in a group of cells in the PVs, whereas ventricular ectopic beats happen because of increased automaticity in Purkinje fibers. Automaticity alterations typically present “warming up” and “cooling down” mechanics, as they originate by gradually increasing their frequency until they stabilize at a certain rate and, when they stop, they return gradually to normal sinus rhythm. Moreover, they can occasionally trigger serious events such as ventricular fibrillation (VF), leading to systolic heart failure [8, 9].

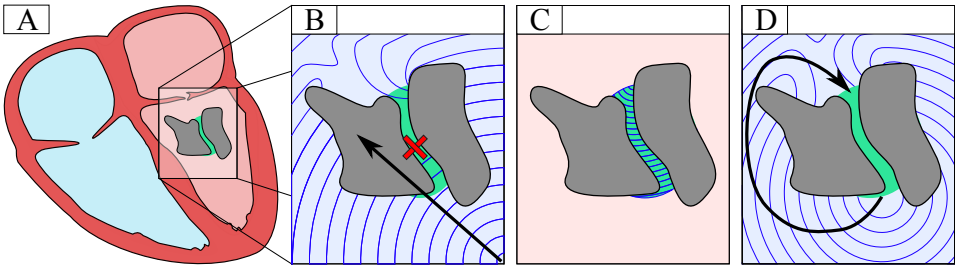


Figure 1.6: Schematic representation of a reentry mechanism. (A) location of a lesion (grayed area) in the left ventricle, which contains a region of partially viable tissue (green). (B) normal conduction in the healthy myocardium, which depolarizes the myocardium but causes an unidirectional block within the partially viable tissue (red cross). (C) continuation of the electrical impulse in the partially viable tissue, causing retrograde conduction, while the medium is repolarizing (red background). (D) initiation of an electrical impulse not mediated by the electrical conduction system of the heart, giving rise to a tachycardia.

Reentry mechanisms occur when a depolarization wavefront is continually propagated alongside a circuit, without getting extinguished, competing with the SA-mediated rhythm. The wavefront causes the rest of the myocardial tissue to be activated passively, causing untimely mechanical contraction. For reentrant tachycardias to occur, a part of the circuit must have repolarized (excitable gap), which is usually (although not always) associated with the existence of an anatomical substrate with unidirectional blocks. That is, there must be: a) an unexcitable area (e.g., fibrotic or scar tissue) with at least two conduction channels surrounding it (bundles of partially viable tissue that conduct electricity at a slower rate than healthy tissue); and b) an unidirectional block when stimulated in the direction of a normal wavefront. When the wavefront extinguishes, there is a secondary wavefront originating from the retrograde direction of the unidirectional block, which is perpetuated. Certain anatomical structures, such as accessory pathways, present in pathologies like the Wolff-Parkinson-White (WPW) syndrome, or dysfunctions of the AV node can act as substrate for tachycardias [8, 9]. Figure 1.6 shows a graphical representation of the mechanism of action of an anatomical substrate.

1.1.5. Surface and intracavitary electrocardiography

The depolarization wavefront traverses the myocardium to cause mechanical contraction, which is an invaluable tool to diagnose CVDs affecting the ECS, and can be measured with a polygraph. In current clinical practice, two main ways are employed to explore diseases affecting the ECS: the surface electrocardiogram (ECG)

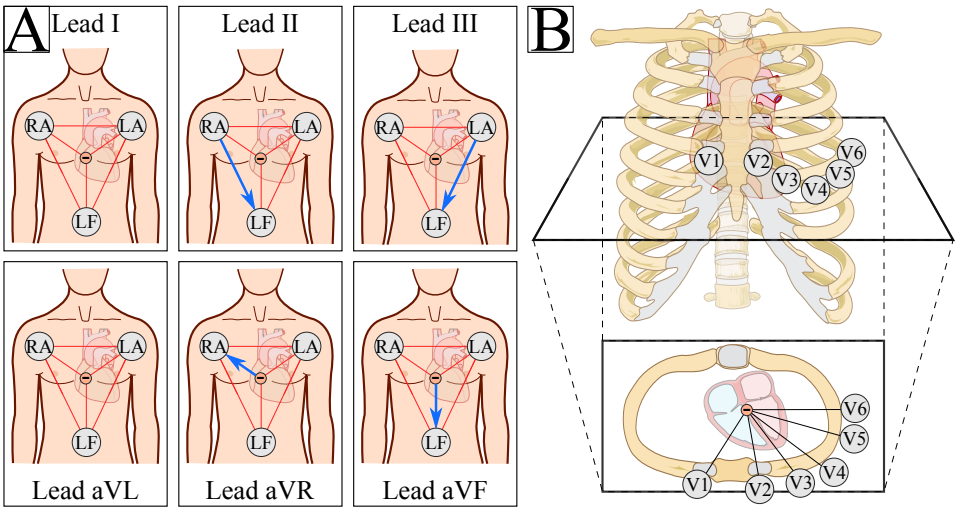


Figure 1.7: Placement of he electrodes for a standard, 12-lead electrocardiographic exam. (A) details the location of the patches that compose the limb leads (I, II, III, aVL, aVR, aVF), where the blue arrow represents the direction of the bipole. (B) represents the spatial disposition of the precordial lead patches. Partially adapted from `https://commons.wikimedia.org/`.

and the intracavitary electrocardiogram (iECG), also known as intracavitary electrogram (EGM).

Surface ECG

Surface ECG are recordings of the summation of action potentials in a patient's skin through electrodes located in different positions. The information captured in the electrodes has directional meaning, as usual practice displays the bipole formed by pairs of electrodes (lead); thus, the direction and relative intensity of the depolarization wavefront can be represented. Thus, whenever a wavefront is moving in the direction of the bipole, the time-voltage representation of the ECG will show a positive deflection [8]. The different leads that comprise an ECG act as alternative "views" of the patient's cardiac function alongside different directions, thus being complementary to one another. Although different lead configurations exist, current clinical practice employ 12 leads: six limb leads (I, II, III and the augmented limb leads: aVR, aVL and aVF), which record electrical information of the frontal plane, and six precordial leads (V1, V2, V3, V4, V5 and V6), which capture information of the axial plane [11]. A graphical representation of the position of the leads on the chest is depicted in Figure 1.7.

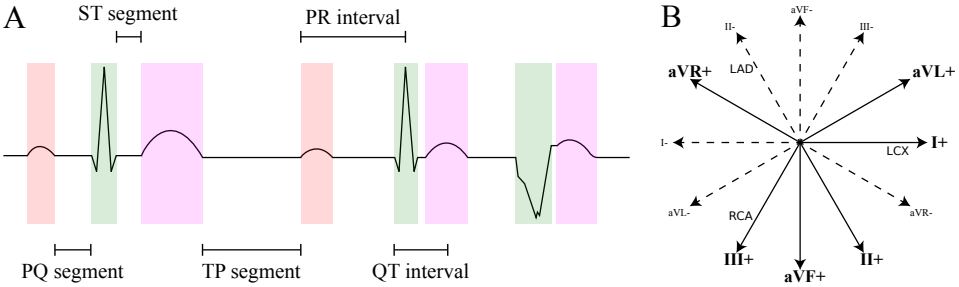


Figure 1.8: (A) Representation of the waves composing an electrocardiographic recording, where the red, green and magenta overlays represent the P, QRS and T waves, respectively (corresponding to atrial depolarization, ventricular depolarization and ventricular repolarization), and where the PQ, TP and QT intervals are displayed. A schematic representation of a ventricular-mediated beat shown, missing a P wave, displaying an abnormally wide QRS complex and a change of cardiac axis. (B) Hex-axial reference system, usually employed to determine the cardiac axis.

The morphology of an ECG signal is an accurate portrayal of cardiac function, as its waveform precedes the different phases of the cardiac cycle. Three main events from the cardiac cycle are visible in the surface ECG as different groups of deflections (or "waves") that have distinct morphologies: atrial depolarization (P wave, corresponding to atrial contraction), ventricular depolarization (QRS complex, composed of Q, R and S waves, corresponding to ventricular contraction) and ventricular repolarization (T wave, corresponding to ventricular relaxation). Atrial repolarization (relaxation) is usually masked within the QRS complex and is often indistinguishable. Between the waves there is an electrical "silence", or region of lower amplitude. Silence segments are conjointly named as the waves they isolate: the silence between the P wave and the QRS complex is named the PQ segment; between the QRS complex and the T waves, ST segment; and finally, the TP segment divides the T and P waves. Thus, a full cardiac cycle consists of the P, PQ, QRS, ST, T and TP segments [11]. Figure 1.8 (A) depicts an example of a cardiac cycle and Figure 1.2 correlates the phases of the cardiac cycle with ECG events.

With the ECG's partially overlapping information of different waveforms across several leads, clinicians perform differential diagnosis. Given that healthy individuals usually present a relatively stable ECG morphology, deviations from that morphology can be inferred to have clinical meaning. Most common clinical markers resort to interpreting the heart rate, heart rhythm, cardiac axis, cardiac intervals and waveform analysis:

- **Heart rate** measures the number of heartbeats per minute. The easiest way to measure it is to take the point in the QRS complex with the longest amplitude and average the time to the next QRS for a given amount of time (e.g., 10 or 30 seconds).

- The heart rhythm is a rough description of the rate at which the heart is beating, characterizing events such as the presence of tachycardia/bradycardia (accelerated/decelerated heart rates, respectively), ectopic beats or normal AV coordination (a QRS complex always follows a P wave, i.e., there is no QRS complexes with P waves missing), among others.

- **Cardiac axis analysis** gives a rough approximation of the direction of a ventricular depolarization, and is calculated by correlating the angle in the hex-axial reference system (Figure 1.8 (B)) to the direction of the depolarization wavefront as computed from the amplitudes of the QRS complex in different leads, which indicate whether the wavefront approaches or departs from the ECG electrode. In normal sinus rhythm patients, the ventricular depolarization is performed starting from the AV node and spreading throughout the myocardium, so the global waveform lies mostly in the (positive) direction of lead II (Figure 1.7). Depending on the patient, the QRS axis would lie within -30 and +90 degrees.

- **Interval analysis** provides information of the duration of the ECG segments. They are usually performed on the limb leads for easier comparison with the scientific literature, as those are the leads with most prominent and easy-to-measure waves. Although the measurements of interest are pathology specific, the PR interval, QRS duration and QT interval are usually measured in clinical practice.

- **Waveform analysis** consists in the interpretation of the relationship between an ECG morphology and its clinical significance. Although it is a very broad concept, it will be illustrated with the following example: in patients who suffer ventricular tachycardia (VT) episodes, the ECG presents several peculiarities: 1) there is usually no visible P wave and, when it is visible, it is not coordinated with the ventricular response; 2) the QRS wave is abnormally long and has a large amplitude; 3) the cardiac axis of the QRS complex usually differs from the normal sinus rhythm axis. Markers such as the amplitude of the QRS complex, the existence of delta or J waves or the presence of fractionation in the QRS complex are computed due to their diagnostic value.

Electrocardiographic exams present many advantages: it is an accessible, inexpensive and versatile test, and is capable to reliably represent a very large array of pathologies. Moreover, it is the only way to register the functioning of the ECS in a non-invasive manner, making it one of the most performed test to diagnose CVDs.

The ECG, however, is limited by two main factors. Firstly, due to its more distant relation to cardiac anatomy: deviations from normal sinus rhythm have to be interpreted under the light of a very wide array of possible contributing factors to the ECG morphology, such as age, gender, race or any other patient-specific factors such as cardiac axis deviations [8, 9]. Moreover, ECG traits can go unnoticed to non-specialists or even to trained cardiologists, especially in stress-related situations or when analysing multiple leads for several heart cycles for very long registries, such as Holter recordings [12]. Secondly, ECG analysis is often reliant on the definition of a set of structured measurements in the shape of various intervals and segments of the ECG, among others [12]. These measurements (and their interpretation) are, however, not fully automated to date. Computational methods can help unburden physicians and researchers by providing objective measurements over clinical data [13] through the development of robust and well-performing algorithms. Moreover, they can help in discovering new clinical markers that aid in patient and risk stratification [14–18]. The development of computational methods to aid in signal quantification and interpretation is the main focus of the thesis.

Intracavitary Electrograms

Intracavitary electrograms, hereinafter EGM, are recordings of time-voltage information from action potentials directly on the myocardial surface. EGMs are recorded invasively, introducing a catheter in the patient's body, most commonly through a vein (usually a femoral access, although the vena cava and jugular veins are used on occasion) [8]. Catheters employed in EP procedures are equipped with small electrodes used to probe the local voltages, which are visualized as summations of waveforms occurring at different frequencies, in the endocardial or epicardial surface. Recorded waveforms are influenced by the distance between the source of the signal and the electrode, as high-frequency components lose energy faster than their low-frequency counterparts. Thus, EGMs can be employed to visualize nearby and distant activations, which have markedly distinct frequential responses: close activations, usually described as local field (LF), display a mixture of high- and low-frequency components; whereas distant activations, usually coined far field (FF), only retain their low-frequency components [19]. Examples of LFs and FFs are schematically depicted in Figure 1.9.

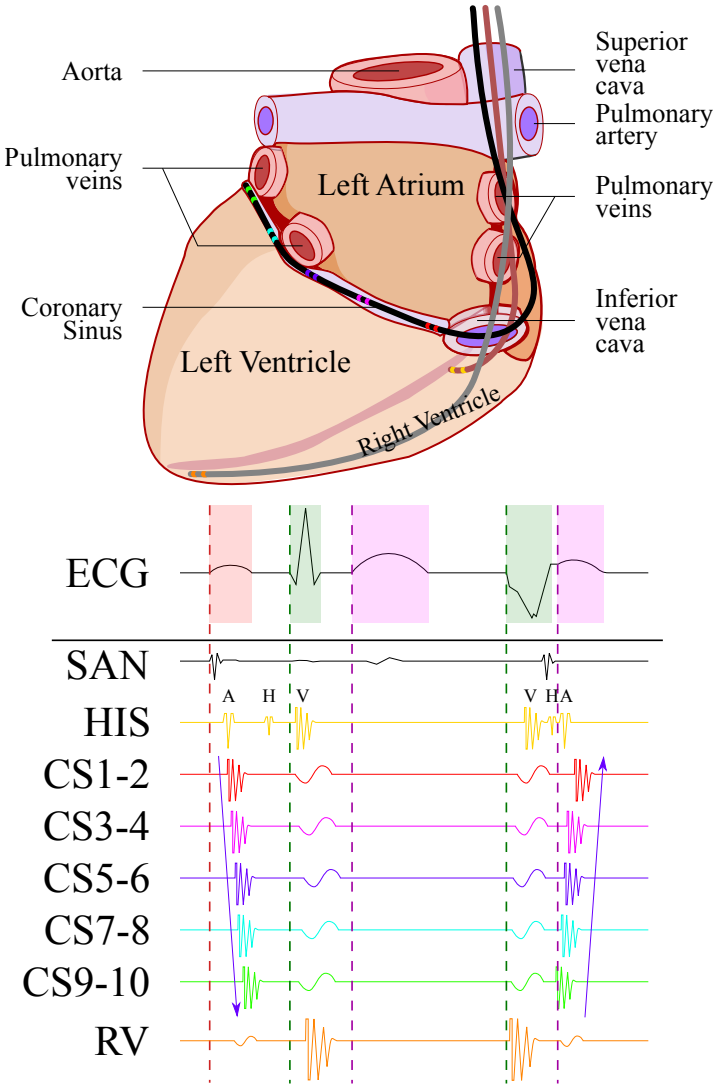


Figure 1.9: Diagram of catheter positions and recorded electrograms in an electrophysiology procedure. The Coronary Sinus (CS), right ventricular (RV) and His Bundle (HIS) catheters are color-coded with the different bipolar electrograms that produce the depicted signals. The signals depict responses for a physiological beat (left, with P wave) and of ventricular origin (right, no P wave). In a beat mediated by the sinoatrial node (SAN), the generated impulse depolarizes the atria (CS1-10), travels through the HIS and finally produces a ventricular response (RV). In a ventricular-mediated response, the earliest activation is recorded in the RV catheter, traverses the HIS and produces an atrial depolarization.

EGM information is local and directional. On the first hand, the information captured by the catheter depends on the electrode size: the larger, the least sensible it is to capture very small perturbations that might be of diagnostic interest, as potentials of interest can be masked in larger activations. On the other hand, the information provided by the catheters is directional, as EGMs record the voltage difference between two electrodes: the amplitude of an EGM is maximal when the direction of the depolarization wave is parallel to the axis of the electrodes, and zero if the wave travels perpendicularly [19].

Diagnosis through catheterization

The possibility opened by EP procedures of exploring the myocardial electrical activity allows for interpreting local depolarization patterns to identify areas with abnormal EGMs or unexpected behaviours of clinical significance. The local nature of EGM information is especially useful in radio frequency ablation (RFA) procedures, in which special catheters are used to deliver high-energy waves to damage the patient's tissue to correct conduction disorders. The objective of RFA procedures is to prevent arrhythmias from initiating by physically damaging strands of tissue that participate in arrhythmia perpetuation or present increased automaticity.

In general, clinical diagnosis of cardiac arrhythmias consists in recognizing their mechanism of action through the analysis of an EGM's direction of depolarization and the complexity of its waveform. Additionally, its time or voltage relationship to micro- (EGMs, e.g., time earliest activation) or macro- (surface ECG, e.g., peak time of QRS complex) events is measured, which allows the quantification of important clinical markers such as the local activation time (LAT) within the cardiac cycle [3]. Local alterations can take shape in many ways, as the ECS is complex and is composed of intertwined contractile tissue, areas of blocked conduction (such as the cardiac skeleton) and preferential electrical pathways, giving room for a large array of CVDs to arise. As an example, if a dysfunction of the AV node is suspected, a catheter can be employed to assess whether the electrical impulse has effectively reached the His bundle.

Waveform analysis

Waveform complexity plays an important role in diagnosing substrate-mediated arrhythmias, i.e., arrhythmias which occur due to the existence of a proarrhythmic modification of the local tissue composition (e.g., fibrosis, necrosis) which alters its conductive properties. Tissues with substrate display very complex electro-mechanical behaviour, presenting decremental properties that can provoke reentry-mediated arrhythmias. Currently, the exploration of substrate-based tachycardias

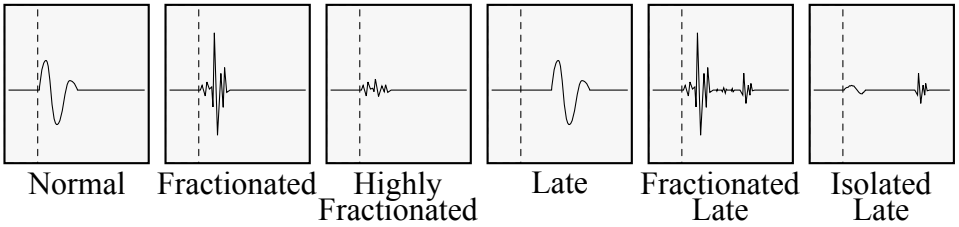


Figure 1.10: Schematic representation of common electrograms (EGM) found in electrophysiological studies with respect to a reference (dotted line). Normal electrograms are often high-frequency waves representing passing electrical impulses in the catheter's tip. If the catheter is placed in a fibrotic region, the electrogram can display fractionation/fragmentation. If the tissue presents decremental properties, late potentials might arise. Combinations of some characteristic EGMs might be found during studies and are of diagnostic interest [20].

is based on the identification of EGMs with local abnormal ventricular activities (LAVA) with respect to the expected waveform.

Different LAVAs have been identified and are currently under exploration in the scientific literature: late potential (LP) (sometimes subdivided into isolated late potential, ILP, and into fractionated late potential, FRLP), decrement-evoked potential (DEEP) (also known as hidden slow conduction, HSC [21]), fractionated potential (FRP) (sometimes subdivided into highly fractionated potential, HFRP, and into FRLP) [20, 22]. The LP group is characterized by the existence of EGMs that extend beyond or at the end of the QRS complex [23, 24]. FRP consist in EGMs with a large number of fractionations, which can be isolated or non-isolated. Finally, DEEPs are locations in the myocardium that present hidden decremental conduction. During EP procedures, a standard diagnosis tool is the application of extrastimuli sourcing from a catheter's electrode. Those extrastimuli are on-demand activations performed artificially by electrophysiologists that initiate a controlled depolarization at some known location in the myocardium. Given decremental properties of anatomical substrates and other cardiac structures, the application of extrastimuli can be used to evoke decremental conduction at a certain location in the myocardial surface, which is visualized as a late potential that was not present before the stimulation [24–26]. Although different clinical markers are available, the latest expert consensus in VT ablation recommends the usage of DEEPs for anatomical substrate ablation [27]. Some examples of LAVAs are shown in Figure 1.10.

The joint information required to diagnose and treat a CVD is, however, difficult to interpret due to the large amount of EGMs to be analyzed, with the added difficulty

of their location-dependence [3]. The large amount of recorded information causes intervention times and long-term success of some complex procedures, such as VT, to be improvable [28]. High recurrence rates are still hindering RFA procedures, which can be related to the development of new viable tissue within the substrate (substrate remodelling) or due to incomplete ablation procedures. Moreover, manually tagging DEEP potentials, especially in multi-electrode catheters, is prohibitively labour-intensive and time consuming [21, 28].

For simplifying EGM visualization and interpretation, 3D electroanatomical mapping systems (EAM) have been developed and are used in some EP procedures. The purpose of EAM systems is to visualize the cardiac anatomy alongside the electrical activation patterns, which is recorded with special catheters that are able to triangulate its spatial position within a coordinate system, while also storing the EGM's morphology for its posterior analysis [29]. However, despite recent advances in computational techniques for simplifying EGM analysis, EAM systems still perform relatively basic signal processing.

1.2. Computer-based analysis of cardiac signals

As it has been stated in the previous sections, ECG and iECG are the main diagnostic tools for assessing the electro-mechanical function of the heart. The ECG is the most accessible tool for cardiac function diagnosis in a non-invasive manner, being one of the main diagnostic test performed in usual clinical care and in emergency rooms. On the other hand, iECGs are relatively expensive to acquire but are the only available tool for evaluating local electrical patterns, enabling in-depth diagnosis through the evaluation of any deviations in the normal functioning of the ECS.

Despite their advantages and non-interchangeability with other cardiac assessment tools (such as computed tomography, CT, or magnetic resonance imaging, MRI, among others), the analysis of cardiac signals lacks automatization as compared to imaging modalities, even though it predates them. The medical image analysis community is very active, with hundreds of segmentation, classification and synthetic data generation algorithms being published yearly [30]. Of special interest are algorithms related to objective data quantification; as opposed to those employed in classification, data quantification algorithms reduce clinical workload, can aid in the development of new biomarkers for risk stratification and allow for more precise diagnostic criteria, instead of concentrating efforts on producing highly performing, albeit difficult to explain systems [15, 18].

The cardiac signal analysis community has been more focused on automated diag-

nosis algorithms than on automatic quantification [12, 31–33]. The unavailability of all-purpose quantification algorithms for cardiac signal analysis has left large portions of ECG data analysis to electrophysiologists, who oftentimes must perform quantifications during the procedures as part of the clinical guidelines leading to precise diagnosis. Apart from the inconvenience and time-consuming nature of in-procedure quantification, the process is very operator-dependent. Furthermore, most of the processing has been based on the application of digital signal processing (DSP) algorithms that allow the computation of higher-level clinical markers such as rhythm, rate and surrogates of wave morphology [12, 34, 35]. Only recently more modern algorithms based on artificial intelligence (AI), specifically deep learning (DL), have been employed for fully automatizing the cardiac signal analysis pipeline. DL-based algorithms have the advantage of incorporating data pre-processing as a step to be automatically tuned in a data-driven manner, precluding the need for intensive and difficult-to-tune data pre-processing. In the following sections, existing solutions will be summarized to give an overview of the state of the art.

1.2.1. Digital signal processing

Approaches based on digital signal processing (DSP) are often used for performing ECG/iECG data pre-processing, such as delineation or feature extraction. DSP-based algorithms transform the data into easier-to-process surrogates of the signal which include, but are not limited, to the computation of the signal's derivative/integral [36–38], wavelet transform (WT) [35, 39, 40], Fourier transform (FT) [38], phasor transform (PT) [41] or Hermite transform (HT) [34]. Moreover, most works apply some type of filtering [35–37, 42]. The applied data transformations are usually coupled with a posterior, experimentally calibrated logic layer that allows the computation of the final output, such as rule-based or adaptive threshold-based methods [35, 36, 39, 40]. Figure 1.11 exemplifies the usual processing performed with DSP-based approaches for ECG delineation.

Most DSP-based algorithms are focused of cardiac signal detection or delineation. Detection algorithms aim at localizing the occurrence of important cardiac events, whereas delineation algorithms seek to identify the onset and offset of every specific wave for every specific beat within a recording; i.e., in the case of the ECG, the computation of the P_{on}, P_{off}, QRS_{on}, QRS_{off}, T_{on} and T_{off} fiducials within the signal. One of the earliest available algorithms is "ecgpuwave", which consists in the application of a QRS detector algorithm (Pan-Tompkins, [43], based on derivative filtering) alongside a combination of band-pass filtering, signal derivative filtering, zero-crossing location and adaptive amplitude-valued thresholds to locate

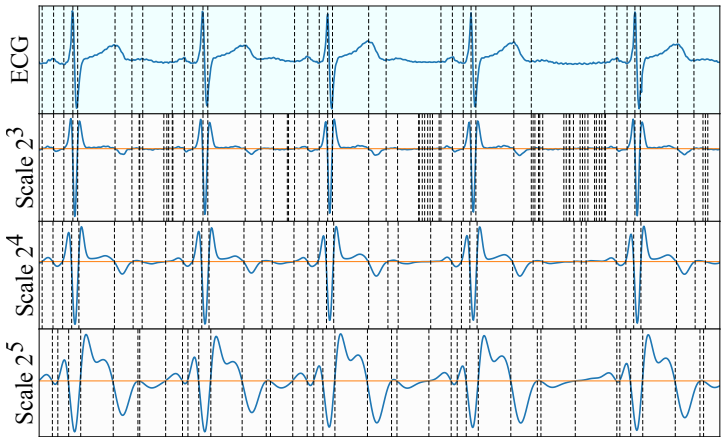


Figure 1.11: Usual processing in electrocardiogram (ECG) delineation with digital signal processing. The ECG is transformed (in the example, with the wavelet transform, WT) to an easier-to-process proxy for the computation of the fiducials. Similarly to [35, 40], some information is computed from the data transformation (in the example, zero crossings), and then aggregated to produce a prediction, usually through a rule-based algorithm.

the signal's fiducials [36]. A more streamlined approach was devised by Martínez *et al.* [35] employs the wavelet transform in place of the derivative operation to better characterize the zero crossings that would later be used as onsets and offsets for each individual wave, and is still used to date [34, 44]. A similar approach was applied by Alcaine *et al.* [39, 40] for the identification of wave fiducials in iECG signals. Many other similar DSP-based alternatives for ECG delineation exist based on the above described template, involving signal filtering, zero crossing approximation and decision rules to locate the final fiducials [35, 36, 45].

Other approaches have focused on diagnosis through the computation of data transformations. Palacios *et al.* [44] use the WT and the Phase-Rectified Signal Averaging for exploring periodic repolarization dynamics. Lyon *et al.* [34] employ the Hermite transform on the QRS complex as a surrogate of its morphology to explore biomarkers for the characterization of hypertrophic cardiomyopathy (HCM) patients through Holter recordings, attempting at linking specific ECG criteria to the diagnosis of HCM. However, articles that perform data transformations usually depend on quality signal delineation algorithms, as they are usually employed as pre-processing step for producing the algorithm's input.

1.2.2. Machine learning

Machine learning is a sub-field of AI comprising a set of algorithms that are able to adjust their behaviour through "learning" from a set of inputs. The model is "trained" through tuning the model's parameters, which govern a set of decision functions that learn an input-output mapping, process after which the model can be used to produce predictions on upcoming data. With the recent rise of data-driven approaches to the forefront of data analysis tools [32], computer-aided diagnosis algorithms for cardiac signal analysis has also shifted from its DSP-based origins since the consolidation of deep learning (DL) algorithms.

Machine learning algorithm types

To reduce the very broad scope of available computational solutions, two main types of algorithms will be explored with respect to their formulation and their data pre-processing requirements: "classical" machine learning (hereinafter ML) and DL algorithms. The former fit decision functions with a relatively small number of parameters and a relatively straight-forward formulation that allow the optimization of the decision function, often minimizing a concept of distance. The latter is a sub-branch of ML based on stacking operations, called "artificial neurons", which perform the weighted sum through learnable weights over the different attributes of an observation, and is usually performed jointly with a non-linear operation. An artificial neural network (ANN) is an algorithm that stacks several artificial neurons, so that the input of a certain neuron is the output of the previous neuron. Stacking a large quantity of trainable, non-linear operations allow DL algorithms to act as universal approximators of decision functions [46], and usually out-perform classical ML algorithms by a wide margin. A disadvantage of DL with respect to other ML algorithms is that the latter can sometimes guarantee that the algorithm has reached the global minimum of the optimization function, whereas the former can only assure that a minimum has been reached due to its stochastic optimization procedure.

Independently from the type of algorithm, ML algorithms can be employed for two main sub-groups: supervised or unsupervised learning. In supervised algorithms the model is trained through explicitly defining the expected output of each observation, so that the model learns a decision function that minimizes the mismatch between its prediction and the ground truth (GT). The specific task to be performed with supervised algorithms gives rise to different sub-branches such as classification, segmentation (known as delineation for cardiac signals), registration, regression or pattern recognition, among others. In unsupervised algorithms, the intrinsic variability contained in the training data is leveraged to obtain a data representation

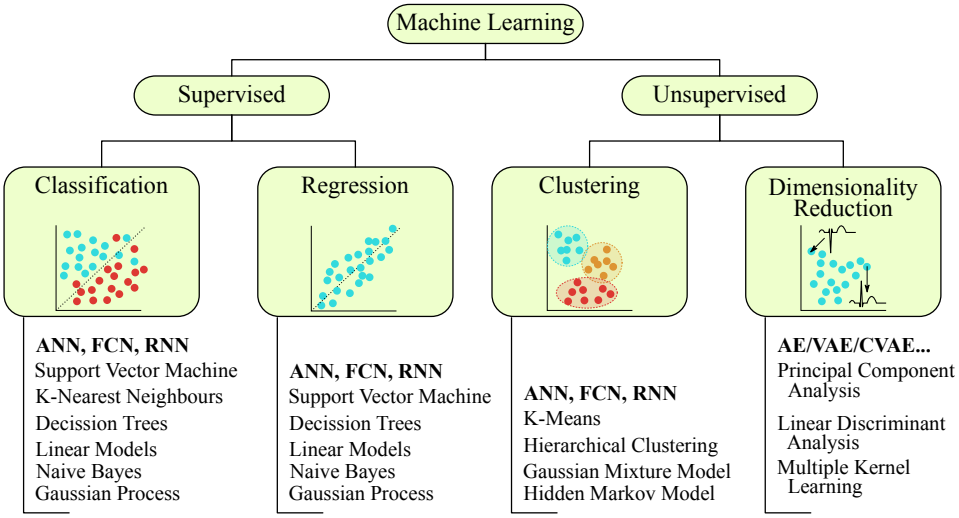


Figure 1.12: Categorization of machine learning algorithms. A first division is produced between supervised and unsupervised models, stemming from the usage of labels for optimizing the model's weights. Supervised models branch into classification, with discrete labels (e.g., *True* or *False*), and regression, with continuous labels. Unsupervised models are split into clustering methods, which group observations according to a similarity metric, and dimensionality reduction algorithms, which structure information for easier observation comparison. Deep learning algorithms (Artificial Neural Networks, ANN; Fully Convolutional Networks, FCN; Recurrent Neural Networks, RNN; Autoencoders, AE; Variational AE, VAE; Conditional VAEs, CVAE) can be used for a variety of tasks if a matching loss function is employed.

that attempts at minimizing the distance between observations that are "similar" (as measured by a metric) while maximizing the distance between "dissimilar" observations. A sub-branch of unsupervised learning, namely dimensionality reduction (DR) algorithms, are interesting for identifying differences contained within populations. DR algorithms take observations, for which pairwise distances might be difficult to measure, and convert them into an alternative representation that is able to condense the original information. The resulting low-dimensional space allows for comparing observations more easily, determining which samples are "similar". Figure 1.12 schematically depicts the different existing algorithms of importance for the thesis.

1.2.2.1. Classical machine learning

Many classical ML algorithms have been used for ECG data processing. Most of the literature concentrated in employing ML for classification, either using linear discriminators, support vector machine (SVM), decision trees, hidden Markov models (HMM) or many other algorithms [38, 47–51]. To increase the predictive power of ML algorithms, pre-processing steps are usually employed to facilitate model training and reduce the variability in the input domain. Such pre-processing is known as "feature engineering", and is based on the extraction of characteristics (or features) of the input population which allow the algorithm's developer to introduce expert knowledge into the model at the cost of being domain-specific. In the case of ECG analysis, features are extracted through DSP-based algorithms [38], which are the most straightforward way to manually reduce the complexity of the input signal. Thus, the ML model substitutes the developer-defined logic layer that was employed in DSP-based approaches, such as the adaptive threshold.

Three main tasks exist in the literature that employ ML for ECG classification. Firstly, heartbeat classification is often performed, where a single heartbeat is isolated (either using a window from a detection fiducial or using a full ECG delineation) and classified in the type of heartbeat (e.g., ventricular ectopic or bundle branch block, among many others). Secondly, for fully automated ECG diagnosis, either with wearables or in clinic [47, 51, 52], where whole ECG registries are analyzed to diagnose arrhythmias or other CVDs. Finally, other approaches have focused of ECG delineation, in the shape of Gaussian mixture model (GMM) or hidden Markov models (HMM) [12, 53, 54]. However, classification methods scale poorly when trained on large datasets, generally underperform when compared to DSP- and DL-based algorithms and require laborious definition and extraction of hand-crafted features which hinder their applicability in real-world scenarios.

Other wide application of ML analysis on cardiac signals is the application of unsupervised/dimensionality reduction (DR) algorithms [17, 34, 47]. DR algorithms employ the raw signal or hand-crafted features extracted from the signal as input to a group of algorithms, such as principal component analysis (PCA), multiple kernel learning (MKL) or linear discriminant analysis (LDA), to explore the distribution of observations in the output space. DR algorithms operate by transforming observations into a low-dimensional embedding throught the usage of similarity metrics, which are computed to estimate the distances between observations in the input space. Then, the algorithm optimizes a projection into an output space so that the relative distances between observations are preserved. DR algorithms are especially useful when exploring clinical markers or features that can aid in risk

stratification or in biomarker discovery [15, 18, 55].

On the other hand, given the difficulty in acquiring and annotating iECG databases [56], not many ML-based approaches exist in the literature. Most ML algorithms employed in iECG analysis have focused on localizing reentry mechanisms, potential ablation sites or propagation patterns [50, 51, 56–59].

1.2.2.2. Deep learning

Deep learning is a branch of ML which uses the artificial neuron (or perceptron) as the main processing unit/operation, consisting in a non-linear operation performed over the weighted sum of the algorithm's inputs. Multiple artificial neurons form "layers" of relatively simple operations, which are then embedded in multi-layered structures (also known as "architectures" or "networks") that are able to produce much more complex input-output mappings. These multi-layered algorithm, when accounting for the sufficient number of layers (or sufficiently "deep"), are able to act as universal approximators, and have the ability to process raw data as input [46, 60, 61]. The artificial neurons can be organized in different topologies, which give rise to different components that make the most commonly used building blocks of the most common networks types: the feedforward neural network (FNN) employs solely artificial neurons; the convolutional neural network (CNN) uses a convolutional operation as the main building block, which uses a trainable weight (or "kernel") whose size is smaller than the input size (known as "kernel size") and convolves the kernel with the input; and the recurrent neural network (RNN), which employs a neuron that produces an output for every time-step, taking as input both the current time-step's input as well as the output for the previous time-step. A schematic representation of artificial neurons is depicted in Figure 1.13.

Once the network's topology has been defined, the model's weights are trained through stochastic gradient descent: for any given input, the model predicts the output according to its current weight values, which is employed to compute an error metric or "loss function" and then used for iteratively adjusting the weights of the model through a process called backpropagation. Loss functions might be supervised (e.g., computing the Dice score of a segmentation, an L_2 error term or any other metric that uses the expected value of the algorithm as a reference) or unsupervised (e.g., computing the "separation" of similar samples such as the Kullback–Leibler divergence, which measures the difference between two probability distributions). Backpropagation, on its behalf, computes the gradient of the loss function, allowing to estimate whether an update in an single neuron's weights would result in a reduced global error [60]. Given the large scale of datasets em-

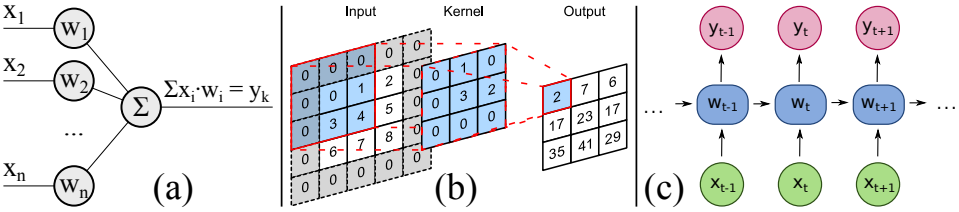


Figure 1.13: Schematic representation of the most widespread artificial neurons before applying the non-linear operation. (a) depicts the operation performed in feedforward neural networks. (b) represents convolutional operations found in convolutional neural networks. (c) shows the basic operation performed in recurrent neural networks. In all examples, w_i represents the i-th weight; x_j, the j-th input; and y_k, the k-th output.

ployed in current data analysis, batch-based, stochastic algorithms have been developed for approximating the network's gradient [62–64].

Convolutional Neural Networks

Convolutional layers have several important properties that are especially desirable for biomedical data analysis, making convolutional-based architectures the most employed in medical image/signal analysis. Firstly, the convolutional layer weights are the same at all positions when convolving the input, which makes convolutional operations more parameter-efficient and force the weights to generalize as much as possible. Secondly, the formulation of the convolutional layer can be extended to N-dimensional inputs, which allow for the analysis of 2D and 3D inputs. Finally, their formulation can be thought of being analogous to the design of digital filters, providing a solid foundation for comparing the behaviour of CNNs with carefully hand-crafted DSP or image processing solutions. It has been demonstrated that earlier layers are generally responsible for computing derivative filters for input pre-processing, such as detecting edges, and latter layers agglomerate the filtered input into larger structures [65]. Figure 1.14 shows a schematic representation of a CNN.

ECG and iECG analysis is, as was the case with classical ML, mostly based on classification. DL-based ECG analysis focuses on similar tasks as their ML counterpart, with heartbeat classification and arrhythmia detection as the main focus [31, 32, 52, 66–71]. These works mostly employ some sort of ResNet variant [72], which uses convolutional layers alongside feedforward artificial neurons and a classification loss function (usually cross-entropy) to produce a categorical classification of the input data. iECG analysis, on its behalf, is centered around au-

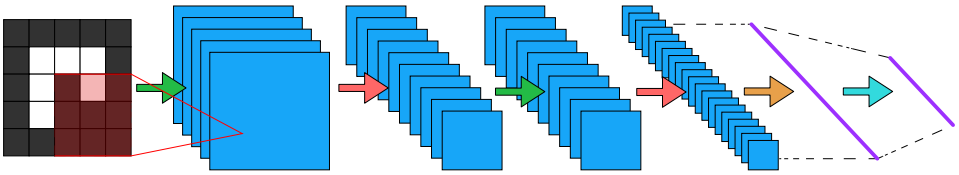


Figure 1.14: Convolutional Neural Network processing of an image (left). Blue rectangles and purple lines represent intermediate results after operations, and arrows represent operations: convolution (green), downsampling (red), global average pooling (purple) and fully connected layer (cyan).

tomatic ablation site location through propagation patterns [73–76], employing similar architectures.

Other approaches have focused on automatic ECG quantification, delineating the onsets and offset of important waveforms in the recorded trace. Some approaches have employed some sort of classifier (CNN or RNN) to estimate the probability of a single sample to be identified as a waveform's onset or offset [77–79]. Other approaches have profited from the medical image segmentation literature, applying the U-Net [80] (a type of fully-convolutional network, FCN, specifically designed for biomedical image segmentation; Figure 1.15) or some variation of it [81–85]. A corpora of works have attempted at producing ECG delineation directly employing images or spectral representations as inputs as opposed to 1D traces, but have also been largely been restricted to classification [86]. Finally, to the best of our knowledge, no DL-based iECG delineation approach exists to date.

1.2.3. Limitations of machine-learning-based algorithms for cardiac signal analysis

Although the application of ML for the analysis of cardiac signals has grown greatly in the last years due to the resurgence of DL methods and to the large increase in availability of annotated datasets, current state-of-the-art presents several drawbacks preventing their widespread adoption. The drawbacks gravitate towards two main points: the small size of open datasets and the black-box nature of some algorithms, which prevents model explainability.

From the data availability perspective, most existing datasets before the explosion of DL counted with a relatively small number of patients (usually below 100). Although small-scale datasets were sufficient for DSP- and classical ML-based approaches, where agglomerating information from too large data sources was computationally challenging or impractical, DL-based approaches usually require much larger datasets (e.g., ImageNet contains 1,281,167 tagged images [87]). In

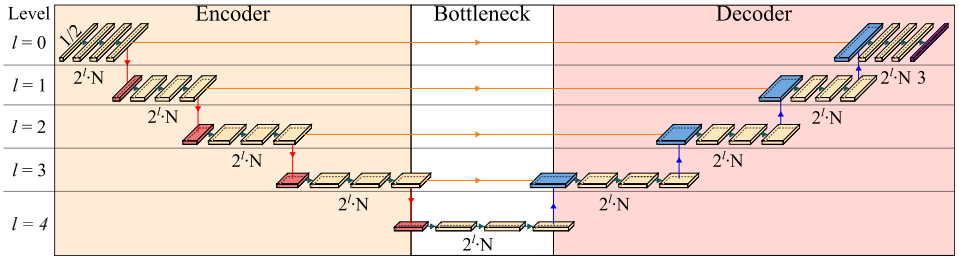


Figure 1.15: Example of a fully convolutional network (FCN) topology, which employ convolutional operations (green arrows), up/downsampling operations (blue/red arrows) and concatenation operations (orange arrows). Many FCNs are organized in an encoder/decoder structure, where the encoder processes the input to obtain an abstracted representation of the input data (bottleneck) whereas the decoder transforms the abstract representation and merges it with the high-frequency information from the unaltered input to produce an output (purple block). In the figure, N represents the initial number of channels to perform convolution upon.

recent years, however, institutions have started collecting large databases for classification that greatly outnumber the amount of records contained in public access initiatives such as PhysioNet [88]. Moreover, although some databases for cardiac signal quantification exist [89, 90], they face the usual drawbacks of open databases, namely the small intra- and inter-patient variability; or, in the case of iECG, no open dataset exists, neither large nor small. To give an overview of the magnitude difference between public and private study sizes, a open dataset comprising 66,101 annotated recordings was released as part of the PhysioNet Challenge 2020 [52], thus becoming one of the largest public annotated datasets available for ECG classification. Parallelly, the Telehealth network of Minas Gerais was recently able to gather a private dataset of 2,470,424 ECG recordings, obtaining a specificity larger than 99% with a relatively simple CNN [91]. The existence of large-scale datasets is especially relevant due to performance plateauing with respect to model capacity whilst improvements could still be achieved with larger data collections, which has been coined as the "unreasonable effectiveness of data" [92]. However, large datasets remain inaccessible to researchers, forcing scientific research to manage with limited datasets.

From the explainability point of view, many ML algorithms, most notably DL-based approaches, lack easy interpretation of the decision function that ultimately maps an input to an output due to several factors. On the one hand, existing datasets are designed for CVD classification, which many times are used alongside DL algorithms due to the relative cheapness of annotating labels for classification as opposed to individual markings for data quantification. Thus,

most published works applying DL in cardiac signals focus on classification [52, 66, 70, 71, 88, 91, 93, 94]. On the second hand, the lack of interpretability of DL algorithms becomes a fundamental drawback in data-sensitive scenarios. Applying computational solutions in the clinical practice should not aim at providing closed, uninterpretable diagnostic proposals but at reducing their workload and at supporting decision-making with increasingly more automated tools [12, 13]. The inability to rationalize the set of characteristics that are internally used by the model causes a dismissal of the diagnostic responsibility to an algorithm, which raises the concern of accountability in cases of misdiagnosis or negligence due to the shift in the evidence-based norms that stand at the core of the diagnostic process [95]. A middle ground between harnessing the high performance of DL algorithms and the interpretability of ML- or DSP-based algorithms is to employ DL models for immediately interpretable tasks such as signal delineation. The existence of quality quantifiation algorithms would unburden clinical practitioners from labour-intensive tasks attaining the ability to be easily checked for errors, effectively extending the capabilities of the users.

The focus of the thesis is to work on these limitations from a computational point of view. Firstly, our work concentrates on DL-based data quantification by producing high-quality delineations of ECG data. Two paths are explored. In chapter 2, a FCN-based network was extensively tested in the various aspects that alter the model's capacity (network's depth, several state-of-the-art architectural modifications) to assess its capability to produce high-quality delineations. Given its limitations due to the reduced dataset size, chapter 3 describes the development of a synthetic data augmentation (DA) methodology for artificially increasing the dataset size, alongside other state-of-the-art architectural modifications. Although some works exist for synthetic GT production, they are based on simulating biophysical models of the heart, which are generally too slow for massive sample generation [96], or on the usage of data-driven approaches such as generative adversarial networks (GAN) for generating samples in real time, which both cannot extend beyond the input data's manifold and reduces the control on the generation parameters. The approaches' pros and cons will be discussed in length in chapter 3. The combined approach proved sufficient to produce high-quality ECG delineation, which was then employed for iECG recordings in chapter 4. Finally, chapter 5 applies the methodology developed for ECG delineation for performing dimensionality reduction on a population of HCM patients, with the objective of identifying clinical markers that aid in risk stratification.

1.3. Motivation and objectives

Cardiovascular diseases are the leading cause of death worldwide, with 20 million new cases in European countries in 2017, accounting for up to 30% of global deaths worldwide [47] and for the highest expenditure in healthcare in Europe (up to 19% of the total spending in high-income countries) [97]. Clinical practice requires the quantification of biomarkers during diagnosis for an objective assessment of a patient's condition. Specifically, the analysis of ECG signals plays a fundamental role for screening, diagnosis and treatment of conditions affecting the heart's ECS. The ECG is one of the main cardiac diagnosis tool, with 300 million annual exams reported in 1999 [98], and is often the first performed test to assess cardiac function. The iECG, on its behalf, is the main diagnostic tool preceding and during RFA procedures, which the most recommended treatment in many cardiac arrhythmias for non-responders to drug therapy [99–101]. Despite their importance, computational solutions for ECG quantification are error prone in complex waveforms, such as in VTs, and leave room for improvement. In the case of iECGs, no delineation algorithm exists in the literature, with current solutions solely performing basic detection of the local field component with the highest amplitude. Both ECGs and iECGs are mostly analyzed with algorithms based on DSP, but these algorithms are laborious to develop and to adapt to the wide variability present in routine clinical practice. On the other hand, the rise of DL has opened a new landscape for the analysis of biomedical data [65], but the application of said algorithms for cardiac signal analysis is mainly focused in classification.

The objective of the thesis is to develop high-performing ECG and iECG delineation algorithms to reduce quantification-related workload in clinical environments. For this purpose, DL models were trained for ECG and iECG signals, obtaining the onsets and offsets pairs related to each main wave in the analysed registries. Firstly, an algorithm was developed for ECG delineation, based on retrieving the temporal markers corresponding to the onsets and offsets of the P, QRS and T waves that compose a heartbeat. Given the small scale of available databases for delineation, a custom synthetic data augmentation algorithm was developed for maximizing input variability. Secondly, the developed methodology was applied for the delineation of iECG recordings, localizing individual local activations at specific locations in the myocardial wall with the objective of identifying different types of pathological EGMs. Concentrating on quantification tasks allows for obtaining better data representations, producing versatile algorithms for their usage in different scenarios. Thus, as a final contribution, the developed ECG delineation tool and an unsupervised learning algorithm were employed on rest ECGs

with the objective of characterizing and developing new ECG-based biomarkers for risk stratification in a population of patients with HCM.

1.3.1. Contributions

- Development of an algorithm for the automatic delineation of ECG recordings, irrespective of their configuration (number of leads, sampling frequency).
- Development of an algorithm for the automatic delineation of iECG recordings, irrespective of their configuration (number of electrodes in the catheter, sampling frequency).
- Development of novel synthetic data generation strategy for cardiac signals (ECGs and iECGs), based on composing traces from wave fundamentals.
- Generation of ground truth for an open-source ECG database of patients with outflow tract ventricular arrhythmias and a private dataset of iECG recordings.
- Delineation of a 12-lead ECG dataset of patients with HCM, alongside a dimensionality reduction algorithm, for performing risk stratification.

1.3.2. Outline of the thesis

The rest of the document is organized as follows:

Chapter 2 describes the development of a DL-based algorithm for the automatic quantification of surface ECG signals.

Chapter 3 describes an improvement over Chapter 2's delineator, in the shape of a synthetic DA algorithm and architectural modifications.

Chapter 4 describes the application of the developed methodology for ECG delineation for the automatic quantification of iECG signals.

Chapter 5 describes the development of an algorithm based on DR for relating sarcomeric mutations with ECG phenotypes.

Chapter 6 summarizes the main ideas, contributions, limitations and future directions of the thesis.

CHAPTER

2

Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks

2.1. Introduction

Surface electrocardiogram (ECG) is the main cardiac diagnostic and monitoring tool in clinical practice due to its widespread accessibility and ease of use. Usually, physicians perform visual inspection of the ECG in order to diagnose a patient, interpreting potential pathological deviations in the waveform. However, these markers might go unnoticed to non-specialists or even to trained cardiologists, especially when analysing multiple leads for several heart cycles or in stress-related situations. Moreover, this analysis is often reliant on the definition of a set of structured measurements in the shape of various intervals and segments of the ECG, among others [12].

Computational methods can help unburden physicians of these problems by providing objective measurements over clinical data [13] or by aiding in the discovery of potential biomarkers [47, 102]. For these purposes, ECG detection and delineation (hereinafter delineation) is often a prerequisite step, aiding in data structuring [47]. ECG delineation consists in computing the onset and offset locations for

This chapter is adapted from: **Jimenez-Perez, G.**, Alcaine, A. & Camara, O. Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks. *Sci Rep* **11**, 863 (2021). https://doi.org/10.1038/s41598-020-79512-7

each ECG wave (P, QRS and T waves). Delineation can be performed directly on all available leads (multi-lead) or on individual leads (single-lead).

Several computational methods exist in the literature for ECG data processing. To the best of our knowledge, digital signal processing (DSP) algorithms using the wavelet transform (WT) and rule-based adaptive thresholds are often cited as state-of-the-art for ECG delineation [35, 45], reaching high precision and recall values of 95%, 99% and 98% for the P, QRS and T waves. However, these methods require laborious rule adaptation when extended to morphologies outside the development dataset; moreover, these algorithms were fine-tuned using the whole dataset, compromising their generalization.

Although DSP algorithms have historically been used for this purpose, machine learning (ML) tools are gaining momentum for biomedical applications. Nonetheless, and in spite of their good performance, ML methods on the ECG are scarce and have mainly focused on classification [31]. As suggested by Pinto et al. [33], this can be caused by the lack of large, manually annotated databases for ECG analysis, usually including less than a hundred patients.

Classical ML algorithms, in the shape of Gaussian mixture models [53] or hidden Markov models (HMM) [54], have been applied for ECG delineation. However, these methods might scale poorly when trained on large datasets and generally underperform when compared to DSP- and other ML-based algorithms. Deep learning (DL) algorithms, a branch of machine learning capable of assimilating large amounts of data, have also been used for delineation. Specifically, convolutional neural networks (CNN) [77, 79], long short-term memory (LSTM) networks [78] and fully-convolutional networks (FCN) [81, 83]. However, some of these works solely delineate the QRS wave [77], whereas others only validate their performance on sinus rhythm [83] or show reduced performance compared to DSP-based approaches [78, 79, 81].

In this work we present the adaptation of the U-Net architecture [80], the most successful FCN for biomedical image segmentation, for ECG delineation. For this purpose, the U-Net was adapted to one-dimensional data and delineation was framed as a segmentation task. The developed methodology was tested on the PhysioNet's QT database [89], which holds approximately 3,000 two-lead beats annotated by expert cardiologists having both leads in sight. Given the difficulties posed by the small dataset, with high intra-recording beat redundancy and large patient variability, several regularization strategies were applied, consisting in developing ECG-tailored data augmentation (DA) such as baseline wander or powerline noise, in performing semi-supervised pre-training with automatic labels and in adding in-built regularizers such as spatial dropout (SDr) and batch nor-

malization (BN) in the architecture. A large array of architectural variations were tested for completeness.

The rest of the chapter is organized as follows. Section 2.2 describes the employed database. Section 2.3 details the methodology followed in this work. Section 2.4 addresses the results obtained by this work. Section 2.5 discusses about the obtained results and their implications on the feasibility of applying this pipeline in the clinical practice. Finally, Section 2.6 summarizes this work's conclusions. A preliminary version of this work has been reported in [81].

2.2. Materials

The QT database was employed for model training and evaluation [89]. The QT database is comprised of 105 ambulatory, two-lead recordings of 15 minutes at 250 Hz representing a variety of pathologies, comprising arrhythmia, ischemic and non-ischemic ST episodes, slow ST level drift, transient ST depression and sudden cardiac death. Two label sets exist per recording: a manual annotation performed by an expert cardiologist ("manual", hereinafter) consisting of approximately 30 fully delineated beats per recording, and an automatic delineation ("automatic") performed on every beat of each recording [37]. The automatic ground truth is produced in a single-lead manner, whereas the manual dataset is annotated in a multi-lead fashion. Each annotation set holds nine fiducials per beat: the P (if present), QRS and T wave detection markers and their respective onsets and offsets.

Some recordings in the manual dataset had to be partially re-annotated, as they contained extrasystolic beats that were neither detected nor delineated. Specifically, 112 beats in recordings *sel102*, *sel213*, *sel221*, *sel308*, *sel44* and *sel820* were added. Isolated delineations in the manual dataset were also excluded, as they were unusable for training the algorithm. Records *sel232*, *sel233* and *sel36* were discarded given that the annotations were incomplete. A single recording, *sel35*, was discarded due to being the only recording in atrial flutter, making it impossible to abstract this morphology with a single example. A total of $3,246$ manual beats and $135,170$ automatic beats were available for training.

Lastly, all fiducials for a recording r and lead ℓ were transformed into binary masks B:

$$B^{(r,\ell)}[n] = \begin{cases} 1 & \text{if } n \in \left[w_{\text{on}}^{(r,\ell)}[m], w_{\text{off}}^{(r,\ell)}[m] \right]_{0..M} \\ 0 & \text{otherwise} \end{cases}, \tag{2.1}$$

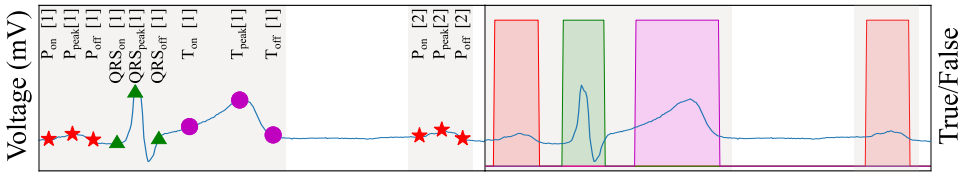


Figure 2.1: Database's manual ground truth as fiducials (left) and as three overlapped binary masks (right). Red mask/stars: P wave. Green mask/triangles: QRS wave. Magenta mask/circles: T wave. The figure depicts the employed notation.

where $W_i^{(r,\ell)}$ represent the fiducials, with $W \in \{P, QRS, T\}$, $m \in [0, M]$ are the number of annotated fiducials, and n is the sample number. Given the existence of three main waves, the information of the different waves was encoded into separate channels in the label tensor. Figure 2.1 depicts both the original fiducials and the binary masks.

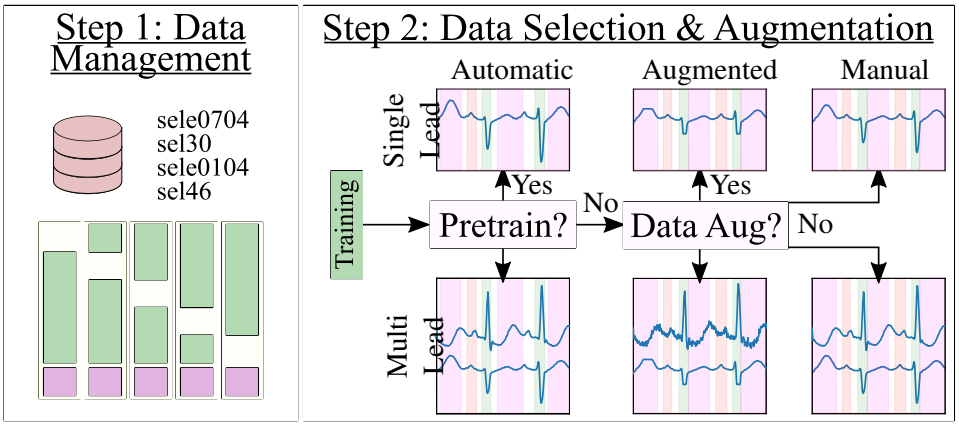
2.3. Methods

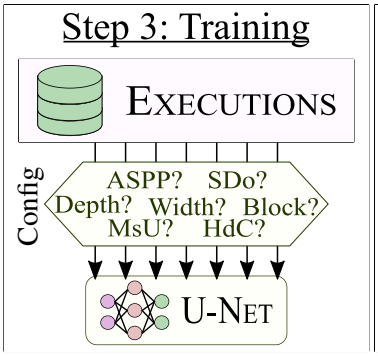
The developed methodology for ECG delineation is depicted in Figure 2.2. The first step describes the data splitting and management (Section 2.3.1). The second step outlines data selection and augmentation methodologies (Section 2.3.1 and Section 2.3.2). The third step summarizes the base architecture and its additions (Section 2.3.3). The fourth step details the evaluation methodology (Section 2.3.4). The configurations tested in the third step are listed in Section 2.3.5. We have made our code publicly available in `https://github.com/guillermo-jimenez/ECGDelNet`.

2.3.1. Data management and selection

In ML algorithms, data instances are usually divided into train, test and validation sets in a non-overlapping manner. However, given the high intra-recording and inter-lead beat similarity of ECG signals, a higher risk of performing an incorrect data splitting is incurred, assigning similar representations of the same entity to different sets. Models trained with this flawed splitting incur the risk of memorizing the data instead of inferring abstract patterns over it, especially in the case of high capacity models such as DL. According to Faust *et al.* [102], although undesirable, this practice is widespread in ECG-based machine learning procedures. For avoiding this, 5-fold cross-validation subject-wise splitting was performed.

Single-lead and multi-lead prediction strategies were attempted to address the multi-view nature of ECG. When using single-lead annotations, the algorithm would





Step 4: Metrics

		P	QRS	T
Pr (%)		90.12	99.14	98.25
Re (%)		98.73	99.94	99.88
On (ms)	mean	-1.5	-0.1	21.6
	STD	22.9	8.4	66.3
Off (ms)	mean	0.3	3.5	5.9
	STD	16.0	12.6	31.1

Figure 2.2: Developed pipeline. Step 1: random shuffle split of subjects (green: train, purple: test) in 5 folds. Step 2: the (single- or multi-lead) dataset is either used for training with automatic labels or with manual annotations. In the latter case, ECG-tailored data augmentation is optionally applied (augmented). Step 3: a U-Net is instantiated per fold with a selection of execution parameters and trained with the selected data. Step 4: the fold-wise test sets are predicted and evaluated against the manual annotations. The employed metrics are precision and recall for detection and mean $\pm$ standard deviation (SD) of the onset and offset markers.

be inputted one lead at a time, producing a segmentation for every lead separately. For multi-lead prediction, a single mask would be generated when inputting all available leads as different channels. To alleviate data scarcity and low intra-recording beat variability, three different training strategies were attempted: training with manual data, semi-supervised pre-training with automatic labels and applying a custom DA over manual labels. Semi-supervised pre-training was performed by training the model from scratch using only data annotated with an algorithm in the literature, ECGpuwave [37], without DA. These decisions are schematized in the second step of Figure 2.2.

2.3.2. Data augmentation

Data augmentation improves a network's generalization by adding realistic noise sources to the input data, learning noise-insensitive representations [103], acting as a *de facto* regularizer. In this work, we developed six different ECG-tailored noise sources, computed to have a specific signal-to-noise ratio (SNR) with respect to an input signal, comprising additive white Gaussian noise (AWGN), random periodic spikes (RS), amplifier saturation (AS), powerline noise (PN), baseline wander (BW) and pacemaker spikes (PS):

- $AWGN[n] = \mathcal{N}\left(0, \sqrt{\tilde{P}_n}\right)$
- $RS[n] = \sqrt{\frac{\tilde{P}_n}{f}} \sum_{k=-\infty}^{k=\infty} (\delta * Sp)\left[n - k\frac{1}{f}\right]$
- $AS[n] = \begin{cases} -x[n] + S_v & if\, x[n] \geq S_v \\ -x[n] - S_v & if\, x[n] \leq -S_v \\ 0 & otherwise \end{cases}$
- $PS[n] = \begin{cases} \sqrt{\frac{\tilde{P}_n}{f}} & if\, n \in QRSon \\ 0 & otherwise \end{cases}$
- $PN/BW[n] = \sqrt{2\tilde{P}_n}\cos\left(\frac{2\pi f}{f_s} n\right)$,

Where $\mathcal{N}$ is the normal distribution, $\tilde{P}_n = P_s/10^{\mathrm{SNR}/10}$ is the noise power, P_s is the input signal power, f_s is the sampling frequency, $S_v = p\max|\mathbf{x}|$ is the saturation value, $(a * b)[n]$ indicates the convolution operation, $Sp = [0, 0.15, 1.5, -0.25, 0.15]^T + \mathcal{U}(-0.25, 0.25)$ is a custom filter with uniform noise that models pacemaker spikes and δ is the impulse function.

The first five noise sources were engineered to represent usual and observed variations in the dataset. Pacemaker spikes were designed to avoid misidentifying spike-like noise near QRS complexes and for completeness. Powerline and baseline noises share the same formulation but are instantiated with different hyperparameters ($f = 50Hz$ and $f = 0.5Hz$, respectively). Some noise in the generating hyperparameters was added upon generation for maximizing input variability, given $p^{(i)} = p + \mathcal{U}(\pm SNR/10)$, where $\mathcal{U}$ is the uniform distribution. Figure 2.3 depicts an example of the developed noise sources.

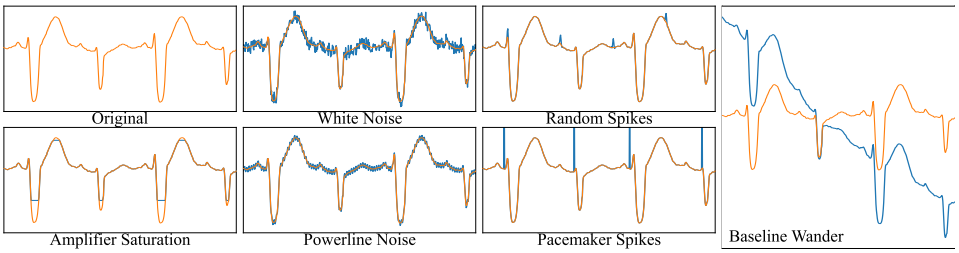


Figure 2.3: Data augmentation strategy example for an ECG recording. Re-execution results in slight signal-to-noise ratio (SNR) and frequency (f) variations, altering the final shape of the computed noise.

2.3.3. U-Net architecture

The employed architecture is based on the U-Net [80], which consists in an encoder, a bottleneck and a decoder with skip connections between the encoder and the decoder, as seen in Figure 2.4. The encoder extracts increasingly abstract representations of the input data through several levels of stacked convolutional operations and downsampling blocks. The decoder recovers information from the bottleneck, the nexus between the encoder and the decoder, through convolutional and upsampling blocks. Skip connections allow for crisper segmentation at the object boundaries by direct information transmission from the encoder. In the U-Net, the number of convolutional filters is doubled after each downsampling block and halved after each upsampling block. For a clearer exposition, we have grouped operations in the U-Net architecture into "blocks", which form "levels". We define a "block" as an ordered composition of operations on a tensor $\mathbf{x}$ and a "level" as a set of operations whose results have compatible tensor size. The considered blocks are convolutional, downsampling, upsampling, and skip connection blocks. The ordering of operations in the blocks were defined to agree with the image segmentation literature [72, 104].

In this work, convolutional (C) and separable convolutional (S) operations were considered, paired with ReLU non-linearities (NL) and regularizers (R) and pointwise additions $A(\cdot,\cdot)$. The following blocks were independently explored:

- "Vanilla" [80]: $\mathbf{y} = C(R(N(C(R(N(\mathbf{x}))))))$
- Residual [72]: $\mathbf{y} = A(C(R(N(C(R(N(\mathbf{x})))))),\mathbf{x})$
- XCeption [104]: $\mathbf{y} = A(S(R(N(S(R(N(\mathbf{x})))))),\mathbf{x})$

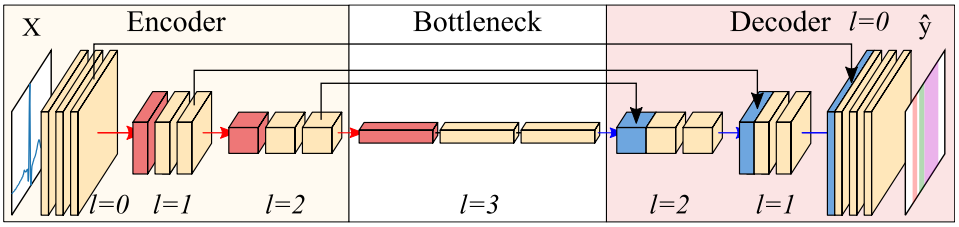


Figure 2.4: Base U-Net instantiated with 4 levels and 2 convolutional blocks per level. Blocks represent output tensors, whereas arrows indicate operations. Yellow: convolutions; red: pooling; blue: upsampling; black: concatenation. Convolutional operations extract $2^l N$ channels per level, whereas pooling and upsampling have a kernel size of 2.

To comply with the requirements posed by ECG data and the task at hand, the U-Net convolutions were replaced by 1D operations, zero padding was applied to keep input resolution, and a stem (one extra convolutional module right after the input) was included to mimic classification architectures [104–106]. Lastly, given an initial model testing phase, the need for stronger regularization than batch normalization was apparent. We opted to apply SDr [107], which randomly drops entire tensor channels during training, as opposed to standard dropout, where neurons are dropped in an unstructured manner, as well as semi-supervised pre-training and data augmentation, described in Sections 2.3.1 and 2.3.2, respectively.

2.3.4. Evaluation

The evaluation is inspired by the metrics used in state-of-the-art DSP-based algorithms [35] for comparison purposes. A correspondence matrix H of the correspondence between the true (w) and predicted ($\hat{w}$) wave fiducials can be computed as:

$$H_{ij}^{(r,\ell)} = \begin{cases} 1 \text{ if } (\hat{w}_{\text{fid}}^{(r,\ell)}[j] \in [w_{\text{on}}^{(r,\ell)}[i], w_{\text{off}}^{(r,\ell)}[i]]) \\ \quad \text{or } (w_{\text{fid}}^{(r,\ell)}[j] \in [\hat{w}_{\text{on}}^{(r,\ell)}[i], \hat{w}_{\text{off}}^{(r,\ell)}[i]]) \\ 0 \text{ otherwise} \end{cases}, \tag{2.2}$$

where r is the subject's recording, ℓ is the lead within the recording, w_{fid} (with fid $\in$ {on, peak, off} and $w \in$ {P, QRS, T}) are the onset, peak and offset information for a specific wave, and $i \in [0, M]$ and $j \in [0, \hat{M}]$ are the total true and predicted fiducials, respectively.

The individual lead information for a specific wave is then combined into a single correspondence matrix through a logical "OR" operator $\bar{H} = OR(H^{(0)}, ..., H^{(L)})$. The true positives (TP) for a given recording r were considered as $TP_r = \sum \bar{H}_{ij}$.

False-positives (FP), on their behalf, are elements of $\tilde{\mathbf{w}}_{\text{fid}}^{(r,\ell)}$ that did not correspond to any true fiducial ($FP_r = \hat{M} - card(\{(i,j) \mid \bar{H}_{ij} = 1\})$). Finally, a false-negative (FN) is considered when the ground truth displays a beat that is not captured by a TP (corresponding to $FN_r = M - card(\{(i,j) \mid \bar{H}_{ij} = 1\})$). The precision ($Pr$) and recall ($Re$) for each r and ℓ were computed, reporting in this work the overall performance for all recordings and leads for conciseness. Additionally, the F1 score is also computed for comparing different architectural variations, as a single figure of merit.

The delineation metrics were computed for the TPs (cases where $\bar{H}_{ij} = 1$), as no onset/offset correspondences between the GT and the prediction exist otherwise. The relative error of the segmentation was computed through the mean (M) and standard deviation (SD) of the difference of the actual and predicted onsets or offsets of the correspondences found in Eq. 2.2:

$$\min_{i,j,\ell} \; w_{\text{fid}}^{(r,\ell)}[i] - \tilde{w}_{\text{fid}}^{(r,\ell)}[j] \qquad \text{s.t. } \bar{H}_{ij} = 1. \tag{2.3}$$

For comparison purposes, the delineation errors were compared to the within-dataset bias, in the shape of two different sources of error: the inter-observer and the inter-lead variability. The first accounts for the difference in criteria used by the first observer (O_1) and the second (O_2) when annotating a sub-set of the input data. The latter accounts for the difference in criteria used by one observer when delineating morphologically similar beats. For this purpose, a running cross-correlation (normalized so that autocorrelations at displacement 0 equal 1) between all pairs of delineated waves was computed for all possible overlapping positions. Those pairs that shared a 99% or higher similarity were marked as true positives (for the comparison) and accounted for the delineation error. For producing a fair comparison, a window of 40 ms in the onset and offset of the compared delineations was added.

2.3.5. Experiments

For testing the model's robustness, this work features a series of variations on the data level and in the network's topology. Data-level variations aimed at alleviating data scarcity through the application of in-built SDr regularization ($p = 0.25$) and batch normalization (Section 2.3.3), DA strategies (Section 2.3.2) and semi-supervised pre-training (Section 2.3.1) for both single- and multi-lead inference strategies. Given our limited computational budget, no DA was applied during semi-supervised pre-training.

For each of the data-level variations, a set of topological changes were independently tested. These changes took shape in the type of convolutional block em-

		Single-Lead	Multi-Lead	Martínez [35]	Camps [77]	Sodmann [79]		
						10	50	150
P wave	Pr	90.12	94.17	91.03	N/A	79.6	84.6	90.0
	Re	98.73	94.70	98.87	N/A	86.8	92.2	98.1
	OnE	1.54 ± 22.89	-1.72 ± 17.83	2.0 ± 14.8	N/A	N/A	N/A	N/A
	OffE	0.32 ± 15.99	4.01 ± 16.08	1.9 ± 12.8	N/A	N/A	N/A	N/A
QRS wave	Pr	99.14	99.40	99.86	N/A	93.0	98.5	99.9
	Re	99.94	99.28	99.80	N/A	92.2	97.7	99.1
	OnE	-0.07 ± 8.37	-3.83 ± 14.64	4.6 ± 7.7	-2.6 ± 10.8	N/A	N/A	N/A
	OffE	3.64 ± 12.55	5.39 ± 16.77	0.8 ± 8.7	4.4 ± 15.2	N/A	N/A	N/A
T wave	Pr	98.25	96.36	97.79	N/A	80.2	87.4	97.7
	Re	99.88	99.09	99.77	N/A	80.7	87.9	98.3
	OnE	21.57 ± 66.29	19.10 ± 66.51	N/A	N/A	N/A	N/A	N/A
	OffE	4.55 ± 31.11	9.93 ± 46.33	-1.6 ± 18.1	N/A	N/A	N/A	N/A

Table 2.1: Metrics of our best performing single-lead and multi-lead models as compared to other approaches. The considered metrics are the precision (Pr, %), recall (Re, %), onset error (OnE, milliseconds in mean $\pm$ standard deviation) and offset error (OffE, milliseconds in mean $\pm$ standard deviation). Sodmann [79] reports results for different window sizes (10, 50, 150 milliseconds), considering a true positive if their prediction is contained within the window. N/A stands for "not applicable".

ployed ("vanilla", residual, XCeption), the network's depth ($L \in [4,5,6,7]$) and the number of convolutional blocks per level ($CB \in [2,3,4,5,6]$) were attempted. In total, 201 executions were performed testing various configurations, with training times ranging from 6 hours for the smallest models without pre-training to several days. The executions were performed in a high performance computing environment where each configuration was assigned to a single NVIDIA 1080Ti or NVIDIA Titan Xp GPU. To ensure reproducibility, the same random seed was employed in all executions. Some aspects were kept constant in all executions, such as the nonlinearity (ReLU for all blocks and sigmoid for the last block), convolutional kernel (3) and pooling (2) sizes, loss function (Jaccard), optimizer (Adam [64]) and random seed (1234).

2.4. Results

2.4.1. Model selection

This section describes the performance comparisons of independent design decisions tested in Section 2.3.5. The pipeline mainly benefited from the application of SDr regularization approach, reaching improvements of 1.98%, 3.27% and 7.21% F1 score in the detection of the P, QRS and T waves, and a reductions of $-2.79 - 0.83$ ms, -0.73 ± 1.43 ms and $8.06 + 3.90$ ms in onset error and of $-2.07 - 1.18$ ms, $+1.88 \pm 2.90$ ms and $-2.81 - 1.32$ ms in offset errors in the

Spatial Dropout

Metric	ΔF_1	$\Delta OnE\,(M \pm STD, ms)$	$\Delta OffE\,(M \pm STD, ms)$
P wave	+ **1.98**%	$-2.79 - 0.83$	$-2.07 - 1.18$
QRS complex	+ **3.27**%	$-$**0.73**$-$ **1.43**	+**1.88**$-$ **2.90**
T wave	+ **7.21**%	+8.06 + 3.90	$-2.81 - 1.32$

Semi-supervised learning

Metric	ΔF_1	$\Delta OnE\,(M \pm STD, ms)$	$\Delta OffE\,(M \pm STD, ms)$
P wave	+ 1.85 %	$-$**0.18**$-$ **4.31**	$-0.64 - 2.51$
QRS complex	+ 1.07 %	+0.08 $-$ 0.74	$-0.23 - 1.56$
T wave	+ 1.65 %	+**1.53**$-$ **3.88**	$-$**0.36**$-$ **4.87**

Data augmentation

Metric	ΔF_1	$\Delta OnE\,(M \pm STD, ms)$	$\Delta OffE\,(M \pm STD, ms)$
P wave	+ 1.22 %	+0.86 $-$ 2.10	$-$**0.42**$-$ **3.23**
QRS complex	+ 0.63 %	+0.26 $-$ 0.90	+0.61 $-$ 1.01
T wave	+ 0.80 %	+3.70 $-$ 0.39	$-0.90 + 3.63$

Table 2.2: Performance gain comparisons of applying spatial dropout, semi-supervised learning and data augmentation, expressed as median difference values in F1 score (%), onset and offset error (OnE and OffE, ms). A positive F1 score indicates a performance increase of the design decision, whereas negative onset/offset mean or STD errors indicate more precise fiducial location. Bold values represent best independently performing approaches.

P, QRS and T waves, respectively. Such generalized improvement is also seen in semi-supervised pre-training and when using DA. Summarized results can be visualized in Table 2.2. The model performance degraded consistently at higher capacity models (6 and 7 levels of depth and over 4 blocks per level) whenever SDr was not applied, but performed very similarly to other model definitions when it was. Other additions such as the type of convolutional block, width and depth of the network showed comparable performance throughout all executions.

2.4.2. Best performing model

Both single-lead and multi-lead best performing models feature strong regularization techniques in the shape of pre-training with automatic labelled data and SDr of 25%. The best performing single-lead model, in accordance to the results expressed above, consists in a model with 5 levels and 3 blocks per level employing the "vanilla" convolutional block, with P, QRS and T wave precisions of 90.12%, 99.14% and 98.25% and recalls of 98.73%, 99.94% and 99.88% for detection. The delineation performance shows errors of 1.54 ± 22.89 ms, -0.07 ± 8.37 ms and 21.57 ± 66.29 ms in the onset and of 0.32 ± 4.01 ms, 3.64 ± 12.55 ms and

4.55 ± 31.11 ms in the offset for delineation, with Dice scores of 88.99%, 92.05%, 88.40% for the P, QRS and T waves, respectively.

The best multi-lead model features 4 levels and 6 blocks per level employing the "XCeption" convolutional block, reaching P, QRS and T precisions of 94.17%, 99.40% and 96.36% and recalls of 94.70%, 99.28% and 99.09% for detection. The delineation performance deviated from the ground truth 1.54 ± 22.89 ms, 1.54 ± 22.89 ms and 1.54 ± 22.89 ms in the onset and 4.01 ± 16.08 ms, 5.39 ± 16.77 ms and 9.93 ± 46.33 ms in the offset, reaching Dice scores of 88.19%, 92.14%, 89.33% for the P, QRS and T waves, respectively. The optimal network configuration for both single- and multi-lead is detailed in Table 2.1, whereas Figures 2.5-2.7 depict several samples from the single-lead and multi-lead approaches.

2.4.3. Inter- and intra-observer bias

The comparison of the different sources of bias within the database have been summarized in Table 2.3. The inter-observer bias shows larger or comparable bias for both single- and multi-lead scenarios in the QRS onsets (3.84 ± 14.17 ms vs. -0.07 ± 8.37 ms and -3.83 ± 14.64 ms) and offsets (2.74 ± 16.94 ms vs. 3.64 ± 12.55 ms and 5.39 ± 16.77 ms). The T wave is generally delineated more precisely by the human operator in its onsets (-9.52 ± 44.85 ms vs. 21.57 ± 66.29 ms and 19.10 ± 66.51 ms) and offsets onsets (5.84 ± 39.84 ms vs. 4.55 ± 31.11 ms and 9.93 ± 46.33 ms).

The intra-observer bias, on the other hand, is consistently larger in the database as compared to the model's predictions in the P onset (2.26 ± 42.93 ms vs. 1.54 ± 22.89 ms and -1.72 ± 17.83 ms), P offset (7.76 ± 23.28 ms vs. 0.32 ± 15.99 ms and 4.01 ± 16.08 ms), QRS onset (2.56 ± 22.95 ms vs. -0.07 ± 8.37 ms and -3.83 ± 14.64 ms), QRS offset (2.02 ± 21.9 ms vs. 3.64 ± 12.55 ms and 5.39 ± 16.77 ms) and T onset (-51.96 ± 105.88 ms vs. 21.57 ± 66.29 ms and 19.10 ± 66.5 ms1) for both single- and multi-lead scenarios, respectively. The T offset in the multi-lead prediction strategy is the only metric that shows worse performance with respect to the inherent dataset bias (4.53 ± 42.71 ms vs. 4.55 ± 31.11 ms and 9.93 ± 46.33 ms), and is within 1 sample (4 ms) difference.

2.5. Discussion

Deep learning techniques show improved performance upon classical approaches for supervised tasks given sufficient training data [108]. These models can be used to improve and automate tasks in the medical domain such as image (or signal)

Inter-observer bias

	Onset Error		Offset Error	
	M (ms)	SD (ms)	M (ms)	SD (ms)
P wave	N/A	N/A	N/A	N/A
QRS wave	3.84	14.17	2.74	16.94
T wave	-9.52	44.85	5.84	39.84

Intra-observer bias

	Onset Error		Offset Error	
	M (ms)	SD (ms)	M (ms)	SD (ms)
P wave	2.26	42.93	7.76	23.28
QRS wave	2.56	22.95	2.02	21.9
T wave	-51.96	105.88	4.53	42.71

Table 2.3: Onset and offset errors (mean, M $\pm$ standard deviation, SD) of the inter- and intra-observer bias within matching delineations in the ground truth. No P wave annotations were produced by the second observer.

segmentation, aiding in clinical decision-making [109]. Moreover, these models can be re-trained on newly acquired data to produce a positive feedback loop that enhances performance upon usage.

Under this context, this work presents a FCN-based approach for ECG delineation by framing the problem as a segmentation task. Our work exhibits good detection and delineation performance, with metrics comparable to DSP-based methods while presenting competitive advantages and increased performance over ML-based works. A summary of the performance comparison to other approaches can be seen in Table 2.1. The network has an excellent detection performance, in both single-lead and multi-lead scenarios, even if trained on a small dataset. The model was thoroughly explored in its hyperparameters to assess its performance under different training conditions. Many of these model variations showed inconsistent performance gains. Only explicit regularization strategies such as SDr, pre-training on automatic labels and DA consistently improved overall detection and delineation performance, increasing F1 scores and decreasing onset and offset errors. These variations demonstrated good detection and delineation performance in all executions.

We have compared our method with state-of-the-art methods in either DSP and ML methods. DSP-based approaches, such as Martínez *et al.* [35] provide a high delineation performance and are considered state-of-the-art. Our best performing approach performed on par with these methods in detection (Table 2.1, [35]) with differences in precision and recall lesser than 1%. Our model, however, produced higher delineation errors compared to the state-of-the-art, especially in the T wave

(Table 2.1). The multi-lead approach showed consistently worse performance at delineation, with up to 15 ms difference in the T wave offset, which can be explained in two ways. Firstly, our approach is data-driven, so any bias in the QT database will be learnt by the network. Therefore, the inter- and intra-observer bias were computed (Table 2.3), demonstrating that our error was below in general (or comparable in T offset) to the intrinsic data variability. Some examples of extreme intra-observer criterion differences are shown in Figure 2.7. Moreover, it is noteworthy that one mere sample difference is equivalent to 4 ms error, so large errors are equivalent to small sample differences.

Secondly, the results reported by [35] do not allow for testing generalizability on other datasets, as they adjusted their algorithm to produce globally minimal metrics on the whole dataset without using a separate test set, thus their method's generalizability remaining untested [12]. A fair numerical comparison between the approaches is challenging, since the proposed DL-based method produces metrics on data that the algorithm has not been previously trained on (cross-validation) and, therefore, no global error minimization is made. The used training method, however, hints at the ability of our algorithm to generalize on unseen data –which is a more desirable scenario.

Overall, our approach provides comparable results to DSP-based methods, in spite of being trained on small amounts of highly biased annotations, while applying cross-validation with strict subject-wise splitting for ensuring generalization and obtaining smaller delineation errors than those in the dataset. On the other hand, DSP-based algorithms require laborious rule re-calibration when extended to other morphologies, which is heavily time consuming. In this sense, DL-based approaches such as the proposed model can more easily assimilate newly annotated data to enhance delineation performance and shorten development once the right design decisions have been modeled, thus arising as an alternative to DSP-based methods with great potential.

In contrast with DSP-based approaches, DL methods use a variety of architectures, providing a good framework for comparison. It is noteworthy, however, that the majority of the compared literature does not detail how the train/test splitting is made, leading to potentially misleading model performance, as noted in Section 2.3.1. In general, the proposed method clearly outperforms all other data-driven approaches found in the literature, obtaining clearly higher detection values and lower delineation errors. In the following paragraphs, the different works are analyzed, grouped by type of network.

Although limited bibliography of CNN-based delineation methods exists, these compare unfavourably to FCN-based approaches. Camps *et al.* [77] delineated

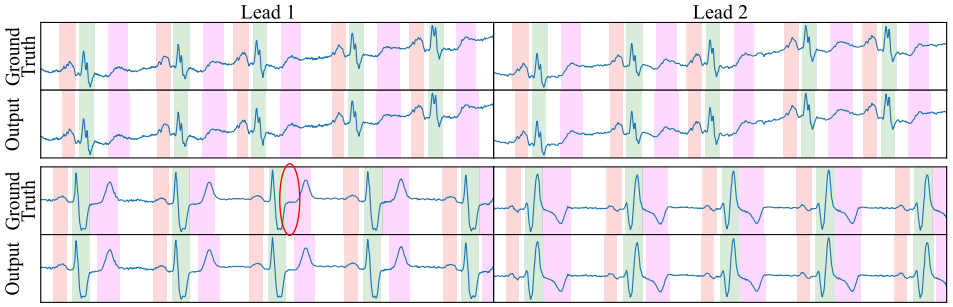


Figure 2.5: Examples of correctly predicted samples, depicting samples from sudden cardiac death (top) and ST change (bottom). Red mask: P wave. Green mask: QRS wave. Magenta mask: T wave. Representative examples have been encircled.

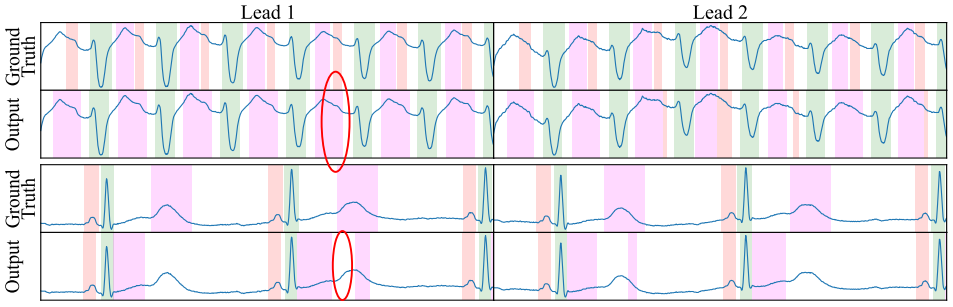


Figure 2.6: Examples of incorrectly predicted samples, featuring fused T and P waves (top) and severe bradycardia (bottom). Red mask: P wave. Green mask: QRS wave. Magenta mask: T wave. Representative examples have been encircled.

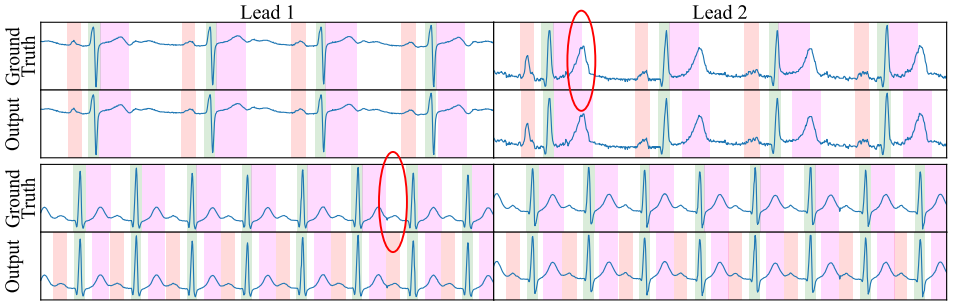


Figure 2.7: Examples of incorrectly annotated ground truth, demonstrating incorrect T offset location (top) and missed P waves (bottom). Red mask: P wave. Green mask: QRS wave. Magenta mask: T wave. Representative examples have been encircled.

solely the QRS wave, neglecting P and T waves, while attaining delineation performance of -2.6 ± 10.8 ms and 4.4 ± 15.2 ms for the QRS onset and offset. The authors did not report precision or recall metrics, difficulting direct performance comparison. Sodmann *et al.* [79] directly predicted the fiducial's sample of occurrence (w) through fully convolutional layers. However, their work suffers from performance pitfalls, achieving differences in performance up to 10% with respect to DSP-based approaches even with large ($\sim$50 ms) tolerance windows, while disregarding detections with error higher than 250 ms. Moreover, the authors excluded 23 recordings of the QT database.

A single recurrent formulation employing LSTM has been proposed in the literature by Abrishami *et al.* [78]. However, their work featured relatively low precision for the QRS and T waves (94% and 90%, respectively) and overall low recall (91%, 94% and 91% for the P, QRS and T waves, respectively), and the authors did not report delineation performance metrics. This work, however, merits from having performed subject-wise splitting.

Lastly, Tison *et al.* [83] published a U-Net based model for the delineation of 12-lead ECGs, similar to our initial attempt published in [81]. Tison *et al.* presented an asymmetric U-Net which featured an appended structure at the base level for producing 12-lead fusion and direct 8-fold upsampling from level 5 to level 2 in the decoder. The authors reported a high Dice score (P wave: 91 ± 3 %; QRS wave: 94 ± 4 %; T wave: 92 ± 5 %). The authors, however, employed a private ECG database, discarded recordings with large errors in a downstream task and heavily rely on HMM-based post-processing for refining the results. Moreover, their work is restricted to sinus rhythm recordings, compromising its generalizability on harder-to-delineate pathological beats. Although direct comparison is difficult in this case, our network needs no post-processing, has been tested against a standard database and was based on a well-founded architecture.

Besides the competitive delineation performance, we learnt several lessons for processing ECG data with DL-based techniques. Firstly, when working with ECG data, strong regularization techniques such as SDr and DA are of utmost importance, as the network easily overfitted the training data on the first epoch while stagnating the validation loss. Secondly, semi-supervised pre-training tags provides largely increased performance, plausibly due to the larger input data variability. Given the scarcity of manually annotated data in the QT database, automatically annotated data can give sufficient samples to learn better abstractions, acting as a regularizer via data [110]. Finally, the application of ECG-based DA methodologies seemed to increase overall performance of the network, when access to a larger database or to low-quality labelling is not possible.

An especially interesting result is drawn from the comparison of the model performance when producing single-lead and multi-lead predictions. The multi-lead fiducial computation suffers 1.89% and 4.03% drops in T wave precision and P wave recall, as well as large (up to 15.22 ms of difference) in onset and offset standard deviations. These gaps are partially due to the employed evaluation method, which compares the ground truth to the best predicted fiducial, irrespective of the lead on which it has been produced. This methodology, adapted from [35] for comparison with DSP-based methods, is a double-edged sword: while it decouples the performance of the delineation to the specific lead fusion strategy, it also masks the error it would inevitably produce. A second reason for the difference in performance is that the multi-lead scenario has half the samples to learn a representation from an input space that is doubly as large, as two leads are used as input.

A last point raised by the results, the model model performance degradation on higher capacity models, can be explained by the great imbalance between the network's capacity and the small amount of annotated data. This imbalance makes it more plausible to fall into local minima without stronger regularization techniques. In this sense, this work adds to the growing evidence that prior knowledge imposition, such as the applied regularization strategies (SDr, DA, pre-training) can be more effective than architectural modifications.

Although our network compares positively to other methods in the literature, many can be improved upon. One of the main limitations of this work is the lack of more up-to-date databases, containing a higher variability for a wider array of pathologies. Despite ECG usually being the first information registered of the patient's cardiac condition, not many large annotated databases for ECG delineation exist. This might partially explain the systematic errors in the T wave in a reduced amount of recordings, where the network predicted rises and falls as independent waves for very long T waves (Figure 2.6). With the current database, counting with 105 different represented ECG morphologies, our approach remains a proof of concept. This data scarcity could be alleviated through the application of clever data techniques, such as further semi-supervised on large, unannotated databases [111] or through realistic simulation of ECG traces.

Besides the apparent database-related shortcomings, some improvements could be made in the architecture. Temporal dependencies explicitly modeled with RNNs [112] or attention-based models [113, 114]. Efficiency-based modifications could also be explored, such as MobileNet [115], and model compression [116] for deploying the model in CPU-only computers. Finally, domain-specific prior imposition, model pre-training or alternative segmentation loss development would

further improve performance. Other DA schemes such as varying the heart rate, isoelectric line, specific wave shapes (e.g., voltage or width of P, QRS or T waves within a beat), DA hyperparameter tuning and composing new beats from components would also help, but the executions were made to keep an assumable computational budget.

2.6. Conclusions

Despite its potential, DL for cardiac signal analysis is not well established in the community [12, 47]. Some influencing factors are the lack of large-scale, quality databases (such as UK BioBank [117] in the imaging community), lack of digital support (many hospitals still print ECGs), lack of per-beat annotated data and high waveform variability due to, among others, pathological conditions, uncertainty in lead positioning, body composition, gender or noisy recordings. This work attempts at helping boost research in the signal-based cardiovascular field by providing measurements over clinical data, facilitating further downstream tasks by augmenting clinical decision-making without providing black-box diagnostic solutions. By bridging the gap between the imaging and the signal communities for cardiovascular diseases, we demonstrate that a DL model, properly trained and with an adequate objective function, can provide good delineation with good generalization. The existing limitations hinder, however, the application of this model into the clinical practice. Besides the need for accessing a larger pool of data, prediction efficiency and model compression constraints must be met for its embedding in clinical systems. Possible directions for expansion would be to apply more extensive semi-supervised pre-training or data augmentation methodologies.

Data Availability

The dataset is publicly available at `https://physionet.org/content/qtdb/1.0.0/`.

CHAPTER

3

Generalizing electrocardiogram delineation: training convolutional neural networks with synthetic data augmentation

3.1. Introduction

The electrocardiogram (ECG) is one of the main measurement tools in clinical practice given its rich insight into the cardiac electrophysiology, its ease of use and its relative inexpensiveness compared to other diagnostic methods. The ECG reflects the electrical activity of the heart, which can be logically grouped as a set of waves corresponding to different phases in the cardiac cycle. Thus, the P wave corresponds to atrial depolarization, the QRS corresponds to ventricular depolarization and the T wave corresponds to ventricular repolarization [118]. Extracting these waves (and their corresponding ST, PQ and TP segment pauses) allows the quantification of objective measurements of the heart's electrophysiological function, which can be used to characterize many pathological deviations from normal sinus rhythm (i.e., absence of P wave in ventricular rhythms or ST segment elevation/depression in myocardial infarction) [118]. Moreover, these measurements can be, in turn, employed in algorithms for diagnosis [119], either as clinical thresholds that indicate deviations from normality or in machine learning algorithms as extracted features for training and testing models [12]. In the case of the

This chapter is adapted from: **Jimenez-Perez, G.**, Acosta, J., Alcaine, A. & Camara, O. Generalizing electrocardiogram delineation: training convolutional neural networks with synthetic data augmentation. (*Submitted*).

ECG, accurately and automatically measuring the different waves could aid in the development of more precise decision support systems or monitorization tools by aggregating information in multiple-lead registries for several heart cycles, which is a highly time-consuming task that hampers the cardiologists' workflow [12].

Many computational approaches exist for the automatic quantification of the ECG. Most of these produce delineation of the electrocardiogram [35, 45, 53, 54, 77–79, 81–83]. Delineation methods can be divided in two main groups: digital signal processing (DSP) and machine learning (ML) based methods. The latter can be further subdivided into deep learning (DL) and non-DL (hereinafter "hand-crafted") methods.

Digital signal processing methods [35, 36, 45, 120] have the advantage of explicitly imposing priors on the biomarker extraction process, as they usually consist in a series of data-transformation steps (i.e., application of the wavelet transform (WT)) that reveal the cutoff points more clearly and a posterior rule-based algorithm for aggregating this partial information. These methods, however, often generalize poorly to unseen morphologies given their dependence on the production of robustly engineered transformation and rule-based aggregation steps [12], thus becoming more difficult to maintain.

Machine learning methods, on the other hand, have different associated problems that hinder their widespread adoption. Hand-crafted ML algorithms, in the shape of Gaussian mixture models [53] or hidden Markov models (HMM) [54], are difficult to train when using large amounts of annotated samples –which are becoming commonplace in current state-of-the-art– and usually provide reduced performance as compared to well-tuned DSP-based or DL-based solutions. The reason for this is that feature engineering, a key step in hand-crafted ML-based solutions, is costly and difficult to produce in a robust, fast and comprehensive manner [47]. DL-based methods [78, 79, 81–83, 85, 121], on their behalf, provide black-box solutions that are difficult to verify, require large amounts of annotated data, have difficulties leveraging *a priori* information and need quality loss functions for obtaining sensible data representations [12, 13, 60, 122]. Specifically, works found in the state-of-the-art employ convolutional neural networks (CNN) [77, 79], long short-term memory (LSTM) networks [78] and fully-convolutional networks (FCN) [81, 83, 85, 121]. However, some of these works solely delineate the QRS wave [77], whereas others only validate their performance on sinus rhythm [83, 85, 121] or show reduced performance compared to DSP-based approaches [78, 79, 81, 85, 121]. Moreover, both hand-crafted- and DL-based algorithms face difficulties when learning ECG data, given its high beat-to-beat morphological similarity and the small size of current ECG databases for their

usage in data-driven approaches.

For addressing the aforementioned issues with DL-based data analysis, a fully-convolutional network (FCN) model was trained on the QT [89] and Lobachevsky University (LU) [90] databases with a focus on direct data quantification, i.e., for producing ECG delineations. Using the U-Net [80] as the base architecture, some of the previously mentioned DL-based hindrances are addressed in three ways. Firstly, we developed a novel synthetic data augmentation (DA) method for increasing the database size with *a priori* information on normal and pathological ECG behaviour. Secondly, two loss functions were developed: the BoundaryLoss, which provide enhanced pixel accuracy close to the segmentation borders and is similar to other approaches in the literature [123, 124], and the F_1-InstanceLoss, which promotes cohesiveness in the predicted pixels regions. Lastly, we explored different modifications on the base architecture, namely different connectivity patterns such as the W-Net [125, 126], attention-based mechanisms [127] and different number of pooling operations. To the best of our knowledge, these improvements have not been explored in the literature for ECG analysis yet. A more rudimentary delineator has previously been published [82]; however, the current approach displays key components that allow the algorithm to generalize better against a wider array of morphologies, such as the application of the synthetic DA, the novel loss functions and a much wider array of architectural variation exploration.

The rest of the chapter is organized as follows. Section 3.2 describes the databases and methodology employed. Section 3.3 summarizes the main results. Finally, Section 3.4 discusses the obtained results in their context.

3.2. Materials and methods

This section firstly describes the used databases in Section 3.2.1 for then defining the methodology employed for their analysis. The proposed methodology, on its behalf, can be divided into several steps. The first step consists in the synthetic ECG generation from fundamental segments from a probabilistic rule-based algorithm (Section 3.2.2). The second step involved the definition and training of a DL architecture, which is subdivided into the description of the architecture itself (Section 3.2.3) and the employed loss functions (Section 3.2.4). Finally, the evaluation metrics are described in Section 3.2.5. A final section was added for detailing the specific experiments performed (Section 3.2.6). Our code will be made publicly available in `https://github.com/guillermo-jimenez/DelineatorSwitchAndCompose`.

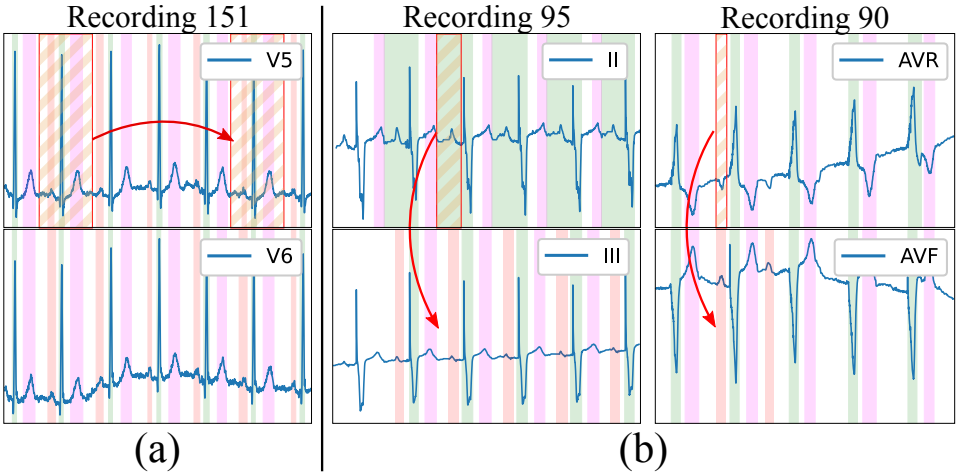


Figure 3.1: Limitations of existing delineation databases for training deep learning models. Examples in the LU database: a) High beat-to-beat redundancy within recordings; b) Incorrectly-annotated ground-truth (top lead) and correction (bottom lead). Color code: P wave (red), QRS wave (green) and T wave (magenta). Stripped segments highlight the errors.

3.2.1. Databases

The QT [89], the LU [90] and the database from the Ningbo First Hospital of Zhejiang University [128] (hereinafter the "Zhejiang" database) were employed for model training and evaluation. Specifically: the QT database was used for synthetic DA and model training; the LU database, for synthetic DA and evaluation; and the Zhejiang database, for model testing. The QT database contains 105 two-lead ambulatory recordings of 15 minutes at 250 Hz, representing different pathologies (arrhythmia, ischemic/non-ischemic ST episodes, slow ST level drift, transient ST depression and sudden cardiac death, SCD) as well as normal sinus rhythm. The LU database is composed of 200 12-lead recordings of 10 seconds of length, sampled at 500 Hz, comprising sinus and abnormal rhythms as well as a variety of pathologies. The Zhejiang database, on its behalf, includes 334 12-lead outflow tract ventricular arrhythmia (OTVA) recordings of variable size ($2.8-22.6$ seconds), sampled at $2,000$ Hz, and was originally devised for identifying the site of OTVAs, containing no delineation annotations. These databases are an appropriate sample for testing generalizability, since they present heterogeneity in their represented pathologies, sampling rates, lead configurations (Holter and standard 12-leads) and centers of acquisition.

The existing delineation databases have certain characteristics that hinder the de-

velopment of reliable delineation algorithms. On the first hand, although they contain a relatively big amount of delineated cardiac cycles (3,528 and 1,830 annotated beats in the QT and LU databases, respectively), these present a high intra- and inter-patient redundancy (i.e., very similar morphologies in different patients for certain pathologies or during normal sinus rhythm and very stable ECG beat--to-beat morphology in the same trace), which complicates model training due to reduced population variability (Figure 3.1a). Moreover, given the difficulty and time-consuming process of delineating an ECG, some registries present delineation errors such as skipped beats or inconsistent onset/offset predictions for similar morphologies, among others. Those problems were addressed in two ways. Firstly, those outlier beats were reannotated when necessary with the help of an expert cardiologist. Secondly, new ground truth (GT) was generated for the Zhejiang database, which was not annotated for delineation purposes, and reserved for algorithm testing as an independent set. Some examples of annotation corrections can be seen in Figure 3.1b.

The data and GT, either real or synthesized, were then represented as binary masks for their usage in DL-based segmentation architectures, where a mask of shape $\{0,1\}^{3\times N}$ was *True*-valued whenever a specific sample $n \in N$ was contained within a P, QRS or T wave (indices 0, 1 and 2, respectively) and *False*-valued otherwise [82], bridging the gap with the imaging literature. Finally, the joint training database was split into 5-fold cross-validation with strict subject-wise splitting, not sharing beats or leads of the same patient in the training and validation sets [82, 102]. Given that the proposed method employs synthetic DA, the generated ECGs were also generated using data uniquely from the training set for each fold, ensuring no cross-fold contamination.

3.2.2. Synthetic data generation

The structure of an ECG can be regarded as a combination of the P, QRS and T segments, alongside the PQ, ST and TP pause segments, which represent different phases of the electrical activation of the heart. The ECG is able to represent in its trace many pathological and non-pathological changes, reflecting slight deviations in its different constituting segments. The resulting "modular" structure can be leveraged in data-driven approaches for generating synthetic data.

The developed generation pipeline, depicted in Figure 3.2, consisted in two main stages: a pre-processing step that prepared the data for its posterior usage and a DA step that created synthetic ECG traces through composing independently generated cardiac cycles. The data pre-processing step, on its behalf, involved cropping the delineated GT (in this case, the QT and LU databases) in its con-

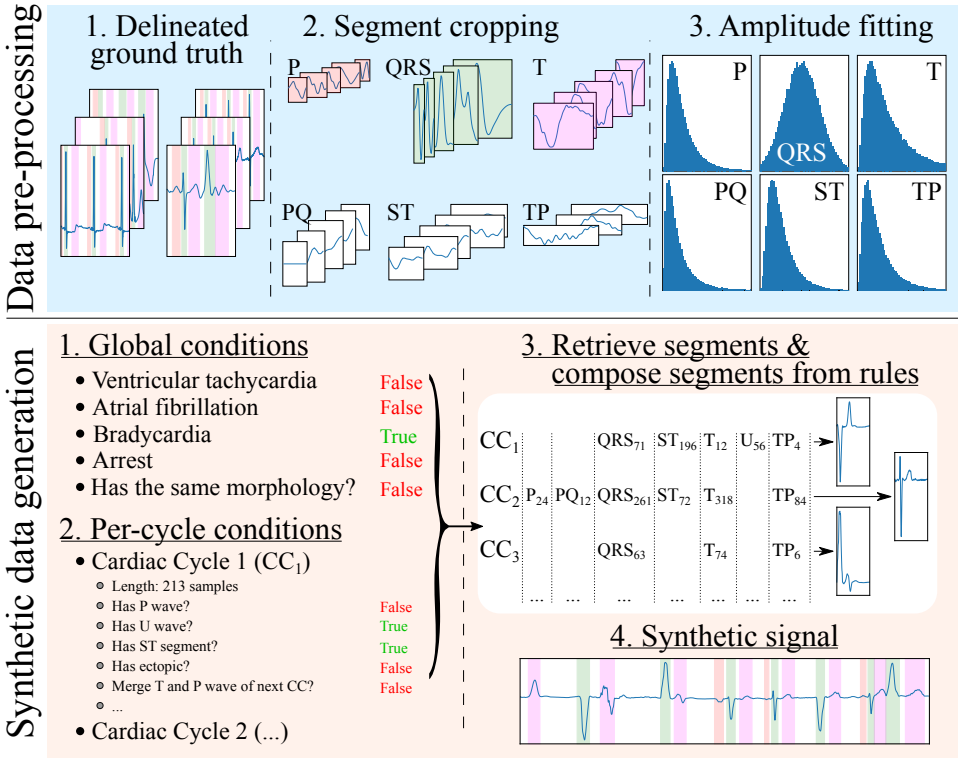


Figure 3.2: Synthetic electrocardiogam generation pipeline. The data pre-processing step consists in: (1) delineating the ground truth; (2) cropping the different beats contained in the ground truth into their constituent segments (P, PQ, QRS, ST, T and TP), normalizing the QRS segment to have an absolute amplitude of 1, and normalizing the rest of the segments' as the amplitude fraction with respect to their (normalized) relative QRS; and (3) fitting the amplitudes to a normal distribution for the QRS wave (fraction with respect to its original amplitude) and log-normal distributions for the rest of the segments (fraction with respect to the QRS' amplitude). The synthetic data generation step, on its behalf, involves (1) producing a set of global rules that will be common for all synthesized cardiac cycles (in the example, the registry has bradycardia), (2) producing a set of rules that will affect each cardiac cycle individually (in the example, the first cardiac cycle, CC_1, skips its P wave to simulate a ventricularly-mediated beat or a very low amplitude P wave), (3) retrieving the specific segments and their amplitude relationships from the "bags" of cropped segments for their composition into independent cardiac cycles and (4) concatenating the segments into the final synthetic trace.

stituent segments and into separate "pools" of segments from which to draw in subsequent stages. Additionally, the segment's amplitude (relative to their associated QRS) was fitted into independent log-normal distributions, which would be sampled from in the generation step to relate the amplitude of each segment to the amplitude of the QRS in each cardiac cycle. The QRS segment amplitude was normalized with respect to the maximum QRS amplitude in the whole registry (comprising all leads).

The synthetic DA step has several sub-steps. First, a set of global generation rules that affect all generated cardiac cycles were probabilistically generated for each sample. These have been limited to ventricular tachycardia (VT), atrial fibrillation (AF), atrioventricular (AV) blocks, sinus arrest (and its duration) and ST elevation/depression as a proof of concept. Second, a set of per-cardiac-cycle rules were generated, such as the presence or absence of each specific segment (P, QRS+T, PQ, ST, TP and U), whether the cycle corresponded to a ventricular ectopic (larger QRS amplitude and duration, absence of P wave) or whether there was wave merging (P with QRS, QRS with T, T with the next cycle's P). In the first and second steps, the rules were defined by drawing samples from a uniform distribution and applying the associated operation (global in the first case and per-cycle in the latter) in case they surpassed a pre-defined threshold.

Third, a set of segments were randomly selected from the segment "pools". A set of operations were then applied when extracting the segments from the pools as well as on the resulting cardiac cycles to comply with the global and per-cycle conditions. In particular, these operations comprised setting the segment's amplitude, interpolating the segment to a randomized number of samples to enforce as much variability as possible, cropping the segment, merging of the segment with the next (e.g., merging the T and the P waves, thus enforcing TP segment suppression), sign-correcting the segment to match other cardiac cycles or applying per-segment elevation/depression.

At last, the final synthetic signal and the GT were composed from the individual cardiac cycles. A set of post-composition operations were added to further increase the generated signal's variability, consisting in adding baseline wander noise, interpolating to slower or faster rhythms, adding flat line noise at the signal's edge, setting the global amplitude (multiplying the amplitude by a factor) and defining the trace's starting segment.

An important aspect to synthetic ECG generation is efficiency, as the samples were generated online rather than offline to avoid restricting the approach to a fixed set of previously drawn samples. This is, however, only relevant during the training phases of the model, but can limit the options of operations that can be performed

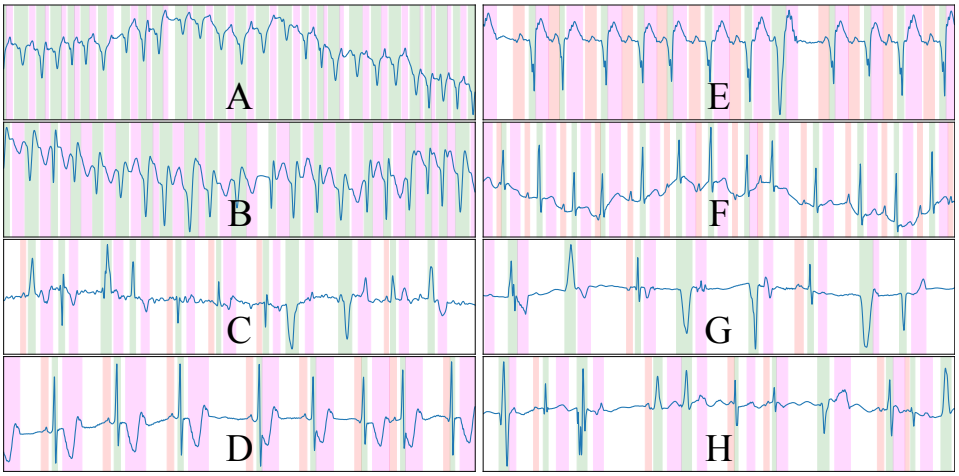


Figure 3.3: Randomly drawn samples from the developed synthetic data generator. The generator is able to produce samples of a variety of conditions such as ventricular tachycardia or atrial fibrillation, among others. The samples presented display ventricular ectopics (C, G, H), sinus rhythm (D, E, F), atrial fibrillation (C) and ventricular tachycardia (A, B), and are generated alongside their ground truth. Color code: P wave (red), QRS wave (green) and T wave (magenta).

on the algorithm; in fact, many of the chosen additions were limited in their scope by this constraint, being restricted sometimes to oversimplified operations that offer close-enough approximations of the underlying represented cardiac conditions. Some randomly drawn samples from the synthetic data generator are shown in Figure 3.3.

3.2.3. Architecture

The U-Net [80] is a CNN based on an encoder-decoder structure, as depicted in Figure 3.4. The encoder extracts high-level representations of the input data by means of convolutional operations, which transform an input tensor by convolving it with a trainable kernel, and pooling operations, which allow for reducing computational complexity. The decoder, on the other hand, upsamples the high-level encoder tensor to recover the original input's resolution while aggregating partial results obtained in different levels of the encoder. This direct feature aggregation between the encoder and the decoder, in the shape of tensor concatenation, allows for finer border definitions while avoiding gradient vanishing problems [80]. As in the original article, the number of trainable convolutional filters is doubled after every pooling operation and halved after every upsampling operation.

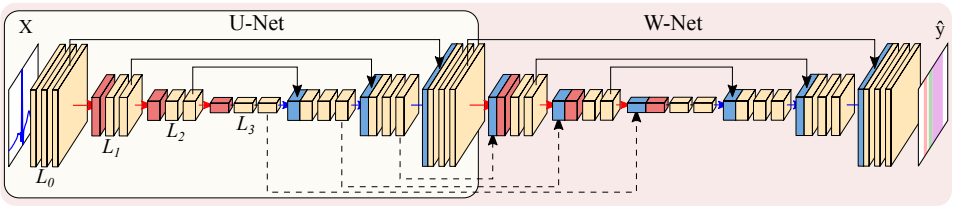


Figure 3.4: Representation of the U-Net (encircled in yellow) and W-Net architectures (encircled in red, containing the U-Net). Both networks are instantiated with 3 levels and 2 convolutional blocks per level. Arrows represent operations, while blocks are indicative of output tensors. Convolutional filters are doubled at each level, so that level L_i has $2^i N$ channels per level (with N being the starting number of channels), whereas pooling and upsampling have a kernel size of 2. Color code: convolutions (yellow), pooling operations (red), upsampling operations (blue), concatenation operations (black).

Many authors have experimented with the hyperparameters governing the U-Net, in the shape of number of convolutional operations (width), the number of upsampling/downsampling pairs (depth), starting number of convolutional filters, type of convolutional operation, type of non-linearity and presence/absence of other post-convolutional operations (batch normalization [129], spatial dropout [107]), among others, which was partially covered in [82] for the QT database.

Other authors have explored refining further architectural changes. Given the myriad of options, we restricted the exploration to the application of the W-Net architecture due to its good performance in other segmentation domains [130] as well as the usage of self-attention mechanisms in the shape of efficient channel attention (ECA). The W-Net [125, 126] involves the application of a second U-Net whose input is the output of the first U-Net, thus approximately doubling the amount of parameters for the same number of initial channels. The W-Net also concatenates the tensors at the decoder of the first U-Net with the encoder of the second, similarly to the connections established between the encoder and the decoder of a "vanilla" U-Net. This secondary structure makes the network deeper, which usually presents increased performance [106]. Self-attention applies the attention mechanism to a tensor, thus allowing different elements of the tensor to evaluate their relative importance for obtaining a certain result. This usually improves overall model performance and explainability [68]. ECA, specifically, is an approach to apply this mechanism to CNNs in an efficient manner [127].

3.2.4. Loss functions

Two novel loss functions, the BoundaryLoss and the F_1-InstanceLoss, were developed with the objective of enhancing the resulting prediction accuracy in two ways: the F_1-InstanceLoss enforces the retrieval of connected structures so that a penalty term is induced if the number of predicted and present structures differ; the BoundaryLoss attempts at adapting more tightly to the target boundary by means of computing the intersection-over-union of a subset of the original samples present in a mask, as opposed to the usual Dice score computation. These losses were based on the application of edge detectors, embedded into convolutional operations (flagged as non-trainable and with fixed weights), allowing automatic differentiation for posterior gradient propagation.

The first step consisted in applying the edge detector along all non-batch and non-channel axes of the input tensors, isolating the segmentation boundary. In the case of the BoundaryLoss, a large kernel size is employed ($K \in \mathbb{R}^n$, n being an hyperparameter), whereas in the F_1-InstanceLoss the kernel size remains small ($K \in \mathbb{R}^3$). In this case the Prewitt operator was employed as the edge detector, which is defined as:

$$\vec{K}_{F_1} = \begin{pmatrix} -1 & 0 & +1 \end{pmatrix}^T, \qquad \vec{K}_{Bound} = \begin{pmatrix} -1 & 0 & \dots & 0 & +1 \end{pmatrix}^T. \tag{3.1}$$

The second step took the absolute of the edge-detected tensors for both the predicted and the GT masks. In the case of the BoundaryLoss, the third step involved the calculation of the Dice coefficient between the resulting tensors. This has the advantage of comparing the mask overlap on a reduced pool of pixels, increasing the precision at the segmentation boundary, as is the case in usual image processing pipelines. In the case of the F_1-InstanceLoss, the third step was based on summing the border activations along each non-batch and non-channel axis separately for both the predicted and GT tensors, obtaining the number of discontinuities present in the binary mask. These discontinuities act as surrogates of the onset/offset pairs of the binary masks, thus allowing the computation of number of predicted and GT elements (P_{elem} and GT_{elem}, respectively) for their usage in usual precision and recall metrics in a fully differentiable manner. The true positives (TP), false positives (FP) and false negatives (FN) metrics are then computed by clamping these values, so that:

$$\begin{aligned} TP &= |GT_{elem} - max(GT_{elem} - P_{elem}, 0)| \\ FP &= max(P_{elem} - GT_{elem}, 0) \\ FN &= max(GT_{elem} - P_{elem}, 0). \end{aligned} \tag{3.2}$$

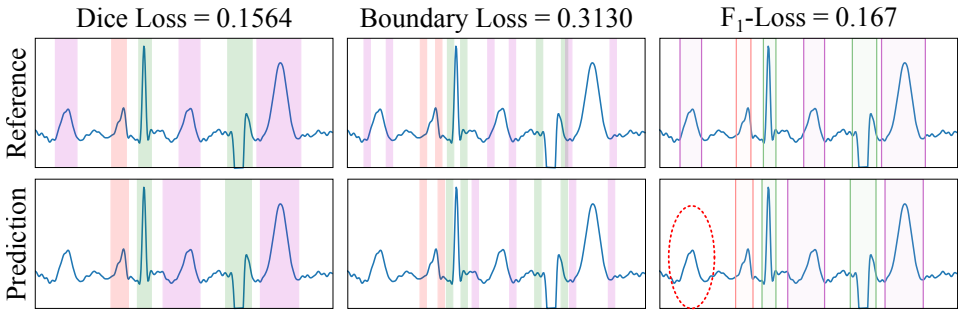


Figure 3.5: Example of loss functions applied to a sample from the LU database. The Dice loss measures the overlap between the ground truth mask (GT, top) and the predicted mask (bottom). The BoundaryLoss computes a secondary mask for isolating samples surrounding the boundaries of the GT and predictions, thus more specifically penalizing onsets/offset mistakes. The F_1-InstanceLoss locates the onsets/offsets pairs of the masks for using these as surrogates of precision and recall metrics. In the example, the ground truth contains three T waves, whereas only four T waves have been predicted; the F_1-score loss for each individual wave is, thus, 0, 0 and 0.167, respectively, given the false negative in the T wave. Color code: P wave (red), QRS wave (green) and T wave (magenta).

Finally, the TP, FP and FN values are then used to compute the smoothed F_1-score between the input and target masks. The computation process of these loss functions is depicted in Figure 3.5.

3.2.5. Evaluation

The model evaluation is based on the computation of detection metrics, i.e., the model's precision, recall and F_1-Score, and delineation metrics, i.e., onset and offset errors on the true positives (mean, M $\pm$ standard deviation, SD). The computation of the metrics consisted in three steps. Firstly, the onset and offset fiducials were retrieved from the predicted binary mask (described in Section 3.2.1) to express the sample of occurrence by retrieving the locations of value change (*False* to *True* or vice-versa). Secondly, the GT and predicted fiducials were matched through a correspondence matrix. Thirdly, the correspondence matrix was used to compute the detection and delineation metrics.

The correspondence matrices between the true (P, QRS and T) and predicted ($\hat{P}$, $\hat{QRS}$ and $\hat{T}$) fiducials were computed as:

$$
\begin{aligned}
P_{ij} &= \begin{cases} 1 \text{ if } \hat{P}_{\text{fid}}[j] \in [P_{\text{on}}[i], P_{\text{off}}[i]] \\ \quad \text{or } P_{\text{fid}}[j] \in [\hat{P}_{\text{on}}[i], \hat{P}_{\text{off}}[i]] \\ 0 \text{ otherwise} \end{cases} \\
QRS_{ij} &= \begin{cases} 1 \text{ if } \hat{QRS}_{\text{fid}}[j] \in [QRS_{\text{on}}[i], QRS_{\text{off}}[i]] \\ \quad \text{or } QRS_{\text{fid}}[j] \in [\hat{QRS}_{\text{on}}[i], \hat{QRS}_{\text{off}}[i]] \\ 0 \text{ otherwise} \end{cases} \\
T_{ij} &= \begin{cases} 1 \text{ if } \hat{T}_{\text{fid}}[j] \in [T_{\text{on}}[i], T_{\text{off}}[i]] \\ \quad \text{or } T_{\text{fid}}[j] \in [\hat{T}_{\text{on}}[i], \hat{T}_{\text{off}}[i]], \\ 0 \text{ otherwise} \end{cases}
\end{aligned}
\tag{3.3}
$$

where fid $\in \{$on, peak, off$\}$ is the specific fiducial to be explored, and $i \in [0, M]$, $j \in [0, N]$ are the total true and predicted fiducials for each of the waves, respectively.

These correspondence matrices were used to obtain the detection and delineation metrics. The detection metrics (TP,FP and FN) were computed as follows: given a correspondence matrix H, TPs were computed as elements that have been matched (TP $= \sum H_{ij}$); FPs were elements of a predicted fiducial that did not match any element in the GT, corresponding to the difference between the number of predicted fiducials and the cardinality of the matches (FP $= N - card(\{(i, j) \mid H_{ij} = 1\})$); and FNs were computed as elements of the GT that did not match any true fiducial, corresponding to the difference between the number of true fiducials and the cardinality of the matches (FP $= M - card(\{(i, j) \mid H_{ij} = 1\})$). The TP, FP and FN were in turn used to compute the model's precision (Pr), recall (Re) and F_1-score. The delineation error, on its behalf, was computed as the M $\pm$ SD of the difference of the actual and predicted onsets and offsets of the TP in the correspondence matrix:

$$
\min_{i,j} \; w_{\text{fid}}[i] - \tilde{w}_{\text{fid}}[j] \qquad \text{s.t. } H_{ij} = 1. \tag{3.4}
$$

These metrics were employed in turn for assessing the performance on the QT, LU and Zhejiang databases. In the case of the QT database, to homogenize the evaluation criteria with the existing literature, the detection and delineation metrics were computed for single-lead and multi-lead approaches, where the single-lead is based on evaluating the performance of both leads in the Holter registry independently, and the multi-lead consists in taking, for each beat, the lead that produces the best adjustment. Contrarily, the LU and Zhejiang databases were evaluated by fusing the individual lead predictions to obtain a single output prediction for,

posteriorly, comparing this delineation with the annotated GT. The final prediction was computed through combining the individual lead results using majority voting of the 12 leads and the different models resulting from training on separate folds of the QT database, forming an ensemble.

Finally, these metrics were also used to define the "best" performing model, which was selected as the one producing good detection performance while attaining the lowest possible delineation error for the QT (in the validation fold), LU and Zhejiang databases. This was addressed through the calculation of two figures of merit: the largest F_1-score as detection performance and the smallest SD of the error as delineation performance for all three databases across all waves, and reported in Section 3.3.1. Moreover, this model ranking was employed for producing ablations of the different modifications (Section 3.2) by isolating a single modified factor while leaving the rest of the hyperparameters unmodified. These have been reported in Section 3.3.2.

3.2.6. Experiments

The model's performance was tested under an array of complementary tests to address the contributions of the different elements to the results. Firstly, the importance of the synthetic DA was addressed by training the same model architecture using real and synthetic data (real+synthetic), synthetic data only and real data only. Identical computational budget was ensured by producing the same number of batches (with identical batch size) for the same number of epochs by oversampling the training database. Secondly, the importance of the BoundaryLoss and F_1-InstanceLoss was addressed also by doubling the number of executions, with and without the proposed losses. The Dice score always remained as a baseline for training in every configuration. Finally, the importance of the architectural modifications was addressed. Several architectures were tested: U-Net for depths 5, 6 and 7; W-Net for depths 5 and 6; and W-Net with ECA for depth 5. In all cases, the number of input channels was kept the same in the W-Net as in its U-Net counterpart, resulting in models with increased number of parameters (capacity). These were selected to have as many candidate architectures as possible but without compromising the computational budget of our equipment. In total, 66 different configurations were tested to address the model's performance.

Some design choices were kept constant to avoid unfeasibly large hyperparameter exploration. All model configurations used the same random seed (123456), leaky rectified linear unit (ReLU) non-linearities, zero padding for preserving tensor shape, kernels of size 3, batch normalization (BN), spatial dropout (SDr) [107] ($p = 0.25$), Adam optimizer [64] ($lr = 0.001$) and the Dice loss alongside the de-

		This work (SL)	This work (ML)	[82], SL	[82], ML	[35]	[77]
P wave	Pr	**99.27**	98.90	90.12	94.17	91.03	N/R
	Re	98.38	**99.72**	98.73	94.70	98.87	N/R
	OnE	-1.2 ± 17.9	**-0.8 ± 13.5**	1.5 ± 22.9	-1.7 ± 17.8	2.0 ± 14.8	N/R
	OffE	1.1 ± 16.6	**-0.6 ± 12.7**	0.3 ± 16.0	4.0 ± 16.1	1.9 ± 12.8	N/R
QRS wave	Pr	99.31	99.24	99.14	99.40	**99.86**	N/R
	Re	99.94	**99.97**	99.94	99.28	99.80	N/R
	OnE	-0.5 ± 11.2	**0.1 ± 7.5**	-0.1 ± 8.4	-3.8 ± 14.6	4.6 ± 7.7	-2.6 ± 10.8
	OffE	3.7 ± 13.1	**1.7 ± 7.8**	3.6 ± 12.6	5.4 ± 16.8	0.8 ± 8.7	4.4 ± 15.2
T wave	Pr	**98.73**	98.24	98.25	96.36	97.79	N/R
	Re	99.78	**99.97**	99.88	99.09	99.77	N/R
	OnE	5.8 ± 39.6	**5.2 ± 31.1**	21.6 ± 66.3	19.1 ± 66.5	N/R	N/R
	OffE	2.4 ± 51.3	3.8 ± 37.2	4.6 ± 31.1	9.9 ± 46.3	**-1.6 ± 18.1**	N/R

Table 3.1: Precision (Pr, %), recall (Re, %), onset error (OnE, mean [M] ± standard deviation [SD], in miliseconds) and offset errors (OffE, M ± SD, in miliseconds) of our best performing single-lead (SL) and multi-lead (ML) models in the QT database. N/R stands for "not reported".

veloped losses. The BoundaryLoss employed a kernel size of 11 samples. The ordering of operations after the convolutional operations was defined to agree with the image segmentation literature (non-linearity → BN → SDr) [72, 104]. All networks were trained using ECG-centered data augmentation, as described elsewhere [82], comprising additive white Gaussian noise (AWGN), random periodic spikes (RPS), amplifier saturation (AS), powerline noise (PN), baseline wander (BW) and pacemaker spikes (PS) to enhance the model's generalizability. All executions were performed at the Universitat Pompeu Fabra's high performance computing environment, assigning the jobs to either an NVIDIA 1080Ti or NVIDIA Titan Xp GPU, and used the PyTorch library [131].

3.3. Results

3.3.1. Best performing model

The best performing model according to the criteria presented in Section 3.2.5 was a self-attention W-Net model with 5 levels, trained with both real and synthetic data, while excluding the F_1-InstanceLoss and the BoundaryLoss. The model obtained an average F_1-score of 99.38% and a average delineation error of 2.19 ± 17.73 ms and 4.45 ± 18.32 ms for the onsets and offsets, respectively, across all waves and databases. The per-database and per-wave metrics of the model (precision, recall, onset error and offset error) are reported in Tables 3.1 and 3.2 for completeness.

		Zhejiang (this work)	**LU** (this work)	LU ([120])	LU ([36], *via* [120])	LU ([85])	LU ([121])
P wave	Pr	**97.57**	**99.62**	98.43	98.43	97.69	90.48
	Re	**98.65**	**99.81**	96.44	96.44	98.01	97.36
	OnE	**2.46 ± 12.58**	8.23± 9.01	**2.2 ± 7.4**	2.8 ± 7.5	-0.6 ± 17.5	3.4 ± 18.4
	OffE	**2.87 ± 12.43**	3.01±10.40	-6.5 ± 10.7	**-7.3 ± 10.1**	-2.4 ± 18.4	-4.1 ± 19.4
QRS wave	Pr	**99.53**	**100.00**	100.0	99.56	99.93	98.27
	Re	**99.87**	**100.00**	99.86	99.86	100.0	99.86
	OnE	**4.72 ± 13.35**	**4.27± 9.75**	15.4 ± 14.6	18.4 ± 14.7	1.5 ± 11.1	1.7 ± 10.0
	OffE	**3.26 ± 11.91**	**4.00± 9.14**	-3.8 ± 13.6	-5.4 ± 14.3	2.0 ± 10.6	-3.4 ± 12.3
T wave	Pr	**98.86**	**100.00**	99.21	99.09	99.37	96.23
	Re	**99.86**	**100.00**	98.85	98.85	99.68	93.51
	OnE	**8.73 ± 28.85**	18.26±18.21	**-1.3 ± 8.8**	-2.6 ± 11.4	2.9 ± 23.7	9.2 ± 28.2
	OffE	**-3.77 ± 24.32**	-8.84±18.05	**-1.2 ± 6.8**	-3.3 ± 7.3	-2.4 ± 30.4	-6.0 ± 25.0

Table 3.2: Precision (Pr, %), recall (Re, %), onset error (OnE, mean [M] ± standard deviation [SD], in miliseconds) and offset errors (OffE, M ± SD, in miliseconds) of our best performing model in the LU and Zhejiang databases, obtained through pixel-wise majority voting of the model developed for each fold trained on the QT database.

3.3.2. Performance comparison of model additions

The best performing additions were the usage of synthetic DA, where using both real and synthetic data reported an increased performance of 0.62% on average with respect to the usage of real-only data. Interestingly, using synthetic data only for model training still produced increased performance over using real-only data, surpassing its performance by 0.35% on average. Boxplots of the models grouped by data source can be visualized in Figure 3.6.

The second-to-best model performance addition was the usage of W-Net, which produced 0.53% less delineation error and a reduction in its SD of 1.83 ms and 2.20 ms for the onset and offset metrics as compared to its U-Net counterpart (see Figure 3.6). The third best model addition was the inclusion of the F_1-InstanceLoss and the BoundaryLoss functions, with added predictive performance of 0.26% F_1-score as well as reducing the offset error in 0.07 ms but increasing the onset error in 0.17 ms (Figure 3.6). The rest of the improvements (usage of self-attention, increase of model capacity) did not show a consistent effect on model performance.

3.4. Discussion

Deep learning commonly displays improved performance over many hand-crafted data analysis methodologies for a wide array of tasks [132], due to its ability to leverage large pools of data, its adaptability to a wide range of tasks, the built-in

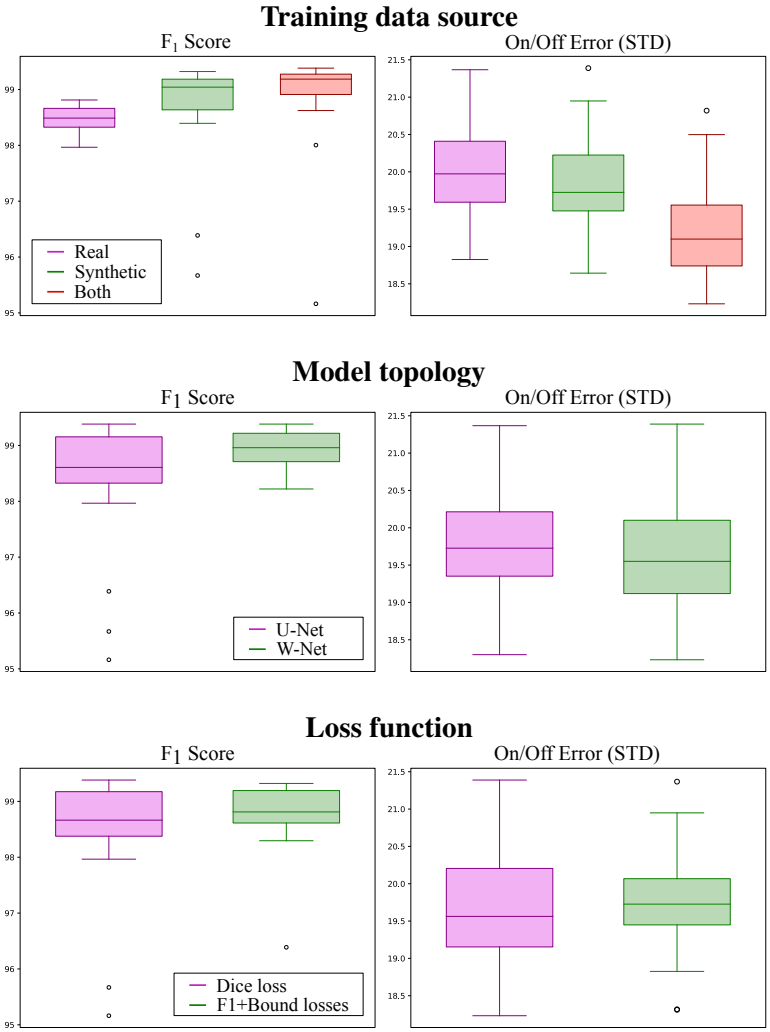


Figure 3.6: Detection (left; higher is better) and delineation (right; lower is better) performance of all models grouped by training data source, model topology and employed loss function. Synthetic-only data (green) showed higher detection and delineation performance than real-only data (magenta), whereas using both sources produced the best results for both detection and delineation performance. The W-Net (green) showed slightly higher detection and delineation performance than the U-Net (magenta). Finally, using the F_1-InstanceLoss and BoundaryLoss (green) resulted in models with higher detection performance but slightly lower delineation performance as compared to using Dice loss only (magenta).

feature engineering and the availability of of open code and large size datasets [60]. However, DL algorithms have a series of drawbacks that hinder their implementation in data-sensitive contexts. Firstly, they have a large dependence on the size of the training data [12], which might be difficult or expensive to acquire and annotate in many contexts such as in clinical environments. Secondly, DL models find difficulties when leveraging data priors, i.e., information that the system's designer knows must be implemented in the system. Some examples for the ECG would be the P wave (which might have imperceptible amplitude or might be masked within a QRS complex) or the fact that no T wave can exist without a QRS complex. Finally, the black-box nature of the models, where most works cannot guarantee that their predictions are not the result from spurious or acausal relations within the data or data leakage [133].

Some solutions exist for addressing these issues. Data scarcity has been addressed in the literature through pseudo labels [134] or through synthetic DA, either using simulations [96] or generative adversarial networks (GAN), but these present efficiency issues (simulations) or face difficulties when extending beyond the training data manifold (data-driven approaches). Data priors, on their behalf, have been enforced either by producing representations that explicitly exclude previously known information via minimizing mutual information [122] or by providing the specific prior as input data (e.g., by including the label as an input to the model, such as in conditional GANs [135]), but the ability to explicitly control data-side priors is still limited. Finally, some authors have attempted to enhance model explainability through filter visualization methods [136] or attention maps [68], but these can only act as surrogates of decision rules and need interpreting on their own. A middle ground for DL on data-sensitive contexts is to employ DL for quantification tasks, such as segmentation. These intermediate networks do not provide an instantaneous prediction (e.g., diagnosis), but present the advantage of being immediately interpretable by an operator [13], avoiding placing confidence in faulty predictions while enabling many downstream tasks [47]. This, however, contrasts with the current scientific production; as an example, in the specific case of DL-based analysis of ECGs, the vast majority of published works focused on classification, which is often praised by the clinical community [137]. Some authors have mentioned data scarcity, expensiveness of data annotation and lack of generalizability of existing solutions as the main hindrances for training DL models for data quantification [12, 13].

This work addresses these issues by producing a network for ECG segmentation. Given the small size of these databases, the models were enriched with a novel synthetic DA strategy, which allowed for imposing expert knowledge through constraining the topology of the generated data. These priors were further en-

forced in the shape of two novel loss functions by minimizing the boundary error with respect to the reference (BoundaryLoss) and by maximizing precision and recall metrics (F_1-InstanceLoss). Some approaches exist for ECG delineation [77, 78, 81–83] but, to the best of our knowledge, no approaches exist in the literature that combine a quantification task through explicit (rule-based synthetic DA) and implicit (application of the BoundaryLoss and F_1-InstanceLoss functions) prior imposition.

Performance-wise, the developed models compare favorably with existing DSP-based and ML-based approaches found in the literature. Our best performing model surpassed the state-of-the-art in the usual precision and delineation metrics. We obtained an average F_1 score of 99.38% and onset and offset errors of 2.19 ± 17.73 ms and 4.45 ± 18.32 ms with respect to the reference for all waves in the QT database, as detailed in Table 3.1. These figures represent a precision gain of up to 8% at some fiducials such as the P wave, recall values nearing 100% and an overall reduced SD, surpassing the best-performing methods in most categories. Some delineation metrics, however, present decreased performance as compared to DSP-based methods, which might be explained by their higher control on the decision boundaries that produce a certain prediction, as tailored to specific datasets; to the difficulties at localizing onsets and offsets in smoothly increasing/decreasing waves such as the P and T waves; and to the fact that, at the 250 Hz sampling rate, each sample accounts for a 4 ms difference.

Secondly, the trained model presented good generalization properties when predicting samples from the QT database (Table 3.1) as well as in the LU [90] and the Zhejiang [128] databases (Table 3.2). Specifically, the performance on the LU database surpasses all approaches in the literature and even displays F_1-scores of 100% for the QRS and T waves, which hint to both the robustness of our model and at the relative simplicity of the represented rhythms in the database. The performance on the more challenging Zhejiang database does not reflect reduced performance, providing metrics that closely resemble those of the QT database and, thus, hinting at the good generalizability of the developed model. A prediction from the Zhejiang database is depicted in Figure 3.7.

As presented in Section 3.3.2, the increased model performance can be directly related to the adopted design decisions, i.e., the addition of synthetic DA, the usage of the W-Net architecture and the implementation of the F_1-InstanceLoss and the BoundaryLoss. The inclusion of synthetic data supposed the best model addition, consistently improving the performance of the model when comparing models trained with and without synthetic DA. Interestingly, models trained exclusively with synthetic data also displayed better metrics than those trained only

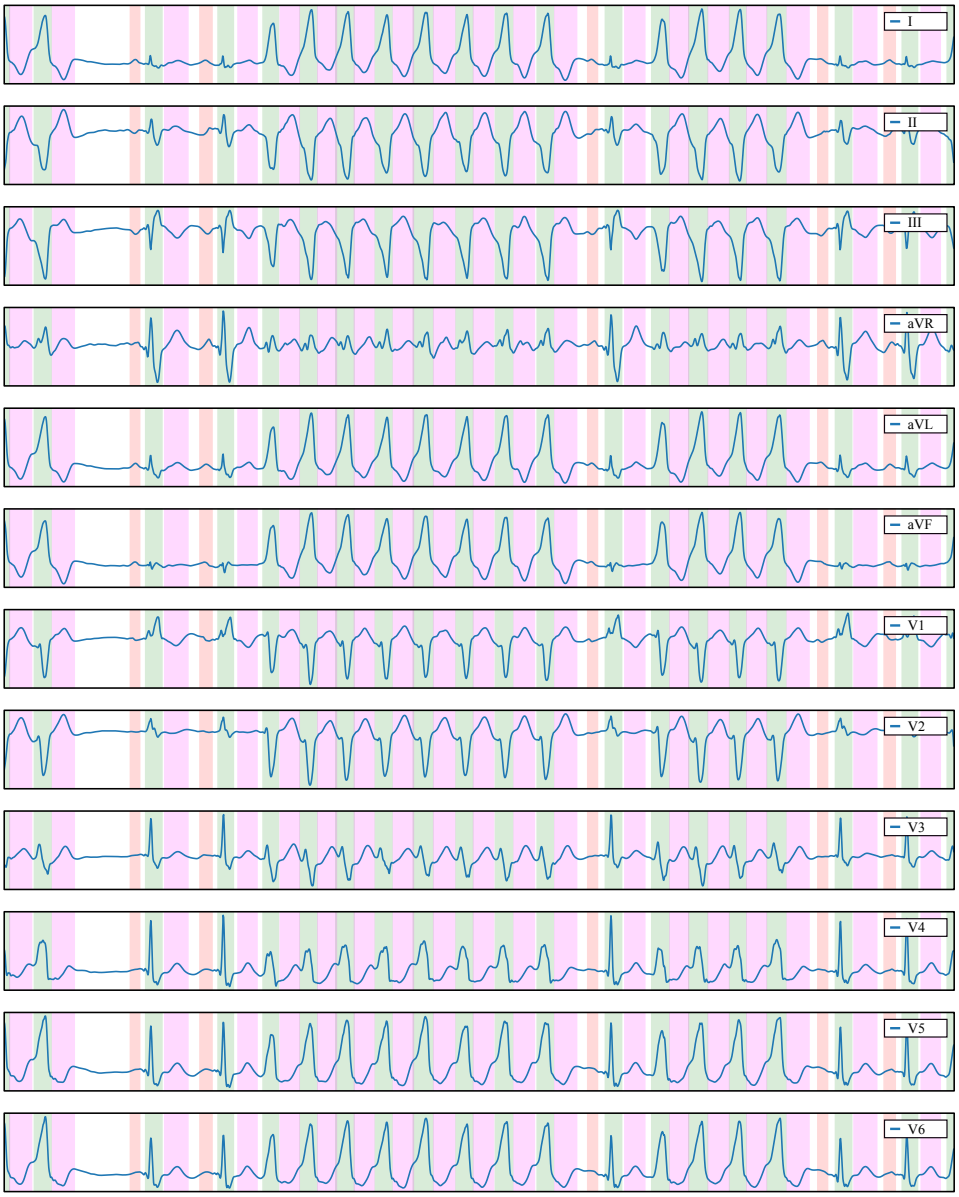


Figure 3.7: Delineation prediction of the sample “922551” of the Zhejiang database, containing a non-sustained ventricular tachycardia. Color code: P wave (red), QRS wave (green) and T wave (magenta).

with real data. The second-to-best model modification was the usage of the W-Net architecture, which can be linked to an increase in model capacity. The third best addition was the usage of novel loss functions. This addition, although it has been shown to improve model performance across all runs, was not present in some of the top models. Further research is needed to assess the mechanism these losses use to increase model performance. The specific F_1 and SD figures for all the aforementioned model additions have been reported in Section 3.3.2.

As it has been discussed, training a delineation model with synthetic DA demonstrates some advantages with respect to the state of the art besides purely enhanced detection and delineation performance. Firstly, it allows for explicit prior imposition in the synthetic DA process, a philosophy that can be extended to many other quantification domains (e.g., medical images or intracavitary electrocardiograms, as in chapter 4). Moreover, the synthetic DA pipeline addresses the lack of generalization of current existing delineation algorithms, reducing the impact of small databases and low inter-sample variance that has hindered both ML-based and DSP-based approaches [12]. An illustrative example is the good performance of the algorithm on the Zhejiang database, which contains rhythms that have been engineered into the synthetic DA but were not part of the training data, such as short bursts of VT. Finally, the predictions produced by the model were much more robust than other state-of-the-art approaches. For instance, some approaches displayed incoherent predictions, where the rise and fall of hyperacute T waves were predicted as two separate waves, completely missed waves, mistagged waves, misplaced onset/offset pairs or noisy activations spanning 1-2 samples [78, 82].

There are, however, some limitations to the presented approach. Firstly, the set of rules developed in the synthetic DA is too narrow. Many more conditions could be represented, and richer modifications over the fundamental ECG segments (cropped P, QRS, T waves) could be applied, such as addition of delta, J or epsilon waves, or atrial/ventricular hypertrophy. Secondly, the rise in computation time tied to the on-the-fly DA, alongside the existing computational and temporal constraints that are common in the DL literature when training large models, have limited our ability to provide exhaustive testing on the contributions of each element to the final result, especially due to the large amount of tunable hyperparameters. The synthetic generator has been employed with some hyperparameters that produced visually credible samples, but a rigorous validation is still lacking. Moreover, despite the efforts placed on generating VT records, and despite the success in a large percentage of predictions, the network still has difficulties finding the onsets and offsets of very fast VTs/ventricular flutter. This, however, is to be expected as even trained physicians have difficulties when delimiting said rhythms. Finally, the network is sensible to input normalization. Given that the

amplitude was normalized for sinus rhythm QRS to take values in the range $[0.5, 1]$ (thus taken larger values for other rhythms, such as extrasystoles), we have opted for normalizing the model's input with a moving average over the signal, with a window of 256 samples. This criterion could also be improved upon.

3.4.1. Conclusions

This work addresses some of the main challenges in ML-based clinical data analysis: the uninterpretability of classification-based models, the reduced database size and the imposition of data priors. For this purpose, we developed a DL-based pipeline for the automatic quantification of the electrocardiogram through novel prior imposition strategies in the shape of synthetic DA and shape regularization losses.

The produced network has demonstrated remarkable detection and delineation metrics, as well as good generalization when predicting a variety of samples of different open source databases. This allows its application to a large number of downstream tasks, allowing the production of automatic and objective metrics over clinical data, thus becoming an enabling technology for further automatization of ECG analysis. It, however, presents some limitations. Firstly, the synthetic DA produces a dependence on input data normalization when predicting input samples, although windowing and normalizing to the median usually performs well. Secondly, a larger array of cardiac conditions in the synthetic DA algorithm and a more in-depth exploration of the generative parameters should be performed. Finally, a more exhaustive testing of the performance gain of each model addition could be explored.

CHAPTER

4

Delineation of intracavitary electrograms for the automatic quantification of hidden slow conduction

4.1. Introduction

As mentioned in chapter 1, the heart is an electro-mechanical organ in which its mechanical contraction (pumping) is triggered by the transmission of an electrical impulse (conduction) throughout the myocardium. The effect of different cardiovascular diseases (CVD) are often reflected in the normal behaviour of the electrical conduction system of the heart (ECS), either caused by disruption of normal conduction patterns or being reflected in them. Thus, understanding any conduction deviations is a key tool for diagnosis, which gave rise to minimally invasive electrophysiology (EP) procedures. These interventions consist in the introduction of catheters, equipped with small electrodes at their tip, which are used to measure sums of myocardial action potentials on small areas in the endocardial or epicardial surfaces. These registries of intracavitary electrocardiogram (iECG) signals are the main diagnosis and treatment endpoint in procedures such as radio frequency ablation (RFA), as electrophysiologists interpret electrical activation patterns on key locations in the myocardial surface to assess whether a conduction

This chapter is adapted from: **Jimenez-Perez, G.**, Arana-Rueda, E., Frutos-López, M., Camara, O., Pedrote, A. & Acosta, J. A Deep-Learning-based electrogram delineator for Electrophysiology procedures. (*In Preparation*).

pathology might be responsible for the perpetuation of certain CVD, such as cardiac arrhythmias (CA) [3, 24].

The depolarization waves in iECG recordings are physically captured as a series of local activation patterns or electrograms (EGM). In normal cardiac tissue, these EGMs are visualized as isolated high-frequency deflections corresponding to a single passing of the depolarization wave. Existence of factors that alter the electro-mechanical properties of cardiac tissue (CVDs such as myocardial infarction, remodelling or genetic factors) can alter the morphology of an EGM and cause depolarization waves to conduct unevenly on the myocardial surface, which is captured by diagnosis catheters as fractionations in the local components (local fields, LF) or as late potentials (LP) in the EGM [20, 22]. Other conditions, such as atrioventricular reentrant tachycardias (AVRT) or accessory pathways (AP), perpetuate due to the existence of reentry circuits that bypass the normal sinoatrial (SA) node impulse origin. These conduction abnormalities are reflected in the morphology of LFs, which is often used as the main diagnostic markers during RFA.

Diagnosis of conduction abnormalities is performed by elucidating which portions of tissue show decremental properties or conduct electricity when they should not. Decrement occurs when local components are delayed with respect to a reference (e.g. the QRS complex) when stimulated at faster frequencies, as compared with the previous cardiac cycle. Decrement might be caused by (a) physiological structures such as the AV node, in which normal conduction entails decremental conduction at the node; and (b) anatomical substrates, where strands of partially viable tissue present decremental properties and unidirectional blocks that might lead to reentrant CA. Current clinical guidelines hint at the diagnostic value of decrement-evoked potentials (DEEP, also known as hidden slow conduction, HSC), which are portions of tissue presenting hidden decremental conduction which are uncovered using extrastimuli in specific locations in the myocardium [21, 28, 138]. Moreover, some conditions such as APs and AVRTs are diagnosed based on the presence or absence of decrement of atrial components as measured in the coronary sinus (CS) with respect to the ventricular components, effectively disambiguating between AV-mediated conduction or otherwise.

Despite the importance of localizing the distinct local EGM components, which could aid in localizing target ablation sites during AF interventions or to compare waveforms in VT procedures to assess the existence of decrement, current software solutions do not perform automatic signal quantification [22]. In simple procedures, the operators must manually perform interval quantification to assess the existence of decrement during pacing or extrastimuli protocols. Even 3D elec-

troanatomical mapping systems (EAM), the state-of-the-art solution for complex RFA procedures, only localize the LFs with the largest deflection within a cardiac cycle for the computation of diagnostic markers such as the local activation time (LAT) [22]. Moreover, the relative simplicity of existing quantification algorithms make them error prone; during RFA procedures, EAM operators must often reassign the detected earliest activation in many locations due to incorrect LAT identification, to avoid erroneous interpretations of the depolarization wave's direction [22]. Some computational solutions for the detection of individual waves exist [42, 139], but even fewer solutions exist for the delineation of local EGM activations [40, 140, 141]. Moreover, existing delineation solutions are based on digital signal processing (DSP) algorithms and suffer the same drawbacks as their electrocardiogram (ECG) counterparts (chapter 2): they are difficult to tune robustly and, given the case, might perform poorly in too complex signals. Moreover, these solutions concentrate on the delineation of the wave with the largest amplitude, as opposed to a full delineation of all EGM components.

In recent times, deep learning (DL) algorithms have gained popularity for the automated analysis of data, given their minimal pre-processing requirements and their high performance on a wide variety of tasks, surpassing the state-of-the-art in most data-driven analyses. In the specific case of cardiac signals, some solutions exist for the automatic quantification of ECG signals [78, 79, 81–83, 85, 121]. However, not many algorithms have been developed for analyzing iECG signals, and they revolve around classification [74]. In this work, and similarly to the setup used in chapter 3, a fully-convolutional network (FCN) was trained on a dataset of signals from the CS, retrieved from the Hospital Universitario Virgen del Rocío in Sevilla, Spain. The dataset was manually annotated for EGM delineation, localizing the onsets and offsets of all local components (LFs, FFs, stimulation artifacts). The U-Net [80] and W-Net [125] were used as base architectures alongside a specialized synthetic data generation method tailored towards iECG signals. Moreover, the Dice loss, which is the usually employed metric for training segmentation algorithms, was contrasted to the usage of novel losses that favoured a high sensitivity (as opposed to maximizing the F_1 score) and minimal delineation error [123, 124]. Finally, the addition of self-attention mechanisms and their effect on model performance was explored [127]. To the best of our knowledge, this is the first developed approach for iECG signal delineation, bridging the gap between the ECG and iECG communities.

The rest of the chapter is organized as follows. Section 4.2 describes the used databases, whereas and methodology employed. Section 4.4 summarizes the main results. Finally, Section 4.5 discusses and contextualizes the obtained results.

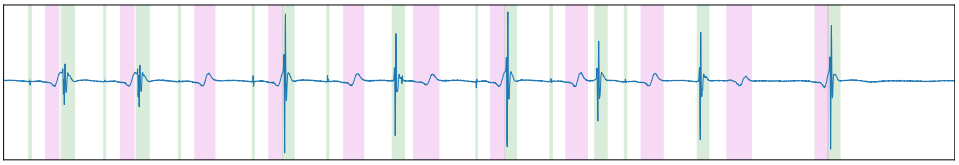


Figure 4.1: Generated ground truth for an intracavitary electrocardiographic recording at the coronary sinus. The green and magenta overlays represent, respectively, local field activations from the coronary sinus and the ventricular far field. The recording presents ventricular pacing and decremental properties.

4.2. Materials

A proprietary dataset for the automatic delineation of EGMs was developed in the Hospital Universitario Virgen del Rocío (Sevilla, Spain). This dataset comprises 312 iECG recordings of variable size taken from 77 distinct patients who underwent AP or AVRT ablation. The LF and FF activations were annotated independently, generating 20,671 local and 13,354 far EGMs. For enabling the prediction of stimulation onsets, 318 stimulation artifacts were annotated. All interventions used a decapolar catheter localized in the CS, which registers 5 bipolar EGMs: CS-P (proximal), CS-1, CS-2, CS-3 and CS-D (distal). The acquisitions were produced during pacing or application of extrastimuli, during which the decremental properties of the AV node were tested for diagnosis. The iECGs were captured with a BARD LABSYSTEM PRO EP Recording System© at 1,000Hz sampling frequency, encoded with 16 bits resolution (signed short int, range [-32,768, 32,767], 2.5 μV/bit resolution), and bandpass-filtered within [30, 500] Hz. Figure 4.1 shows an annotated EGM signal. The data was annotated using a custom-built system, written in Python and using the Bokeh library [142]. The annotation system employed chapter 3's ECG delineator to localize the P, QRS and T waves for aiding in visually discriminating between LF and FF activations during annotation. Moreover, chapter 3's model was also used for localizing the QRS complex's barycenter to find a common fiducial that allowed the isolation of a single cardiac cycle of a single electrode pair, which is used as input to the DL model during training and prediction. The RR' representation disambiguated better between LF and FF components (which might have similar frequential properties) and is the usual processing performed in EAM systems [29].

Due to the lack of externally annotated datasets in the literature and to time constraints, the same database was used for model training and evaluation, while ensuring no patient cross-contamination in different sets. Given the substantial intra-patient similarity (as discussed in chapter 3.2.1), where two consecutive beats of the same patient present high correlation, the dataset was split at the patient level,

comprising all bipolar EGMs of all acquired registries. Thus, 10% of the patients were selected for validation, 25% for testing and the rest for training (8, 20 and 49 patients, respectively). The training and validation sets were solely employed for synthetic data generation, whereas the real data contained in the test set was used for assessing the model's performance. Similarly to chapter 3.2.1, the annotations were represented as binary masks for their usage as optimization targets in the segmentation architectures, where a mask of shape $\{0,1\}^{3\times N}$ was *True*-valued whenever a specific sample $n \in [0,N]$ was contained within a stimulation, LF or FF activation (indices 0, 1 and 2, respectively) [82]. The value of N depended on the duration of the cardiac cycle.

4.3. Methods

This section describes the employed methodology for the analysis of iECGs. The analysis pipeline is very similar to the methodology exposed in chapter 3 with respect to the DL architecture and evaluation metrics. The main difference lies within the synthetic data generation pipeline (Section 4.3.1) and in the performed experiments (Section 4.3.3). The rest of the methodology is shared among the two approaches.

4.3.1. Synthetic data generation

As discussed in Section 1.1.5, the impulses recorded in the myocardial wall can represent local activations (local field, LF), which are captured as high-frequency components, and distant activations (far field, FF), which usually only retain their low-frequency components. Structurally, iECGs are composed of one or several LF activations (local potentials and any isolated LP), might contain a FF activation and segments of electrical silence (or rest) between them. Moreover, some segments might overlap, as is the case with LF/LPs and FF activations. Taking advantage of this modular structure, an algorithm for generating synthetic data was developed in this work. The algorithm has two major steps: data pre-processing, which crops the original signal into its basic components, and signal generation, which draws fragments from the independent set of segments and stochastically composes the signal according to some pre-defined rules. Figure 4.2 schematically represents the synthetic data generation pipeline.

Data pre-processing

As in chapter 3, data pre-processing involved a series of steps on the annotated data. Firstly, the ground truth is cropped in its fundamental segments, separating

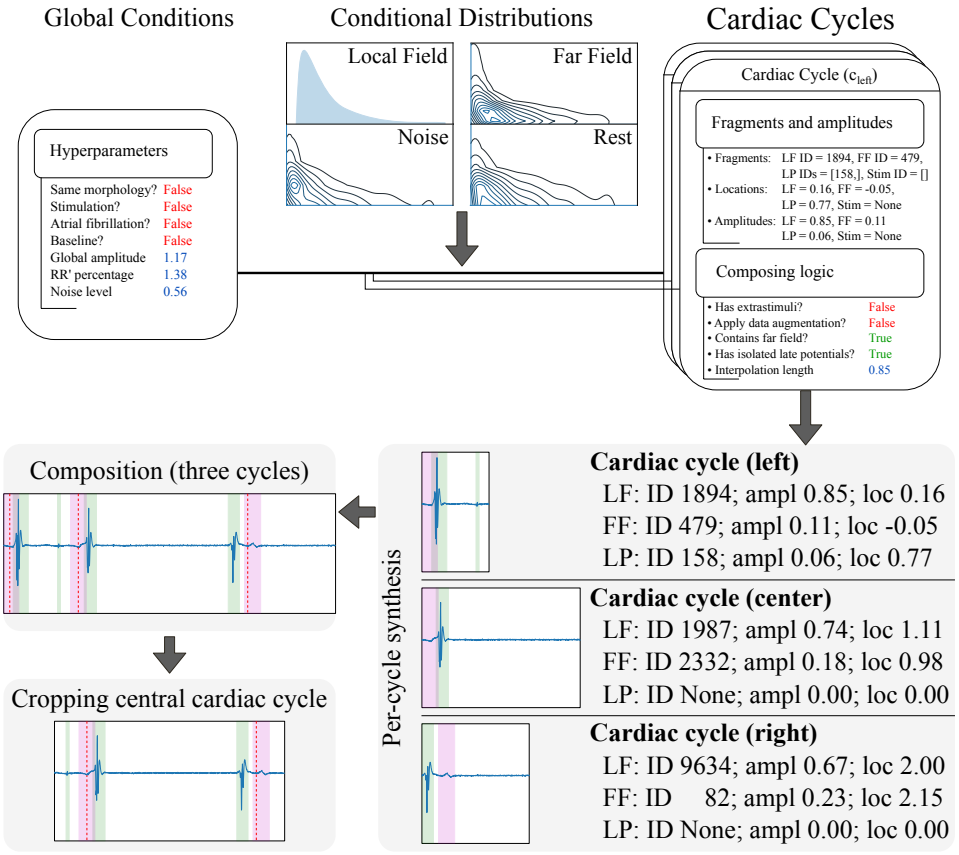


Figure 4.2: Schematic representation of the synthetic data augmentation pipeline. The global conditions govern the selection of fragments and other factors such as the global amplitude or the noise level of the synthetic registry. After global rule generation, three cardiac cycles are generated (c_{left}, c_{center}, c_{right}) and a set of per-cycle rules are generated (e.g., the cycle might not contain a FF due to an atrioventricular block or the cycle contains a stimulation artifact). Then, the specific segments are drawn from the "sets of segments" (local fields, LF; far fields, FF and late potentials, LP) alongside their localization and amplitudes (from the computed conditional amplitude distributions). Thirdly, the three cardiac cycles are independently generated with the rule-based algorithm, localizing each fiducial at specific positions in each cardiac cycle, where c_{left}, c_{center} and c_{right} have locations between [0, 1], [1, 2] and [2, 3], respectively. Finally, the cardiac cycles are summed and cropped to isolate the central cardiac cycle, producing the final synthesized signal (dotted lines indicate approximate positions of the QRS complex. The ground truth is generated through the locations of the different fiducials.

into independent "sets of segments" the LF, FF, LP, stimulation and rest segments. The FF and rest segments were low-pass filtered (100Hz, 2nd order Butterworth filter) to suppress any noise or unsegmented LP in its trace. Moreover, each segment was onset/offset corrected so that the voltage value of the cropped excerpt started and ended in zero, thus avoiding errors when composing the synthetic signal. Then, the LF morphologies were subdivided into LF and LP morphologies according to whether the segment displayed a length shorter than 25 samples as a rule of thumb.

Secondly, once the segments had been cropped and zero-corrected, their amplitudes were fitted into log-normal distributions, which best resembled the original data distribution (Figure 4.3, left). Given the very distinct amplitude relationship of FF and rest segments with respect to the LF's amplitude (Figure 4.3, right; the amplitude of the FF and rest segments is relatively higher in cardiac cycles where the amplitude of the LF is low), the amplitude of the FF and rest segments were fitted relative to the LF's amplitude. For this purpose, the amplitude of the LF was split into 10 bins (dividing the [0,1] interval in increments of 0.1) and, for each bin, the sub-set of FF and rest segments that accompanied each specific LF fragment was selected. Once the splitting had been performed, 10 amplitude distributions were fitted to model the amplitude relationship of the FF and rest segments. The amplitude of the LF and LP segments, on their behalf, were fitted independently of the amplitude of any other fiducial.

Finally, once the amplitudes had been fitted, all segments were normalized to their maximum absolute value ("max abs" scaling), homogenizing their amplitudes in the [-1, 1] range for posteriorly multiplying them with amplitudes drawn from the conditional distributions. The inputs to the data generation algorithm are, thus, the segments as cropped in the different sets of segments and the amplitude distributions, alongside the generation hyperparameters.

Synthetic trace generation

The synthetic generation step is aimed at producing bipolar EGM traces corresponding to single cardiac cycles for algorithm training. Given the dyssynchrony existing in some patients between the location of the QRS complex in the surface ECG and the specific wave locations in the iECG recording (where, e.g., a FF segment can start after the QRS complex in the surface ECG), three cardiac cycles are generated for each patient, and the left- and rightmost are cropped to preserve the context of the central cardiac cycle. Given that some properties might coincide between the three cardiac cycles, some global conditions that affect all cardiac cycles are generated, which comprise:

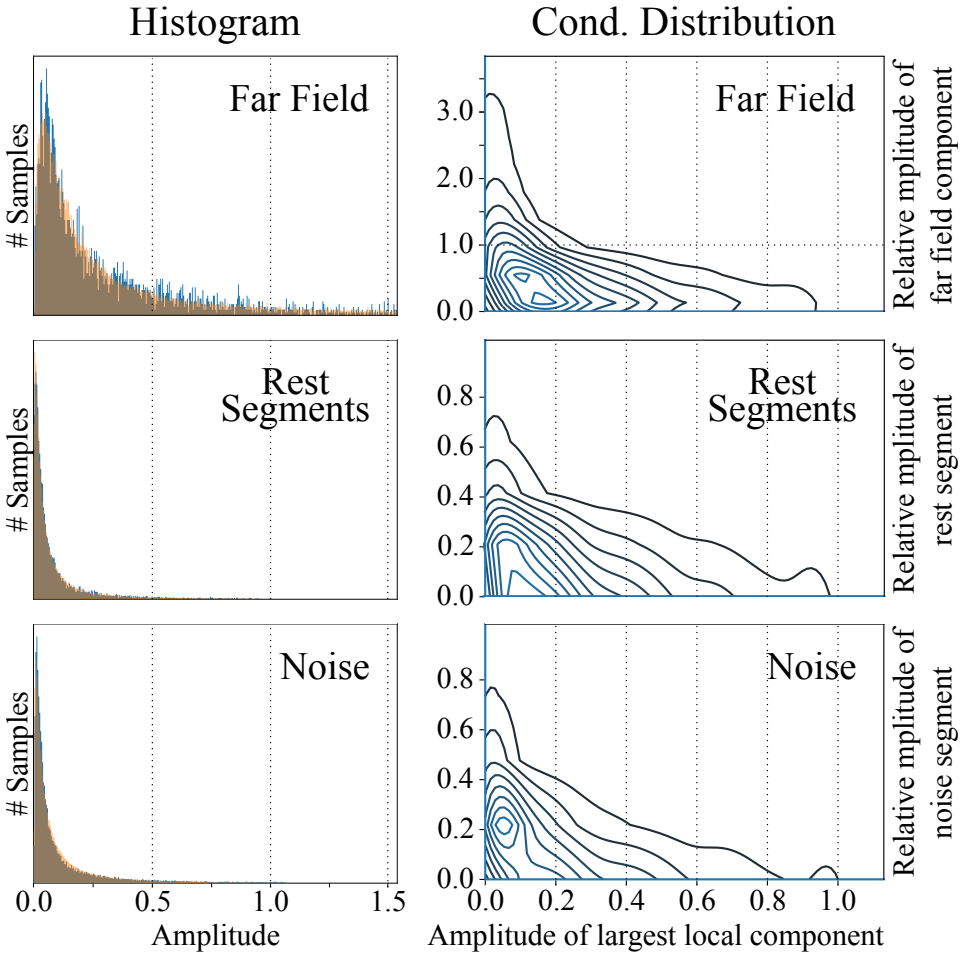


Figure 4.3: Distribution of amplitudes (left) and conditional amplitudes (right) of the cropped segments after normalization. The histograms represent, in blue, the amplitudes of the segments and, overlaid in orange, the result of drawing the same number of samples from a log-normal distribution fitted with the amplitude data, demonstrating a good fit (x axis: amplitude; y axis: samples). On the right, the kernel density estimate distributions of the amplitudes is shown with respect to the amplitude of the local component with the highest amplitude in the cardiac cycle. The plots demonstrate a larger segment amplitude at smaller local field amplitudes (x axis: amplitude of local field; y axis: relative amplitude of segment).

- **(bool) Same morphology**: whether the waveforms retrieved in the independent sets of segments (LFs, FFs, LPs if any and rest) are the same for all cardiac cycles.
- **(float) Global amplitude**: multiplicative factor for the overall registry amplitude, randomly drawn from a normal distribution centered in 1.
- **(int) Beats per minute**: duration of the cardiac cycles, in samples.
- **(float) Interpolation factor**: multiplicative factor (75%-125%) to be used for interpolating the resulting segment. This changes the cardiac cycle duration but also extends the variability in the waveforms of the sets of segments by "compressing" or "expanding" them.
- **(bool) Has stimulation**: whether to include stimulation artifacts in the recording.
- **(float) RR' percentage**: percentage of other cycles to preserve; changes in each execution. E.g. if RR' percentage is 150%, a 25% of the cardiac cycles of the left and right are preserved.
- **(int) Maximum LPs**: maximum number of LPs to include in the cardiac cycle, tied to the RR' percentage.
- **(bool) Has atrial fibrillation**: whether to include more than one "main" LF in the cardiac cycle.
- **(float, float) Noise level, noise factor**: parameters governing random Gaussian noise generation. The noise level (random normal centered at 1) indicates the overall noise amplitude whereas the noise factor acts as a (dividing) normalization term.
- **(bool) Has baseline**: whether to include baseline wander. Currently it is modelled as a sum of sine waves.

Once the rules affecting all cardiac cycles have been determined, each cardiac cycle is independently composed and merged into the final waveform. For this purpose, different segments (LFs, FFs, LFs, LPs and rest segments) and their respective amplitudes are drawn from the sets of segments and amplitude distributions for each cardiac cycle. Given a pre-defined probability, some segments might not be drawn for a specific cardiac cycle (e.g., in the case of AV block, no ventricular activation might take place). If the "same morphology" boolean is toggled, the same segments are drawn for all cardiac cycles, although the amplitudes might

vary. Finally, each segment is given a relative position (in the [0, 3] range) that localizes it within the left ([0,1]), central ([1,2]) or right ([2,3]) cardiac cycles. The leftmost and rightmost cardiac cycles, as mentioned before, are partially cropped to isolate a single cardiac cycle for its analysis.

After the segments, amplitudes and locations are drawn, the signal is composed. Firstly, the rest segments and their respective amplitudes are concatenated to form a baseline upon which to place the rest of the segments. Then, each segment is multiplied by its amplitude and located in the trace according to its selected position. For this purpose, the waveforms (zero-corrected at the onset and offset) are summed to the rest segment baseline starting at a specific index within the baseline, placing them spatially into the registry. The specific location is recorded to generate the ground truth as a binary mask, indicating the precise onset and offset of each accidental. To maximize variability, each segment is given a chance to be interpolated to 75%-125% its original length and a chance to be merged with another waveform using Mixup [143], a data augmentation strategy that produces a linear combination of different waveforms. Finally, once all segments are added into the baseline, the noise and baseline wander are added to the trace and the final segment is cropped according to the value of "RR' percentage" generated in the global conditions.

4.3.2. Architecture

The U-Net [80] is a convolutional neural network (CNN) which is organized as an encoder-decoder structure. The encoder-decoder is a type of artificial neural network (ANN) topology revolving around the usage of an encoder for obtaining highly abstract data representations (usually tied to reducing input complexity), and a decoder to leverage this abstracted information into a specific output [60]. In the case of the U-Net, the encoder and the decoder are conformed of convolutional operations, which act similarly to trainable filters and emphasize local relationships in the data (either spatial or temporal, depending on the data to be analyzed) and pooling/upsampling operations, which reduce/increase the tensor size and allow models to contain a larger amount of trainable filters in the last layers of the encoder. Moreover, the encoder and the decoder are connected by "skip connections", which recover the input information at different levels of abstraction for (a) allowing a more precise definition of segmentation borders which would be lost with the pooling layers, and (b) preventing problems arising from vanishing gradients when optimizing the model's weights [80]. The number of trainable convolutional filters is usually doubled after every pooling operation and halved after every upsampling operation.

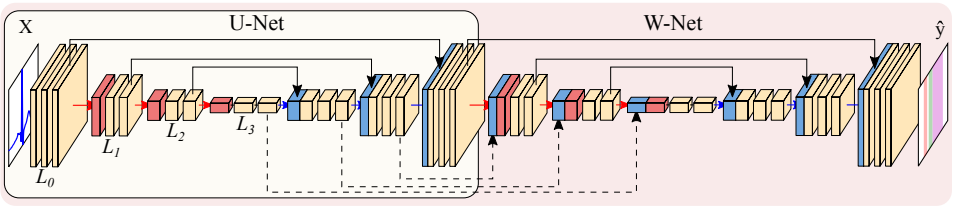


Figure 4.4: Representation of the U-Net (encircled in yellow) and W-Net architectures (encircled in red, containing the U-Net). Both networks are instantiated with 3 levels and 2 convolutional blocks per level. Arrows represent operations, while blocks are indicative of output tensors. Convolutional filters are doubled at each level, so that level L_i has $2^i N$ channels per level (with N being the starting number of channels), whereas pooling and upsampling have a kernel size of 2. Color code: convolutions (yellow), pooling operations (red), upsampling operations (blue), concatenation operations (black).

Many alternatives based on the U-Net exist due to its wide usage and overall high performance [144]. Most works explore altering the model's original design decisions, such as the number of convolutional operations before any pooling operation (hereinafter, model "width"), the number of times the model reduces the input size (model "depth"), number of convolutional filters, employed non-linearity or choice of regularization [82]. Some authors have even developed heuristics for automatically adjusting the model's training parameters and reducing the developer's workload [145]. Other authors have attempted at incorporating state-of-the-art additions such as self-attention mechanisms [114], which allow the weights of an operation to be controlled by a secondary set of weights, effectively controlling feature importance [68]. As an example, in a classification task whose objective is predict whether a traffic light is red, such attention mechanism might outweigh the usage of the red color channel as opposed to the green or blue channels. While some adaptations exist for convolutional operations, this work employs efficient channel attention (ECA) due to its low impact on prediction times [127].

Other works explore expanding the topology, either by embedding the U-Net into another structure [125, 126, 146] or by increasing the model's connectivity (number of times the output tensors from each convolutional operation are used) [147, 148]. In this work, the W-Net architecture [125] was employed given its good performance in other segmentation domains, such as the segmentation of echocardiographic images [130]. The W-Net involves using two U-Nets, where the second network takes as input the output of the first network, and employ "skip connections" not only between each encoder/decoder pair but also between the decoder of the first U-Net and the encoder of the second. This second U-Net

increases the model's capacity, which is usually tied to better performing models [106]. A visual representation of the U-Net and the W-Net are presented in Figure 4.4.

4.3.3. Experiments

The model performance was assessed by training a wide array of configurations to test the contribution of all elements in the model. Firstly, the best architectural configuration, similar to those referred to in chapter 3, was assessed by comparing the performance of the U-Net and W-Net (for depths 5 and 6), both with and without ECA. Secondly, the effect of using a pre-trained model for the task of ECG delineation, taking the weights from the experiments in chapter 3, was tested. Finally, the effect of applying a loss function that forces higher sensitivity was explored by doubling the executions, comprising training models with and without the loss function. The loss function employed the Prewitt operator described in Equation 3.1 for computing the true positives (TP), false positives (FP) and false negatives (FN) as described in Equation 3.2, which are in turn employed for computing the classic sensitivity score: $Se(\%) = TP/(TP+FN)$.

Some aspects were kept constant throughout all experiments. On the one hand, as reported in chapter 2 [82], the application of some regularization strategies such as SDr or certain types of DA was associated with better performance, so these were always applied. A random seed (123456) was employed for reproducibility, the Adam optimizer was used [64], leaky ReLUs [149] were selected as the non-linearities of choice, and the number of base channels was kept the same (32). The goodness of fit was evaluated through the computation of the detection and delineation metrics described in chapter 3. A secondary metric was developed due to the large number of LPs detected within the confines of FF activations, consisting in avoiding counting these as false positives. Finally, due to limitations in time and computational budget, training was solely performed using synthetic data. However, as reported in chapter 3, this was associated with only a slight decrease in performance as compared to using synthetic and real data, and outperformed training the model with real data only. All executions were performed with a NVIDIA Titan Xp GPU, and used the PyTorch library [131].

4.4. Results

The best performing model according to the global F_1 score was a W-Net model with 6 levels, optimized solely with the Dice loss. The model obtained precisions of 76.44%, 74.73% and of 100.0%, and recalls of 94.84%, 95.23% and 100.0% for

	Precision (%)	Recall (%)	Dice (%)	Onset Error (M ± SD)	Offset Error (M ± SD)
Local Field	76.44	94.84	77.37	4.20 ± 13.89	-6.45 ± 19.86
Far Field	74.73	95.23	73.22	3.74 ± 19.26	-5.71 ± 21.91
Local+Far Field	90.02	97.53	83.52	9.04 ± 26.09	-10.65 ± 29.32
Stimulation	100.0	100.0	94.78	-0.68 ± 1.27	-
Local Field (≤ 25ms)	75.04	67.98	45.41	1.51 ± 1.41	-5.69 ± 2.91
Local Field (> 25ms)	80.77	96.18	78.68	4.04 ± 13.42	-3.65 ± 16.67

Table 4.1: Precision (%), recall (%), Dice score (%), onset error (mean [M] ± standard deviation [SD], in miliseconds) and offset errors (M ± SD, in miliseconds) of our best performing model.

localizing LF activations, FF activations and extrastimuli, respectively. The model also attained an average delineation error of 4.20 ± 13.89 and -6.45 ± 19.86 ms when localizing the LF's onsets and offsets, respectively; and of 3.74 ± 19.26 and -5.71 ± 21.91 ms when estimating the onsets and offsets of the FF. The localization of stimulations was very precise, with onset errors of -0.68 ± 1.27 ms. Given the ambiguity between some segments and the errors in the dataset annotations (as will be discussed in Section 4.5, a metric was obtained by merging the binary masks of LF and FF components, which obtained a precision, recall, onset and offset errors of 90.02, 97.53, 83.52, 9.04 ± 26.09 and -10.65 ± 29.32, respectively. A detailed description of the per-wave metrics of the model (precision, recall, Dice score, onset error and offset error) is reported in Table 4.1.

A secondary set of measurements was computed by discarding as false positives any LF that occurred within the confines of a FF, as described in Section 4.3.3. With this secondary metric, the model obtained precisions of 91.28%, 77.78% and of 100.0%, and recalls of 94.86%, 95.25% and 100.0% for localizing LF activations, FF activations and extrastimuli, respectively. The model had an average delineation error of 3.89 ± 14.56 and -6.16 ± 20.25 ms when localizing the LF's

	Precision (%)	Recall (%)	Dice (%)	Onset Error (M ± SD)	Offset Error (M ± SD)
Local Field	91.28	94.86	77.37	3.89 ± 14.56	-6.16 ± 20.25
Far Field	77.78	95.25	73.22	3.47 ± 20.03	-5.44 ± 22.82
Local+Far Field	91.39	97.57	83.52	7.85 ± 28.52	-9.67 ± 31.77
Stimulation	100.0	100.0	94.78	-0.68 ± 1.27	-
Local Field (≤ 25ms)	94.53	67.98	45.41	1.51 ± 1.41	-5.69 ± 2.91
Local Field (> 25ms)	94.06	96.19	78.68	4.0 ± 13.51	-3.6 ± 16.76

Table 4.2: Precision (%), recall (%), Dice score (%), onset error (mean [M] ± standard deviation [SD], in miliseconds) and offset errors (M ± SD, in miliseconds) of our best performing model after discarding small local field activations contained within far field activations.

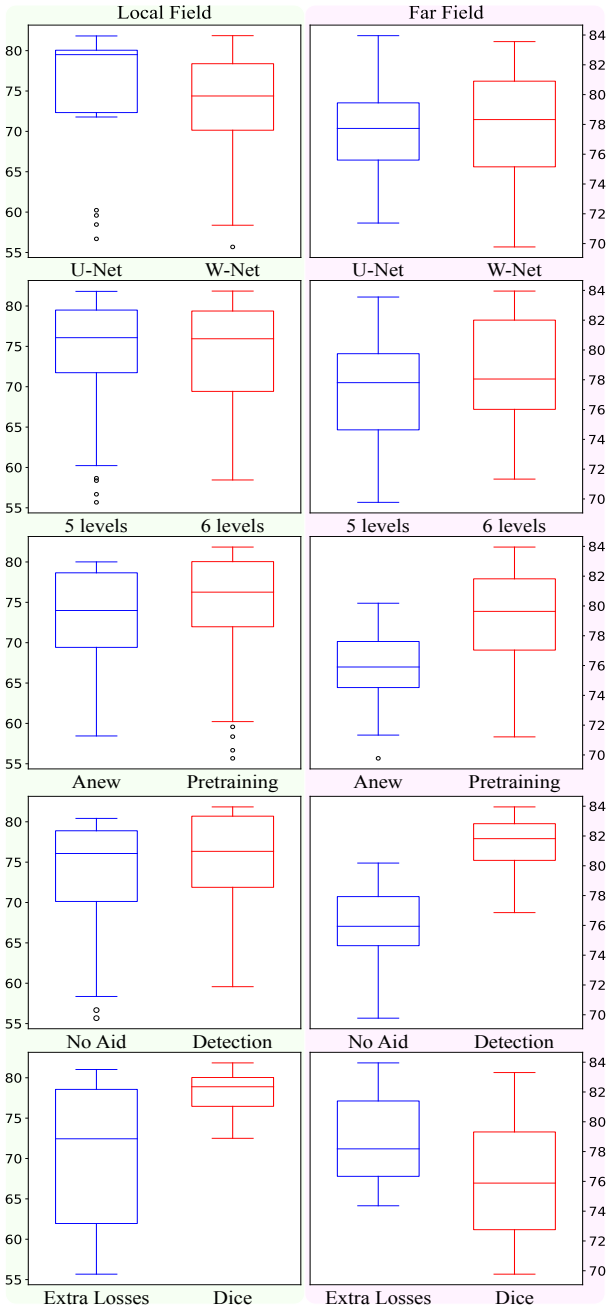


Figure 4.5: Boxplots of the contributions of the different model additions to the overall model performance, divided into the local field (left) and far field (right). Y axis corresponds to the F_1 score.

onsets and offsets, respectively; and of 3.47 $\pm$ 20.03 and -5.44 $\pm$ 22.82 ms in the FF. A more in-depth report of the per-wave metrics of the model is reported in Table 4.2.

With respect to the model additions, the increased model capacity (either with W-Net or with more model depth) and model pre-training with the weights resulting from chapter 3. Other effects, such as the addition of custom data losses, were generally detrimental for model performance. Figure 4.5 summarizes the effect of the different model additions. Finally, some representative examples of the best performing model's performance have been plotted in Figure 4.6 and Figure 4.7. To aid in the discussion, the samples were grouped according to the different types of errors produced by the network (or absence of). These can be divided into four main categories: good samples (Figure 4.6, top), annotation errors in the database (Figure 4.6, bottom), errors due to increased model sensitivity with respect to the ground truth (Figure 4.7, top) and true network errors (Figure 4.7, bottom).

4.5. Discussion

Electrogram segmentation is a crucial task for advancing in the automatization of EP procedures. Currently, physicians must manually produce basic measurements when performing interventions such as AVRT or AP ablation for determining decremental properties or to measure basic intervals. The development of more advanced quantification tools is a must for reducing physician workload, intervention times, complications and intervention cost [22]. However, even state-of-the-art EAM systems solely perform basic detection of the most salient wave within a cardiac cycle, which is in turn used as a single figure of merit for derived metrics such as the LAT. The incapability of locating each individual wave's onset and offset, or simply detecting each wave individually, is limiting; it hampers the system's ability to compare wave morphologies, which has become of utmost importance in recent times due to the development of better diagnostic markers for catheter ablation such as decrement-evoked potentials. DEEPs are detected comparatively through the analysis of portions of myocardial tissue that produce LFs or LPs that are delayed on subsequent cardiac cycles after applying increasingly faster extrastimuli. However, these components might neither be the most salient, nor the first/last in a cardiac cycle, which is the criteria usually employed by EAM systems.

Despite their relative importance, the state-of-the-art for EGM signal analysis mostly focuses on detection rather than on delineation [42, 139]. The employed algorithms are based on the computation of data transformations, such as the Fourier

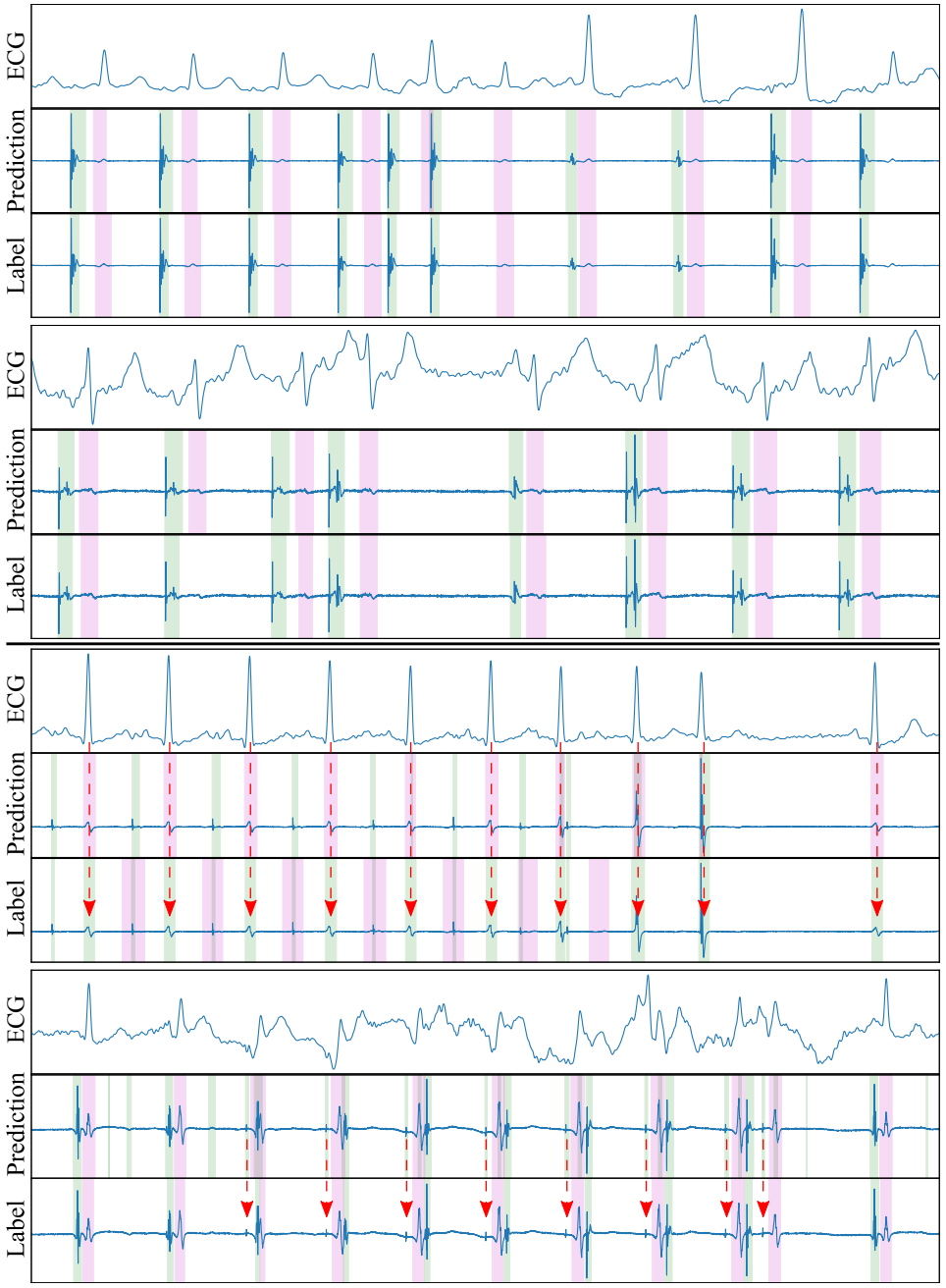


Figure 4.6: ECG reference, predicted fiducials and ground truth from four registries in the annotation dataset, depicting good predictions (top two) and annotation errors (bottom two).

transform, the wavelet transform or filtering, which aid in reducing data complexity for producing robust signal detection. In [42], Osorio *et al.* produced an algorithm based on filtering iECG recordings to discard high-frequency components, aiding in the robust localization of local components in AF. In [139], Felix *et al.* used the WT and a threshold criteria to a similar end, for estimating LFs. In [150], Faes *et al.* proposed to estimate LAT from the barycenter of LFs in bipolar EGMs, after filtering and adaptive thresholding. In [151], Hajimolahoseini *et al.* used a Gaussian mixture model (GMM) for the analysis of the natural logarithm of the signal. Finally, to the best of our knowledge, a single work has been published which directly attempted at EGM delineation. In [39, 40], Alcaine *et al.* firstly delineated onsets and offsets of the QRS complex in the surface ECG, which were then used as detection windows. Then, the iECG was pre-processed by extracting its signal envelope and then the WT was used to produce a more robust representation of the input data. Then, a rule-based algorithm was employed to determine the onset/offset pair of the LFs associated with the delineated QRS complex, reaching delineation errors of 2.2 ± 12.4 and -1.9 ± 7.5 ($M \pm SD$) over the 2138 correctly delineated QRS complexes. This approach, however, is unable to detect and delineate isolated LPs and would fail in the identification of extra LFs in patients with AF, preventing its usage as a general purpose tool for the analysis of iECG recordings. Neither of the aforementioned works in the literature produce detections of individual waves outside the most salient component and only one work computed the onsets and offsets of the predicted wave.

The work presented here builds upon the existing detection and delineation literature by advancing towards an all-purpose iECG analysis system. Similarly to the approach proposed in chapter 3, a DL model was trained for automatic data quantification; focusing on quantification counterbalances the drawbacks of DL algorithms with an application that is immediately interpretable by the operator. Given the lack of large-scale iECG datasets annotated for delineation, two main design decisions were made. Firstly, the model was trained with synthetically generated data from a modest private dataset of 312 iECG recordings taken from 77 distinct patients, with ground truth generated for localizing independent LF and FF activations (no ventricular repolarization). This synthetic dataset greatly improves model performance in scenarios where data is scarce, and has been proven to be even more performant than training on real samples if the data is scarce (chapter 3). Secondly, the prediction pipeline was designed to analyse iECG excerpts of individual cardiac cycles, whose onsets and offsets were localized with the QRS complex's barycenter in the surface ECG, as delineated with chapter 3's model. Cropping the iECG recordings into individual cardiac cycles allowed the model to adjust the prediction of a specific waveform according to whether the LFs (high

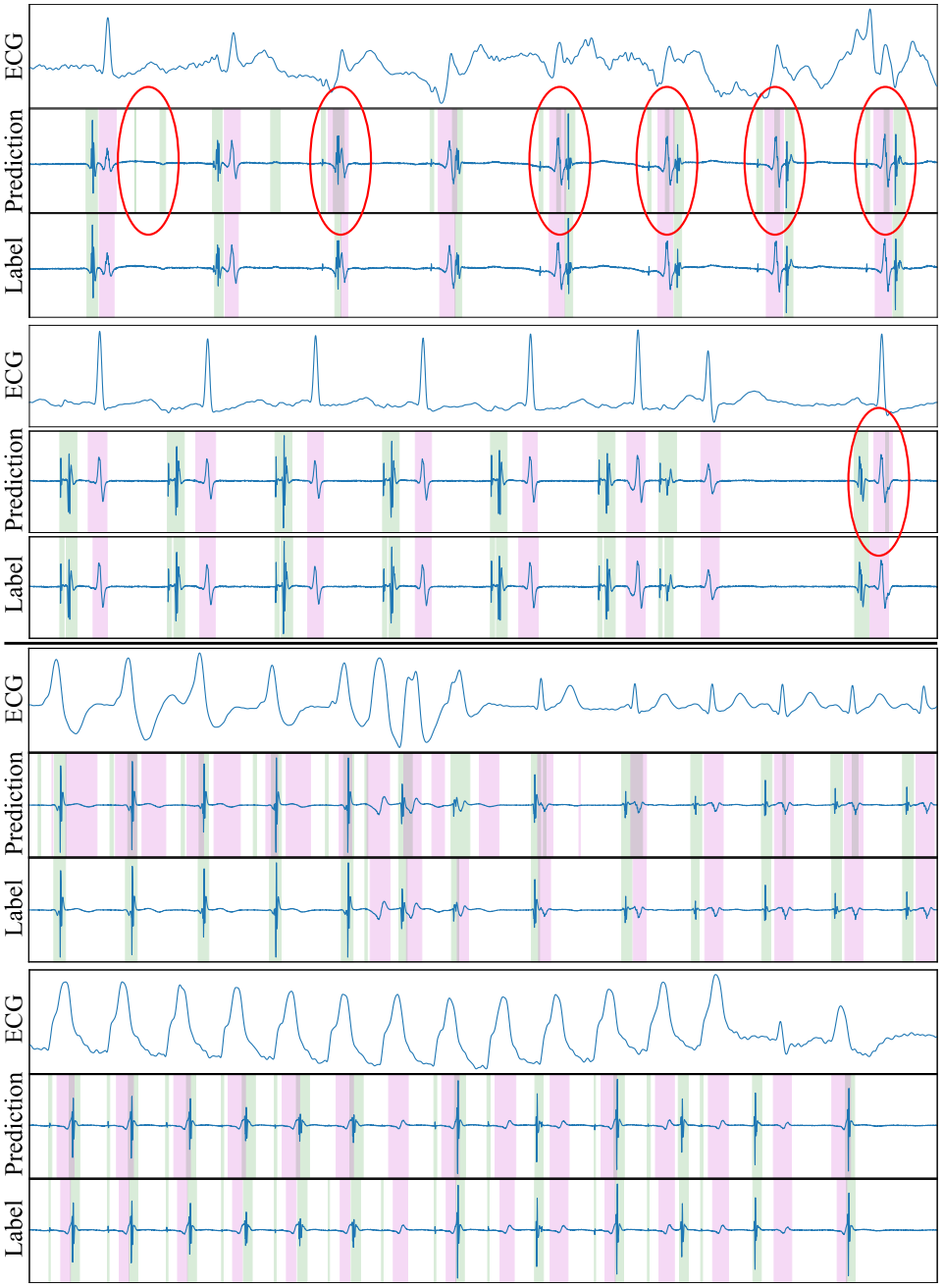


Figure 4.7: ECG reference, predicted fiducials and ground truth from four registries in the annotation dataset, depicting annotation due to increased model sensibility (top two) and errors attributable to the neural network (bottom two).

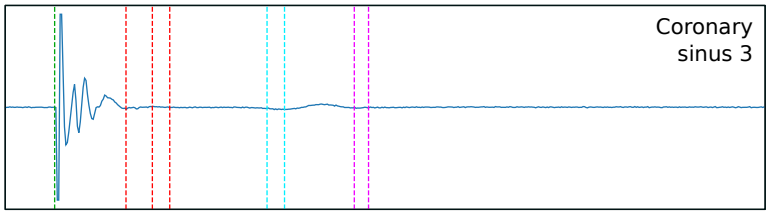


Figure 4.8: The smoothness of the wave complicates the definition of the local field's offset (red dashed line) and the far field's onset and offset (cyan and magenta dashed lines, respectively). Multiple possible onsets/offsets are marked.

frequency components) occurred before or during ventricular depolarization. The combination these design decisions allowed to alleviate the main limitations found in initial approaches, producing more versatile networks.

The produced models demonstrated high sensitivity but moderate precision (around 95% and 75%, respectively, for both LF and FF activations in a held-out test set). With respect to the onset/offset localization, the models provided a good fit with respect to the reference (errors of 3.89 $\pm$ 14.56 and -6.16 $\pm$ 20.25 ms when estimating the LF's onsets and offsets, respectively; and of 3.47 $\pm$ 20.03 and -5.44 $\pm$ 22.82 ms at the FF components). Comparing the proposed approach to the existing literature gives the impression of a reduced algorithm performance: some DSP-based methods reach precision and recall figures nearing 100% [42, 139] and half the SD in onset/offset localization [40]. This, however, is misleading for several reasons. Firstly, existing algorithms are only concerned with locating a single LF activation for each cardiac cycle and disregard any other type of activation (e.g. LP or FF), which difficults direct comparison between methodologies. Secondly, all datasets are private, preventing a fair comparison of methods; the dataset collected for this work consists of real clinical data and no compromises have been made with respect to signal quality or difficulty. Thirdly, models that are more sensible than specific were sought for, and distinguishing subtle LPs from noise is a challenging task. Finally, the larger delineation errors are to be expected given smoothness at signal initiation and termination (Figure 4.8) and the lack of an unified criterion for their definition. Moreover, the overall morphology and location of specific components is more important for downstream applications than a very precise definition of a set of onsets/offsets.

Despite performance considerations, the proposed approach has two main advantages. Firstly, and most importantly, a full delineation of all important fiducials in the registry is performed as opposed to the localization of a single activation [40, 42, 139]. This is of capital importance for advancing in the quantification of iECG recordings, as many routine tasks such as the elucidation of decremental

conduction can only be performed by comparing the localization of specific components in subsequent cardiac cycles. Extending the capabilities of what current quantification algorithms are able to measure can only reduce intervention times and clinical workloads. Secondly, using a DL model coupled with a synthetic data generation algorithm allows to control the conditions that would trigger the prediction of a local component (either LFs or LPs), which is beneficial to the task at hand: the low specificity is a design decision resulting from lowering the threshold at which a perturbation in the trace can be recognized as a local component (Figure 4.7, top). Thus, the system predicts a large amount of high frequency deflections, compatible with local components, that are unannotated in the ground truth while not necessarily erroneous. The difference between the ground truth and the predictions might represent a limitation of the ground truth rather than of the developed model.

The proposed approach is, however, faced with some limitations that are unique to EGMs as opposed to surface ECG processing. Firstly, expressing the ground truth as a binary mask delimitating each local component, which is the traditional DL-based approach, might clash with some scenarios where the individual local components should not be merged; difficulties might arise when analyzing highly fractionated potentials, where predicting a continuous *True*-valued mask spanning the whole fractionation might not be useful for posterior analyses. Secondly, a compromise with respect to the architectural choice might be of need, as the model is too computationally complex to be employed on real time (7.88 ms per cardiac cycle and lead). This, however, might be circumvented by processing the iECG recordings while the catheter changes position, while the EAM system waits for respiration cues or as single execution algorithm that allows filtering acquired points to retrieve those that meet some criteria. Finally, a series of tests could not be performed due to time limitations at the time of writing this thesis. On the one hand, the model could not be trained leveraging real data, partially due to the necessity to improve the quality of the ground truth annotations: many waves were not correctly delineated and accounted for false positives (Figure 4.6, top), requiring re-annotation, and more prevalence of fractionated potentials is needed to assess the generalizability of our approach. On the other hand, the current state of the rule-based synthetic DA algorithm remains somewhat basic, requiring the inclusion of more real-world casuistry in it to enhance performance.

4.5.1. Conclusions

The proposed methodology for the analysis of iECG recordings has proven to be useful in other signal analysis tasks such as ECG delineation, hinting at the

feasibility of a good-performing, all-purpose EGM annotation tool. Current results show great promise while being, to the best of our knowledge, the first tool in the literature allowing the delineation of all local components present in a recording. The algorithm, based on an encoder-decoder DL architecture, was trained solely with synthetic data according to a rule-based algorithm that allows for controlling the generation process.

The algorithm is, however, in a developmental state. Several limitations exist in the dataset, data generation pipeline and data representation, which need to be solved to reduce the high number of false positives. Nevertheless, the development of an all-purpose EGM delineation model is a key tool for unlocking a wide array of downstream tasks, ranging from the automatic identification of myocardial portions of scar presenting DEEPs to the exploration of morphological indicators that might aid in diagnosis or risk stratification, similarly to the work presented in chapter 5.

CHAPTER

5

Characterization of electrocardiographic patterns relating genotype and phenotype in patients with hypertrophic cardiomyopathy

5.1. Introduction

Hypertrophic cardiomyopathy (HCM) is a genetic condition that is the leading cause of sudden cardiac death (SCD) in adolescents and young adults [152], with an annual incidence of 0.3 to 0.5 per 100,000 [153]. HCM is mostly originated by mutations in the sarcomere protein genes although, in some cases, its source might be other genetic diseases (metabolic/neuromuscular diseases, chromosomic abnormalities) or other conditions that are similar in behaviour, such as amyloidosis [153]. HCM causes the myocardium to contract inefficiently due to the presence of left ventricular outflow tract obstruction (LVOTO), mitral regurgitation (MR), diastolic dysfunction, ischaemia, arrhythmias or dysautonomia. This disease has a highly increased risk of SCD due to the interplay between these conditions and the occurrence of ventricular arrhythmias [34, 154]. The most common sarcomeric mutations are in the beta-myosin heavy chain (MYH7) and myosin-binding protein C (MYBPC3) genes, although other genetic markers exist (TNNT2, TNNI3, TPM1, ACTC1, MYL2, and MYL3) and many others are under scrutiny to address their association with the disease [153, 155].

This chapter is adapted from: **Jimenez-Perez, G.**,*et al*. Characterization of electrocardiographic patterns relating genotype and phenotype in patients with hypertrophic cardiomyopathy. (*In Preparation*).

Patients with risk of SCD are considered for implantable cardioverter defibrillator (ICD), which can help treat life-threatening ventricular arrhythmia (VA) episodes. However, ICD device implantation is hampered by the disassociation between the disease's aetiology and its pathophysiology, generating a set of criteria with low specificity [156]. The state-of-the-art risk stratification method suggested by clinical guidelines is the HCM Risk-SCD score [157], which takes parameters such as age, maximal left ventricular (LV) wall thickness, left atrial (LA) diameter, LV outflow tract gradient, family history of SCD, non-sustained ventricular tachycardia, and unexplained syncope, most of which are assessed with echocardiography or MRI [153, 154]. Other markers such as the presence of fibrosis, assessed with late gadolinium enhancement (LGE) MRI, are being evaluated for their clinical use [158].

Despite its wide usage in clinical practice, the ECG is not used for HCM due to a lack of clear understanding between the disease's pathophysiology and specific ECG-derived clinical markers. In HCM, 90% of patients present ECG anomalies such as left ventricular hypertrophy (LVH), left axis deviation (LAD), prominent Q waves, ST segment displacement or T wave inversions (TWI) [152, 159]. Nevertheless, these clinical markers are non-specific for HCM, which prevents their usage as a unified set of criteria for diagnosis and complicates risk stratification [34, 152, 154]. Additionally, the markers explored in the literature are unreliable, as some studies hint at their diagnostic usefulness whereas others do not [154]. Given these issues, the ECG is preferentially used for population screening, requiring diagnosis confirmation with the aforementioned cardiac imaging modalities [153, 154].

Some authors have explored computational approaches for the diagnosis or risk stratification of HCM, despite the small cohorts usually present in clinical analysis. However, most of these approaches have been employed for classification. In [152], Ko *et al*. trained a convolutional neural network (CNN) with 2,448 patients diagnosed with HCM and a control group of 51,153 non-HCM patients for the purpose of automatic diagnostic. Although the authors do not disclose implementation details, they report an area under the curve (AUC) of 0.96, with a sensitivity of 87% and a specificity of 90%. Rahman *et al*. [48, 49] employed support vector machine (SVM) and random forest classifiers over a population of 221 HCM patients and 541 controls comprising patients with ischemic or non-ischemic cardiomyopathies. The authors extracted ECG markers derived from temporal intervals and the raw morphology of delineated ECG recordings, downsampled to 10 samples per lead, generating 504 features per patient and reaching accuracies nearing 85%. The classification approach followed by these works, however, causes a disconnect between the diagnosis and a concrete set of symptoms, disrupting the

normal hypothetico-deductive method [12, 13, 122].

On the other hand, some authors have attempted at directly addressing this relationship between aetiology and pathophysiology. Lyon *et al*. [34] demonstrated the capability of robust feature extraction coupled with principal component analysis (PCA), an unsupervised dimensionality reduction (DR) method, for clustering together patients with similar ECG-based phenotypes. These "phenogroups" were then related with the presence/absence of HCM through MRI-based markers. In a population of 85 HCM patients, they concluded that using the QRS morphology alongside T wave markers allowed the identification of four distinct phenotypes (or phenogroups): a group displaying short R and deep S waves in lead V4; a group with short R and long S waves in V4-V6; a group with normal QRS complex and non-inverted T waves; and a group with TWI and normal QRS morphology. The last group presented higher HCM Risk-SCD scores by almost a factor of two, and displayed coexisting septal and apical hypertrophy ($p \leq 0.0001$). This work, however, had three main limitations that hampered its applicability in the general case: it lacked patient genotyping, had a relatively low event rate and was based on the extraction of surrogates of morphological markers through digital signal processing (DSP)-based approaches.

In this work, we take over from Lyon *et al*. and introduce a novel methodology for the analysis of fully genotyped HCM patients with ECG recordings. For this purpose we employ multiple kernel learning (MKL), a more flexible DR algorithm that allows aggregating raw information from the 12 leads simultaneously. In Section 5.2, information about the study population is provided. In Section 5.3, the data pre-processing, the methodology for DR and the evaluation criteria are detailed. In Section 5.4, the results from the analysis pipeline are described. In Section 5.5, the results obtained in the previous section are discussed under the light of their relevance for HCM analysis and their comparison with similar works in the literature. And, finally, Section 5.6 offers a brief summary of the benefits and limitations of this work.

5.2. Materials

The study population is composed of 103 patients with HCM and 50 screening patients with family history of HCM or SCD. The patients were recruited from the Hospital Universitario Puerta de Hierro, Majadahonda, Madrid. All patients underwent full diagnostic procedure, comprising a 12-lead ECG recording, an echocardiographic exam and, in some cases, a LGE-MRI for assessing fibrosis. Then, echocardiographic markers were extracted for the calculation of the HCM

Risk-SCD score – age, maximal LV wall thickness, LA diameter, LV outflow tract gradient. Additionally, other markers were extracted for assessing their usefulness in diagnosing HCM, mostly related to lifestyle (smoker/nonsmoker), general markers (body mass index, blood pressure, etc.), functional level (NYHA index), family history of events [HCM, SCD, heart failure (HF), arrhythmias], clinical history (unexplained syncope, history of VT, etc.) and LGE-MRI-derived markers (presence of fibrosis, maximum wall thickness, etc.). Given the small occurrence of SCD in the population (2 patients), more inclusive events were considered as endpoints, including the occurrence of VT, unexplained syncope, aborted SCD, atrial fibrillation (AF) or HF. Every patient was genotyped to assess the presence of sarcomeric mutations, and categorized as their expressed genotype (MYBPC3, MYH7, other pathological, variant of uncertain significance (VUS) and the overall positive/negative). A full description of the extracted clinical markers is shown in Table 5.1.

5.3. Methods

The employed methodology consisted in several phases, each of which will be explored in a different section. In Section 5.3.1, the ECG is pre-processed to obtain a common reference system for patient representation. In Section 5.3.2, a DR algorithm employed to obtain a low-dimensional embedding for easier characterization, MKL, is described. Finally, Section 5.3.3 describes the methodology for grouping observations with similar phenotype (phenogroups) and their analysis.

5.3.1. Data preprocessing

Working with ECG presents several singularities that difficult their usage in ML algorithns. On the one hand, important information for diagnosis might be spread over the course of several heartbeats: patients who present atrioventricular (AV) blocks might display some beats in which the impulse failed to communicate to the ventricles, displaying a P wave but neither a QRS complex nor a T wave. On the other hand, events in the ECG might occur at different time intervals, so segments are misaligned in two different patients. This causes comparing two ECG recordings from different patients to be a non-trivial matter, where computing the "similitude" between two heartbeats requires creating a common reference system that handles these mismatches.

A joint representation was designed to compare ECG morphologies aligned to the main segments of interest (P, PQ, QRS, ST, T and TP). For this purpose, the full ECG trace was delineated using the methodology described in chapter 3 [81, 82],

	Screening	HCM	p-value
GENERAL INFORMATION			
Age	42.0 (37.0 - 49.0)	49.0 (42.0 - 60.0)	< **0.0001**
Female	30 (61.22)%	40 (38.83)%	< **0.0001**
Body Mass Index	24.16 (22.0 - 28.0)	26.0 (23.7 - 29.0)	0.0148
Systolic blood pressure (mmHg)	130.0 (123.0 - 130.0)	127.5 (120.0 - 140.0)	0.9209
Diastolic blood pressure (mmHg)	77.71 $\pm$ 10.4	79.17 $\pm$ 10.67	0.9842
MAGNETIC RESONANCE IMAGING/ECHOCARDIOGRAPHIC ANALYSIS			
Heart rate (echo)	66.0 (61.0 - 77.5)	61.0 (55.0 - 68.0)	< **0.0001**
Maximal wall thickness	9.9 (8.3 - 10.3)	18.45 (16.9 - 22.0)	< **0.0001**
Ejection Fraction	50.82 (44.83 - 56.61)	52.37 (44.23 - 57.49)	0.4774
Fibrosis (Gadolinium Enhancement)	2 (11.76)%	41 (65.08)%	< **0.0001**
Left atrial diameter	35.0 (33.0 - 37.0)	43.0 (39.0 - 48.0)	< **0.0001**
COMORBIDITIES			
Hypertension	4 (8.16)%	23 (22.33)%	0.0002
Diabetes	3 (6.12)%	5 (4.85)%	0.5709
Chronic kidney disease	0 (0.0)%	2 (1.94)%	0.0890
Sleep apnea syndrome	0 (0.0)%	4 (3.88)%	0.0155
Coronary disease	0 (0.0)%	2 (1.94)%	0.0890
History of heart failure	0 (0.0)%	8 (7.77)%	0.0005
History of arrhythmias	0 (0.0)%	27 (26.21)%	< **0.0001**
Previous cardiac intervention	0 (0.0)%	27 (26.21)%	< **0.0001**
GENOTYPE			
Genotype positive	21 (42.86)%	103 (100.0)%	< **0.0001**
Genotype negative	15 (30.61)%	0 (0.0)%	< **0.0001**
Genotype MYBPC3	16 (32.65)%	55 (53.4)%	< **0.0001**
Genotype MYH7	5 (10.2)%	33 (32.04)%	< **0.0001**
Genotype VUS	13 (26.53)%	0 (0.0)%	< **0.0001**
Genotype other	0 (0.0)%	15 (14.56)%	< **0.0001**
EVENTS/RISK FACTORS			
Implantable Cardioverter Defibrillator	0 (0.0)%	23 (22.33)%	< **0.0001**
Myocardial infarction	0 (0.0)%	0 (0.0)%	1.0000
Unexplained syncope	2 (4.08)%	12 (11.65)%	0.0090
Aborted Sudden Cardiac Death (SCD)	0 (0.0)%	2 (1.94)%	0.0890
Atrial Fibrillation (AF)	0 (0.0)%	12 (11.65)%	< **0.0001**
Ventricular Tachycardia (VT)	0 (0.0)%	6 (5.83)%	0.0028
History of non-sustained VT (NSVT)	0 (0.0)%	15 (14.56)%	< **0.0001**
Family history (FHx) of HCM or SCD	50 (100.0)%	88 (85.44)%	< **0.0001**
Massive LVH (>30mm)	0 (0.0%)	5 (4.81%)	0.1150
LVOT obstruction (>30 mmHg)	0 (0.0)%	28 (26.92%)	< **0.0001**
Calcification Mitral Valve (MV)	0 (0.0)%	12 (11.65)%	< **0.0001**
Calcification Aortic Valve (AV)	0 (0.0)%	6 (5.94)%	0.0047
RISK SCORES			
Event 1 (VT, syncope, SCD)	2 (4.08)%	34 (33.01)%	< **0.0001**
Event 2 (VT, syncope, SCD, AF, HF)	2 (4.08)%	46 (44.66)%	< **0.0001**
HCM risk	0.95 (0.72 - 1.3)	2.42 (1.92 - 4.42)	< **0.0001**
Cornell Criteria	1 (2.04)%	41 (39.81)%	< **0.0001**
Sokolow Criteria	1 (2.04)%	19 (18.77)%	< **0.0001**
NYHA score I	48 (96.0%)	63 (60.58%)	< **0.0001**
NYHA score II	2 (4.0%)	28 (26.92%)	0.0008
NYHA score III/IV	0 (0.0%)	13 (12.5%)	0.0090

Table 5.1: Clinical variable distribution of the hypertrophic cardiomyopathy (HCM) and screening populations. In brackets, mean, range or standard deviation, depending on the distribution of the variable.

	Lead	Screening	HCM	p-value
P wave amplitude	I	0.09 (0.08 - 0.11)	0.11 (0.08 - 0.14)	0.0031
	II	0.16 ± 0.05	0.17 ± 0.05	0.1468
	III	0.11 (0.09 - 0.13)	0.11 (0.08 - 0.15)	0.0618
	aVR	0.12 ± 0.04	0.13 ± 0.04	0.0549
	aVL	0.07 (0.05 - 0.08)	0.08 (0.06 - 0.1)	0.0002
	aVF	0.13 ± 0.04	0.13 ± 0.05	0.2415
	V1	0.1 (0.08 - 0.12)	0.13 (0.11 - 0.17)	< **0.0001**
	V2	0.08 (0.06 - 0.1)	0.1 (0.08 - 0.13)	< **0.0001**
	V3	0.1 (0.09 - 0.11)	0.12 (0.09 - 0.15)	< **0.0001**
	V4	0.1 ± 0.03	0.11 ± 0.04	< **0.0001**
	V5	0.09 (0.07 - 0.11)	0.1 (0.08 - 0.12)	0.0290
	V6	0.09 ± 0.03	0.09 ± 0.03	0.1415
QRS wave amplitude	I	0.66 (0.5 - 0.87)	0.99 (0.71 - 1.25)	< **0.0001**
	II	1.21 (0.89 - 1.57)	1.24 (0.89 - 1.93)	0.0123
	III	0.8 (0.61 - 1.19)	1.06 (0.72 - 1.86)	< **0.0001**
	aVR	0.91 (0.78 - 1.04)	0.96 (0.76 - 1.3)	0.0024
	aVL	0.57 (0.32 - 0.79)	0.85 (0.54 - 1.24)	< **0.0001**
	aVF	0.89 (0.66 - 1.33)	1.02 (0.7 - 1.85)	0.0003
	V1	0.99 (0.67 - 1.16)	1.44 (0.92 - 2.01)	< **0.0001**
	V2	1.29 (1.12 - 1.84)	2.08 (1.46 - 2.77)	< **0.0001**
	V3	1.31 (0.89 - 1.63)	2.2 (1.67 - 2.91)	< **0.0001**
	V4	1.33 (1.01 - 1.81)	2.05 (1.45 - 2.68)	< **0.0001**
	V5	1.28 (0.99 - 1.64)	1.56 (1.13 - 2.34)	< **0.0001**
	V6	1.13 (0.91 - 1.3)	1.19 (0.91 - 1.55)	0.0015
T wave amplitude	I	0.22 (0.15 - 0.26)	0.21 (0.14 - 0.29)	0.1658
	II	0.32 (0.24 - 0.38)	0.23 (0.16 - 0.33)	< **0.0001**
	III	0.16 (0.11 - 0.22)	0.2 (0.12 - 0.28)	0.0031
	aVR	0.27 (0.2 - 0.32)	0.19 (0.15 - 0.27)	**0.0001**
	aVL	0.1 (0.08 - 0.14)	0.16 (0.1 - 0.24)	< **0.0001**
	aVF	0.22 (0.17 - 0.3)	0.19 (0.13 - 0.27)	0.0409
	V1	0.15 (0.1 - 0.19)	0.25 (0.17 - 0.36)	< **0.0001**
	V2	0.29 (0.21 - 0.42)	0.38 (0.26 - 0.64)	< **0.0001**
	V3	0.29 (0.17 - 0.52)	0.39 (0.22 - 0.64)	0.0002
	V4	0.28 (0.2 - 0.48)	0.28 (0.18 - 0.53)	0.3831
	V5	0.3 (0.19 - 0.41)	0.25 (0.15 - 0.41)	0.5344
	V6	0.27 (0.18 - 0.37)	0.21 (0.15 - 0.31)	0.0677
Axis P wave (degrees)		50.0 (36.0 - 64.0)	45.0 (27.0 - 62.0)	0.1124
Axis QRS wave (degrees)		48.52 ± 39.55	17.11 ± 51.44	< **0.0001**
Axis T wave (degrees)		50.0 (34.5 - 66.0)	40.0 (20.0 - 60.0)	0.0091
T wave inversion (TWI)		5 (3.4)%	152 (49.19)%	< **0.0001**
P duration (ms)		108.0 (100.0 - 112.0)	116.0 (108.0 - 124.0)	< **0.0001**
PQ duration (ms)		52.0 (44.0 - 62.0)	56.0 (44.0 - 68.0)	0.0152
QRS duration (ms)		88.0 (84.0 - 92.0)	96.0 (88.0 - 104.0)	< **0.0001**
ST duration (ms)		96.0 (90.0 - 108.0)	108.0 (100.0 - 120.0)	< **0.0001**
T duration (ms)		196.0 (184.0 - 208.0)	220.0 (204.0 - 240.0)	< **0.0001**
TP duration (ms)		292.0 (216.0 - 422.0)	360.0 (296.0 - 452.0)	0.0011

Table 5.2: Electrocardiographic descriptors distribution for the hypertrophic cardiomyopathy and screening populations. The extracted features are grouped by amplitudes of the main waves (P, QRS, T) and other characteristics, comprising axis deviations, T wave inversions and segment durations.

marking the beginning and end of the P, QRS and T waves. Next, each heartbeat was independently cropped in PP' (from the P_{on} of a beat to the P_{on}' of the next). After this, the length of every segment was independently measured for the whole dataset, and the maximum value was retrieved independently for each segment type (88, 100, 94, 80, 204 and 414 samples for the P, PQ, QRS, ST, T and TP segments, respectively). Then, each segment of each heartbeat was interpolated to the computed fixed segment length, storing as extra information the ratio between the real length and the interpolation length. After this process, all heartbeats are registered so that every segment coincides in size and waves occur at the same time. The finished ECG traces were amplitude-normalized with respect to the calibration pulse to facilitate posterior incorporation of new data.

Given the input size limitations of the employed DR algorithm (Section 5.3.2), the number of cardiac cycles had to be reduced from the original 1397 heartbeats. For this purpose, the cross-correlation between different cycles was computed for each patient independently. The cross-correlation (averaged throughout all leads) served as a ranking for addressing the similarity between beats. The two beats with the largest mean cross-correlation were selected for each patient, alongside the beat that presented the highest dissimilarity. Outliers presenting distortion due to movement/respiration or excessive noise were manually removed. After this sub-set selection, a total of 459 heartbeats were included for training the DR algorithm and its posterior analysis. A visual representation of the whole data pre-processing pipeline is shown in Figure 5.1.

5.3.2. Multiple kernel learning

Multiple kernel learning is an algorithm employed to extract a low-dimensional embedding from a dataset that contains multiple descriptors per observation (e.g., different leads in the ECG) while preserving the relative distances between different observations. This low-dimensional space can be useful to assess whether different elements in a population share any common set of traits, allowing to explore the existence of different groups of similar phenotype ("phenogroups") and to describe their behaviour in posterior analyses. The MKL algorithm consists in formulating the dimensionality reduction as a graph embedding problem [160] and then optimizing the algorithm's parameters. Formally, given a high-dimensional dataset $X = \left\{\mathbf{x}_i \in \mathbb{R}^{d_N}\right\}^N$, the objective of MKL is to find a low-dimensional embedding $Y = \left\{\mathbf{y}_i \in \mathbb{R}^{M-1}\right\}$, where N is the number of different data descriptors (e.g., the different leads in an ECG, with $N = 12$) and $\mathbf{x}_i$, with $i \in 1..M$, is the vector corresponding to the ith observation, of size $\mathbb{R}^{d_N}$ (the vector's size need not coincide in different descriptors).

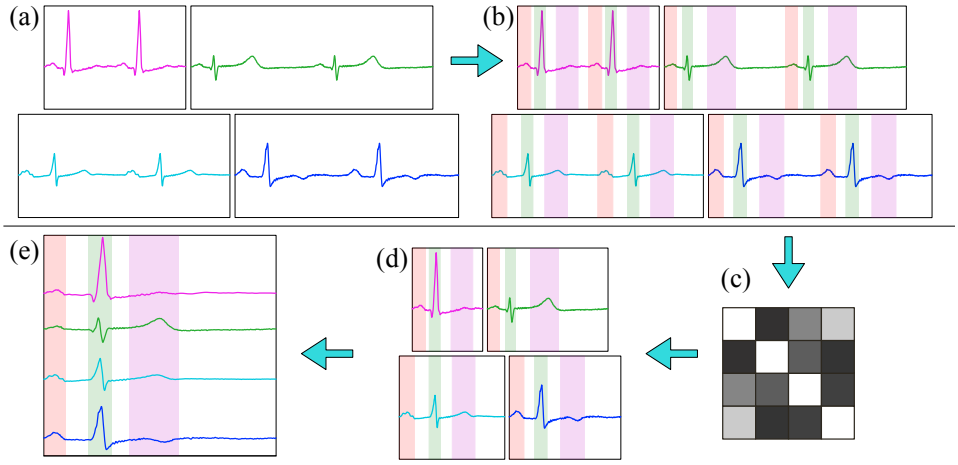


Figure 5.1: Pre-processing pipeline exemplified with electrocardiographic (ECG) recordings of four different patients (only lead I is depicted). Starting from the unprocessed signal (a), the ECG is delineated (b) using the delineator described in chapter 4. Then, the cross-correlation between cardiac cycles, taken from P to the next P (P'), is computed (c). Then, of each registry, the two most similar and the most dissimilar beats are selected and isolated (d). Finally, the segment length for each individual segment (P, PQ, QRS, ST, T, TP) is computed, and the 99th percentile is selected as the interpolation length. Finally, each cardiac cycle is interpolated into a common reference system. Each individual cardiac cycle has different segment size and onset/offset fiducial locations (red: 251 ms; green: 255 ms; magenta: 149 ms; cyan: 223 ms; blue: 236 ms). After registration, all segments have the same length (490 ms) and the onset/offsets coincide (e).

A single-descriptor formulation of MKL can be formulated for easier understanding of multiple descriptors. Consider an input data matrix $X = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$ (where $\mathbf{x}_i \in \mathbb{R}^m$) which is to be embedded into a lower dimensional space $\mathbf{y} = [y_1, y_2, ..., y_N]$ (where $y_i \in \mathbb{R}$). The relationship between elements in the input data matrix X can be expressed as a full (weighted) graph, where the vertices are each observation in the space ($V = \{\mathbf{x}_i,\ \forall i = 1..N\}$) and the edges represent the pairwise similarity (measured according to a predefined metric) between two sets of points ($E = \{(x_i, x_j) \rightarrow \mathbb{R}, \forall i, j = 1..N\}$). Thus, the adjacency matrix W represents the relative distances between each pair of observations, and the degree (D) and Laplacian (L) symmetric matrices can be computed as:

$$D_{ii} = \sum_{i \neq j} w_{ij},\ \forall i\ , \qquad L = D - W. \tag{5.1}$$

For producing a low-dimensional embedding of the input data, the coordinates of

the low-dimensional space $\mathbf{y}$ must be estimated from the input data X while preserving the relative similarities that are present in X; i.e., for a large similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$, the distance between y_i and y_j should be small and vice-versa [160]. If the projection was linear ($\mathbf{y} = X^T\mathbf{v}$, with $\mathbf{v} \in \mathbb{R}^m$ being the unitary projection vector), this intuition is formally expressed as the following optimization problem:

$$\begin{aligned}
\mathbf{v}^* &= \underset{\mathbf{v}XBX^T\mathbf{v}=1}{\operatorname{argmin}} \sum_{i\neq j} \left\|\mathbf{v}^T\mathbf{x}_i - \mathbf{v}^T\mathbf{x}_j\right\|^2 w_{ij}, = \underset{\mathbf{v}XBX^T\mathbf{v}=1}{\operatorname{argmin}} \mathbf{v}^T XLX^T\mathbf{v} \\
&\text{or, equivalently,} \\
\mathbf{v}^* &= \operatorname{argmin} \sum_{i,j=1}^{N} \left\|\mathbf{v}^T\mathbf{x}_i - \mathbf{v}^T\mathbf{x}_j\right\|^2 w_{ij}, \\
&\text{subject to} \sum_{i=1}^{N} \left\|\mathbf{v}^T\mathbf{x}_i\right\|^2 d_{ii} = 1.
\end{aligned} \tag{5.2}$$

This optimization problem can solved by resolving the eigenvalue problem $XLX^T\mathbf{v} = \lambda XL'X^T\mathbf{v}$ which, in turn, allows for: (a) extracting up to $N-1$ dimensions from the input data, as many eigenvectors can be extracted using this formulation ($V = [\mathbf{v}_1, \mathbf{v}_1, ..., \mathbf{v}_{N-1}]$, so $Y = X^TV$); and (b) sorting the low-dimensional embedding into decreasingly contributing elements, providing a sense of importance [14, 160].

A non-linear extension to this solution is to project the data into a high-dimensional Hilbert space ($\phi : \mathscr{X} \to \mathscr{F}$) through the usage of kernel functions ($k : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$), so that $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$) [160]. These functions are used in many dimensionality reduction algorithms, such as kernel-PCA [161] and can take many shapes, as long as they represent an inner product space and the matrix K of the pairwise distances in the data X is positive definite. An example of such kernel functions, and the one that will be employed in this work, is the Gaussian kernel (also known as the radial basis kernel):

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \tag{5.3}$$

where σ is a free parameter, which can be computed as the average distance between each point $\mathbf{x}_i$ and its k-nearest neighbors (k-NN) [15]. The similarity/affinity/kernel Gram matrix K can be thus computed with the pairwise kernel distances between each point, such that $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$. With this, we can

assume that the mapping lies in the basis produced by the (kernelized) input data, so that:

$$\begin{aligned} \mathbf{v} &= \sum_{i=1}^{N} \alpha_i \phi(\mathbf{x}_i) = \boldsymbol{\alpha}^T \boldsymbol{\phi}(X), \qquad K = \boldsymbol{\phi}(X)\boldsymbol{\phi}(X)^T, \\ \mathbf{K}_i &= [\phi(\mathbf{x}_i)\phi(\mathbf{x}_1), \phi(\mathbf{x}_i)\phi(\mathbf{x}_2), ..., \phi(\mathbf{x}_i)\phi(\mathbf{x}_N)]. \end{aligned} \tag{5.4}$$

These terms can be introduced in Equation 5.2, reformulating it to offer a kernelized version of the algorithm:

$$\begin{aligned} &\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \sum_{i,j=1}^{N} \left\| \boldsymbol{\alpha}^T \mathbf{K}_i - \boldsymbol{\alpha}^T \mathbf{K}_j \right\|^2 w_{ij} \\ &\text{subject to} \sum_{i=1}^{N} \left\| \boldsymbol{\alpha}^T \mathbf{K}_i \right\|^2 d_{ii}. \end{aligned} \tag{5.5}$$

Extension to Multiple Descriptors

The multiple descriptor formulation operates by computing independent similarity matrices for each descriptor [14, 15]. The dataset would consist in one matrix per descriptor $\mathscr{X} = \{X \in \mathbb{R}^{N \times M_d}\}_{d=1}^{D}$, the output space would remain the same ($Y = \{\mathbf{y}_i \in \mathbb{R}^{M-1}\}$) and the affinity matrix would correspond to $\mathbb{K} \in \mathbb{R}^{N \times N \times D}$. A variety of kernel functions can be used different descriptors, allowing for the computation of the kernel that better suits the descriptor; moreover, the dimensions of the features should be consistent for different samples of the same descriptor but does not need to coincide in different descriptors ($d_i \neq d_j$).

The kernels are employed in two ways. Firstly, a coefficient vector $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_D]$ governs the contribution of each descriptor to the global ensemble. Secondly, the affinity matrices K_d are then used for computing a global affinity matrix, which measures the relative distance between samples ($W = \sum_{d=1}^{D} (K_d)^{1/\alpha_d}$), where α_d is a regularization term that prevents descriptors with large variabilities to outweigh the other descriptors [15]. Finally, this formulation can be used to modify Equation 5.4, so that:

$$\mathbf{v}^T \phi(\mathbf{x}_i) = \sum_{j=1}^{N} \sum_{d=1}^{D} \alpha_j \beta_d K_{d_{ij}} = \boldsymbol{\alpha}^T \mathbb{K}^{(i)} \boldsymbol{\beta}, \tag{5.6}$$

where $\mathbb{K}^{(i)} \in \mathbb{R}^{N \times D}$ corresponds to the kernel distances of an observation $\mathbf{x}_i$ with respect to all other observations in the dataset and all descriptors. Plugging these

terms into Equation 5.2, the MKL formulation becomes:

$$\begin{aligned} &\underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i,j=1}^{N} \left\| \boldsymbol{\alpha}^T \mathbb{K}^{(i)} \boldsymbol{\beta} - \boldsymbol{\alpha}^T \mathbb{K}^{(j)} \boldsymbol{\beta} \right\|^2 w_{ij}, \\ &\text{subject to} \sum_{i=1}^{N} \left\| \boldsymbol{\alpha}^T \mathbb{K}^{(i)} \boldsymbol{\beta} \right\|^2 d_{ii}, \\ &\beta_d \geq 0, \sum \beta_d = 1 \end{aligned} \tag{5.7}$$

This equation can be used, as described in the single-descriptor case, for the computation of multiple dimensions in the projection ($A = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_N]$), while sharing the kernel coefficients $\boldsymbol{\beta} \in \mathbb{R}^D$, becoming:

$$\begin{aligned} &\underset{A,\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i,j=1}^{N} \left\| A^T \mathbb{K}^{(i)} \boldsymbol{\beta} - A^T \mathbb{K}^{(j)} \boldsymbol{\beta} \right\|^2 w_{ij}, \\ &\text{subject to} \sum_{i=1}^{N} \left\| A^T \mathbb{K}^{(i)} \boldsymbol{\beta} \right\|^2 d_{ii}, \\ &\beta_d \geq 0, \sum \beta_d = 1 \end{aligned} \tag{5.8}$$

Given the difficulty of solving the above optimization problem, Lin *et al.* [14] devised a two-step optimization algorithm, by means of optimizing the system's A and $\boldsymbol{\beta}$ independently, after initializing $\boldsymbol{\beta}$ uniformly to $1/D$. The first step consists in computing A by fixing the value of $\boldsymbol{\beta}$, and is performed by eigendecomposition. The second step optimizes the values of $\boldsymbol{\beta}$ while fixing A. Once the values of A and $\boldsymbol{\beta}$ are optimized, the samples can be projected in the low-dimensional embedding by computing:

$$Y = A^T \sum_{d=1}^{D} K_d \beta_d, \tag{5.9}$$

where $Y \in \mathbb{R}^{N \times N-1}$ contains the projected samples in the output space, each column corresponding to a different extracted dimensions.

5.3.3. Low-dimensional embedding analysis

Contrarily to supervised approaches, MKL does not cluster together data according to a specific label (e.g., presence/absence of a disease) but based on the similarity between observations according to a pre-defined metric. The benefit of unsupervised approaches is that the distribution of labels in the output space is explored *a posteriori*, allowing the characterization of correlations found in the data

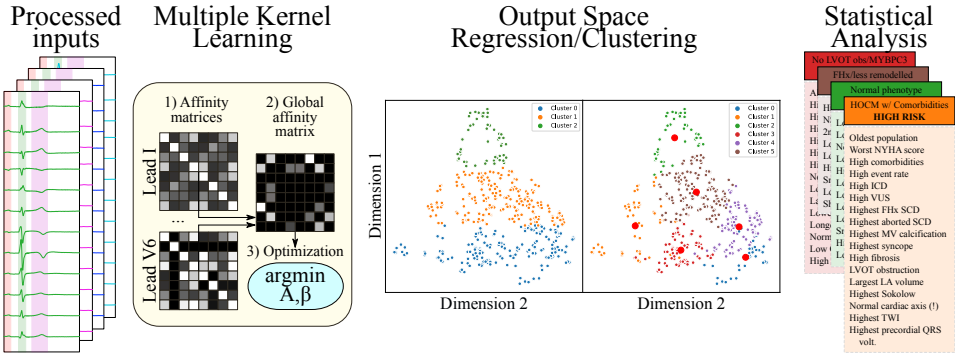


Figure 5.2: Analysis pipeline. The pre-processed inputs are projected into a low-dimensional embedding through the usage of Multiple Kernel Learning. The observations are then projected into the output space for their characterization, via multiscale kernel regression, clustering and through clustering along a single dimension. Finally, the found phenogroups are correlated to clinical variables of interest, providing insight into the characteristics of the input dataset.

that might correspond with clinical variables. In this work, this analysis was performed in three ways: a coarse clinical data exploration, clustering and statistical analysis. An overview of the analysis pipeline is depicted in Figure 5.2.

Output space pre-processing

The data from the populations of HCM patients and patients selected for screening was projected into a low-dimensional embedding, consisting in 458 dimensions ($3N-1$ dimensions due to taking three cardiac cycles from each patient), and sorted with respect to their eigenvalues. Firstly, the output space was rotated in its first and second dimensions to facilitate interpretation of the output space, as depicted in Figure 5.3. Secondly, given the high correlation between dimensions in the low-dimensional embedding, an algorithm was employed for determining the space's self-correlation when introducing an increasing number of dimensions, effectively performing feature selection. The algorithm considers the number of dimensions that are required so that each observation's neighbours remains the same up to a given correlation threshold, result that is used throughout the whole data analysis. In this work, a 95% correlation was employed (Figure 5.4, left).

Coarse exploration

An initial coarse exploration of the output space was performed. For that purpose, the clinical labels of interest, such as events or specific sarcomeric mutations, were

projected to coarsely assess their distribution and relative density in the space. Furthermore, the output space was explored through the usage of multiscale kernel regression (MKR) [162, 163], a regression algorithm that allows for reconstructing the expected morphology of an observation in the original space at specific positions in the low-dimensional embedding. For the purposes of this work, the ECG morphology was reconstructed at high density positions in the output space, according to the cloud point's mean (M) and standard deviation (SD): [M-2SD, M-SD, M, M+SD, M+2SD].

Output space clustering

The patients were grouped using two main methods. Firstly, a simple unsupervised clustering algorithm, k-means (kM) [164] was employed. This algorithm iteratively adjusts the positions of k cluster centers, assigning the same label to all points that lie closest to a specific cluster center. The algorithm minimizes the within-cluster variances, attempting at producing compact clusters (in an Euclidean sense) that span the minimum amount of space with respect to the other clusters. Although this algorithm is sensitive to the choice of k, the Silhouette score [165] was used as heuristic to set the number of clusters that had the lowest intra-cluster distance while attaining the highest inter-cluster distance. A secondary clustering was performed by dividing the output space in equally sized blocks with respect to the first dimension in the output space, which accounted to most of the variability in the population. The Silhouette score was also employed to determine the number of divisions that produce the most compact partitions (Figure 5.5, left).

Clinical variables exploration

Finally, once the clusters were computed, the distribution of clinical markers in each cluster was analyzed to characterize their phenotype and their relationship to the clinical markers of interest. This was performed in two ways: firstly, the variability within each cluster was assessed by visualizing the cluster's general ECG morphology, also selecting a representative example of each cluster to better characterize the cluster's general morphology. Secondly, the effects of the clustering were assessed by performing statistical analysis of the distribution of the clinical variables within each cluster. For this purpose, and similarly to the work by Lyon *et al.* [34], the Kruskal-Wallis test was performed for non-normally distributed data, the Chi-square test was used for categorical data and one-way ANOVA was used for normally distributed data. A statistical significance threshold of 0.0001 was employed.

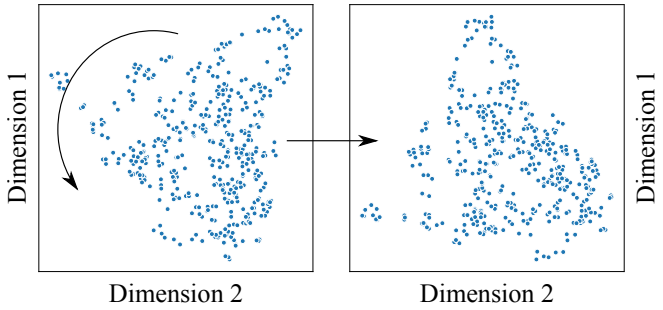


Figure 5.3: Rotation of the first two dimensions of the output space, allowing for a vertical vs. horizontal interpretation of the low-dimensional embedding.

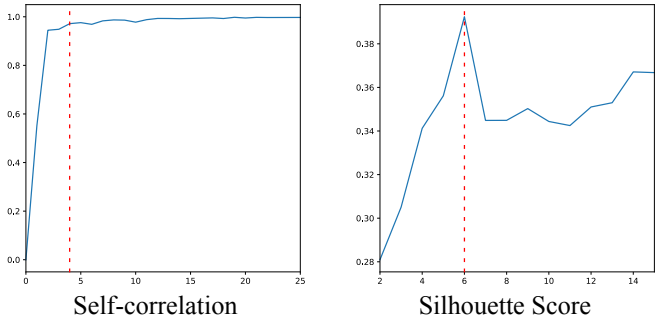


Figure 5.4: Self-correlation (left) and Silhouette score (right) of the low-dimensional embedding. Dotted in red, the optimal values chosen for the analysis: 4 dimensions for the self-correlation, surpassing the 95% threshold, and 6 clusters, corresponding to the largest Silhouette score.

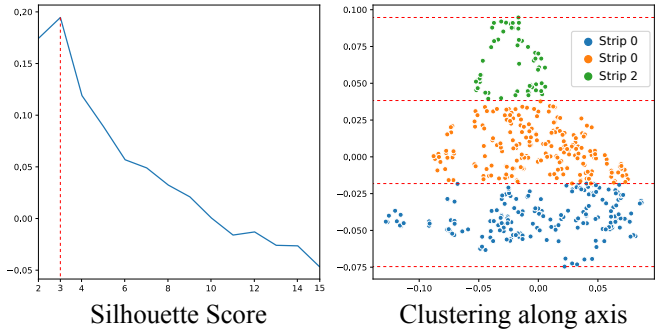


Figure 5.5: Division of the low-dimensional embedding into equisized strips along dimension 1. The Silhouette score indicates an optimal data distribution (left, dashed line indicates highest value) comprising three strips (right, dashed lines correspond to division values).

5.4. Results

5.4.1. Baseline population characteristics

The acquired clinical variables were analyzed to assess their contribution to distinguishing HCM patients ($n = 103$) from patients selected for screening ($n = 50$); the results are presented in Tables 5.1 and 5.2. The HCM patients were slightly older (49 vs. 42 years, $p < 0.0001$) and more distinctly male (61.17 vs. 38.78%, $p < 0.0001$); they also demonstrated a larger LV wall thickness (18.45 vs. 9.9 mm, $p < 0.0001$) and LA diameter (43 vs. 35 mm, $p < 0.0001$); and the highest amount of comorbidities and fibrosis. Moreover, all HCM patients were genotype positive, with 55 patients displaying the MYBPC3 variant (53.4%, $p < 0.0001$), 33 presenting the MYH7 variant (32.04%, $p < 0.0001$) and 15 having other pathological variants (14.56%, $p < 0.0001$). The population had a low event rate (2 patients had a SCD episode, prevented by an ICD; not statistically significant), so other markers were employed to assess risk, comprising VT or AF episodes, as well as HF or unexplained syncopes. The screening population, however, only accounted for two patients with unexplained syncope, whereas 46 HCM patients demonstrated different events (4.08% vs. 44.66%, $p < 0.0001$). The distribution of some clinical variables in the output space is depicted in Figure 5.6.

5.4.2. Low-dimensional embedding exploration

After output space rotation and subset selection (Figure 5.4, left), MKR was employed for determining the variability contained in the main four dimensions. For this purpose, five representative points for each dimension were defined in the output space, located at M-2SD, M-SD, M, M+SD, and M+2SD. These five points were then used for performing MKR, regressing its morphology in the input space. Figure 5.7 depicts the morphologies associated to these points. The analysis of these regressions show that most changes concentrate on the morphology of the QRS, ST and T segments. Dimension 1 (and, to a lesser extent, dimensions 2 and 4) reflects TWI inversions, which are most prevalent in leads I, II, aVR and V4-V6 when the values in the embedding grow smaller. Dimension 1 also displays changes in QRS morphology compatible with LAD, with aVF and III becoming more negative. Finally, dimension 1 also shows a later precordial transition (the polarity changes from negative to positive from V3-V4, red line in Figure 5.7, to V4-V5, blue line), ST depression/elevation in most leads, and a wide variability in segment duration. Dimension 2 mainly represents changes in voltage and in QRS axis, with smaller values in the embedding representing larger voltages and a more deviated axis. Moreover, TWI is seen in some leads (I, aVL, V4-V6) and

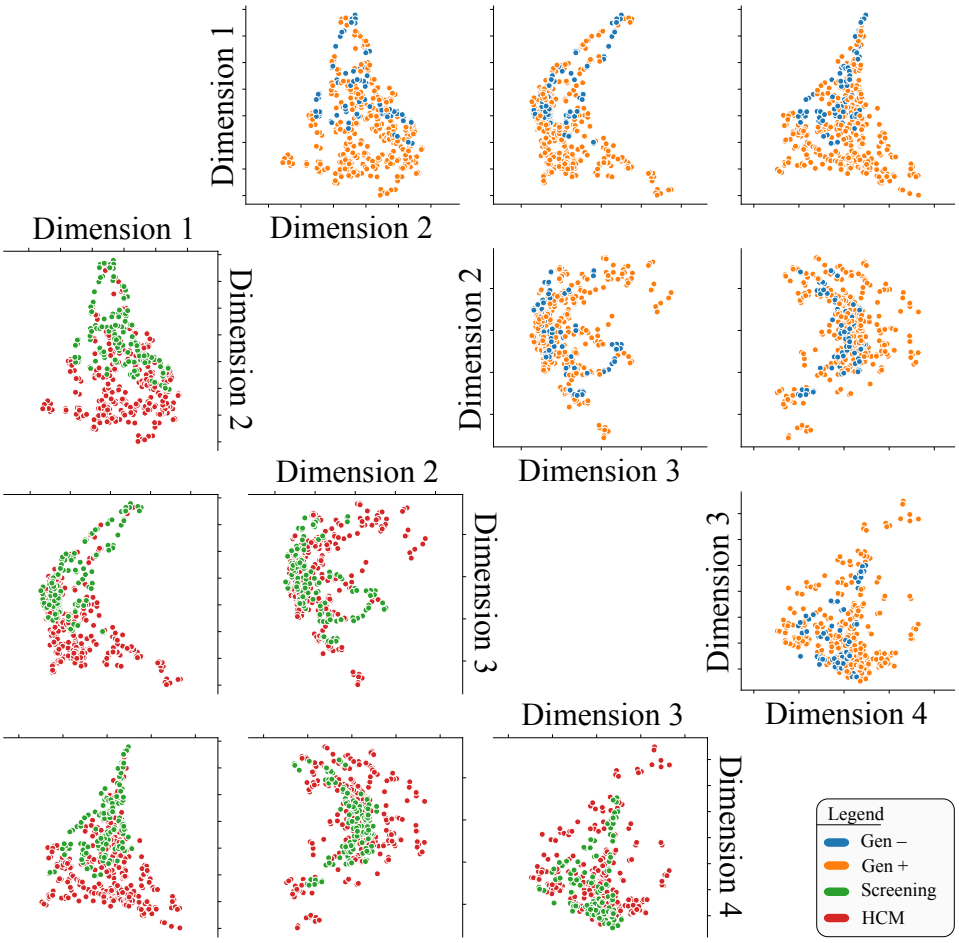


Figure 5.6: Pairplot of the low-dimensional embedding, color-coded according to incidence of genotype positive (upper diagonal) and belonging to screening group (lower diagonal; HCM stands for hypertrophic cardiomyopathy). The pairplot matrix is symmetric with respect to the diagonal.

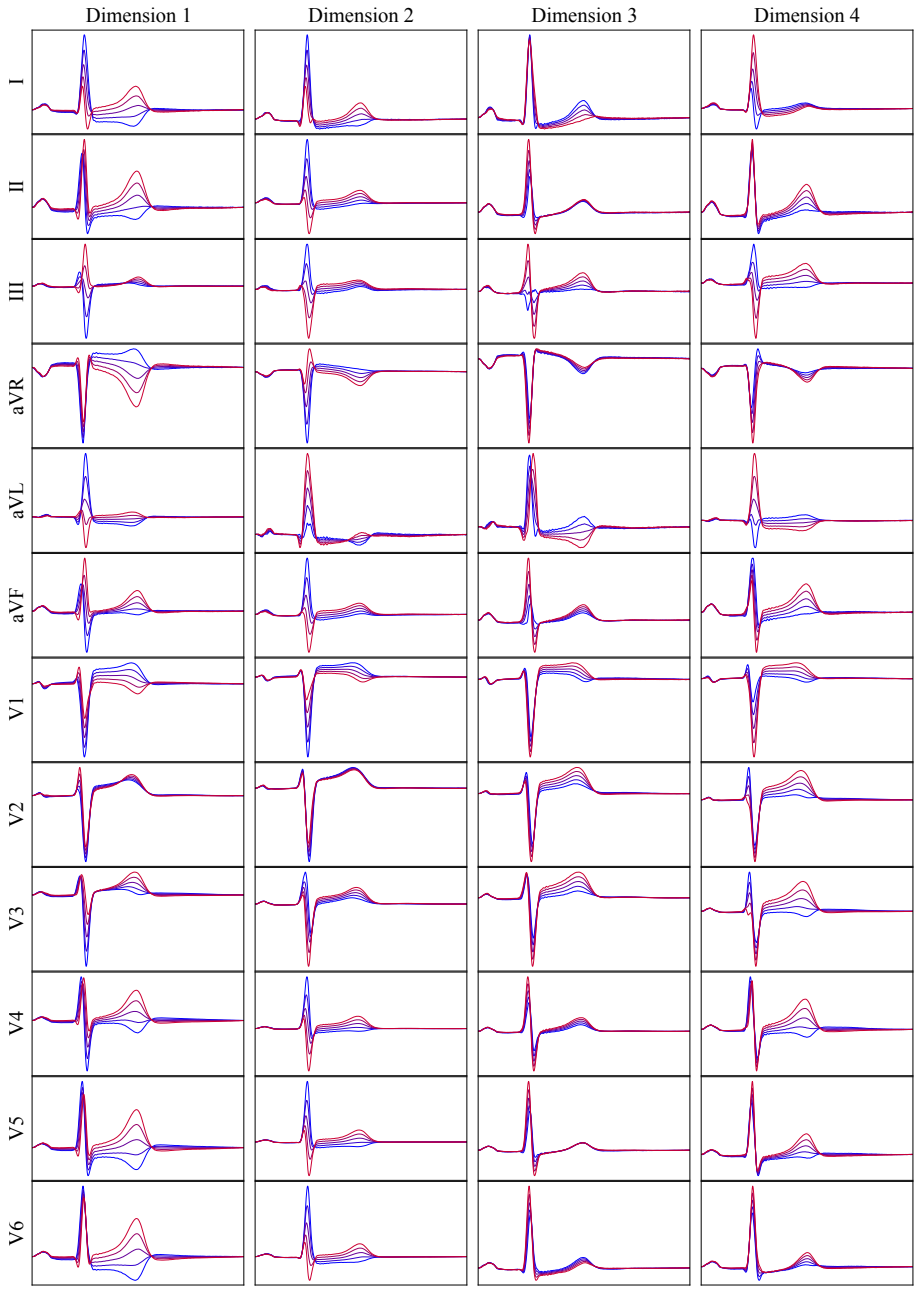


Figure 5.7: Variability captured in the low-dimensional embedding, color-coded as a progression between M − 2SD (blue) and M + 2SD (red), in increments of one SD in the low-dimensional embedding per line. Mean: M; standard deviation: SD.

ST elevations are present, albeit less pronounced. Dimension 3 principally depicts morphology changes in the QRS complex in some leads, with prime examples in leads II and aVF, and TWI in leads aVL and V1. Finally, dimension 4 represents changes in amplitude in the T and QRS waves, and with LAD deviation. This dimension, however, is highly correlated to the first three dimensions, so its contribution is less pronounced.

Some ECG features were also extracted and analyzed to characterize the HCM population, as summarized in Table 5.2. The HCM patients had larger P and QRS wave amplitudes, especially in the precordial leads (V1-V6). The T wave amplitudes, on their behalf, were smaller for some leads (most notably in some limb leads) in HCM patients while larger in others (precordial leads). Moreover, HCM patients displayed a LAD in the QRS (17.11 vs. 48.52 degrees, $p < 0.0001$) and T waves (50 vs 40 degrees, $p < 0.0001$), a high amount of T wave inversions (49% vs. 4%. $p < 0.0001$). Finally, HCM patients presented longer overall intervals, especially in the P (116 vs. 108 ms), QRS (96 vs. 88 ms), ST (108 vs. 96 ms) and T (220 vs. 196 ms) segments, all statistically significant ($p < 0.0001$).

5.4.3. Electrocardiographic-based clustering

The optimal number of clusters for kM were 6 according to the silhouette score, as depicted in Figure 5.4 (right). The unsupervised cluster distribution, depicted in Figure 5.8, was then employed to explore stratification of clinical variables, as summarized in Table 5.3. Clusters 0, 1, 3 and 4 (blue, orange, red and purple in Figure 5.8) were located in the zone with smaller values in dimension 1 and were generally correlated with high or intermediate risk and higher event rate. Cluster 2 (green) was associated with normal phenotype patients, with the lowest HCM Risk-SCD score, wall thickness, rate of events, comorbidities and genotype positive patients. Cluster 5 (brown), predominantly male, corresponded to the second lowest event rate, wall thickness and HCM Risk-SCD score, but with high family history of SCD. Cluster 4 (purple) displayed the highest risk of arrhythmias and aborted SCD, presenting a high percentage of genotype positive, family history of SCD, LAD, TWI and high wall thickness. Cluster 0 (blue), on its behalf, corresponded to HCM patients presenting obstruction, with the largest wall thickness, LVOT obstruction and fibrosis (as explored with LGE-MRI) but, interestingly, with the lowest family history of events. These patients presented interesting ECG features, with the largest TWI prevalence and the highest overall limb lead amplitude. All patients in Cluster 0 (blue) were genotype positive, with the highest percentage of MYBPC3. Cluster 3 (red) had the highest event rate but did not present any aborted SCD, with all patients being genotype positive (pre-

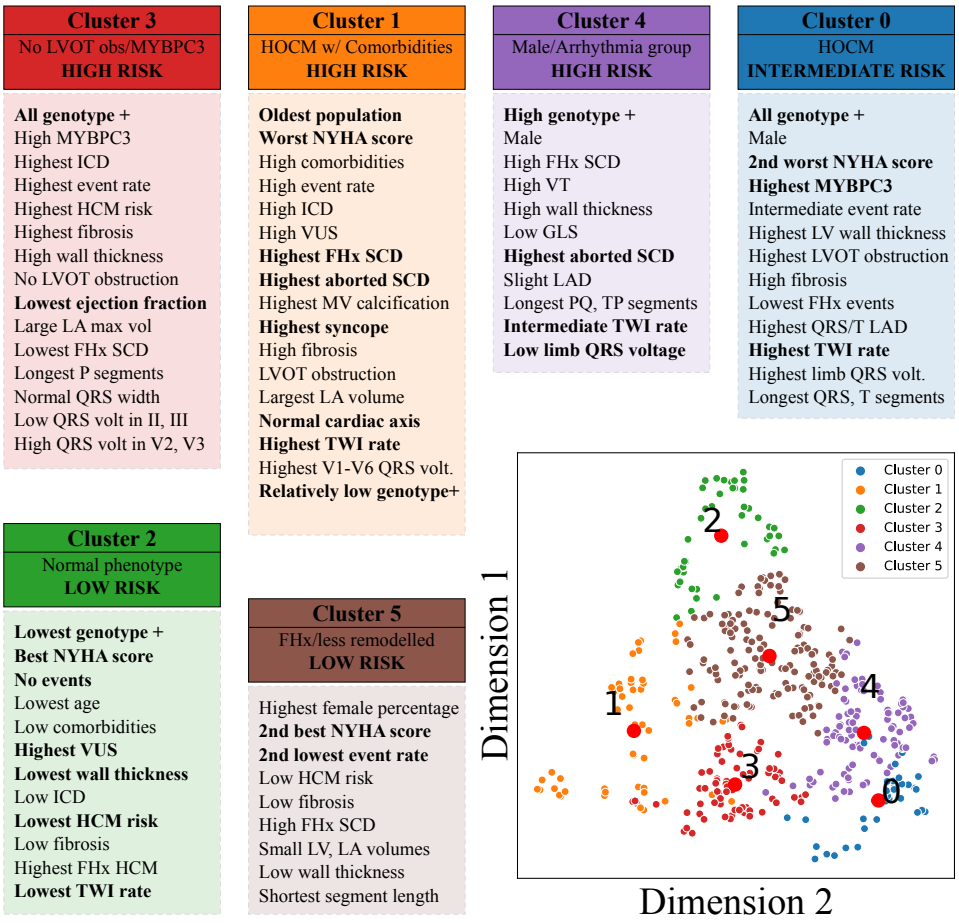


Figure 5.8: Cluster distribution and main characteristics of the extracted phenogroups. Lower values in dimension 1 are correlated with higher event rates and poorer prognosis. Clusters 2 (green) and 5 (brown) account for the lowest risk group, whereas Clusters 0, 1, 3 and 4 represent high-risk patients, with high percentages of genotype positive patients and high event rates. ICD: implantable cardioverter defibrillator; HCM: hypertrophic cardiomyopathy; LVOT: left ventricular outflow tract; FHx: family history; SCD: sudden cardiac death; VUS: variant of uncertain significance; TWI: T wave inversion; LA: left atrial; LV: left ventricular; MV: mitral valve; LAD: left axis deviation; VT: ventricular tachycardia; NYHA: New York Heart Association; GLS: global longitudinal strain.

	Cluster 0 (n = 37)	Cluster 1 (n = 56)	Cluster 2 (n = 45)	Cluster 3 (n = 73)	Cluster 4 (n = 98)	Cluster 5 (n = 147)	p-value
GENERAL INFORMATION							
Age	48.0 (47.0 - 58.0)	53.0 (46.0 - 64.5)	42.0 (40.0 - 49.0)	49.0 (40.0 - 62.0)	46.0 (40.0 - 56.5)	48.0 (38.0 - 58.0)	0.0090
Female	12 (32.43)%	29 (51.79)%	23 (51.11)%	36 (49.32)%	26 (26.53)%	84 (57.14)%	< **0.0001**
BSA	1.95 (1.8 - 2.1)	1.8 (1.7 - 1.93)	1.9 (1.6 - 1.9)	1.8 (1.7 - 1.9)	1.8 (1.8 - 2.05)	1.8 (1.65 - 2.0)	**0.0001**
BMI	25.0 (22.5 - 29.7)	26.0 (24.0 - 30.0)	25.0 (23.0 - 28.0)	25.0 (22.0 - 29.0)	26.0 (24.7 - 28.4)	25.0 (22.22 - 28.0)	0.0287
Systolic BP	128.0 (118.0 - 142.5)	130.0 (120.0 - 145.0)	130.0 (121.0 - 140.0)	125.0 (120.0 - 130.0)	125.0 (115.0 - 144.0)	130.0 (122.5 - 134.5)	0.7094
Diastolic BP	80.12 ± 11.08	77.41 ± 6.89	70.18 ± 12.21	74.94 ± 7.98	83.76 ± 13.96	82.57 ± 7.05	< **0.0001**
MAGNETIC RESONANCE IMAGING/ECHOCARDIOGRAPHIC ANALYSIS							
Heart rate (echo)	65.0 (59.0 - 69.0)	60.0 (56.0 - 61.0)	63.0 (61.0 - 79.0)	63.0 (57.0 - 67.0)	64.0 (55.0 - 72.0)	62.0 (58.0 - 74.0)	0.0066
Maximal WT (echo)	18.0 (17.0 - 19.6)	17.0 (10.0 - 19.8)	10.3 (9.8 - 13.3)	20.4 (16.3 - 24.0)	18.05 (11.5 - 21.7)	11.3 (9.5 - 17.95)	< **0.0001**
EF	49.23 (43.01 - 58.39)	53.1 (41.99 - 55.27)	57.42 (50.64 - 58.81)	47.58 (42.01 - 53.17)	51.29 (45.01 - 55.74)	53.87 (46.25 - 57.54)	0.0001
LGE	16 (59.26)%	18 (62.07)%	6 (28.57)%	37 (78.72)%	34 (61.82)%	18 (29.51)%	< **0.0001**
LA diameter	43.0 (37.0 - 48.0)	41.0 (35.0 - 47.0)	37.0 (33.0 - 44.0)	42.0 (38.0 - 47.0)	39.0 (35.0 - 43.5)	38.0 (35.0 - 43.0)	< **0.0001**
COMORBIDITIES							
Hypertension	6 (16.22)%	17 (30.36)%	6 (13.33)%	15 (20.55)%	11 (11.22)%	26 (17.69)%	0.0752
Diabetes	3 (8.11)%	6 (10.71)%	0 (0.0)%	9 (12.33)%	3 (3.06)%	3 (2.04)%	0.0033
CKD	3 (8.11)%	3 (5.36)%	0 (0.0)%	0 (0.0)%	0 (0.0)%	0 (0.0)%	**0.0001**
SA syndrome	0 (0.0)%	6 (10.71)%	0 (0.0)%	3 (4.11)%	3 (3.06)%	0 (0.0)%	0.0008
Coronary disease	0 (0.0)%	0 (0.0)%	0 (0.0)%	4 (5.48)%	2 (2.04)%	0 (0.0)%	0.0160
History of HF	3 (8.11)%	5 (8.93)%	0 (0.0)%	3 (4.11)%	7 (7.14)%	6 (4.08)%	0.3150
History of arrhythmias	4 (10.81)%	13 (23.21)%	0 (0.0)%	26 (35.62)%	14 (14.29)%	24 (16.33)%	< **0.0001**
Previous intervention	1 (2.7)%	12 (21.43)%	3 (6.67)%	29 (39.73)%	13 (13.27)%	23 (15.65)%	< **0.0001**
GENOTYPE							
Genotype positive	37 (100.0)%	44 (78.57)%	25 (55.56)%	73 (100.0)%	81 (82.65)%	112 (76.19)%	< **0.0001**
Genotype negative	0 (0.0)%	3 (5.36)%	9 (20.0)%	0 (0.0)%	12 (12.24)%	21 (14.29)%	0.0005
Genotype MYBPC3	28 (75.68)%	18 (32.14)%	15 (33.33)%	42 (57.53)%	36 (36.73)%	74 (50.34)%	< **0.0001**
Genotype MYH7	9 (24.32)%	21 (37.5)%	6 (13.33)%	16 (21.92)%	32 (32.65)%	30 (20.41)%	0.0231
Genotype VUS	0 (0.0)%	9 (16.07)%	11 (24.44)%	0 (0.0)%	5 (5.1)%	14 (9.52)%	< **0.0001**
Genotype other	0 (0.0)%	5 (8.93)%	4 (8.89)%	15 (20.55)%	13 (13.27)%	8 (5.44)%	0.0029

Table 5.3: Clinical variable distribution according to K-means clustering. Abbreviated clinical variables are shown; for a more detailed description of the clinical variables, refer to Table 5.1.

	Cluster 0 (n = 37)	Cluster 1 (n = 56)	Cluster 2 (n = 45)	Cluster 3 (n = 73)	Cluster 4 (n = 98)	Cluster 5 (n = 147)	p-value
EVENTS/RISK FACTORS							
ICD	1 (2.7)%	12 (21.43)%	3 (6.67)%	23 (31.51)%	10 (10.2)%	20 (13.61)%	< **0.0001**
MI	0 (0.0)%	0 (0.0)%	0 (0.0)%	0 (0.0)%	0 (0.0)%	0 (0.0)%	1.0000
Syncope	3 (8.11)%	9 (16.07)%	0 (0.0)%	9 (12.33)%	12 (12.24)%	9 (6.12)%	0.04442
SCD	0 (0.0)%	3 (5.36)%	0 (0.0)%	0 (0.0)%	3 (3.06)%	0 (0.0)%	0.0202
AF	6 (16.22)%	8 (14.29)%	0 (0.0)%	3 (4.11)%	10 (10.2)%	9 (6.12)%	0.0206
VT	1 (2.7)%	4 (7.14)%	0 (0.0)%	5 (6.85)%	7 (7.14)%	1 (0.68)%	0.0361
NSVT	4 (10.81)%	9 (16.07)%	0 (0.0)%	12 (16.44)%	2 (2.04)%	18 (12.24)%	0.0023
FHx HCM/SCD	28 (75.68)%	53 (94.64)%	45 (100.0)%	60 (82.19)%	87 (88.78)%	138 (93.88)%	0.0004
Massive LVH	1 (2.7%)	0 (0.0%)	0 (0.0%)	6 (8.22%)	6 (6.12%)	2 (1.36%)	0.0212
LVOT obs.	9 (24.32%)	15 (26.79%)	3 (6.67%)	15 (20.55%)	22 (22.45%)	17 (11.56%)	0.0174
Calcification MV	3 (8.11)%	14 (25.0)%	4 (8.89)%	3 (4.11)%	7 (7.14)%	5 (3.4)%	< **0.0001**
Calcification AV	1 (2.7)%	3 (5.66)%	0 (0.0)%	0 (0.0)%	7 (7.37)%	7 (5.11)%	0.1551
SAM	18 (48.65)%	15 (28.3)%	6 (13.95)%	15 (21.43)%	35 (35.71)%	22 (15.07)%	< **0.0001**
RISK SCORES							
Event 1	7 (18.92)%	16 (28.57)%	0 (0.0)%	32 (43.84)%	23 (23.47)%	30 (20.41)%	< **0.0001**
Event 2	13 (35.14)%	24 (42.86)%	0 (0.0)%	35 (47.95)%	33 (33.67)%	39 (26.53)%	< **0.0001**
HCM risk	2.19 (1.96 - 3.04)	2.3 (1.32 - 4.73)	1.07 (0.92 - 1.75)	2.38 (1.21 - 4.38)	2.28 (1.44 - 3.62)	1.5 (0.97 - 2.49)	0.0012
Cornell	30 (81.08)%	10 (17.86)%	3 (6.67)%	36 (49.32)%	30 (30.61)%	17 (11.56)%	< **0.0001**
Sokolow	7 (18.92)%	24 (42.86)%	5 (11.11)%	21 (28.77)%	3 (3.06)%	1 (0.68)%	< **0.0001**
NYHA score I	22 (59.46%)	20 (35.71%)	39 (86.67%)	52 (71.23%)	71 (72.45%)	126 (85.71%)	< **0.0001**
NYHA score II	12 (32.43%)	30 (53.57%)	6 (13.33%)	12 (16.44%)	12 (12.24%)	15 (10.2%)	< **0.0001**
NYHA score III	3 (8.11%)	6 (10.71%)	0 (0.0%)	9 (12.33%)	15 (15.31%)	6 (4.08%)	0.0089

Table 5.3: (continued) Clinical variable distribution according to K-means clustering. Abbreviated clinical variables are shown; for a more detailed description of the clinical variables, refer to Table 5.1.

dominantly MYBPC3). This group had the highest rate of ICD, HCM Risk-SCD score and fibrosis; and high overall wall thickness and low ejection fraction. On the other hand, the group did not show LVOT obstruction and had a low family history of SCD. Finally, Cluster 1 (orange) was the oldest group and included patients with the highest comorbidities. Additionally, it was the group with the second highest event rate, and coincided with the highest amount of aborted SCD, unexplained syncope, family history of SCD and mitral valve calcification; and an overall high fibrosis and LVOT obstruction. Finally, the group had also the largest TWI percentage and overall precordial lead QRS amplitude but normal QRS axis. Figure 5.9 shows the Cluster centroids for this partition, illustrating some of the aforementioned differences between groups.

5.4.4. Phenotyping along the dimension with highest variability

A secondary analysis explored the distribution of clinical variables with respect to clusters formed alongside dimension 1, which accounted for the majority of the variability of the output space (55.17%) as compared to dimensions 2, 3 and 4 (39.29%, 0.4% and 2.32%, respectively). For this purpose, the space was partitioned in regions of equal size, dividing the space into N strips. The Silhouette score was computed, establishing the optimal number of divisions to 3 (Figure 5.5). The exploration of the clinical variables associated to this partitioning is summarised in Table 5.4, where it can be seen that the strip with the smallest values in dimension 1 presented a larger amount of events (36.96 vs. 16.74 vs. 7.02% for Strip 0, 1 and 2, respectively, and from bottom to top; $p < 0.0001$), genotype positive patients (97.28 vs. 73.49 vs. 61.4%), TWIs (65.22 vs. 16.74 vs. 1.75%, $p < 0.0001$), fibrosis (67.5 vs. 43.3 vs. 26.09%, $p < 0.0001$), wall thickness (18.3 vs. 15.3 vs. 10.3 mm, $p < 0.0001$) and HCM Risk-SCD score (2.32 vs. 1.57 vs. 1.11, $p < 0.0001$), among others. A similar analysis performed with a larger number of points ($N = 10$) yielded similar results, but in a more granular manner.

5.5. Discussion

Although ECG features are not usually employed for risk stratification in patients with HCM, unsupervised DR algorithms have shown great promise at characterizing ECG phenotypes of high SCD risk [34]. This work contributes to this objective through the development of a pipeline based on automatic ECG delineation, projection to a common reference system, and the posterior utilization of the raw signal in DR algorithms. By employing MKL, different descriptors are agglomerated into a single low-dimensional embedding [14, 15, 160], leveraging the whole ECG trace to explore how clinical variables relate to the variability found in the multiple

	Strip 0 (blue)	Strip 1 (orange)	Strip 2 (green)	p-value
GENERAL INFORMATION				
Age	48.0 (41.5 - 60.5)	47.0 (38.5 - 57.0)	42.0 (40.0 - 49.0)	0.0184
Female	67 (36.41)%	116 (53.95)%	27 (47.37)%	0.0021
BMI	25.0 (23.0 - 29.0)	26.0 (23.0 - 28.34)	24.16 (23.0 - 27.0)	0.1932
SBP	125.0 (120.0 - 140.0)	130.0 (118.5 - 138.5)	130.0 (123.0 - 140.0)	0.9160
DBP	78.41 $\pm$ 9.98	82.03 $\pm$ 9.97	71.89 $\pm$ 12.58	0.0015
MAGNETIC RESONANCE IMAGING/ECHOCARDIOGRAPHIC ANALYSIS				
Heart rate	61.0 (57.0 - 68.0)	62.0 (58.0 - 74.0)	63.0 (59.0 - 78.0)	0.0069
MWT	18.3 (16.4 - 21.7)	15.3 (9.9 - 19.0)	10.3 (9.1 - 13.4)	$<$ **0.0001**
EF	49.81 (43.25 - 56.05)	52.45 (45.12 - 56.88)	56.33 (48.16 - 58.81)	0.0048
Fibrosis	81 (67.5)%	42 (43.3)%	6 (26.09)%	$<$ **0.0001**
LA diameter	42.0 (38.0 - 47.0)	38.0 (35.0 - 44.0)	37.0 (35.0 - 43.0)	$<$ **0.0001**
COMORBIDITIES				
Hypertension	35 (19.02)%	39 (18.14)%	7 (12.28)%	0.4983
Diabetes	12 (6.52)%	12 (5.58)%	0 (0.0)%	0.1500
CKD	6 (3.26)%	0 (0.0)%	0 (0.0)%	0.0112
SA syndrome	6 (3.26)%	6 (2.79)%	0 (0.0)%	0.3973
Coronary dis.	6 (3.26)%	0 (0.0)%	0 (0.0)%	0.0112
Hx of HF	13 (7.07)%	11 (5.12)%	0 (0.0)%	0.1122
Hx of arrhy.	50 (27.17)%	28 (13.02)%	3 (5.26)%	$<$ **0.0001**
Previous CI	44 (23.91)%	34 (15.81)%	3 (5.26)%	0.0033
GENOTYPE				
Gen+	179 (97.28)%	158 (73.49)%	35 (61.4)%	$<$ **0.0001**
Gen-	5 (2.72)%	28 (13.02)%	12 (21.05)%	$<$ **0.0001**
MYBPC3	99 (53.8)%	90 (41.86)%	24 (42.11)%	0.0442
MYH7	54 (29.35)%	55 (25.58)%	5 (8.77)%	0.0071
VUS	0 (0.0)%	29 (13.49)%	10 (17.54)%	$<$ **0.0001**
Other	26 (14.13)%	13 (6.05)%	6 (10.53)%	0.0258
EVENTS/RISK FACTORS				
ICD	38 (20.65)%	28 (13.02)%	3 (5.26)%	0.0089
MI	0 (0)%	0 (0)%	0 (0)%	1.0000
Syncope	24 (13.04)%	17 (7.91)%	1 (1.75)%	0.0240
Aborted SCD	3 (1.63)%	3 (1.4)%	0 (0.0)%	0.6342
AF	24 (13.04)%	12 (5.58)%	0 (0.0)%	0.0014
VT	11 (5.98)%	7 (3.26)%	0 (0.0)%	0.0995
NSVT	26 (14.13)%	16 (7.44)%	3 (5.26)%	0.0380
FHx HCM/SCD	154 (83.7)%	203 (94.42)%	54 (94.74)%	0.0008
Massive LVH	10 (5.43%)	5 (2.33%)	0 (0.0%)	0.0732
LVOT obst.	43 (23.37%)	32 (14.88%)	6 (10.53%)	0.0270
Calcified MV	19 (10.33)%	11 (5.12)%	6 (10.53)%	0.1153
Calcified AV	9 (5.06)%	9 (4.48)%	0 (0.0)%	0.2585
RISK SCORES				
Event 1	68 (36.96)%	36 (16.74)%	4 (7.02)%	$<$ **0.0001**
Event 2	87 (47.28)%	53 (24.65)%	4 (7.02)%	$<$ **0.0001**
HCM risk	2.32 (1.81 - 4.54)	1.57 (0.92 - 2.83)	1.11 (0.92 - 2.02)	$<$ **0.0001**
Cornell	99 (53.8)%	24 (11.16)%	3 (5.26)%	$<$ **0.0001**
Sokolow	50 (27.17)%	5 (2.33)%	6 (10.53)%	$<$ **0.0001**
NYHA I	113 (61.41%)	166 (77.21%)	51 (89.47%)	$<$ **0.0001**
NYHA II	47 (25.54%)	34 (15.81%)	6 (10.53%)	0.0102
NYHA III/IV	24 (13.04%)	15 (6.98%)	0 (0.0%)	0.0046

Table 5.4: Variable distribution of line clustering on dimension 1. A more detailed description of clinical variables (including events) is included in Table 5.1.

leads of standard ECGs. MKL allowed the identification of several phenogroups with distinct risk profiles, from phenotype-negative patients (and mostly genotype--negative) to high-risk patient who suffer obstructive HCM or groups with high prevalence of arrhythmias. This analysis, however, presented some challenges when performing data pre-processing, as different cardiac cycles differ in the relative time allocation within the cardiac cycle (e.g., longer T waves or shorter PQ segments), causing mismatches in the input data when comparing the morphology of different patients. To avoid these mismatches, the 12-lead ECG was firstly cropped in a PP' interval, and then registered to segments of fixed length, matching every phase of the cardiac cycle (Figure 5.1). Finally, for each patient, the 12-lead ECG and a vector containing the original segment lengths in the cardiac cycle was retrieved and processed with MKL. This registration to a common reference system allowed the direct morphological comparison among observations while retaining temporal information associated to them.

An initial exploration of the output space showed a clear gradation of HCM and screening patients (Figure 5.8). As analyzed in Section 5.4.2, the HCM and screening populations mostly differed in the morphology of the QRS complex, T wave and ST segment, as is usually reported in the literature. Other segments did not seem to vary greatly in the input data (Figure 5.7). Moreover, analyzing the captured variability of each dimension allowed the identification of their most important contributions to the overall morphology. Dimension 1 was mainly related with TWI, higher P, QRS and T wave amplitudes, PQ and ST elevation and the majority of the variability in segment durations (P, QRS, ST, T and TP). Dimension 2 also reflected changes in QRS amplitude and ST elevation, but a clearer gradation in QRS axis can be observed. Besides this initial exploration of the output space, the ECG characteristics of the screening and HCM populations were analyzed, which hinted at some factors that differentiated both groups (Table 5.2). Out of the extracted ECG characteristics, LAD in the QRS complex, TWIs, representative changes in the QRS, ST and T durations and overall increased amplitude in the QRS complex and in the T wave in precordial leads depicted changes of statistical significance between the sub-populations. Other imaging markers such as increased LV wall thickness, more prevalence of fibrosis or increased LA diameter were also significant. The differences found in our work are consistent with the findings in the population selected by Lyon *et al.* [34].

This initial exploration was followed by phenogrouping of the output space. Two types of clustering were performed: using kM and dividing the space in equisized strips alongside dimension 1, which encoded the largest variability. Firstly, kM was used to distinguish groups in the output space in a data-driven manner, where the optimal number of clusters (6) was selected using the Silhouette score. These

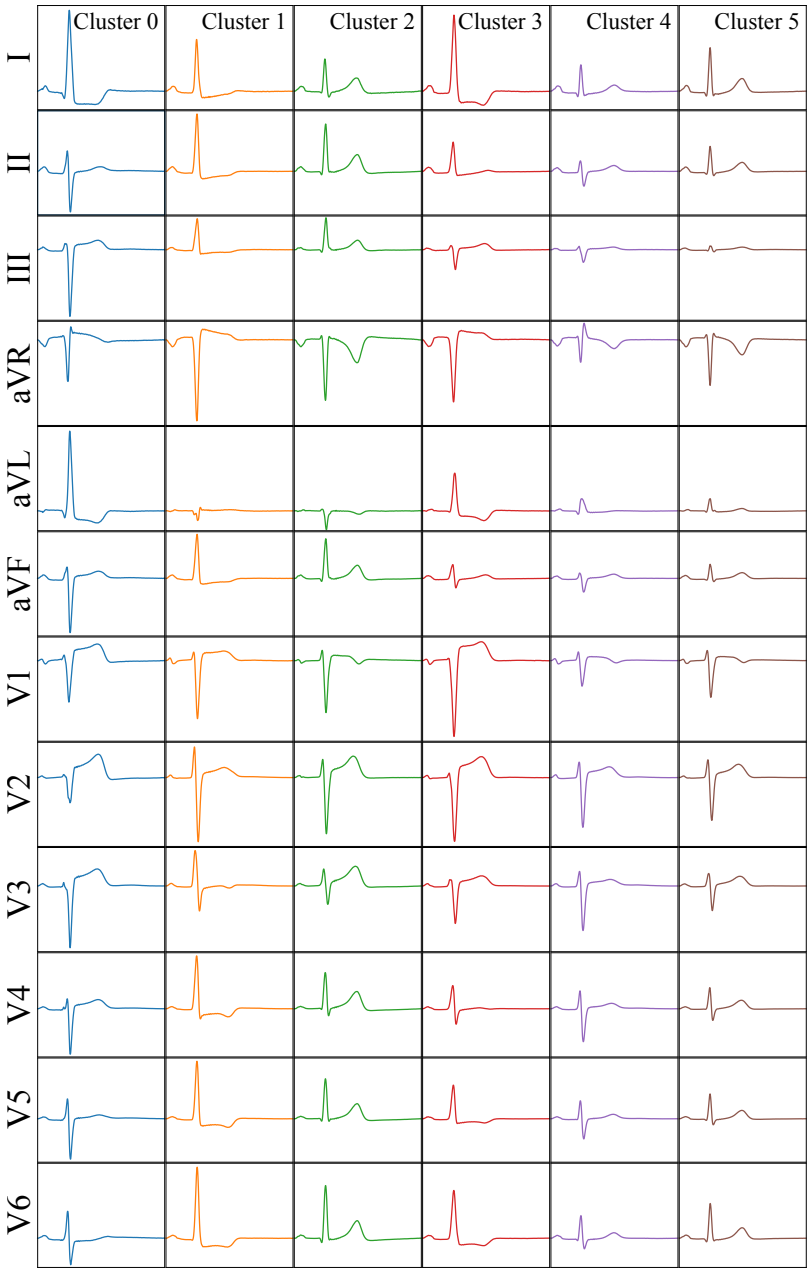


Figure 5.9: Cluster centroids, color-coded according to their associated cluster. In low-risk groups (Clusters 2, 5), low overall QRS amplitude is observed, as well as no T wave inversions. In high-risk groups (Clusters 0, 1, 3 and 4), a higher prevalence of T wave inversions, axis deviations and ST depressions.

clusters yielded two low-risk groups and four high-risk groups, as depicted in Figure 5.8. Some salient features could be identified in these groups according to the patient's clinical history. Cluster 4 (purple) characterized patients with high risk of arrhythmias. Cluster 1 corresponded with patients with a high rate of comorbidities. Cluster 5 (brown, low risk) was morphologically similar to Cluster 2 (green, normal phenotype) in most leads, but presented slight left axis deviation (polarity inversions in leads III, aVL, aVF), overall reduced T wave amplitude and a higher prevalence of TWI. The groups were also correlated with the relative amount of genotype positive patients, with Cluster 2 (green) accounting for normal phenotype and the lowest amount of genotype positive patients, whereas Clusters 0 and 3 (blue, red) contained genotype positive-only patients. Clustering also allowed for the differentiation between groups of increased hypertrophy, fibrosis and HCM Risk-SCD score, with high-risk groups depicting generally increased risk as compared to low-risk groups, which is compatible with the disease's aetiology [153]. As in other works [34], high-risk clusters contained patients that did not reach classic ECG-based criteria for HCM, such as Cornell or Sokolow scores [166].

A secondary analysis was performed, dividing dimension 1 into N strips of equal size (Table 5.4). The Silhouette score was also employed as a single figure of merit to distinguish the best number of divisions, saturating at three strips (Figure 5.4). A good match could be found between the Strips and the Clusters found using kM: Strip 0 roughly corresponded to the high-risk clusters (Clusters 0, 1, 3 and 4); Strip 1 approximately corresponded to Cluster 5 (low risk, 2nd lowest event rate); and Strip 2 had a good match with Cluster 2 (normal phenotype, lowest event rate). The computed Strips correlated strongly with HCM Risk-SCD score, with smaller values in the 1st dimension presenting worse scores (2.32%, 1.57% and 1.11% for Strips 0, 1 and 2, respectively), and more events (36.96%, 16.74% and 7.02%). Moreover, the gradation along dimension 1 was associated with statistically significant changes in QRS amplitude, QRS and T wave axis, QRS, T and ST segment duration and ST elevation ($p < 0.0001$). This demonstrates the effectiveness of the presented approach to stratify patients according to ECG features, and further solidified dimension 1 as the main axis of variability, corresponding with the most important ECG factors for HCM exploration. However, further research would be needed to address whether the changes can be related to HCM or any comorbidities.

Although not many computational approaches exist in the literature, Lyon *et al.* [34] developed a similar methodology by utilizing another dimensionality reduction algorithm (PCA) alongside hand-crafted features extracted from the delineated ECG trace. The hand-crafted features consisted in the first four basis function coefficients from the Hermite transform, applied to the QRS complex mor-

phology, and the presence/absence of TWI over the averaged ECG from 30 minute excerpts. The authors reported that the generated output space allowed the identification of four main clusters of data, as computed by DBSCAN, another unsupervised algorithm. Posterior analyses demonstrated good separation with respect to HCM Risk-SCD scores (1A: 4.0%, 1B: 1.8%, 2: 2.1%, 3: 2.5%, p = 0.0001) over their 85 HCM/38 healthy controls population, suggesting the viability of ECG for risk stratification. Their findings suggested that TWI in patients with normal QRS as well as QRS abnormalities in precordial leads V4-V6 (in patients with and without TWI) had a higher risk of arrhythmia and SCD. Thus, the authors hint at the role of ECG to capture ionic and structural abnormalities in its trace, being reflected in the depolarization and repolarization phases. In a similar note to Lyon *et al*. [34], our findings seemed to relate clearly the occurrence of TWI, QRS axis deviations and increased precordial QRS amplitudes to higher risk of HCM; even coinciding to identify high-risk groups with TWI and normal QRS axis. Moreover, our approach also provided clear separation of populations with respect to the HCM Risk-SCD scores (Table 5.3): high-clusters 0, 1, 3 and 4 had scores of 2.19%, 2.3%, 2.38% and 2.28%, respectively, whereas low-risk clusters 2 and 5 obtained scores of 1.07% and 1.5%, respectively ($p < 0.0001$). Our approach, however, was able to incorporate the full ECG trace in the analysis, uncovering significant changes in ST elevation, segment durations and late precordial transitions as interesting clinical markers to explore. Furthermore, the computed output space allowed to establish a gradation of the disease, where patients slowly transition to areas of higher risk over time, as opposed to including binary markers such as TWI [34] or to all-or-nothing approaches such as classification approaches [48, 49, 152].

Limitations and future work

This work, however, presents some limitations. Firstly, a more in-depth characterization of the output space is needed, as rationalizing the features that distinguish a cluster from another is labour-intensive. Although this has partially been performed, the characterization is incomplete, as some contributing factors are statistically significant but poorly explored (e.g., many P wave markers showed statistically significant changes, but these have not been analysed yet), or require more derived ECG features that would be able to distinguish pathological waves from normal (e.g., precordial transition, biphasicity of the wave, etc.), which would make the posterior analysis easier. Another limitation is the employed algorithm: although MKL has proven its usefulness in similar approaches and in a wide variety of tasks [15, 18, 55], it scales poorly to large amounts of data, needing to optimize for large tensors that increase computational time as the cube of the number

of observations. Although clinical problems that are often studied with small-scale datasets are abundant (due to the difficulty in recruiting a large amount of patients), a general solution for large-scale problems is of key importance to bridge the gap between interpretable solutions and the predictive power of deep learning (DL) algorithms. A final limitation is the dataset itself, where the population is faced with large amounts of comorbidities and a small amount of SCD-related events, which might limit the applicability of our approach. Multi-centric approaches for recruiting a more comprehensive population to address the generalizability of our approach are required [167]. Moreover, as mentioned by [34], many ECG test are still performed in non-digital machines, which limit the applicability of this approach to the general case.

Besides these limitations, this work could be extended to offer a more comprehensive analysis of HCM. Firstly, we are working towards exploring the effects of the discovered phenogroups in other data descriptors, such as Doppler curves. The analysis of raw Doppler curves have shown to be challenging, masking the relationships in the data to produce interpretable results. Producing a similar pre-processing for echocardiographic data or including ECG information might aid in disentangling the information present in Doppler traces. Another extension possibility is the application of sequence analysis [18], which could aid in addressing changes over time in patients who are suspected of suffering MKL. The analysis of trajectories in a low-dimensional embedding might further aid in risk stratification, concentrating on personalized evolution rather than on the analysis of decontextualized cardiac cycles.

5.6. Conclusions

In this work, a pipeline was developed for the analysis of standard ECG recordings, addressing the modality's capability to aid in patient risk stratification. Multiple kernel learning was employed for this purpose, which is a multi-descriptor dimensionality reduction algorithm that comparing samples directly in a low-dimensional embedding. The multi-descriptor formulation of the algorithm allows the usage of the raw ECG traces as input, as opposed to hand-crafted surrogates of ECG morphology and function, and allows for exploring the variability within the input database. This permitted a better characterization of the output space, allowing the identification of phenogroups that share some morphological and clinical characteristics, clearly discriminating between high-risk and low-risk groups. Moreover, the space allows to examine the disease's progression due to its continuous nature, providing a gradation of the characteristics indicative of disease.

CHAPTER

6

General discussion and conclusions

This thesis has delved into the development of computational algorithms for reducing clinical workload when processing and interpreting cardiac signals. Specifically, deep learning (DL) algorithms were selected as the main analysis tool due to their high performance under a wide array of tasks. This work, however, deviates from the utilization of data-driven approaches for classification, as the uninterpretable nature of these algorithms cause a disconnect in the clinical hypothetico-deductive method, disrupting the link between the specific set of symptoms that lead to a specific diagnosis. To depart from these methods while profiting from state-of-the-art computational methods, this thesis has concentrated on creating high-quality and robust quantification systems. The developed algorithms can, in turn, be used for a wide variety of downstream tasks, such as decision support systems, research of risk stratification algorithms and automatization of the extraction of routine clinical markers.

The main tools employed in cardiac electrophysiology (EP) studies are the electrocardiogram (ECG) and intracavitary electrocardiograms (iECG). The ECG is the main cardiac diagnostic, screening and risk stratification tool, with million of exams performed yearly throughout the world. The ECG contains valuable information about the normal functioning of the heart during the different phases of the cardiac cycle, which is reflected on the different waves represented in its trace: the P, QRS and T waves. iECG recordings, on their behalf, contain valuable local information regarding the relative timing, amplitude and morphology of local acti-

vation patterns arising in the myocardial surface as captured with a catheter when the depolarization wave traverses its electrodes. This information can aid at localizing areas that present proarrhythmic properties, such as partially viable tissue in patients with previous myocardial infarctions (MI) or conduction after pulmonary veins (PVs) isolation in atrial fibrillation (AF) procedures. This work, thus, aims at reducing clinical workload by performing cardiac signal delineation. In chapter 2 and chapter 3, an algorithm was developed for the robust delineation of ECG recordings. In chapter 4, an iECG delineation tool was developed with a similar methodology employed for surface ECG delineation.

Chapter 2 performed a first exploration of the computational possibilities of DL algorithms to produce high-quality surface ECG delineation. For that purpose, a large-scale model ablation was performed around the U-Net, a DL model usually employed in medical image segmentation, to assess the different design decisions that would lead to better performing models. To the best of our knowledge, ours was the first approach to transfer the well-performing algorithms arising in the medical imaging community to the cardiac signal analysis community [81]. However, attaining a model with optimal delineation performance proved difficult given the limitations of the development dataset, which presented high intra- and inter-patient redundancy and a relatively small sample size. A solution for that problem was sought in chapter 3. In that work, a novel synthetic data augmentation (DA) algorithm was developed alongside custom loss functions and more advanced DL architectures existing in the literature. The main methodological contribution was the synthetic DA algorithm, consisting in composing plausible traces of cardiac cycles given a pool of annotated data and a rule-based algorithm for its final arrangement. This algorithm allowed the generation of data samples that reinforced the algorithm's performance while extending the scope of the training data, as the composition was flexible enough to allow for extending beyond the original data manifold, and facilitating explicit prior imposition, thanks to the rule-based composition algorithm. This methodology proved useful for producing a high--quality algorithm, greatly reducing delineation errors in ECG samples outside the development dataset in a large number of target applications. The developed model has proven useful in the analysis of patients with hypertrophic cardiomyopathy (HCM), tetralogy of Fallot, long QT syndrome, as well as in patients with diverse cardiac pathologies undergoing radio frequency ablation or in pre-processing Doppler images for locating cardiac cycles.

The developed methodology for the analysis of surface ECG recordings was adapted for iECG delineation, as detailed in chapter 4. However, this work posed its own challenges. Firstly, no open delineation dataset exists for this modality, so the work was accompanied with ground truth generation for a private dataset. Sec-

ondly, iECG recordings present waveforms of interest that might overlap, which is not the case in ECG signals: a local field (LF) activation corresponding to a late potential (LP) might coincide with a far field (FF) activation; some conditions can cause the occurrence of escape rhythms, in which the electrical impulse is generated in the atrioventricular node, causing simultaneous atrial and ventricular depolarization, which is registered as simultaneous LF and FF activations. Thirdly, there is ambiguity when distinguishing LF and FF activations in some cases and their relationship with the underlying phase in the cardiac cycle, as the catheter probes the electrical activation at specific portions of myocardial tissue that have different roles in the heart's electrical conduction system of the heart. Finally, a delineation that is more sensible than specific is sought for, as delineations would be compared in a beat-to-beat basis and posterior analysis can aid at discriminating false positives. Despite these challenges, the model developed with synthetic DA demonstrated a good sensitivity when localizing local activations and great promise at localizing areas with decremental activity.

Finally, in chapter 5, the developed ECG delineator was employed for the characterization of a population of patients with HCM, exploring signal-based patterns that might aid in risk stratification. multiple kernel learning (MKL) was employed, a dimensionality reduction (DR) algorithm employed for unsupervised exploration of similitude between observations in a population. Analyzing raw signals, however, posed a problem that had to be circumvented to draw similarities between samples: ECG signals present mismatches due to inter- and intra-patient differences, as asynchronies might arise during the cardiac cycle. This was addressed through delineating the ECG recordings, for posteriorly registering the different phases of the cardiac cycles to segments of fixed length, virtually creating a common reference system. This common reference system permitted to draw inferences from direct morphological comparisons between samples, grouping the data into pools of patients that shared similar morphologies. Several cluster analyses of the resulting low-dimensional embedding allowed the identification high-risk phenogroups, and the characterization of T wave inversion (TWI), left axis deviation (LAD), ST elevation and late precordial transitions as interesting ECG markers for risk stratification.

Limitations and future work

This work, however, presents several limitations that require addressing for improving performance and assessing the real generalizability of the developed analysis pipelines.

Firstly, the difficulty in generating high-quality delineation approaches for cardiac

signals are due to the lack of large corpora of annotated data. This is especially the case in public entities such as universities and hospitals. If public initiatives such as the creation of coordinated data gathering centers by the European Union [168] yield results, the need for innovative data augmentation strategies such as those developed in this work would be barren, as appropriate population variability could be captured through the retrieval of high-quality annotated data. Moreover, project-specific data augmentation strategies are difficult to port to other data modalities or clinical problems. As an example, whereas composing a synthetic signal in one-dimensional datasets is a feasible (and even computationally inexpensive), the computational cost of "squaring" the complexity in two-dimensional images might render these approaches unfeasible in other data domains. Nevertheless, computational solutions such as self- or semi-supervised approaches are already being applied to medical image analysis [134, 169]. Besides the difficulties posed by the small-data nature of these approaches, the developed delineation solutions face other limitations. On the first hand, the DL field evolves so rapidly that many modifications could be performed to the baseline architecture, or even its complete replacement by better-performing models. On the other hand, obvious limitations exist in the synthetic DA approaches themselves; although they present, in our opinion, advantages over current data-driven approaches such as generative adversarial networks, due to their ability to extend outside the input data's manifold and to control explicit prior imposition, a rule-based approach is time-consuming to produce and fine-tune, blurring the gap between digital signal processing (DSP) and DL-based approaches. Despite these challenges, many projects are employing the currently developed delineation tool with a high success rate, comprising the analysis of the site of origin (SoO) of outflow tract ventricular arrhythmias or the development of tools for the analysis of patients with AF, tetralogy of Fallot, Long QT syndrome in neonates and Brugada syndrome. Moreover, the delineation tool is also being used to pre-process Doppler images to locate beginnings and ends of cardiac cycles, simplifying downstream analysis tasks.

Secondly, in the specific case of iECG delineation, the proposed approach might require a more in-depth adaptation to the data modality further than re-coding the DA pipeline. A first problem is the existence of overlapping LF activations (or LPs), although clearly distinguishable, such as in waves that have very different frequential profile. This has been partially covered in this work, with LFs being detected within FFs, but further efforts are needed to better localize these waves. Reformulating the problem as an instance segmentation problem, where the segmentations might be of the same class but pertain to distinct objects, producing non-overlapping segmentations. A second problem consists in the difficulty even for human operators to distinguish isolated local components in certain patholo-

gies with very high firing rate or with highly fractionated electrograms. A lack of consensus on how to annotate a signal might hinder the adaptation of a fully working solution. These limitations are, however, consistent with the very experimental phase of this work, and different solutions will be attempted in the following months, comprising the implementation of state-of-the-art segmentation architectures, dataset re-annotation, real data incorporation and an overall improvement in the synthetic data generation pipeline.

Finally, the pipeline developed for HCM risk stratification shows promising results. However, the specific links between ECG findings and the disease's pathophysiology need to be further studied, through proposing specific electrophysiological mechanisms compatible with such morphological changes. Moreover, a larger patient cohort needs to be recruited to verify the generalizability of the proposed approach, and the findings would benefit from their comparison to findings in imaging modalities such as echocardiography or magnetic resonance imaging (MRI). Other markers besides purely ECG-based could also be interesting to analyze, e.g. by including Doppler traces. Finally, the employed DR algorithm would also benefit from improving the formulation to one that is able to take into account much larger cohorts of patients, even comprising other cardiovascular diseases. Algorithms that agglomerate large data pools will improve the generalizability of the approach by producing richer low-dimensional embeddings or latent spaces, where the captured variability is maximal and the inter-dimensional correlation is minimized. For this purpose, extending the MKL approach by adapting it to a formulation compatible with batch-based approaches and stochastic optimization is a promising venture for expansion. Other approaches such as variational autoencoderss will also be explored.

Bibliography

[1] Paul A. Iaizzo. *Handbook of Cardiac Anatomy, Physiology, and Devices*. Springer International Publishing, Cham, 2015.

[2] K.N. Hor and A.J. Trask. Cardiac Anatomy, Physiology, and Pathophysiology. In Ramachandran S. Vasan and Douglas B. Sawyer, editors, *Encyclopedia of Cardiovascular Research and Medicine*, pages 373–383. Elsevier, Oxford, 2018.

[3] Ziad F. Issa, John M. Miller, and Douglas P. Zipes. Braunwald's Heart Disease Family of Books. In Ziad F. Issa, John M. Miller, and Douglas P. Zipes, editors, *Clinical Arrhythmology and Electrophysiology (Third Edition)*, pages xi–xii. Elsevier, Philadelphia, third edition edition, 2019.

[4] Robert J. Laird and Scot Irwin. Chapter 1 - Cardiovascular Structure and Function. In Scot Irwin and Jan Stephen Tecklin, editors, *Cardiopulmonary Physical Therapy (Fourth Edition)*, pages 3–38. Mosby, Saint Louis, fourth edition edition, 2004.

[5] Kim E. Barrett, Susan M. Barman, Scott Boitano, and Heddwen L. Brooks. Chapter 31 - The Heart as a Pump. In Kim E. Barrett, Susan M. Barman, Scott Boitano, and Heddwen L. Brooks, editors, *Ganong's Review of Medical Physiology (Twenty-Third Edition)*, pages 507–520. McGraw Hill Medical, twenty-third edition edition, 2010.

[6] Mark A. Hill. Cardiovascular System - Heart Histology. In Mark A. Hill, ed-

itor, *Embryology*. University of New South Wales, 22 July 2021. Retrieved from https://embryology.med.unsw.edu.au/embryology/index.php/Cardiovascular_System_-_Heart_Histology.

[7] Andrew L Wit, Hein J Wellens, and Mark E Josephson. *Electrophysiological Foundations of Cardiac Arrhythmias: A Bridge Between Basic Mechanisms and Clinical Electrophysiology*. Cardiotext Publishing, 2017.

[8] Afzal Sohaib. *Decoding Cardiac Electrophysiology: Understanding the Techniques and Defining the Jargon*. Springer Nature, 2019.

[9] Sandeep Prabhu and Afzal Sohaib. The basic language of cardiac electrophysiology—an introduction to intracardiac electrograms and electrophysiology studies. In *Decoding Cardiac Electrophysiology*, pages 3–19. Springer, 2020.

[10] Gail D. Baura. Chapter 2 - Electrocardiographs. In Gail D. Baura, editor, *Medical Device Technologies*, pages 39–57. Academic Press, Oxford, 2012.

[11] Brent G Petty. *Basic electrocardiography*. Springer Nature, 2020.

[12] Ana Mincholé, Julià Camps, Aurore Lyon, and Blanca Rodríguez. Machine learning in the electrocardiogram. *Journal of Electrocardiology*, 57:S61–S64, 2019.

[13] Ana Mincholé and Blanca Rodriguez. Artificial intelligence for the electrocardiogram. *Nature Medicine*, 25(1):22–23, Jan 2019.

[14] Yen-Yu Lin, Tyng-Luh Liu, and Chiou-Shann Fuh. Multiple Kernel Learning for Dimensionality Reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(6):1147–1160, 2011.

[15] Sergio Sanchez-Martinez, Nicolas Duchateau, Tamas Erdei, Alan G. Fraser, Bart H. Bijnens, and Gemma Piella. Characterization of myocardial motion patterns by unsupervised multiple kernel learning. *Medical Image Analysis*, 35:70–82, 2017.

[16] P. Garcia-Canadilla, S. Sanchez-Martinez, F. Crispi, and B. Bijnens. Machine Learning in Fetal Cardiology: What to Expect. *Fetal Diagnosis and Therapy*, 47(5):363–372, 2020.

[17] Maja Cikes, Sergio Sanchez-Martinez, Brian Claggett, Nicolas Duchateau, Gemma Piella, Constantine Butakoff, Anne Catherine Pouleur, Dorit Knappe, Tor Biering-Sørensen, Valentina Kutyifa, Arthur Moss, Kenneth Stein, Scott D. Solomon, and Bart Bijnens. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *European Journal of Heart Failure*, 21(1):74–85, 2019.

[18] Mariana Nogueira, Mathieu De Craene, Sergio Sanchez-Martinez, Devyani Chowdhury, Bart Bijnens, and Gemma Piella. Analysis of nonstandardized

stress echocardiography sequences using multiview dimensionality reduction. *Medical Image Analysis*, 60:101594, 2020.

[19] Usha B. Tedrow and William G. Stevenson. Recording and interpreting unipolar electrograms to guide catheter ablation. *Heart Rhythm*, 8(5):791–796, May 2011.

[20] Jana M. Schwarzl, Ruben Schleberger, Ann-Kathrin Kahle, Alexandra Höller, Michael Schwarzl, Benjamin N. Schaeffer, Paula Münkler, Julia Moser, Ruken Ö. Akbulak, Christian Eickholt, Leon Dinshaw, Jannis Dickow, Philippe Maury, Frederic Sacher, Claire A. Martin, Tom Wong, Heidi L. Estner, Pierre Jaïs, Stephan Willems, and Christian Meyer. Specific electrogram characteristics impact substrate ablation target area in patients with scar-related ventricular tachycardia—insights from automated ultrahigh-density mapping. *Journal of Cardiovascular Electrophysiology*, 32(2):376–388, 2021.

[21] Juan Acosta, David Andreu, Diego Penela, Mario Cabrera, Alicia Carlosena, Viatcheslav Korshunov, Francesca Vassanelli, Roger Borras, Mikel Martínez, Juan Fernández-Armenta, Markus Linhart, José M Tolosana, Lluis Mont, and Antonio Berruezo. Elucidation of hidden slow conduction by double ventricular extrastimuli: a method for further arrhythmic substrate identification in ventricular tachycardia ablation procedures. *EP Europace*, 20(2):337–346, 12 2016.

[22] Katja Zeppenfeld and Andreu Porta-Sánchez. Automated Functional Substrate Mapping: Further Hurdles to Be Cleared. *JACC: Clinical Electrophysiology*, 6(14):1794–1796, 2020.

[23] David F. Briceño, Jorge Romero, Carola Gianni, Sanghamitra Mohanty, Pedro A. Villablanca, Andrea Natale, and Luigi Di Biase. Substrate Ablation of Ventricular Tachycardia: Late Potentials, Scar Dechanneling, Local Abnormal Ventricular Activities, Core Isolation, and Homogenization. *Cardiac Electrophysiology Clinics*, 9(1):81–91, Mar 2017.

[24] Andreu Porta-Sánchez, Nicholas Jackson, Peter Lukac, Steen Buus Kristiansen, Jan Moller Nielsen, Sigfus Gizurarson, Stéphane Massé, Christopher Labos, Karthik Viswanathan, Benjamin King, Andrew C.T. Ha, Eugene Downar, and Kumaraswamy Nanthakumar. Multicenter Study of Ischemic Ventricular Tachycardia Ablation With Decrement-Evoked Potential (DEEP) Mapping With Extra Stimulus. *JACC: Clinical Electrophysiology*, 4(3):307–315, 2018.

[25] Sachin Nayyar, Eugene Downar, Abhishek P. Bhaskaran, Stéphane Massé, and Kumaraswamy Nanthakumar. Signature signal strategy:

Electrogram-based ventricular tachycardia mapping. *Heart Rhythm*, 17(11):2000–2009, 2020.

[26] Neil T Srinivasan and Alex Cambridge. Ventricular tachycardia ablation. In *Decoding Cardiac Electrophysiology*, pages 99–107. Springer, 2020.

[27] Edmond M. Cronin, Frank M. Bogun, Philippe Maury, Petr Peichl, Minglong Chen, Narayanan Namboodiri, Luis Aguinaga, Luiz Roberto Leite, Sana M. Al-Khatib, Elad Anter, Antonio Berruezo, David J. Callans, Mina K. Chung, Phillip Cuculich, Andre d'Avila, Barbara J. Deal, Paolo Della Bella, Thomas Deneke, Timm-Michael Dickfeld, Claudio Hadid, Haris M. Haqqani, G. Neal Kay, Rakesh Latchamsetty, Francis Marchlinski, John M. Miller, Akihiko Nogami, Akash R. Patel, Rajeev Kumar Pathak, Luis C. Saenz Morales, Pasquale Santangeli, John L. Sapp Jr., Andrea Sarkozy, Kyoko Soejima, William G. Stevenson, Usha B. Tedrow, Wendy S. Tzou, Niraj Varma, and Katja Zeppenfeld. 2019 HRS/EHRA/APHRS/LAHRS expert consensus statement on catheter ablation of ventricular arrhythmias. *Heart Rhythm*, 17(1):e2–e154, Jan 2020.

[28] Abhishek Bhaskaran, John Fitzgerald, Nicholas Jackson, Sigfus Gizurarson, Kumaraswamy Nanthakumar, and Andreu Porta-Sánchez. Decrement Evoked Potential Mapping to Guide Ventricular Tachycardia Ablation: Elucidating the Functional Substrate. *Arrhythmia & electrophysiology review*, 9(4):211–8, 2020.

[29] Jason V. Garcia and Daniel K. Y. Wan. *The Kit: Access, Catheter Placement, Transeptal Puncture, Ablation Technology, 3D Mapping*, pages 21–39. Springer International Publishing, Cham, 2020.

[30] Geert Litjens, Francesco Ciompi, Jelmer M. Wolterink, Bob D. de Vos, Tim Leiner, Jonas Teuwen, and Ivana Išgum. State-of-the-Art Deep Learning in Cardiovascular Image Analysis. *JACC: Cardiovascular Imaging*, 12(8_Part_1):1549–1565, 2019.

[31] Borui Hou, Jianyong Yang, Pu Wang, and Ruqiang Yan. LSTM-Based Auto-Encoder Model for ECG Arrhythmias Classification. *IEEE Transactions on Instrumentation and Measurement*, 69(4):1232–1240, 2020.

[32] Saman Parvaneh, Jonathan Rubin, Saeed Babaeizadeh, and Minnan Xu-Wilson. Cardiac arrhythmia detection using deep learning: A review. *Journal of Electrocardiology*, 57:S70–S74, 2019.

[33] João Ribeiro Pinto, Jaime S. Cardoso, and André Lourenço. Evolution, Current Challenges, and Future Possibilities in ECG Biometrics. *IEEE Access*, 6:34746–34776, 2018.

[34] Aurore Lyon, Rina Ariga, Ana Mincholé, Masliza Mahmod, Elizabeth

Ormondroyd, Pablo Laguna, Nando de Freitas, Stefan Neubauer, Hugh Watkins, and Blanca Rodriguez. Distinct ECG Phenotypes Identified in Hypertrophic Cardiomyopathy Using Machine Learning Associate With Arrhythmic Risk Markers. *Frontiers in Physiology*, 9:213, 2018.

[35] J.P. Martinez, R. Almeida, S. Olmos, A.P. Rocha, and P. Laguna. A wavelet-based ECG delineator: Evaluation on standard databases. *IEEE Transactions on Biomedical Engineering*, 51(4):570–581, 2004.

[36] Pablo Laguna, Raimon Jané, and Pere Caminal. Automatic Detection of Wave Boundaries in Multilead ECG Signals: Validation with the CSE Database. *Computers and Biomedical Research*, 27(1):45–60, 1994.

[37] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[38] Lucie Maršánová, Marina Ronzhina, Radovan Smíšek, Martin Vítek, Andrea Němcová, Lukas Smital, and Marie Nováková. ECG features and methods for automatic classification of ventricular premature and ischemic heartbeats: A comprehensive experimental study. *Scientific Reports*, 7(1):11239, Sep 2017.

[39] Alejandro Alcaine, David Soto-Iglesias, David Andreu, Juan Fernandez-Armenta, Antonio Berruezo, Pablo Laguna, Oscar Camara, and Juan Pablo Martínez. Wavelet-based electrogram onset identification for ventricular electroanatomical mapping. In *Computing in Cardiology 2013*, pages 615–618, 2013.

[40] Alejandro Alcaine, David Soto-Iglesias, Mireia Calvo, Esther Guiu, David Andreu, Juan Fernández-Armenta, Antonio Berruezo, Pablo Laguna, Oscar Camara, and Juan Pablo Martínez. A Wavelet-Based Electrogram Onset Delineator for Automatic Ventricular Activation Mapping. *IEEE Transactions on Biomedical Engineering*, 61(12):2830–2839, 2014.

[41] Arturo Martínez, Raúl Alcaraz, and José Joaquín Rieta. Application of the phasor transform for automatic delineation of single-lead ECG fiducial points. *Physiological Measurement*, 31(11):1467–1485, sep 2010.

[42] Diego Osorio, Raúl Alcaraz, and José J Rieta. A fractionation-based local activation wave detector for atrial electrograms of atrial fibrillation. In *2017 Computing in Cardiology (CinC)*, pages 1–4, 2017.

[43] Jiapu Pan and Willis J. Tompkins. A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236,

1985.

[44] Saúl Palacios, Enrico G. Caiani, Federica Landreani, Juan Pablo Martínez, and Esther Pueyo. Long-Term Microgravity Exposure Increases ECG Repolarization Instability Manifested by Low-Frequency Oscillations of T-Wave Vector. *Frontiers in Physiology*, 10:1510, 2019.

[45] S. Banerjee, R. Gupta, and M. Mitra. Delineation of ECG characteristic features using multiresolution wavelet analysis method. *Measurement*, 45(3):474–487, 2012.

[46] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

[47] Aurore Lyon, Ana Mincholé, Juan Pablo Martínez, Pablo Laguna, and Blanca Rodriguez. Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. *Journal of The Royal Society Interface*, 15(138):20170821, 2018.

[48] Quazi Abidur Rahman, Larisa G. Tereshchenko, Matthew Kongkatong, Theodore Abraham, M. Roselle Abraham, and Hagit Shatkay. Identifying hypertrophic cardiomyopathy patients by classifying individual heartbeats from 12-lead ECG signals. In *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 224–229, 2014.

[49] Quazi Abidur Rahman, Larisa G. Tereshchenko, Matthew Kongkatong, Theodore Abraham, M. Roselle Abraham, and Hagit Shatkay. Utilizing ecg-based heartbeat classification for hypertrophic cardiomyopathy identification. *IEEE Transactions on Nanobioscience*, 14(5):505–512, July 2015. Publisher Copyright: © 2002-2011 IEEE.

[50] Max Falkenberg McGillivray, William Cheng, Nicholas S. Peters, and Kim Christensen. Machine learning methods for locating re-entrant drivers from electrograms in a model of atrial fibrillation. *Royal Society Open Science*, 5(4):172434, 2018.

[51] Albert K. Feeny, Mina K. Chung, Anant Madabhushi, Zachi I. Attia, Maja Cikes, Marjan Firouznia, Paul A. Friedman, Matthew M. Kalscheur, Suraj Kapa, Sanjiv M. Narayan, Peter A. Noseworthy, Rod S. Passman, Marco V. Perez, Nicholas S. Peters, Jonathan P. Piccini, Khaldoun G. Tarakji, Suma A. Thomas, Natalia A. Trayanova, Mintu P. Turakhia, and Paul J. Wang. Artificial Intelligence and Machine Learning in Arrhythmias and Cardiac Electrophysiology. *Circulation: Arrhythmia and Electrophysiology*, 13(8):e007952, 2020.

[52] Erick A Perez Alday, Annie Gu, Amit J Shah, Chad Robichaux, An-K-

wok Ian Wong, Chengyu Liu, Feifei Liu, Ali Bahrami Rad, Andoni Elola, Salman Seyedi, Qiao Li, Ashish Sharma, Gari D Clifford, and Matthew A Reyna. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. *Physiological Measurement*, 41(12):124003, jan 2021.

[53] Rémi Dubois, Pierre Maison-Blanche, Brigitte Quenet, and Gérard Dreyfus. Automatic ECG wave extraction in long-term recordings using Gaussian mesa function models and nonlinear probability estimators. *Comput. Methods Prog. Biomed.*, 88(3):217–233, December 2007.

[54] S. Graja and J.-M. Boucher. Hidden Markov tree model applied to ECG delineation. *IEEE Transactions on Instrumentation and Measurement*, 54(6):2163–2168, 2005.

[55] Filip Loncaric, Pablo-Miki Marti Castellote, Sergio Sanchez-Martinez, Dora Fabijanovic, Loredana Nunno, Maria Mimbrero, Laura Sanchis, Adelina Doltra, Silvia Montserrat, Maja Cikes, Fatima Crispi, Gema Piella, Marta Sitges, and Bart Bijnens. Automated Pattern Recognition in Whole–Cardiac Cycle Echocardiographic Data: Capturing Functional Phenotypes with Machine Learning. *Journal of the American Society of Echocardiography*, 2021.

[56] Ben Ng, Sachin Nayyar, and Vijay S. Chauhan. The Role of Artificial Intelligence and Machine Learning in Clinical Cardiac Electrophysiology. *Canadian Journal of Cardiology*, 2021.

[57] Mahmood Alhusseini, Firas Abuzaid, Paul Clopton, Albert Rogers, Miguel Rodrigo, Tina Baykaner, Paul Wang, Wouter-Jan Rappel, and Sanjiv Narayan. Machine Learning identifies sites where ablation terminates persistent atrial fibrillation. *Journal of the American College of Cardiology*, 73(9_Supplement_1):301–301, 2019.

[58] Jonathan J.C. Nicolet, Juan F. Restrepo, and Gastón Schlotthauer. Classification of intracavitary electrograms in atrial fibrillation using information and complexity measures. *Biomedical Signal Processing and Control*, 57:101753, 2020.

[59] Alexander M. Zolotarev, Brian J. Hansen, Ekaterina A. Ivanova, Katelynn M. Helfrich, Ning Li, Paul M.L. Janssen, Peter J. Mohler, Nahush A. Mokadam, Bryan A. Whitson, Maxim V. Fedorov, John D. Hummel, Dmitry V. Dylov, and Vadim V. Fedorov. Optical Mapping-Validated Machine Learning Improves Atrial Fibrillation Driver Detection by Multi-Electrode Mapping. *Circulation: Arrhythmia and Electrophysiology*, 13(10):e008249, 2020.

[60] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015.

[61] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[62] Herbert Robbins and Sutton Monro. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951.

[63] Léon Bottou, Frank E. Curtis, and Jorge Nocedal. Optimization Methods for Large-Scale Machine Learning. *SIAM Review*, 60(2):223–311, 2018.

[64] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015*, 2015.

[65] Matthew D. Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 818–833, Cham, 2014. Springer International Publishing.

[66] Awni Y. Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H. Tison, Codie Bourn, Mintu P. Turakhia, and Andrew Y. Ng. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1):65–69, Jan 2019.

[67] Jingshan Huang, Binqiang Chen, Bin Yao, and Wangpeng He. ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network. *IEEE Access*, 7:92871–92880, 2019.

[68] Eedara Prabhakararao and Samarendra Dandapat. Myocardial Infarction Severity Stages Classification From ECG Signals Using Attentional Recurrent Neural Network. *IEEE Sensors Journal*, 20(15):8711–8720, 2020.

[69] Zahra Ebrahimi, Mohammad Loni, Masoud Daneshtalab, and Arash Gharehbaghi. A review on deep learning methods for ECG arrhythmia classification. *Expert Systems with Applications: X*, 7:100033, 2020.

[70] Sarah W.E. Baalman, Florian E. Schroevers, Abel J. Oakley, Tom F. Brouwer, Willeke van der Stuijt, Hidde Bleijendaal, Lucas A. Ramos, Ricardo R. Lopes, Henk A. Marquering, Reinoud E. Knops, and Joris R. de Groot. A morphology based deep learning model for atrial fibrillation detection using single cycle electrocardiographic samples. *International Journal of Cardiology*, 316:130–136, 2020.

[71] Xiaoxi Yao, Rozalina G. McCoy, Paul A. Friedman, Nilay D. Shah, Barbara A. Barry, Emma M. Behnken, Jonathan W. Inselman, Zachi I. Attia, and Peter A. Noseworthy. ECG AI-Guided Screening for Low Ejection

Fraction (EAGLE): Rationale and design of a pragmatic cluster randomized trial. *American Heart Journal*, 219:31–36, 2020.

[72] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[73] Miguel Rodrigo, Albert J Rogers, Prasanth Ganesan, Mahmood Alhusseini, JUSTIN XU, and Sanjiv M Narayan. Abstract 14738: Classification of Atrial Fibrillation by Deep Learning of Electrogram Shapes versus Rate and Regularity. *Circulation*, 142(Suppl_3):A14738–A14738, 2020.

[74] Miguel Rodrigo, Albert Rogers, Prasanth Ganesan, Chayakrit Krittanawong, Mahmood Alhusseini, and Sanjiv Narayan. Classification of individual atrial intracardiac electrograms by Deep Learning. *Journal of the American College of Cardiology*, 77(18_Supplement_1):3217–3217, 2021.

[75] Chih-Min Liu, Shih-Lin Chang, Hung-Hsun Chen, Wei-Shiang Chen, Yenn-Jiang Lin, Li-Wei Lo, Yu-Feng Hu, Fa-Po Chung, Tze-Fan Chao, Ta-Chuan Tuan, Jo-Nan Liao, Chin-Yu Lin, Ting-Yung Chang, Cheng-I Wu, Ling Kuo, Mei-Han Wu, Chun-Ku Chen, Ying-Yueh Chang, Yang-Che Shiu, Henry Horng-Shing Lu, and Shih-Ann Chen. The Clinical Application of the Deep Learning Technique for Predicting Trigger Origins in Patients With Paroxysmal Atrial Fibrillation With Catheter Ablation. *Circulation: Arrhythmia and Electrophysiology*, 13(11):e008518, 2020.

[76] Shun Liao, Don Ragot, Sachin Nayyar, Adrian Suszko, Zhaolei Zhang, Bo Wang, and Vijay S. Chauhan. Deep Learning Classification of Unipolar Electrograms in Human Atrial Fibrillation: Application in Focal Source Mapping. *Frontiers in Physiology*, 12:1162, 2021.

[77] Julià Camps, Blanca Rodríguez, and Ana Mincholé. Deep Learning Based QRS Multilead Delineator in Electrocardiogram Signals. In *2018 Computing in Cardiology Conference (CinC)*, volume 45, pages 1–4, 2018.

[78] Hedayat Abrishami, Chia Han, Xuefu Zhou, Matthew Campbell, and Richard Czosek. Supervised ECG interval segmentation using LSTM neural network. In *BIOCOMP*, pages 71–77, 2018.

[79] Philipp Sodmann, Marcus Vollmer, Neetika Nath, and Lars Kaderali. A convolutional neural network for ECG annotation as the basis for classification of cardiac rhythms. *Physiological Measurement*, 39(10):104005, oct 2018.

[80] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors,

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pages 234–241, Cham, 2015. Springer International Publishing.

[81] Guillermo Jimenez-Perez, Alejandro Alcaine, and Oscar Camara. U-Net Architecture for the Automatic Detection and Delineation of the Electrocardiogram. In *2019 Computing in Cardiology (CinC)*, volume 46, pages Page 1–Page 4, 2019.

[82] Guillermo Jimenez-Perez, Alejandro Alcaine, and Oscar Camara. Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks. *Scientific Reports*, 11(1):863, Jan 2021.

[83] Geoffrey H. Tison, Jeffrey Zhang, Francesca N. Delling, and Rahul C. Deo. Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery. *Circulation: Cardiovascular Quality and Outcomes*, 12(9):e005289, 2019.

[84] Abdolrahman Peimankar and Sadasivan Puthusserypady. DENS-ECG: A deep learning approach for ECG signal delineation. *Expert Systems with Applications*, 165:113911, 2021.

[85] Viktor Moskalenko, Nikolai Zolotykh, and Grigory Osipov. Deep learning for ECG segmentation. In Boris Kryzhanovsky, Witali Dunin-Barkowski, Vladimir Redko, and Yury Tiumentsev, editors, *Advances in Neural Computation, Machine Learning, and Cognitive Research III*, pages 246–254, Cham, 2020. Springer International Publishing.

[86] Fatma Murat, Ozal Yildirim, Muhammed Talo, Ulas Baran Baloglu, Yakup Demir, and U. Rajendra Acharya. Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review. *Computers in Biology and Medicine*, 120:103726, 2020.

[87] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[88] Sulaiman Somani, Adam J Russak, Felix Richter, Shan Zhao, Akhil Vaid, Fayzan Chaudhry, Jessica K De Freitas, Nidhi Naik, Riccardo Miotto, Girish N Nadkarni, Jagat Narula, Edgar Argulian, and Benjamin S Glicksberg. Deep learning and the electrocardiogram: review of the current state-of-the-art. *EP Europace*, 02 2021. euaa377.

[89] P. Laguna, R.G. Mark, A. Goldberg, and G.B. Moody. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In *Computers in Cardiology 1997*, pages 673–676, 1997.

[90] Alena I. Kalyakulina, Igor I. Yusipov, Viktor A. Moskalenko, Alexander V.

Nikolskiy, Konstantin A. Kosonogov, Grigory V. Osipov, Nikolai Yu. Zolotykh, and Mikhail V. Ivanchenko. LUDB: A New Open-Access Validation Tool for Electrocardiogram Delineation Algorithms. *IEEE Access*, 8:186181–186190, 2020.

[91] Antonio Luiz P. Ribeiro, Gabriela M.M. Paixão, Paulo R. Gomes, Manoel Horta Ribeiro, Antônio H. Ribeiro, Jéssica A. Canazart, Derick M. Oliveira, Milton P. Ferreira, Emilly M. Lima, Jermana Lopes de Moraes, Nathalia Castro, Leonardo B. Ribeiro, and Peter W. Macfarlane. Tele-electrocardiography and bigdata: The CODE (Clinical Outcomes in Digital Electrocardiography) study. *Journal of Electrocardiology*, 57:S75–S78, 2019.

[92] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 843–852, 2017.

[93] Antônio H. Ribeiro, Manoel Horta Ribeiro, Gabriela M. M. Paixão, Derick M. Oliveira, Paulo R. Gomes, Jéssica A. Canazart, Milton P. S. Ferreira, Carl R. Andersson, Peter W. Macfarlane, Wagner Meira Jr., Thomas B. Schön, and Antonio Luiz P. Ribeiro. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature Communications*, 11(1):1760, Apr 2020.

[94] B Pyakillya, N Kazachenko, and N Mikhailovsky. Deep Learning for ECG Classification. *Journal of Physics: Conference Series*, 913:012004, oct 2017.

[95] Thomas Grote and Philipp Berens. On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, 46(3):205–211, 2020.

[96] Ruben Doste, Rafael Sebastian, Juan Francisco Gomez, David Soto-Iglesias, Alejandro Alcaine, Lluis Mont, Antonio Berruezo, Diego Penela, and Oscar Camara. In silico pace-mapping: prediction of left vs. right outflow tract origin in idiopathic ventricular arrhythmias with patient-specific electrophysiological simulations. *EP Europace*, 22(9):1419–1430, 07 2020.

[97] Adam Timmis, Nick Townsend, Chris P Gale, Aleksandra Torbica, Maddalena Lettino, Steffen E Petersen, Elias A Mossialos, Aldo P Maggioni, Dzianis Kazakiewicz, Heidi T May, Delphine De Smedt, Marcus Flather, Liesl Zuhlke, John F Beltrame, Radu Huculeci, Luigi Tavazzi, Gerhard Hindricks, Jeroen Bax, Barbara Casadei, Stephan Achenbach, Lucy Wright, Panos Vardas, and European Society of Cardiology. European Society of Cardiology: Cardiovascular Disease Statistics 2019. *European Heart Jour-*

nal, 41(1):12–85, 12 2019.

[98] Holger Holst, Mattias Ohlsson, Carsten Peterson, and Lars Edenbrandt. A confident decision support system for interpreting electrocardiograms. *Clinical Physiology*, 19(5):410–418, 1999.

[99] Gerhard Hindricks, Tatjana Potpara, Nikolaos Dagres, Elena Arbelo, Jeroen J Bax, Carina Blomström-Lundqvist, Giuseppe Boriani, Manuel Castella, Gheorghe-Andrei Dan, Polychronis E Dilaveris, Laurent Fauchier, Gerasimos Filippatos, Jonathan M Kalman, Mark La Meir, Deirdre A Lane, Jean-Pierre Lebeau, Maddalena Lettino, Gregory Y H Lip, Fausto J Pinto, G Neil Thomas, Marco Valgimigli, Isabelle C Van Gelder, Bart P Van Putte, Caroline L Watkins, and ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *European Heart Journal*, 42(5):373–498, 08 2020.

[100] Josep Brugada, Demosthenes G Katritsis, Elena Arbelo, Fernando Arribas, Jeroen J Bax, Carina Blomström-Lundqvist, Hugh Calkins, Domenico Corrado, Spyridon G Deftereos, Gerhard-Paul Diller, Juan J Gomez-Doblas, Bulent Gorenek, Andrew Grace, Siew Yen Ho, Juan-Carlos Kaski, Karl-Heinz Kuck, Pier David Lambiase, Frederic Sacher, Georgia Sarquella-Brugada, Piotr Suwalski, Antonio Zaza, and ESC Scientific Document Group. 2019 ESC Guidelines for the management of patients with supraventricular tachycardiaThe Task Force for the management of patients with supraventricular tachycardia of the European Society of Cardiology (ESC): Developed in collaboration with the Association for European Paediatric and Congenital Cardiology (AEPC). *European Heart Journal*, 41(5):655–720, 08 2019.

[101] Edmond M. Cronin, Frank M. Bogun, Philippe Maury, Petr Peichl, Minglong Chen, Narayanan Namboodiri, Luis Aguinaga, Luiz Roberto Leite, Sana M. Al-Khatib, Elad Anter, Antonio Berruezo, David J. Callans, Mina K. Chung, Phillip Cuculich, Andre d'Avila, Barbara J. Deal, Paolo Della Bella, Thomas Deneke, Timm-Michael Dickfeld, Claudio Hadid, Haris M. Haqqani, G. Neal Kay, Rakesh Latchamsetty, Francis Marchlinski, John M. Miller, Akihiko Nogami, Akash R. Patel, Rajeev Kumar Pathak, Luis C. Saenz Morales, Pasquale Santangeli, John L. Sapp Jr., Andrea Sarkozy, Kyoko Soejima, William G. Stevenson, Usha B. Tedrow, Wendy S. Tzou, Niraj Varma, and Katja Zeppenfeld. 2019

HRS/EHRA/APHRS/LAHRS expert consensus statement on catheter ablation of ventricular arrhythmias. *Heart Rhythm*, 17(1):e2–e154, Jan 2020.

[102] Oliver Faust, Yuki Hagiwara, Tan Jen Hong, Oh Shu Lih, and U Rajendra Acharya. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine*, 161:1–13, 2018.

[103] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning, 2017.

[104] Francois Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[105] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[106] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

[107] Jonathan Tompson, Ross Goroshin, Arjun Jain, Yann LeCun, and Christoph Bregler. Efficient object localization using convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[108] Sergio Sánchez-Martínez, Oscar Camara, Gemma Piella, Maja Cikes, Miguel Angel Gonzalez Ballester, Marius Miron, Alfredo Vellido, Emilia Gomez, Alan Fraser, and Bart Bijnens. Machine Learning for Clinical Decision-Making: Challenges and Opportunities. *Preprints*, 2019.

[109] Jonathan H. Chen and Steven M. Asch. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *New England Journal of Medicine*, 376(26):2507–2509, 2017. PMID: 28657867.

[110] Jan Kukačka, Vladimir Golkov, and Daniel Cremers. Regularization for Deep Learning: A Taxonomy, 2017.

[111] Veronika Cheplygina, Marleen de Bruijne, and Josien P.W. Pluim. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical Image Analysis*, 54:280–296, 2019.

[112] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 11 1997.

[113] Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *3rd International*

Conference on Learning Representations, ICLR 2015, jan 2015.

[114] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[115] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. MobileNets: Efficient convolutional neural networks for mobile vision applications, 2017.

[116] Yu Cheng, Duo Wang, Pan Zhou, and Tao Zhang. A Survey of Model Compression and Acceleration for Deep Neural Networks, 2020.

[117] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, and Rory Collins. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, 12(3):1–10, 03 2015.

[118] Kenneth T Macleod. *An essential introduction to cardiac electrophysiology*. World Scientific Publishing Company, 2013.

[119] David Andreu, Juan Fernández-Armenta, Juan Acosta, Diego Penela, Beatriz Jáuregui, David Soto-Iglesias, Vladimir Syrovnev, Elena Arbelo, José María Tolosana, and Antonio Berruezo. A QRS axis-based algorithm to identify the origin of scar-related ventricular tachycardia in the 17-segment American Heart Association model. *Heart Rhythm*, 15(10):1491–1497, Oct 2018.

[120] Genlang Chen, Maolin Chen, Jiajian Zhang, Liang Zhang, and Chaoyi Pang. A crucial wave detection and delineation method for twelve-lead ECG signals. *IEEE Access*, 8:10707–10717, 2020.

[121] Iana Sereda, Sergey Alekseev, Aleksandra Koneva, Roman Kataev, and Grigory V. Osipov. ECG segmentation by Neural Networks: Errors and Correction. In *International Joint Conference on Neural Networks, IJCNN 2019 Budapest, Hungary, July 14-19, 2019*, pages 1–7. IEEE, 2019.

[122] Byungju Kim, Hyunwoo Kim, Kyungsu Kim, Sungjin Kim, and Junmo Kim. Learning not to learn: Training deep neural networks with biased data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[123] Kristine Aavild Juhl, Rasmus Reinhold Paulsen, Anders Bjorholm Dahl,

Vedrana Andersen Dahl, Ole de Backer, Klaus Fuglsang Kofoed, and Oscar Camara. Guiding 3D U-nets with signed distance fields for creating 3D models from images. In *Medical Imaging with Deep Learning, (MIDL) 2019*, pages 1–4, 2019.

[124] Bowen Cheng, Ross Girshick, Piotr Dollar, Alexander C. Berg, and Alexander Kirillov. Boundary IoU: Improving Object-Centric Image Segmentation Evaluation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15334–15342, June 2021.

[125] Xide Xia and Brian Kulis. W-net: A deep model for fully unsupervised image segmentation, 2017.

[126] Lu Xu, Mingyuan Liu, Zhenrong Shen, Hua Wang, Xiaowei Liu, Xin Wang, Siyu Wang, Tiefeng Li, Shaomei Yu, Min Hou, Jianhua Guo, Jicong Zhang, and Yihua He. DW-Net: A cascaded convolutional neural network for apical four-chamber view segmentation in fetal echocardiography. *Computerized Medical Imaging and Graphics*, 80:101690, 2020.

[127] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[128] Jianwei Zheng, Guohua Fu, Kyle Anderson, Huimin Chu, and Cyril Rakovski. A 12-Lead ECG database to identify origins of idiopathic ventricular arrhythmia containing 334 patients. *Scientific Data*, 7(1):98, Mar 2020.

[129] Sergey Ioffe and Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, page 448–456. JMLR.org, 2015.

[130] Lu Xu, Mingyuan Liu, Jicong Zhang, and Yihua He. Convolutional-Neural-Network-Based Approach for Segmentation of Apical Four-Chamber View from Fetal Echocardiography. *IEEE Access*, 8:80437–80446, 2020.

[131] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran

Associates, Inc., 2019.

[132] Nima Tajbakhsh, Laura Jeyaseelan, Qian Li, Jeffrey N. Chiang, Zhihao Wu, and Xiaowei Ding. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, 63:101693, 2020.

[133] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176):1203–1205, 2014.

[134] Hieu Pham, Zihang Dai, Qizhe Xie, and Quoc V. Le. Meta Pseudo Labels, June 2021.

[135] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets, 2014.

[136] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.

[137] Toby R. Engel. Diagnosis of Hypertrophic Cardiomyopathy: Who Is in Charge Here—The Physician or the Computer? *Journal of the American College of Cardiology*, 75(7):734–735, 2020.

[138] Juan Acosta, David Soto-Iglesias, Beatriz Jáuregui, Juan Fernández Armenta, Diego Penela, Manuel Frutos-López, Eduardo Arana-Rueda, Alonso Pedrote, Lluís Mont, and Antonio Berruezo. Long-term outcomes of ventricular tachycardia substrate ablation incorporating hidden slow conduction analysis. *Heart Rhythm*, 17(10):1696–1703, 2020. Focus Issue: Sudden Death.

[139] J Felix, R Alcaraz, and JJ Rieta. Adaptive wavelets applied to automatic local activation wave detection in fractionated atrial electrograms of atrial fibrillation. In *2015 Computing in Cardiology Conference (CinC)*, pages 45–48, 2015.

[140] J. Hashemi, M. H. Shariat, and D. Redfearn. Intracardiac electrogram envelope detection during atrial fibrillation using fast orthogonal search. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3162–3165, 2017.

[141] Philip Hoyland, Néfissa Hammache, Alberto Battaglia, Julien Oster, Jacques Felblinger, Christian de Chillou, and Freddy Odille. A Paced-ECG Detector and Delineator for Automatic Multi-Parametric Catheter Mapping of Ventricular Tachycardia. *IEEE Access*, 8:223952–223960, 2020.

[142] Bokeh Development Team. *Bokeh: Python library for interactive visual-*

ization, 2018.

[143] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond Empirical Risk Minimization. In *6th International Conference on Learning Representations, ICLR 2018*, 2018.

[144] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[145] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, Feb 2021.

[146] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 833–851, Cham, 2018. Springer International Publishing.

[147] Yuanyi Zeng, Xiaoyu Chen, Yi Zhang, Lianfa Bai, and Jing Han. Dense-U-Net: densely connected convolutional network for semantic segmentation with a small number of samples. In Chunming Li, Hui Yu, Zhigeng Pan, and Yifei Pu, editors, *10th International Conference on Graphics and Image Processing, ICGIP 2018*, volume 11069, pages 665 – 670. International Society for Optics and Photonics, SPIE, 2019.

[148] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571, 2016.

[149] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical Evaluation of Rectified Activations in Convolutional Network. *Computing Research Repository (CoRR)*, abs/1505.00853, 2015.

[150] L. Faes, G. Nollo, R. Antolini, F. Gaita, and F. Ravelli. A method for quantifying atrial fibrillation organization based on wave-morphology similarity. *IEEE Transactions on Biomedical Engineering*, 49(12):1504–1513, 2002.

[151] Habib Hajimolahoseini, Javad Hashemi, Saeed Gazor, and Damian Redfearn. Inflection point analysis: A machine learning approach for extraction of IEGM active intervals during atrial fibrillation. *Artificial Intelligence in Medicine*, 85:7–15, 2018.

[152] Wei-Yin Ko, Konstantinos C. Siontis, Zachi I. Attia, Rickey E. Carter, Suraj Kapa, Steve R. Ommen, Steven J. Demuth, Michael J. Ackerman, Bernard J. Gersh, Adelaide M. Arruda-Olson, Jeffrey B. Geske, Samuel J. Asirvatham, Francisco Lopez-Jimenez, Rick A. Nishimura, Paul A. Friedman, and Peter A. Noseworthy. Detection of Hypertrophic Cardiomyopathy Using a Convolutional Neural Network-Enabled Electrocardiogram. *Journal of the American College of Cardiology*, 75(7):722–733, 2020.

[153] Authors/Task Force members, Perry M. Elliott, Aris Anastasakis, Michael A. Borger, Martin Borggrefe, Franco Cecchi, Philippe Charron, Albert Alain Hagege, Antoine Lafont, Giuseppe Limongelli, Heiko Mahrholdt, William J. McKenna, Jens Mogensen, Petros Nihoyannopoulos, Stefano Nistri, Petronella G. Pieper, Burkert Pieske, Claudio Rapezzi, Frans H. Rutten, Christoph Tillmanns, Hugh Watkins, Additional Contributor, Constantinos O'Mahony, ESC Committee for Practice Guidelines (CPG), Jose Luis Zamorano, Stephan Achenbach, Helmut Baumgartner, Jeroen J. Bax, Héctor Bueno, Veronica Dean, Christi Deaton, Çetin Erol, Robert Fagard, Roberto Ferrari, David Hasdai, Arno W. Hoes, Paulus Kirchhof, Juhani Knuuti, Philippe Kolh, Patrizio Lancellotti, Ales Linhart, Petros Nihoyannopoulos, Massimo F. Piepoli, Piotr Ponikowski, Per Anton Sirnes, Juan Luis Tamargo, Michal Tendera, Adam Torbicki, William Wijns, Stephan Windecker, Document Reviewers, David Hasdai, Piotr Ponikowski, Stephan Achenbach, Fernando Alfonso, Cristina Basso, Nuno Miguel Cardim, Juan Ramón Gimeno, Stephane Heymans, Per Johan Holm, Andre Keren, Paulus Kirchhof, Philippe Kolh, Christos Lionis, Claudio Muneretto, Silvia Priori, Maria Jesus Salvador, Christian Wolpert, Jose Luis Zamorano, Matthias Frick, Farid Aliyev, Svetlana Komissarova, Georges Mairesse, Elnur Smajić, Vasil Velchev, Loizos Antoniades, Ales Linhart, Henning Bundgaard, Tiina Heliö, Antoine Leenhardt, Hugo A. Katus, George Efthymiadis, Róbert Sepp, Gunnar Thor Gunnarsson, Shemy Carasso, Alina Kerimkulova, Ginta Kamzola, Hady Skouri, Ghada Eldirsi, Ausra Kavoliuniene, Tiziana Felice, Michelle Michels, Kristina Hermann Haugaa, Radosław Lenarczyk, Dulce Brito, Eduard Apetrei, Leo Bokheria, Dragan Lovic, Robert Hatala, Pablo Garcia Pavía, Maria Eriksson, Stéphane Noble, Elizabeta Srbinovska, Murat Özdemir, Elena Nesukay, and Neha Sekhri. 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: The Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). *European Heart Journal*, 35(39):2733–2779, 08 2014.

[154] Steve R. Ommen, Seema Mital, Michael A. Burke, Sharlene M. Day, Anita Deswal, Perry Elliott, Lauren L. Evanovich, Judy Hung, José A. Joglar, Paul Kantor, Carey Kimmelstiel, Michelle Kittleson, Mark S. Link, Martin S. Maron, Matthew W. Martinez, Christina Y. Miyake, Hartzell V. Schaff, Christopher Semsarian, and Paul Sorajja. 2020 AHA/ACC Guideline for the Diagnosis and Treatment of Patients With Hypertrophic Cardiomyopathy. *Circulation*, 142(25):e558–e631, 2020.

[155] Jodie Ingles, Jennifer Goldstein, Courtney Thaxton, Colleen Caleshu, Edward W. Corty, Stephanie B. Crowley, Kristen Dougherty, Steven M. Harrison, Jennifer McGlaughon, Laura V. Milko, Ana Morales, Bryce A. Seifert, Natasha Strande, Kate Thomson, J. Peter van Tintelen, Kathleen Wallace, Roddy Walsh, Quinn Wells, Nicola Whiffin, Leora Witkowski, Christopher Semsarian, James S. Ware, Ray E. Hershberger, and Birgit Funke. Evaluating the Clinical Validity of Hypertrophic Cardiomyopathy Genes. *Circulation: Genomic and Precision Medicine*, 12(2):e002460, 2019.

[156] Martin S. Maron, Ethan J. Rowin, and Barry J. Maron. The ESC Risk Score Is Less Reliable than ACC/AHA Risk Factors in Hypertrophic Cardiomyopathy: When Sensitivity Trumps Specificity. *Canadian Journal of Cardiology*, 35(12):1626–1628, Dec 2019.

[157] Constantinos O'Mahony, Fatima Jichi, Menelaos Pavlou, Lorenzo Monserrat, Aristides Anastasakis, Claudio Rapezzi, Elena Biagini, Juan Ramon Gimeno, Giuseppe Limongelli, William J. McKenna, Rumana Z. Omar, Perry M. Elliott, and for the Hypertrophic Cardiomyopathy Outcomes Investigators. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM Risk-SCD). *European Heart Journal*, 35(30):2010–2020, 10 2013.

[158] Anna Wålinder Österberg, Ingegerd Östman-Smith, Robert Jablonowski, Marcus Carlsson, Henrik Green, Cecilia Gunnarsson, Petru Liuba, and Eva Fernlund. High ECG Risk-Scores Predict Late Gadolinium Enhancement on Magnetic Resonance Imaging in HCM in the Young. *Pediatric Cardiology*, 42(3):492–500, Mar 2021.

[159] Neal K. Lakdawala, Jens Jakob Thune, Barry J. Maron, Allison L. Cirino, Ole Havndrup, Henning Bundgaard, Michael Christiansen, Christian M. Carlsen, Jean-François Dorval, Raymond Y. Kwong, Steven D. Colan, Lars V. Køber, and Carolyn Y. Ho. Electrocardiographic Features of Sarcomere Mutation Carriers With and Without Clinically Overt Hypertrophic Cardiomyopathy. *American Journal of Cardiology*, 108(11):1606–1613, Dec 2011.

[160] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-jiang Zhang, Qiang Yang,

and Stephen Lin. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007.

[161] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201, 2001.

[162] Amit Bermanis, Amir Averbuch, and Ronald R. Coifman. Multiscale data sampling and function extension. *Applied and Computational Harmonic Analysis*, 34(1):15–29, 2013.

[163] Nicolas Duchateau, Mathieu De Craene, Marta Sitges, and Vicent Caselles. Adaptation of multiscale function extension to inexact matching. Application to the mapping of individuals to a learnt manifold. In *SEE International Conference on Geometric Science of Information (GSI 2013)*, volume 8085 of *SEE International Conference on Geometric Science of Information (GSI)*, pages 578–586, Paris, France, 2013.

[164] J. Macqueen. Some methods for classification and analysis of multivariate observations. In *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[165] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

[166] Maurice Sokolow and Thomas P. Lyon. The ventricular complex in left ventricular hypertrophy as obtained by unipolar precordial and limb leads. *American Heart Journal*, 37(2):161–186, 1949.

[167] Christopher M. Kramer, Evan Appelbaum, Milind Y. Desai, Patrice Desvigne-Nickens, John P. DiMarco, Matthias G. Friedrich, Nancy Geller, Sarahfaye Heckler, Carolyn Y. Ho, Michael Jerosch-Herold, Elizabeth A. Ivey, Julianna Keleti, Dong-Yun Kim, Paul Kolm, Raymond Y. Kwong, Martin S. Maron, Jeanette Schulz-Menger, Stefan Piechnik, Hugh Watkins, William S. Weintraub, Pan Wu, and Stefan Neubauer. Hypertrophic Cardiomyopathy Registry: The rationale and design of an international, observational study of hypertrophic cardiomyopathy. *American Heart Journal*, 170(2):223–230, 2015.

[168] European Commission. Proposal For A Regulation Of The European Parliament And Of The Council On European Data Governance (Data Governance Act). Directive COM/2020/767 final, 2020.

[169] Shekoofeh Azizi, Basil Mustafa, Fiona Ryan, Zachary Beaver, Jan Freyberg, Jonathan Deaton, Aaron Loh, Alan Karthikesalingam, Simon Korn-

blith, Ting Chen, Vivek Natarajan, and Mohammad Norouzi. Big Self-Supervised Models Advance Medical Image Classification, 2021.

Curriculum Vitae

Guillermo Jiménez Pérez received his B.Sc. degree in Biomedical Engineering from the Universidad de Sevilla in 2015, and his M.Sc. degree in Logic, Computation and Artificial Intelligence from the same institution in 2016. He joined the PhD program at the Universitat Pompeu Fabra under the supervision of Prof. Oscar Camara. His work is focused on the processing of electrocardiographic data with artificial intelligence algorithms, with a focus on automatic data quantification.

Publications

Journal papers

1. **Jimenez-Perez, G.**, Alcaine, A., & Camara, O. (2021). Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks. Scientific reports, 11(1), 1-11.

2. Lampreave, P., **Jimenez-Perez, G.**, Sanz, I., Gomez, A., & Camara, O. (2020). Towards assisted electrocardiogram interpretation using an AI- enabled Augmented Reality headset. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 1-8.

3. **Jimenez-Perez, G.**, Alcaine, A., & Camara, O. Generalizing ECG Delineation: Training Convolutional Neural Networks with Synthetic Data Generation (submitted).

4. **Jimenez-Perez, G.***, Jáuregui, B.*, Alcaine, A. Soto-Iglesias, D., Berruezo, A. & Camara, O. Machine-Learning-based Site of Origin Identification of Ventricular Tachycardias from Electrocardiographic Recordings (in preparation).

5. **Jimenez-Perez, G.**, Loncaric, F., Marti Castellote, P., Ramos Jovani, M., Gonzales Lopez, E., Gonzales Mirelis, J., Camara, O., Piella, G., Garcia-Alvarez, A., Garcia Pavia, P., Sitges, M. & Bijnens, B. Deep learning on

electrocardiograms to phenotype hypertrophic cardiomyopathy – automated multi-beat, whole-cardiac cycle, ECG morphology analysis and ECG-based clustering validated with echocardiography, cardiac magnetic resonance and genotyping (in preparation).

6. Aviles, J., Mejía Córdova, M., Garrido-Oliver, J., Dux-Santoy, L., Morales Perez, X., **Jimenez-Perez, G.**, Ruiz-Munoz, A., Teixido-Tura, G., Maso Talou, G.D., Romero, D., Ferdian, E., Gilbert, K., Elsayed, A., Young, A.A., Evangelista, A., Ferreira, I., Rodriguez-Palomares, J., Camara, O., Guala, A. Machine learning for automatic assessment of aortic rotational flow and wall shear stress from 4D flow cardiac magnetic resonance imaging (submitted, European Heart Journal: Cardiovascular Imaging).

Peer-reviewed conference papers

1. **Jimenez-Perez, G.**, Alcaine, A., & Camara, O. (2019, September). U-Net architecture for the automatic detection and delineation of the electrocardiogram. In 2019 Computing in Cardiology (CinC) (pp. Page-1). IEEE.

2. Morales, X., Mill, J., Juhl, K. A., Olivares, A., **Jimenez-Perez, G.**, Paulsen, R. R., & Camara, O. (2019, October). Deep learning surrogate of computational fluid dynamics for thrombus formation risk in the left atrial appendage. In International Workshop on Statistical Atlases and Computational Models of the Heart (pp. 157-166). Springer, Cham.

3. Doste, R., Lozano, M., **Jimenez-Perez, G.**, Mont, L., Penela, D., Camara, O. & Sebastian, R. Machine Learning Models Trained with Simulations Predict the Site of Origin of Outflow Tract Ventricular Arrhythmias from Multi-centric Databases. In 2021 Computing in Cardiology (CinC) (pp. Page-1). IEEE.

4. Isabel, A., **Jimenez-Perez, G.** & Camara, O. (2021, September). Mobile app for the digitisation and deep-learning-based classification of electrocardiogram printed records. In 2021 Computing in Cardiology (CinC) (pp. Page-1). IEEE.

5. Gassa, N., Sacristan, B., Zemzemi, N., Laborde, M., Garrido Oliver, J., Matencio Perabla, C., **Jimenez-Perez, G.**, Camara, O., Ploux S., Strik M., Bordachar, P. & Dubois, R. (2021, September). Benchmark of Deep Learning algorithms for the automatic screening in electrocardiograms transmitted by

implantable cardiac devices. In 2021 Computing in Cardiology (CinC) (pp. Page-1). IEEE.

6. Mejia Cordova, M., Guala, A., Morales, X., **Jimenez-Perez, G.**, Dux-Santoy, L., Ruiz-Munoz, A., ... & Camara, O. (2021). Reinforcement machine learning-based aortic anatomical landmarks detection from phase-contrast enhanced magnetic resonance angiography. European Heart Journal - Cardiovascular Imaging, 22(Supplement_1), jeaa356-286.

7. Guala, A., Mejia Cordova, M., Morales, X., **Jimenez-Perez, G.**, Dux-Santoy, L., Aviles, J., ... & Rodriguez-Palomares, J. (2021). Machine learning to automatically detect anatomical landmarks on phase-contrast enhanced magnetic resonance angiography. European Heart Journal - Cardiovascular Imaging, 22(Supplement_2), jeab090-122.

Conference abstracts

1. **Jimenez-Perez, G.**, Alcaine, A. Soto-Iglesias, D., Berruezo, A. & Camara, O. (2018). Localización automática del lugar de origen de las taquicardias ventriculares por cicatriz miocárdica a partir del análisis de electrocardiogramas de 12 derivaciones. Revista Española de Cardiología. 71(Supl 1):399.

2. **Jimenez-Perez, G.**, Loncaric, F., Martí Castellote, P., Gonzales, E., Gonzales Mirelis, J., Piella, G., Garcia Pavia, P, Sitges, M, Camara, O. & Bijnens, B. Deep learning on electrocardiograms to phenotype hypertrophic cardiomyopathy – automated multi-beat, whole-cardiac cycle, ECG morphology analysis and ECG-based clustering validated with echocardiography, cardiac magnetic resonance and genotyping. European Heart Journal - Cardiovascular Imaging (accepted; EuroEcho).