TESIS DOCTORAL

Data Analytics for Supporting Clinical Decision on Patients with Implantable Cardioverter Defibrillator

Autor:
José María Lillo Castellano

Directores:
Dra. Inmaculada Mora Jiménez
Dr. José Luis Rojo Álvarez

Programa de Doctorado Interuniversitario en Multimedia y Comunicaciones

Escuela Internacional de Doctorado

2017

To my father. . .

Contents

Resumen

Antecedentes

Todo trastorno de la actividad eléctrica del corazón que induce una frecuencia cardiaca anormalmente rápida, lenta o errática es definido como *arritmia cardiaca*. En la actualidad, las arritmias son la enfermedad cardiovascular más común. Según la región anatómica donde éstas se originan, las arritmias se categorizan en supraventriculares (aurículas) o ventriculares (ventrículos). Por norma general, las arritmias supraventriculares son trastornos no graves. Por el contrario, ventriculares son potencialmente dañinas ya que, si no se controlan en un espacio breve de tiempo, a menudo derivan en la muerte súbita del paciente. Actualmente, los pacientes con alto riesgo de sufrir este tipo de arritmias son tratados con un Desfibrilador Automático Implantable (DAI). Un DAI es un pequeño dispositivo electrónico, de memoria y batería limitadas, que se incrusta bajo la piel y que, mediante unos pequeños cables posicionados directamente sobre la superficie del corazón, es capaz de detectar episodios arrítmicos, registrarlos y, si es necesario, aplicar terapias eléctricas para revertir un hipotético episodio arrítmico.

De un tiempo a esta parte, el apresurado desarrollo de las tecnologías de computación en la *nube* ha favorecido la aparición de nuevas plataformas para el almacenamiento masivo de episodios arrítmicos registrados por DAIs. Un claro ejemplo de ello es el sistema integrado en el proyecto *SCOOP*, un proyecto colaborativo lanzado en 2011 por la empresa Medtronic® para dar soporte a un repositorio masivo multicéntrico de registros de DAIs. La gran variabilidad, volumen y heterogeneidad de los datos incluidos en este repositorio está abriendo un amplio abanico de posibilidades para desarrollar nuevas herramientas que presten soporte en la práctica clínica.

Este nuevo entorno digital, el cual no tiene precedentes en el ámbito de la cardiología, es el que ha motivado la realización de esta Tesis Doctoral. De forma específica, la Tesis se ha enfocado hacia la solución de dos retos específicos: (1) la clasificación automática de episodios arrítmicos registrados por los DAIs; (2) la determinación de un umbral de seguridad para DAIs que asegure un bajo riesgo de infrasensado (no detección) de episodios arrítmicos fatales tales como Fibrilación Ventricular (FV).

Respecto al primer reto, aunque actualmente los DAIs son dispositivos altamente sensibles y fiables detectando episodios arrítmicos fatales, presentan una baja especificidad detectando y clasificando otro tipo de episodios. Como consecuencia, los cardiólogos tienen que analizar cada episodio detectado, clasificarlo y, además, chequear si tanto la detección como la posible terapia fueron adecuadas. Por tanto, la clasificación automática de episodios arrítmicos continúa siendo un campo de investigación abierto para la mejora. Es por ello que el diseño de nuevos enfoques para la clasificación, que se sustenten sobre un repositorio de datos masivos de alta calidad como es el incluido en SCOOP, podría ayudar en el desarrollo de herramientas que prestaran soporte a cardiólogos en sus decisiones clínicas.

En relación al segundo reto, la detección correcta de las ondas-R ha sido siempre un punto crítico en los DAIs. El motivo de esto deriva de que los algoritmos de detección de episodios arrítmicos en DAIs están principalmente basados en la frecuencia cardíaca instantánea, la cual es estimada a partir de la diferencia de dos ondas-R consecutivas detectadas. En la actualidad, la amplitud media (voltaje) de las ondas-R durante un Ritmo Basal (RB) es el mejor indicador para asegurar un bajo riesgo de infrasensado de episodios arrítmicos fatales tales como FV. Las guías clínicas recomiendan que, al menos durante el implante del DAI, la amplitud media de las ondas-R sea de 7 mV. Sin embargo, un valor mínimo, es decir un umbral de seguridad, aún no ha sido establecido. Como resultado, cuando la amplitud media de la onda-R de un paciente decrece hasta valores inferiores a 5 mV, los cardiólogos suelen chequear el funcionamiento del DAI mediante inducciones controladas de fibrilación en el paciente; inducciones que incrementan las complicaciones. En este sentido, la determinación de un umbral de seguridad sobre las ondas-R, obtenido a partir de los episodios de un repositorio como el de SCOOP, podría ser altamente beneficioso para definir nuevas recomendaciones de guías clínicas orientadas hacia un retardo o incluso una suspensión de procesos clínicos adversos como las inducciones de desfibrilación.

Objetivos

Esta Tesis Doctoral tiene como objetivo principal proporcionar una solución computacional a dos retos:

O1. La clasificación automática de episodios arrítmicos registrados por los DAIs.

O2. La determinación de un umbral de seguridad sobre las ondas-R durante un RB que garantice bajo riesgo de infrasensado de episodios arrítmicos de FV en los DAIs.

Metodología

La metodología propuesta para garantizar la consecución de los dos objetivos definidos se ha basado en el uso de técnicas de Data Analytics o Análisis de Datos (AD). El término AD se define como el proceso de aplicar sistemas computacionales al análisis de grandes cantidades de datos digitales para descubrir información útil que preste soporte a los procesos de toma de decisión. Hasta la fecha, este proceso ha sido expresado bajo una multitud de términos tales como *Inteligencia Artificial* (IA), *Minería de Datos* (MD), *Análisis Estadístico* (AE), *Inteligencia Empresarial* (IE), *Aprendizaje Máquina* (AM) o *Big Data* (BD). De forma específica, IA hace referencia a los procesos en los que cualquier dispositivo o máquina percibe su entorno y toma acciones para maximizar sus posibilidades de éxito sobre algún objetivo. MD cubre todos los aspectos del modelado y descubrimiento de patrones para propósitos predictivos. La combinación del análisis aplicado y descriptivo para descubrir características de los datos que validen hipótesis se conoce comúnmente como AE. Cuando el AE es aplicado para la extracción de información empresarial, el término utilizado es IE. Por último, AM y BD han sido recientemente los más usados cuando se hace referencia a técnicas que incluyen el pronóstico y clasificación predictiva, la regresión, y el procesado computacional de datos digitales de estructura heterogénea. Sin embargo, el término AD está destacando ya que intrínsecamente engloba técnicas de todas las áreas previamente mencionadas. Es por ello que AD también es considerado como un campo interdisciplinar que combina métodos y procedimientos de una gran variedad de dominios tales como estadística, teoría de la señal, reconocimiento de patrones, o inteligencia computacional (entre otros).

En relación con este trabajo y más específicamente con el objetivo *O1*, el diseño de una nueva metodología para la clasificación automática de episodios arrítmicos registrados por DAIs fue abordado mediante la combinación de técnicas procedentes de los dominios del MD, AE, y AM. Primeramente, se realizó un análisis descriptivo de los datos de los registros disponibles en el repositorio SCOOP, con el objetivo de diseñar una metodología lo más eficiente posible (en términos de coste y procesado computacional). En este repositorio, episodios arrítmicos de diferente duración e incluso con intervalos en blanco (periodos en los que no se registra señal) estuvieron presentes. Además, los datos que el DAI proporciona cuando registra un episodio arrítmico también estaban disponibles. Por tanto, técnicas basadas en la *Compresión de Datos* (CD) fueron utilizadas como una solución alternativa a los enfoques convencionales (tales como segmentación, filtrado, interpolación, o detección de ondas) para cuantificar la similitud entre pares de episodios. Estas técnicas permitieron abordar la clasificación mediante la definición de medidas de similitud, manteniendo además bajo coste computacional (característica crítica en sistemas de tiempo real). Específicamente, la metodología estima la similitud entre dos episodios arrítmicos haciendo uso de información de: (1) configuración del registro de la señal y resultados de la detección y terapia (si la hubiera); y (2) características arrítmicas ampliamente utilizadas en la práctica clínica tales como eventos de activación, frecuencia cardíaca, y forma de onda de la señal. Para abordar el problema de clasificación, de naturaleza potencialmente no lineal, para diseñar el clasificador se aplicó una técnica de AM basada en métodos núcleo fue aplicada para el diseño del clasificador. Por último, para mejorar iterativamente y hacer más eficiente este diseño, se propuso un enfoque de *Aprendizaje Activo* (AA).

Con respecto al objetivo *O2*, se aplicaron técnicas de AD procedentes de AE y AM para determinar el umbral de seguridad sobre las ondas-R. Primeramente, dos expertos llevaron a cabo una inspección visual de los episodios para seleccionar los más apropiados para el estudio. Seguidamente, se aplicaron técnicas para filtrar, rectificar, y detectar saturación sobre cada señal registrada por los DAIs, adaptándolas para seleccionar las amplitudes de ondas-R. Para determinar el umbral de seguridad se definió un procedimiento que maximiza una función objetivo. Finalmente, el valor del umbral fue validado usando técnicas de validación cruzada. Asimismo, las diferencias poblacionales en el valor del umbral fueron estadísticamente identificadas haciendo uso de test estadísticos.

Resultados

En referencia al primer reto planteado en la Tesis Doctoral, el enfoque propuesto para la clasificación de episodios arrítmicos alcanzó altas prestaciones en los experimentos realizados. Los resultados sobre un conjunto de 6233 episodios procedentes de 599 pacientes mostraron una tasa de acierto (y coeficiente kappa) en test con pacientes cercana al 78% (0.6) y 90% (0.8) para esquemas desbalanceados de 8 y 3 categorías, respectivamente. Estas tasas fueron obtenidas utilizando la clasificación de episodios proporcionada por un comité de expertos cardiólogos.

Se evaluó el impacto que sobre las prestaciones tiene la inclusión de cada uno de los bloques que conforman la metodología. Los resultados evidenciaron una mejora de hasta 18 y 26 puntos de porcentaje (esquema de 8 y 3 categorías, respectivamente) al incluir todos los bloques. Por otra parte, para estimar la similitud entre episodios arrítmicos se evaluaron tres medidas de similitud basadas en CD. Los resultados no mostraron diferencias significativas entre medidas. Por otra parte, un análisis en profundidad de las tasas de acierto por categoría mostraron que las capacidades de clasificación de la metodología son notablemente altas a la hora de discriminar entre episodios arrítmicos auriculares y ventriculares. Sin embargo, la metodología también

mostró dificultades en la discriminación de clases minoritarias del esquema de 8 categorías. El motivo de ello fue, en gran medida, el desbalanceo entre categorías.

Por último, los resultados en experimentos con técnicas de AA mostraron que estas herramientas pueden utilizarse para optimizar el proceso de etiquetado de episodios llevado a cabo por cardiólogos. Se compararon cuatro estrategias de selección, obteniendo que únicamente un 46% y un 22% (esquemas de 8 y 3 categorías) de los episodios contenían información relevante para la clasificación. De hecho, cuando estas prestaciones fueron analizadas por categorías, los resultados evidenciaron que la tasa de redundancia/no relevancia llegaba a ser mucho mayor.

Con respecto al segundo reto, sólo 229 de episodios de FV procedentes de 83 pacientes fueron útiles para los experimentos. En estos episodios se extrajeron amplitudes de onda-R para describir su comportamiento y relación con RB durante un episodio de FV. Los resultados confirmaron o mostraron que:

- La amplitud de onda-R durante un episodio de FV claramente decrece con respecto a RB.
- Cuanto mayor es la desviación de la amplitud de una onda-R con respecto a RB, mayor es la probabilidad de que la onda-R sea infrasensada.
- La tasa de infrasensado de ondas-R en un episodio de FV está exponencialmente relacionada con la distribución de las desviaciones con respecto a RB para la que: cuanto más negativa es la densidad de desviaciones durante un episodio de FV, mayor es la tasa de infrasensado de ondas-R en el episodio de FV.
- Una amplitud de onda-R durante un RB por debajo de 2.47 mV implica un alto riesgo de infrasensado de episodios de FV en DAIs.

Conclusiones

Esta Tesis Doctoral proporciona una solución computacional basada en AD a los dos retos planteados: (1) la clasificación automática de episodios arrítmicos registrados por los DAIs; y (2) la determinación de un umbral de seguridad sobre las ondas R que asegure bajo riesgo de infrasensado de episodios arrítmicos de FV. La conclusión general alcanzada para el primero de ellos es que la metodología propuesta para la clasificación presenta una estructura alternativa a la planteada en enfoques convencionales, alcanzando altas prestaciones en esquemas desbalanceados de 8 y 3 categorías. Además, la combinación efectiva de conceptos de AD ha permitido que la metodología:

- requiera un mínimo coste computacional;
- opere con episodios de diferente duración con intervalos sin registro;
- considere simultáneamente eventos de activación del corazón y formas de onda de la señal.

En un nuevo entorno de clasificación como el propiciado por la plataforma SCOOP, el análisis y etiquetado de episodios llevado a cabo por cardiólogos puede reducirse de forma considerable si se hace uso de metodologías como la aquí planteada. Esta conclusión ha sido alcanzada tras evaluar este efecto mediante un enfoque de AA. Los resultados mostraron que la metodología puede ser útil para seleccionar episodios relevantes para la clasificación. De hecho, las técnicas de AA evidenciaron que un alto número de los episodios incluidos en el repositorio de SCOOP son, o bien redundantes, o no contienen información relevante.

Por otra parte, la conclusión alcanzada para el segundo reto es que el umbral de seguridad sobre las ondas-R determinado mediante técnicas de AD puede influir considerablemente en las recomendaciones de las guías clínicas actuales. El valor obtenido como mínimo seguro es mucho menor al recomendado por las guías durante el implante del DAI (7 mV). Específicamente, los resultados han mostrado que amplitudes inferiores a 2.47 mV (rango mínimo-máximo de 2.27-2.62 mV) podrían dar lugar a situaciones de riesgo potencial para el infrasensado de episodios de FV. Por ello, procedimientos tales como las inducciones controladas de fibrilación podrían ser retrasados e incluso evitados si se tuvieran en consideración estas conclusiones. Asimismo, para la estimación del umbral fue necesario:

- describir el comportamiento de la tasa de infrasensado de ondas-R durante los episodios de FV;
- y definir la relación de la amplitud de onda-R entre un RB y FV.

Por último, esta Tesis Doctoral ha contribuido a la literatura científica proporcionando nuevas perspectivas para el desarrollo de nuevas herramientas basadas en técnicas de AD que puedan prestar un soporte en la práctica clínica. Los resultados obtenidos en este trabajo, los cuales están sustentados sobre un repositorio masivo de alta calidad como SCOOP, abren un amplio abanico de posibilidades para la mejora del diagnóstico cardíaco y toma de decisión. Esta Tesis representa un pequeño paso en este nuevo escenario sin precedentes que cambiará, con toda seguridad, las recomendaciones aportadas en guías clínicas y los sistemas de apoyo en cardiología.

Abstract

NOWADAYS, the massive storage of cardiac arrhythmic episodes from Implantable Cardioverter Defribrillators (ICDs) is opening up a new range of opportunities for electrophysiological knowledge extraction. Large and high quality databases are increasingly encouraging the development of new Data Analytics (DA) tools supporting cardiologists on their clinical decisions. Within this new context, this Thesis aims to provide a computational solution for two current challenges in cardiology: (1) the automatic classification of cardiac arrhythmic episodes recorded by ICDs; and (2) the determination of a safety threshold on R-wave amplitudes during a normal Baseline Rhythm (BR) for ensuring a low risk of undersensing fatal arrhythmic episodes in ICDs.

On the one hand, current ICDs are highly reliable detecting fatal arrhythmic episodes. However, the accurate automatic classification into specific classes remains a field for improvement. As a result, cardiologists still need to manually analyze each episode and to check whether the ICD detection and treatment were adequate. Therefore, a novel DA-based methodology for the automatic classification of ICD arrhythmic episodes is proposed in this Thesis. The methodology is defined to be potentially used in real-world ICD scenarios, since: (1) it requires minimal signal preprocessing due to memory and battery constraints; and (2) it deals with episodes of different duration in the presence of non-recording intervals. Likewise, the proposed methodology emulates the know-how of expert cardiologists, for which it simultaneously considers heart activation events and signal waveforms. Results on a set of 6,233 actual ICD detected episodes from 599 patients showed test accuracy rate (and kappa coefficient) close to 78% (0.6) and 90% (0.8) in both 8 and 3-class imbalanced schemes, respectively.

On the other hand, R-wave amplitudes during normal BR are the current indicators used to subjectively characterize the risk of undersensing Ventricular Fibrillation (VF) episodes in ICDs. However, a minimum value (or safety threshold) has not been established yet. Clinical guidelines recommend R-wave amplitudes of at least 7 mV at the ICD implantation. When the amplitude is lower, it is usual the induction of defibrillation tests in the clinical practice to ensure that undersensing does not occur. The drawback of this is those inductions increase the patient complications, and the efficacy and safety of ICD therapies are not improved by using these data as secondary information. In order to tackle this challenge, a DA-based procedure for estimating a safety threshold on BR R-wave amplitudes is proposed in this Thesis. To define this procedure: (1) the behavior and undersensing rate of R-wave amplitudes during VF episodes are defined; and (2) the R-wave amplitude relationship between BR and VF episodes is determined. Results on a set of 229 actual VF episodes from 83 patients showed that R-wave amplitudes lower than 2.47 mV can lead to potentially risk situations of non or late detection of VF episodes.

This Thesis contributes to scientific literature by offering new insights for the development of new DA-based tools to support cardiologists during the follow-up of patients with an ICD. Results for both raised challenges convincingly demonstrate that the new generation of large and high quality clinical databases plays a major role in future trends in cardiology.

Figures

Tables

Acronyms

AF **A**trial **F**ibrillation.

AI **A**rtificial **I**ntelligence.

AL **A**ctive **L**earning.

AR **A**ccuracy **R**ate.

AUC **A**rea **U**nder the ROC **C**urve.

BD **B**ig **D**ata.

BI **B**usiness **I**ntelligence.

BR **B**aseline **R**hythm.

BT **B**reaking **T**ies.

CL **C**ycle **L**ength.

CSM **C**ompression-based **S**imilarity **M**easure.

DA **D**ata **A**nalytics.

DC **D**ata **C**ompression.

DM **D**ata **M**ining.

EGM **E**lectro**G**ra**M**.

ICD **I**mplantable **C**ardioverter **D**efibrillator.

ID **I**nformation **D**istance.

IQR **I**nter-**Q**uartile **R**ange.

IU **I**nformation **U**nit.

JSD **J**accard **S**imilarity of **D**ictionaries.

JSWD **J**accard **S**imilarity of **W**eighted **D**ictionaries.

KC **K**olmogorov **C**omplexity.

KM **K**ernel **M**ethod.

LOPOCV **L**eave **O**ne **P**atient **O**ut **C**ross **V**alidation.

LU-D **L**evel **U**ncertainty with **D**iversity.

LZMA **L**empel-**Z**iv-**M**arkov **A**lgorithm.

LZW **L**empel **Z**iv **W**elch.

ML **M**achine **L**earning.

MMR **M**inimum-**M**aximum **R**ange.

MS **M**argin **S**ampling.

NCD **N**ormalized **C**ompression **D**istance.

NDD **N**ormalized **D**ictionary **D**istance.

NDS **N**ormalized **D**ictionary **S**imilarity.

NID **N**ormalized **I**nformation **D**istance.

NS **N**oi**S**e.

OM **O**ur **p**roposed **M**ethodology.

PD **P**atient **D**iversity.

PDF **P**robability **D**ensity **F**unction.

PPM **P**rediction by **P**artial **M**atching.

PPV **P**redictive **P**ositive **V**alue.

PVC **P**remature **V**entricle **C**omplex.

PWE **P**arzen **W**indow **E**stimator.

RMT **R**emote **M**onitoring **T**ransmission.

RS **R**andom **S**election.

SA **S**tatistical **A**nalysis.

SCOOP **S**cientific **COO**peration **P**latform.

SMVT **S**ustained **M**onomorphic **V**entricular **T**achycardia.

SPVT-VF **S**ustained **P**olymorphic **V**entricular **T**achycardia or **V**entricular **F**ibrillation.

ST **S**inus **T**achycardia.

SVM **S**upport **V**ector **M**achine.

SVT **S**upra**V**entricular **T**achycardia of flutter.

TWO **T**-**W**ave **O**versensing.

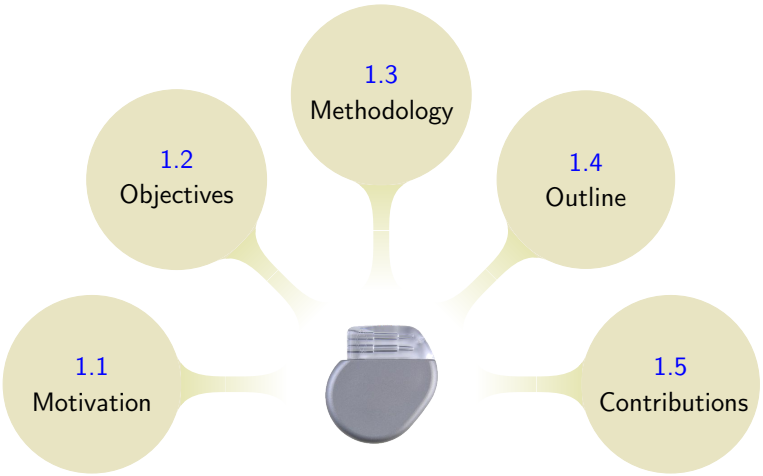
UST **U**ncertain **S**upraventricular **T**achycardia.

VF **V**entricular **F**ibrillation.

VT **V**entricular **T**achycardia.

WA **W**avelet **A**lgorithm.

Chapter 1



Introduction

THIS CHAPTER provides an overview of the Thesis. Firstly, the motivation behind the research developed in the Thesis is presented. Next, the objectives set are described. Eventually, an outline of this document and the main contributions to scientific literature are detailed.

1.1 Motivation and background

Cardiovascular diseases are the primary cause of death worldwide and they remain directly or indirectly responsible for more than 30% of reported deaths [2]. Nowadays, the most common cardiac condition is the *arrhythmia*, defined as an abnormally fast, slow, or erratic electrical disorder of the heart beating. Depending on the anatomical region of origin, arrhythmias are classified into supraventricular (atria) or ventricular (ventricles). The first ones are usually nonfatal disorders in the short term, whereas the second ones are severe disorders that, if not quickly controlled, can often result in a patient's sudden cardiac death [3].

Increasingly, patients at high risk of suffering from sudden cardiac death are treated with an **I**mplantable **C**ardioverter **D**efibrillator (ICD), a battery-powered device placed under the patient's skin with limited memory and computational resources. This device is equipped with thin wires (leads) directly positioned in the right ventricular apex (sometimes in the septum) and additionally in the right atrium or the left ventricular epicardium. These leads allow the ICD to record electrical digital signals, known as intracardiac **E**lectro**G**ra**M**s (EGMs), which are used to automatically detect cardiac arrhythmic episodes for applying, if necessary, therapies such as pacing, cardioversion, or defibrillation [4]. A real-world functional diagram of an ICD system is sketched in Fig. 1.1. Besides, when an arrhythmic episode is detected, ICDs store the corresponding record and, when the patient visits the cardiologist (or even during a remote follow-up transmission), this record can be downloaded into a database. Thus, cardiologists can later analyze each arrhythmic episode and determine whether detection and treatment were adequate or need to be improved [4].

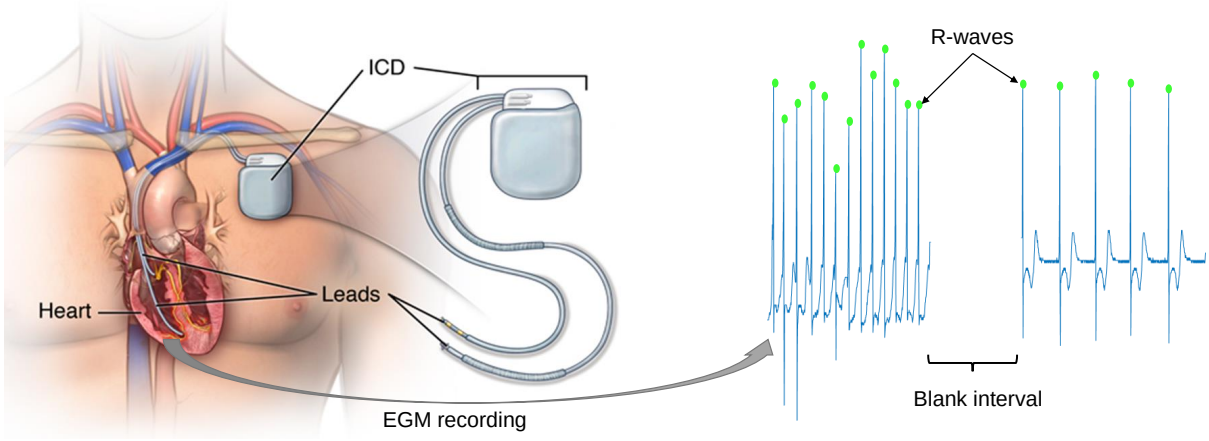


Figure 1.1: Functional diagram of an ICD system.

In recent years, the rapid development of the cloud computing technology is favoring the emergence of new platforms for automatically collecting massive digital data. In the ICD cardiology field, a model to be followed by research associations is the cloud computing system integrated in the **S**cientific **COO**peration **P**latform (SCOOP) project, a joint venture launched in 2011 by Medtronic® company [5] for cooperatively supporting a massive repository of cardiac arrhythmic episodes recorded by ICDs. In addition to its large size, SCOOP can be considered as a high quality repository since each episode is manually reviewed, analyzed, and labeled into an arrhythmic class by a scientific committee of expert cardiologists. This allows SCOOP to provide researchers with raw quality data for generating knowledge in the cardiac electrophysiology field [6], thus enhancing the support of clinical decisions (and even the treatment) on patients with an ICD. To date, SCOOP involves more than 50 clinical centers and contains more than 20,000 ICD arrhythmic episodes. The high variability, volume, and heterogeneity of these data are opening up a whole new range of possibilities for developing new analysis tools. This unprecedented situation in the cardiology field can be very useful in a large number of active cardiac research areas. On the evidence, SCOOP is the root of more than 60 current scientific research projects, and its results are beginning to be published in relevant cardiology research journals [6–9].

In this new context of massive ICD data, ***the motivation behind this Thesis is focused on two specific cardiac challenges***: (1) the automatic classification of cardiac arrhythmic episodes recorded by ICDs; and (2) the determination of a safety threshold ensuring a low risk of undersensing fatal arrhythmic episodes by ICDs.

Regarding the first challenge, current ICDs are highly reliable and sensitive in the detection of fatal arrhythmic episodes. However, their low specificity[1] and generalization capability for classifying arrhythmic episodes from different patients is open for improvement [10]. As a result, cardiologists have to manually analyze each detected episode and check whether detection and treatment were adequate. To date, algorithms developed for the automatic classification of arrhythmic episodes have been usually focused on the use of [10]: (1) databases with records from a low number of patients; (2) records labeled without following an appropriate gold standard; and (3) signal preprocessing and feature extraction stages conditioned (in most cases) by the specific ICD. In addition, EGMs of different time duration and the presence of blank intervals (as the example represented in Fig. 1.1) have not been usually considered [10], which is a key point in the real-world of ICDs. Hence, the design of new automatic classification approaches developed from a high quality database as SCOOP might be highly useful to develop new tools supporting cardiologists on their clinical decisions.

[1]The specificity is defined as the probability of correctly detecting/classifying a specific class.

Regarding the second challenge, the correct detection of R-waves [2] in EGMs (as green dots in Fig. 1.1) has always been a critical point of ICDs. The reason for this is that the ICD algorithms for detecting fatal arrhythmic episodes are mainly based on the instantaneous heartbeat rate [11], which is estimated from the difference time of two consecutive detected R-waves. Currently, the R-wave amplitude (voltage) during a normal patient's **B**aseline **R**hythm (BR) is the best indicator ensuring a low undersensing of fatal arrhythmic episodes such as the **V**entricular **F**ibrillation (VF) ones. Clinical guidelines recommend an R-wave amplitude of at least 7 mV at the ICD implantation [12, 13]. However, a minimum value, i.e. a safety threshold, has not been established yet. Several publications [14–16] have suggested that patients with R-wave amplitudes lower than 5 mV should undergo VF induction (defibrillation test) to ensure that undersensing does not occur. However, a defibrillation test increases the patient's complications and the efficacy and safety of ICD therapies are not improved [17]. Hence, the setting of a safety threshold on R-wave amplitudes determined from a high quality database as SCOOP could be highly useful to define new recommendations in current clinical guidelines for supporting cardiologists, leading to delay or even to avoid harmful therapies, such as defibrillation tests.

1.2 Objectives

The topic of this work falls within the scope of computational clinical applications for cardiology. In this context, ***the main objective of this Thesis is to develop new tools for supporting clinical decisions on patients treated with an ICD***. Specifically, this research aims at providing a solution for two challenges:

O1. The automatic classification of cardiac arrhythmic episodes recorded by ICDs.

O2. The determination of a safety threshold on BR R-wave amplitudes for ensuring a low risk of undersensing VF episodes in ICDs.

For the first challenge, a new computational classification methodology aiming at providing a real-world solution is proposed. The methodology is characterized by:

- requiring minimal computational burden, to be used in current systems of massive data;
- dealing with episodes of different time duration.
- managing episodes with blank intervals, since this is the actual digital structure of episodes recorded by ICDs;
- considering both heart activation events and EGM waveforms to emulate the classification know-how of expert cardiologists.

Regarding the second challenge, a procedure for estimating a safety threshold on BR R-wave amplitudes is proposed. To achieve this, three goals are required:

- describing the behavior and undersensing rate of R-wave amplitudes during VF episodes;
- analyzing the R-wave amplitude relationship between BR and VF episodes;
- evaluating results among patient populations to avoid an over-fitting of the procedure.

[2] An R-wave refers to cardiac electrical activity associated with the ventricle depolarization in a heartbeat. This activity is key in the detection of arrhythmic episodes in the common EGM representation because its maximum vale, or R-wave amplitude, allows to estimate the heart rate.

1.3 Methodology

To guarantee the achievement of the two objectives of this Thesis, the use of **D**ata **A**nalytics (DA) techniques is proposed as mechanism for the methodology. In literature, the term DA is defined as the process of applying computational systems to the analysis of large digital data to discover useful information that suggests conclusions for supporting decision-making [18]. To date, this process has been named under different terms. Thus, for example, **A**rtificial **I**ntelligence (AI) refers to any device or machine perceiving its environment and takes actions for maximizing its chance of success at some goal. **D**ata **M**ining (DM) covers all aspects of modeling and knowledge discovering from data for predictive purposes. The combination of descriptive and applied analysis for discovering data features that validate one hypothesis is commonly included in **S**tatistical **A**nalysis (SA). When SA is specifically applied for extracting business information, the term used is **B**usiness **I**ntelligence (BI). **M**achine **L**earning (ML) and **B**ig **D**ata (BD) have been more recently used to refer to techniques including predictive forecasting or classification, regression, and computational processing of unstructured digital data. However, the DA term is increasingly taking a leading role on a large number of different fields (such as health-care [19], marketing [20], education [21], or business [22]) since it intrinsically encompasses techniques from all mentioned areas. This is why DA is also considered as a very interdisciplinary field combining methods and procedures from many science domains such as statistics, signal theory, pattern recognition, or computational intelligence, among others.

In relation to this work and more specifically to objective *O1*, a set of techniques coming from DM, SA, and ML domains were used to design a new methodology for automatic classifying cardiac arrhythmic episodes recoded by ICDs. Firstly, a descriptive analysis of the SCOOP database was carried out to extract its main characteristics in order to design a methodology as efficient as possible. In this database, arrhythmic episodes of different duration and even with blank intervals were present. Besides, data provided by the ICD when recording each arrhythmic episode were also available. In this way, **D**ata **C**ompression (DC) techniques were used as an alternative solution to conventional approaches (such as segmentation, filtering, interpolation, or wave detection) for quantifying similarity between pairs of episodes. These techniques allow us to address this complex classification problem as a similarity issue, also keeping a low computational burden (what is a critical feature in real time systems). Specifically, the proposed methodology estimates the similarity between two arrhythmic episodes using information about: (1) lead configuration and detection/therapy results; and (2) arrhythmic features broadly used in the clinical practice and in the literature [14, 23–26], such as heart activation events or the EGM waveform. Then, aiming at addressing the arrhythmic classification of potential non-linear nature, a ML technique based on **K**ernel **M**ethods (KMs) [27] was used to design a classifier. Additionally, an **A**ctive **L**earning (AL) approach was also proposed for iteratively streamlining the classifier design.

Regarding objective *O2*, DA techniques coming from SA and ML were used to determine a safety threshold on BR R-wave amplitudes. Firstly, a visual inspection of the potentially useful episodes for the study was carried out by two experts to select the most appropriate ones. Next, techniques for filtering, rectification, and clipping detection were applied to each EGM in order to adapt them to the selection of R-wave amplitudes. Then, the behavior of amplitudes during a VF episode with respect to a normal BR was statistically described. Finally, a safety threshold was set maximizing a cost function and validated using ML techniques such as **L**eave **O**ne **P**atient **O**ut **C**ross **V**alidation (LOPOCV) [28]. Besides, the Student's t test [29] SA approach was also used for identifying statistical differences among data groups.

1.4 Thesis outline

This Thesis is structured in 4 more chapters. The content of each chapter is the following:

- **Chapter 2** presents SCOOP, the project for cooperatively supporting of massive ICD data that has allowed to achieve the objectives of this work. Features about the platform and the labeling process followed by the committee of expert cardiologists are detailed. A description of the repository of ICD cardiac arrhythmic episodes is presented.
- **Chapter 3** describes the computational solution proposed for the automatic classification of ICD arrhythmic episodes. Specifically, a new classification methodology based on DA tools such as DC techniques and KMs is presented. Performance results and the main features of this methodology are detailed. Likewise, a proposal for designing classification models based on an AL approach that streamline the manual and time-consuming labeling process is also presented.
- **Chapter 4** details the analysis carried out to find the model describing the relationship of R-waves amplitudes during BR and VF. The procedure for determining a safety threshold on BR R-wave amplitudes ensuring a low risk of undersensing VF episodes in ICDs is presented. Likewise, results are finally detailed.
- **Chapter 5** presents conclusions and future work.

1.5 Contributions

The work described in this Thesis has been previously published on journals indexed in the Journal Citation Reports service and presented in international conferences. Similar techniques to those presented in this document were developed for other research contributions which topic is not related to the Thesis. All contributions are enumerated next.

1.5.1 International journals

Thesis contributions

(1) **J.M. Lillo-Castellano**, I. Mora-Jiménez, R. Santiago-Mozos, F. Chavarría-Asso, A. Cano-González, A. García-Alberola, and J.L. Rojo-Álvarez, "Symmetrical compression distance for arrhythmia discrimination in cloud-based big data services," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1253–1263, 2015. Impact factor: 2.093. Category: *Computer Science, Information Systems*. Ranking: 29/144.

(2) **J.M. Lillo-Castellano**, M. Marina-Breysse, I. Mora-Jiménez, J.L. Rojo-Álvarez, D. Filgueiras-Rama *et al*, "Safety threshold of R-wave amplitudes in patients with implantable cardioverter defibrillator," *Heart*, vol. 102, no. 20, pp. 1662–1670, 2016. Impact factor: 5.693. Category: *Cardiac & Cardiovascular Systems*. Ranking: 15/124.

(3) **J.M. Lillo-Castellano**, J.L. Rojo-Álvarez, F. Chavarría-Asso, A. García-García, M. Martín-Méndez, A. García-Alberola, and I. Mora-Jiménez, "Classifying cardiac arrhythmic episodes via data compression," *IEEE Transactions on Biomedical Engineering*, vol.-, no.-, pp.-, *Submitted*, 2017. Impact factor: 2.468. Category: *Engineering, Biomedical*. Ranking: 22/76.

Other contributions

(4) **J.M. Lillo-Castellano**, I. Mora-Jiménez, R. Santiago-Mozos, J.L. Rojo-Álvarez, J. Ramiro-Bargueño, and A. Algora-Weber, "Weaning outcome prediction from heterogeneous time series using normalized compression distance and multidimensional scaling," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1737-1747, 2013. Impact factor: 2.981. Category: *Computer Science, Artificial Intelligence*. Ranking: 19/130.

(5) **J.M. Lillo-Castellano**, I. Mora-Jiménez, C. Figuera-Pozuelo, and J.L. Rojo-Álvarez, "Traffic sign segmentation and classification using statistical learning methods," *Neurocomputing*, vol. 153, pp. 286–299, 2015. Impact factor: 2.392. Category: *Computer Science, Artificial Intelligence*. Ranking: 31/130.

(6) O. Barquero-Pérez, R. Santiago-Mozos, **J.M. Lillo-Castellano**, B. García-Vizuete, R. Goya-Esteban, A.J Caamaño, J.L. Rojo-Álvarez, and C. Martín-Caballero, "Fetal heart rate analysis for automatic detection of perinatal hypoxia using normalized compression distance and machine learning," *Frontiers in Physiology*, vol. 8, 2017. Impact factor: 4.031. Category: *Physiology*. Ranking: 14/83.

Contributions (1) and (3) are related to objective *O1* (Chapter 3); whereas (2) is related to objective *O2* (Chapter 4).

1.5.2 International conferences

Thesis contributions

(1) M. Marina-Breysse, **J.M. Lillo-Castellano**, J.B. Martínez-Ferrer, J. Bautista-Alzueta, L. Pérez-Álvarez, A. García-Alberola, I. Fernández-Lozano and D. Filgueiras-Rama, "Umbral de seguridad de la onda R en el seguimiento de pacientes con desfibrilador," in *Spanish Society of Cardiology Congress*, Bilbao, Vizcaya, Spain, 22-24 October 2015.

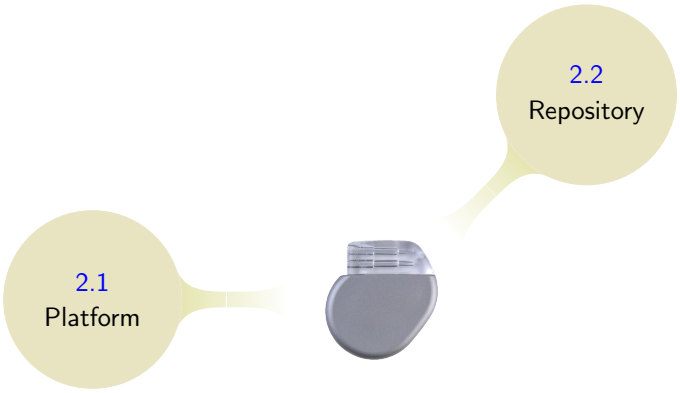
(2) **J.M. Lillo-Castellano**, I. Mora-Jiménez, R. Moreno-González, M. Montserrat-García-de Pablo, A. García-Alberola, and J.L. Rojo-Álvarez, "Big-data analytics for arrhythmia classification using data compression and kernel methods," in *Computing in Cardiology*, Nize, France, 6-9 September 2015.

(3) **J.M. Lillo-Castellano**, M. Marina-Breysse, J.B. Martínez-Ferrer, J. Bautista-Alzueta, L. Pérez-Álvarez, A. García-Alberola, I. Fernández-Lozano, A. Rodríguez, R. Porro and D. Filgueiras-Rama, "R-wave amplitude during follow-up of implantable cardioverter defibrillators to predict undersensing upon spontaneous ventricular fibrillation," in *European Society of Cardiology Congress*, London, Inner London, United Kingdom, 29 August - 2 September 2015.

Other contributions

(4) R. Santiago-Mozos, B. García-Vizuete, **J.M. Lillo-Castellano**, J.L. Rojo-Álvarez, C. Martín-Caballero, "On the early detection of perinatal hypoxia with Information-Theory based methods," in *Computing in Cardiology*, Zaragoza, Spain, 22-25 September 2013.

Contribution (2) is related to objective *O1*; whereas (1) and (3) are related to objective *O2*.



Chapter 2

The SCOOP Project

THE GROWING INTEGRATION of the cloud computing technology into systems of a high number of organizations is currently paving the way for the emergence of new computational environments responding to the technological needs. A current example in the cardiology field is the cloud computing system integrated in the research project named SCOOP. This project is supported by a joint-venture coordinated by Medtronic® company and aims to massively collect ICD data by means of multicenter cooperations. The feasibility of SCOOP and the high quality of its data have formed the basis of the work developed in this Thesis. Next, a description of this project and its repository of ICD arrhythmic episodes are presented.

2.1 Cooperation platform

In recent years, the massive storage of digital data is opening up a new range of opportunities for DA. It is becoming increasingly common the emerge of new systems for supporting massive data such as those coming from current ICDs. An example of this is SCOOP, a research project supported by a joint venture coordinated by Medtronic® company, which was launched in 2011 for massively collecting ICD data that generate new electrophysiological knowledge. As a result, a new scientific BD tool based on cloud computing technology was developed in order to automatically incorporate records of ICD arrhythmic episodes into a multi-center database. Currently, more than 50 centers cooperate nowadays in an altruistic way in favor of cardiology knowledge. Besides, to ensure the legal regulations in the scientific data exploitation as well as the patient's privacy, SCOOP project was included within an observational research study, named UMBRELLA, which is registered in the U.S. National Institutes of Health [30].

Nowadays, the SCOOP database is continuously increasing and more than 20,000 ICD records are currently included in it. This fact is possible because the SCOOP platform is technically based on the remote monitoring capabilities of ICDs. This allows an automatic transmission of the information during the patient's follow ups, without the need of additional manual task. Likewise, to ensure a fair cooperation among the participating centers, an agreement was defined by Medtronic®. The agreement provides that, when a research group aims to carry out a specific study using SCOOP data, the group has to firstly present a proposal within the an-

nual periods established for that purpose. Then, a specialized committee of experts from the Medtronic® team analyzes each proposal and decides about it. Thus, when a proposal is approved, Medtronic® provides the group with the necessary and available SCOOP data to carry out the specific study.

2.2 Repository of cardiac arrhythmic episodes

The most important element of the SCOOP project is its digital repository of ICD cardiac arrhythmic episodes. In SCOOP, each episode is manually analyzed and labeled into an arrhythmic category by a scientific committee of 6 expert cardiologists (reviewers) with deep ICD background. For this task, a systematic clinical procedure was defined by the Medtronic® team aiming at ensuring that the labeling of one expert was not conditioned by the labeling provided by another expert. This procedure (sketched in Figure 2.1) is followed for each episode and specifically involves 4 stages:

1. Committee members are randomly separated into 2 groups of 3 experts. For each group, 2 experts perform a first classification and the remaining expert is suitable for reserve.

2. The episode is assigned to the first group and labeled by the 2 experts. If their independent labeling match, then the episode is labeled and the process ends. Otherwise, the expert in reserve labels the episode and, if his/her labeling matches one of the two previous labellings, the episode is labeled and the process ends.

3. If the episode has not been labeled yet (i.e., every expert in the first group provides a different label), then the episode is assigned to the second group and stage 2 is repeated.

4. Finally, if the episode is not labeled by any group, an external review is required and the episode is jointly labeled by the 6 experts of the committee in an external joint face review.

This sequential procedure was defined with the aim of ensuring that at least two independent labellings match, conferring to this process a great independence.

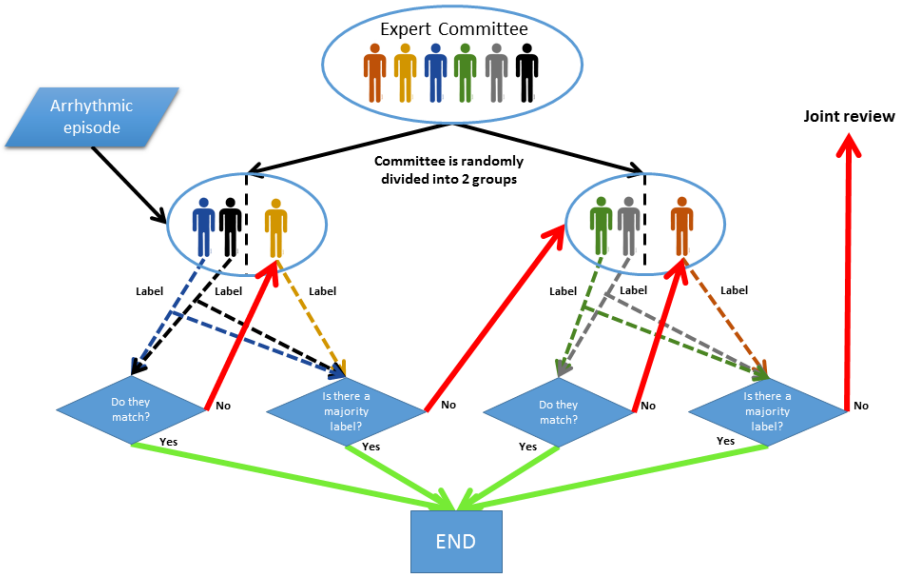


Figure 2.1: Diagram for the labeling process of an arrhythmic episode in SCOOP.

Likewise, to provide committee with a common classification criterion, a guide (summarized in Table 2.1) was previously defined according to the main factors of the arrhythmic episodes. This guide categorizes entire episodes and considers 8 arrhythmic categories (the 8-class scheme) based on: (1) the arrhythmia origin; (2) the heart activation events; and (3) the signal EGM waveform. Besides, the first criterion groups categories into three major sets (the 3-class scheme: atrial, ventricular, and other) broadly used in clinical practice. In Fig. 2.2, the relative frequency of each category of SCOOP database for both 8 and 3-class schemes is represented. Note that the clear imbalance among categories will greatly affect in the development of DA tools.

Origin	**Category** (acronym)	**Description**
Atrial	**S**inus **T**achycardia (ST)	Normal P and R-waves, regular P-R intervals, and heart activation rate in the range of 100 to 180 beats per minute (bpm).
	Atrial **F**ibrillation (AF)	Variable P-R intervals, irregular P-P intervals, uneven P-wave waveform, and a minimum atrial activation rate of 400 bpm.
	Supra**V**entricular **T**achycardia of flutter (SVT)	Regular P-P intervals, stable P-wave waveform, ir/regular ventricular activation rate, and atrial activation rate >100 bpm in flutter (usually between 250 and 400 bpm).
	Uncertain **S**upraventricular **T**achycardia (UST)	Ir/regular ventricular activation rate >100 bpm and unavailable atrial information. Assumption of arrhythmic episode with atrial origin.
Ventricular		Ventricular events that either last longer than 30 seconds or meet ICD programmed criteria finished by an electrical therapy, these being:
	Sustained **M**onomorphic **V**entricular **T**achycardia (SMVT)	Regular and sustained rhythms.
	Sustained **P**olymorphic **V**entricular **T**achycardia or **V**entricular **F**ibrillation (SPVT-VF)	Variable waveform and/or chaotic rhythm.
Other	**T**-**W**ave **O**versensing (TWO)	Wrong detection of the T-waves by the ICD, which interprets T-waves as R-waves and estimates the heart activation rate with a double real value during a normal sinus rhythm.
	Noi**S**e (NS)	Irregular and chaotic rhythm with short duration and heart activation rate > 500 bpm, which starts and finishes spontaneously.

Table 2.1: Arrhythmic categories defined by the SCOOP scientific committee as a guide for the classification of arrhythmic episodes.

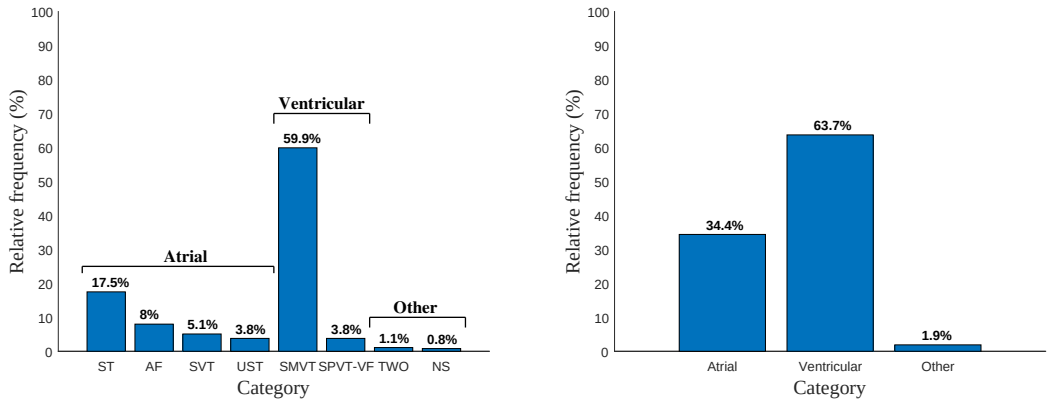
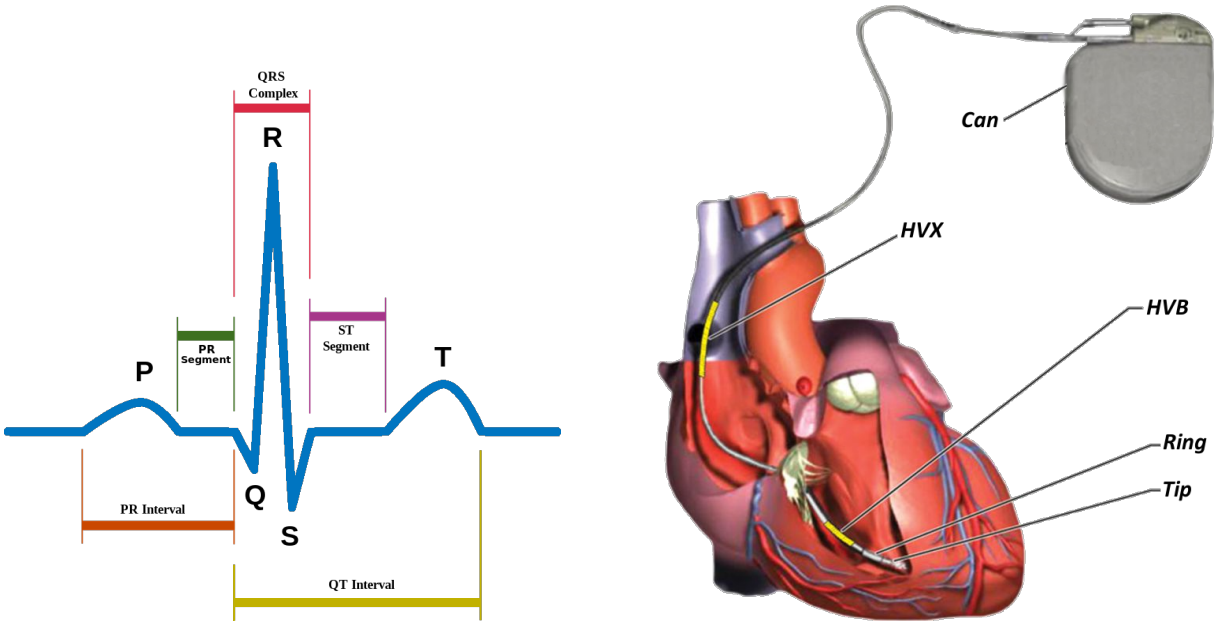


Figure 2.2: Bar graph representing the relative frequency of the arrhythmic categories in the SCOOP database for both 8 (a) and 3-class (b) schemes.

On the other hand, the heart activation events and the EGM waveform are based on the electrical activity propagation during a heartbeat. In a normal sinus rhythm, three parts are differentiated in this activity (which standard representation is shown in Fig. 2.3a): the P-wave, originated by the atria depolarization; the QRS complex[1], associated with the ventricle depolarization and atria reporalization; and the T-wave, originated by the ventricle repolarization. The waveform of this activity and specially the time intervals among electrical events form the basis for the definition of arrhythmic categories. More details can be consulted in [31].



(a) (b)

Figure 2.3: (a) Standard EGM waveform of a heartbeat in normal sinus rhythm. This waveform changes depending on the position of the sensing electrodes. (b) Diagram of an ICD system composed of a single lead positioned in the right ventricular apex. This image was extracted from [1] and later modified. *Can*, *HVX*, *HVB*, *Ring*, and *Tip* represent sensing electrodes.

In the SCOOP database, episodes have a variable duration which depends on the specific arrhythmic event and the ICD model and programming. Besides, episodes can present blank intervals of no-recording. Specifically, the storage structure of each episode consists of two EGMs and an associated set of time-notes (known as *markers*). On the one hand, EGMs correspond with electrical signals (voltage difference) stored as digital data with a sampling frequency of 128Hz and 1 byte of resolution per sample. Specifically, EGMs were recorded from an ICD system composed of two/three leads positioned as follows: the first one in the right ventricular apex or septum; the second one in the right atrium; and the third one (if available) in the lateral or posterolateral wall of the left ventricle (see more details in [4]). For clarification purposes, a diagram of an ICD system composed of a single lead positioned in the ventricular apex is represented in Fig. 2.3b. The sensing electrodes are also represented, where *Can*, *HVB*, *HVX*, and *A(V)tip/ring* denote the can of the ICD, the right ventricle coil, the superior vena cava coil, and the atrial (ventricular) tip/ring electrodes, respectively. Note that multiple combinations of electrode pairs (which are programmable in the ICD) are possible to record an EGM. Therefore, multiple pairs of EGM configurations are possible in a recorded arrhythmic episode. Specifically, 8 configurations (*AtipToAring*, *AringToHVB*, *VtipToVring*, *VtipToHVB*, *CanToVring*, *CanToHVB*, *CanToHVX*,

[1]The QRS complex actually includes three separate waves: Q, R, and S.

HVBToHVX) and 16 different pairs of configurations are present in the SCOOP database. However, two configuration pairs, [*AtipToAring*, *VtipToVring*] and [*CanToHVB*, *VtipToVring*], are the most common and they include 94% of total.

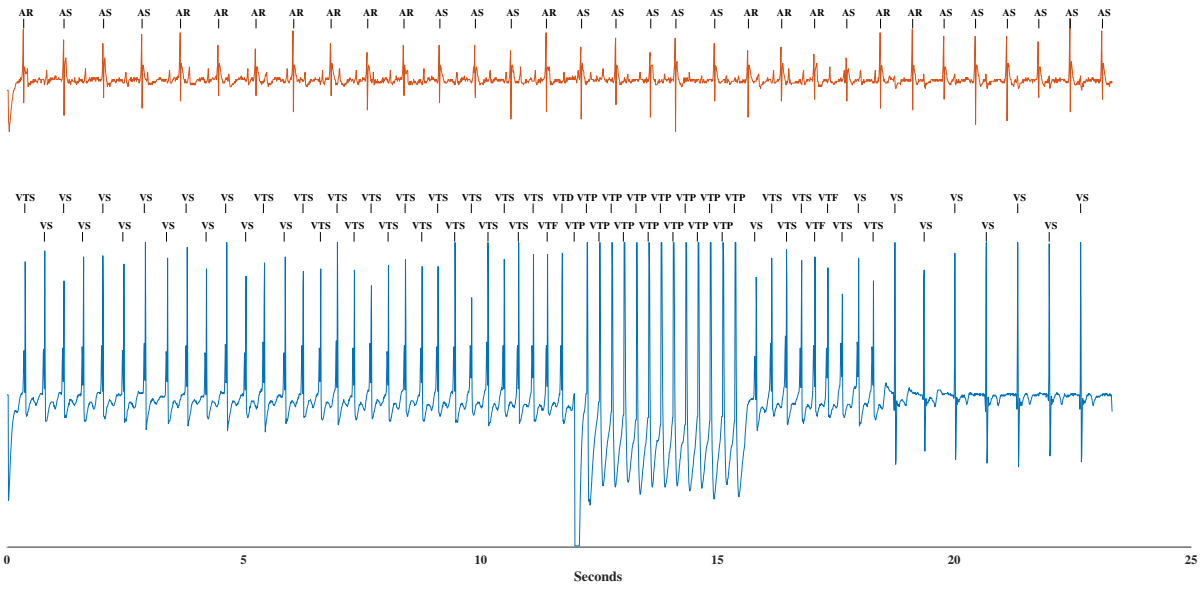
On the other hand, markers denote relevant ICD pacing and detection events. In ICDs, markers are stored as text strings with their associated time-stamps, resulting in 23 possible categories. Besides, they are just available for *AtipToAring* and *VtipToVring* configurations. The five most frequent categories (more than 80% total) are VS, AS, VTS, VTP, and AR; denoting V ventricular, A atrial, S sensing, T tachycardia, P pacing [2], and R refractory [3]. In Table 2.2 a description of each marker and its relative frequency is summarized.

Lead configuration	Type	Marker	Description	Ocurrence (%)
AtipToAring (Atrial)	**Sensing**	AS	Activation.	11.6
		ASPVAB	Activation during post-ventricular atrial blanking.	6.1
		ATS	Activation during tachycardia.	<0.1
	Pacing	AP	Normal stimulation.	1.5
		APP	Proactive stimulation.	0.1
		ATP	Anti-tachycardia stimulation.	<0.1
	Detection	AR	Refractory.	6.8
		ATD	Tachycardia.	0.1
		AFD	Fibrillation.	<0.1
VtipToVring (Ventricular)	**Sensing**	VS	Activation.	39.7
		VTS	Activation during tachycardia.	14.8
		VFS	Activation during fibrillation.	3.5
	Pacing	VP	Normal stimulation.	1.6
		BVP	Bi-ventricular stimulation.	3.8
		VTP	Anti-tachycardia stimulation.	6.4
	Detection	VR	Refractory.	<0.1
		VTD	Tachycardia.	0.6
		VFTD	Fast-tachycardia.	0.1
		VFD	Fibrillation.	0.2
		VTMD	Tachycardia during monitory zone (no therapies applied).	<0.1
-	**Parameters**	CE	End of charging of ICD discharge capacitors.	0.1
		VCD	Ventricular cardioversion defibrillation.	0.1
		MS	Shift of the ICD algorithm mode.	<0.1

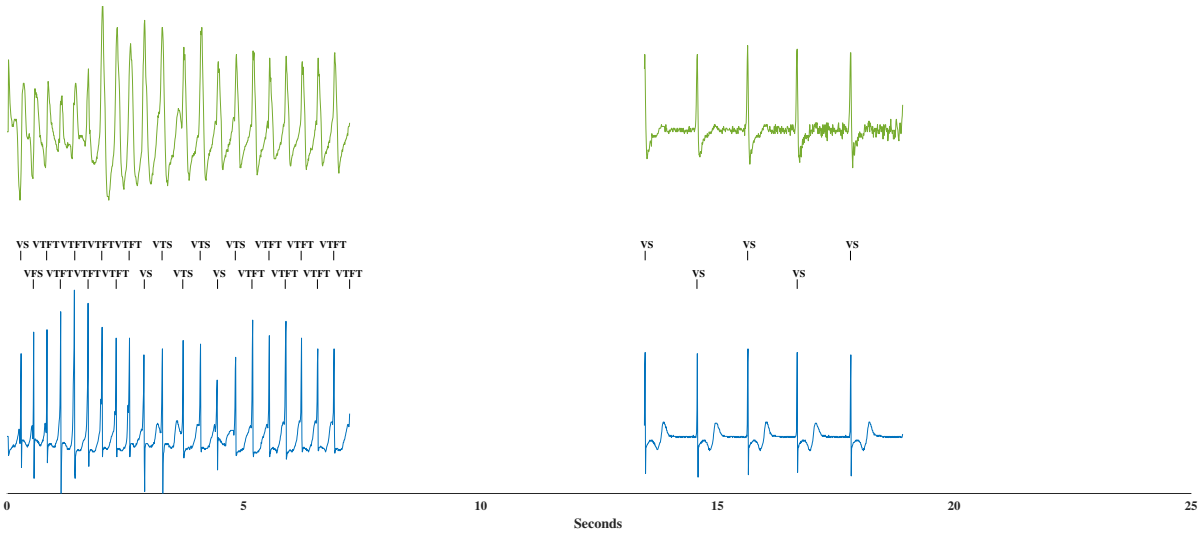
Table 2.2: Markers associated to arrhythmic episodes stored in the SCOOP database.

Finally, arrhythmic episodes in the SCOOP database have an average length of 25±17 seconds (mean±standard deviation), and a median length of 20 seconds with an **I**nter-**Q**uartile **R**ange (IQR) of 14 seconds. The average number of episodes per patient is 10±20, and the median number of episodes is 5 with an IQR of 10. As an example, two randomly chosen episodes (one of each more common configuration) are represented in Fig. 2.4. Signals in the lower panels (in blue) represent *VtipToVring* EGM, whereas those in the upper panels (in red and green) represent *AtipToAring* and *CanToHVB* EGMs in Fig. 2.4a and Fig. 2.4b, respectively. Note the difference in the episode durations and the presence of a blank interval in the second one. Ticks and text labels over EGM waveforms represent markers.

[2] Stimulated electrical impulse.
[3] Non-repetition time period.



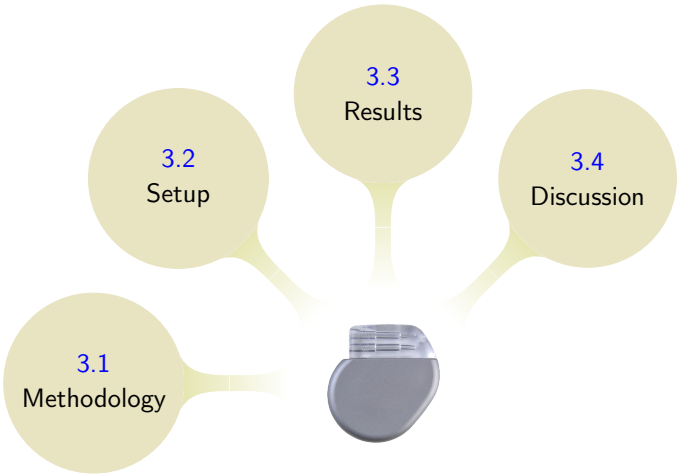
(a)



(b)

Figure 2.4: Example of two actual arrhythmic episodes randomly selected from the SCOOP database. Lower panels correspond to *VtipToVring* EGMs. Upper panels are *AtipToAring* (a) and *CanToHVB* (b) EGMs. Ticks and text labels over EGM waveforms represent ICD markers. Empty middle zone in (b) corresponds to a blank interval.

Chapter 3



Classification of Cardiac Arrhythmic Episodes

CURRENTLY, automatic classification of cardiac arrhythmic episodes remains an open active research field that aims to improve the performance of ICDs constrained by low memory, processors, and batteries. ICDs are highly reliable in the detection of fatal arrhythmic episodes but, by contrast, they do not correctly detect other specific ones. As a result, cardiologists have to manually analyze and classify each ICD detected episode and check whether detection and treatment were adequate. In this chapter, a new methodology for classifying cardiac arrhythmic episodes recorded by ICDs is proposed. The methodology design has been focused on: (1) addressing the current ICD limitations on classification after detecting an arrhythmic episode; and (2) ensuring a low computational burden. Firstly, the methodology stages and DA techniques are described. Next, a proposal for iteratively streamlining the design of the classification scheme is also presented. Experimental setup and results are detailed below. Results show the potential of the proposed methodology as a tool supporting cardiologist in their diagnosis, forecasting its future use in ICDs and related systems. A discussion is finally summarized at the end of the chapter.

3.1 Methodology for the classification

The general way for annotating relevant time instants by current ICDs is by means of markers. A marker denotes: (1) the detection of a heart activation event; (2) the detection of an arrhythmic episode; or (3) the application of an electrical therapy. Together with the EGM waveform, markers are highly informative because the clinical classification criteria of arrhythmic episodes are mostly based on both features. Taking this information into account, we propose a new methodology of three stages for classifying cardiac arrhythmic episodes recorded by ICDs. The main characteristics of this methodology are: (1) it preprocesses EGMs to enhance the search of relevant patterns based on the heart activation events and signal waveform; (2) it considers both the lead configuration and the marker information; (3) it is based on DC and KM techniques; and (4) it requires a low computational burden. We describe next the stages of the proposed methodology.

3.1.1 EGM preprocessing

The first stage transforms the original EGM in another signal aiming to emphasize heart activation events and EGM waveform. As a result, EGM redundancy is increased and DC algorithms can take advantage of it for estimating the similarity among episodes in a better way. This stage is organized in three blocks, as explained below.

- Firstly, the *differentiation* block refers to the application of the derivative operator to emphasize the temporal electrical variations. They are very relevant to identify the EGM fragments corresponding to the heart activation events.
- Next, the application of the *thresholding* block sets to 0 EGM variations lower than a threshold *th*, hence increasing the subsequent coding redundancy.
- Finally, the *sign function* block emphasizes the EGM fragments according to both heart activation events and waveform criteria. This will increase the redundancy of the waveform, make easier the search of characteristic patterns, and reduce the computational burden.

The diagram of the EGM preprocessing stage using a fragment of a real-world EGM is sketched in Fig. 3.1. Note how the preprocessed EGM changes when thresholding block is applied, mitigating the effect of emphasizing in the same way both low and high EGM variations. Besides, if th is suitably fixed (as in Fig. 3.1), heart activation events and waveform information from the original EGM are preserved. Then, DC techniques are next applied to compute similarity between preprocessed episodes and implicitly define patterns which help to classify each arrhythmic episode into a specific category.

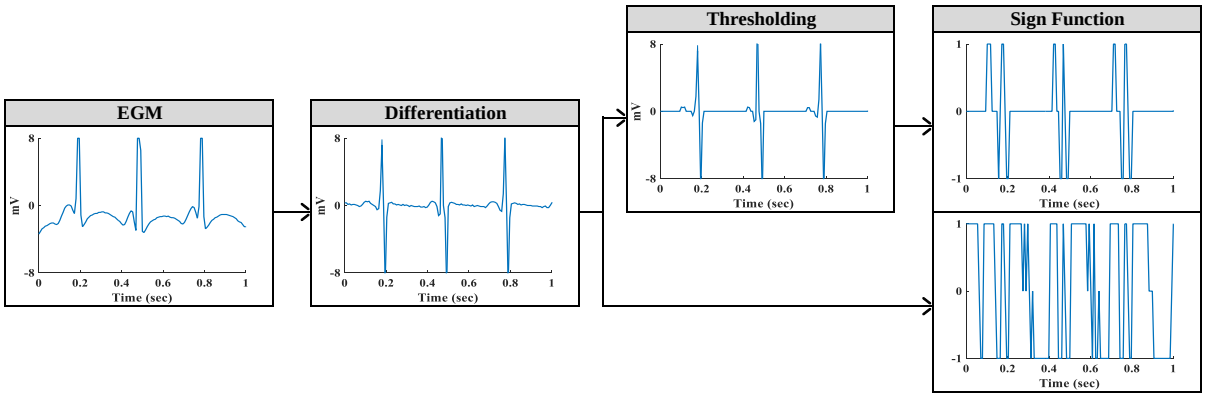


Figure 3.1: Diagram of the EGM preprocessing stage and the resulting signal when thresholding ($th = 0.125$ mV) is (or not) considered.

3.1.2 Similarity estimation via data compression

The second stage extracts relevant patterns from the arrhythmic episode by means of DC. Lossless DC algorithms (also called *compressors*) reduce (compress) the size of a dataset with no information loss. That is, any dataset can be exactly recovered from the compressed dataset. Lossless DC techniques are increasingly being used for defining similarity measures between digital entities, addressing methodologies for clustering and classification tasks (among others). These approaches have been applied to data of diverse nature such as images, text documents, genetics, or biomedical signals [32–35].

For a compressor, information is represented by means of standard digital units, known as **I**nformation **U**nits (IUs), which basic element is the bit. From the bit, other IUs can be defined using sequences of bits, such as the byte (sequence of 8 bits). A **C**ompression-based **S**imilarity **M**easure (CSM) is characterized by its ability to exploit the information shared by two digital entities using compressors [36]. The mathematical basis of a CSM comes from the Information Theory field, where the **K**olmogorov **C**omplexity (KC) concept defined for strings of bits is used. Given a bit string $\boldsymbol{x}$, the KC of $\boldsymbol{x}$ (denoted by $K_{\boldsymbol{x}}$) is the number of bits of the shortest computer program of the fixed reference computing system capable of producing $\boldsymbol{x}$ [37] (i.e., the ultimate lossless compressed version of $\boldsymbol{x}$). Using this definition, the **I**nformation **D**istance (ID) between two strings $\boldsymbol{x}$ and $\boldsymbol{y}$, denoted by $ID_{\boldsymbol{xy}}$, is also defined [37] and it expresses the length of the shortest program computing $\boldsymbol{x}$ *from* $\boldsymbol{y}$,

$$ID_{\boldsymbol{xy}} = K_{\boldsymbol{xy}} - min\{K_{\boldsymbol{x}}, K_{\boldsymbol{y}}\}, \tag{3.1}$$

where $K_{\boldsymbol{xy}}$ is the KC of string $\boldsymbol{xy}$ (concatenated pair $\boldsymbol{x}$ and $\boldsymbol{y}$), and $min\{\cdot,\cdot\}$ denotes the minimum operator. It can be shown that $ID_{\boldsymbol{xy}}$ is actually a metric which depends on the strings length [38]. As a consequence, if the value of $ID_{\boldsymbol{xy}}$ between two short strings is large with respect to their lengths, then both strings are very different; however, if the same value of $ID_{\boldsymbol{xy}}$ is obtained for two long strings, then those strings are very similar. Thus, a relative measure was defined [39] to express similarity regardless of the string length, called **N**ormalized **I**nformation **D**istance (NID), given by

$$NID_{\boldsymbol{xy}} = \frac{ID_{\boldsymbol{xy}}}{max\{K_{\boldsymbol{x}}, K_{\boldsymbol{y}}\}}, \tag{3.2}$$

where $max\{\cdot,\cdot\}$ denotes the maximum operator. Note that, due to $K_{\boldsymbol{xy}} \leq K_{\boldsymbol{x}} + K_{\boldsymbol{y}}$, NID is in the range from zero to one [39].

In practice, the KC is not computable [37] and data compressors are used to approximate its value. For a given compressor C, let us denote by $C_{\boldsymbol{x}}$ and $C_{\boldsymbol{xy}}$ the length, in bits, of the compressed versions for strings $\boldsymbol{x}$ and $\boldsymbol{xy}$, respectively. Using this approximation in Eq. (3.2), the **N**ormalized **C**ompression **D**istance (NCD) can be defined [39] as

$$NCD_{\boldsymbol{xy}} = \frac{C_{\boldsymbol{xy}} - min\{C_{\boldsymbol{x}}, C_{\boldsymbol{y}}\}}{max\{C_{\boldsymbol{x}}, C_{\boldsymbol{y}}\}}. \tag{3.3}$$

Ideally, NCD values close to zero (one) correspond to similar (non-similar) strings. In practice and when real-world compressors are used, $C_{\boldsymbol{xy}} \neq C_{\boldsymbol{yx}}$ and the symmetry property is not fulfilled by the NCD and a high computational burden is required [39]. Due to its low computational burden without degradations in performance, a branch of the lossless DC algorithms based on the use of dictionaries [32] is increasingly taking advantage in the definition of similarity measures. A *dictionary* is a table including associations between patterns (strings of IUs known as *words*) and their descriptions (known as *codes*). Depending on the compressor, different word-code assignment strategies are used. All strategies characterize codes by their size in terms of digital storage of IUs, which is inversely proportional to the occurrence frequency of the associated word [40]. Thus, the compressed dataset is obtained by replacing words with codes.

A dictionary-based CSM between two strings is essentially determined by their common words. This concept is known as *dictionary matching* and is the basis of different CSM definitions. Let us consider two strings $\boldsymbol{x}$ and $\boldsymbol{y}$, and their associated sets of words $W_{\boldsymbol{x}}$ and $W_{\boldsymbol{y}}$, respectively, extracted from their dictionaries. Topological relations between $W_{\boldsymbol{x}}$ and $W_{\boldsymbol{y}}$ can be represented by a Venn diagram (see Fig. 3.2) which allows to easily visualize: (1) the total space of elements, i.e., $W_U = W_{\boldsymbol{x}} \cup W_{\boldsymbol{y}}$, referring to the total number of words (elements inside the thick line); (2)

the shared information, i.e., $W_I = W_{\boldsymbol{x}} \cap W_{\boldsymbol{y}}$, referring to the shared words (in gray zone). These relations allow to define different CSMs based on the dictionary matching concept. The most usual compression measure is the **N**ormalized **D**ictionary **D**istance (NDD) [41], a dissimilarity metric in the interval $[0, 1]$. In terms of similarity, the NDD complementary measure (1 - NDD) is the **N**ormalized **D**ictionary **S**imilarity (NDS) [41] and it is defined as

$$NDS_{\boldsymbol{xy}} = \frac{|W_I|}{max\{|W_x|, |W_y|\}} \tag{3.4}$$

where $|\cdot|$ denotes the cardinality operator, which expresses the number of elements in a set.

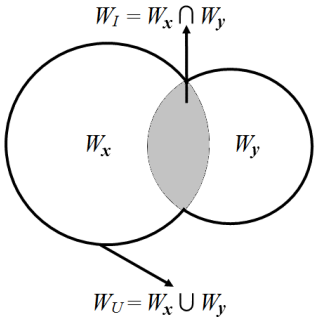


Figure 3.2: Venn diagram illustrating the topological relationship among words of two dictionaries ($W_{\boldsymbol{x}}$ and $W_{\boldsymbol{y}}$). The intersection set (W_I) is shown as a gray zone, and black thicker line delimits the union set (W_U).

When dealing with sets, the *Jaccard Similarity* [42] is very useful. If this similarity is extrapolated to the dictionary case, the **J**accard **S**imilarity of **D**ictionaries (JSD) is defined as

$$JSD_{\boldsymbol{xy}} = \frac{|W_I|}{|W_U|} \tag{3.5}$$

Note that both NDS and JSD just account for the cardinality of word sets and do not consider other information potentially relevant, such as the cardinality of each word (referring to the number of IUs contained in the word), which could be useful to determine the similarity between strings. Taking this consideration into account and extrapolating the JSD, in this work we propose the **J**accard **S**imilarity of **W**eighted **D**ictionaries (JSWD), which is defined as

$$JSWD_{\boldsymbol{xy}} = \frac{\sum_{i=1}^{M} |W_I^i|}{\sum_{i=1}^{P} |W_U^i|} \tag{3.6}$$

where W_I^i and W_U^i denote the i-th word included in W_I (out of a total of M words) and W_U (out of a total of P words), respectively.

The main objective of using CSMs in the proposed methodology is to estimate similarities between arrhythmic episodes. As described later, each arrhythmic episode is characterized with a word set (relevant patterns) extracted from its compression which is used in the CSMs computation. Some of the most common dictionary-based compressors are **L**empel **Z**iv **W**elch (LZW), Zip, GZip, **L**empel-**Z**iv-**M**arkov **A**lgorithm (LZMA), or **P**rediction by **P**artial **M**atching (PPM) [40]. In our experiments, we have used the LZW compressor because it is a lossless compression algorithm, present a low computational burden, and its performance has broadly been studied in literature [40]. For a simple compression example of the LZW algorithm (extracted from [43]), let us consider the string of bits $\boldsymbol{x}$ = 0110011001. The resulting dictionary after applying LZW to $\boldsymbol{x}$ is shown in Table 3.1. The cardinality of this dictionary is $|W_{\boldsymbol{x}}| = 8$ and

$W_{\boldsymbol{x}}^1$	$W_{\boldsymbol{x}}^2$	$W_{\boldsymbol{x}}^3$	$W_{\boldsymbol{x}}^4$	$W_{\boldsymbol{x}}^5$	$W_{\boldsymbol{x}}^6$	$W_{\boldsymbol{x}}^7$	$W_{\boldsymbol{x}}^8$
0	1	01	11	10	00	011	100

Table 3.1: Dictionary obtained with LZW algorithm on $\boldsymbol{x} = 0110011001$.

its set of words gives us an idea of the information contained in $\boldsymbol{x}$. Note that a parameter to consider in this algorithm is the IU choice (bit in the example), which defines the compression alphabet and, as represented in the example, determines the number of words found. Next, we describe this point in detail.

Alphabet for compression

The set of states represented by an IU is known as *alphabet*. For a compressor, the choice of the IU is crucial. If the IU used by the compressor cannot represent the number of states required by the problem, the search of patterns (words, strings of IUs) will be suboptimal. In this case, the compressor will extract words, but they will not represent the information properly.

As indicated in Section 2.2 of Chapter 2, an arrhythmic episode consists of 2 EGMs (voltage signals in a range of ±8 mV) recorded using a byte per sample, and a set of markers (23 categories) recorded as strings with an associated time-stamp. In this way, 256 amplitude values and 23 marker categories are possible. It seems reasonable to think that EGMs coming from different lead configurations and markers should be encoded using different ranges of states. Thus, patterns extracted during the compression that come from different "sources" could be differentiated and the classification approach would be better. This is because the same pattern will represent a different cardiac activity when it is registered by different lead configurations. The same comment applies to the marker value, which codifies time notes events instead of amplitude values.

Taking the above considerations into account, a source encoding was carried out in this work to characterize each type of lead configuration and marker by a specific range of alphabet states. After the EGM preprocessing, only 33 different states were possible: 9 associated to EGMs (3 per each ICD lead configuration: *CanToHVB*, *VtipToVring* and *AtipToAring*), and 23 + 1 values to markers (23 possible categories + no marker value). Hence, an IU of 6 bits is enough to represent those states since an alphabet of 64 potential states is represented. In our experiments, LZW compressor was set up with an IU of 6 bits and ranges of values in [1,24] and [25,33] were specifically used for markers and EGMs, respectively. When other classification schemes are considered (e.g. omitting some blocks of the proposed methodology, see Section 3.3), more than 64 states can be required. In these cases, IUs of more bits and other different ranges of values were used.

Support words of an arrhythmic episode

Figure 3.3 sketches the methodology proposed for extracting the characteristic patterns of an arrhythmic episode composed of two EGMs (E1 and E2) and a set of markers (M). Essentially, this methodology corresponds to a transformation of the information contained in the arrhythmic episode to enhance the search of patterns based on heart activation events and EGM waveform. On the one hand, EGMs are preprocessed and subsequently encoded depending on their ICD lead configuration. On the other hand, markers and their associated time-stamps are used to compute a temporal sequence of the same length as that of he EGMs. This sequence has a zero value when no marker is available, and a different state otherwise. Resulting EGM signals

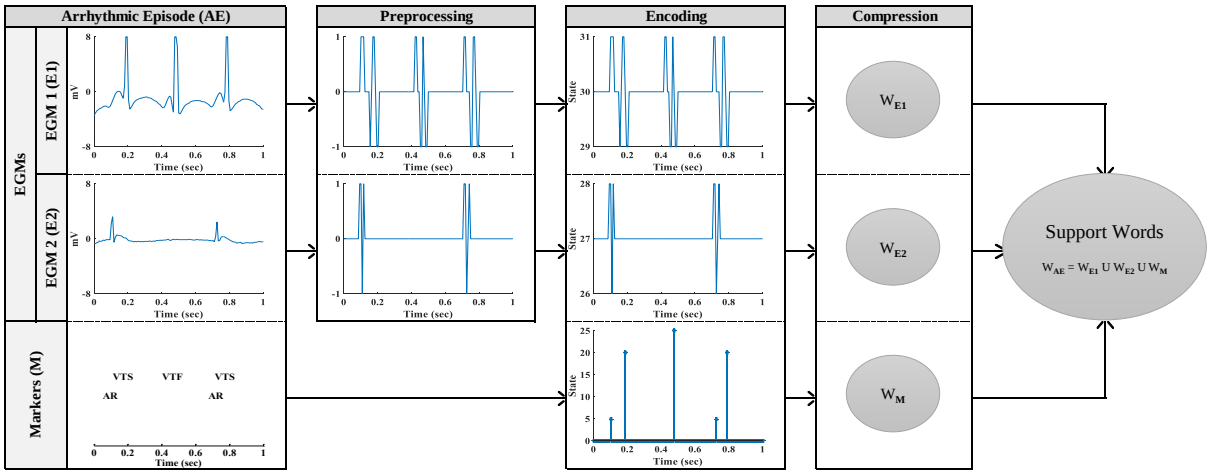


Figure 3.3: Diagram of the proposed methodology to extract the support words of an arrhythmic episode recorded by an ICD.

and marker sequence are individually compressed to obtain three sets of words per arrhythmic episode, namely, $W_{\mathbf{E1}}$, $W_{\mathbf{E2}}$, and $W_{\mathbf{M}}$. These sets are jointly considered to create the set W_{AE}, which contains the *support words* of an arrhythmic episode, such that $W_{AE} = W_{\mathbf{E1}} \cup W_{\mathbf{E2}} \cup W_{\mathbf{M}}$.

From W_{AE} sets, similarities between episodes were estimated computing two CSMs used in the literature (NDS and JSD) and a new one define in this work (JSWD). These measures allowed to compute matrices which were considered as kernels by the following classification scheme.

3.1.3 Classification using a compression-based kernel

The last stage is a ML classifier (classification model). As previously indicated, our problem corresponds to a multi-categories task in which each arrhythmic episode (also named instance or sample) is a multi-signal structure of variable time duration and even with blank intervals. To address this problem of potential non-linear nature, we propose the use of a nonlinear KM for the classification task. This is because KMs allow to develop general frameworks for performing pattern analysis on many types of data [44]. Besides, when KMs are suitably adjusted to data, complex relations can be detected and simplified by combining their non-linear performance and the efficiency of linear algorithms [44].

Let us consider the dataset $\mathcal{S}$ of N instances labeled in two categories, $\mathcal{S} = \{\mathbf{x}_i, y_i\}_{i=1}^{N}$, where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{\pm 1\}$ coming from an unknown distribution $p(\mathbf{x}, y)$. A KM allows to map the original dataset defined on input space $\mathcal{X}$ (i.e., $\{\mathbf{x}_i\}_{i=1}^{N} \subseteq \mathcal{X}$), onto a Hilbert space $\mathcal{H}$ with a higher (possibly infinite) dimension, in accordance with the Cover's Theorem [45]. The mapping is carried out by means of a function denoted as $\phi(\cdot) : \mathcal{X} \to \mathcal{H}$. However, this operation can be computed using the *kernel trick*, such that the similarity between samples from the original dataset is computed using pairwise inner products between mapped samples. Thus, $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ ($\langle \cdot, \cdot \rangle$ expresses the inner product operator) and it is used to compute the kernel matrix and design linear algorithms in the mapped space. The most common kernels used in the literature are: (1) the linear kernel, $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$; (2) the polynomial kernel, $(\langle \mathbf{x}_i, \mathbf{x}_j \rangle + 1)^p$, with $p \in \mathbb{N}$; and (3) the radial basis function kernel, $exp\left(\frac{-||\mathbf{x}_i - \mathbf{x}_j||^2}{2\sigma^2}\right)$, where $\sigma \in \mathbb{R}^+$ is the Gaussian width and $||\cdot||$ denotes the norm operator. These kernels present a sample length dependence (i.e., all samples must have the same length) and, as in the polynomial and radial basis cases, they require the tunning of free-parameters (p for (2) and σ for (3)). Although the

last two kernels are the most common ones, new kernels can be defined using similarity measures that fulfill the Mercer's Theorem [46], which can be summarized in the following two points:

- The similarity measure must be a symmetric continuous mapping function.
- The kernel matrix computed from the similarity measure must be positive semi-definite.

The CSMs used in this work fulfill both conditions. Firstly, as defined in Eqs. (3.4) - (3.6), the three CSMs are symmetric measures. Secondly, in each experiment (later defined in Section 3.2) we checked that each CSM matrix was positive semi-definite, i.e., all eigenvalues from its singular value decomposition were non-negative [47]. Hence, we considered each CSM in the way $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \text{CSM}(\mathbf{x}_i, \mathbf{x}_j)$.

Due to its implicit relationship with kernels, a **S**upport **V**ector **M**achine (SVM) [48] has been used as classification model in the proposed KM. In a linearly separable problem of two categories, the SVM basic idea is to find the decision boundary maximizing the margin, defined as the distance between the boundary and its closest samples. Using KMs, this idea is readily extrapolated for nonlinear problems by applying the SVM onto the mapped space. Thus, the SVM minimizes the following *primal* problem:

$$\min_{\mathbf{w},\xi_i,b} \left(\frac{1}{2} \left\| \mathbf{w} \right\|^2 + C \sum_{i=1}^{N} \xi_i \right) \tag{3.7}$$

with constraints

$$y_i(\phi(\mathbf{x}_i)\mathbf{w} + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad \forall i = 1, \ldots, N \tag{3.8}$$

where $\mathbf{w}$ and b (bias) define the linear classifier (hyperplane). Parameter C provides a trade-off between training error and margin, which is the distance between well classified samples and the separation hyperplane, whereas ξ_i are positive slack variables enabling to deal with permitted training errors. Then, a binary linear classifier is defined in $\mathcal{H}$ for any testing instance $\mathbf{x}_*$ as $f(\mathbf{x}_*) = sgn\left(\langle \mathbf{w}, \phi(\mathbf{x}_*) \rangle + b\right)$, where $sgn(\cdot)$ denotes the sign function. The *primal* problem can be solved by maximizing its *dual* counterpart [48]

$$\max_{\alpha_i} \left(\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) \right) \tag{3.9}$$

with constraints $0 \leq \alpha_i \leq C \quad \forall i = 1, \ldots, N$ and $\sum_{i=1}^{N} \alpha_i y_i = 0$, where α_i are Lagrange multipliers corresponding to restrictions (3.8). Finally, the decision function for any test vector $\mathbf{x}_*$ is formulated as

$$f(\mathbf{x}_*) = sgn \left(\sum_{i=1}^{N} \alpha_i y_i \mathcal{K}(\mathbf{x}_i, \mathbf{x}_*) + b \right) \tag{3.10}$$

The SVM is mainly characterized by: (1) the choice of the kernel; and (2) the setting of free-parameters to achieve a good generalization performance. In our experiments, the LibSVM software [49] was used for designing the SVM classifier. This implementation also allowed us: (1) to use precomputed kernels (in our case, the CSMs); and (2) to tackle multi-class schemes. For these schemes, the LibSVM software implements the "one-against-one" approach [50]. For k classes, this approach: (1) constructs $k(k-1)/2$ SVM classifiers training each one from data of two different classes; and (2) classifies each sample assigning it to the majority class among the $k(k-1)/2$ SVMs. A detailed performance analysis for this approach is presented in [51].

3.1.4 Streamlining via active learning

In general, classification models are designed from ML techniques and they are usually based on similarity measures among samples. However, when those models are aimed at screening massive datasets, a point becomes crucial: the selection of a reduced set containing relevant and representative samples. The reason of this comes mainly from two issues: (1) the exponential increase of computational memory and burden as samples are incorporated in the design of this type of models, which makes technically unfeasible their performance; and (2) the need of labels to train, which in the heath-care DA requires time consuming by physicians.

In DA, the approach of building training sets by iteratively improving the model performance through smart sampling is known as AL. The basic idea of AL is that a classification model trained with a reduced set of well-chosen samples can perform as well as another one trained using a larger set of randomly chosen samples [52]. To achieve this objective, AL focuses on the iterative interaction human operator and model, where: (1) the model returns to the human those samples with uncertainty in the classification outcome; and (2) the human provides the model with his/her knowledge by labeling samples with uncertainty, which in turn reinforce the model performance. In this way, the resulting classification model is optimized by the available information (samples), which is the most difficult task for the model. To date, the effectiveness of AL techniques has been demonstrated in a large number of current applications, such as remote sensing [53–56], improvement of medical tools [57–59], processing of digital data streams [60], vehicle detection [61], or robotic agriculture [62].

For the proposed classification methodology, AL is very suitable for streamlining the SVM classifier as SCOOP database will increase in the future. Currently, committee of expert cardiologists continue labeling episodes which, in most cases, tend to be redundant or non-relevant for classification. Hence, to evaluate how AL can improve the classification performance of the proposed methodology and simultaneously reduce the number of episodes to be labeled, an AL approach is proposed.

Let us consider a initial labeled dataset $\mathcal{X}^0 = \{\mathbf{x}_i, l_i\}_{i=1}^{I}$ used to design a classification model, where samples $\mathbf{x}_i \in \mathbb{R}^d$, labels $l_i \in \{1, \ldots, L\}$, and L is the number of classes. Let us also consider an unlabeled dataset $\mathcal{U}^0 = \{\mathbf{x}_j^u\}_{j=1}^{J}$, known as *pool of candidates*, with samples $\mathbf{x}_j^u \in \mathbb{R}^d$ and $J \gg I$. The goal of an AL algorithm is to iteratively select those samples from the unlabeled dataset which could improve the classification performance of the model designed with labeled dataset. In the AL literature, the selection criterion is known as *heuristic* [52] and it is usually based on the discovery of *uncertain* or *low confidence* regions. This selection is based on the assumption that if the model would provide an adequate classification in those regions, performance would be greatly improved. In short, the general learning procedure of an AL algorithm can be described as follows [52]: (1) for a given iteration ϵ, $\mathcal{S}^\epsilon$ composed of C_ϵ candidates, $\mathcal{S}^\epsilon = \{\mathbf{x}_j^u\}_{j=1}^{C_\epsilon}$, is selected from $\mathcal{U}^\epsilon$ following a predefined heuristic; (2) the selected candidates are labeled by the human, i.e., $\{l_j\}_{j=1}^{C_\epsilon}$ are determined; (3) the labeled set $\mathcal{S}^\epsilon$ is added to $\mathcal{X}^\epsilon$, i.e., $\mathcal{X}^{(\epsilon+1)} = \mathcal{X}^\epsilon \cup \mathcal{S}^\epsilon$, and removed from $\mathcal{U}^\epsilon$; (4) the model is retrained; (5) the process is repeated until a stopping criterion is met.

Hence, an AL algorithm requires the interaction between humans and models. The human provides his/her domain knowledge through the labeling, whereas the model provides its computational capacity. The selection strategy defining an heuristic is crucial for the AL algorithm and sets the success of the learning process. To date, different AL heuristics have been proposed. At least four AL heuristic families are differentiated [52], those based on: committees, large-margin, posterior probability and clusters. Due to the nature of the classification model used

(a SVM) in the proposed methodology, an AL algorithm based on large-margin heuristics was evaluated in this study. Evidences of the performance improvement using this type of heuristics were previously reported in [53] (even in problems of selective reduction of training sets [63]).

An AL algorithm based on large-margin is characterized by quantifying the *uncertainty* depending on the distance to the separating hyperplane. Classification schemes such as SVMs are naturally tailored for this type of AL algorithms. In these schemes, the distance to the separating hyperplane can be used as a direct way for estimating the confidence on the labeling of a sample. As described in Section 3.1.3, a SVM is conceptually based on the sparse representation of the training set defining the classification model. In this setting, relevant samples are those with a nonzero associated coefficient (i.e., the support vectors). Hence, samples from $\mathcal{U}$ lying within the margin of the current model are potentially containing relevant information, which would improve the classification performance by making the classification hyperplane shift after their inclusion in the training set. By using this fact, different large-margin based heuristics can be defined depending on the classification problem [52]. In the following, we describe the most common ones and those used in this work:

Margin Sampling (MS): this heuristic is also known as *most ambiguous* [64], and is characterized by taking the SVM geometrical properties into account. Specifically, it considers the fact that unbounded support vectors are labeled samples on the margin whose decision function value is exactly one [65]. In this way, candidates are selected by minimizing the following cost function:

$$\mathcal{S}^{\epsilon}_{MS} = \underset{\mathbf{x}^u_j \in \mathcal{U}^{\epsilon}}{argmin} \left\{ \underset{c}{max} \left\{ d_c(\mathbf{x}^u_j) \right\} \right\} \tag{3.11}$$

where $d_c(\mathbf{x}^u_j)$ is the distance of sample $\mathbf{x}^u_j$ to the hyperplane defined for class $c \in \{1, \ldots, L\}$ in a one-against-all multi-class schemes.

Breaking Ties (BT): this heuristic, proposed by Luo *et al* [66], is characterized by dealing with the most uncertain class. Thus, the difference of sample-hyperplane distances between the first and the second most probable classes are now considered, i.e., the following cost function is minimized:

$$\mathcal{S}^{\epsilon}_{BT} = \underset{\mathbf{x}^u_j \in \mathcal{U}^{\epsilon}}{argmin} \left\{ d_{c1}(\mathbf{x}^u_j) - d_{c2}(\mathbf{x}^u_j) \right\} \tag{3.12}$$

where $d_{c1}(\mathbf{x}^u_j) = \underset{c}{max} \left\{ d_c(\mathbf{x}^u_j) \right\}$, $d_{c2}(\mathbf{x}^u_j) = \underset{c \setminus c_1}{max} \left\{ d_c(\mathbf{x}^u_j) \right\}$, and c_1 is the class associated to d_{c1}.

Level Uncertainty with Diversity (LU-D): the concept of data *diversity* essentially concerns the ability of selecting samples as different as possible to each other, within a predefined constraints (diversity). The goal of this is to avoid selecting similar samples in a single iteration. To date, the data diversity has been widely studied in the AL field and it depends on each specific problem [52]. In the scenario of arrhythmic episodes, the diversity can be basically defined using the patient information, i.e., the patient who suffered from the arrhythmic episode. Thus, aiming to analyze the effect of this **P**atient **D**iversity (PD), in this work we defined two new heuristics, the MS-PD and the BT-PD. For both heuristics, it is necessary to consider an unlabeled dataset $\mathcal{V}^0 = \{\mathbf{x}^v_j, p^v_j\}_{j=1}^{J}$, where patient identifiers $p^v_j \in \{1, \ldots, P\}$, and P is the number of different patients. Then, candidates in both MS-PD and BT-PD heuristics are selected by minimizing the (3.11) and (3.12) cost functions, respectively, but subject to the constraint:

$$p^v_j \neq p^{v'}_j \quad \forall\, \mathbf{x}^v_j, \mathbf{x}^{v'}_j \in \mathcal{S}^{\epsilon} \tag{3.13}$$

ensuring that PD takes priority over episode uncertainty. In other words, at each iteration the most uncertain episodes are selected but ensuring to sample all patients.

3.2 Experimental setup

3.2.1 Database

The work here developed is associated with a SCOOP study approved in December 2013. Then, following the established agreement, a set of 6,233 arrhythmic episodes from 599 patients (those available to that date) was provided by Medtronic® for this study. Episodes had been registered between January 2012 and December 2013. As mentioned in Section 2.2, the storage structure of each episode consisted of two EGMs and an associated set of markers. Episodes had been recorded either using the configuration pair [*CanToHVB*, *VtipToVring*] or [*AtipToAring*, *VtipToVring*] (3,334 and 2,899 episodes from 320 and 279 patients, respectively). Their lengths were mean±std of 23.5±14.6s, median of 19.4s and IQR of 13.3s. The number of episodes per patient was 10.4±21.2, median of 4, and IQR of 10. The relative frequencies of the arrhythmic categories for both 8 and 3-class schemes are shown in Table 3.2.

Scheme					
3-class	**8-class**	**3-class**	**8-class**	**3-class**	**8-class**
Category		**Number of Episodes**		**Frequency** (%)	
Atrial	ST	2,164	896	34.72	14.38
	AF		710		11.39
	SVT		297		4.76
	UST		261		4.19
Ventricular	SMVT	3,955	3,692	63.45	59.23
	SPVT-VF		263		4.22
Other	TWO	114	80	1.83	1.28
	NS		34		0.55

Table 3.2: Relative frequency of the arrhythmic categories in the database used in this work.

3.2.2 Figures of merit

The most common figure of merit used to quantify the performance of a classifier is the **A**ccuracy **R**ate (AR) [67], which expresses the success rate in the classification. However, AR may lead to wrong conclusions with imbalanced data, i.e., when there is a quite different number of instances per category. In imbalanced scenarios, categories with more instances tend to weight higher and "saturate" the classifier, what affects to the classification of instances belonging to minority categories. Then, the term *baseline accuracy* is frequently used to define the AR obtained when all instances are classified as belonging to the majority category (i.e., to the category with the highest a priori probability). Thus, when imbalance is present, better performance interpretations than those obtained with AR can be achieved by considering other figures of merit such as the sensitivity, the **P**redictive **P**ositive **V**alue (PPV), the area under the **A**rea **U**nder the ROC **C**urve (AUC), the F-Score, or the Cohen's Kappa coefficient, among others [67].

On the one hand, the Cohen's Kappa coefficient (κ) [67] has been used in this work to assess the classifier performance in both 8 and 3-class imbalanced schemes, thus complementing the AR. The Kappa coefficient was originally defined for measuring the agreement between two classifiers and has been broadly used to assess the classifier performance in multi-category imbalanced scenarios [68]. It expresses how reliable the performance of a classifier is, quantifying in a range of [0,1] the degree of chance in the classification [67]. It is assumed that low κ values are associated to a classification by chance, while high values are associated to a good classification in all categories (different than what would be expected by chance). Moreover, κ has shown high correlation with other figures of merit also defined for the multi-category case, such as the AUC weighted by a priori category or the mean F-Measure [67].

On the other hand, sensitivity and PPV have been jointly used to assess the classifier performance for each specific category. The sensitivity of a category, also called the true positive rate, measures the rate of instances of a category that are correctly classified as such. Conversely, the PPV of a category (also called precision) expresses the rate of instances classified as such that are true instances of the category. The joint use of both figures of merit helps to better interpret the classifier performance for specific categories, which allows to identify potential improvements in classification.

3.2.3 Performance assessment

A strategy based on random resampling without replacement was used to assess the generalization performance of the proposed methodology. The original dataset of episodes $\mathcal{E}$ was randomly divided 1,000 times (runs) into 2 subsets $\mathcal{X}$ and $\mathcal{T}$, such that $\mathcal{E} = \mathcal{X} \cup \mathcal{T}$. For each run r we defined: (1) a training set $\mathcal{X}^{(r)}$ containing episodes from 70% of patients in $\mathcal{E}$; and (2) a test set $\mathcal{T}^{(r)}$ with episodes of the remaining 30% of patients. The 70/30 ratio was experimentally checked as suitable for the number of episodes included in our database. To find the most appropriate set of hyperparameters $\boldsymbol{p}$ in each run, a patient 5-fold cross validation strategy [28] was performed on $\mathcal{X}^{(r)}$. In this way, the consideration of different runs for training/test allows to get the confidence interval of the figures of merit and perform statistical comparisons among different scenarios. Note also that the performance is estimated when no episodes of the patient who is tested are considered for training. Otherwise, the performance may be overestimated and patient-dependent. Specifically, the process followed for each run r was:

1. To divide $\mathcal{X}^{(r)}$ in 5 patient independent folds of episodes, i.e., $\mathcal{X}^{(r)} = \mathcal{X}_{F_1}^{(r)} \cup \mathcal{X}_{F_2}^{(r)} \cup \cdots \cup \mathcal{X}_{F_5}^{(r)}$
2. For each round i of the 5-fold:
   - To build the design set $\mathcal{X}_D^{(r,i)} = \bigcup_{j \neq i} \mathcal{X}_{F_j}^{(r)}$ and the validation set $\mathcal{X}_V^{(r,i)} = \mathcal{X}_{F_i}^{(r)}$
   - To compute the dictionary of each episode in $\mathcal{X}_D^{(r,i)}$
   - To compute the CSM matrix of $\mathcal{X}_D^{(r,i)}$, named $\text{CSM}_D^{(r,i)}$
   - To design the SVM classifier for each set of hyperparameters $\boldsymbol{p}$ using $\text{CSM}_D^{(r,i)}$, getting $\text{SVM}^{(r,i,\boldsymbol{p})}$
   - For each episode e of $\mathcal{X}_V^{(r,i)}$:
     - To compute the dictionary of e
     - To compute the vector of CSMs between e and $\mathcal{X}_D^{(r,i)}$, $\text{CSM}_V^{(r,i,e)}$
     - To label e using $\text{CSM}_V^{(r,i,e)}$ as input to each $\text{SVM}^{(r,i,\boldsymbol{p})}$
   - To build vector $\mathcal{Y}_V^{(r,i,\boldsymbol{p})}$ with labels provided by $\text{SVM}^{(r,i,\boldsymbol{p})}$
3. To build vector of labels of $\mathcal{X}^{(r)}$ for each $\boldsymbol{p}$ as $\mathcal{Y}^{(r,\boldsymbol{p})} = \bigcup_{i=1}^{5} \mathcal{Y}_V^{(r,i,\boldsymbol{p})}$
4. To choose the most appropriate set of hyperparameters $\boldsymbol{p}^*$ according to the figure of merit evaluated using $\mathcal{Y}^{(r,\boldsymbol{p})}$
5. To compute the dictionary of each episode in $\mathcal{X}^{(r)}$
6. To compute the CSM matrix of $\mathcal{X}^{(r)}$, $\text{CSM}^{(r)}$
7. To design the SVM classifier, named $\text{SVM}^{(r,\boldsymbol{p}*)}$, using $\text{CSM}^{(r)}$ and $\boldsymbol{p}^*$

8. For each episode t in $\mathcal{T}^{(r)}$:
   - To compute the dictionary of t
   - To compute the vector of CSMs between t and $\mathcal{X}^{(r)}$, $\text{CSM}^{(r,t)}$
   - To label t using $\text{CSM}^{(r,t)}$ as input to $\text{SVM}^{(r,\boldsymbol{p}*)}$
9. To build the set of labels of $\mathcal{T}^{(r)}$ to obtain figures of merit

As described in Section 3.1, the extraction of relevant patterns depends on three components: (1) the source encoding, (2) the EGM preprocessing, and (3) the inclusion of marker information. In order to assess their influence, AR and κ were assessed as each component was progressively incorporated to the methodology, defining a total of 7 classification scenarios. For each scenario, three CSMs (NDS, JSD, and JSWD) were considered using different sets of support words (W_{AE} sets) depending on the specific scenario. In the (O) scenario, W_{AE} is obtained from the compression of the two concatenated original EGMs. This scenario corresponds to no EGM preprocessing and it is included for comparison purposes. In the (E) scenario, support words were obtained according to the diagram depicted in Fig. 3.3, but without considering the preprocessing stage and markers. The remaining (D), (T), and (S) scenarios incrementally incorporate the differentiation, thresholding, and sign function blocks of the EGM preprocessing stage. Results for the full methodology (diagram of Fig. 3.3) are presented in the final (F) scenario. Moreover, the classification capability when using just maker information is presented in the (M) scenario. In this case, support words were extracted just from the compression of the marker sequence (i.e., only last row of diagram depicted in Fig. 3.3).

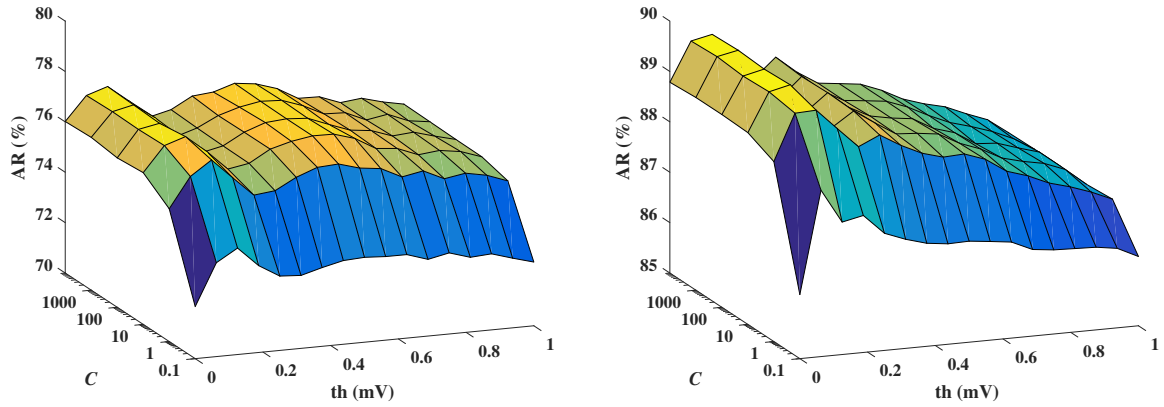
As previously indicated in Section 3.1.2, 33 states are required when all blocks are included: 9 associated to EGMs, and 23 + 1 to markers. Note that, depending on the number of blocks included, a different number of states can be required. For example, the (O) scenario does not consider any difference between the 3 ICD lead configurations (*CanToHVB*, *VtipToVring* and *AtipToAring*), so only 256 states are required since a byte of resolution is used by the ICD to record EGMs. However, in the (E) scenario, the number of states is triplicated since 256 states per ICD lead configuration are necessary (i.e., a total of $256 \times 3 = 768$ states). If the differentiation block is further included, the number of states is duplicated to tackle both positive and negative differences. Finally, when thresholding and sign function blocks are included, the number of states is reduced to 9 possible states (3 per lead configuration). Table 3.3 shows the number of states required on each scenario and the IU used for compression in each case (aiming to reduce computational requirements).

Scenario	Number of States	IU
(O)	256	8 bits
(E)	768	10 bits
(D)	1,536	11 bits
(T)	1,536	11 bits
(S)	9	4 bits
(F)	33	6 bits
(M)	24	5 bits

Table 3.3: Number of states required and IU used in the 7 classification scenarios.

3.2.4 Settings

Regarding the classification assessment, the proposed methodology requires choosing two hyperparameters depending on the specific classification scenario: (1) C parameter for the SVM classifier in all scenarios; and (2) th parameter for the thresholding block in (T), (S), and (F) scenarios. Ranges of $[10^{-1},10^{4}]$ and $[0,1]$ mV were experimentally explored in each scenario for C (resolution of one order of magnitude) and th (resolution of 0.0625 mV), respectively. As mentioned, for each run a 5-fold cross validation strategy was performed to choose the most appropriate C and th (when necessary). Specifically, hyperparameters were set to those ones that provided the best median AR on 5-fold cross validation. Fig. 3.4 shows the AR surfaces obtained from validation considering 1,000 runs for the (F) scenario with JSWD in both 8 and 3-class schemes (Figs. 3.4a and 3.4b, respectively). Regardless the scenario, C did not show performance differences for values greater than 100, whereas th provided a progressive decrease in AR for values greater than 0.125 mV.



(a) (b)

Figure 3.4: Influence of set of hyperparameters $\boldsymbol{p} = \{th, C\}$ on the (F) scenario with JSWD. Validation AR median values considering 1,000 runs in both (a) 8 and (b) 3-class schemes.

Regarding the evaluation of AL heuristics, the labeled $\mathcal{X}$ dataset was initialized by randomly selecting from the training set 1% of the episodes, considering the remaining 99% of them as the unlabeled $\mathcal{U}$ set. For each heuristic and ϵ iteration, a 1% of episodes of the training set were selected from $\mathcal{U}$ as candidates, i.e. $\mathcal{S}^{\epsilon}$, and added to $\mathcal{X}$. Thus, 100 iterations were computed for each heuristic and run. Besides, aiming to analyze the PD, the percentage of patients who included at least one episode in $\mathcal{X}$ was also evaluated. For comparison purposes, two bounds on the classification performance were defined: (1) an upper bound considering a classifier trained on the whole training dataset, i.e., $\mathcal{X} \cup \mathcal{U}$; and (2) a lower bound assessing a classifier in which the $\mathcal{S}^{\epsilon}$ candidates were randomly selected at each iteration, the **R**andom **S**election (RS) heuristic.

3.3 Results

3.3.1 Classification performance

Performance results achieved by our methodology are presented in Table 3.4. Each cell refers to a specific classification scenario and expresses the median value of test AR (first value, in %) and κ (second value) obtained considering 1,000 runs. Specifically, the 7 classification scenarios for the 3 CSMs in both 8 and 3-class schemes were evaluated. Columns are associated with the incorporation of blocks in the proposed methodology, whereas rows refer to the 3 CSMs. The column on the right of the double dashed line is associated to the only use of marker information. The best performance for each classification problem is denoted in bold.

Scenario	(O)	(E)	(D)	(T)	(S)	(F)	(M)
NDS	67.90	68.86	70.72	71.19	73.26	76.79	71.41
	0.44	0.47	0.49	0.50	0.54	0.60	0.50
JSD	67.26	68.10	70.11	71.01	73.17	77.22	72.86
	0.44	0.46	0.48	0.50	0.54	0.61	0.53
JSWD	67.13	68.14	70.07	71.51	73.75	**77.71**	71.54
	0.44	0.45	0.48	0.51	0.55	**0.62**	0.51

(a)

Scenario	(O)	(E)	(D)	(T)	(S)	(F)	(M)
NDS	84.96	85.82	86.06	86.05	87.24	88.78	82.66
	0.68	0.70	0.70	0.70	0.73	0.76	0.63
JSD	84.97	85.95	86.12	87.08	88.54	89.73	84.65
	0.68	0.70	0.70	0.72	0.76	0.77	0.67
JSWD	85.04	85.94	86.16	87.18	88.55	**89.74**	83.83
	0.68	0.70	0.70	0.73	0.76	**0.78**	0.65

(b)

Table 3.4: Each cell contains test AR (first, in %) and κ (second) median values considering 1,000 runs in both 8 (a) and 3-class (b) schemes. Results for the three CSMs (rows) and the 7 classification scenarios (columns) are presented to assess the influence of each CSM and block. Best performances are in bold.

A relevant result from Table 3.4 is the increase in performance as each block is incorporated. AR performance comparison between (O) and (F) scenarios showed that (F) outperformed (O) about 10 percentage points for the 8-class scheme and 5 percentage points for the 3-class scheme. Note also that this is also supported by an increase in κ for the (F) scenario: values greater than 0.5 suggest an evident relationship between the classification carried out by the expert committee (our gold standard) and our methodology. Besides, these results are more consistent for the 3-class scheme, since κ outperformed the 0.5 value in all scenarios. This can be justified because, when our methodology fails in the 8-class scheme, it mainly fails among different subcategories of the same "Atrial" or "Ventricular" category. However, the latter does not occur for subcategories of the "Other" category, since they can be more similar to other chaotic episodes as AF or VF.

It is also remarkable the robustness of the methodology to the imbalance among categories, which is far from being by chance. When AR performance for (F) is compared with the baseline accuracy (59.23% for 8-class and 63.45% for 3-class schemes), the baseline is exceeded about 18 and 26 percentage points, respectively. Besides, the classification capability of marker sequences is also notable. Note that for the (M) scenario, results even outperformed those in the (O) scenario for the 8-class scheme, which evidences the importance of markers and their intrinsic relationship with the expert committee classification.

In order to provide greater detail in results for imbalanced schemes, performance can be scrutinized for each specific category by means of both sensitivity and PPV. Table 3.5 shows both values for each category, as well as the median number of episodes used for their computation on the test set. Values in Table 3.5 refer to the (F) scenario with JSWD. Note that, in the 3-class scheme, performance is remarkably high for "Atrial" and "Ventricular" categories since sensitivities about 84% and 93% and PPVs about 86% and 90% were achieved. These figures of merit are lower for the "Other" category (sensitivity about 22%, PPV about 45%). This clearly evidences the capabilities of the proposed methodology to discriminate between atrial and ventricular episodes, and the difficulties for classifying episodes in the "Other" category. The main reason of this comes from the fact that there is only a median of 11 "Other" episodes on the test set. These results can be extrapolated to the 8-class scheme. Note again that higher performance is mainly achieved for "Atrial" and "Ventricular" subcategories such as ST-AF and SMVT. This confirms that, as previously mentioned, when the methodology fails in the classification of atrial or ventricular episodes, it mainly fails among subcategories of atrial or ventricular ones. Our approach could be useful in the clinical practice as a tool for providing a preliminary classification of episodes supporting cardiologists.

Category	**ST**	**AF**	**SVT**	**UST**	**SMVT**	**SPVT-VF**	**TWO**	**NS**
Sensitivity (%)	71.59	63.08	17.65	11.11	92.08	60.00	0	40.00
PPV (%)	64.63	69.57	25.00	19.44	85.35	76.67	0	100
Median number of episodes	148	146	28	27	717	39	4	3

(a)

Category	**Atrial**	**Ventricular**	**Other**
Sensitivity (%)	84.41	92.57	22.22
PPV (%)	86.35	90.24	45.00
Median number of episodes	375	734	11

(b)

Table 3.5: Test sensitivity and PPV median values per category considering 1,000 runs in both 8 (a) and 3-class (b) schemes. Values were obtained for the (F) Scenario with JSWD.

It is interesting to analyze performance for TWO and NS categories in the 8-class scheme. Median values of sensitivity/PPV are 0/0% for TWO and 40/100% for NS. The poor performance for TWO is totally expected because it is a specific case where the ICD algorithm fails, detecting T-waves as R-waves. Hence, ICD sometimes records episodes which waveform and frequency are quite similar to a normal sinus rhythm or ST. In fact, classification for these specific episodes was ST in most cases. Note also that the median number of episodes on the test set for this category is very low. Regarding performance in the NS category, results are explained by the similarity of NS to other chaotic arrhythmic episodes such as AF or VF. However, despite this and the clear imbalance, the proposed methodology provides 40% of sensitivity and 100% of PPV. This shows a clear potential of our approach for classifying NS episodes, although the median number of episodes on test set is very low and these results must be taken with caution.

With respect to the use of CSMs, results showed no significant differences among NDS, JSD and JSWD. For more detail, Fig. 3.5 represents the box-plots for AR and κ on the 1,000 runs used to obtain values of Table 3.4. Mesh, white, and gray boxes are associated with NDS, JSD and JSWD measures, respectively. Note again the mentioned increase in performance as each block is incorporated, and how the confidence intervals overlap in every scenario. A more in-depth analysis shows a slightly better median performance for JSD and JSWD in comparison

to NDS, excepting (O), (E) and (D) scenarios in the 8-class scheme (i.e., before thresholding application). To explain this, the number of support words per episode and the set of words used to normalize the three CSMs (cardinality in denominator of Eqs. (3.4), (3.5), and (3.6)) have been scrutinized in Fig. 3.6.

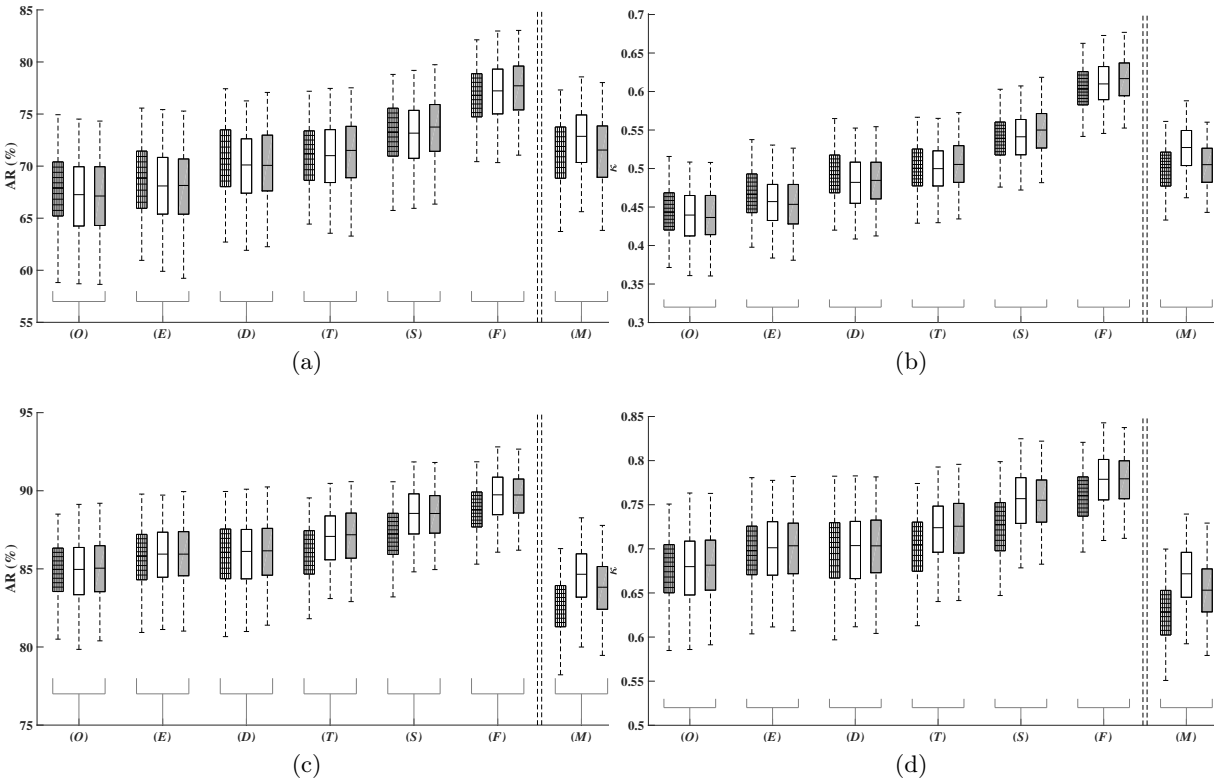


Figure 3.5: Box-plots of test AR ((a) and (c)) and κ ((b) and (d)) values considering 1,000 runs in both 8 and 3-class schemes (upper and down panels, respectively). The 7 classification scenarios are represented for NDS (mesh), JSD (white) and JSWD (gray solid). Center lines, boxes and dashed lines denote medians, IQRs, and [2.5%, 97.5%] ranges, respectively.

On the one hand, the box-plots for the number of support words in the 7 classification scenarios are represented in Fig. 3.6a. Note that after applying the thresholding block (from (T) scenario), the number of support words decreases considerably but, conversely, performance increases. These results suggest that support words have a higher classification capability when thresholding is applied. Besides, this capability can be scrutinized by categories. Figs. 3.6b and 3.6c represent the number of support words per category in both 8 and 3-class schemes for (O) and (F) scenarios (white and grey boxes, respectively). Note that, in the 8-class scheme, there are no significant differences among categories, which is probably due to the great imbalance in minority categories. However, in the 3-class scheme, it is evidenced a difference between the "Other" and "Atrial"/"Ventricular" categories. For the "Other" category, a greater number of support words is necessary to characterize episodes. The reason of this comes specifically from the episodes included in this category (TWO and NS). As previously mentioned, TWO and NS are quite varied categories since they come from ICD failures or have a more chaotic frequency and waveform.

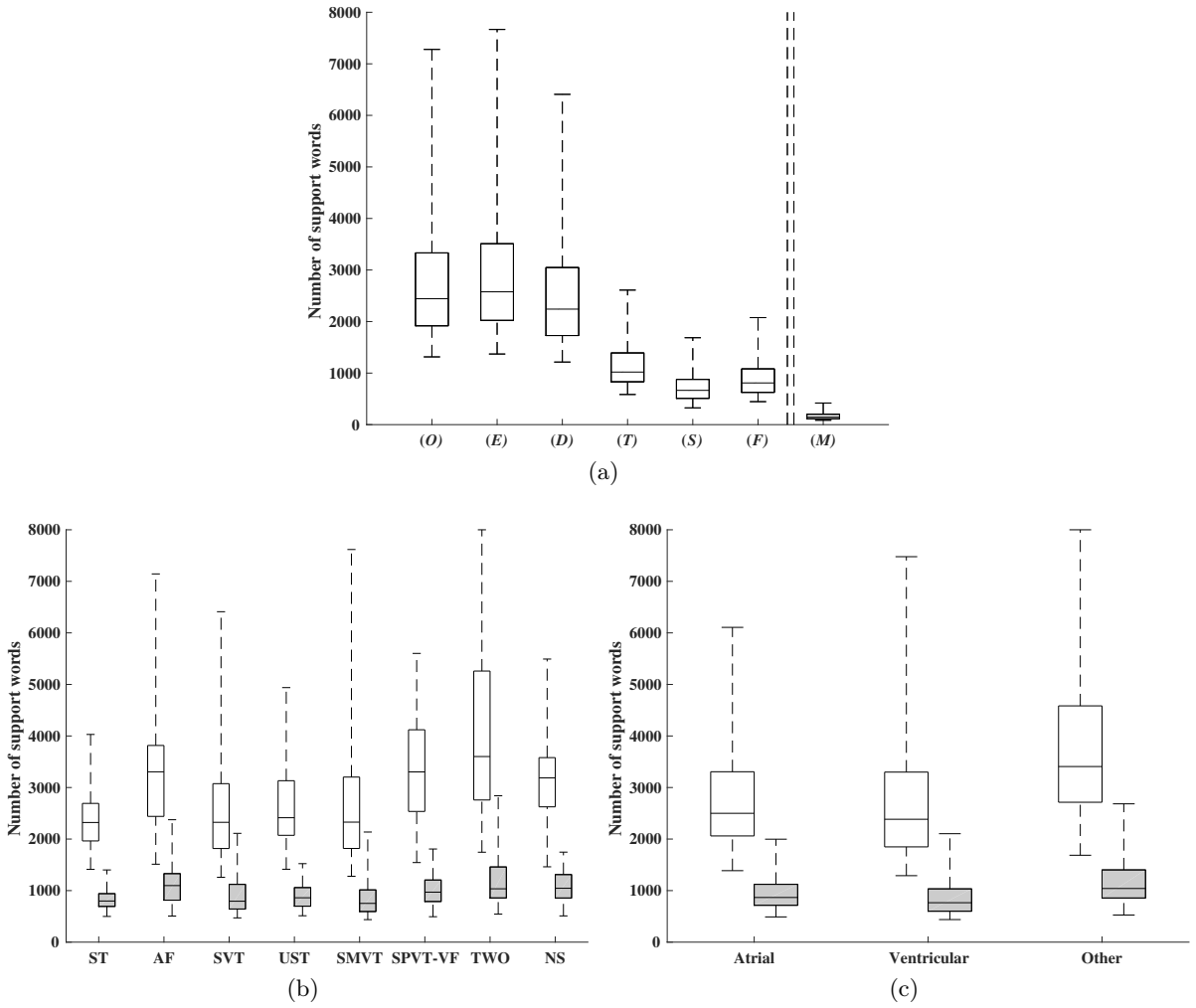


Figure 3.6: Box-plots of the number of support words considering the 6,233 SCOOP arrhythmic episodes in (a) the 7 classification scenarios, and in categories of both (b) 8 and (c) 3-class schemes for the (O) and (F) scenarios (white and grey boxes, respectively). Center lines, boxes and dashed lines denote medians, IQRs, and [2.5%, 97.5%] ranges, respectively.

On the other hand, the normalization of JSD and JSWD is carried out using the union set of support words, whereas the maximum cardinality of each individual set of support words is used for NDS. When both normalizations are compared, the use of the maximum cardinality (as in NDS) involves considering a lower number of support words in the similarity estimation. As support words after thresholding have a higher classification capability, considering a lower number of support words has a negative effect on the similarity estimation among episodes. Thus, a better performance is achieved for JSD and JSWD when thresholding is applied.

In this work, results have been obtained using arrhythmic episodes recorded by Medtronic® ICDs. Currently, these devices are equipped with the **W**avelet **A**lgorithm (WA) [10], a classification algorithm based on the EGM waveform differentiating between supraventricular and ventricular arrhythmic episodes. However, the activation of this algorithm in each ICD depends on clinical criteria and it only provides a label in an episode when it is active. Hence, a comparison between the WA and **O**ur **p**roposed **M**ethodology (OM) can be carried out using those episodes with a WA label in a new 2-class scheme (supra and ventricular categories). Specifically, a total of 3,610 episodes from 361 patients in our SCOOP database were useful for this comparison. In Fig. 3.7, box-plots for AR and κ with the 1,000 runs using the 70/30 patient

resampling assessment are represented considering as a gold standard the labels from SCOOP expert committee. Note also that in this assessment, WA labels were collected for each test set, whereas for OM, test episodes were labeled using the SVM designed from the training set. Median values of 77.6% and 90.2% for AR and 0.47 and 0.78 for κ were achieved by the WA and OM, respectively. This result evidences the classification capabilities of our proposal against those algorithms implemented in current ICDs.

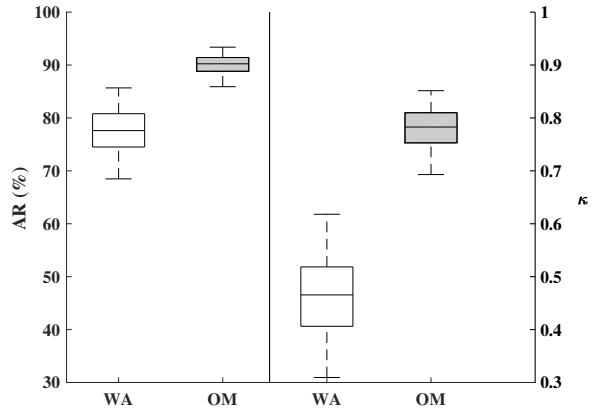


Figure 3.7: Box-plots of test AR (left) and κ (right) values achieved by OM (white box) and the WA (gray box) considering 1000 runs in the 2-class scheme (supra and ventricular) defined for the comparison. Center lines, boxes, and dashed lines denote medians, IQRs, and [2.5%, 97.5%] ranges, respectively.

3.3.2 Active learning evolution

Performance and comparisons among AL heuristics are presented in Fig. 3.8. A total of 5 heuristics (4 based on large-margin and 1 associated to RS) were evaluated in both 8 and 3-class schemes. In Figs. 3.8a-3.8d, the median value of performance (AR and κ) for the lower (RS heuristic) and upper (gold standard) bounds are represented by means of a red curve and a horizontal black line, respectively. The 4 curves between both bounds (gray zone) are associated with the median values of performance achieved by the four large-margin based AL heuristics: MS (yellow), BT (purple), MS-PD (blue), and BT-PD (green).

As expected, performances of MS, BT, MS-PD, and BT-PD were significantly better than those obtain by RS. As samples were selected considering their uncertainty, the gold standard was quickly reached when approximately 50/20% (8/3-class schemes) of episodes in the training set have been included. These results are relevant because they indicate us that a high volume of episodes contain either redundant or non-relevant information for the classification. In general, BT heuristics performed better than MS ones, because they reached high performance using a lower number of episodes. However, if the percentage of episodes included in the model is analyzed when the AL heuristics reach the gold standard, differences are not significant. In Fig. 3.8, this percentage is represented by dashed vertical black lines. Specifically, when the 46/22% of episodes from the training set have been included in the model, the four heuristics reached the gold standard. Predictably, the higher the number of categories, the higher the number of episodes needed with relevant information. In this setting, models more effective and also computationally faster can be designed.

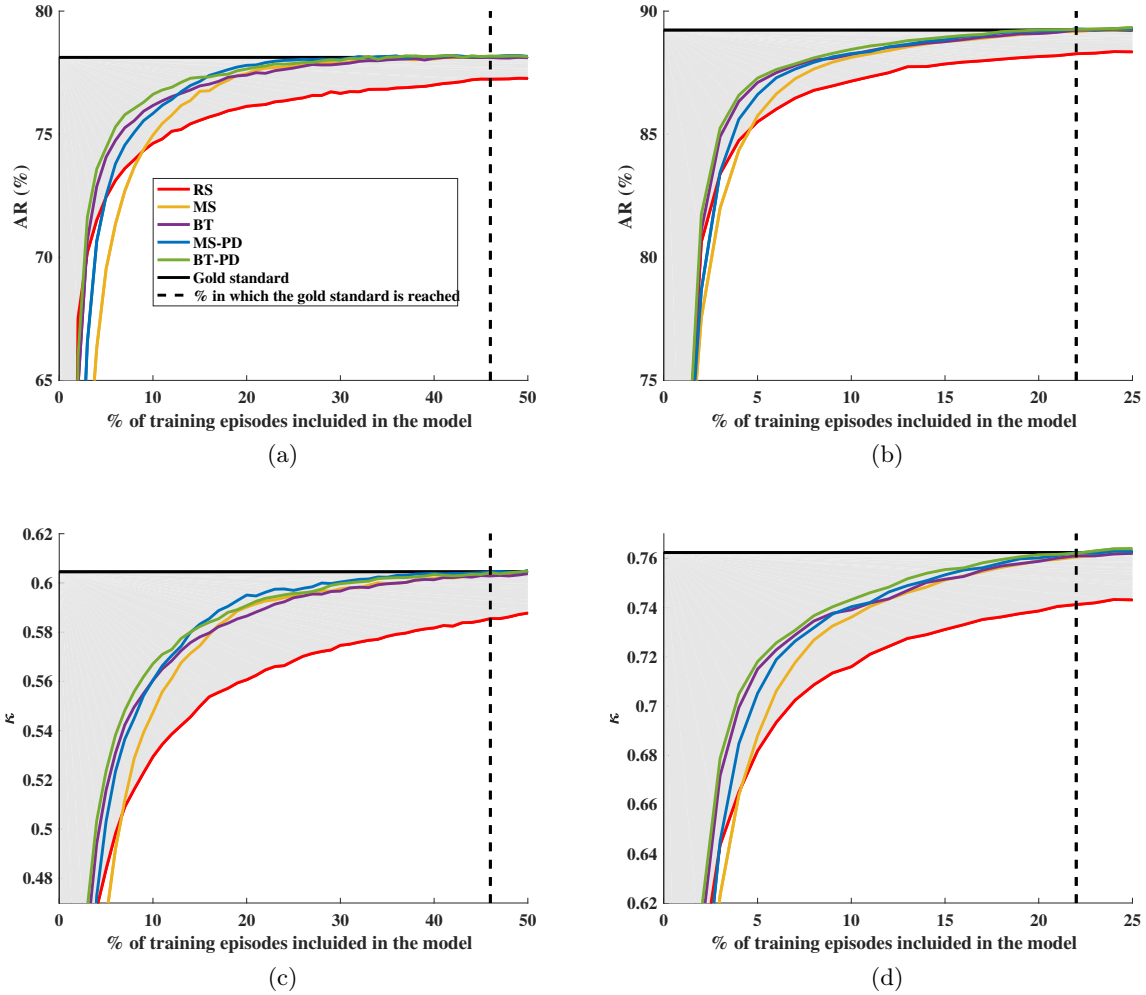


Figure 3.8: Evolution of the test AR ((a) and (b)) and κ ((c) and (d)) median values considering 1,000 runs when the proposed methodology is AL-streamlined in both 8 ((a) and (c)) and 3-class ((b) and (d)) schemes. Horizontal Black lines denote the gold standard. Dashed vertical black lines indicate the percentage for which the gold standard is reached by the four AL heuristics different from RS (red curves).

With respect to PD, it was analyzed by evaluating the percentage of patients who included in the model at least one episode per iteration. This evolution is represented in Figs. 3.9a and 3.9b for both 8 and 3-class schemes, respectively. Note that evolution is nonlinearity correlated with the inclusion of episodes in the model for all heuristics (even for the RS one). The reason of this effect is twofold: (1) patients have not associated the same number of episodes; and (2) the category imbalance. Moreover, there is a clear difference among the 4 large-margin heuristics and the RS one. As expected, the MS-PD and BT-PD heuristics include more patients in their classifiers because they have a PD constraint. However, differences with the other two heuristics (MS and BT) are not significant (close to 1-2%), which indicates that uncertainty is not mainly determined by the type of episode but by the patient. Specifically, when approx 85/60% of patients (8/3-class schemes, respectively) from the training set are included in the model, gold standard performance is reached. This result is very useful for systems as SCOOP, since the manual labeling could be optimized towards the search of episodes from different patients. This even would avoid the labeling of non-representative episodes.

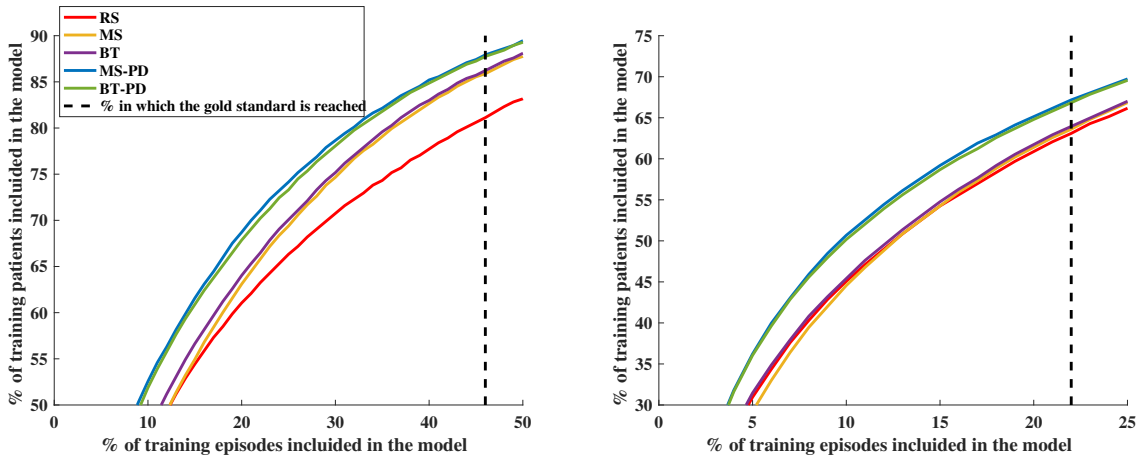


Figure 3.9: Evolution of the percentage of patients included in the model for each heuristic in both 8 (a) and 3-class (b) schemes. Dashed vertical black lines indicate the percentage for which the gold standard is reached by the four AL heuristics different from RS (red curves).

A relevant point of the AL is how performance in each category evolves along training. In Fig. 3.10, this evolution is shown for the BT-PD heuristic (similar results were obtained for the rest of heuristics excepting RS) in both 8 and 3-class schemes. Note that for the 3-class scheme, when 5% of episodes are included in the model, the maximum learning for the "Atrial" and "Ventricular" categories is practically achieved. However, for the "Other" category, there is a decrease and a subsequent continuous increase in performance until approx 20% of episodes are included. This value matches with the minimum number of episodes (dashed black line) required for reaching the gold standard. This result suggests us that: (1) the percentage of relevant episodes is lower than the that expected according to Fig. 3.8 probably because of imbalance; (2) the minority category largely condition the learning. This conclusion can be also be extended for the 8-class scheme, where imbalance is more accentuated. Specifically, SPVT-VF and NS categories condition the learning, and in a less degree SVT, UST and TWO categories. These results suggest that AL should be oriented towards the labeling of arrhythmic episodes belonging to minority categories, what would optimize even more the model learning process.

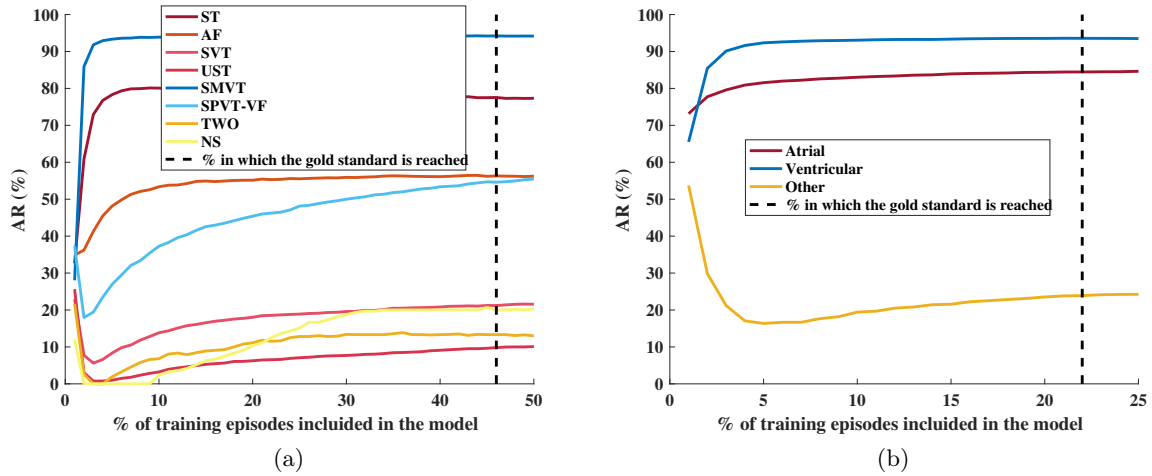


Figure 3.10: Evolution of the test AR median values considering 1,000 runs when the proposed methodology is AL-streamlined in both 8 (a) and 3-class (b) schemes. BT-PD heuristic was used as for the AL streamlining. Dashed vertical black lines denote the gold standard performance value represented in Fig. 3.8.

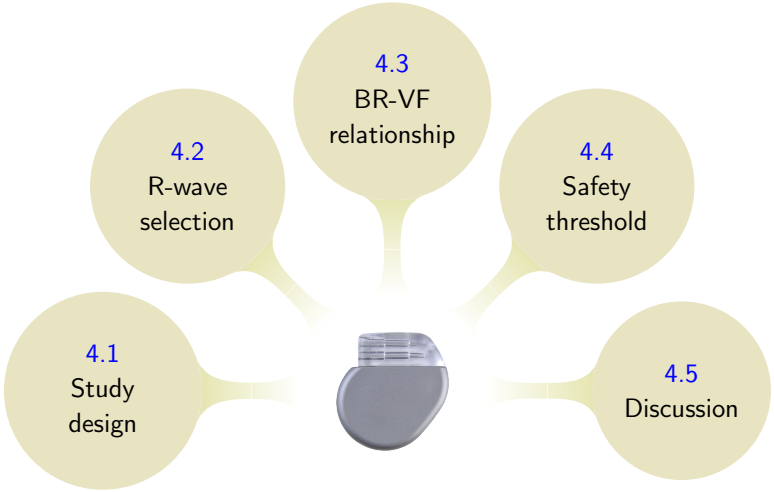
3.4 Discussion

The classification of arrhythmic episodes remains an active research field because the improvement of specificity for supra-ventricular episodes and the discrimination of noise are an open issue in current ICDs. The backbone of the well-known ICD classification criteria has been the analysis of **C**ycle **L**ength (CL) intervals, for which algorithms for the automatic detection of heartbeats are continuously proposed [69–71]. However, the classification using CL intervals is limited and needs to be complemented with other features, such as those extracted from signal waveform. By using the potential of dual-lead ICD configurations, classification algorithms based on signal waveform have been proposed to improve the performance (including some of them in today ICDs). In these algorithms, features such as peak amplitudes, polarities, number and order of peaks of ventricular complexes, or even wavelet transforms are analyzed [10]. However, results do not seem to have improved the classification performance [72].

The work here presented proposes a new methodology for classifying ICD arrhythmic episodes with a high classification performance. The proposed methodology requires minimal EGM preprocessing, deals with episodes of different duration with blank intervals, and simultaneously considers heart activation events, EGM waveform features, and ICD marker information. In the clinical practice, two scenarios could potentially benefit from our proposal. Firstly, the follow-up of most patients treated with and ICD is currently performed via a remote monitoring system. When an arrhythmic episode is detected by the ICD, a warning is sent to the cardiologist in order to confirm the diagnosis and allow a quick therapeutic intervention, if needed. Our methodology could help in the search of those episodes requiring an immediate action on the patient and give them a higher priority for the cardiologist consideration. Specifically, cardiologists could be provided with a preliminary classification of Atrial/ventricular episodes (scenario for which our approach works considerably well) for supporting diagnoses. However, the classification of specific cases such as SPVT-VF (which are critical episodes) is still open for improvement, since sensitivities/PPV close to 60/75% may be not acceptable in the clinical practice. Secondly, in an environment with continuous data growing such as the SCOOP platform (which in the near future is expected to contain more than 60,000 arrhythmic episodes), the analysis and labeling of the arrhythmic episodes by an experts' committee is an increasingly cumbersome task. AL results have shown how our approach could help for selecting the most relevant episodes aiming at reducing the massive checking of episodes by cardiologists. In fact, our results evidenced that more than a half of all episodes included in SCOOP are not relevant for the classification.

The methodology proposed in this work could be extrapolated to other ICD manufacturers or even to signals coming from electrocardiogram or heart sound recordings [26, 73, 74]. Eventually, though no classify entire episodes, another statistical technique as deep learning is beginning to be applied in the cardiology field [75, 76]. In this regard, the adaptation of our methodology to be applied with a deep learning scheme would open the road towards research on new classification algorithms, aiming to include them in future ICDs.

Chapter 4



Safety Threshold on R-Wave Amplitudes

NOWADAYS, a safety threshold on R-wave amplitudes during a normal BR ensuring a low probability of undersensing fatal arrhythmic episodes in ICDs has not been established in clinical guidelines yet. In this chapter, a procedure for overcoming this issue from SCOOP data is proposed. Firstly, the design criteria for the study and the data preprocessing are presented. Next, the process followed to select the R-waves useful to carry out the study is detailed. DA techniques used to set the safety threshold are then described. Finally, obtained results are detailed together with a summarized discussion. The value herein obtained here as the safety threshold can be highly useful to define new clinical recommendations to delay or even avoid harmful therapies.

4.1 Study design and data preprocessing

Currently, the most common indicator used by cardiologists for determining a possible undersensing of VF episodes in ICDs is the R-wave amplitude during a normal BR. Together with its associated **R**emote **M**onitoring **T**ransmission (RMT)[1], this value is periodically provided to cardiologists by ICDs during the patient's follow up. Current clinical guidelines recommend at least an amplitude of 7 mV at the ICD implantation [12, 13]. However, neither the relationship between the BR and VF R-wave amplitudes, nor the probability of undersensing R-waves during a VF episode, have been elucidated yet. Both issues are the key to set a minimum value on BR R-wave amplitudes ensuring a low undersensing risk of VF episodes in ICDs. Therefore, the objective of this study is twofold:

1. to determine the BR-VF R-wave amplitude relationship;
2. to set a safety threshold on BR R-wave amplitudes.

[1] An EGM record of 10 seconds duration.

In order to reach these goals, a total of 11,780 arrhythmic episodes were extracted from SCOOP. Episodes had been recorded from August 2011 to September 2014, and they came from 2,507 patients with an average follow-up of 2.7±2.6 years (mean±standard deviation). Episodes classified as SPVT/VF by the SCOOP committee were selected from this set and then reviewed by 2 independent experts, in order to choose those ones including at least one VF segment (a third expert took the final decision in case of disagreement). In this review process, VF was defined as a rapid[2], grossly irregular ventricular rhythm with variable morphology and marked fragmentation in the *VtipToVring* EGM. VF episodes were excluded when they were associated with defibrillation testing and mono/poli-morphic ventricular tachycardias preceding VF segments.

Then, a data preprocessing stage was designed to fairly compare the BR and VF R-wave amplitudes, and to define a safety threshold according to the ICD restrictions. On the one hand, and due to the low memory capabilities, ICDs do not store its original recorded EGM (which is used by their implemented algorithms), but instead another signal with lower dynamic range (±8 mV). This fact conditions the extraction of the actual R-wave amplitudes, because when the amplitude in the original EGM is very high, clipped segments are present in the stored EGM. On the other hand, and according to the guidelines provided by the Medtronic® team, ICD algorithms use the *VtipToVring* EGM to detect fatal arrhythmic episodes, such as VF. To do this, ICDs process the *VtipToVring* EGM, applying a first order Butterworth band-pass filter [77] with cut-off frequencies of 15 and 35 Hz and a subsequent positive rectification (negative values are set to positive). The resulting EGM is used to determine the time position of each R-waves by applying an algorithm of auto-adjusting an exponentially decaying threshold [78]. This algorithm sets an ICD sensitivity threshold[3] of 0.3 mV (on the processed EGM) to avoid considering electrical interferences as cardiac activity. As later described in Section 4.4, this value will be key in the setting of a safety threshold. Once R-waves are detected, a moving-window that includes the last 24 CLs (or 32, depending on the ICD programming) is applied to detect VF episodes, provided that the 25% of CLs in the window (18 or 24) are below a programmed threshold (typically $\leq$ 250 mS). That is why undersensing more than of 25% of R-waves during a VF episode might lead to a high risk of non/late detection of fatal arrhythmic episodes.

Taking the previous information into account, the VF episodes selected from SCOOP were processed as follows. Firstly, clipped segments were automatically detected and their values in the original recorded EGM were estimated. To this end, an interpolation process was proposed. Specifically, a new point (crossing point) was obtained in each clipped segment from crossing two lines: the first one was obtained from the two previous-segment points; an the second one was obtained from the two post-segment points. Then, the first/second line and the crossing point were used to interpolate values of the EGM between the starting/final-segment and the crossing position. Due to the non-availability of the original EGM, interpolation performance could not be adequately evaluated. However, as presented in Section 4.2, this performance was suitable at least for R-wave amplitude values. An example of a resulting EGM from this estimation process is represented in Fig. 4.1.

Secondly, and by using a custom-made Matlab® [79] tool, VF and BR (prior to the VF) segments were manually selected in the estimated *VtipToVring* EGMs. Besides, when available, the *CanToHVB* EGM was used to tune the selection (see an example in Figs. 4.2a and 4.2b). In case of no available BR segments prior to VF, BR segments were obtained from the nearest-time RMT collected during the patient's follow-up within a ±1-year window (the reason for choosing

[2] CL intervals ≤220 ms during at least part of the recording.
[3] Minimum value (mV) considered by ICD as cardiac electrical activity.

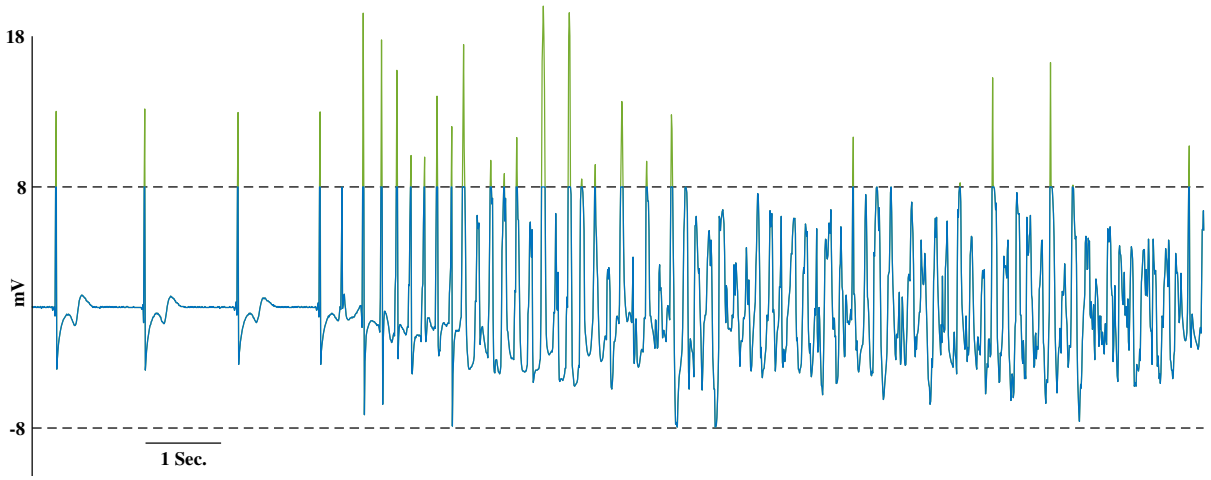


Figure 4.1: Example of a resulting EGM after estimating its values in clipped segments. The ICD dynamic range (±8 mV) is represented as horizontal dashed lines. Blue and green represent recorded and estimated EGM segments, respectively.

this time window is described in Section 4.2). In this BR selection, any **P**remature **V**entricle **C**omplex (PVC) or activation pacing events was excluded aiming to obtain a better estimation of the patient's physiological BR R-wave amplitude. After this process, a total of 229 VF episodes from 83 patients were useful for the study. Then, estimated *VtipToVring* EGMs were processed by replicating the process in the ICD (filtering and rectification) to select the R-waves amplitudes useful for the study. This selection process is next described.

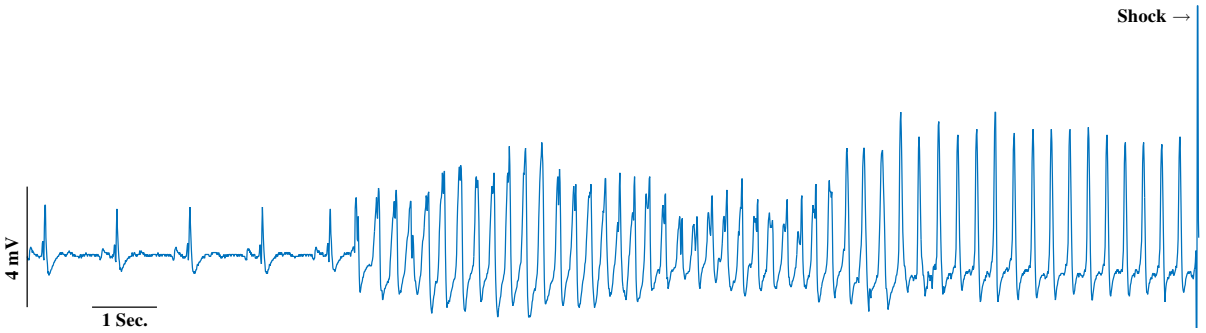
4.2 Selection of the R-wave amplitudes

The most important stage of the study is the correct selection of R-wave amplitudes (as the example shown in Fig. 4.2c) that are useful for setting the safety threshold. Stored ICD makers associated to R-wave detections (black stars in Fig. 4.2c) were extracted to automatically identify R-waves amplitudes. Manual inspection by two independent experts (the same ones mentioned in Section 4.1) ensured:

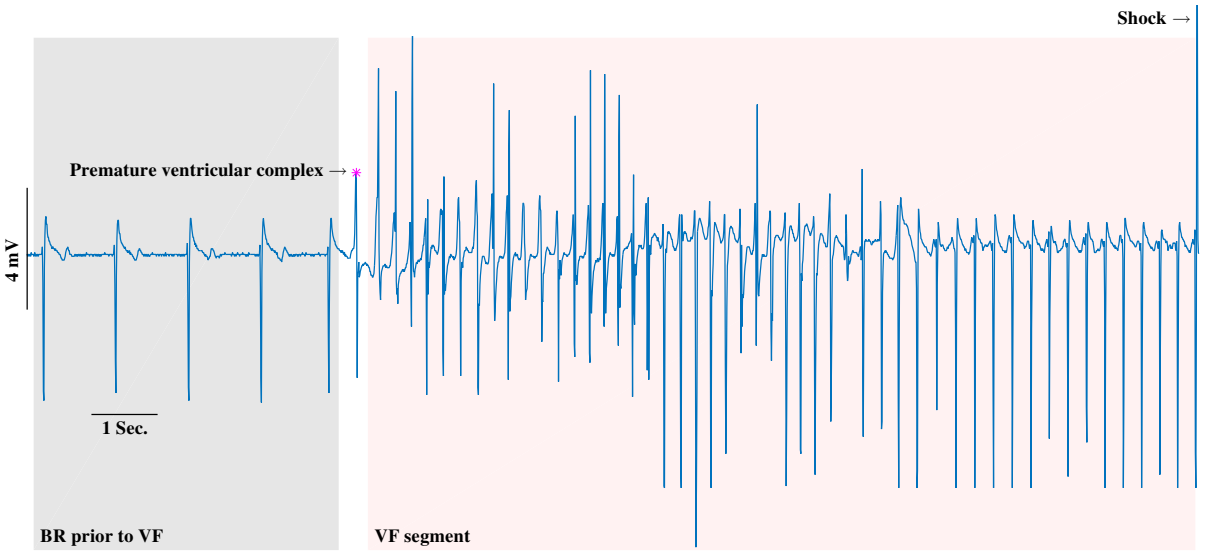
1. an adequate BR and VF R-wave detection (i.e., excluding pacing events and PVCs);
2. a manual selection of R-waves occurring during the *blanking period*[4] at least 80 ms apart from the R-wave previously detected;
3. a manual selection of undersensed R-waves after the blanking period, either above or bellow the ICD sensitivity threshold of 0.3 mV.

Accurate voltage values for each BR or VF R-wave were determined by using an automatic custom-made Matlab® [79] tool that corrected the time position of each maker to the nearest peak position within an 80 ms time-window (green points in Fig. 4.2c). The length of this window was set according to previous studies [80–82], and it was crucial during VF fragmented signals to properly mark undersensed and blanked R-waves. Any other lower amplitude peak within the window was considered as signal fragmentation.

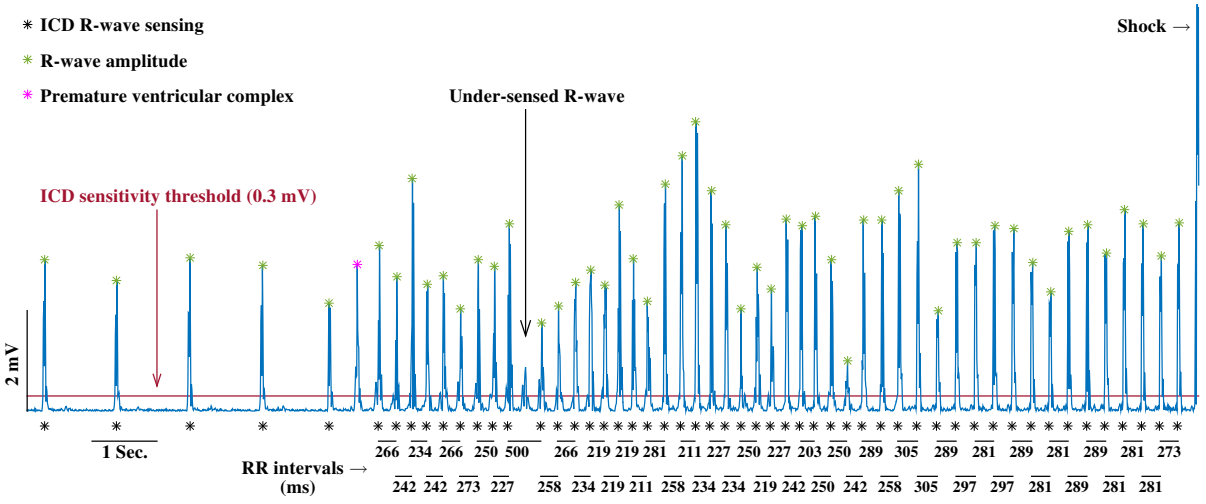
[4]Time duration after a detected R-wave in which ICD does not detect another peak, fixed in 120 ms.



(a)



(b)



(c)

Figure 4.2: Selection process of BR and VF segments in an actual SCOOP arrhythmic episode: (a) *CanToHVB* EGM; (b) *VtipToVring* EGM; and (c) filtered and rectified signal from (b), used for comparisons between BR and VF R-wave amplitudes. In (c), the ICD sensitivity threshold and an example of undersensed VF R-wave are represented.

Then, a reference BR R-wave amplitude was associated with each VF episode aiming at determining the BR-VF R-wave amplitude relationship. Each reference value was computed as the median of the BR R-wave amplitudes in the filtered and rectified EGM, either from the segment just prior to the VF, or from the nearest RMT within a ±1-year window. The limit of this time window was set after analyzing the patient's physiological variability of the reference BR R-wave amplitude in different time windows. Specifically, the reference BR R-wave amplitude variability was evaluated in a 6-month, 1-year, and >1-year time-windows. Only 17 patients were used in this analysis because both BR R-wave amplitudes prior to VF initiation and RMTs in the three time-windows were simultaneously available. For each window and patient, R-waves amplitudes were compared with the corresponding median value in the window. Differences (in %) are plotted in Fig. 4.3 for each time-window. Note that the larger the time-window, the greater the number of amplitudes (n) and variability, since more RMT per patient are available. Besides, for the 6-month and 1-year time-windows, the vast majority of amplitudes felt within a ±30% of difference. This value is acceptable since ICDs can provide a BR R-wave amplitude difference of 30% in their measures, according to guidelines provided by the Medtronic® team. This was also reported in [14]. Hence, a 1-year time window was used as a limit to associate a reference BR R-wave amplitude with the VF episodes without a BR prior to the VF initiation.

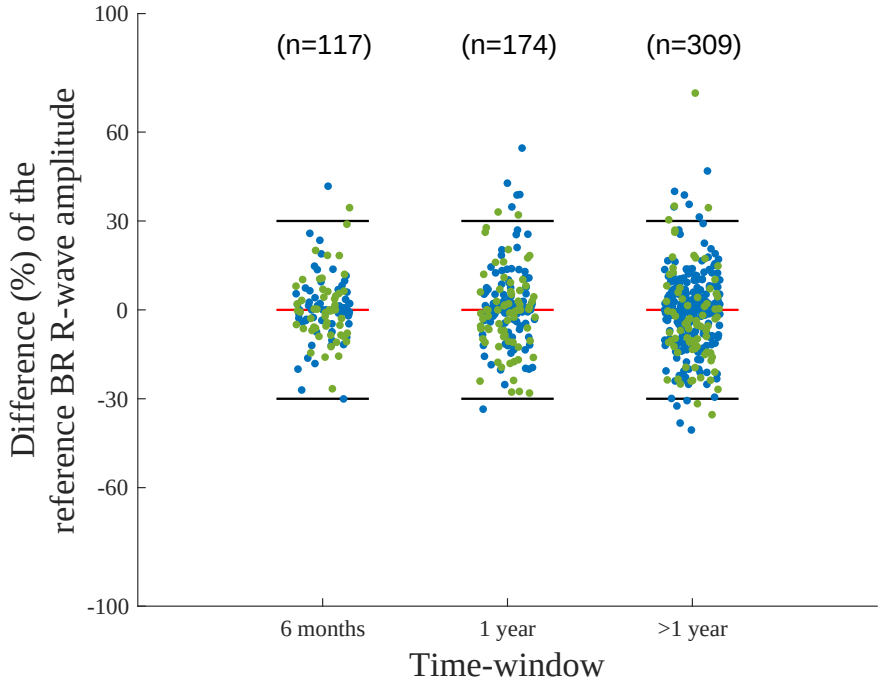
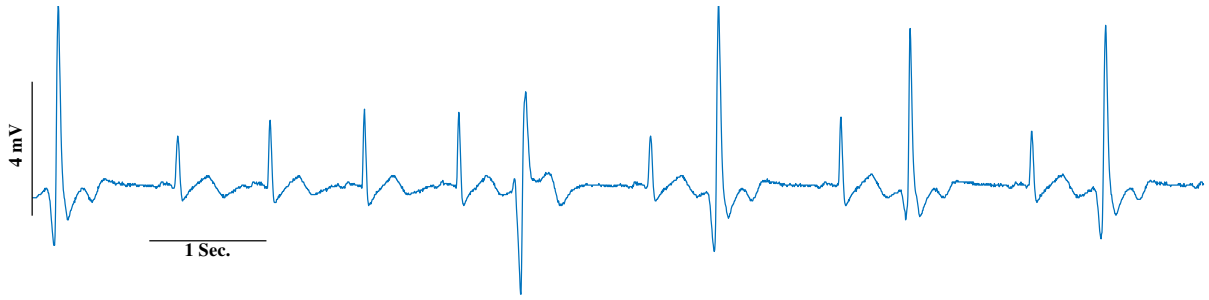


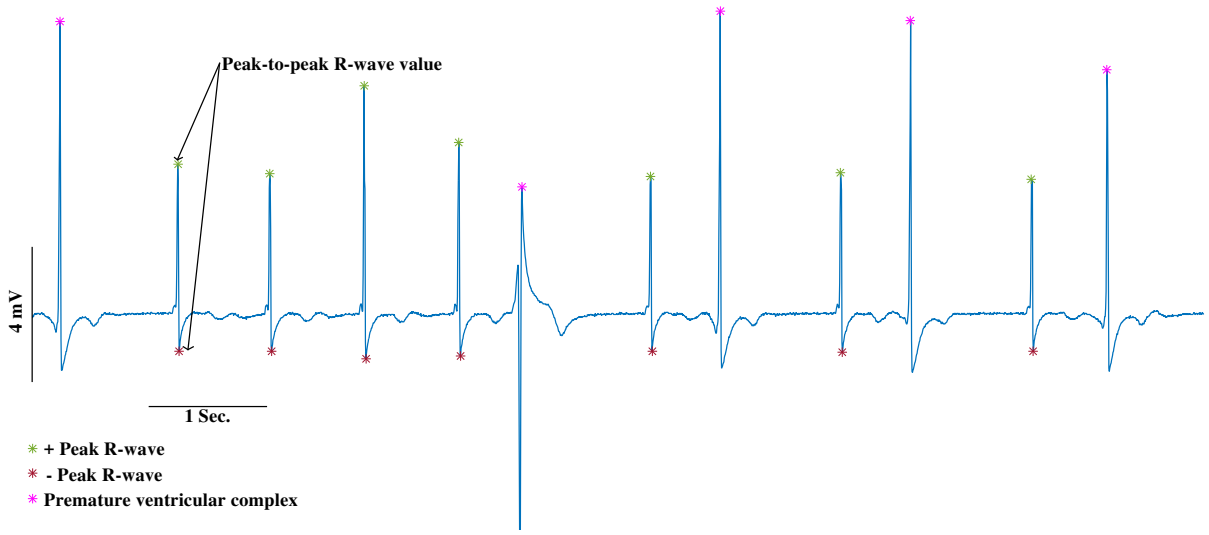
Figure 4.3: Variability of the difference between the reference BR R-wave amplitude and the median value of the patient in different time-windows. Blue and green dots are associated with reference values computed from RMT or segments prior to VF, respectively. Red and black lines denote difference of 0% and ±30%, respectively.

As mentioned in Section 4.1, during the patient's follow up, ICDs periodically provide cardiologists with RMTs associated to a BR R-wave amplitude (the clinical BR R-wave amplitude). These amplitudes are the measures used by cardiologists to intuitively determine the quality of the ICD working. In this setting, the definition of a safety threshold on this clinical BR R-wave amplitude could be highly useful.

Unlike the reference BR R-wave amplitude, the clinical one is directly computed by the ICD from its original EGM (i.e., no clipping is present). Specifically, the ICD computes a median value of the peak-to-peak no-pacing R-wave amplitudes in the *VtipToVring* EGM of a RMT (see example in Fig. 4.4). Note also that this value is computed from a non-clipped, filtered, and rectified EGM. This involves that a conversion process to obtain the clinical BR R-wave amplitude from the reference one is required to define the BR-VF R-wave relationship in a clinical way. To achieve this, we firstly confirmed that peak-to-peak amplitudes provided by ICD were similar to those computed from their associated RMT after estimating their values of clipped segments (detailed in Section 4.1).



(a)



(b)

Figure 4.4: Example of a RMT. (a) *CanToHVB* EGM. (b) *VtipToVring* EGM used for estimating the clinical BR R-wave amplitude from peak-to-peak R-wave values (amplitude difference between green and red points). PVC were excluded.

Specifically, a set of 290 RMTs and ICD clinical R-wave amplitudes were used. For each RMT, a clinical BR R-wave amplitude was computed in the same way that the ICD does. However, in contrast to ICDs, PVC were manually excluded using (when available) the *CanToHVB* EGM. On the one hand, in Fig. 4.5a is represented the scatter plot between the ICD-provided and the computed clinical BR R-wave amplitudes. On the other hand, the scatter plot between the ICD-provided and the reference (computed from the estimated, filtered and rectified EGM) BR R-wave amplitudes is represented in Fig. 4.5b. Note that deviations between ICD-provided and computed values were not greater than 30% in most cases (mean±standard deviation were 19.8%±15.6%). Note also that there is a linear relationship between the ICD-provided and the reference amplitudes. These results confirmed two issues: (1) the estimation of the original EGM

by linear regression in clipped segments is suitable (at least for R-waves amplitudes); and (2) a conversion of the clinical BR R-wave amplitude from the reference one is feasible. Specifically, the linear regression result between the clinical BR R-wave amplitude ($R_{Clinical}$) and the reference one (R_{BR}) was:

$$R_{Clinical} = 2.23 \cdot R_{BR} + 1.38 \tag{4.1}$$

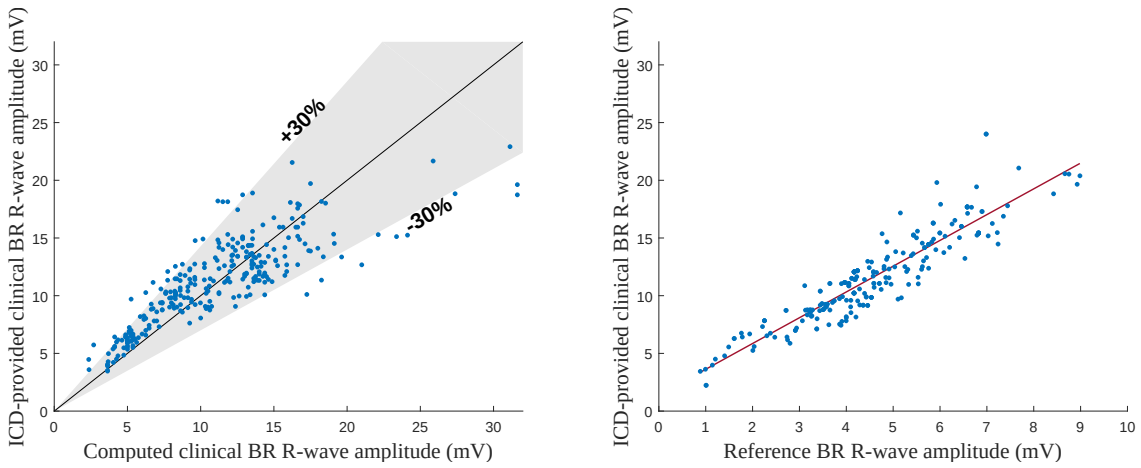


Figure 4.5: Scatter plots of BR R-wave amplitudes: (a) computed clinical vs. ICD-provided clinical; (b) reference vs. ICD-provided clinical. Red line in (b) represents the linear regression between both represented variables.

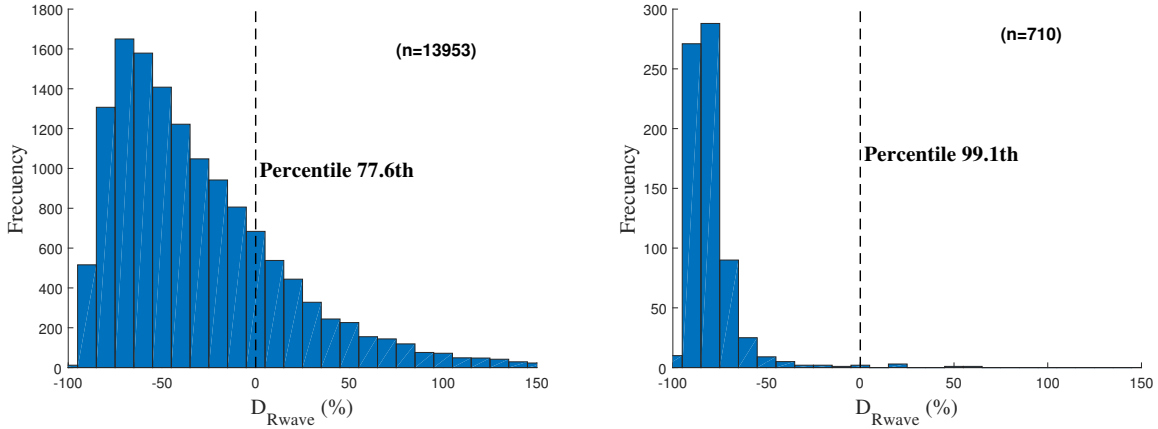
4.3 Relationship between the BR and VF R-wave amplitudes

As previously mentioned in Section 4.1, the setting of a safety threshold on BR R-wave amplitudes requires to determine the relation between the R-wave amplitudes during a normal BR and a VF episode. To do this, a normalized amplitude deviation was firstly computed for all the selected VF R-waves (detected, undersensed, and those occurring during the blanking period), aiming at comparing R-waves coming from different episodes/patients in a relative way. Specifically, for each i-th R-wave amplitude of each e-th VF episode ($R^{e,i}_{VF}$), a normalized amplitude deviation ($D^{e,i}_{Rwave}$) from the reference BR amplitude of the episode (R^{e}_{BR}) was computed as follows:

$$D^{e,i}_{Rwave} = \frac{R^{e,i}_{VF} - R^{e}_{BR}}{R^{e}_{BR}} \tag{4.2}$$

A total of 13,953 D_{Rwave} values were computed from the 229 VF episodes. In Fig. 4.6a, the D_{Rwave} histogram built from equal-size bins[5] of 10% deviation is represented. Note that a 77,6% of values are lower than zero (when $R_{VF} = R_{BR}$). If this D_{Rwave} histogram is just represented for undersensed R-waves (which were manually selected), the percentage of deviations lower than zero is even higher (99,1%, see Fig. 4.6b). Note also the low rate of R-wave undersensing ($\approx 5.1\%$). These D_{Rwave} distribution results confirmed two assumptions: (1) the clear decrease in the R-wave amplitude when a VF episode occurs; and (2) the probability of undersensing a VF R-wave increases with the negativity of the D_{Rwave}. Evidences of both issues were previously reported in [83], however, a BR-VF R-wave amplitude relationship was not elucidated.

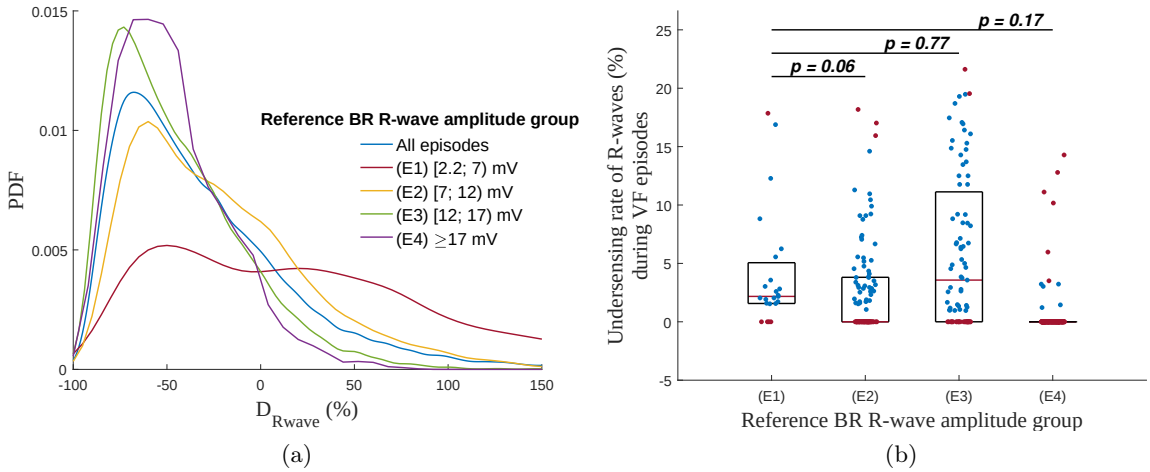
[5]Intervals that divide the entire range of possible values of a variable.



(a) (b)

Figure 4.6: D_{Rwave} histogram for: (a) all VF R-waves; and (b) undersensed VF R-waves. "n" denotes de number of D_{Rwave} values in the histogram.

Despite the variability among episodes and the amplitude normalization, a dependence of the D_{Rwave} distribution on the reference BR R-wave amplitude can be intrinsically assumed. This is why, the lower the reference BR R-wave amplitude, the greater the relative chances of VF R-waves. Therefore, potential differences in the D_{Rwave} distribution were evaluated for episodes with associated reference value ≥7 mV. This value was chosen since it is the amplitude that is recommended at ICD implantation [12, 13, 84]. Firstly, four groups of episodes with a similar reference amplitude at ≈5 mV intervals were defined: (E1) [2.2; 7) mV; (E2) [7; 12) mV; (E3) [12; 17) mV; and (E4) ≥17 mV. Then, for all episodes and each group a D_{Rwave} **P**robability **D**ensity **F**unction (PDF) (represented in Fig. 4.7a) was computed from the D_{Rwave} histogram, by using the **P**arzen **W**indow **E**stimator (PWE) [28], which is described bellow.



(a) (b)

Figure 4.7: (a) D_{Rwave} PDF for all episodes (blue line) and the four groups of reference BR R-wave amplitude. (b) Box-plots including the undersensing rates of episodes of groups. Red lines, black boxes, and red dots denote medians, IQRs, and outliers at least twice the IQR from the median, respectively. Values did not show significant differences (p value > 0.05) among the undersensing rates of groups.

Let us consider a finite set of N D_{Rwave} values, $\mathcal{D} = \{d_i \in \mathbb{R} \,|\, d_i \geq -1\}_{i=1}^{N}$. The PWE places a kernel function $\mathcal{K}$ on each d_i and then sums all the kernels considering a scaling factor. Specifically, the expression for estimating the D_{Rwave} PDF from $\mathcal{D}$ is:

$$PDF_{D_{Rwave}}(d_i) = \frac{1}{N}\sum_{j=1}^{N}\mathcal{K}(d_i, d_j) \tag{4.3}$$

As mentioned in Subsection 3.1.3 of Chapter 3, there are several common kernels used in the literature. In this study, a radial basis function kernel $exp\left(\frac{-||\mathbf{x}_i-\mathbf{x}_j||^2}{2\sigma^2}\right)$ was used and Silverman's rule [85] was followed to fix the σ parameter. Then, to estimate the D_{Rwave} PDF of each group, only D_{Rwave} values associated of each group were used by the PWE in each case. Note in Fig. 4.7a that D_{Rwave} PDF showed a progressive attenuation in negative deviations from (E4) to (E1) groups. Besides, differences among group distributions were evaluated using the Kolmogorov-Smirnov test [86], and they were only statistically significant for (E1).

Likewise, the PDF of a variable is directly related to the concept of *probability*, which can be interpreted as the relative frequency of a specific event. In the case of continuous variables, the probability is defined for intervals and it refers to the case when the variable can take a value in a predefined interval. For our specific case of D_{Rwave} values, the probability of a deviation lying in an interval [a,b] is defined as:

$$\Pr\left[a < D_{Rwave} \leq b\right] = \int_{a}^{b} PDF_{D_{Rwave}}(d) \cdot d_d \tag{4.4}$$

Once the D_{Rwave} distribution has been defined for undersensed VF R-waves, the relation between the D_{Rwave} distribution and the undersensing rate of R-waves during a VF episode was also determined. This relationship was useful to characterize the behavior of R-waves during a VF episode depending on the reference BR R-wave amplitude. The undersensing rate of R-waves was computed for each VF episode (i.e., 229 rates were obtained). Firstly, undersensing rates were evaluated for each reference BR amplitude group to discard that they were different among groups. Box-plots including undersensing rates of episodes groups are represented in Figure 4.7b. In contrast to D_{Rwave} distributions, no statistical differences among groups were present in this case. Next, episodes were grouped again, but this time into groups with similar undersensing rate of R-waves (represented in Fig. 4.8a). Specifically, six groups were defined: (G1) [0; 1]%; (G2) (1; 2.5]%; (G3) (2.5; 4]%; (G4) (4; 8]%; (G5) (8; 14]%; and (G6) ≥14%. The number of groups was set for maintaining at least ≥20 episodes per group. Then, each group was characterized with a normalized D_{Rwave} distribution, obtained as follows:

1. D_{Rwave} values were divided at 1% intervals.
2. The density value of each D_{Rwave} interval was computed as the number of episodes of the group for which that interval was included in their IQR of deviations.
3. Density values were normalized by the total number of episodes included in the group.

Therefore, each group was defined by a specific D_{Rwave} distribution and covered a range of undersensing rates. Resulting joint representation of both values is represented in Fig. 4.8b. Note that an exponential relationship is evidenced between both values, for which: the higher the density of negative D_{Rwave} values during a VF episode, the higher the undersensing rate. This relates directly the behavior of R-wave amplitudes during a VF episode and the undersensing rate, which conditions the undersensing of VF episodes in an ICD.

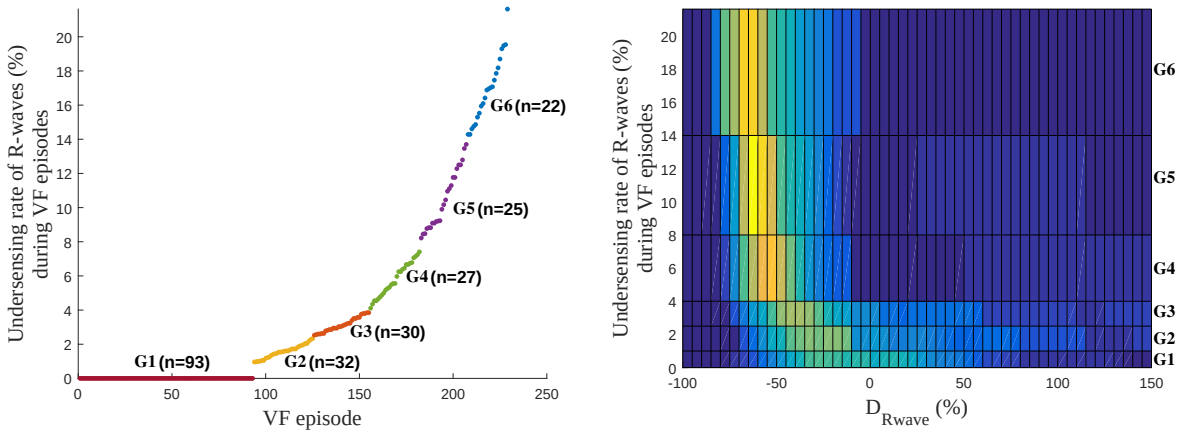


Figure 4.8: (a) Groups of VF episodes with similar undersensing rate. (b) Undersensing rate of VF R-waves depending on the D_{Rwave} distribution. Horizontal lines between groups denote the undersensing rate range covered by each group. Yellow and blue colors represent high and low density of R-waves, respectively.

4.4 Setting of the safety threshold

As mentioned in Section 4.1, any voltage not exceeding 0.3 mV in the filtered and rectified EGM is not considered as cardiac electrical activity by ICDs. This minimum value is critical in the detection of fatal arrhythmic episodes because, if the EGM amplitude is too low, R-waves can be masked as non-cardiac electrical activity. Besides, when the undersensing rate of R-waves during a VF episode exceeds 25%, the risk of undersensing or late detecting fatal arrhythmic episodes increases significantly. During a VF episode, two possibilities of undersensing are possible:

1. R-waves not considered as cardiac electrical activity, i.e., R-waves with an amplitude bellow the ICD sensitivity threshold of 0.3 mV;

2. R-waves not detected by the ICD.

Taking this information into account, the D_{Rwave} PDF can be used to compute dos both rates as probabilities by selecting:

1. R-waves with an amplitude bellow a specific d deviation, i.e., $\Pr[D_{Rwave} \leq d]$;

2. R-waves with an amplitude above a specific d deviation but undersensed by the ICD, i.e., $\Pr[D_{Rwave} > d \wedge D_{Rwave} \text{ is undersensed}]$

If d is assigned to the ICD sensitivity threshold, the deviation associated to the safety threshold can be set from a trade-off of 25% between both probabilities. Then, as both events are mutually exclusive, the probability of the global event of undersensing R-waves during a VF episode, $UP(\cdot)$, can be expressed for a specific d as

$$UP(d) = \Pr[D_{Rwave} \leq d] + \Pr[D_{Rwave} > d \wedge D_{Rwave} \text{ is undersensed}] \tag{4.5}$$

Therefore, the process for setting the safety threshold is composed of three stages:

1. to find the deviation D_{ST} for which $UP(D_{ST}) \leq 0.25$, by maximizing the following cost function,
$$D_{ST} = \underset{d}{argmax}\,\{UP(d)\} \quad subject\,to \quad UP(d) \leq 0.25 \tag{4.6}$$

2. to assign D_{ST} to the ICD sensitivity threshold for determining the reference BR R-wave amplitude (R_{BR}), i.e., the safety threshold on reference BR R-wave amplitudes. This value is obtained by replacing D_{Rwave} in Eq. (4.2) with D_{ST}, and R_{VF} with the ICD sensitivity threshold of 0.3 mV,
$$ST_{BR} = \frac{0.3}{1 + D_{ST}} \quad (mV) \tag{4.7}$$

3. to extrapolate ST_{BR} to the clinical scenario to obtain the $ST_{Clinical}$, i.e., the safety threshold on clinical BR R-wave amplitudes. This value is obtained by means of the linear relationship defined in Eq. (4.1),
$$ST_{Clinical} = 2.23 \cdot ST_{BR} + 1.38 \quad (mV) \tag{4.8}$$

Following this process and using the LOPOCV validation strategy (defined in Section 3.2 of Chapter 3), the safety threshold was experimentally set from the 229 VF episodes used in the study. In Fig. 4.9, the evolution of UP with respect to deviation d_i is represented when all episodes are used (Fig. 4.9a) and for the four groups of episodes with a similar reference BR R-wave amplitude (Figs. 4.9b - 4.9e). In these figures, blue lines, filled colors, and red dashed lines, represent (in each scenario) the median value estimated for UP, the **M**inimum-**M**aximum **R**ange (MMR) estimated for UP, and the median value estimated for D_{ST}, respectively. Note that, in contrast to current guidelines, the safety threshold obtained in all scenarios is much lower than the amplitude of 7 mV recommended at the ICD implantation. Specifically, for the worst scenario of reference BR R-wave amplitude, i.e. (E1), R-wave amplitudes $\leq$2.47 mV (MMR: 2.27–2.62 mV) may lead to $\geq$25% of undersensed VF R-waves. Besides, as represented in Fig. 4.9, the lower the reference BR R-wave amplitude, the lower the safety threshold. Only for (E4) scenario this progression does not occur. However, that result and the dispersion of the MMR in (E1) can be explained by the low number of deviations (n) used for the estimation ($\approx$1250 values in both scenarios). Results here obtained evidence the security of ICD detecting fatal arrhythmic episodes. As explained in the next section, it presents relevant clinical implications.

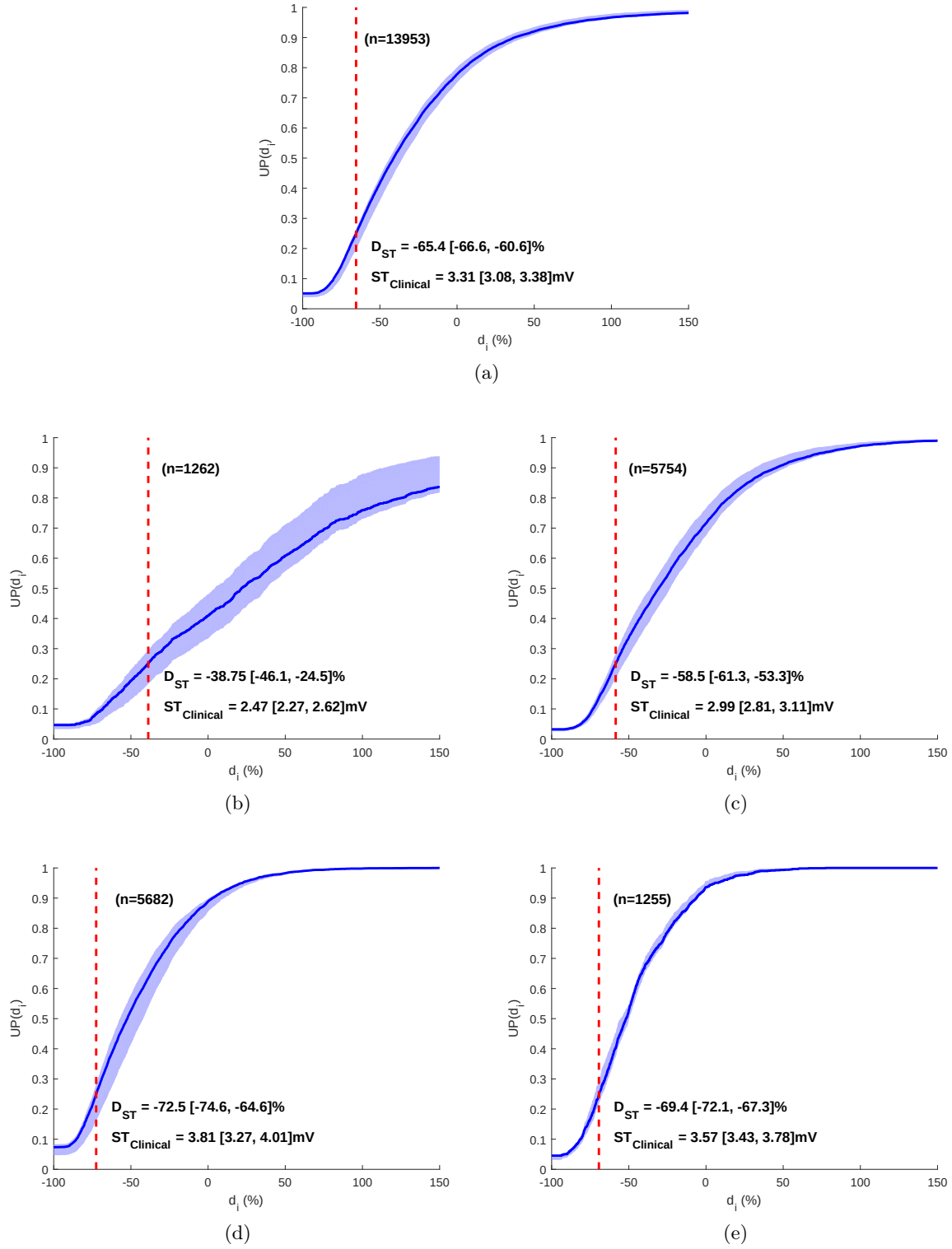


Figure 4.9: Evolution of the UP with respect to the deviation d_i when all episodes are used (a) and for the four groups of episodes with a similar reference BR R-wave amplitude (b)-(d). Blue lines, filled colors, and red dashed lines represent the estimated UP median, UP MMR, and D_{ST} median used for setting the safety threshold, respectively. Here, n denotes de number of D_{Rwave} values used in each scenario.

4.5 Discussion

The main findings of this chapter are the following:

1. R-wave amplitudes during a VF episode decrease with respect to a normal BR.
2. The probability of undersensing a VF R-wave increases with the negativity of the D_{Rwave}.
3. The rate of undersensing R-waves in a VF episode is exponentially related to the D_{Rwave} distribution: the higher the density of negative D_{Rwave} values during a VF episode, the higher the undersensing rate.
4. BR R-wave amplitudes $\leq$ 2.47 mV (MMR: 2.27–2.62 mV) have the potential to undersense high rates ($\geq$ 25%) of R-waves in a VF episode.

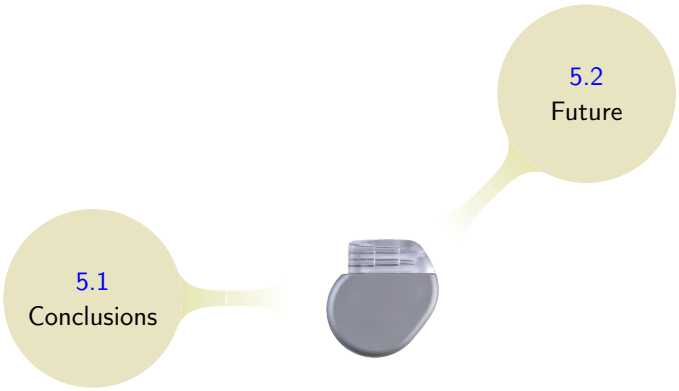
Previously, a decrease in amplitude between BR and VF R-waves had been described in patients undergoing intra-operative induced VF [87, 88]. Both works and the one presented by Leitch *et al* in [89] reported that BR R-wave amplitude decreased by an average ≈-15% to 50% after induced VF/polymorphic ventricular tachycardia. Hence, the different VF criteria, signal recording and processing may explain the differences with the results of our study. Using current digital tools for analysis and strict VF criteria, we observed negative median amplitude deviations ≈-40.9% (IQR=-64.9%, -5.4%) during spontaneous VF episodes compared with BR R-wave amplitudes. However, rather than mean amplitude values during VF, rapid fluctuations in amplitude between consecutive VF R-waves may lead to higher undersensing rates due to sensitivity adaptation during VF [90, 91]. In fact, undersensed intrinsic deviations during VF often followed a single high amplitude R-wave [83]. Thus, higher amplitude differences between consecutive VF R-waves explain higher undersensing rates.

Likewise, our data did not show statistically significant differences in undersensing rates of VF R-waves among the four groups of reference BR R-wave amplitude. Ruetz *et al* [83] reported similar results using BR R-waves and comparisons with both induced and spontaneous VF episodes. However, we only included spontaneous VF episodes using rigorous selection criteria and wide consensus among experts, to avoid analyzing both monomorphic and polymorphic VFs. Thus, we included 229 VF episodes for analysis compared with 596 spontaneous VF episodes by Ruetz *et al*, despite a much larger sample size in our data (2507 vs 259 patients, respectively). Besides, the mean of VF-CL in our data was 189.6±29.1 ms, whereas the most common VF cut-off rates in the work by Ruetz *et al* was 300–320 mS. This clearly shows the differences among episodes included in both studies. Such criteria are especially relevant to avoid incorrect estimation of the safety threshold, since amplitude differences at longer CL and more organized rhythms may be substantially different, as reported by Ellenbogen *et al* [87]. In this work, both low variability and slightly decreased mean amplitude values (-5% to -15%) in patients with **V**entricular **T**achycardia (VT) episodes ranging in CL between 240 and 360 ms were documented. Ruetz *et al* [83] also reported that VF amplitude did not decrease significantly between 5 and 10 sec., which may be explained by the inclusion of episodes with considerable longer CL.

Interestingly, there were no differences in the D_{Rwave} distribution among the four groups (E1)-(E4) of reference BR R-wave amplitude. Such a phenomenon has important implications to calculate the safety threshold. Thus, the safety threshold would be considerably different using the distribution of each group: (E1) $ST_{Clinical}$ = 2.47 mV (MMR: 2.27–2.62 mV); (E2) $ST_{Clinical}$ = 2.99 mV (MMR: 2.81–3.11 mV); (E3) $ST_{Clinical}$ = 3.81 mV (MMR: 3.27–4.01 mV); and (E4) $ST_{Clinical}$ = 3.57 mV (MMR: 3.43–3.78 mV). In fact, the safety thresholds obtained from (E4) to (E1) groups would be less consistent with current clinical evidence, showing insignificantly higher undersensing rates when BR R-wave amplitudes are ≥3 mV [83, 91].

Otherwise, episodes used in this study come from patients with a high-risk of spontaneous VF episodes. Therefore, the MMR of safety threshold may assist physicians during follow-up of this type of patients for determining the right balance between lead replacement at the time of generator change and the undersensing risk of VF. Lead removal is not exempt from life-threatening complications, and such decision is especially difficult in patients with chronically low BR R-wave amplitudes [92]. Our data suggest that chronically low BR R-wave amplitudes may be stable. Thus, 15 out of 83 patients with generator replacement showed 1.3%±12.8% variability in BR R-wave amplitudes (follow-up 2.0±1.5 years). The safety threshold also provides additional information to physicians, before attempting a defibrillation test in patients with severe systolic dysfunction and high risk of complications [93]. In fact, despite performing a defibrillation test, spontaneous VF episodes might differ in amplitude from induced episodes, which would make the test useless.

Finally, the study here presented is encompassed in a much more multidisciplinary clinical work. More extensive clinical results are detailed in [94], in which clinical variables were also analyzed and differences in the safety threshold among groups of cardiomyopathies were evaluated. Likewise, only patients with Medtronic® devices were included in the study, which precludes the extrapolation of our results to other situations and manufacturers.



Chapter 5

Conclusions and Future Work

PREDICTIVE DA is increasingly taking a leading role in clinical practice providing health-care professionals with support for their daily decisions. Currently, advances in technologies for DA are being significant and they are allowing to develop new tools solving common problems in the clinical practice. The research work developed in this Thesis is a clear example. Within the cardiology scope, a DA solution to two current challenges has been proposed, aiming to support the clinical decision on patients with ICD. Conclusions and active research lines for future work are described bellow, ending this Thesis.

5.1 Conclusions

This Thesis provides a computational solution based on DA techniques for two current challenges in cardiology: (1) the automatic classification of cardiac arrhythmic episodes recorded by ICDs; and (2) the determination of a safety threshold on BR R-wave amplitudes for ensuring a low risk of undersensing VF episodes in ICDs.

Regarding the first challenge, a novel methodology for classifying ICD arrhythmic episodes has been proposed. The methodology allowed us to design classifiers yielding high classification performance. Besides, all previously defined requirements were successfully fulfilled. This was possible thanks to the effective combination of DA concepts, which allowed to design a methodology that:

- *requires minimal computational burden*;
- *deals with episodes of different time duration*;
- *manages episodes with blank intervals*;
- *and considers both heart activation events and EGM waveforms.*

As described in Chapter 3, the proposed methodology is composed of three stages: preprocessing, similarity estimation, and classification. The requirement that mainly conditioned its design was the irregularity of time duration among arrhythmic episodes. Specifically, the use of DC techniques allowed to quantify similarity (by means of CSMs) between pairs of arrhythmic episodes even with blank intervals. Besides, a minimal computational burden was maintained because a minimal specific preprocessing was applied to: (1) emphasize heart activation events and EGM waveform to emulate the know-how of expert cardiologists; and (2) increase the EGM redundancy aiming to DC algorithms could take advantage in the episode similarity estimation. Likewise, within the preprocessing stage, fast computational tasks were used, such as differentiation, sign detection, and thresholding. Next, the ICD markers were also included in the similarity estimation due to their high capability for classification. Finally, the classification task was addressed by using KMs with a SVM classification scheme.

On the one hand, the general conclusion is that the proposed methodology for classifying cardiac arrhythmic episodes recorded by ICDs allows to design high performance classifiers. Specifically, the methodology yielded a test AR (and κ coefficient) close to 78% (0.6) and 90% (0.8) in both 8 and 3-class imbalanced schemes. However, results were influenced by this imbalance situation, and performance in minority classes is open for improvement. On the other hand, in an environment as the SCOOP platform, the analysis and labeling of the arrhythmic episodes by an experts committee can be widely reduced using the proposed methodology. This effect was also evaluated by means of AL techniques. Results showed that the methodology could be also used for selecting the most relevant episodes for classification. In fact, AL techniques evidenced that a high number of episodes included in SCOOP are not relevant or contain redundant information for classification tasks. The effect of the PD for the classification was also evaluated. Two new AL heuristics were defined to this end, which evidenced that the uncertainty for the classification is not mainly determined by the type of episode but rather by the patient.

Regarding the second challenge, a procedure for estimating a safety threshold on BR R-wave amplitudes has been proposed. In order to estimate the threshold, two goals were specified, namely:

- *describing the behavior and undersensing rate of R-wave amplitudes during VF episodes*;
- *and defining the R-wave amplitude relationship between BRs and VF episodes.*

As described in Chapter 4, both goals were successfully reached. For the fist one, detected R-waves amplitudes were extracted from VF episodes and undersensed ones were manually selected. Normalized deviations of amplitudes regarding to a normal BR were computed to obtain a normalized distribution of R-waves amplitudes during VF episodes. This allowed: (1) to describe the behavior of VF R-wave amplitudes; (2) to define the R-wave amplitude relationship between BR and VF episodes; and (3) to estimate the undersensing rate of R-waves during a VF episode. Using this information, a 3-stage process was defined for setting the safety threshold, where: (1) a cost function was maximized to obtain the safety deviation (D_{ST}), for which the rate of undersensed R-waves during a VF episode was lower than 25%; (2) the D_{ST} was assigned to the ICD sensitivity threshold; and (3) the safety threshold to be used in a clinical scenario was determined. From this process, our results confirmed or showed that:

1. The R-wave amplitude during a VF episode clearly decreases regarding a normal BR.
2. The higher the deviation of a VF R-wave amplitude with respect to a normal BR (D_{Rwave}), the higher the probability of undersensing the VF R-wave.

3. The rate of R-wave undersensing in a VF episode is exponentially related to the D_{Rwave} distribution, so that the higher the density of negative D_{Rwave} values during a VF episode, the higher the undersensing rate in that VF episode.

Likewise, the safety threshold on BR R-wave amplitudes was evaluated among 4 different patient populations. Results obtained can have high clinical impact because the obtained value is much lower than one recommended by current clinical guidelines at ICD implantation. Specifically, BR R-wave amplitudes lower than 2.47 mV (MMR 2.27–2.62 mV) can lead to potentially risk situations of high R-wave undersensing rates ($\geq$ 25%) during a VF episode. These situations might lead to high risk of non or late detection of fatal arrhythmic episodes.

We conclude that this Thesis has contributed to scientific literature by providing with new insights into the development of DA-based tools for the two proposed challenges in the cardiology field. Results on a database of high variability, volume, and heterogeneity have shown that DA opens up a whole new range of possibilities for improving the clinical practice, and more specifically the cardiac diagnosis. The work here presented represents a small step of that unprecedented situation, which is starting to change current clinical guidelines and other related systems.

5.2 Future work

Advances in technologies for DA are allowing the development of new tools for supporting healthcare professionals in their clinical practice. The work developed in this Thesis makes progress in the understanding of the two raised challenges in the cardiology field. However, a wide range of new possibilities to extend, improve, and continue the research work are now open. A summary of the main active lines for future work are next described.

Classification of cardiac arrhythmic episodes

- *To improve the performance of the methodology in minority arrhythmic classes such as NS or TWO.* New approaches, such as Deep Learning could be also explored by introducing a *priori* information. Besides, more advanced DA techniques based on the combination of multiple classifiers, such as Bagging or Adaboost [95], could be tested.

- *To analyze new approaches towards the search of more discriminating patterns for classification.* This would allow to define more efficient CSMs.

- *To extend the methodology to a wider number of arrhythmic categories and EGMs.* As described in Section 2.2, due to the memory and battery constraints in ICDs, the storage structure of a given arrhythmic episode in SCOOP database consists of two EGMs and an associated set of markers. However, the storage structure is being expanded to a format including as much as possible information. In this setting, the extension of the methodology to this scenario would eventually allow to improve performance and even to validate conclusions on a wider scope.

- *To extrapolate the methodology to other ICD manufacturers or even to other scenarios of signal classification.* The proposed methodology was developed using episodes from patients with a Medtronic® ICD, which precludes its direct extrapolation to other manufacturers and classification scenarios. However, this issue represents a challenge for future work where: (1) results here obtained can be validated using similar databases; and (2)

the methodology can be adapted to similar classification scenarios such as signals coming from electrocardiograms or heart sound recordings [26, 73, 74].

- *To develop new AL heuristics oriented towards the labeling of episodes from minority arrhythmic classes.* In particular, the selection of arrhythmic classes (such as NS or TWO) would optimize even more the model learning process. Besides, this could dramatically reduce in a large extent the labeling of the arrhythmic episodes by expert committees.

Safety threshold on BR R-wave amplitudes

- *To increase the patients database useful for the study.* The safety threshold was estimated using episodes from a population of 83 patients. Hence, the increase in patients could more strongly confirm the statistical significance of the obtained results.

- *To validate the safety threshold in exitus ICD patients with a potential VF undersensing.* The use of ICD data from this population could help to evaluate the safety threshold estimated in scenarios of actual VF undersensing.

- *To extrapolate the safety threshold to other ICD manufacturers.* The safety threshold was developed using episodes coming from patients with a Medtronic® ICD. Current ICD manufactures have implemented similar R-wave and VF detection algorithms. In this setting, a specific safety threshold could be estimated depending on the manufacturer.

- *To analyze the variability of safety threshold in terms of the patient population.* As SCOOP database increases, more specific populations of patients can be categorized according to clinical variables, heart diseases, or drug treatments. Hence, a future work would be to determine a specific safety threshold for each population oriented towards a specialized diagnosis.

Finally, the work presented in this Thesis could form the basis for future systems alerting cardiologists to the possibility of further patients at risk for complications. As technology has evolved, ICD constraints has been reduced. Therefore, we can think that future ICDs will include or interact with DA-based tools such as those presented here. Results have convincingly demonstrated that large and high quality databases are going to play a major role in the development of future trends in cardiology. The best is yet to come.

Acknowledgment

THANKS to the researchers and contributors in SCOOP platform and to the Medtronic Ibérica® S.A. team for their kind and professional support. Special thanks to Fernando Chavarría Asso and Juan Benito Tur Santolaria for their valuable technical assistance with Medtronic® devices. Major thanks to Drs. Ricardo Santiago Mozos, Devis Tuia, Manuel Marina Breysse, and David Filgueiras Rama for their fruitful ideas, discussions, and collaboration on this work.

This research work has been mainly and partly supported by FPU grant FPU13/03134 and [TEC2013-48439-C4-1-R, TEC2013-48439-C4-2-R, TEC2016-75161-C2-1-R, and TEC2016-75361-R] projects from Spanish Government, respectively.

References

[1] C. D. Swerdlow, S. J. Asirvatham, K. A. Ellenbogen, and P. A. Friedman, "Troubleshooting implanted cardioverter defibrillator sensing problems I," *Circulation: Arrhythmia and Electrophysiology*, vol. 7, no. 6, pp. 1237–1261, 2014.

[2] World Health Organization. (2015) The top 10 causes of death (Fact Sheet no. 310). Accessed on April 1st, 2017. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs310/en/

[3] T. B. García and G. T. Miller, *Arrhythmia recognition: The art of interpretation*, 1st ed. Jones & Bartlett Learning, 2004.

[4] R. X. Stroobandt, S. S. Barold, and A. F. Sinnaeve, *Implantable cardioverter defibrillators step by step: An illustrated guide*, 1st ed. Wiley-Blackwell, 2009.

[5] Medtronic Ibérica. (2017) Medical devices and technology. Accessed on April 1st, 2017. [Online]. Available: http://www.medtronic.es/

[6] A. Fontenla, M. López-Gil, J. Martínez-Ferrer *et al.*, "Clinical profile and incidence of ventricular arrhythmia in patients undergoing defibrillator generator replacement in Spain," *Revista Española de Cardiología (English Edition)*, vol. 67, no. 12, pp. 986–992, 2014.

[7] A. Ruiz-Salas, A. Barrera-Cordero, X. Vi nolas *et al.*, "Usefulness of antitachycardia pacing in arrhythmogenic right ventricular dysplasia/cardiomyopathy," *International Journal of Cardiology*, vol. 181, pp. 172–173, 2015.

[8] I. Anguera, P. Dallaglio, J. Martínez-Ferrer *et al.*, "Shock reduction with multiple bursts of antitachycardia pacing therapies to treat fast ventricular tachyarrhythmias in patients with implantable cardioverter defibrillators: A multicenter study," *Journal of Cardiovascular Electrophysiology*, vol. 26, no. 7, pp. 774–782, 2015.

[9] V. Bertomeu-González, J. Moreno-Arribas, J. Castillo-Castillo *et al.*, "Etiology and programming effects on shock efficacy in ICD recipients," *Pacing and Clinical Electrophysiology*, vol. 39, no. 1, pp. 73–80, 2016.

[10] E. Aliot, R. Nitzsché, and A. Ripart, "Arrhythmia detection by dual-chamber implantable cardioverter defibrillators. A review of current algorithms," *Europace*, vol. 6, no. 4, pp. 273–286, 2004.

[11] L. R. C. Dekker, T. A. M. Schrama, F. H. L. Steinmetz *et al.*, "Undersensing of VF in a patient with optimal R wave sensing during sinus rhythm," *Pacing and Clinical Electrophysiology*, vol. 27, no. 6, pp. 833–834, 2004.

[12] M. Glikson, D. Luria, P. A. Friedman *et al.*, "Are routine arrhythmia inductions necessary in patients with pectoral implantable cardioverter defibrillators?" *Journal of Cardiovascular Electrophysiology*, vol. 11, no. 2, p. 127–135, 2000.

[13] P. Panotopoulos, D. Krum, K. Axtell *et al.*, "Ventricular fibrillation sensing and detection by implantable defibrillators: Is one better than the others? A prospective, comparative study," *Journal of Cardiovascular Electrophysiology*, vol. 12, no. 4, p. 445–522, 2001.

[14] C. D. Swerdlow, "Implantation of cardioverter defibrillators without induction of ventricular fibrillation," *Circulation*, vol. 103, no. 17, pp. 2159–2164, 2001.

[15] C. D. Swerdlow, A. M. Russo, and P. J. Degroot, "The dilemma of ICD implant testing," *Pacing and Clinical Electrophysiology*, vol. 30, no. 5, p. 675–700, 2007.

[16] J. D. Day, R. N. Doshi, P. Belott *et al.*, "Inductionless or limited shock testing is possible in most patients with implantable cardioverter defibrillators / cardiac resynchronization therapy tefibrillators," *Circulation*, vol. 115, no. 18, pp. 2382–2389, 2007.

[17] J. S. Healey, S. H. Hohnloser, M. Glikson *et al.*, "Cardioverter defibrillator implantation without induction of ventricular fibrillation: A single-blind, non-inferiority, randomised controlled trial (SIMPLE)," *The Lancet*, vol. 385, no. 9970, p. 785–791, 2015.

[18] T. A. Runkler, *Data analytics: Models and algorithms for intelligent data analysis*, 2nd ed. Springer, 2016.

[19] A. Kankanhalli, J. Hahn, S. Tan, and G. Gao, "Big data and analytics in healthcare: Introduction to the special section," *Health Information Science and Systems*, vol. 18, no. 2, pp. 233–235, 2016.

[20] Z. Xu, G. L. Frankwick, and E. Ramirez, "Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective," *Journal of Business Research*, vol. 69, no. 5, pp. 1562–1566, 2016.

[21] B. K. Daniel, *Big data and learning analytics in higher education: Current theory and practice*, 1st ed. Springer, 2016.

[22] G. Shmueli, P. C. Bruce, and P. N. R, *Data mning for business analytics: Concepts, techniques, and applications with XLMiner*, 1st ed. Springer, 2016.

[23] J. Gillberg, "Detection of cardiac tachyarrhythmias in implantable devices," *Journal of Electrocardiology*, vol. 40, no. 6, pp. 123–128, 2007.

[24] J. Toquero, J. Alzueta, L. Mont *et al.*, "Morphology discrimination criterion wavelet improves rhythm discrimination in single-chamber implantable cardioverter-defibrillators: Spanish register of morphology discrimination criterion wavelet (REMEDIO)," *Europace*, vol. 11, no. 6, pp. 727–733, 2009.

[25] S. Shome, D. Li, J. Thompson, and A. McCabe, "Improving contemporary algorithms for implantable cardioverter-defibrillator function," *Journal of Electrocardiology*, vol. 43, no. 6, pp. 503–508, 2010.

[26] D. Mahajan, Y. Dong, L. A. Saxon *et al.*, "Performance of an automatic arrhythmia classification algorithm: Comparison to the ALTITUDE electrophysiologist panel adjudications," *Pacing and Clinical Electrophysiology*, vol. 37, no. 7, pp. 889–899, 2014.

[27] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*, 1st ed. Cambridge University Press, 2004.

[28] R. Duda, P. Hart, and D. Stork, *Pattern classification*, 2nd ed. Wiley-Interscience, 2001.

[29] Student, "The probable error of a mean," *Biometrika Trust*, vol. 6, no. 1, p. 1–25, 1908.

[30] Medtronic Bakken Research Center. (2012) UMBRELLA - Incidence of arrhythmias in Spanish population with a Medtronic implantable cardiac defibrillator implant. Identifier: NCT01561144. Accessed on April 1st, 2017. [Online]. Available: https://clinicaltrials.gov/

[31] D. Dubin, *Rapid interpretation of EKG's*, 6th ed. Cover, 2000.

[32] D. Cerra and M. Datcu, "A fast compression-based similarity measure with applications to content-based image retrieval," *Visual Communication and Image Representation*, vol. 23, no. 2, pp. 293–302, 2012.

[33] A. Granados, K. Koroutchev, and F. de Borja-Rodriguez, "Discovering data set nature through algorithmic clustering based on string compression," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 3, pp. 699–711, 2015.

[34] N. J. Hudson, L. Porto-Neto, J. W. Kijas, and A. Reverter, "Compression distance can discriminate animals by genetic profile, build relationship matrices and estimate breeding values," *Genetics Selection Evolution*, vol. 47, no. 1, pp. 78–92, 2015.

[35] I. Capurro, F. Lecumberry, F. Martín *et al.*, "Efficient sequential compression of multichannel biomedical signals," *IEEE Journal of Biomedical and Health Informatics*, vol. -, no. -, pp. –, 2016.

[36] P. Vitányi, "Compression-based similarity," in *International Conference on Data Compression, Communications and Processing*, Palinuro, Cilento Coast, Italy, 21-24 June 2011, pp. 111–118.

[37] M. Li and P. Vitányi, *An introduction to Kolmogorov complexity and its applications*, 3rd ed. Springer, 2008.

[38] C. H. Bennett, P. Gács, M. Li *et al.*, "Information distance," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1407–1423, 1998.

[39] M. Li, X. Chen, X. Li *et al.*, "The similarity metric," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3250–3264, 2004.

[40] D. Salomon, *Data compression. The complete reference*, 4th ed. Springer, 2007.

[41] A. Macedonas, D. Besiris, G. Economou, and S. Fotopoulos, "Dictionary based color image retrieval," *Journal of Visual Communication and Image Representation*, vol. 19, no. 7, pp. 464–470, 2008.

[42] J. Leskovec, A. Rajaraman, and J. Ullman, *Mining of massive datasets*, 2nd ed. Cambridge University Press, 2014.

[43] A. Gersho and R. M. Gray, *Vector quantization and signal compression*, 1st ed. Springer, 1992.

[44] J. L. Rojo-Álvarez, M. Martínez-Ramón, J. Muñoz Marí, and G. Camps-Valls, *Digital signal processing with kernel methods*, 1st ed. Wiley, 2016.

[45] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Transactions on Electronic Computers*, vol. EC-14, no. 3, pp. 326–334, 1965.

[46] J. Mercer, "Functions of positive and negative type, and their connection with the theory of integral equations," *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, vol. 209, no. 441, pp. 415–446, 1909.

[47] S. Banerjee and A. Roy, *Linear algebra and matrix analysis for statistics*, 1st ed. Chapman & Hall-CRC, 2014.

[48] V. Vapnik, *The nature of statistical learning theory*, 2nd ed. Springer Verlag, 2000.

[49] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

[50] S. Knerr, L. Personnaz, and G. Dreyfus, *Single-layer learning revisited: a stepwise procedure for building and training a neural network.* Springer Berlin Heidelberg, 1990, pp. 41–50.

[51] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

[52] B. Settles, "Active learning," *Synthesis lectures on artificial intelligence and machine learning*, vol. 6, no. 1, pp. 1–114, 2012.

[53] D. Tuia, F. Ratle, F. Pacifici *et al.*, "Active learning methods for remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2218–2232, 2009.

[54] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4085–4098, 2010.

[55] E. Pasolli, F. Melgani, D. Tuia *et al.*, "SVM active learning approach for image classification using spatial information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 4, pp. 2217–2233, 2014.

[56] M. Babaee, S. Tsoukalas, G. Rigoll, and M. Datcu, "Visualization-based active learning for the annotation of SAR images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 10, pp. 4687–4698, 2015.

[57] D. H. M. Nguyen and J. D. Patrick, "Supervised machine learning and active learning in classification of radiology reports," *Journal of the American Medical Informatics Association*, vol. 21, no. 5, pp. 893–901, 2014.

[58] N. Nissim, M. R. Boland, N. P. Tatonetti *et al.*, "Improving condition severity classification with an efficient active learning based framework," *Journal of Biomedical Informatics*, vol. 61, pp. 44–54, 2016.

[59] D. Mahapatra, F. M. Vos, and J. M. Buhmann, "Active learning based segmentation of crohns disease from abdominal MRI," *Computer Methods and Programs in Biomedicine*, vol. 128, pp. 75–85, 2016.

[60] I. Žliobaité, A. Bifet, B. Pfahringer, and G. Holmes, "Active learning with drifting streaming data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 1, pp. 27–39, 2014.

[61] S. Sivaraman and M. M. Trivedi, "Active learning for on-road vehicle detection: a comparative study]," *Machine Vision and Applications*, vol. 25, no. 3, pp. 599–611, 2011.

[62] X. Eirini-Pantazi, D. Moshou, and C. Bravo, "Active learning system for weed species recognition based on hyperspectral sensing," *Biosystems Engineering*, vol. 146, pp. 193–202, 2016.

[63] I. Mora-Jim'enez and A. Figueiras-Vidal, "Improving performance of neural classifiers via selective reduction of target levels," *Neurocomputing*, vol. 72, no. 13, pp. 3020–3027, 2009.

[64] G. Schohn and D. Cohn, "Less is more: Active learning with support vectors machines," in *International Conference on Machine Learning*, Standford University, California, USA, 29 June - 2 July 2000, pp. 839–846.

[65] B. Schölkopf and A. J. Smola, *Learning with kernels*, 1st ed. The MIT Press Cambridge, 2002.

[66] T. Luo, K. Kramer, S. Samson *et al.*, "Active learning to recognize multiple types of plankton," in *International Conference on Pattern Recognition*, vol. 3, 26 August 2004, pp. 478–481.

[67] C. Ferri, J. Hernández-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," *Pattern Recognition Letters*, vol. 30, no. 1, pp. 27–38, 2009.

[68] C. Aggarwal, *Data classification: Algorithms and applications*, 1st ed. Chapman & Hall, 2014.

[69] I. R. Legarreta, P. S. Addison, N. Grubb *et al.*, "R-wave detection using continuous wavelet modulus maxima," in *Computing in Cardiology*, Thessaloniki, Greece, 21-24 September 2003, pp. 565–568.

[70] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a kalman filter," *Physiological Measurement*, vol. 29, no. 1, p. 15–32, 2008.

[71] I. Silva, B. Moody, J. Behar *et al.*, "Robust detection of heart beats in multimodal data," *Physiological Measurement*, vol. 36, no. 8, pp. 1629–1644, 2015.

[72] M. R. Gold, D. A. Theuns, B. P. Knight *et al.*, "Head-to-head comparison of arrhythmia discrimination performance of subcutaneous and transvenous ICD arrhythmia detection algorithms: The START study," *Journal of Cardiovascular Electrophysiology*, vol. 23, no. 4, pp. 359–366, 2012.

[73] F. Morgado, E and Alonso-Atienza, R. Santiago-Mozos *et al.*, "Quality estimation of the electrocardiogram using cross-correlation among leads," *BioMedical Engineering OnLine*, vol. 14, no. 1, pp. 59–78, 2015.

[74] I. Silva, G. B. Moody, and L. Celi, "Classification of normal/abnormal heart sound recordings: the Physionet/Computing in Cardiology challenge," in *Computing in Cardiology*, Vancouver, British Columbia, Canada, 11-14 September 2016, pp. 609–612.

[75] S. Kiranyaz and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2016.

[76] M. M. Al-Rahhal, Y. Bazi, H. AlHichri *et al.*, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, no. 1, pp. 340–354, 2016.

[77] J. G. Proakis and D. K. Manolakis, *Digital signal processing: Principles, algorithms, and applications*, 4th ed. Pearson, 2007.

[78] B. U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 1, pp. 42–57, 2002.

[79] Mathworks. (2017) Makers of matlab and simulink. Accessed on April 1st, 2017. [Online]. Available: http://www.mathworks.com/

[80] K. Nademanee, J. McKenzie, E. Kosar *et al.*, "A new approach for catheter ablation of atrial fibrillation: mapping of the electrophysiologic substrate," *Journal of the American College of Cardiology*, vol. 43, no. 11, pp. 2044–2053, 2004.

[81] H. Oral, A. Chugh, E. Good *et al.*, "Radiofrequency catheter ablation of chronic atrial fibrillation guided by complex electrograms," *Circulation*, vol. 115, no. 20, pp. 2606–2612, 2007.

[82] M. Nair, S. Nayyar, S. Rajagopal *et al.*, "Results of radiofrequency ablation of permanent atrial fibrillation of >2 years duration and left atrial size >5cm using 2-mm irrigated tip ablation catheter and targeting areas of complex fractionated atrial electrograms," *The American Journal of Cardiology*, vol. 104, no. 5, pp. 683–688, 2009.

[83] L. L. Ruetz, J. L. Koehler, M. L. Brown *et al.*, "Sinus rhythm r-wave amplitude as a predictor of ventricular fibrillation undersensing in patients with implantable cardioverter-defibrillator," *Heart Rhythm*, vol. 12, no. 12, pp. 2411–2418, 2015.

[84] B. L. Wilkoff, L. Fauchier, M. K. Stiles *et al.*, "2015 HRS/EHRA/APHRS/SOLAECE expert consensus statement on optimal implantable cardioverter-defibrillator programming and testing," *Journal of Arrhythmia*, vol. 32, no. 1, p. 1–28, 2016.

[85] B. W. Silverman, *Density estimation for statistics and data analysis*, 1st ed. Chapman & Hall/CRC, 1986.

[86] S. W. Wayne, *Applied nonparametric statistics*, 2nd ed. Duxbury, 2000.

[87] K. A. Ellenbogen, M. A. Wood, B. S. Stambler *et al.*, "Measurement of ventricular electrogram amplitude during intraoperative induction of ventricular tachyarrhythmias," *The American Journal of Cardiology*, vol. 70, no. 11, pp. 1017–1022, 1992.

[88] B. I. Michelson, D. A. Igel, and B. L. Wilkoff, "Adequacy of implantable cardioverter-defibrillator lead placement for tachyarrhythmia detection by sinus rhythm electrogram amplitude," *The American Journal of Cardiology*, vol. 76, no. 16, pp. 1162–1166, 1995.

[89] J. W. Leitch, R. Yee, G. J. Klein *et al.*, "Correlation between the ventricular electrogram amplitude in sinus rhythm and in ventricular fibrillation," *Pacing and Clinical Electrophysiology*, vol. 13, no. 9, p. 1105–1109, 1990.

[90] A. Natale, J. Sra, K. Axtell *et al.*, "Undetected ventricular fibrillation in transvenous implantable cardioverter-defibrillators," *Circulation*, vol. 93, no. 1, pp. 91–98, 1996.

[91] U. Birgersdotter-Green, L. L. Ruetz, K. Anand *et al.*, "Automated vulnerability testing identifies patients with inadequate defibrillation safety margin," *Circulation: Arrhythmia and Electrophysiology*, vol. 5, no. 16, pp. 1073–1080, 2012.

[92] M. K. Bie, D. A. Fouad, C. J. W. Borleffs *et al.*, "Trans-venous lead removal without the use of extraction sheaths, results of >250 removal procedures," *Europace*, vol. 14, no. 1, pp. 112–116, 2011.

[93] D. Birnie, S. Tung, C. Simpson *et al.*, "Complications associated with defibrillation threshold testing: The canadian experience," *Heart Rhythm*, vol. 5, no. 3, pp. 387–390, 2008.

[94] J. M. Lillo-Castellano, M. Marina-Breysse, A. Gómez-Gallanti *et al.*, "Safety threshold of R-wave amplitudes in patients with implantable cardioverter defibrillator," *Heart*, vol. 102, no. 20, pp. 1662–1670, 2016.

[95] M. N. Murty and V. S. Devi, *Pattern recognition. An algorithmic approach.*, 1st ed. Springer, 2011.