Ph.D. Thesis

Computational Tools for Atrial Arrhythmia Detection on ECG Signals: The Contribution of Different Leads

Guadalupe García Isla

SUPERVISORS:

Prof. Valentina Corino

Prof. Luca Mainardi

Ph.D. in Bioengineering

December, 2022
XXIX Cycle

"Todos los niños nacen artistas.
El problema es cómo seguir siendo artistas al crecer"
- Pablo Picasso

Summary and conclusions

Atrial fibrillation (AF) is the most common sustained arrhythmia. It affects around 0.5% of global population. The risk of suffering from AF increases with age and with the ageing of global population, its incidence is continuously increasing. Although AF itself does not represent a life-threatening condition, it is associated with higher mortality and morbidity levels.

Despite its high incidence, life-long clinical consequences and high associated costs, we are still lacking with the tools to properly predict and screen the general population. One of the main problems associated with AF is that it increases stroke risk up to a 25%. If a patient is properly diagnosed it undergoes an anticoagulant treatment to prevent such event. However, in many cases AF is diagnosed after a cryptogenic stroke has taken place.

With the development of technology and smartphones, more affordable solutions to the general population are being made available, representing a powerful tool for AF screening and monitoring. These devices generate an enormous amount of information daily that cannot be manually annotate by clinicians as typically done in short electrocardiogram (ECG) recordings.

There exists a need to develop corresponding software solutions to extract the relevant information from all the data that is being gathered daily. Moreover, it is also imperative to understand what information of these data should be stored. For so the clinical relevance of some cardiac events still needs to be clarified (i.e. premature atrial complexes (PAC).

The development of automatic algorithm for ECG signals annotation could play a double role: that of the monitoring of a patients' cardiac activity and that of the annotation of large ECG databases for the study of the clinical relevance of different cardiac events and their relationship with AF.

In this thesis algorithms for the automatic annotation of ECG signals of different nature were developed. In a fist stage a methodology for the

detection of AF that could be generalized to other arrhythmia based on the *RR* interval (RRi) analysis was presented. The methodology was based on the observation that each cardiac rhythm produces an specific pattern on the Poincaré Plot and that such pattern can be captured and used to classify further Poincaré Plot patterns. For so, a binarized version of the Poincaré Plot was introduced, the Poincaré Image. Each rhythm's pattern was captured in a Poincaré Atlas which was used to classify further Poincaré Images. Using 4 public PhysioNet databases, the methodology was tuned and tested for AF, atrial bigeminy (AB) and normal sinus rhythm (NSR) in *RR* signals of 120 s, 60 s, 30 s and 20 s length. The final methodology used normalized mutual information (NMI) and 2D correlation for the classification of new Poincaré Images using a pixel dimension of [40x40] ms. The proposed method could be suitable for its implementation on constant monitoring devices as it was only based on RRi and required a low computational power.

Nevertheless, work for the prediction of first-time occurrence of AF still ought to be done to enable the application of prophylactic measures. For so, the role of other atrial abnormal events suspected to be an indicator of future development of AF need of further study. Such is the case of PACs that have been linked in some studies with AF appearance. In order to help elucidate their role we have developed an automatic beat classifier that can detect PACs with high sensitivity. This detector could annotate long-term recording signals which cannot be manually analyzed give their length, and enable the study of their relationship with AF. The proposed model was a random forest ensemble that used morphological and rhythm information from 2-lead ECG Holter signals. Moreover, a complementary detector was developed able to identify ECG segments containing ventricular or atrial beats. It was developed with the aim of reducing the amount of false positives, a problem common to all classifiers present in literature. The segment-wise classifier used only *RR* intervals which made it more resistant to noise than those classifiers using morphological information. *RR* intervals were transformed into Poincaré Images and then introduced into a convolutional neural network (CNN). This classifier could aid not only in reducing the false positives on automatic signal annotation, but also in the manual data revision as it could reduce the amount of signal segments that should be analyzed.

Finally, artificial intelligence (AI) solutions were explored for the automatic annotation of ECG signals of different lengths and leads acquired in a

clinical setting. Two approaches were taken: one based on classic machine learning and another one integrating deep learning with machine learning. For the machine learning model, a random forest ensemble was trained to detect 7 different ECG abnormalities using morphological and temporal ECG features. Regarding the deep learning model, 26 different cardiac abnormalities were targeted. In addition the best training strategy for integrating machine learning features into a deep learning model was explored. Both models were evaluated following the challenge metric proposed by the Computing in Cardiology PhysioNet Challenges of 2020-2021. Automatic detection of numerous cardiac disorders could enable the annotation of large hospital databases. A high amount of cardiac information remains unused because their manual annotation is highly time consuming. However, these data could help in answering diverse clinical questions if analyzed properly.

To conclude this thesis has contributed to the automatic annotation of cardiac signals in three different scopes and signal types. The first analysis based on RRi segments transformed into Poincaré Images provided a new methodology that could be applicable for the detection of AF and extended to other arrhythmia. The second analysis instead targeted single beats and provided a tool for the detection of PACs on 2-lead Holter ECGs with the aim of aiding on the study of their role with respect to first-time appearance of AF and stoke. Finally, the third analysis instead provided a tool for the automatic annotation of 26 different cardiac disorders on signals acquired in a clinical setting with variable number of leads. Such tool could be use to take advantage of the huge amount of information that remains unused in hospital's unannotated databases.

Keywords: Electrocardiogram; Signal Processing; Implantable Cardioverter Defibrillator; Atrial Fibrillation; Premature Atrial Complex; Ectopic Beats; Machine Learning; Deep Learning.

Riepilogo e conclusioni

La fibrillazione atriale (AF) Colpisce circa lo 0,5% della popolazione mondiale. È l'aritmia sostenuta più comune. Il rischio di soffrire di fibrillazione atriale aumenta con l'età e, con l'invecchiamento della popolazione mondiale, la sua incidenza è in continuo aumento. Sebbene la AF di per sé non rappresenti una condizione di pericolo di vita, è associata a livelli di mortalità e morbilità più elevati.

Nonostante l'elevata incidenza, le conseguenze cliniche che si protraggono per tutta la vita e gli alti costi associati, mancano ancora gli strumenti per prevedere e sottoporre a screening in modo adeguato la popolazione generale. Uno dei problemi principali associati alla AF è che aumenta il rischio di ictus fino al 25%. Se un paziente viene diagnosticato correttamente, viene sottoposto a un trattamento anticoagulante per prevenire tale evento. Tuttavia, in molti casi la AF viene diagnosticata dopo che si già è verificato un ictus criptogenetico.

Con lo sviluppo della tecnologia e degli smartphone, sono disponibili soluzioni sempre più accessibili alla popolazione generale, che rappresentano un potente strumento per lo screening e il monitoraggio della AF. Questi dispositivi generano quotidianamente un'enorme quantità di informazioni che non possono essere annotate manualmente dai medici, come avviene di solito nelle registrazioni di brevi elettrocardiogrammi (ECG).

Esiste la necessità di sviluppare soluzioni software corrispondenti per estrarre le informazioni rilevanti da tutti i dati che vengono raccolti quotidianamente. Inoltre, è indispensabile capire quali informazioni di questi dati debbano essere memorizzate. Per questo motivo, la rilevanza clinica di alcuni eventi cardiaci deve ancora essere chiarita (ad esempio, i complessi atriali prematuri (PAC).

Lo sviluppo di un algoritmo automatico per l'annotazione dei segnali ECG potrebbe svolgere un duplice ruolo: quello del monitoraggio dell'attività

cardiaca del paziente e quello dell'annotazione di grandi database ECG per lo studio della rilevanza clinica di diversi eventi cardiaci e della loro relazione con la AF.

In questa tesi sono stati sviluppati algoritmi per l'annotazione automatica di segnali ECG di diversa natura. In una prima AF se è stata presentata una metodologia per il rilevamento della fibrillazione atriale, che potrebbe essere generalizzata ad altre aritmie, basata sull'analisi dell'intervallo RR (RRi). La metodologia si basava sull'osservazione che ogni ritmo cardiaco produce un modello specifico sul diagramma di Poincaré e che tale modello può essere catturato e utilizzato per classificare altri modelli di diagrammi di Poincaré. Per questo motivo è stata introdotta una versione binarizzata del diagramma di Poincaré, l'immagine di Poincaré. Il pattern di ogni ritmo è stato catturato in un Atlante di Poincaré che è stato utilizzato per classificare altre immagini di Poincaré. Utilizzando 4 database pubblici di PhysioNet, la metodologia è stata messa a punto e testata per la fibrillazione atriale, la bigeminia atriale (AB) e il ritmo sinusale normale (NSR) in segnali RR di 120 s, 60 s, 30 s e 20 s di lunghezza. La metodologia finale ha utilizzato la mutua informazione normalizzata(NMI) e la correlazione 2D per la classificazione di nuove immagini di Poincaré utilizzando una dimensione dei pixel di [40x40] ms. Il metodo proposto potrebbe essere adatto all'implementazione su dispositivi di monitoraggio continuo, poiché si basa solo sulle serie RRi e richiede una bassa potenza di calcolo.

Tuttavia, è ancora necessario lavorare alla previsione della prima comparsa di AF per consentire l'applicazione di misure profilattiche. Per questo motivo, il ruolo di sporadici eventi anomali atriali sospettati di essere un indicatore del futuro sviluppo della AF, necessita di essere studiato. È il caso delle PAC, che in alcuni studi sono state collegate alla comparsa di AF. Per contribuire a chiarire il loro ruolo, abbiamo sviluppato un classificatore automatico di battiti in grado di rilevare le PAC con un'elevata sensibilità. Questo rilevatore potrebbe annotare i segnali di registrazione ECG a lungo termine che non possono essere analizzati manualmente, data la loro lunghezza, e consentire lo studio della loro relazione con la AF. Il modello proposto è un random forest che utilizza informazioni morfologiche e ritmiche provenienti da segnali ECG Holter a 2 derivazioni. Inoltre, è stato sviluppato un rilevatore complementare in grado di identificare i segmenti ECG contenenti battiti ventricolari o atriali. È stato sviluppato con l'obiettivo di ridurre la quantità di falsi positivi, un problema comune a tutti i classificatori presenti in letteratura. Il classificatore di segmenti ECG uti-

lizza solo gli intervalli *RR* che lo rendono più resistente al rumore rispetto ai classificatori che utilizzano informazioni morfologiche. Gli intervalli *RR* sono stati trasformati in immagini di Poincaré e quindi introdotti in una rete neurale convoluzionale (CNN). Questo classificatore potrebbe aiutare non solo a ridurre i falsi positivi nell'annotazione automatica del segnale, ma anche nella revisione manuale dei dati, in quanto potrebbe ridurre la quantità di segmenti di segnale da analizzare.

Infine, sono state esplorate soluzioni di intelligenza artificiale (AI) per l'annotazione automatica di segnali ECG di diverse lunghezze e derivazioni acquisiti in ambito clinico. Si sono adottati due approcci: uno basato sull'apprendimento automatico classico e un altro che integra deep learning con machine learning. Per il modello di machine learning, è stato addestrato un random forest per rilevare 7 diverse anomalie ECG utilizzando caratteristiche ECG morfologiche e temporali. Per quanto riguarda il modello di deep learning, sono state individuate 26 diverse anomalie cardiache. Inoltre, è stata esplorata la migliore strategia di addestramento per integrare le caratteristiche di machine learning in un modello di deep learning. Entrambi i modelli sono stati valutati in base alla metrica proposta dal Computing in Cardiology PhysioNet Challenges del 2020-2021. Il rilevamento automatico di numerosi disturbi cardiaci potrebbe consentire l'annotazione di grandi database ospedalieri. Un'elevata quantità di informazioni cardiache rimane inutilizzata perché la loro annotazione manuale richiede molto tempo. Tuttavia, se analizzati correttamente, questi dati potrebbero aiutare a rispondere a diverse domande cliniche.

In conclusione, questa tesi ha contribuito all'annotazione automatica dei segnali cardiaci in tre diversi ambiti e tipi di segnale. La prima analisi, basata su segmenti RRi trasformati in immagini di Poincaré, ha fornito una nuova metodologia che potrebbe essere applicata per il rilevamento della AF ed estesa ad altre aritmie. La seconda analisi, invece, si è rivolta a singoli battiti e ha fornito uno strumento per il rilevamento di PAC su ECG Holter a 2 derivazioni, con l'obiettivo di contribuire allo studio del loro ruolo rispetto alla prima comparsa di AF e stoke. Infine, la terza analisi ha fornito uno strumento per l'annotazione automatica di 26 diversi disturbi cardiaci su segnali acquisiti in ambiente clinico con un numero variabile di derivazioni. Tale strumento potrebbe essere utilizzato per sfruttare l'enorme quantità di informazioni che rimangono inutilizzate nei database non annotati degli ospedali.

Parole chiave: Elettrocardiogramma; elaborazione del segnale; defibrillatore cardioverter impiantabile; fibrillazione atriale; complesso atriale prematuro; battiti ectopici; machine learning; deep learning.

Resumen y conclusiones

La fibrilación auricular (AF) es la arritmia sostenida más común. Afecta alrededor del 0,5% de la población mundial. El riesgo de padecer AF aumenta con la edad y, con el envejecimiento de la población mundial, su incidencia está incrementando. Aunque la AF en sí misma no representa una afección que ponga en peligro la vida, se asocia a niveles de mortalidad y morbilidad elevados.

A pesar de su extendida incidencia, de sus consecuencias clínicas de por vida y de los elevados costes asociados, aún carecemos de las herramientas necesarias para predecir y examinar adecuadamente a la población general. Uno de los principales problemas asociados a la AF es que aumenta el riesgo de ictus hasta en un 25%. Si un paciente es diagnosticado adecuadamente, se somete a un tratamiento anticoagulante para prevenir dicho evento. Sin embargo, en muchos casos la AF se diagnostica después de que ya se haya producido un ictus criptogénico.

Con el desarrollo de la tecnología y de los smartphones, se están poniendo a disposición de la población general soluciones más asequibles que representan una poderosa herramienta para el cribado y la monitorización de la AF. Estos dispositivos generan diariamente una enorme cantidad de información que no puede ser anotada manualmente por los clínicos, como suele hacerse en los registros de electrocardiogramas cortos (ECG).

Existe la necesidad de desarrollar las correspondientes soluciones de software para extraer la información relevante de todos los datos que se recogen a diario. Además, también es imprescindible comprender qué información de estos datos debe almacenarse. Por ello, la relevancia clínica de algunos eventos cardíacos aún debe ser aclarada (por ejemplo, los complejos auriculares prematuros (PAC)). El desarrollo de algoritmos automáticos para la anotación de señales de ECG podría desempeñar un doble papel: el de la monitorización de la actividad cardíaca de los pacientes y el de la anotación

de grandes bases de datos de ECG para el estudio de la relevancia clínica de diferentes eventos cardíacos y su relación con la AF.

En esta tesis se han desarrollado algoritmos para la anotación automática de señales de ECG de diferente naturaleza. En una primera etapa se presentó una metodología para la detección de la AF basada en el análisis del intervalo RR (RRi) que podría generalizarse a otras arritmias. La metodología se basó en la observación de que cada ritmo cardíaco produce un patrón específico en el Poincaré Plot y que dicho patrón puede ser capturado y utilizado para clasificar otros patrones del Poincaré Plot. Para ello, se introdujo una versión binarizada del Poincaré Plot, la Imagen Poincaré. El patrón de cada ritmo se capturó en un Atlas Poincaré que se utilizó para clasificar otras Imágenes Poincaré. Utilizando 4 bases de datos públicas de PhysioNet, la metodología se ajustó y probó para la FA, el bigeminismo auricular (AB) y el ritmo sinusal normal (NSR) en señales RR de 120 s, 60 s, 30 s y 20 s de duración. La metodología final utilizó la información mutua normalizada (NMI) y la correlación 2D para la clasificación de las nuevas imágenes Poincaré utilizando una dimensión de píxel de [40x40] ms. El método propuesto podría ser adecuado para su implementación en dispositivos de monitorización constante, ya que sólo se basaba en RRi y requeriría de una baja potencia computacional.

Sin embargo, todavía hay que trabajar en la predicción de la primera aparición de la AF para poder aplicar medidas profilácticas. Por ello, es necesario seguir estudiando el papel de otros eventos anormales auriculares que se sospecha que son un indicador del desarrollo futuro de la AF. Tal es el caso de los PACs que se han relacionado en algunos estudios con la aparición de AF. Para ayudar a dilucidar su papel, hemos desarrollado un clasificador automático de latidos que puede detectar los PACs con alta sensibilidad. Este detector podría anotar señales de registro de larga duración que no pueden ser analizadas manualmente dada su longitud, y permitir el estudio de su relación con la AF. El modelo propuesto es un conjunto de bosques aleatorios que utiliza información morfológica y de ritmo de las señales de ECG Holter de 2 derivaciones. Además, se desarrolló un detector complementario capaz de identificar segmentos de ECG que contienen latidos ventriculares o auriculares. Se desarrolló con el objetivo de reducir la cantidad de falsos positivos, un problema común a todos los clasificadores presentes en la literatura. El clasificador por segmentos utiliza únicamente intervalos RR, lo que lo hace más resistente al ruido que los clasificadores que utilizan información morfológica. Los intervalos RR se transformaron en

imágenes de Poincaré y luego se introdujeron en una red neuronal convolucional (CNN). Este clasificador podría ayudar no sólo a reducir los falsos positivos en la anotación automática de la señal, sino también en la revisión manual de los datos, ya que podría reducir la cantidad de segmentos de señal que deberían analizarse. En cuanto a la detección de taquiarritmias, todavía es necesario comprender el papel de los latidos individuales.

Por último, se exploraron soluciones de inteligencia artificial (AI) para la anotación automática de señales de ECG de diferentes longitudes y derivaciones adquiridas en un entorno clínico. Se adoptaron dos enfoques: uno basado en el aprendizaje automático clásico y otro que integraba el deep learning con el aprendizaje automático clásico (machine learning). Para el modelo de aprendizaje automático, se entrenó un conjunto de random forest para detectar 7 anomalías diferentes en el ECG utilizando características morfológicas y temporales del ECG. En cuanto al modelo de aprendizaje profundo, se dirigió a 26 anormalidades cardíacas diferentes. Además, se exploró la mejor estrategia de entrenamiento para integrar características de aprendizaje automático en un modelo de aprendizaje profundo. Ambos modelos se evaluaron siguiendo la métrica de los retos propuestos por el Computing in Cardiology PhysioNet Challenges de 2020-2021. La detección automática de numerosos trastornos cardíacos podría permitir la anotación de grandes bases de datos hospitalarias. Una gran cantidad de información cardiaca permanece sin utilizar porque su anotación manual requiere mucho tiempo. Sin embargo, estos datos podrían ayudar a responder a diversas cuestiones clínicas si se analizan adecuadamente.

Para concluir, esta tesis ha contribuido a la anotación automática de señales cardíacas en tres ámbitos y tipos de señales diferentes. El primer análisis basado en segmentos de RRi transformados en imágenes de Poincaré proporcionó una nueva metodología que podría ser aplicable para la detección de la AF y extenderse a otras arritmias. El segundo análisis se centró en latidos individuales y proporcionó una herramienta para la detección de PACs en ECGs Holter de 2 derivaciones con el objetivo de ayudar en el estudio de su papel con respecto a la aparición por primera vez de la AF y el paro. Por último, el tercer análisis proporcionó una herramienta para la anotación automática de 26 trastornos cardíacos diferentes en señales adquiridas en un entorno clínico con un número variable de derivaciones. Esta herramienta podría servir para aprovechar la enorme cantidad de información que queda sin utilizar en las bases de datos no anotadas de los hospitales.

Palabras clave: Electrocardiograma; Procesamiento de señales; Desfibrilador cardioversor implantable; Fibrilación auricular; Complejo auricular prematuro; Latidos ectópicos; Machine Learning; Deep Learning.

Acknowledgments

This thesis would not have been possible without the trust of my supervisors Prof. Luca Mainardi and Prof. Valentina Corino. I feel very grateful for the opportunity and the time dedicated. I have also to thank all the MY-ATRIA network which has served as inspiration and support all along these years.

Thanks to all my B3 Lab colleagues, those that were there since the beginning and those that arrived at the end. Thank you for making going to the lab a reason for joy. Thank you for the music, the aperitivi, the complicity, the support and the laughter. Life is colourful with you.

Thank you to all my friends in Italy outside the Lab. Thanks to Paula and Gabriele who were my home in Milano most of the time these years. Thanks to my Spanish friends who, from the distance, have always made me feel accompanied and backed up.

Finalmente, quiero dar las gracias a mi familia. Han sido años revueltos. Hemos tenido que aprender a aceptar el cambio y a crecer con él. Gracias a mis tíos y mis primos, me siento más cercana a vosotros que nunca. Gracias a mi padre, mi madre, mi hermano, mi hermana (y a mi perro Congo!). Gracias a cada uno de vosotros de una manera diferente. Gracias por el apoyo y la guia.

A los que están y a los que
se fueron a mitad camino.

Contents

List of Figures

List of Tables

List of Acronyms

AAMI Association for the Advancement of Medical Instrumentation

AB Atrial bigeminy

AP Action potential

AF Atrial fibrillation

AFDB MIT-BIH Atrial Fibrillation Database

AFL Atrial flutter

AV Atrioventricular

AT Atrial tachycardia

AI Artificial intelligence

AVNRT AV Nodal Re-entry Tachycardia

AVRT Atrio-Ventricular Reentry Tachycardia

ANN Artificial neural network

Brady Bradycardia

BVP Blood volume pulse

BPM Beats per minute

CVDs Cardiovascular diseases

CNN Convolutional neural network

CM Challenge metric

DWT Discrete wavelet transform

ECG Electrocardiogram

ESC European Society of Cardiology

FDA Food and Drug Administration

FPR False positive rate

GHDx Global Health Data Exchange

GEH Global Electric Heterogeneity

HRV Heart rate variability

ICD Implantable cardioverter defibrillator

IRBBB Incomplete right bundle branch block

ILR Implantable loop recorder

LA Left atrium

LAA Left atrial appendage

LAnFB Left anterior fascicular block

LAD Left axis deviation

LDC Linear discriminant classifier

LBBB Left bundle branch block

LTSTDB Long-term ST Database

LQRSV Low QRS voltages

LV Left ventricle

LD Linear discriminant

LTAFDB Long-term Atrial Fibrillation Database

LTSTDB Long-term ST Database

LPR Prolonged PR interval

LQT Prolonged QT interval

MI Myocardial infarction

MAT Multifocal atrial tachycardia

MI Mutual information

MITDB MIT-BIH Arrhythmia Database

NMI Normalized mutual information

NSR Normal sinus rhythm

NSIVCB Nonspecific intraventricular conduction disorder

PAC Premature atrial complex

PR Pacing rhythm

PAF Paroxysmal atrial fibrillation

PPG Photoplethysmography

PPi Inter-systolic series

PPV Positive predictive value

PTT Pulse transit time

PV Pulmonary vein

PVC Premature ventricular complex

PWV Premature ventricular complex

QAb Q-wave abnormal

RAD right axis deviation

RRi RR interval

RA Right atrium

RF Random forest

RV Right ventricle

ROC Receiver operating characteristic

RMMSD Root mean square of successive differences

SA Sinus arrhythmia

SB sinus bradycardia

SE Shannon entropy

SVG Spatial ventricular gradient

SNR Signal-to-noise ratio

SVPB Supraventricular premature beats

SVDB Supraventricular Database

SVM Support vector machine

STach Sinus tachycardia

STE ST elevation

STD ST depression

SEB Squeeze-and-Excitation Block

TVCF Time varying coherence functions

TAb T-wave abnormal

TInv T-wave inversion

WRN Wide residual network

WT Wavelet transform

IAVB 1st degree AV block

Chapter 1

Introduction

1.1 Motivation

Atrial tachyarrhythmia comprehend an heterogeneous group of electrophysiological disorders affecting the depolarization pattern of the atrial tissue. Atrial fibrillation (AF) is the most common prevalent cardiac arrhythmia. It affects about 0.51% of global population according to the data analyzed from the Global Health Data Exchange (GHDx) database by [1]. It is characterized by a disorganized and abnormal depolarization of the atrial tissue and by irregularly irregular ventricular contractions. Although AF itself does not represent a life-threatening condition, it induces myocardial tissue degeneration and can lead to serious conditions as ischemic stroke or heart failure.

AF has gained a lot of attention in the past years due to its already high incidence and its expected increase over the next years. While AF prevalence in under 49 years-old population is estimated to be 0.12%-0.16%, it increases with age reaching up to 10%-17% in individuals of 80 years or older [2]. AF prevalence has suffered a 33% increase in the past 20 years [1] and with the ageing of the population due to the extended life expectancy, predictions are that it will continue to do so.

Typical symptoms of AF include palpitations (42%-55%), asthenia (15%-49%), dyspnea (24%-49%) and angina (10%-20%). About 10%-25% of AF cases remain asymptomatic. Untreated AF can produce blood clotting inside the atria which in the worst case may lead to stroke. Despite applying the correct treatments with all the currently available technology, AF recurrence rate ranges from 40% to 50% [3]. The high prevalence and incidence of AF, the life-long treatment, high recurrence rate and the clinical sequelae demands a high amount of resources and loads an important burden on healthcare systems.

Classic understanding of AF and stroke is that the impeded atrial contraction due to the aberrant electric impulse conductance on the myocardial tissue disrupts normal atria haemodynamics, favouring thrombi formation. However, it has been suggested that AF can also be a marker for thrombus risk associated with a diseased atrium rather than a direct cause of it. Although higher AF burden has been related with higher stroke risk the relationship between the minimum AF episode length necessary to increase stroke risk is unclear. Guidelines identify an AF diagnosis if an episode of a least 30 seconds is detected. However, this measure is not given by any clinically relevant reason but rather by a limitation of the available technology. The relation of AF 30 s with stroke is still controversial, with some studies as [4] suggesting they increase the risk for thromboembolic events.

Other atrial rhythm disturbances, from milder to more severe, include premature atrial complexes (PACs), focal atrial tachycardia (AT) multifocal atrial tachycardia (MAT) and atrial flutter (AFL). Although it is less common and its relation with stroke is not as established, AFL typically follows a similar treatment than AF. MAT and AT treatment instead, varies depending on the symptomatology and frequency. Differently, PACs have been typically considered benign and clinically irrelevant. While the focus of attention is on AF, the relationship of the rest of supraventricular tach-

yarrhythmia with stroke and myocardial tissue degeneration is still mostly unexplored.

About 25% to 30% of ischemic strokes remain unexplained (cryptogenic) [5]. One of the possible causes is that the thromboembolic events are caused by occult or silent AF. When AF is present without any perceived symptoms that enable its diagnosis, it is denominated *silent* AF. Prediction of the appearance of these episodes of AF could reduce the incidence rate of cryptogenic strokes. Despite typically being considered benign, several late studies link frequent PACs to first time appearance of AF [6–9]. Others have studied PACs as the possible direct reason for stroke [10]. Furthermore, frequent PACs have been studied as a measure of cardiac tissue deterioration [7, 10–12] and as a possible cause for left ventricular remodelling [13].

AF induces an electrophysiological, structural and mechanical remodelling since an early stage. It alters the atrial effective refractory periods (ERPs) increasing their dispersion, disrupts intra-atrial conduction and reduces its contractile function. Structural remodeling is initiated since the first week of AF, inducing progressively different changes as more homogeneous distribution of chromatin and a decrease in the myocardial protein cardiotin, among others. Even if some of the AF-induced changes seem reversible, studies as [14] pointed out at the effect of another cumulative factor induced by the altered atrial rhythm promoting AF persistence. The unknowns behind the promotion the arrhythmogenic substrate triggers doubts about how other atrial events as PACs may affect atrial myocardium.

Currently, patient treatment focuses on the prevention of thrombi formation and rhythm and rate restoration, but only once AF is detected. Assuming that the first AF episode detected could occur several months later than the first effective AF event, once treatment is initiated, important atrial remodeling has already taken place. Although much of the AF-induced atrial electrical and structural remodeling has been shown to be reversible, some structural changes may be irreversible. In addition, still there exist unknown factors involved in the development of permanent AF. Prediction of first time appearance of AF could lead to a prompt treatment, impeding further myocardial tissue deterioration and thromboembolic events.

In order to study the long term relationship of other atrial events with AF, there is a need to develop the adequate tools to monitor atrial tachyarrhythmia in order to asses and understand their interaction with AF and stroke. Typically, cardiac electrical activity is evaluated using the electro-

cardiogram (ECG). Different devices are present in the market for registering cardiac electrical activity in the short, medium and long term. Each of them have to compromise the quality of the recordings and the amount of information with the recording time. Normally longer recording periods imply less number of leads in order to increase portability and patient compliance. Short-term recordings include classic 12-lead ECG typically performed in a clinical setting for a duration of 30 seconds as well as event recorders. Medium-term includes 24-48 hour 2 or 3-lead Holter monitoring. Long-term recording is currently carried out majorly using 1-lead implantable loop recorders (ILR). Novel monitoring devices keep emerging in the market as biopatches, smartwatches and wristbands using photoplethismograph technology to monitor volumetric variations of blood circulation.

Given the low costs associated with the acquisition of ECG signals a high amount of these recordings are acquired everyday in the clinical setting. This includes short 30 second 12-lead ECGs and 24-48 h Holters. Hospitals normally store wide repositories with recordings of different types. However, although these signals contain a high amount of information, they require of manual annotation of trained personnel which is highly time consuming. Therefore, this annotation process has a high associated cost. As a result only a few complete databases have been made publicly available as those in PhysioNet repository [15]. This has limited the research activity to either private datasets which impedes replicability and is only available to a limited number of research groups. Other datasets have been made available recently [16–22] thanks to the PhysioNet Challenges of 2021-2022. Nevertheless, data and annotation strategy homogenization is not trivial and opens an important challenge for the drawing of meaningful conclusions from the potential studies on these databases.

Different number of leads provide different information: the higher the number the more the points of view on the cardiac electrical activity. In addition, the level of noise present in the signals vary depending if the recordings were acquired in an ambulatory or in a controlled clinical setting. Both scenarios provide information that could serve for different purposes and each of them needs of specific computational tools to be able to automatically analyze the high amount of data automatically and alleviate the workload from trained professionals. From an AF perspective, medium and long term recordings could serve to study the interaction of different atrial events and AF episodes. Many algorithms are present in literature to provide an automatic AF detection for single-lead ECGs [23–33].

However, most methods target only AF and miss other supraventricular events. The development of a technique capable of detecting other types of tachyarrhythmia adapted to the noise and computational limitations of constant monitoring devices could serve in the study of the occurrence of these episodes in relation to AF.

Given the noise limitation in single lead recordings, most of the techniques are based on the *RR* intervals. This impedes the study of single beats as PACs as their discrimination with respect to premature ventricular complexes (PVCs) based solely on the *RR* interval is rather challenging. In order to study single beats with relation to AF, a compromise should be reached between augmenting the number of leads and signal quality and between reducing the time of recording and increasing computational capacity, perhaps by performing an off line annotation. An automatic classifier capable of identifying single PACs and runs of supraventricular beats could enable the study of their relationship with AF and stroke. A better understanding of early abnormal atrial events as PACs and their role or maker information in the development of an arrhythmogenic substrate could mean a change of paradigm in how AF treatment is confronted, by the application of preventive therapies.

In this thesis different techniques have been developed for the automatic detection of AF together with other tachyarrhythmia. As mentioned above, different signals entail different information, therefore in this work we present a contribution for automatic tachyarrhythmia detection on signals acquired with one, two and twelve leads as well as providing tools for short, medium and long-term recordings. The first method intends to develop a classifier based solely on *RR* intervals, suitable for a constant single-lead monitoring device. The second methodology intends to provide a technique for automatically annotate ambulatory 2-lead Holter recordings to enable PAC automatic annotation and the study of their occurrence with relation to paroxysmal AF (PAF) and stroke on retrospective studies. The third method, focuses on short recordings acquired in a clinical setting. The information of signals acquired in a controlled setting with lower noise and higher number of leads is higher than that acquired in ambulatory conditions. Therefore, it is possible to develop techniques able to detect several types of cardiac disturbances. An automatic detector capable of annotate extensive databases could also enable the study of AF and other tachyarrhythmia in relation to other coexisting cardiac disorders.

1.2 Background: The Heart

1.2.1 Cardiac Anatomy

The *heart* is a muscular organ located in the mediastinum, between the lungs whose principal function is to pump blood into the circulatory system. It is formed by a bundle of muscle divided into four main chambers; two upper chambers named *atria* and two lower chambers named *ventricles*. While the right and left atria (RA, LA) are in charge of receiving blood from the vena Cava and the pulmonary veins (PVs), respectively, the right and left ventricles (RV, LV) pump it into the pulmonary artery and the aorta, respectively. The combination of left and right atria and ventricles constitute two independent pumps supplying blood to the *systemic* and *pulmonary* circulatory systems, respectively. The RA is connected to the RV through the tricuspid valve and the RV to the pulmonary artery through the pulmonary valve. On the other hand, the LA is connected to the LV through the mitral valve and the LV to the aorta through the aortic valve. The LV and LA are separated from the RV and RA by a muscular wall named *septum*.

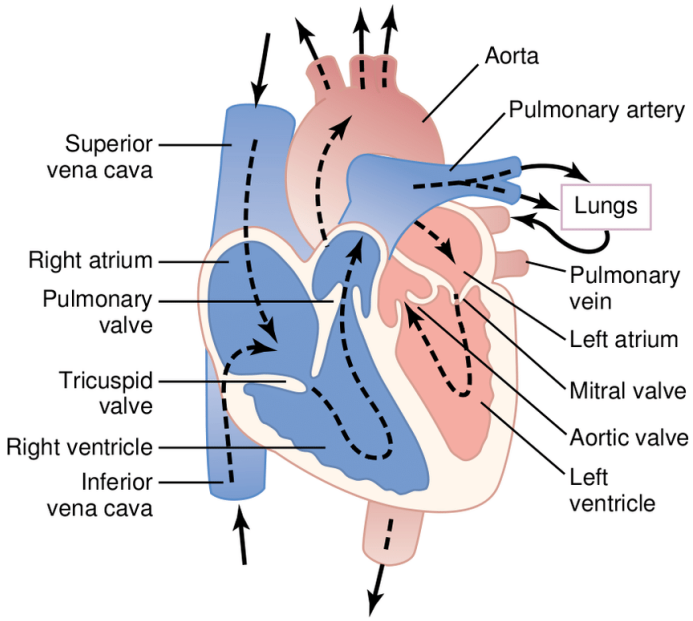

Figure 1.1: Schematic representation of cardiac anatomy (reproduced from [34]).

The heart is enclosed by the pericardium which holds it in place. Inside the pericardium, the cardiac wall is composed on three layers, the epicardium, myocardium and endocardium. While the epicardium and the endocardium play a structural role, the myocardium is formed by the cardiac muscle in charge of pumping blood into the circulatory system. Located beneath the endocardium, the conductive system plays an important role in propagating and coordinating the cardiac contraction. The conductive system is composed by a set of specialized conductive fibers. In addition, it contains to key elements in cardiac electrophysiology: the sinoatrial (SA) node, in charge of originating the electrical impulse, and the atrioventricular (AV) node, electrically connecting the atria and ventricles, the His bundle and the Purkinje fibers spreading through the ventricles. The AV node is connected to the His bundle which is then divided into two branches, the right and left bundles which innervate the RV and LV, respectively.

1.2.2 Cardiac Electrophysiology

The heart's capacity to coordinately contract is given by the organized propagation of an electric impulse. The main elements contributing to the conduction of this impulse are the cells composing the myocardium, the *myocytes*, and the conductive fibers. The cardiac impulse, named *action potential* (AP), is originated in the SA node and propagated at a high pace through the conductive fibers and at a slower rate through myocytes. Conductive fibers act as an AP "highway", enabling it to arrive to the different regions of the myocardium at a precise time to allow a coordinated contraction of the cardiac muscle. Myocytes react to this impulse by a fast depolarization of their membrane leading to their contraction and followed by a progressive repolarization.

Cellular level

In individual cells the AP is given by the movement of ion charges between the intracellular and extracellular space. During resting conditions the cell is negatively charged, generating a negative transmembrane potential which can be expressed as $V_m = V_{in} - V_{out}$. The main ion currents taking place in the membrane polarization and AP are given by Na^+, Ca^{+2} and K^+ ions. To describe process a cardiomyocyte goes through during the AP it can be divided into five different phases:

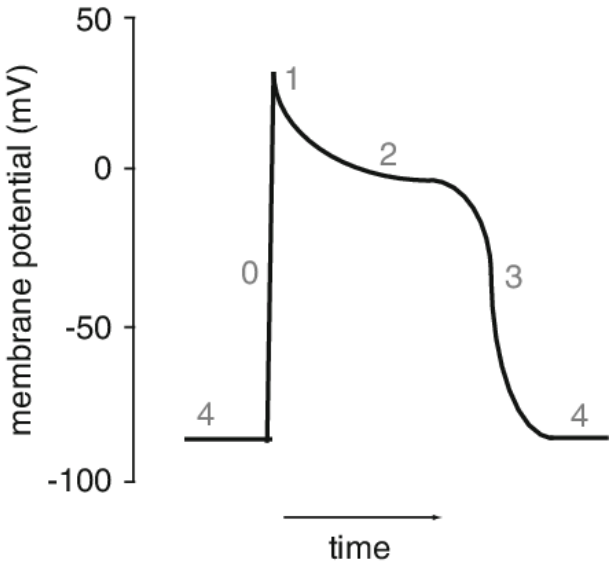

Figure 1.2: Representation of the four phases of a cardiac myocyte action potential (reproduced from [35]).

- Phase 4: During the diastolic phase, the membrane is at its negative resting potential, constant between -80 to -95 mV.
- Phase 0: In this phase cellular depolarization occurs. AP spreads through gap junctions arriving to the myocyte. Voltage-gated fast Na^+ channels open producing a rapid Na^+ inflow leading to membrane depolarization.
- Phase 1: Voltage-gated fast Na^+ channels close and a short repolarization is triggered. K^+ exits the cell creating and outward current. Both phenomena together lead to the appearance of a sharp brief negative voltage slope.
- Phase 2: This phase is denominated the Plateau phase. L-type voltage-mediated Ca^{2+} channels open generating a Ca^{2+} influx. This influx counterbalances the K^+ outflow, leading to a plateau at +50mV. Ca^{2+} influx induces calcium release from the sarcoplasmic reticulum. Ca^{2+} binds to troponin for actin-myosin cross-bridge activation.
- Phase 3: Repolarization takes place by closing of voltage-mediated Ca^{2+} channels and K^+ outflow through the delayed rectifier K^+ channels.

Organ level

In a healthy heart, the initial electrical impulse leading to the AP wave is initiated spontaneously in the SA node. AP propagates through the atria until reaching the AV node. The AV node is a complex structure containing two excitatory pathways that allow the transmission of the impulse. It acts as a gateway and filter of the incoming impulses from the atria. During its transition through the AV node, the electric impulse suffers a delay allowing the atria to contract and pump blood into the ventricles. The impulse then travels to the bundle of His, located in the septum dividing both ventricles, divides in both the right and left bundle branches and arrives to the Purkinje fibers. Purkinje fibers radiate throughout the ventricular tissue to generate a coordinated contraction of the ventricles. Given the high extension of the ventricles, without the fast conduction of the Purkinje fibers, a synchronize ventricular myocardium activation would not be possible. Ventricular depolarization thus initiates from the heart apex up to the atriventricular groove.

Cardiac cells can self stimulate and generate an AP in absence of an external stimuli. This property is called *automaticity*. Sinus node cells have a slightly higher discharge rate than other cardiac cells and thus typically initiate the AP wave. The rhythm marked by the SA node is named *sinus rhythm* (SR). The SA node is influenced by the *autonomic nervous system* (ANS) by *sympathetic* and *parasympathetic* innervation. The ANS regulate different cardiac functions, being *heart rate* one of them. While the sympathetic nervous system forms part of the "fight-or-flight" reaction increasing HR, the parasympathetic system is responsible for the "rest and digest" state, decreasing HR.

Due to the automaticity property of cardiac cells, occasionally a region of the myocardium other than the SA node develops a more rapid activation rate and takes over the pacemaker role of the SA node. The region initiation the AP is denominated *ectopic focus* and depending on its location, it can lead to an inefficient pumping of the blood. If the ectopic beat is located in the supraventricular region it is denominated *premature atrial contraction* or *supraventricular premature beats*. On the other hand, if the ectopic focus is located in the ventricles it is denominated *premature ventricular contractions*. Automaticity plays an important role in acting as a save net in case the SA node and or the AV node fail to initiate the stimulus. In that case an ectopic focus takes the role of initiating the AP. In this case the beat

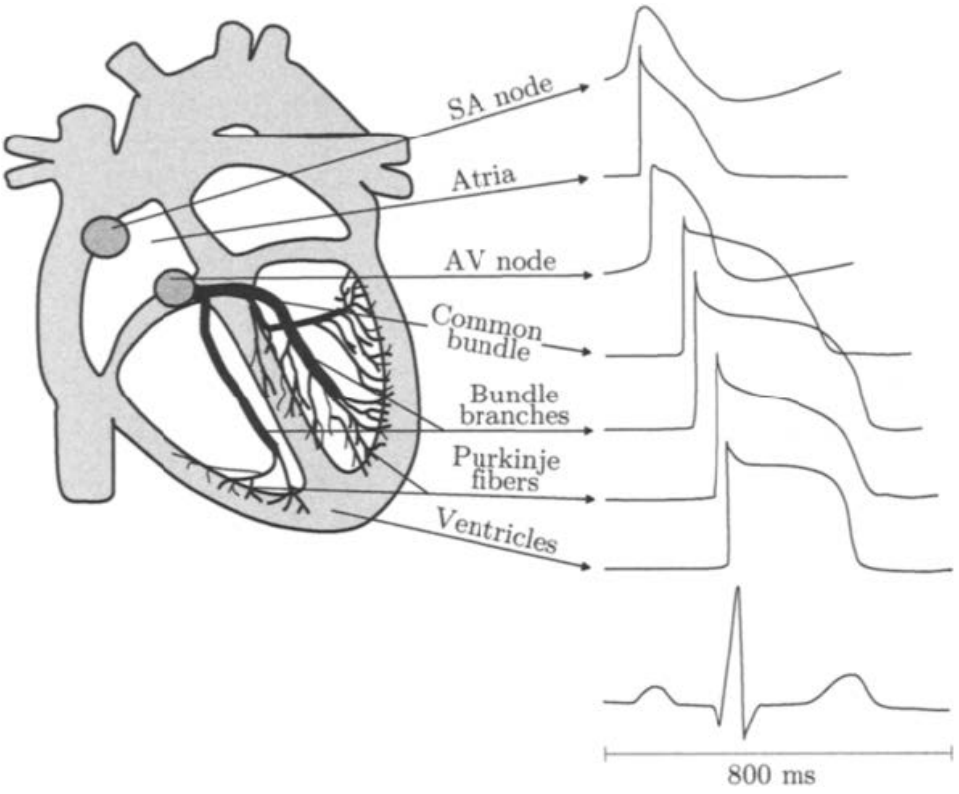

Figure 1.3: Cardiac conductive system and the action potential generated by each of its different parts (reproduced from [36]).

originated is called atrial or ventricular *escape beat*, depending it is located in the supraventricular or ventricular region.

1.2.3 The Electrocardiogram

The electrocardiogram (ECG) signal captures and describes the cardiac electrical activity on the body surface by the placing of electrodes on the skin. These electrodes measure the addition of all voltage variations occurring in the cardiac tissue over time. The resulting ECG signal is characterized by determined waves and segments representative of the different structures of cardiac depolarization and repolarization. Both the timing and the morphology of those waves and segments convey information regarding the heart's electrical activity and reflect possible diseases that induce distortions on the signal.

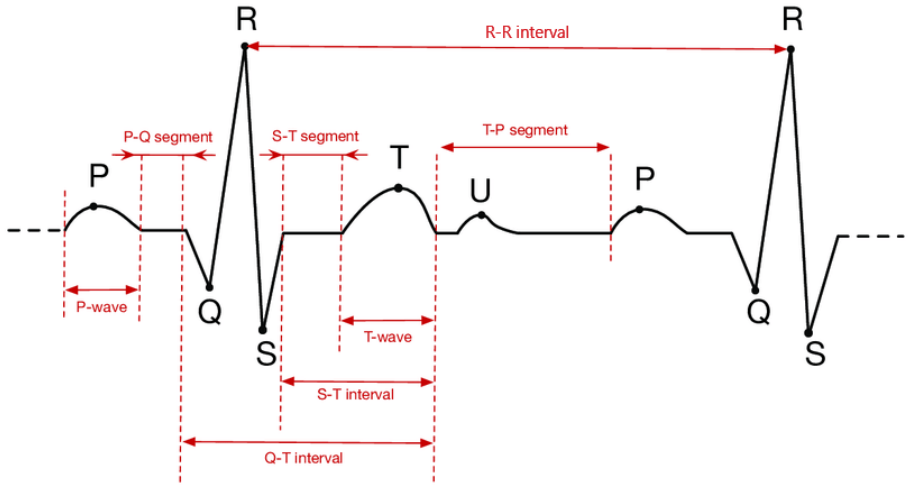

Figure 1.4: Graphic description of the different parts conforming an ECG signal during two normal sinus rhythm heartbeats (reproduced from [37]).

Depending of the information of interest, a different number of electrodes are used, and the measured potentials are combined so to obtain the ECG *lead*. Each lead provides a point of view of cardiac electrical activity. The morphology and polarity of each wave in the ECG signal depends on the electrodes position. A depolarizing wavefront moving towards the electrode would be translated as a positive deflection on the signal, while a current moving away would generate a negative slope.

In contrast, during repolarization the opposite scenario is found. Repolarization has an opposite polarity than depolatization so when moving towards an electrode, it generates a negative deflection and a positive one when moving away. As the resultant waves are the result of the addition of all cells during a certain point of the AP, the wave generated by the ventricles has a greater amplitude than that generated by the atrial, given their superior mass and dimension.

The typical physiological NSR beat is described by the P wave, the PQ interval, QRS complex, J point, ST segment, QT interval, T wave and RR interval as displayed in Figure 1.4. Each of these elements reflect an specific event:

- P wave: atrial depolarization. Typically presents an amplitude lower than 300 μV and a duration of less than 120 ms

- PR segment: delay between atrial and ventricular depolarization. It measures the speed of the AP transition through the AV node. A normal PQ interval ranges between 0.12 to 0.22 s
- QRS complex: ventricular depolarization. Typically of a duration less than 0.10 s. During this phase atrial repolarization also takes place but it is masked by the much larger amplitude of ventricular depolarization.
- J point: onset of ventricular repolarization
- ST segment and T wave: ventricular repolarization. There is no cardiac muscle activity during this wave.

Lead systems

Each ECG lead captures the electrical activity of the heart from a different axis projection. Consequently, each lead gathers a different spatial perspective or cardiac's electical activity. Two types of leads are used to register the ECG, the *unipolar* lead and the *bipolar* lead. While unipolar leads measure the voltage variation of a single electrode with respect to a reference, bipolar leads measures voltage differences between two different electrodes.

The choice of the number of electrodes and their position depends on the purpose and duration of the exploration. While in a clinical setting the most common configuration is the 12-lead ECG to perform a cardiac exploration for typically 10 seconds, for a prolonged ambulatory recording a reduced number of electrodes are used.

The Standard 12-Lead ECG

The 12-lead ECG is the most common setup used for cardiac electrical activity recording. It is acquired using 10 electrodes positioned in the limbs and chest. From the 12 leads, 6 are precordial or chest leads and the other 6 frontal leads. While the frontal leads measure the cardiac electrical activity on the frontal plane, precordial leads do it on the transversal plane.

Frontal lead consist on three bipolar limb leads and three augmented unipolar limb leads. The three bipolar limb leads are denoted as I, II and III. These leads are obtained by measuring the voltage differences between the left arm (V_{LA}), right arm (V_{RA}), and left leg (V_{LL}) following the relations:

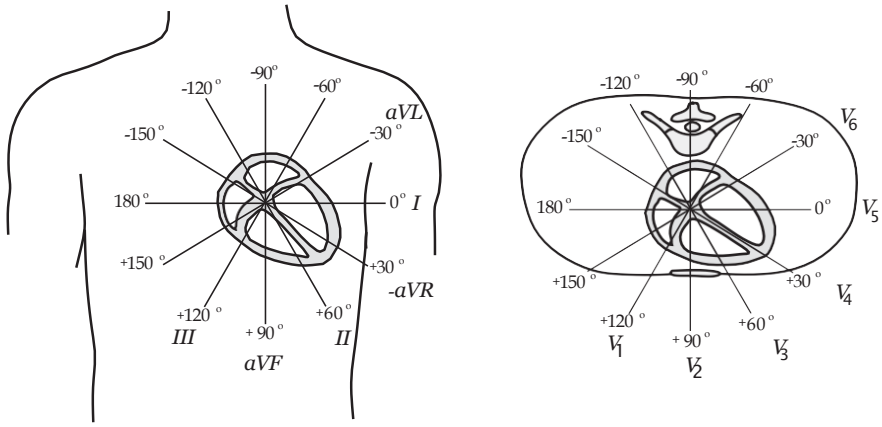

Figure 1.5: Lead angles reference system for the frontal (left) and horizontal (right) planes (reproduced from [36]).

$$I = V_{LA} - V_{RA} \tag{1.1}$$

$$II = V_{LL} - V_{RA} \tag{1.2}$$

$$III = V_{LL} - V_{LA} \tag{1.3}$$

These three leads form what is denominated the "Einthoven's triangle", which positions the heart at the centre of a triangle. Lead I observes the heart's activity from 0^{o}, lead II from $+60^{o}$ and lead III from $+120^{o}$ as observed in Figure 1.5.

Augmented unipolar leads instead are aVF, aVL and aVR are denoted as:

$$aVF = V_{LL} - \frac{V_{LA} + V_{RA}}{2} \tag{1.4}$$

$$aVL = V_{LA} - \frac{V_{RA} + V_{LL}}{2} \tag{1.5}$$

$$aVR = V_{RA} - \frac{V_{LA} + V_{LL}}{2} \tag{1.6}$$

They complement unipolar limb leads by providing a shifted 30^{o} angle view from each of them. Their corresponding angles are 90^{o}, -30^{o} and -150^{o} for aVF, aVL and aVR, respectively.

Precordial leads are all unipolar leads that measure the potential difference from one point in the chest to an imaginary central point denominated *Wilson Central Terminal*. This central point is calculated by averaging V_{LA}, V_{RA} and V_{LL}. Leads V1 and V2 face right's ventricle surface, they are positioned in the 4^{th} intercostal space to the right and left hand side of the sternum, respectively. V4 is placed in the 5^{th} intercostal space in the nipple line. V3 is placed between V2 and V4. V6 is placed in the midaxillary line, in the same horizontal line than V4.

Ortogonal Leads

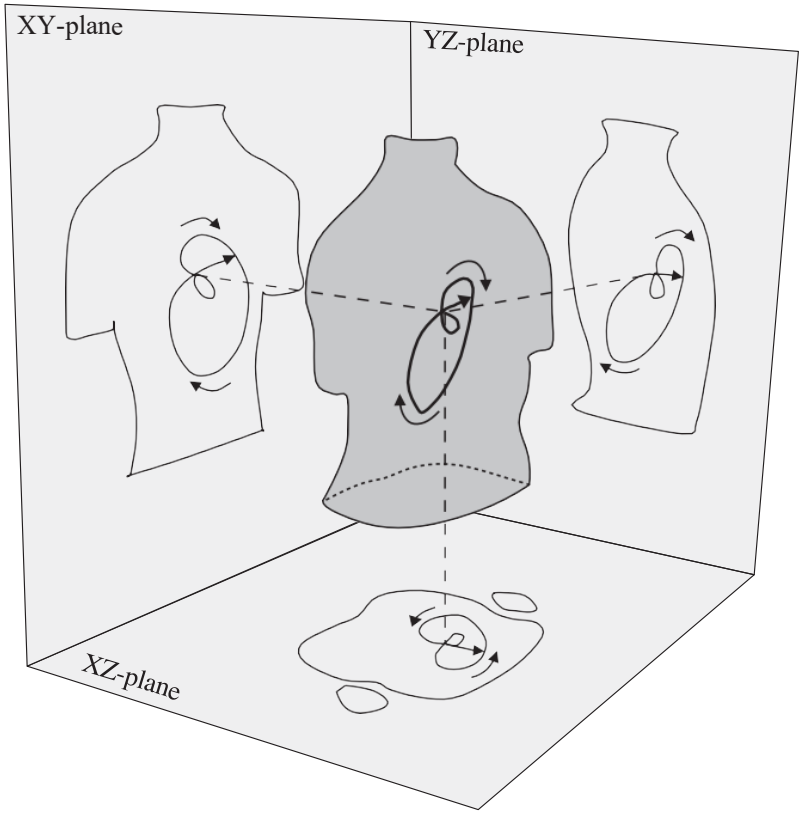

Figure 1.6: Ortogonal projections of the vectocardiographic loop (reproduced from [36]).

The ortogonal lead system projects the heart's electrical activity into the X, Y and Z planes. The combination of the different leads produce a 3D loop representation also referred to as *vectocardiogram* (VCG). The loop

is traced by the tip of the vector describing the dominant direction of the cardiac's wavefront.

The Frank lead system is the most common electrode disposition for the acquisition of the orthogonal projections. To obtain a left side, below and front view of the heart, this lead system uses 7 electrodes positioned in the chest, neck, back and left foot.

1.2.4 Atrial Tachyarrhythmias

Atrial tachyarrhythmias on the ECG

Atrial tachyarrhythmias comprehend a different number of pathological depolarization patterns in the atria. They comprehend AT, AFL, AF. Other supraventricular tachycardias also involve the AV node: AV Nodal Re-entry Tachycardia (AVNRT), Atrio-Ventricular Reentry Tachycardia (AVRT)- orthodromic and AV junctional tachycardia.

The simplest abnormal atrial event is the single premature atrial complex (PAC). PACs are beats triggered in a site of the atria other than the sinus node. They have been classically considered benign although some studies point at the possible relationship of frequent PAC, first time AF appearance and stroke. PACs in the ECG can be identified by an RR interval and a P-wave morphology alteration. The RR previous to the PAC is shortened while the subsequent RR interval is elongated. When two consecutive PACs are alternated with a normal sinoatrial beat it is denominated *atrial bigeminy* (AB). If instead three PACs are alternated with one sinoatrial beat it is denominated *atrial trigeminy*.

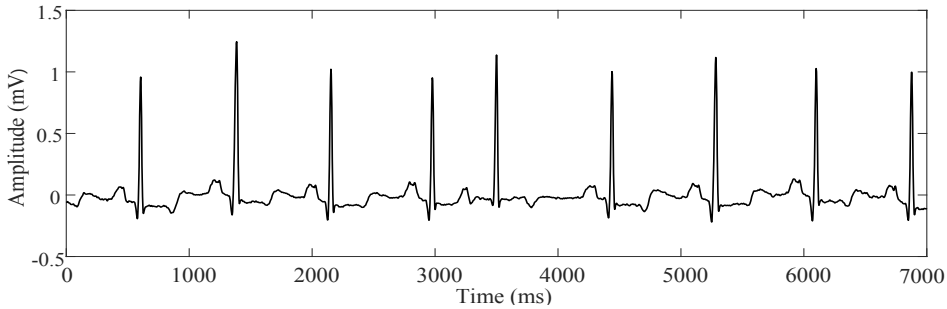

Figure 1.7: NSR ECG signal with single PAC. Data extracted from the MIT-BIH Arrhythmia database [15]

Atrial tachycardia (AT) is given when beats are triggered from a site other than the sinus node. It is characterized by a regular heart rate higher than 100 BPM. If there exist more than one foci in the atria triggering cardiac depolarization, it is denominated multifocal atrial tachycardia (MAT) and can lead to variable ventricular frequencies and P-wave morphologies.

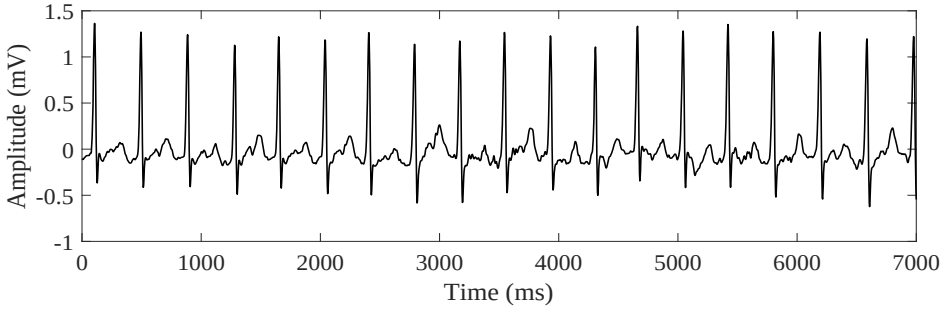

Figure 1.8: AT ECG signal. Data extracted from the MIT-BIH Arrhythmia database [15].

Atrial flutter (AFL) is characterized by a regular macroreentrant depolarization mechanism. During AFL the atria depolarizes following an organized circular pathway. Atrial rate ranges between 250 and 350 beats per minute (BPM) while ventricular rate ranges from 75-150 bmp. The AV node acts as filter, blocking atrial impulses in a 2:1, 3:1 or 4:1 ratio. During an AFL episode different block ratio can be alternated. The resulting pattern in the ECG is a sawtooth-like baseline oscillation caused by the atrial reentry mechanism. RR intervals regularity are subjected to the frequency of block ratio changes, in many cases leading to apparently irregular heart beats.

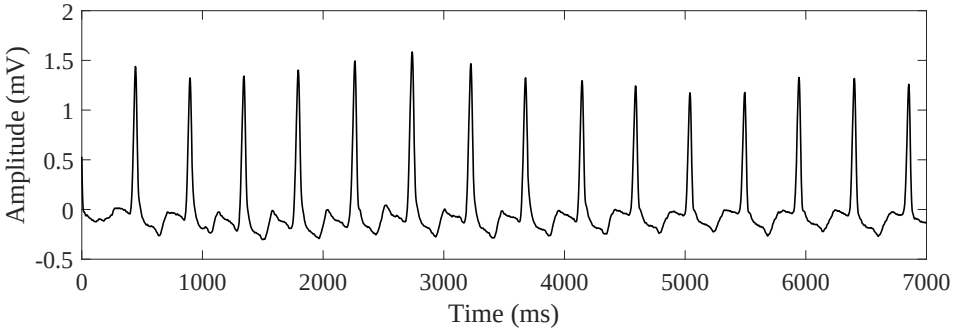

Figure 1.9: AFL ECG signal. Data extracted from the MIT-BIH Arrhythmia database [15].

Atrial fibrillation (AF) is the most common cardiac arrhythmia. During AF the atria depolarizes following a chaotic pattern, impeding proper atrial contraction. Atria depolarization rates can reach up to 600 bmp. During an AF episode, electrical impulses arrive fastly and disorganized to the AV node, that filters these APs resulting in a irregularly irregular RR intervals. Cardiac stroke volume is reduced during AF due to the rapid heart rate and the impeded atrial contractions. AF can be divided into different categories based on its duration and recurrence.

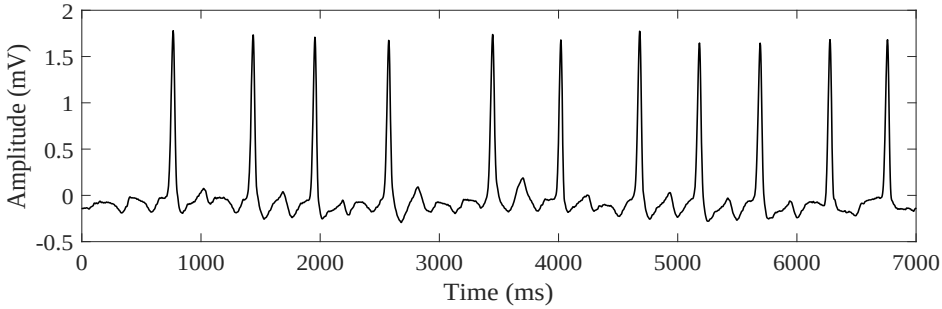

Figure 1.10: AF ECG signal. Data extracted from the MIT-BIH Arrhythmia database [15].

- New AF onset: 1^{st} episode occurrence regardless of the duration and symptomatology.
- Paroxysmal AF: recurrent episodes that self-terminate in a period shorter than 7 days, typically within 24 h.
- Persisting AF: episodes longer than 24 h that require of external intervention for their termination.
- Long-standing persistent AF: continuous incidences of AF that have lasted for at least 12 months. Rhythm strategy is initiated.
- Permanent AF: sinus rhythm cannot be restored and therapy aims at rate control.

Myocardial tissue remodeling

AF is known to worsen over time. Several studies point at AF itself as a cause for inducing changes in the atrial function and structure. Atrial tissue remodeling occurs at an electrical, contractile and structural level.

Electrical remodelling is given by a shortening of the atrial refractory period which increases AF stability. It was firstly introduced in 1995 by inducing AF through rapid atrial pacing in animal models. They found out that refractory period was reduced up to a 15%. Higher AF susceptibility was attributed to atrial impulse wavelength shortening. This may enable small intra-atrial conduction regions to serve as a reentry initiation site, increasing AF vulnerability. In addition, wavelength shortening may contribute to increase AF stability by allowing more wavefronts to coexist in the atrial tissue. The change in AF complexity due to the increased electrical remodelling was first distinguished by [38]. The ionic mechanisms behind this were posteriorly described in [39–41] among others. Changes in the refractory period including a missadaptation to rate changes were found to occur during the first 24 h, reaching a new stable state after a 2-3 days period. A reduced or poor rate adaptation of the refractory period was a marker for a *cryptic* atrial pathology, linked to atrial tachyarrhythmia susceptibility [42]. In humans, the ADP_{90} of patients suffering from AF or AFL was 130-150 ms shorter than that of healthy subjects. Nevertheless these electrical changes were found to be completely reversible, meaning that electrical remodelling is not itself the responsible for AF occurrences 1 week after sinus rhythm restoration. An still not well defined second factor could be involved in the long time development of AF substrate [14, 43]. Gap junctional protein expression changes or tissue fibrosis may be involved in this AF substrate creation. Heterogeneities in connexins may create microscopic obstacles for the depolarization wavefront promoting its fragmentation

Atrial contractile function is also affected during and after an AF episode. The reasons behind the depressed contractile function after an AF episode are still not well understood. Contractile dysfunction after a short AF episode may be given by cellular metabolic changes, whereas long-term episodes may induce additional changes contributing to the perpetuation of the contractility loss [42].

AF structural-induced changes in the atrial myocytes comprehend: cell size increase, glycogen perinuclear accumulation, sarcomeres central loss, connexin expression alteration, mitochondrial shape modification, sarcoplasmic reticulum fragmentation, nuclear chromatin homogeneous distribution and quality and position of structural cellular proteins changes. However, signs of AF irreversible changes causing cell apoptosis are not present in chronic lone AF [44]. Nevertheless, increased interstitial fibrosis levels are found in chronic AF patients.

The three AF-induced remodeling factors interact in a positive feedback loops. Since the first minutes of AF, already metabolic changes start taking place, reaching a new atrial refractory period steady state. Contractility is also altered starting with an hypercontractility state that is fastly transformed into an hypocontractility. Reversal of these effects takes a couple of days. The atrial refractory period is shortened during the first days of AF in about 40%-60%. Electrical remodeling facilitates AF stability. These changes have been shown to be reversible in a 3-day period of NSR. Atrial cellular structural remodeling goes on between the weeks 1 and 4 of the arrhythmia and after 4 months the amount of connective tissue has not increased but myocyte enlargement has led to a higher amount of intracellular connective tissue. Some of these structural changes were found to be reversible, whereas even after 16 weeks of NSR myocytes were still myolytic. In an attempt to study the involvement of a "second factor" on AF progression from paroxysmal to persistent AF episodes were induced consecutively leaving inbetween periods in which the atria was allowed to return to control. The time for electrical remodeling was found to be maintained in all episodes. However, the time needed for AF to become persistent was reduced after every AF episode [14].

Whereas electrical remodeling seems to be reversible, it structural remodeling appears might leave irreversible changes in the atrium. Although AF effects on the atrium have been more extensively studied, the effect of other atrial events as PAC runs and other AT in the develop if this arrhythmogenic substrate are still controversial. Nevertheless, several studies have found that frequent PACs are related to first time appearance of AF [5,7,10].

Atrial Tachyarrhythmia and Stroke

AF is known to be related to stroke, increasing up to 25% the risk of suffering a thromboembolic event. Most of these thrombi are thought to be developed in the left atrial appendage (LAA). However, the exact mechanism that influence in the development of clots inside the atrium during AF are still controversial. The classic theory relates changes in heamodynamics inside the atrium during AF in comparison to that of NSR. During AF, atrial contractions are impeded, leading to possible blood stagnation. Due to its morphology, the LAA is the atrial structure more prone for such stagnation. Thromboembolic risk is maintained even days after cardioversion perhaps

due to the maintained atrial contractility depression which takes some days to fade out.

In contrast, some suggest AF could be a marker for high stroke risk rather than a direct cause of it. The minimum episode length considered to increase stroke risk is controversial. While guidelines point at a minimum episode length of 30 s so that AF is diagnosed, this length is not based in any clinical reasons but rather a limitation of the available recording devices. Still about 25%-30% of strokes remain unexplained, some authors argue that many of these strokes could be caused by occult paroxysmal AF (PAF) episodes. However, other atrial events as frequent PACs have also been related to stroke in several studies. This relation remains still unclear and currently no measures are taken to monitor or treat PACs. A better understanding of the interaction of AF and stroke would entail also an study of how other atrial abnormal events, that could be an early marker of a diseased atrium, may influence thrombi formation.

Currently, AF is frequently diagnosed after a thromboembolic stroke. Given the severe consequences of this event further efforts should be made to be able to initiate anticoagulant treatment when risk for AF is detected. Still we are lacking the knowledge to be able to predict the first time occurrence of AF. Monitoring and a more in depth study of early signs of abnormal atrial activity (PACs) could provide an starting point for developing preventive therapies and avoid the life-long sequelae strokes may cause.

1.3 State of the Art: Atrial Tachyarrhythmia Detection

1.3.1 Recording Devices

The rapid development of new technologies has enabled the creation of different solutions to provide a cardiac monitoring adapted to different time and signal quality requirements: from short-term recordings of few minutes or seconds as 12-lead ECGs, to medium-term recordings from hours to a few days as Holters or biopatches, to long-term recordings that can monitor cardiac activity for years as implantable devices.

Monitoring and early AF diagnosis continues to gain attention given its increasing incidence. AF early diagnosis could help in the prevention of

thromboembolic events and lead to early treatment. A prolonged cardiac monitoring can enable the study of AF occurrence over time, its burden and its evolution. However, typically during ambulatory monitoring the signal quality is degraded as a lower number of leads are used and a higher number of noise sources are present, providing less information than signals acquired in a clinical setting. On the other hand, short-term monitoring can only capture cardiac events taking place during a very slim time window, and risk to obtain only a partial picture of the true cardiac health.

Different types of devices can fulfill a different objective to understand the cardiac electrical system state. Compromise between signal quality, comfortability for proper patient compliance, battery life, memory and computational power must be made for each device available.

Short-term (*seconds to minutes*)

Standard resting ECG
As detailed in Section 1.2.3, the standard resting ECG is composed of 12-leads and it is normally acquired in a clinical setting. It typically has a high signal quality from which both morphological and RR interval information can be extracted. However, given its short recording time (10-30 s) it can miss PAF episodes, PACs and other arrhythmia. Some examples of the devices used to acquire the 12-lead ECG are: Alirus 12-Lead ECG, Alirus USA, Nasiff CardioCard Mobile^{TM} Resting ECg, Nasiff Associates, Inc., USA, SmartHeatpro, SHL Telemedicine USA, ECG Glove AMD-3300, AMD Global Telemedicine, USA.

Handheld recorders
Handheld recorders normally have a recording time of 30 s approximately. They are portable devices that can be used by the patient routinely or when symptoms appear. The signal is acquired with a single lead and tends to have a low quality. These devices are useful to perform patient checks on a daily basis and register symptomatic episodes. However, as the monitoring time is limited, PAF asymptomatic episodes might not be detected. Some handheld recorders include MyDiagnostick, Zenico, Omro HCG-801, Merlin ECG recorder [45].

Smartphone-based devices
In 2020 it was estimated that there were 5.32 billion mobile phone users worldwide. Given their widespread use, their high computational resources,

portability, communication and data transmission characteristics, numerous solutions continue to emerge to adapt their use for healthcare purposes. According to the US Food and Drug Administration (FDA) [46] a medical application can transform a mobile platform into a medical device [47].

The first cardiac activity monitoring using smartphone was the acquisition of blood volume pulse (BVP) through photoplethysmography (PPG) using the smartphone's camera [48]. Other solutions emerged incorporating external electrodes to acquire single-lead ECG signals. The first model to incorporate this was KardiaMobile (AliveCor, Mountain View, CA, USA). Other models followed as ECG Check (Cardiac Designs, Round Rock, TX, USA). In both cases, the ECG was acquired on demand for 30 seconds and a signal similar to the one corresponding to Lead I was obtained. However, in the case of KardiaMobile, a software for automatic AF detection approved by the FDA was implemented, reporting results up to 98% and 97% sensitivity and specificity. Nevertheless, sensitivity values lowered to 55% and 79% in two independent studies with geriatric patients. The usefulness of using KardiaMobile for detecting undiagnosed AF was demonstrated in [49] in which 1.5% of the patients from a cohort of 1,000 pharmacy customers were newly diagnosed. Recently, a new model from AliveCor (KardiaMobile 6L) has been commercialized able to acquire 6-lead ECG signals using a portable device with 3 electrodes.

In the past years also smartwatches and wristbands have started to appear. Mainly two different types can be distinguished: those that use PPG technology and those that can acquire one-lead ECG signals. The first to appear were PPG wristbands able to constantly monitor heart rate, although due to the high content of noise, no FDA-cleared automatic AF detection algorithm have been yet implemented for constant monitoring. Only the Apple Watch recorded tachogram (plot of the RR series) has been granted the FDA approval for AF detection, although the PPG signal must be acquired at rest [50]. The smartwatches able to register single-lead ECG signals are the Apple Watch 4 (and subsequent versions), the Withings Scanwatch and ECG move and the Samsung Galaxy Watch 3 [45]. They register 30 s ECG on demand and have an FDA-approved algorithm for AF detection. Although these devices still cannot provide a continuous monitoring for AF detection, they are compact and portable and enable the user to perform more than one daily measurements or to check for AF if any symptom is perceived.

Other types of sensors can be found in the market as smart T-shirts or chest bands. Wearable technology continues to evolve rapidly and given its availability and affordability offers interesting perspectives for the screening and monitoring of AF patients.

Medium-term (*hours to days*)

Standard ambulatory monitoring

Standard ambulatory monitoring was initially introduced by Normal Holter in 1961. Holter monitoring consist on a portable device with typically 2 or 3 lead configuration that is used to monitor cardiac activity for a 24-48 hour time. After ECG acquisition period, the signals are analyzed offline for AF detection using commercial softwares. Even if these softwares have improved during the years still they require of manual revision by a trained professional to reject false negatives and positives.

Biopatches

Biopatches are a relatively new solution base on a one-lead ECG. In addition, they are able to collect other types of data as accelerometer data, skin temperature and respiration. They have been designed following a minimalist style, leadless and with reduced dimensions. They are water-resistant and are designed for a prolonged monitoring of several days. Due to their reduced size they can be placed in different torso regions, favouring noise reduction. Biopatches are single-use devices that acquire data for several days that is then analyzed offline or in semi-real time by the transmission of data into a gateway device that then sends data to the manufactures for analysis.

Two main biopatches are currently in the market, the Zio-Patch and the NUVANT system. The Zio-Patch offers two types of devices: one for at risk patients that creates a daily report and the standard version that provides only with a report at the end of the monitoring time. Both versions offer a maximum monitoring time of 14 days. Once data is collected, it is sent to the manufacturer for data analysis. Signals are first analyzed using heart rate and signal morphology and then the detected episodes are reviewed by a technician for false positives elimination.

The NUVANT system instead analyzes ECG signals in real-time. As the high risk Zio-Patch solution, the NUVANT system consist on a single-lead recording system and an independent transmitter that sends data to

the monitoring center. Whenever a rhythm abnormality is detected, data is transmitted and reviewed by certified technicians.

Some studies as [51, 52] proved the higher comfortability perception by patients and the higher rate of AF episodes detection with respect to classic ambulatory Holter due to the longer registration time. The improvement in design enable a higher patient compliance and signal quality, increases the AF episodes and cases detected. However, still the cost-effectiveness of these devices should be further investigated.

Cardiac event recorders
Cardiac event recorders are external small portable devices that typically record one-lead ECG signals. They can be continuous loop recorders or symptom event recorders.

Continuous loop recorders are constantly acquiring ECG signals and overwriting information due to memory constrains. If the algorithm or patient detect the onset of an event, the device stores the onset and end of the episode. Symptom event recorders can be worn temporarily when the patient detects arrhythmia symptoms. This makes them unsuitable for arrhythmia onset and asymptomatic events detection.

Cardiac event recorders have been demonstrated to be very prone to false positives due to ectopic beats. It was shown that in average each NSR patient had 5 false positives due to ectopic beats during a 24 hours recording time [53].

Long-term (*months to years*)

Implantable loop recorders
Implantable loop recorders (ILR) are an invasive alternative that provide a continuous monitoring for even several years. In many cases, ILR is implanted to monitor recurrences after atrial ablation or after a cryptogenic stroke. ILRs are being demonstrated to be superior to noninvasive techniques for PAF detection. However, given their high cost and the invasive procedure required, mass implantation of such devices is unrealistic. Therefore even if powerful, their benefits remain restricted to a reduced risk population.

Some ILR with AF detection algorithms are the Reveal XT (Medtronic, Minneapolis, MN, USA), SJM Confirm (St, jude Medical, St.Paul, MN,

USA) and Sleuth (Transoma Medical St.Paul, MN, USA). These devices measure cardiac activity subcutaneously using a single lead. Given their reduced size, their subcutaneous implantation and the invasiveness of the procedure, they have to deal with memory, energy and battery constrains. As continuous loop recorders, ILR are constantly recording and overwriting information. They can automatically detect AF episodes or be patient-activated when symptoms appear.

Reveal XT can also be programmed to AT detection in addition to AF. However, a very high rate of false positives were registered when this function was activated [54]. The minimum episode detected can be also tuned, being 2 min the shortest one. In an study with 247 subjects, sensitivity and positive predictive value increased from 88.2% and 73.5% to 92.1% and 79.6% respectively, when the minimum AF episode detected was increased from 2 to 6 minutes. Therefore, the 6-minute option is typically used in a clinical setting. False positives were given by pectoral muscle activity, ectopic beats, false QRS detection and T wave oversensing in a 35%, 15%, 4% and 1.5% respectively [55].

Pacemakers
Implantable pacemakers are originally used to induce ventricular depolarization at the right time. They deliver an electrical stimuli causing ventricular contraction. However these devices can also be used to monitor cardiac activity directly measuring intracardiac signals.

Cardioverters-defibrillators
Implantable carioverters-defribrillators (ICD) are invasive devices designed to directly treat ventricular tachyarrhythmia. It constantly monitors cardiac activity and when an arrhythmia is detected an electric shock is delivered to guarantee blood supply from ventricular contractions. As pacemakers, these devices can monitor atrial activity by measuring intracavitari atrial signals.

1.3.2 Detection and Classification Computational Techniques

Automatic ECG annotation has attracted many attention since the development of ambulatory recording devices. Manual signal annotation is highly time consuming and in some cases as in ILR it must actually be done automatically.

Cardiac abnormalities can affect the ECG's waves and peaks morphology or intervals. The most studied interval is the RR one, as it represents the rate and regularity at which ventricles depolarize. In addition, given the larger R peaks amplitude with respect to other ECG segments, automatic R peak detection is more resistant to noise than other ECG regions. Although wave morphology entail more detailed information about cardiac conductive system, it is easily corrupted by noise and in many cases inaccessible.

In the case of atrial tachyarrhythmia both a P-wave morphology distortion and an RR interval disruption is produced. During AF, P-waves are substituted by f-waves and RR intervals follow an irregularly irregular pattern. During AFL, P-wave is substituted by a sawtooth like patter. RR intervals are regular although may have alternating blocks. AT induce morphology changes in the P-wave and induce fast regular ventricular contractions. In contrast with AFL and AF, AT morphological differences with respect to NSR are more subtle. As AT, individual PACs typically induce subtle P-wave modifications and a disruption of the RR interval.

A wide variety of algorithms are present in literature for arrhythmia detection, specially for AF. Depending on the objective of the detection: long-term constant monitoring, offline Holter annotation, retrospective database annotation for clinical research or population mass screening. Different objectives entail the use of different signals with their corresponding noise levels and different computational constrains. We can divide the algorithms present in literature into those using only the RR interval analysis to detect tachyarrhymia and those that also use morphological information.

Comparison between the different algorithms present in literature is not trivial. To do so algorithms should be tested in the same database, in order to provide a common framework. In addition, other factors as minimum episode length and their capacity of detecting other cardiac events ought to be taken into consideration.

PhysioNet open source databases [15] establish that common framework that allow research groups all over the globe to test their methodologies and compare them with the state of the art. The Association for the Advancement of Medical Instrumentation (AAMI) guidelines set the MIT-BIH Arrhythmia database available at Physionet as a common framework for reporting performance as it is the only one that contains the five superclasses of arrhythmias.

Rhythm-based Algorithms

As P-wave and f-wave analysis is difficult at low signal-to-noise ratios (SNRs), most AF detectors are based on parameters describing RR intervals regularity. In addition, the computational costs of RR information extraction is much lower than that of morphological data extraction. Specially, one-lead constant monitoring devices take advantage of this. Most classifiers are designed using simple parameters to describe AF irregularity. However, while they have turned out to be very effective in AF detection, most of them are unable to detect other types of tachyarrhythmia. Comparison among AF detectors has been typically done by evaluating their performance on the AFDB [15]. However, the MIT-BIH Atrial Fibrillation Database (AFDB) contains just annotations for AF and in a low proportion AV junctional rhythm and AFL. This hinders the evaluation of the different algorithms in presence of other tachyarrhythmias.

As in [56], we can divide the algorithms developed for AF detection using RR intervals into those using statistical dispersion coefficients [23–26, 33], entropy [24, 27–31], symbolic dynamics [29, 57], Poincaré plot-based parameters [54, 58, 59] and time-varying coherence function [28]. Just a few studies attempted to detect both AF and other tachyarrhythmia [27, 60] reporting poor results. In [54] different parameters were calculated from the Poincaré plot representation of the RR. The performance reported went from a mean, 5th percentile, and lowest value of 99.0%, 98.0% and 95.0% for duration sensitivity (DS_n) on a dataset containing only AF and NSR episodes, to a mean, 5th percentile, and lowest value of 80.3%, 68.0% and 48.0% for DS_n in a dataset containing also AT episodes. In [27] they used the coefficient of sample entropy and expressed their impossibility in detecting other tachyarrythmia as AFL. They considered AF and AFL of the AFDB as the same category and they showed how the are under the receiver operating characteristic (ROC) curve of their model increased when the AFL patients were removed from the test set from 0.928 to 0.955.

In contrast ICDs have been shown to be able to differentiate between AF and AT [61–63]. This is because the combination of an intracavitary atrial electrode and a ventricular one, enables to study the atrial-vetricular activation time which enables the discrimination between AF and AT. Nevertheless, this analysis is restricted to patients with ICDs, patients whose therapy depends on the automatic annotation of atrial tachyarrhythmias as ILR still cannot take other tachyarrhythmia into consideration. Some stud-

ies [64] have pointed out at the importance of AT monitoring specially in risk patients, still a detector able to monitor effectively several tachyarrhythmia is missing.

Morphology-based Algorithms

Single beat classification

The addition of morphological information enables also to perform a more detailed annotation enabling even to target beats rather than AT episodes. Several beat classifiers can be found in literature that divide into supraventricular (S), ventricular (V) and normal (N) beats [65–67, 67–72]. However, the understanding and comparison of their performance is not trivial due to the unbalance nature of the available open-source databases and the differences on how these are used in each study. Many other classifiers present in literature cannot even be taken into consideration as the same patients' beats were included both in the training and in the test set and [66] explored how this strongly biased results.

Several strategies have been used for beat classification, combining RR features with morphological features. Some of these are gathered in Table 1.1. [67] used RR intervals and morphological information of the segmented ECG as features and linear discriminant (LDs) models as classifier. They obtained an S sensitivity of 75.9%, a positive predictive value (PPV) of 38.5% and a false positive rate (FPR) of 4.7% on a subset of the MIT-BIH Arrhythmia database (DB2). In [68] they used features proposed in literature and applied a support vector machine (SVM) algorithm, artificial neural networks (ANNs), Bayesian classifier and an optimum path forest (OPF) for classification. The algorithm achieving the highest S sensitivity was a SVM achieving a sensitivity of 12.3% and 72.4% and a specificity of 95.5% and 88.8% for S and V respectively in the DB2 subset of the MIT-BIH Arrhythmia database. [69] used a combined SVM with RR and morphological information obtaining an overall accuracy of 86.66%, sensitivities of 88.94%, 79.06% and 85.48% and PPVs of 98.98%, 35.98% and 92.75% for N, S and V, respectively. In two studies, [65, 66] developed a classifier including RR interval and morphological features from different scales of the discrete wavelet transform (DWT) and a LD classifier (LDC). They obtained an S sensitivity of 89% and a PPV of 88% in the DB2 of the MIT-BIH Arrhythmia database. However, these performances worsen considerably when the algorithm was tested in the full database, obtaining

a sensitivity and a PPV of 76% and 43% and further degraded when tested in other databases as the Long-term ST Database (LTSTDB), where they obtained 50% and 8% sensitivity and PPV.

Table 1.1: Performance of state-of-the-art beat classifiers.

Study	Year	Database	S			V		
			Se.	Sp.	PPV	Se	Sp	PPV
De Chazal et al. [67]	2004	DB2	75.9%	-	38.5%	77.7%	-	81.9%
Luz et al. [68]	2013	DB2	18.3%	95.5%	-	72.4%	88.8%	-
Zhang et al. [69]	2014	MITDB	79.06%	-	35.98%	85.48%	-	92.75%
Llamedo et al. [66]	2012	DB2	89%	-	88%	90%	-	97%
Llamedo et al. [66]	2012	MITDB	76%	-	43%	80%	-	82%
Llamedo et al. [66]	2012	LTSTDB	50%	-	8%	66%	-	36%

where DB2 is the DB2 of the MIT-BIH Arrhythmia database and MITDB is the full MIT-BIH Arrhythmia database.

The performance changes pointed by [66] point at the biased use of just a subset of the MIT-BIH Arrhythmia to evaluate beat classification, specially for S detection where performance is lower.

Tachyarrhythmia episodes detection
Morphological information adds relevant data to tachyarrhythmia detectors as it can account for P-wave shape and presence of f-waves among others. A considerable amount of papers can be found in literature addressing AF detection. In addition, morphological data enabled others as [32, 33], to study atrial signals and distinguish between NSR, AFL and AT. In [73] an accuracy of 89%, 87% and 91% was obtained for AFL, AT and NSR in 22 AFL and 10 AT patients with 10-hours Holter monitoring.

However, morphological data is more easily corrupted by noise and in many cases the introduction of morphological data for AF detection has not lead to performances superior to the ones reported using just RR intervals [74–77]. The introduction of noise detection techniques could lead to an increase in performance.

Multiple cardiac disorders detection
With the fast development of deep learning, new solutions are being proposed that aim to automatically annotate not only cardiac tachyarrhythmia, but a whole range of cardiac disorders. The problem behind a deep learning approach is that complex networks require of huge amount of data to be

properly trained. Current open access annotated databases either do not contain enough signals for each pathology, or are composed by a low number of patients. However, this has changed with the PhysioNet challenges of 2020 and 2021 were large annotated datasets were made public [16–22].

These databases contained annotations for 111 categories, although the algorithms developed as a result of the challenge targeted only 24: 1st degree AV block (IAVB), AF, AFL, bradycardia (Brady), incomplete right bundle branch block (IRBBB), left anterior fascicular block (LAnFB), left axis deviation (LAD), left bundle branch block (LBBB), low QRS voltages (LQRSV), nonspecific intraventricular conduction disorder (NSIVCB), pacing rhythm (PR), PAC, PVC, prolonged PR interval (LPR), prolonged QT interval (LQT), Q-wave abnormal (QAb), right axis deviation (RAD), sinus arrhythmia (SA), sinus bradycardia (SB), normal sinus rhythm (NSR), sinus tachycardia (STach), T-wave abnormal (TAb) and T-wave inversion (TInv) In contrast with the other classifiers discussed above, the detection of multiple cardiac disorders inducing subtle morphological distortions in the ECG traces requires of higher quality signals.

Scores resulting of algorithms developed for the challenge, followed a metric specifically developed to evaluate an algorithm's performance for cardiac disorders detection. The scope of the metric was to provide a true clinical representation of the model's output by providing full credit to correct diagnosis and partial credit to misclassifications with similar risk as the true diagnosis. Considering $C = [c_i]$ as a collection of diagnoses, the multiclass confusion matrix $A = [a_{ij}]$ can be computed. Where a_{ij} is the number of recordings in a database classified as class c_i but actually belonging to class c_j. Different weights $W = [w_{ij}]$ were assigned to the different positions of the matrix A. The final score is computed as:

$$s = \sum_{ij} w_{ij} a_{ij} \tag{1.7}$$

The score metric was then normalized so that 1 represented that all the true classes were detected and 0 that only the normal rhythm was detected.

In [78] they used a custom ResNet with large convolutional sizes and a multi-head attention mechanism. The output vector was concatenated with a binary ECG indicator and classified by fully connected layers. The CM obtained in the validation and test set was 0.69 and 0.58 respectively,

achieving the 1^{st} position of the challenge. In [79] they used a wide residual network (WRN) with 14 convolutional layers and 1 widening factor integrating a squeeze and excitation block. Demographic data (age and sex) was included in the dense layer of the output stem. They achieved the 2^{nd} challenge position with a validation and test CM of 0.594 and 0.570. In [80] they used two different inputs: the preprocessed ECG and their fast fourier transform introduced in independent convolutional neural networks (CNNs) their outputs were concatenated and introduced into a fully connected block that provided with the final output. The obtained the 3^{rd} position in the challenge with a CM of 0.622 and 0.546 in the validation and test sets respectively.

1.4 Objectives and Outline

Technology is rapidly evolving to provide users with different solutions for cardiac monitoring. However, work still needs to be done to analyze all the information gathered and to understand its relevance. In this thesis different algorithms have been developed with the aim of providing tools for the analysis of signals containing a variable degree of information. This work has confronted the problematic of developing an automatic classifier for a constant monitoring device, the offline single beat annotation of Holter signals and the offline automatic annotation of short ECG signal segments with variate number of leads for different cardiac disturbances, both temporal and permanent.

Chapter 2:
This study focused on the development of a technique for the detection of AF and other tachyarrhythmia based only on the *RR* sequence. It posed a new approach towards cardiac tachyarrhythmia classification by transforming RRi into images. Differently than current methodologies, the proposed model could be generalized to different rhythms and, in contrast with AF-focused methods, it did not define rhythm-specific parameters and thresholds.

- **G. García-Isla**, V. D. A. Corino, and L. T. Mainardi, "Cardiac Tachyarrhythmia Detection by Poincaré Plot-Based Image Analysis", in Computing in Cardiology, 2019.

- **G. García-Isla**, V. D. A. Corino, and L. T. Mainardi, "Poincaré Image-Based Atrial Fibrillation Detection for Photoplethysmography Signals," in Gruppo Nazioneale di Bioingegneria (GNB) proceedings, 2021.
- **G. García-Isla**, V.D.A. Corino, and L. T. Mainardi, "Poincaré Plot Image and Rhythm-Specific Atlas for Atrial Bigeminy and Atrial Fibrillation Detection", in IEEE Journal of Biomedical and Health Informatics, 2020

Chapter 3:
In this chapter, we focused on the occurrence of single beats instead of on the detection of whole tachyarrhythmia episodes. While the development of a tachyarrhythmia detector aims to monitor and detect an abnormal rhythm already occurring in the heart, the study of single abnormal beats triggered in a site other than the sinus node could aid in the understanding and prediction of the development of such tachyarrhythmia. In this work we used artificial intelligence techniques to develop a beat classifier with emphasis on PAC detection. Two complementary classifiers were developed, one segment-wise using deep learning and one beat classifier that used a machine learning ensemble model.

- **G. García-Isla**, L. T. Mainardi, and V. D. A. Corino, "A Poincaré Image-Based Detector of ECG Segments Containing Atrial and Ventricular Beats" 2021 Computing in Cardiology. IEEE, 2021.
- **G. García-Isla**, L. T. Mainardi, and V. D. A. Corino. "A Detector for Premature Atrial and Ventricular Complexes." Frontiers in Physiology 12 (2021).

Chapter 4:
This chapter aimed to develop artificial intelligence solutions for the automatic annotation of multiple cardiac eletrophysiological disturbances on ECG signals. A machine learning model ensemble was developed for the detection of 9 different ECG alterations. Furthermore, a deep learning algorithm was developed for the detection of 26 different cardiac abnormalities in ECG signals of variate lengths and leads. In addition, analysis of the best training strategy for a model integrating both a deep network and classic machine learning features was explored.

- **G. García-Isla**, R. Laureanti,V. D. A. Corino, and L. T. Mainardi, "ECG Morphological Decomposition for Automatic Rhythm Identification." 2020 Computing in Cardiology. IEEE, 2020.
- S. Magni*, A. Sansonetti*, C. Salvi, T. Tabiadon, **G. García-Isla**. "Combining ResNet Model with Handcrafted Temporal Features for ECG Classification with Varying Number of Leads " 2021 Computing in Cardiology. IEEE, 2021.
- **G. García-Isla**, V.D.A. Corino, and L. T. Mainardi, "Poincaré Plot Image and Rhythm-Specific Atlas for Atrial Bigeminy and Atrial Fibrillation Detection", in IEEE Journal of Biomedical and Health Informatics, 2020
- **García-Isla, G.***, Muscato, F.*, Sansonetti, A., Magni, S., Corino, V., Mainardi, L. Ensemble Classification Combining ResNet and Handcrafted Features with Three-Steps Training. Physiological Measurement, 2022

In addition, this work was awarded with:

- Best Poster Award of its Session. Computing in Cardiology 2019. For the work: "Cardiac Tachyarrhythmia Detection by Poincaré Plot-Based Image Analysis,"
- Best Poster Award of the PhysioNet/CinC Challenge 2020 for the work: "ECG Morphological Decomposition for Automatic Rhythm Identification". With the team *Germinating*.
- Best Team Name of the PhysioNet/CinC Challenge 2021. For the work: "Combining ResNet Model with Handcrafted Temporal Features for ECG Classification with Varying Number of Leads". With the Team *PhysioNauts*.

Chapter 2

Arrhythmia Detection on Single-Lead Constant Monitoring Devices

2.1 Motivation

Although supraventricular tachyarrhythmias do not represent an explicit life-threatening risk, they are associated with high mortality and increased morbidity. Atrial tachyarrhythmias like AF, AFL and AT are linked to thromboembolic events and myocardial tissue deterioration [81]. Their monitoring is thus, important to evaluate their occurrence and burden in order to apply the necessary treatment and follow its effectiveness. Nevertheless, many PAF episodes remain asymptomatic, which makes their detection challenging [82]. Atrial tachyarrhythmia mass screening could help diagnose patients on time. The initiation of an appropriate treatment could prevent myocardial tissue deterioration, an unexpected thromboembolic event or even cardiac failure.

Ambulatory ECG devices as 24 or 48 hours Holter monitoring are only used when a risk is suspected or after an anomalous event, as they require of manual data revision by a cardiologist. Furthermore, they also do not guarantee that a paroxysmal event can be captured during the 24 or 48 hours recording [83]. Likewise, event recorders are typically based just on short 30 s registration, which makes the detection of asymptomatic paroxysmal AF episodes rather difficult. ILR are currently the option that offers the longest non-interrupted monitoring. Other devices as single-lead surface recorders have also been developed to attain a more extended monitoring than classical 24-48 hour ambulatory Holters.

Both types of devices are limited either by signal quality or by computational power, memory and battery. Atrial signals, are more susceptible to noise and they typically require of computationally expensive methods for noise and ventricular activity cancellation. Arrhythmia detection algorithms based on wave morphology are still unrealistic for signals' quality and hardware restrictions of currently available devices. Therefore, they normally rely mainly on cardiac rhythm parameters by R peak detection, more resistant against noise interference [27]. On the other hand, the emerging noninvasive PPG technology, integrated in smartphones, wristbands and smartwatches could help in the early detection of PAF episodes accessible to the general public [84]. Being a cheap technology these signals could be used for AF mass screening. After identification of systolic peaks position, the inter-systolic series (PPi) can be derived and analyzed similarly to the RR interval series [85]. While RR intervals measures the time between R peaks, PPi measures the time between each pulse waves. There exists a de-

lay between R and pulse peak denominated pulse transit time (PTT) which is negatively correlated with age, arterial stiffness and blood pressure [86]. Although PPi and RRi share common patterns, their underlying physiological differences lead to substantial HRV and pulse wave variability (PWV) (PPi variability) differences [87].

Many algorithms have been proposed for AF detection using only RRis variability; based on statistical dispersion, entropy, symbolic dynamics, time-varying coherence function and Poincaré plot [83]. Many of these studies have proven to be very effective for the detection of AF. However, all of them focused on characterizing and developing a model specific for AF, unable to detect other atrial rhythms [27, 31, 54, 57, 88]. To the best of our knowledge, there is a lack of methodologies based solely on *RRi*s capable of identifying tachyarrhythmias other than AF. Nevertheless, also AFL is known to be linked with stroke [89, 90] and AFL, AT and MAT with cardiomyopathy caused by rapid atrioventricular conduction. Other atrial electrophysiological disturbances as frequent PAC previously considered clinically irrelevant, have also been linked to higher stroke incidence, myocardial tissue degradation and AF development [7, 10–12]. In addition, maintained AB has been shown to promote left ventricular remodelling and deterioration [13]. These findings suggest to pay more attention to these conditions whose prevalence should not be ignored. Monitoring exclusively AF may underestimate the influence of other types of tachyarrhythmias and lead to an incomplete evaluation of the cardiac state. Furthermore, frequent PACs and other rhythms that alter the *RRi* sequence are known to hinder AF detectors as they are confused with AF, increasing false positive rates [83, 91]. Targeting them directly would improve AF detection.

This study poses a new approach towards cardiac tachyarrhythmia detection by transforming RRi into images. Differently than current methodologies, the proposed model can be generalized to different rhythms and, in contrast with AF-focused methods, it does not define rhythm-specific parameters and thresholds. The Poincaré plot is a graphical representation that describes the *RRi*s non-linear relationships by plotting each *RRi* against its previous one. Depending on the underlying cardiac rhythm, the plot generated by the RR or δRR intervals follows a diverse pattern that can be visually recognized as described previously in [92, 93]. The proposed methodology is based on the observation that different rhythms generate different patterns in the plot, and that these can be captured and used to classify other *RRi*s. Past studies used Poincaré plots for AF detection

for defining parameters and thresholds to describe particular patters in the plot [58–60, 94, 95]. In this study, we introduce two new versions of the Poincaré plot to which we refer as Poincaré Images and Poincaré Atlases and we explore the possibility of identifying and classifying NSR, AB and AF ECG segments. In addition, the possibility of extending the mentioned methodology to PPG signals for AF detection is explored.

2.2 Materials and Methods

The method adopted is based on the analysis of RR intervals and their representation through the Poincaré plots, i.e. the 2-D graph of the current RR versus the previous one. The plot provides a representation of the RR interval dispersion within the RR series and, in a non-linear dynamics view, is the two-dimensional (2-D) reconstructed RR interval phase-space, which is a projection of the reconstructed attractor describing the dynamics of the cardiac system [96, 97]. In this study, the Poincaré plot is treated as an intensity image by introducing the concept of *Poincaré Image* and further computations are performed on it.

2.2.1 The Poincaré Image

Let's consider the pairs of two consecutive RR intervals (RR_n, RR_{n+1}), being $n \in \{1, 2, \ldots, N-1\}$ with N the number of samples in the considered epoch. The Poincaré Image, $P[j,k]$ is a pixelated version of the traditional, continuous Poincaré plot and is defined as follows

$$P[j,k] = \sum_{n=1}^{N-1} \mathbf{1}_{[j,\Delta x]}(RR_n) \cdot \mathbf{1}_{[k,\Delta y]}(RR_{n+1}) \tag{2.1}$$

where $\mathbf{1}_{k,\Delta x}(\cdot)$ is the indicator function

$$\mathbf{1}_{[m,\Delta v]}(x) := \begin{cases} 1 & \text{if } x \in [m\Delta v, (m+1)\Delta v)\,, \\ 0 & \text{if } x \notin [m\Delta v, (m+1)\Delta v)\,. \end{cases} \tag{2.2}$$

and where k, j are the bin indexes and Δx and Δy are the selected bin size, respectively. Thus, $P[j,k]$ is the counting of the number of times RR_n

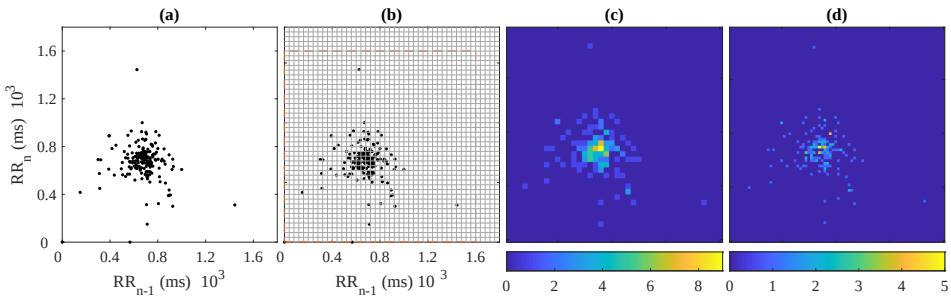

Figure 2.1: Creation of Poincaré Images from Poincaré Plots.(a) Poincaré Plot obtained from a signal segment of 120 s. (b) Grid indicating the bin discretization of the Poincaré Plot. The discontinuos line accounts for the common limits applied for all Poincaré Images. (c) Poincaré Image created using a bin size of [40x40] ms. (d) Poincaré Image of bin size [20x20].

lays in the interval $[k\Delta x, (k+1)\Delta x)$ and, jointly, RR_{n+1} lays in $[j\Delta y, (j+1)\Delta y)$. The values of $P[j,k]$ are computed in a range of interest and for given values of Δx and Δy. It's worth noting that $P[j,k]$ depends on three parameters: $N, \Delta x, \Delta y$. The Poincaré Image differs from the Poincaré plot because it is discretized in bins and each bin has an associated amplitude. A comparison of the two, for the case of RR segment in NSR, is shown in figure 2.1.

2.2.2 Rhythm classification by Poincaré Image

The basic idea behind the proposed classification strategy is that ECG segments with similar rhythm do share similar Poincaré Images. Mutual Information (MI) metric can be used to measure the similarity between two Poincaré Images: it is a commonly adopted metric to measure similarity between medical images (both 2D and 3D) widely used to evaluate image alignment in image registration problems (Viola et al. [98] and Maes et al. [99]). It measures the amount of information that one variable contains about another one. As a similarity measure it has a number of advantages, as it assumes no prior functional relationship between images. Rather, it assumes a statistical relationship that can be captured by analyzing the images' joint entropy. MI can be expressed through diverse mathematical forms such as Shannon entropy (SE), joint entropy or the Kullback–Leibler distance. Each definition contains the same information and can be rewritten into the others. Based on SE, the MI, $I(A,B)$, between two images A and B, can be expressed as

$$I(A,B) = H(B) - H(B/A) \tag{2.3}$$

where $H(B)$ corresponds to the SE of image B based on the probability distribution of its pixels' values. $H(B/A)$ instead represents the conditional entropy computed on the conditional probabilities $p(b/a)$ of pixels from image A and B [100]. As there does not exist a defined upper bound for MI, different normalization strategies have been suggested [101, 102]. Normalized MI (NMI) as an analogue to covariance and related with the geometrical mean as in [102] is expressed as

$$NMI(A,B) = \frac{I(A,B)}{\sqrt{H(A)H(B)}} \tag{2.4}$$

A different approach for measuring dependencies between Poincaré images is the 2D correlation analysis. While MI and NMI evaluate similarities between two images based on their statistical distributions, the 2D correlation coefficient quantifies the energy metric among them. It is computed as

$$Corr(A,B) = \frac{\sum_y \sum_x \left(A_{xy} - \overline{A}\right)\left(B_{xy} - \overline{B}\right)}{\sqrt{\sum_y \sum_x \left(A_{xy} - \overline{A}\right)^2 \sum_y \sum_x \left(B_{xy} - \overline{B}\right)^2}} \tag{2.5}$$

where x and y are the number of pixels per row and column and $\overline{A}$ and $\overline{B}$ correspond to the mean values of image A and B respectively.

Let's now suppose we know the *characteristic* Poincaré Image for a given cardiac rhythm of interest, i.e. a Poincaré Image which represents the typical RR pattern for that cardiac rhythm. We will call it *Poincaré Atlas* (the way these Poincaré Atlas are built will be discussed in the next section). Let's A_r be the Poincaré Atlas for rhythm r, P the Poincaré Image to be classified, and let's suppose we are interested in recognizing R different rhythms (let'say, $r = 1$: NSR, $r = 2$: AF, $r = 3$:AT, ... $r = R$: AB). Let's now consider a new RR segment, whose rhythm has to be identified. A strategy could be to compare its Poincaré Image with all the available Atlases and select the one maximizing the NMI and 2D Correlation coefficient, namely

$$\hat{r}_{NMI} = \underset{0 \leq r \leq R}{\arg\max}\, NMI(A_r, P) \tag{2.6}$$

$$\hat{r}_{Corr} = \underset{0 \leq r \leq R}{\arg\max}\, Corr(A_r, P) \tag{2.7}$$

where $\hat{r}$ is now the estimated cardiac rhythm. Note that Equations (2.6, 2.7) do not require the setting of any threshold to identify a rhythm. Both parameters could be combined as

$$\hat{r} = \underset{0 \leq r \leq R}{\arg\max} \left(\frac{NMI(A_r, P)}{\hat{r}_{NMI}} \cdot \frac{Corr(A_r, P)}{\hat{r}_{Corr}} \right) \tag{2.8}$$

The procedure is described in the following figure

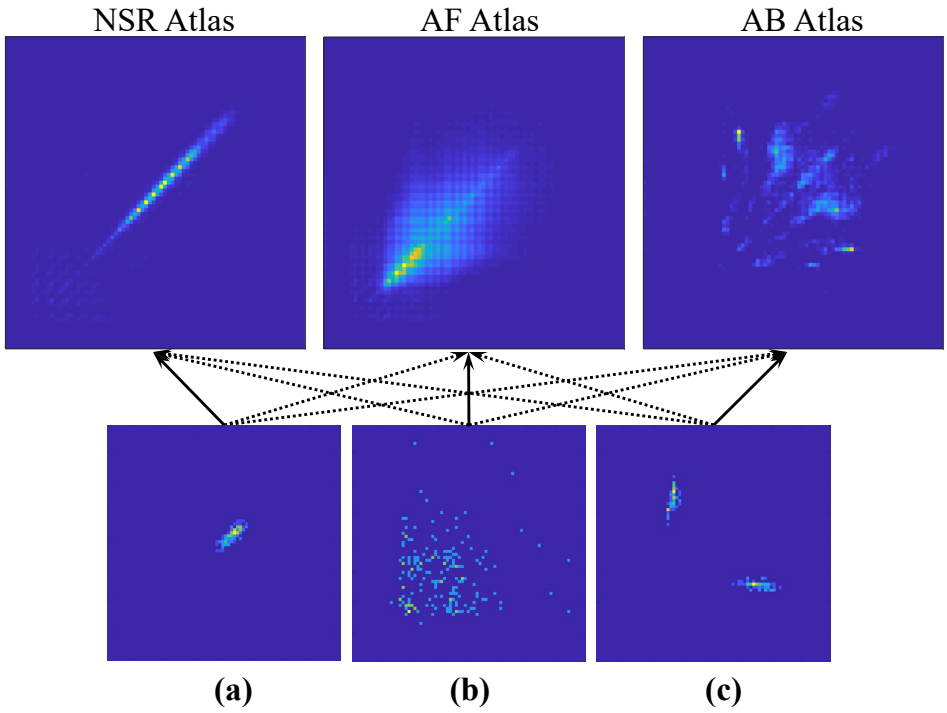

Figure 2.2: Examples of Poincaré Atlases (first row) and Images (second row) of RR configuration for a time window of $120s$ and bin size $20ms$ and rhythms: (a) NSR, (b) AF, (c) AB. Arrows represent the comparison process of NMI and 2D correlation. Solid arrow lines represent the Atlas attaining the maximum value as expressed in 2.8 for (a),(b) and (c), respectively.

In order to apply Equation (2.8) the set of Atlases had to be computed. Atlases were derived as the average among all the Poincaré Image of the same cardiac rhythm. That is

$$A_r = \frac{1}{M_r} \sum_{m=1}^{M_r} \frac{P_m^{(r)}}{N_m^{(r)}} \tag{2.9}$$

where $P_m^{(r)}$ is the Poincaré Image of the m^{th} epoch of rhythm r within the available dataset, $N_m^{(r)}$ the total number of RR points in the Poincaré Image and and M_r is the total number of epochs of rhythm r. Prior to the computation of NMI and $Corr$, images were expressed as probability maps as in Equation (2.9). In addition, in order to discretize values, Poincaré Images and Atlases were scaled to Uint16 before Equation (2.6).

2.2.3 ECG analyzed Datasets and Processing

The MIT-BIH arrhythmia database (MITDB), the AFDB, and the long-term atrial fibrillation database (LTAFDB) were used in this study. The MITDB contains 48 ambulatory half-hour ECG recordings. The AFDB contains 26, 10-hour ambulatory recordings and the LTAFDB 84 recordings of 24–25 hours each [15]. All databases contain beat annotations. From the MITDB and LTAFDB segments labelled as AF, NSR and AB were used. From the AFDB only AF was included. Other rhythms were not considered because either too short episodes or very few episodes were present.

RR sequences were divided into segments of equal length (in s) with 50% overlap. Four window lengths were considered: 120 s, 60 s, 30 s and 20 s. For each segment, three types of Poincaré Images were computed: 1) The RR Poincaré Image, obtained by Equation (2.1), 2) the δRR Poincaré Image, computed by replacing the series of RR intervals in Equation (2.1) with the series of first difference of the RRs, namely $\delta RR_n = RR_{n+1} - RR_n$, and, finally 3) the $RR\delta RR$ Poincaré Images, built by the juxtaposition (side–by–side) of the first two images. All Images were built using the same bin–size on both axes ($\Delta x = \Delta y$). Furthermore, four values of bin–size were tested: 5 ms, 10 ms, 20 ms and 40 ms. In each Image, x– and y-axes ranges were set from 0 ms to 1600 ms for RR Image and from -800 ms to 800 ms for the δRR ones. Images' dimensions (in pixels) varied according to the selected bin–size resulting in a [$320x320$], [$160x160$], [$80x80$] and [$40x40$] matrix for

$\Delta x = 5$ ms, $\Delta x = 10$ ms, $\Delta x = 20$ ms and $\Delta x = 40$ ms, respectively. The combination of the different bin-sizes and window lengths resulted in 16 different configurations that were used to built RR, δRR and $RR\delta RR$ Images. Each Image was labelled as a determined rhythm (NSR, AF, etc...) according to the dominant cardiac rhythm of the relative RR segment (i.e the dominant rhythm was the rhythm that covered more than the 50% of the signal segment used to build the Poincaré Image). From all the images generated, only those labelled as NSR, AF or AB were considered.

2.2.4 ECG performance evaluation

The available Poincaré Image dataset was divided in two subsets by patient–wise division, i.e. each subset contained images of different patients: the training–set (containing 80% of the available patients) and the independent test–set (containing the remaining 20% ones). The training–set Images were used to generate the Atlases for each rhythm by averaging (Equation (2.9)) all the available images belonging to the same cardiac rhythm. Then, for each image of the test–set, the NMI and 2D Correlation coefficient between the image and the Atlases were computed and the rhythm identified according to Equation (2.8). A ten fold cross–validation was performed to avoid bias related to a single split of the dataset. Given the low and unequal distribution of AB images, AB patients were divided as homogeneously as possible in the ten—fold cross–validation sets. At each K–fold, 8 subsets were used for train and 2 for test. Results are expressed as weighted mean and standard deviation with respect to the number of classified segments at each iteration.

2.2.5 Study on the AFDB dataset

With the aim of comparing with other algorithms proposed in the literature. The same technique was applied using only the PhysioNet's open-access MIT-BIH AFDB [15]. The AFDB is formed by 25 long-term ECG recordings of 10 hours duration sampled at 250 samples per second with 12-bit resolution over a range of ± 10 millivolts. Each signal contains rhythm annotations for NSR, AF and AFL. In this substudy AFL segments were also targeted in order to include all signal segments and make the study comparable.

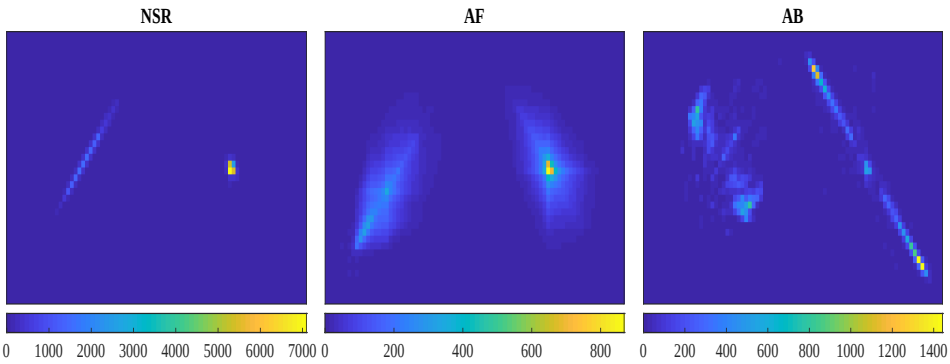

Figure 2.3: One of the 10 different sets of Poincaré Atlas for $RR\delta RR$ images computed during the study representing NSR, AF and AB. Values are probability distributions discretized as uint16.

2.2.6 PPG data

PPG signals included in the study were simulated according to [85]. The model uses RR interval series, obtained from annotated ECG recordings, as input for generating a PPG signal. Each PPG pulse is modelled as a linear combination of two Gaussian functions and two log-normal waveform. The log-normal function defined as

$$\varphi_1(t; m, \sigma_1) = \begin{cases} \frac{1}{t\sqrt{2\pi\sigma_1^2}} e^{-\frac{(ln(t/m))^2}{2\sigma_1^2}}, & t > 0, \\ 0, & t \leq 0 \end{cases} \tag{2.10}$$

and the Gaussian function defined as

$$\varphi_i(t; \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{t^2}{2\sigma_i^2}}, \qquad i = 2, 3, \tag{2.11}$$

The three functions are linearly combined in eq. 2.12 in a time-shifted, time-scaled and weighted version to model a PPG pulse.

$$\varphi_i(t, \theta = w_1\varphi_1(t - \tau_1; m, \sigma_1) + \sum_{i=2}^{3} w_i\varphi_i(t - \tau_i; \sigma_i) + a, \tag{2.12}$$

Parameters are estimated through nonlinear least squares fitting. Three different template pulses were used as according to Dawber et al. [103]. For

more details, the reader is referred to the original publication of Sološenko et al. [85].

The model also accounts for the presence of premature beats by introducing amplitude and time scale factors which modify pulse width and amplitude, thus making it possible to simulate ectopic beats and certain rhythms such as bigeminy, known to cause false alarms in RR interval-based AF detection. 200 PPG signals (100 in AF and 100 in NSR) corresponding to 300 RR intervals were generated. Systolic and diastolic peak positions were identified as in [104]. The inter-systolic series was computed (PPi) and further analyzed.

2.2.7 PPG image generation

In ECG signals, the Poincaré plot is the graphical representation of one RR interval, RR_i, versus the previous one, RR_{i-1}. A different version of the Poincaré plot that follows the same criteria, uses the δRR instead, defined as Eq. 2.13

$$\delta RR_i = RR_{i-1} - RR_i \tag{2.13}$$

Even though Poincaré Plots are typically used for RR data, the same concept can be applied to the PP intervals found in PPG signals. For each PPG signal a sliding window of 20, 30 and 60 s without overlap was used to generate the δPP Poincaré plots from the previously computed PPi. Poincaré Images were computed from the Poincaré plots as previously described in [105]. Each Poincaré plot was discretized into its pixelated version as a 2D histogram. The bin size used to generate the pixels was 20 ms. The counting of the number of $(\delta PPi_n, \delta PPi_{n-1})$ points of the Poincaré plot present inside the limits of a determined bin, accounted for the resultant value of each bin. A common range of (-800, 800) ms was established for all Poincaré plots prior to conversion to Poincaré Images. The dimensions of all the resultant Images were [80x80] pixels. Values in individual Images were then converted into probability maps and transformed into the uint16 scale in order to be discretized.

2.3 Results

2.3.1 ECG Poincaré Images

The total amount of images generated for the three rhythms is displayed in Table 2.1. The last column of the table indicates the total number of different patients from which the images of each rhythm came from.

Table 2.1: Total number of images and patients per rhythm.

Rhythm	*Time Window (s)*				*Patients*
	120	*60*	*30*	*20*	
NSR	52,000	103,800	207,100	309,800	113
AF	41,670	83,490	167,100	250,800	114
AB	211	530	1,303	2,251	30

2.3.2 Image choice

To choose the best image type (RR, δRR and $RR\delta RR$) accuracy, sensitivity and specificity classification values were compared for different window lengths and bin–sizes. As an example the performance for AF classification is shown in Table 2.2. For AF classification, the usage of RR images led to higher specificity but poorer sensitivity values than the usage of δRR images. The combination of both in $RR\delta RR$ led to a more balanced relationship between sensitivity and specificity maintaining similar accuracy values than δRR. Therefore, the following results are shown for $RR\delta RR$ only.

2.3.3 Poincaré Image classification metric

Accuracy, sensitivity and specificity values of the $RR\delta RR$ images classification after the 10 fold cross-validation are gathered in Figures 2.4, 2.5 and 2.6. In each figure, each row of three graphs represents a rhythm's classification results using different time windows and bin–sizes. Each line in a graph represents the results obtained using the same time window with varying bin–sizes along the x-axis.

Table 2.2: Results for all types of AF images, time windows and bin size 40 ms.

	TW (s)	*RR*	*δRR*	*RRδRR*
Accuracy (%)	*120*	82.32±5.27	89.26±3.83	89.96±3.52
	60	83.15±4.79	89.09±3.23	89.78±3.24
	30	82.39±4.20	88.18±2.79	89.12±3.19
	20	82.52±3.21	87.18±3.24	88.60±3.05
Sensitivity (%)	*120*	82.75±8.98	95.61±4.35	94.87±3.71
	60	83.52±9.40	95.15±3.84	94.35±3.68
	30	81.77±6.35	94.43±1.91	93.40±2.95
	20	80.70±6.29	92.20±1.93	91.76±3.03
Specificity (%)	*120*	81.93±4.29	83.39±4.50	85.43±4.06
	60	82.78±7.89	83.45±7.65	85.52±7.46
	30	82.96±7.71	82.36±6.25	85.13±7.46
	20	84.20±6.98	82.49±7.14	85.63±6.83

Values are presented as weighted mean ± standard deviation.

Figures 2.4, 2.5 and 2.6 gather the classification results obtained using NMI, 2D correlation and the combination of both, respectively, distance metrics between Poincaré Images and the rhythm–specific Atlases. The combination of NMI and 2D correlation as represented in Figure 2.6 led to a better performance than both measures independently. Therefore, further results are presented and discussed just using this combined metric as described in Equation 2.8.

A computational cost analysis was performed for each of the metrics of evaluation. For 2D correlation alone the number of computations between a single Poincaré Image of bin size 40x40 ms and a Poincaré Altas was of 2.51 10^4 flops, for NMI was of 4.46 10^7 flops and for both combined 4.46 10^7 flops.

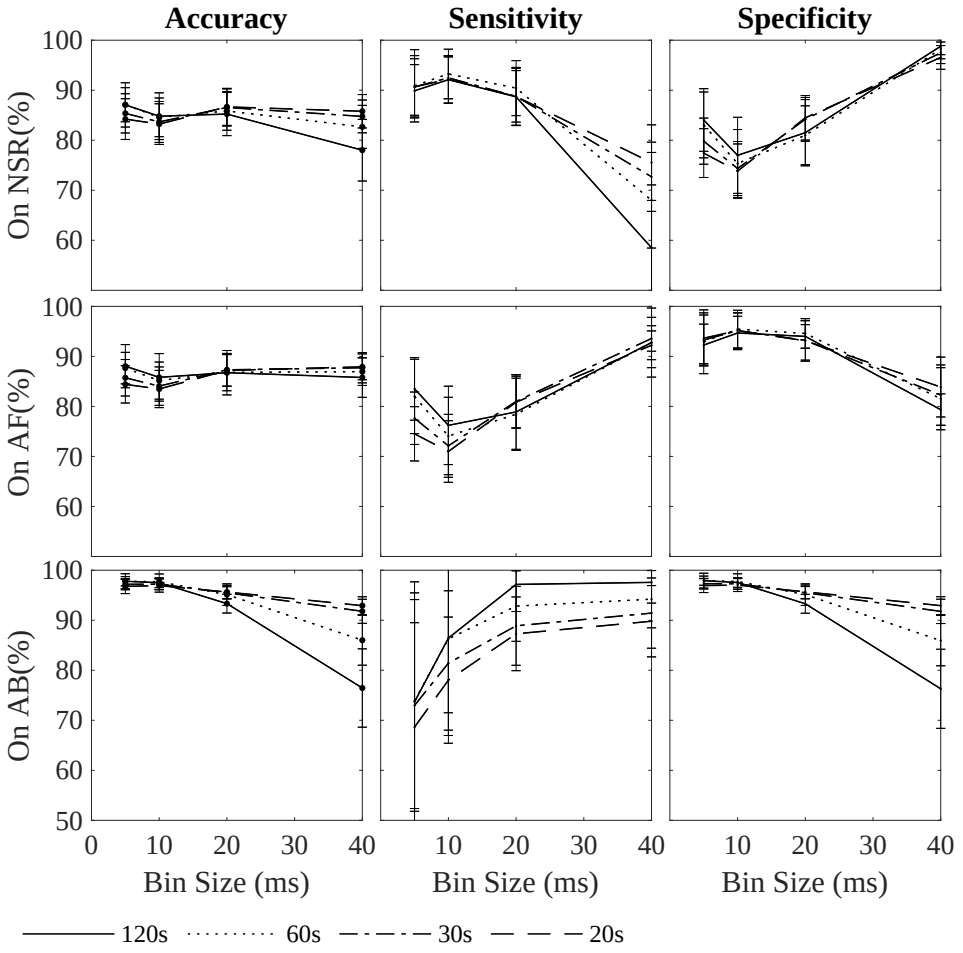

Figure 2.4: Results for NMI as distance metric. Accuracy, sensitivity and specificity values for the classification of NSR, AF and AB as a function of time windows and bin sizes. NSR, AF and AB are shown from the top to the bottom.

2.3.4 Image parameter optimization

Independently of the bin size, the larger the time window, the better the classification performance of AF and NSR. As the time window was reduced, so was the standard deviation for all rhythms.

The underrepresentation of AB, led to very high accuracy values and higher standard deviation with respect to AF and NSR. Sensitivity values for a time window of 120 s showed a poorer performance in comparison to the other time windows used due to the lower amount of AB images. Bin size did not influence substantially results in term of accuracy values for

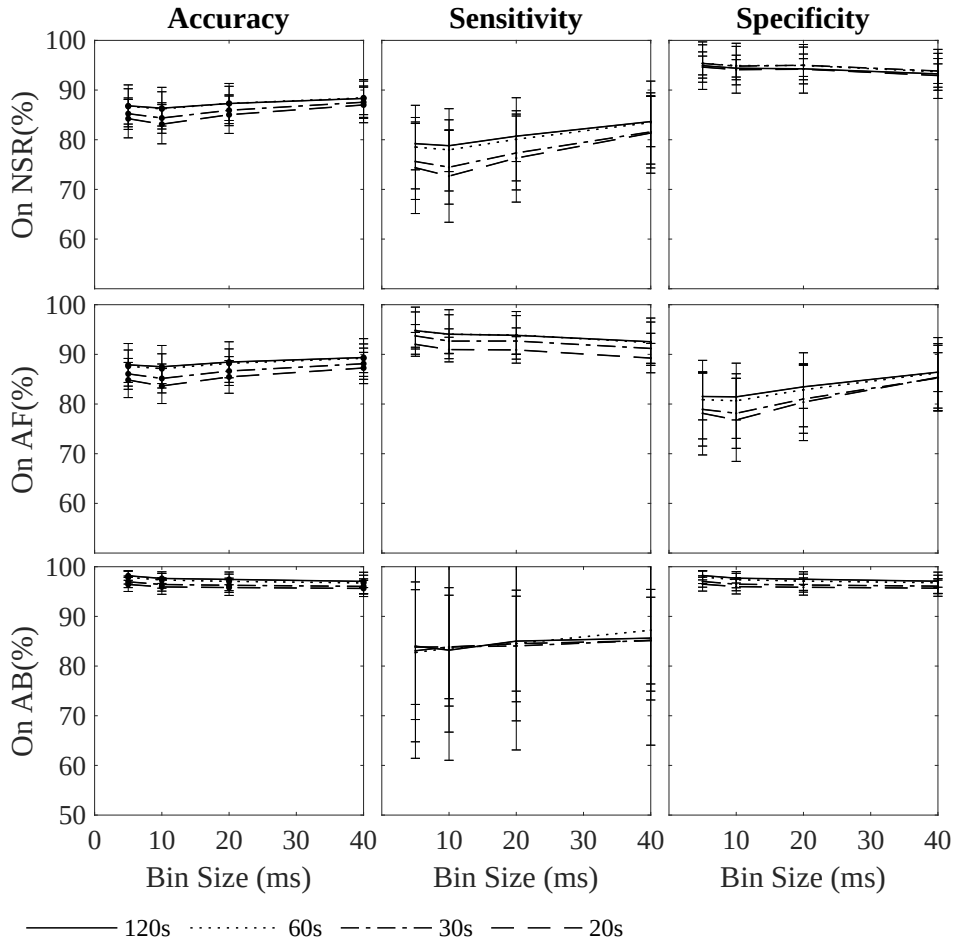

Figure 2.5: Results for 2D correlation as distance metric. Accuracy, sensitivity and specificity values for the classification of NSR, AF and AB as a function of time windows and bin sizes. NSR, AF and AB are shown from the top to the bottom.

AF and NSR. However, it did affect the balance between sensitivity and specificity values of AF and NSR.

2.3.5 Statistical analysis

A two-way ANOVA test was performed to study time window and bin size influence in the classification of NSR, AF and AB. Time window was significant for AF and AB sensitivity and specificity values respectively ($p < 0.05$), 120 s and 60 s the best options for both cases. Bin size was only significant for NSR specificity, being 40 ms the best choice. Based in this

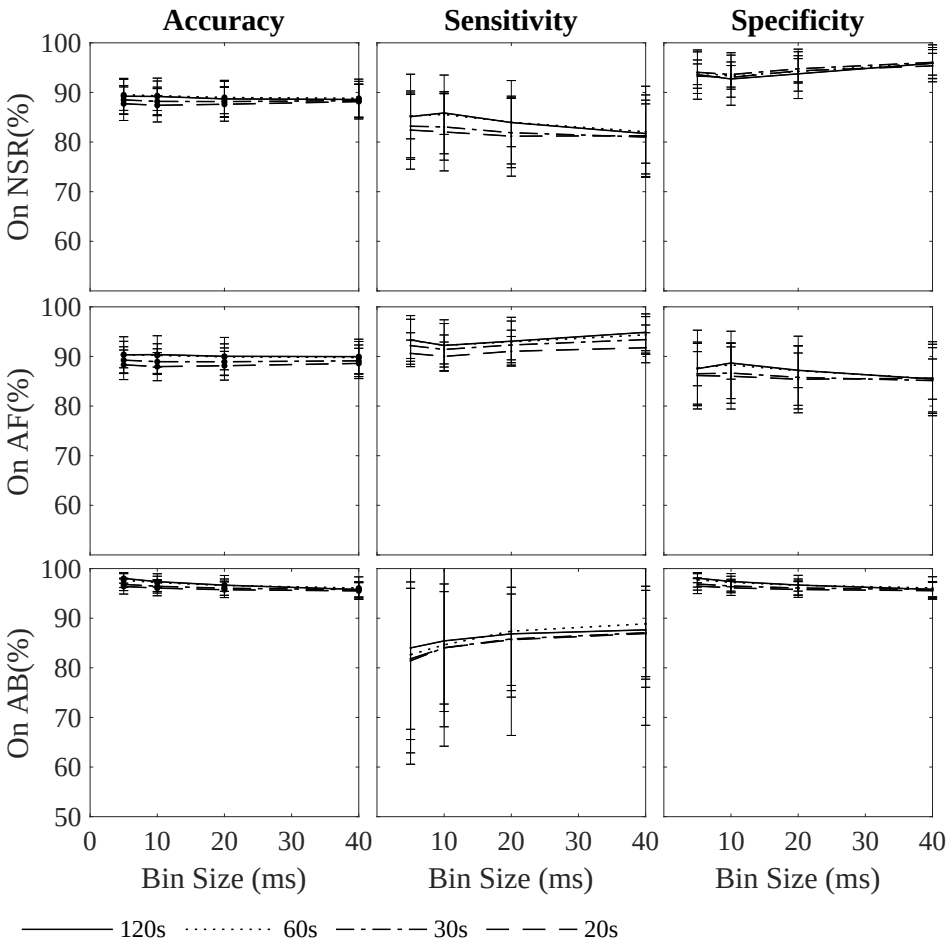

Figure 2.6: Results for NMI and 2D correlation as distance metric. Accuracy, sensitivity and specificity values for the classification of NSR, AF and AB as a function of time windows and bin sizes. NSR, AF and AB are shown from the top to the bottom.

analysis, the best combination of parameters would be a 120 or 60 s time window with 40 ms bin size. Results for this combination are gathered in Table 2.6.

2.3.6 Optimized ECG classification

In Figure 2.7 the classification distribution of the Poincaré Images for the four different time windows and optimal bin–size is shown. Images were labelled as "not classified" when the score resulting from the computation

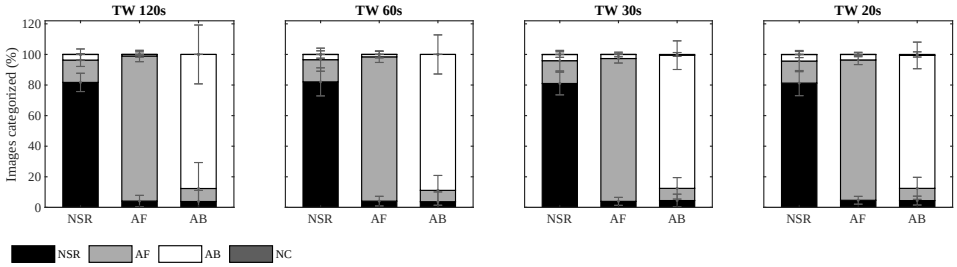

Figure 2.7: Classification results after ten fold cross validation of the three time windows and bin–size [40x40] ms. Each bar represents in percentage, the distribution of the classification of all images with a determined rhythm; NSR, AF or AB, into one of the three rhythms or non classified (NC).

Table 2.3: Classification results with time window 60 s and bin size 40 ms.

Rhythm	*Accuracy (%)*	*Sensitivity (%)*	*Specificity (%)*
NSR	88.81±3.87	82.07±9.18	95.91±3.14
AF	89.78±3.24	94.35±3.68	85.51±7.46
AB	96.08±2.25	88.86±12.78	96.10±2.25

Values are presented as weighted mean ± standard deviation.

of NMI and 2D correlation was the same for more than one Poincaré Atlas. Each bar of the graphs represents the 100% of images of a determined rhythm and the percentage that has been classified into each of the four different categories; NSR, AF, AB or not classified (NC). Regardless of the time window and bin–size used, the classification distribution of each different set of images belonging to the same rhythm followed similar patterns. The misclassified NSR images were always categorized as false AF episodes in a higher percentage than AB episodes. AF false negatives were although classified more often as NSR than as AB. AB classification after ten fold cross-validation attained higher standard deviation values with respect to the other rhythms of study. However, as the time window was shortened and the amount of AB images increased, the standard deviation was reduced.

Table 2.4 gathers the number of images generated for each of the rhythms of study using the three time windows as well as the number of signals used to compute them. An example of one NSR and AF Poincaré Images computed for a time window of 60 s and a bin size of 20 ms is displayed in

Figure 2.8.

Table 2.4: Images generated for AF and NSR for each of the time windows.

Time Window	*NSR*	*AF*	*Signals*
20 s	1137	1032	100
30 s	738	669	100
60 s	347	311	100

2.3.7 Study on AFDB subset

Sensitivity and specificity AF results of the study on the AFDB subset are gathered in Table 2.5. Table 2.5 also contains the best sensitivity and specificity values reported in literature in the past years based only on RRi and applied to the AFDB. In addition, it indicates the number of RRi resolution and the Rhythms targeted. Lake et al. [27] included also AFL in the study but it did not differentiate between AFL and AF.

2.3.8 PPG Poincaré Images

Figure 2.9 represent one of the ten sets of Poincaré Atlases for NSR and AF generated after the ten fold cross-validation for a time window of 60 s and a bin size of 20 ms. Instead, in Figure 2.10 a 3D version of the Poincaré Atlases can be appreciated.

Accuracy, sensitivity and specificity values and standard deviations after the ten fold cross-validation are shown in Table 2.6. The values presented correspond to the classification of the 20% of AF Poincaré Images of each time window. As the time window was lengthened the classification performance increased, yielding a 93.89% of accuracy for 60 s. Specificity values was always higher than sensitivity ones.

2.4 Discussion

Poincaré plots have been used as inspiration for the generation of parameters to describe non–linear RRi patterns, typically focusing on the detection of

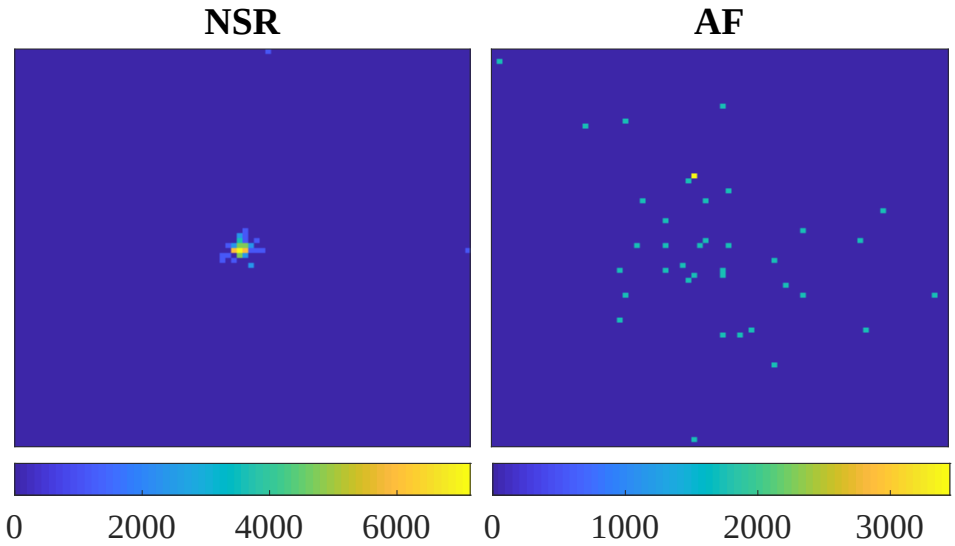

Figure 2.8: Example of one of one NSR and AF Poincaré Image for a time window of 60 s and a bin size of 20ms.

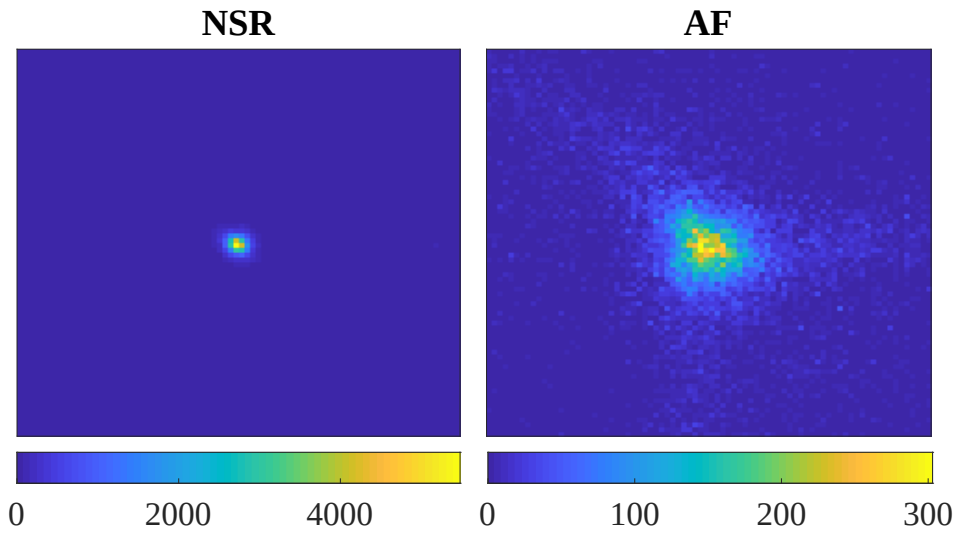

Figure 2.9: Example of one of the ten pairs of Poincaré Atlases generated visualized as a Poincaré Image for NSR and AF. Values are probability maps transformed into Uint16 scale.

Table 2.5: Total number of images used per rhythm and number of patients data were extracted from.

Study	Year	Rhythm	RRI	Se (%)	Sp (%)
Lee et al. [28]	2013	AF	128	98.2	97.7
Zhou et al. [57]	2015	AF	-	97.4	98.4
Petrenas et al. [30]	2015	AF	8	97.1	98.3
Lake et al. [27]	2011	AF,AFL	12	91	94
García-Isla et al.	2019	AF,AFL	(120s)	94.3	97.2

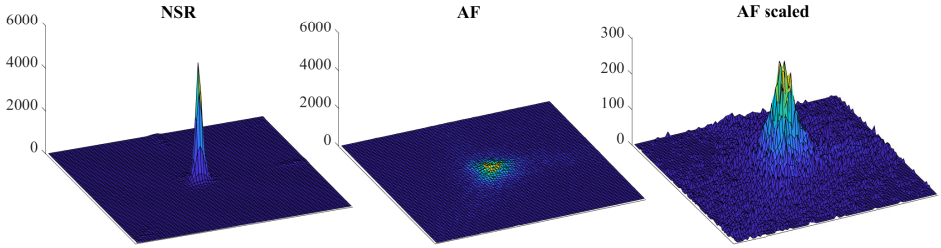

Figure 2.10: Example of one of the ten pairs of Poincaré Atlases generated for NSR and AF displayed in 3D dimensions. Values are probability maps values transformed into Uint16 scale. The scale of the AF Poincaré Atlas has been scaled for proper visualization

Table 2.6: Classification results with time window 60 s and bin size 40 ms.

Time Window	*Accuracy (%)*	*Sensitivity (%)*	*Specificity (%)*
20 s	89.47±3.10	93.57±1.07	85.62±6.24
30 s	93.10±1.47	97.70±0.96	88.92±3.11
60 s	93.89±2.16	100±0.00	88.41±4.01

Values are presented as weighted mean ± standard deviation.

AF [27, 57, 58, 60]. However, to the best of our knowledge, this is the first study that considered the design represented in the plot as a whole. The creation of a reference image specific for each rhythm aimed to capture the general design created by that rhythm as described by [92] even if influenced by the presence of different noise levels, ectopic beats, heart rates or inter–patient variability.

AF, NSR and AB rhythms were used to test the presented methodology. As AB episodes are shorter and more infrequent in the databases than AF and NSR, the number of images and patients was inferior to those of the other two rhythms of study. Different databases were used to maximize variability between the signals employed. Noisy signal segments were not excluded from this study. From the three types of Poincaré Images studied, the combination of RR and δRR in $RR\delta RR$ led to a balanced relationship between AF sensitivity and specificity. Even though both RR and δRR have been used in several past studies [58–60], $RR\delta RR$ has never been reported before. The comparison among the three metrics for Poincaré Images classification resulted in a higher performance of the combination of NMI and 2D correlation. However, the computational cost analysis showed an increment of the number of operations produced by NMI calculation. Using one of the three metrics would be a matter of trade-off between the computational cost and the classifier's accuracy as it can be seen in Figures 2.5, 2.4 and 2.6. Probably in an online monitoring scenario, 2D correlation would be preferred given the high computational burden of NMI.

Time windows of 120 s and 60 s were the best options according to the statistical analysis. This is consistent with the European Society of Cardiology (ESC) AF guidelines which set 30 s as the minimum length for an AF episode to be considered clinically relevant [106]. In the presented methodology, a time window of 60 s would detect >30 s episodes, as only the 50% of arrhythmia prevalence on the time window's samples is needed for its detection. The condition of >30 s AF episodes is based on an agreed convenience due to the limitations of current monitoring techniques rather than on an electrophysiological basis. The significance of shorter episodes still remains unclear. Hence, their monitoring could be of clinical interest. Results obtained for 30 s and 20 s did not differ more than 3 percentage points from those of wider time windows as presented in Table 2.2 and Figure 2.6. Even if better performance was achieved for larger time windows, shorter ones as 20 s and 30 s could be of greater interest from a clinical perspective. The aim of presenting different time windows was not only to optimize this parameter, but also to study how the methodology behaves as it is shortened.

AF and NSR results suggest Poincaré Atlases and Images manage to represent the complex RRi dynamics characteristic of a certain rhythm. Standard deviation values instead, show the dependency of the methodology on the patients used both for classification and for computation of the

Atlases. This may indicate than either more than one Atlas may be needed to successfully capture all the different types of behaviour of AF and NRS or than more patients are required to fully represent the heterogeneity of the RR patterns characteristic of a certain rhythm. AF for instance, is known to present different degrees of organization [107] that may lead to different RRi behaviour. Furthermore, the high standard deviation obtained for AB classification clearly shows that the AB inter–patient variability was not enough to provide a complete representation of AB pattern for all cases. Lower sensitivity values obtained in some K–fold iterations as 56.8%, 65.4% and 70.0% with corresponding specificity 98.3%, 99.4% and 99.0% suggest that the atlases created did not succeed in representing the AB pattern of the test set population. However, high sensitivity values as 97.7%, 96.7% or 94.6% with specificity 98.7%, 96.1% and 96.7% demonstrate that the Poincaré Atlas approach may successfully discriminate among AF, NSR and AB. These results point out also at the necessity of the proposed methodology of a higher number of patients and episodes to build a rhythm–specific Atlas.

Nevertheless, in contrast with current algorithms that are rhythm–specific designed, with the sufficient amount of patients and episodes RRi analysis using Poincaré Images could permit the generalization of the method to different cardiac rhythms.

2.4.1 Related work

Other studies aiming to classify cardiac tachyarrhythmia only through RRi analysis attempted to characterize a particular rhythm behaviour with one or more parameters and set a threshold indicating the presence of such rhythm [24, 27, 28, 30, 31, 57–60, 91]. Those studies targeted a single tachyarrhythmia and proved to be very effective at detecting AF. However, they reported poor results when trying to detect other rhythms as [27, 60].

A comparison of the former approach with the published algorithms is rather challenging as current algorithms have been optimized for the detection of AF while the one presented in this work aims to target other rhythms simultaneously. Contrast among them has been typically done by studying AF detection on the AFDB. Results presented in Table 2.5, show a performance for AF detection lower than some of the state-of-the-art classifiers based exclusively on RRi for AF detection. Only Lake et al. [27] attempted

to detect (but not differentiate with AF) also AFL. Despite obtaining an AF sensitivity lower than other models, still sensitivity and specificity values were kept above 94%. However, the AFDB does not contain annotations for rhythms other than AF, AV junctional rhythm and AFL, being these last two in a very low proportion. Therefore, it is difficult to evaluate how other detector's performance is influenced by the presence of other rhythms and how our results could improve by targeting them.

Lee et al. [28] published a methodology for AF detection based on time varying coherence functions (TVCF) and sensitivity. They reported a sensitivity and specificity of 98.2% and 97.7% respectively on the AFDB. However, when their method was tested on the MIT-BIH arrhythmia database, that contains other arrhythmia, specificity lowered from 97.7% to 81.2%. The thresholds computed for the detection of AF were calculated using the same AFDB database, so performance results using this database could be biased. Zhoe et al [57] used symbolic dynamics and SE for AF detection, obtaining a sensitivity and specificity of 97.4% and 98.4% on the AFDB. Similarly to [28], their specificity dropped from 98.28% in the AFDB to 87.41% in the MIT-BIH arrhythmia database. Petrenas et al. [30] AF detector for an 8-beat window achieved a sensitivity and specificity of 97.1% and 98.3% respectively on the AFDB. One of the key points for the achievement of such performance was the implementation of a bigeminy and ectopic beat suppression algorithm. They also reported a decrease in specificity when applied to the MIT-BIH database from 98.3% to 86.4%.

These differences in performance show how explicitly tackling other rhythms could improve AF monitoring. In addition, they demonstrate that the high specificity rates reported on AFDB might not be representative of AF detection in presence of other arrhythmia. The former study does not intent to present an specific classifier for AF and AB, but to introduce a new type of approach towards the detection of atrial tachyarrhythmia. It aims to put forward a method that could be extended to different types of RRi patterns rather than a model specifically designed to detect a single arrhythmia.

2.4.2 PPG

The main finding of this work is the capability of detecting AF segments on PPG signals using δPPi Poincaré Images. The usage of Poincaré Atlases

applied directly on PPG Poincaré Images managed to translate the same methodology applied previously in the ECG to pulsatile signals. The best classification results in terms of accuracy, sensitivity and specificity were obtained for a time window of 60 s. Performance worsened as the time window was reduced. Nevertheless, high results over 93% accuracy were yet obtained for a time window of 30 s. This work represents one of the few that has attempted to apply the same methodology used for AF detection using RRi to detect AF with PPi. Even thought RRi and PPi share many characteristics, they also entail relevant differences both in the intervals themselves and in the amount of noise present in signals of origin. Results were compatible with those obtained for RRi with a superior percentage in sensitivity. Given that only AF was targeted in the PPi analysis, the increase in performance of the algorithm is coherent.

The proposed classification methodology directly integrates noise into the definition of a rhythm's Poincaré Image pattern by its inclusion in the Poincaré Atlas, relative to each rhythm. By tuning the parameters used to build the Poincaré Images the compromise between sensitivity and specificity could be altered to prevent the hindering effect of noise. Furthermore, arrhythmia detection by Poincaré Image and Atlas analysis could enable its generalization into more rhythms apart from AF and NSR as it was shown on [105, 108]. A wider annotated dataset with other rhythms would be necessary in order to prove this on PPG signals. Being able to capture different rhythms using these signals could provide a cheap and accessible way of cardiac constant monitoring on a broad population and consequently, the premature detection of possible electrophysiological disorders.

Nevertheless, the vulnerability to noise and artifacts of pulsatile morphology compared to electrocardiographic lead to bad–quality signals, which trigger many false positives [83]. Many algorithms obtaining high performance on AF detection were evaluated on controlled datasets with a rather low amount of noise, not applicable to the real-life ambulatory signals [104, 109]. It is thus, important to be able to quantify and control the amount of noise contained in the signals used for the development and testing of the classification algorithm in order to make it as reliable as possible.

2.4.3 Limitations and future work

The major limitation of this study is the need of a considerable amount of tachyarrhythmia episodes and patients to create the Poincaré Atlases. Some data augmentation techniques could help overcome this issue. However, the resultant images should still be physiologically feasible as otherwise the effect could be counterproductive. Geometrical modification techniques implying image transformations as rotation and flipping would not fulfil this condition unless values representing RR intervals are kept withing the physiological range. Although we theorise that the proposed model could be applied to describe other rhythms, the available data limited the study to AF, NSR and AB. Furthermore, the amount of AB data was very low in comparison with AF and NSR, which led to a very high standard deviation values. Ideally, a set of Poincaré Atlases should be built representing all different tachyarrhythmia and made open-source to be used in other studies. In addition, even if the study on PPG signals pointed at the possibility of extending the methodology to this technology, still real ambulatory signals as well as other tachyarrhythmia episodes should be used for its validation.

2.5 Conclusion

This work presents a new approach for cardiac tachyarrhythmia classification introducing the concepts of Poincaré Images and Atlases for the creation of different rhythm reference models. In addition, a combined display of the RR and δRR Poincaré Plot, $RR\delta RR$, was introduced for the joint description of the two-beat and three-beat variability. The proposed method was evaluated for the detection of AF and NSR in PPG signals and in AF, AB and NSR segments in ECG signals. Promising results were obtained on ECG signals showing the capacity of extending it to different rhythms.

Chapter 3

Beat Classification for PAC Detection in Ambulatory 2-lead ECG Holter Recordings

3.1 Motivation

PACs and PVCs account for beats originating in a site different than the sinoatrial node; the atrium and the ventricles, respectively. PACs have

always been considered benign. However, several studies link them to high risk of developing AF and stroke [5, 7, 10].

About 25% to 30% of ischemic strokes remain unexplained (cryptogenic) [5]. One of the possible causes is that the thromboembolic events are caused by occult or silent AF. Prediction of the appearance of these AF episodes could reduce the incidence rate of cryptogenic strokes. Several studies link frequent PACs to first time appearance of AF [6–9]. Others have studied PACs as the possible direct reason for stroke [10]. Furthermore, frequent PACs have been studied as a measure of cardiac tissue deterioration [7, 10–12] and as a possible cause for left ventricular remodelling [13]. Despite the evidence of the possible implications of PACs, no clear relationship between PACs and the enumerated events has been described. Even if frequent PACs have been related to PAF [11], no proper definition of "frequent" PAC has been formulated. Moreover, PAC relationship with stroke is still controversial [10, 12]. The further elucidation of the direct and indirect implications of PACs would require the analysis of long-term ECG recordings and automatic PAC detectors could ease the study of this phenomenon.

All these studies point at the important and undervalued impact that PACs may have on the cardiac electrophysiogy. However, manual beat annotation of long-term ECG recordings is extremely time consuming and requires of specialized professionals. A PAC detector with high sensitivity able to assume this task is still missing. Such a detector would enable to study PAC implications in AF onset and cardiac tissue remodelling. It could be used to monitor patients for the occurrence of frequent PACs and determine stroke risk or possible silent AF or short PAF episodes. In addition, it could enhance the performance of arrhythmia detectors as PAC beats tend to increase AF false positives [83, 91].

Just a few detectors are present in literature explicitly designed for PAC [110, 111], most supraventricular premature beats (SVPB) classifiers are actually beat classifiers [66, 112] that attain low PAC detection sensitivity. Both ventricular (including PVCs) and atrial (including PACs) alter RR intervals (the distance between subsequent R peaks) similarly. However, while ventricular beats induce large QRS complex amplitude distortions, atrial beats cause low amplitude P-wave alterations. The presence of noise, ECG morphological inter-patient variability and P-wave parametrization complexity makes PAC detection a challenging matter. While R peak

detection and RR intervals are robust against noise and easily quantifiable, P-wave morphology study entails a higher complexity. A detector based on RR intervals to detect ECG segments containing any type of atrial or ventricular beat (from now on referred as A/V) could reduce the ECG segments to be analyzed. In addition, if coupled with an already existing automatic beat classifier, it would reduce false positives and ease the identification of false negatives.

In this work we present a beat classifier with emphasis on PAC detection and not requiring of any ECG delineation to extract morphological information. In addition, we present an automatic classifier capable of detecting ECG segments containing A/V beats based on the Poincaré Image representation of RR intervals as described in [108] and CNN. Both methods complement each other. The presented segment classifier enables the detection of regions containing A/V beats, drastically reducing the workload for manual beat annotations. This could ease the study of the PAC implications, currently hampered by the unavailability of extensive annotated databases and the high number of hours required to perform such task by specialized clinicians. Furthermore, the segment classifier could be used as a complementary detector to the beat-to-beat classifier developed and those present in literature, what could lead to more reliant automatic annotations.

3.2 Materials and Methods

3.2.1 Data

Two PhysioNet public databases [15] were used in this study: the Long-term ST Database (LTSTDB) and supraventricular database (SVDB). Signals were 2-lead ECGs acquired at 250 and 128 Hz with a duration of 21-24 hours and 30 minutes for the LTSTDB and SVDB, respectively. These databases were selected because they are the ones containing a higher number of PACs and manual beat annotations. The LTSTDB was originally built so as to represent a wide variety of ST segments. The SVDB contains a high number of supraventricular events. While the LTSTDB contains PACs together with different ST-segment variations, the SVDB contains a high number and variety of different possible PAC occurrences: bigeminy, trigeminy and atrial runs. Both datasets were combined into a single dataset to use their complementary PAC representations for training and testing the model.

Table 3.1 gathers the number of beats per beat class in each database. In this study only 5 categories were originally considered as in [65]: Normal (N), Supraventricular (S), Ventricular (V), Junctional (J) and unclassifiable (Q) beats. These classes were built as follows:

1. We considered as PACs all S beats in which: atrial premature beats (A), aberrated atrial premature beats (a) and PACs (S) annotations were included.

2. V comprehended the categories: premature ventricular contractions (V), fusion of ventricular and normal beats (F) and ventricular escape beats (E).

3. N beats included normal beats N, bundle branch block beats (B) and atrial escape beats (e).

4. J and Q classes were excluded for the successive analysis as they were underrepresented in both databases.

Table 3.1: PhysioNet and simplified beat annotations per database.

Simp. Anns.	N			S			V			J		Q
Beat Anns.	B	N	e	A	a	S	V	F	E	J	j	Q
LTSTDB	88720	6727000	22	5482	29	30820	39840	476	71	1	6	2
SVDB	1	162100	0	0	1	12090	9930	23	0	9	0	80

B stands for bundle branche block beat, N for normal beat, A for atrial premature beat, a for aberrated atrial premature beat, e for atrial escape beat, Q for unclassifiable beat, S for supraventricular premature or ectopic beat (atrial or nodal), V for premature ventricular contraction, F fusion of ventricular and normal beat, E for ventricular escape beat, J for nodal (junctional) premature beat and j for nodal (junctional) escape beat.

3.2.2 Preprocessing

The preprocessing carried out was the same as in [67]. Firstly, all signals were resampled to 250 Hz to homogenize the sampling frequency of the datasets. Secondly, to obtain a baseline corrected signal, two median filters of 200 ms and 600 ms length were applied to obtain the baseline wander estimate which was then subtracted from the original raw one. Thirdly, a

finite impulse response (FIR) low pass filter with cut off frequency of 35 Hz and equal ripple in the pass- and stop-bands was applied to remove powerline and high frequency noise. The full preprocessing was performed on MATLAB 2020a, The Mathworks Inc.

3.2.3 Feature extraction

To classify each beat into one of the three categories considered in section 2.1, a set of features was extracted to describe two main properties of the ECG: the heart rate variability and wave morphology. To compute these features, the first and last 40 beats were not considered

HRV features

For each individual beat, a set of features was computed taking into account the neighbouring beats.

RR intervals are defined as the distance of two consecutive R peaks of each beat. dRRs are instead the series of the difference of consecutive RRs, namely $dRR_n = RR_{n+1} - RR_n$. Both the RR and dRR of the corresponding beat (RR_i, dRR_i), the previous beat (RR_{i-1}, dRR_{i-1}) and the following one (RR_{i+1}, dRR_{i+1}) were analysed. Four different time windows were considered for the extraction of the heart rate variability (HRV) features: 1 or 5 minute windows preceding the current beat; 2 or 10 minutes windows centered on the current beats. From each time window the mean and standard deviation of the RR intervals, along with the standard deviation of the dRR intervals, the percentage of successive interval differences greater than 10, 20, 30, 40 and 50 ms (pNN50) and the root mean square of successive differences (RMMSD) were computed. A total of 41 HRV features were measured.

Morphological features

Morphological information of the P wave, QRS complex, PR segments and the whole beat were extracted using, a fixed window. The window dimensions Using the R peak as reference (i.e. $t = 0$), the considered ECG segments were: [-300, 40] ms for the P wave segment, [-70, 60] ms for the QRS complex, [-288, 0] ms for the PR interval and [-300, 250] ms for the whole

beat [113]. The following segments will be referred as the P wave, QRS complex and PR interval, respectively throughout the rest of the paper. However, it should be noted that as no ECG delineation is performed, the reported segments may not precisely account for these ECG segments (i.e. it is not an exact selection of the onset and offset of the ECG waveform, but rather an approximate estimation). Nevertheless, the scope of this selection is to account for their intra-patient variability rather than to extract any precise parameter describing these ECG regions. Therefore, given that for the same patient the same ECG portion would be extracted for each of the mentioned segments, any variability produced by a premature atrial or ventricular beat, should be detected even if the ECG region is not accurately delineated. Prior extraction of the ECG segments, an intra-patient template was created using the neighbouring beats. Three different templates were computed using 80 (40 prior and 40 posterior the beat of study), 20 (10 prior and 10 posterior the beat of study) and 4 beats(2 prior and two posterior the beat of study) each. Three templates were computed to represent the instant beat differences with respect to the short-term neighbouring beats (4–beat template) and compare each beat with respect to the long-term (80-beat template) and the mid-term (20 beats). While the short-term could be especially useful for the detection of isolated PACs, the long and midterm could be more relevant in identifying PACs in bigeminy, trigeminy or in atrial runs.

The surrounding beats' segments were aligned through cross-correlation and then averaged. Outlier segments, according to the maximum cross-correlation value obtained for alienation, were excluded from the mean and thus, from the computation of the intra-patient template.

Once the intra-patient templates were computed, each beat of the subject is compared with the templates using cross-correlation. At the end of the process, for each beat, the following parameters were extracted and used as features:

- Maximum cross-correlation value of each segment with respect to the different intra-patient templates created with the neighboring beats (80, 20 and 4).
- Lag corresponding to the cross-correlation value described above.
- Median standard deviation of the beats used to create the intra-patient template.

The features enumerated above were computed for each lead of the ECG independently. In addition, these features were computed repeatedly for different intra-patient models that were build taking into consideration 2, 5 and 40 beats before and after the beat of study. A total of 50 features were computed for each beat.

Discrete Wavelet Transform Morphological features were computed also on a filtered version of the ECG obtained through the wavelet transform (WT). The WT for a continuous signal $x(t)$ if defined as follows:

$$W_x(s,b) = \frac{1}{\sqrt{s}} \int_{+\infty}^{-\infty} x(t)\psi\left(\frac{t-b}{s}\right) dt, s > 0 \tag{3.1}$$

This transform maps the input signal into the time-scale plane by means of the prototype wavelet function $\psi(t)$, dependent of the scaling (s) and translation (b) parameters. Low values of s enable the WT to localize fast transitions, whereas higher values localize coarser changes. Instead, the translation parameter b corresponds to their location [114].

A computationally feasible version of the WT is the discrete WT (DWT) which discretizes the time-scale by means of a dyadic sampling i.e. $s = 2^k$ and $b = 2^k l$ for $k, l \in Z$. The same implementation followed in [65] and implemented in [115] was performed. In [65] $b = l$ for $l \in Z$ so as to maintain the same sampling frequency in all scales. In addition, a quadratic spline was used as prototype wavelet $\psi(t)$, retaining ECG information at determined scales [114]. The fourth scale of the DWT ($W_x(4,l)$) retains useful information of the ECG [65]. This ECG decomposition was also used to extract morphological information described above in the *Morphological features* section. The resultant set of morphological features were composed by the same intra-patient cross-correlation information but computed using the filtered ECG signal and the ($W_x(4,l)$) ECG decomposition. A total of 72 features were computed for each beat.

3.2.4 Model definition and training

The selected classification model was a random forest (RF) evaluated in a patient-wise, 10-fold cross-validation i.e. no beats belonging to the same patient were included in the training and test set.

Train-test dataset definition

The dataset was divided into 10 subsets. Given the unbalance occurrence of PAC across the different patients, the different subsets were conformed so as to maintain a similar proportion of PACs in each fold. At each iteration of the cross-validation, 9 of the 10 subsets were used for training and the remaining one was used to test the performances. To prevent patient bias during training, an upper bound limit of 1000 beats of each class was set and used for training the algorithm.

Feature processing

To remove possible outliers an upper and lower bound was set for the features containing RR and dRR information. Values exceeding the established boundaries were reevaluated as the threshold value. Based on the cardiac refractory period, the minimum RR_{min} considered was 250 ms. On the other hand, if we accept 35 beats per minute (BPM) as the lowest possible heart rate (considering an extreme case of bradycardia), the corresponding RR interval would be $60s/35bmp = 1714.3ms$. However, it is known that after a PAC a refractory pause is caused due to the depolarization of the sinoatrial node, and that the maximum this pause can be is double the normal RR. Therefore, the maximum RR considered was $RR_{max} = 3428.6ms$ [56]

The dependence among the different variables was computed using Pearson's correlation coefficient. Correlation sequence were normalized so that the autocorrelations at zero lag were equal to one. The correlation coefficient $\rho_{X,Y}$ between two random variables X and Y with mean μ_X and μ_Y and standard deviation σ_X and σ_Y, respectively is defined as:

$$\rho_{X,Y} = corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{3.2}$$

where E is the expected value operator and cov attains for covariance. Features obtaining a correlation coefficient higher than 90% were excluded.

In addition, features with a variance lower than 0.05 were excluded. Finally, a z-score transformation was applied to the remaining features.

Model definition

RF is a supervised tree-based ensemble machine learning model trained with the "bagging" method. The concept behind bagging is that the combination of several weak simple classifiers can lead to high performance. RF builds a strong classifier by adding together simple decision trees. A strong advantage of this methodology is its resistance towards overfitting which is of great importance to reduce patient and database-dependent bias and ensuring the extrapolation of the model to other scenarios. A first train-test patient-wise split was performed for hyper-parameter tuning. The train set was composed by a random group of patients summing up to the 80% of PACs from the whole dataset, while the test set were all the remaining ones. A first random hyper-parameter search was carried out to prove the most suitable ranges. Grid hyper-parameter search was performed based on the results of the first random search. The best-performing hyper-parameters chosen were: number of estimators = 500, minimum samples for a split = 10, minimum samples for a leaf =2, maximum tree depth =20 and sample replacement in bootstrap aggregation = $False$. RF was implemented in Python 3.8 version using the Scikit-learn library version 0.24.0.

An outline of the full final model pipeline described is shown in Figure 3.1.

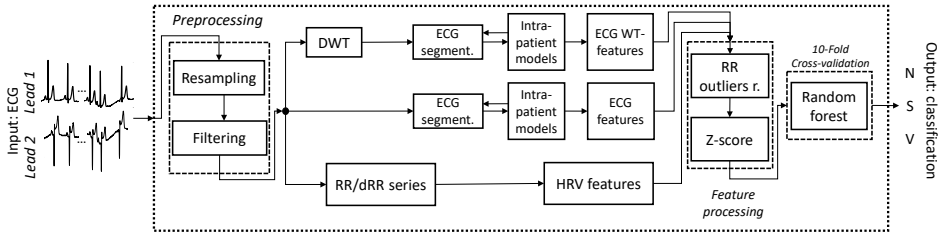

Figure 3.1: Outline of the final working classifier pipeline. The input consists of a single beat on a 2-Lead ECG and the output on the classification of the beat of one of the three categories N, S and V. $Outliers r.$ stands for outliers removal.

The model was both assessed as a PAC detector or binary classifier to discriminate S vs $Other$, and as a multi-class classifier for N, S and V discrimination following the same beat classifiers strategy present in literature. As stated above, S category included PAC with different notations across the two databases used (A, a and S), while $Other$ included both N (formed by B, N and e) and V (formed by V, F and E) categories.

Feature importance

Feature relevance was studied using different techniques. Firstly, as the mean and standard deviation of impurity decrease accumulation within each tree (Gini impurity) was computed for both the binary the muli-class classifier. Mean impurity decrease values were averaged across the models obtained after the 10-fold crossvalidation. The most relevant features were estimated as those achieving the highest mean crossvalidation values. These features corresponded to those with the highest Gini impurity decrease mean accumulation within each tree.

Secondly, permutation importance was computed for train and test sets of every k-fold of the muli-class classifier. This technique estimates feature importance by performing random permutations on the input features and analyzing how the R^2 score of the classifier is degraded. Each feature was randomly permuted 10 times.

Finally shapley additive explanations (SHAP) were computed on the test set of different k-folds of the muli-class classifier. SHAP is a method based in cooperative game theory used to provide explainability to machine learning models. In this work, the adaptation to tree models introduced and implemented by [116] was applied.

3.2.5 Poincaré Images

For the creation of the Poincaré Images of the segment classifier, beats were instead divided into two categories. Table 3.2 displays the beat annotations present in the databases and which of them were considered as A/V.

ECG signals were divided into 30-beat segments with 50% overlap using the QRS annotations provided in each database. From each segment, the RR and the dRR sequences were computed. While the RR sequence is defined as distance between subsequent R peaks i.e. $RR_i = R_i - R_{i-1}$, the dRR stands for the difference of subsequent RR intervals i.e. $dRR_i = RR_i - RR_{i-1}$.

The Poincaré Plot is a graphic representation of one RR_i (or dRR_i) plotted against the previous one, RR_{i-1} (or dRR_{i-1}). Poincaré Images were computed by binarizing the space of the Poincaré Plot for both RR and dRR as in [108]. Following the study in [108] the bin size was set to 40

Table 3.2: PhysioNet and simplified beat annotations per database.

Labelling	N				A/V							
Beat Anns.	B	N	e	Q	A	a	S	V	F	E	J	j
LTSTDB	88720	6727000	22	2	5482	29	30820	39840	476	71	1	6
SVDB	1	162100	0	80	0	1	12090	9930	23	0	9	0

B stands for bundle branche block beat, N for normal beat, A for atrial premature beat, a for aberrated atrial premature beat, e for atrial escape beat, Q for unclassifiable beat, S for supraventricular premature or ectopic beat (atrial or nodal), V for premature ventricular contraction, F fusion of ventricular and normal beat, E for ventricular escape beat, J for nodal (junctional) premature beat and j for nodal (junctional) escape beat.

ms and the Poincaré Images ranges were set to [0, 1600] ms and [-800, 800] ms in both x and y-axis for RR and dRR images, respectively. Poincaré Images containing at least one beat labelled as A/V were labelled as A/V whereas those not containing any beat of the mentioned categories were labelled as N.

3.2.6 Segment-wise classification model

A CNN was used for classification of the described Poincaré Images. Three different types of inputs were tested: RR, dRR and both. In all cases the model was a 2D CNN with 25 layers, kernel size of (3, 3), (1, 1) stride and ReLu activation function, followed by a max pooling layer, a fully connected layer with 100 hidden units and a final sigmoid activation layer. The model was trained using the Adam optimizer and binary cross-entropy as loss function.

3.2.7 Statistical analysis

For analysing the performance of the proposed methodology a 10-Fold cross-validation with patient-wise (i.e. no patient was both included in both the training and test set) train-test division was performed. A Kruskal–Wallis one-way analysis of variance test with Bonferroni Post-Host correction was performed to study the best input option for representing the differences between N and A/V from a patient-wise performance perspective.

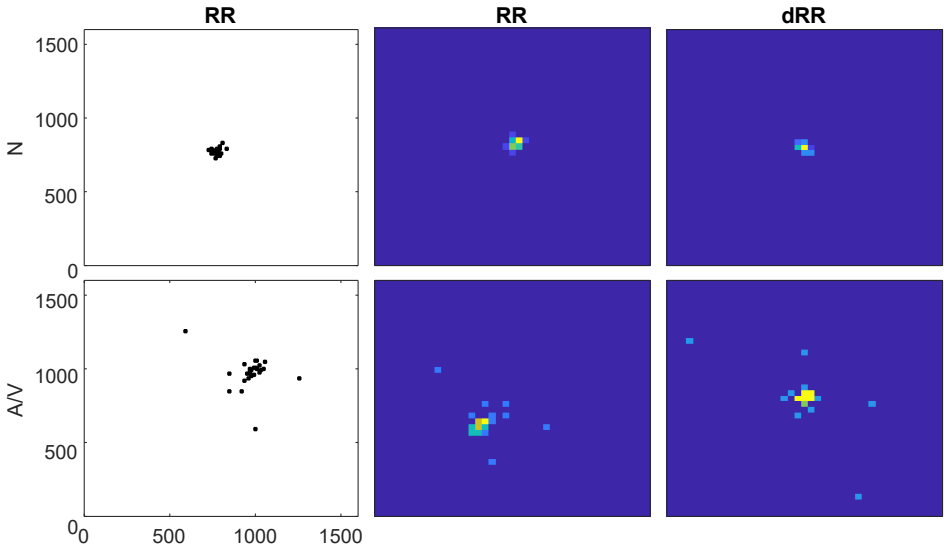

Figure 3.2: Poincaré Plot (first column) and Poincaré Image (second and third column) examples for RR (second column) and dRR (third column) configurations for 30-beat ECG segments for N (first row) and A/V (second row).

3.3 Results

3.3.1 Total features computed

A total of 185 features were computed for 6126250 N, 48032 S and 40312 V beats. For each beat, a total of 48 a intra-patient models were computed: 4 segments (whole beat, P-wave, PR segment, QRS complex), 3 beat windows used to construct each model (80, 20 and 4 beats) and 2 leads for both the raw signal and the $(W_4(l))$ of the DWT. Figure 3.3 presents an example of P-wave intra-patient models computed with 4, 10 and 80 neighbouring beats for the original and the WT of the ECG.

Figure 3.4 shows an example of an N, S and V beat and the corresponding P wave, PR interval and QRS complex. The corresponding intra-patient model built using the 40 anterior and posterior beats is also shown. It can be noted that for the N beat, the P wave, PR segment and QRS complex match almost perfectly the intra-patient model. In contrast, the S beat's P wave differs considerably from the intra-patient model, the PR segment slightly differs and the QRS complex almost matches it. Finally, all V intervals differ from the corresponding intra-patient models. PVCs are not

usually accompanied by a prior P wave. This is the reason why it does not match the neighbouring P waves.

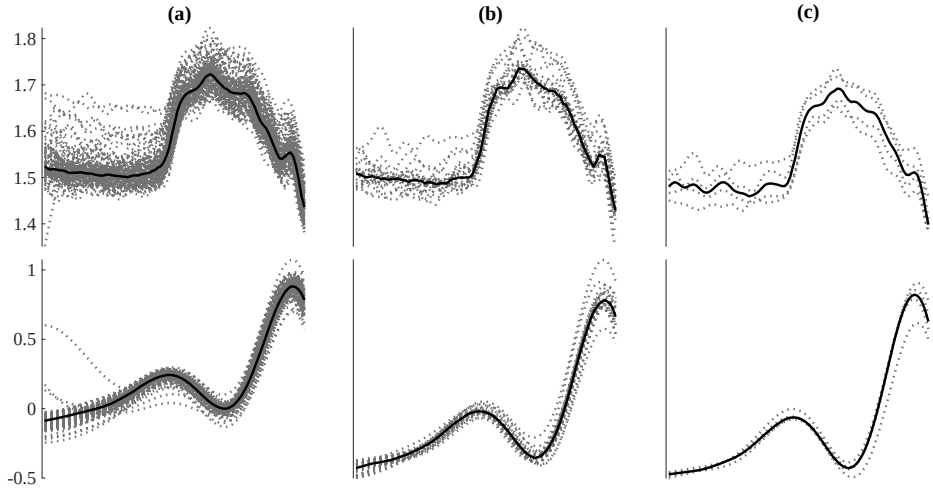

Figure 3.3: Example P-wave intra-patient models built using a different number of surrounding beats for the *raw* signal (top) and the $W_4s(l)$ DWT decomposition. Intra-patient models built using (a) 80 (b) 20 and (c) 4 beats.

3.3.2 Most important features

Feature importance was analyze at each Kfold for the binary and multi-label classification after elimination of the correlated (correlation matrix for k-fold 1 before and after elimination of features with a correlation value > than 90% are printed in Figures 5.10 and 5.11) and low variance ones (variance values are displayed on Figure 5.12 in the Appendix). A high number of correlated features were present, as could have been expected. Features with the highest variance were those accounting for RRis. Regarding the morphological features those with a higher variance were, in decreasing order: P-wave, whole beat, PR segment and QRS complex, as it can be seen in Figure 5.12.

The most important features according to the random forest model Gini impurity after 10 fold cross-validation for each Kfold of the cross-validation are gathered in Table 3.3 for the bi-label and multi-label approach. A total of 18 features conformed the 100 most important features (10 for each Kfold) for the multi-label classification and 16 for the binary one. The most important features were shared between both bi- and multi-label approaches

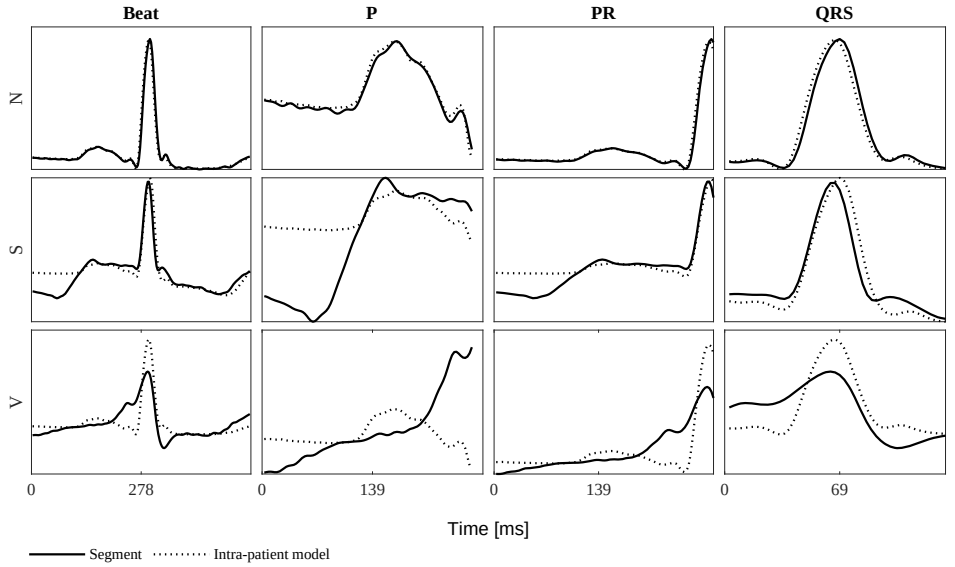

Figure 3.4: Example of each of the beat classes N, S and V for the whole beat, the P wave, PR segment and QRS complex. Each row represent one of the three beat types and each column the mentioned ECG segments. Continuous lines represent the beat segment itself whereas dashed lines represent the corresponding intra-patient models built with the 40 beats before and after the beat of study (80-beat intra-patient model).

with the exception of Beat cross-correlation of Lead 1 (L1) with 4-beat intra-patient template (intr.temp.) and PR cross-correlation of L1 with 80-beat intr.temp. that were only included as the 10 most important for the multi-label classifier and not for the binary one. The most relevant features were those accounting for RR disturbances, QRS complex and Beat morphology for the temporal ECG signal. Only three DWT morphological features were included as top features.

Permutation importance values computed on test and training set features of the multi-class classifier for K-fold 1 are displayed in Figure 3.5 and 3.6, respectively. Permutation values for the rest of k-folds computed on their corresponding test sets are gathered in the Appendix on Figures 5.1 - 5.9. Discrepancy between the most important features was found among the different k-folds and between training and test sets.

Shapley additive explanations (SHAP) computed for the multi-class classifier are displayed in Figure 3.7. Color indicates the actual value of the

Table 3.3: The ten most relevant features for each cross-validation KFold for the bi-label and the multi-label classification.

Bi-label classification	Multi-label classification
dRR_i	dRR_i
Beat cross-corr. DWT, L2, 20-beat intr.temp	Beat cross-corr. DWT, L2, 20-beat intr.temp
Beat cross-corr. L1, 20-beat intr.temp	Beat cross-corr. L1, 20-beat intr.temp
	Beat cross-corr. L1, 4-beat intr.temp
Beat cross-corr. L1, 80-beat intr.temp.	Beat cross-corr. L1, 80-beat intr.temp.
Beat cross-corr. L2, 4-beat intr.temp	Beat cross-corr. L2, 4-beat intr.temp
Beat cross-corr. L2, 80-beat intr.temp.	Beat cross-corr. L2, 80-beat intr.temp.
dRR_{i+1}	dRR_{i+1}
dRR_{i-1}	dRR_{i+1}
P cross-corr. L1, 80-beat intr.temp.	P cross-corr. L1, 80-beat intr.temp
P cross-corr. L2, 80-beat intr.temp.	P cross-corr. L2, 80-beat intr.temp.
	PR cross-corr. L1, 80-beat intr.temp
QRS cross-corr. DWT, L2, 20-beat intr.temp.	QRS cross-corr. DWT, L2, 20-beat intr.temp.
QRS cross-corr. DWT, L2, 4-beat intr.temp.	QRS cross-corr. DWT, L2, 4-beat intr.temp.
QRS cross-corr. L1, 4-beat intr.temp.	QRS cross-corr. L1, 4-beat intr.temp.
QRS cross-corr. L1, 80-beat intr.temp.	QRS cross-corr. L1, 80-beat intr.temp.
QRS cross-corr. L2, 20-beat intr.temp.	QRS cross-corr. L2, 20-beat intr.temp.
QRS cross-corr. L2, 4-beat intr.temp.	QRS cross-corr. L2, 4-beat intr.temp.
QRS cross-corr. L2, 80-beat intr.temp.	QRS cross-corr. L2, 80-beat intr.temp.
RRi0	RRi0

Intr.temp stands for intra-patient template. L1 and L2 refer to leads 1 and 2, respectively.

feature, while the x-axis value how this is related to SHAP value. A high SHAP value is translated into a higher probability for S beat detection. SHAP values for k-fold 2 are presented on Figure 5.13 on the Appendix to provide a comparison within different k-folds.

SHAP values provide a more complex and robust picture of each feature's contribution of the model's predictions. This is shown in the concordance of SHAP values across k-folds (see Figure 5.13 in appendix). Although the 20 most important features displayed in both graphs (Figures 3.7 and 5.13) do not totally coincide, they share a similar pattern. The three main features (accounting in descending order for dRR_i, RR_i and dRR_{i-1}) are shared between both k-folds. After these RRi features, QRS-complex morphology features followed. In both cases, P-wave morphology features were the sec-

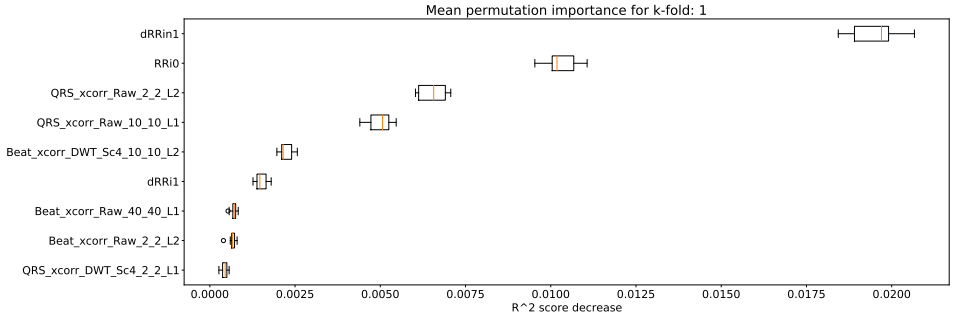

Figure 3.5: Permutation importance values for the 10 most relevant features on k-fold 1 model after 10 random permutations for each train feature. Were morphological features names are constructed as: (ECG part) _(xcorr or lag value)_(Raw(original signal) or 4th decomposition of the DWT)_ (number of beats used prior to beat) _ (number beats after beat, to construct intra-patient model template)_(lead). Temporal features names are instead: (temporal parameter)_(window considered before beat, in minutes)_(window considered after beat, in minuted)

ond most relevant morphological feature. Variables with a clear separation in their magnitude relation to S prediction were dRR_i, dRR_{i-1} and, remarkably, P-wave cross-correlation lag features. High lag values were related to a higher probability of S classification whereas low values to low S probability. In addition, P-wave DWT cross-correlation with intra-patient template of 20 beats ($P\_xcorr\_DWT\_Sc4\_10\_10\_L1$ in Figure 3.7) did also exhibit a clear magnitude separation. Features with mixed values in a single cluster could be due to interactions with other variables and features correlation.

Although with some discrepancies between k-folds, the most relevant features were those accounting for RR disturbances and QRS complex morphology.

3.3.3 Beat classifier performance

The proposed model was evaluated for a binary classification (S vs $Other$) for evaluating explicit PAC detection and for a multi-class classification, evaluating the proposed model as a beat classifier. It should be noted that, the number of PAC per patient varied considerably among patients and thus, not all had the same weight when accounting the classifiers' performance. Therefore, the performance of the model could be assessed in two

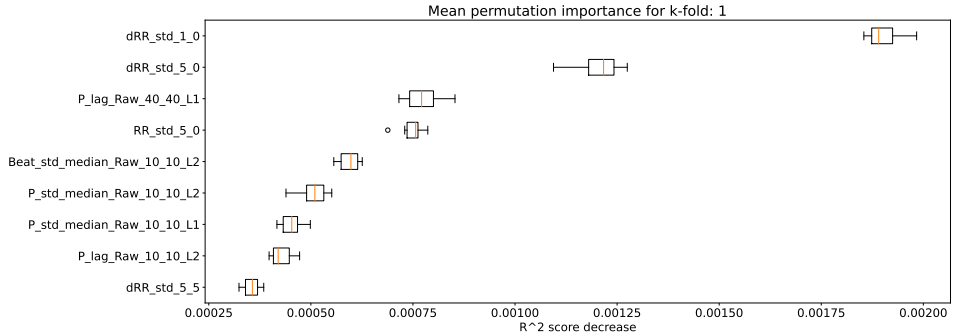

Figure 3.6: Permutation importance values for the 10 most relevant features on k-fold 1 model after 10 random permutations for each test feature. Were morphological features names are constructed as: (ECG part) _(xcorr or lag value)_(Raw(original signal) or 4th decomposition of the DWT)_ (number of beats used prior to beat) _ (number beats after beat, to construct intra-patient model template)_(lead). Temporal features names are instead: (temporal parameter)_(window considered before beat, in minutes)_(window considered after beat, in minuted)

ways: taking into consideration each beat as a separate sample, regardless of the patient (Table 3.5), or by averaging the accuracy, sensitivity and specificity values of every single patient regardless of its number of N, S and V beats (Tables 3.4, 3.6, and 3.7). In addition, patient-wise performance median and percentile values are provided considering a database to enhance comparability with other studies and to provide information about the database dependencies of the results reported.

Table 3.4: Patient-based classifier performance, median(IQR range).

	LTSTDB		SVDB	
	S	V	S	V
#Pat.	63	60	72	66
Acc. (%)	99.76 (99.37 - 99.92)	99.93 (99.68 - 99.99)	97.13 (90.49 - 99.39)	99.76 (98.98 - 99.95)
Se. (%)	95.65 (88.89 -100)	99.81 (96.83 -100)	96.20 (84.70 - 100)	100 (85.74 - 100)
Sp. (%)	99.76 (99.37 - 99.93)	99.95 (99.76 - 99.99)	97.74 (91.46 - 99.54)	99.84 (99.44 - 100)
PPV (%)	13.36 (2.29 - 44.56)	45.83 (2.72 - 91.73)	57.58 (25.03 - 80.73)	93.66 (67.95 - 98.87)
NPV (%)	99.99 (99.99 - 100)	100 (99.99 - 100)	99.84 (99.43 - 100)	100 (99.90 - 100)

Values are expressed as median (25th - 75th) percentile. Acc., Se., Sp. stand for accuracy, sensitivity, specificity, respectively.

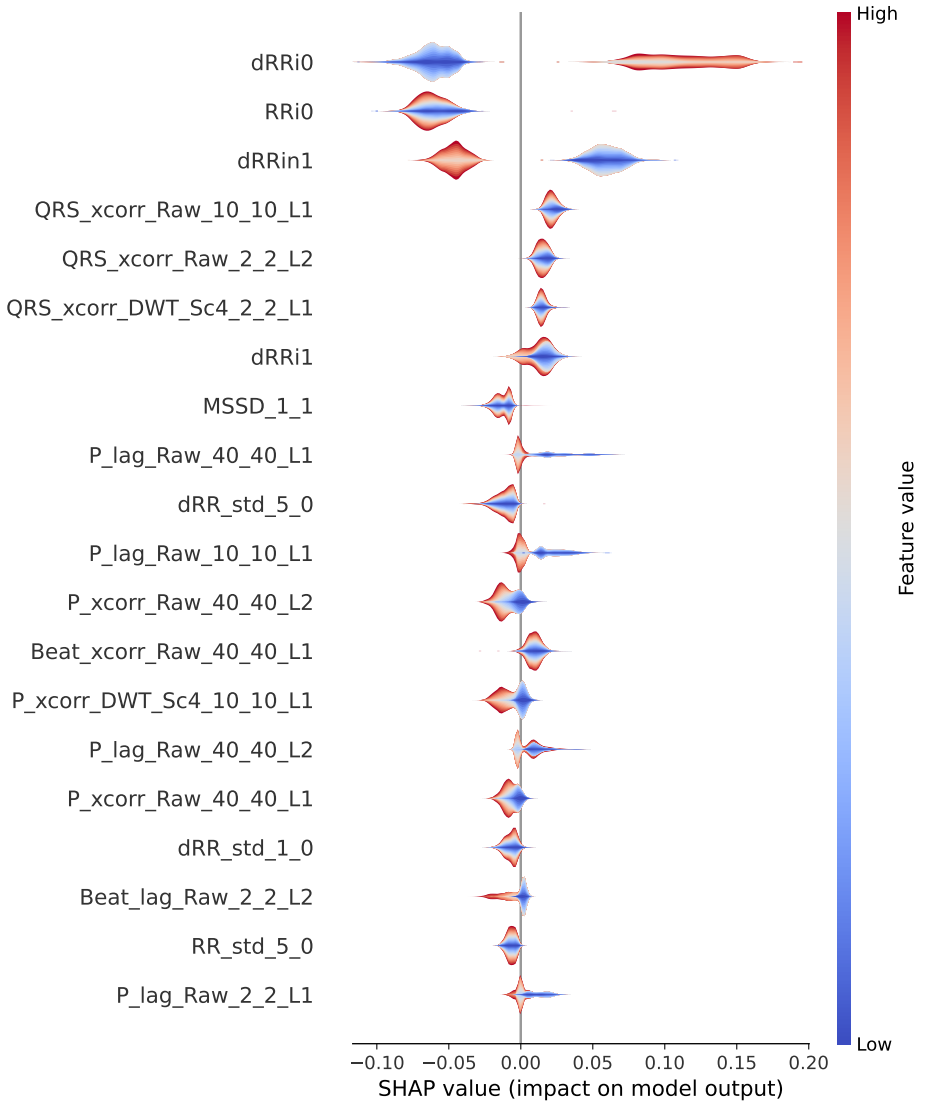

Figure 3.7: SHAP values for S predictions for k-fold 1 model. Were morphological features names are constructed as: (ECG part) _(xcorr or lag value)_(Raw(original signal) or 4th decomposition of the DWT)_ (number of beats used prior to beat) _ (number beats after beat, to construct intra-patient model template)_(lead). Temporal features names are instead: (temporal parameter)_(window considered before beat, in minutes)_(window considered after beat, in minutes)

Binary classification

The first row of Table 3.5 represents the accuracy, sentivity, specificity, positive and predictive value (PPV) value results for S detection from a beat-wise perspective; considering each beat as a sample independently of the patient it came from. Although sensitivity values were slightly lower than those reported in the same table for multi-label classification, the PPV was higher.

Table 3.5: Classifier performance considering single beats regardless of the patient.

	#Beats	Acc.(%)	Se.(%)	Sp.(%)	PPV(%)
S (binary class.)	48032	98.15	89.83	98.78	35.23
Normal (N)	6126250	97.88	97.90	96.57	99.85
Supraventricular (S)	48032	98.30	92.65	98.34	30.30
Ventricular (V)	40312	99.51	95.69	99.53	57.87

Acc., Se., Sp. stand for accuracy, sensitivity, specificity, respectively

Table 3.6 instead presents results from a patient-wise performance. Following the interquartile range (IQR) of the PPV presented in Table 3.6 a high patient-dependent influence can be intuited. Low PPV values even with high sensitivity and specificity are given by the extreme class imbalance of the dataset.

Table 3.6: Patient-based classifier performance, median(IQR range) for the binary classification (S - *Other*).

	LTSTDB (S)	SVDB (S)	LTSTDB+SVDB (S)
#Pat.	63	72	135
Accuracy (%)	99.84 (99.53 - 99.95)	97.82 (92.34 - 99.49)	99.48 (96.31 - 99.87)
Sensitivity (%)	94.12 (85.85 - 99.98)	91.78 (75.68 - 98.05)	92.86 (82.60 - 99.63)
Specificity (%)	99.85 (99.53 - 99.95)	98.63 (94.85 - 99.66)	99.57 (98.13 - 99.91)
PPV (%)	21.21 (3.36 - 45.07)	66.67 (31.19 - 82.88)	40.87 (14.90 - 73.36)

Values are expressed as median (25th - 75th) percentile.

Multi-label classification

Figure 3.8 and Figure 3.9 display the classification distribution for the LTSTDB and the SVDB, respectively. Each of the three sub-graphs shows the classification of one of the three beat categories. Bars represent the classification distribution of individual patients for that specific beat type, in percentage. For example from patient *s*20301, Figure 3.8 shows that all N beats (top subgraph) were correctly classified, around 20% and 10% of S beats (middle sub-graph) were misclassified as N and V, respectively and less than 10% of all V beats (bottom sub-graph) were misclassified as S. From both figures, it can be noted that most of the beats were correctly classified in all patients as it can be also appreciated by the overall results reported in Table 3.7. From results in Figures 3.8 and 3.9 it can be derived that S and V misclassifications have a strong patient-dependent component. Figure 3.10 shows the classification distribution computed for each patient independently as in Table 3.7. The presence of outliers show that even if the classifier attains very high performance for most patients, for some of them it fails to properly classify into the three categories. S sensitivity attained the highest inter-patient variability values.

Results in Table 3.7 show the median and IQR of the accuracy, sensitivity and specificity values for the patient-wise N, S and V classification performance. The three classes attained a sensitivity and specificity higher than 99%, with the exception of S sensitivity that was 95.83%. S sensitivity also attained a higher IQR than the rest of the categories. PPVs for the S class were inferior to those of the other categories, influenced by the presence of false positives and class imbalance. Multi-class S sensitivity was slightly higher than that of the binary classification. PPVs instead were superior and with a lower IQR for the binary classification than for the multi label one.

Finally, Table 3.8 shows the confusion matrix. It can be observed that the majority of false negatives for N and V were S whereas for S most of the false negatives were N.

3.3.4 Poincaré Images

Table 3.9 gathers the total amount of Poincaré Images computed for each database and type of image. In Figure 3.2 an example of N and A/V

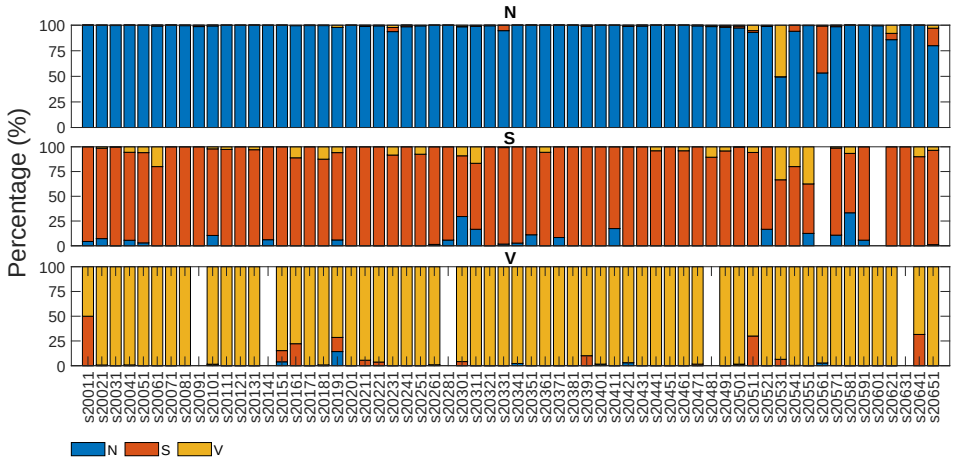

Figure 3.8: Classification percentage of each of the LTSTDB signals' beats. Each sub-graph represents the classification distribution of the beats of the three classes considered:N, S and V. Each bar in each sub-graph represents the total number of beats of that class of a single patient and how they have been classified (in percentage). The x-axis represent the different patient IDs.

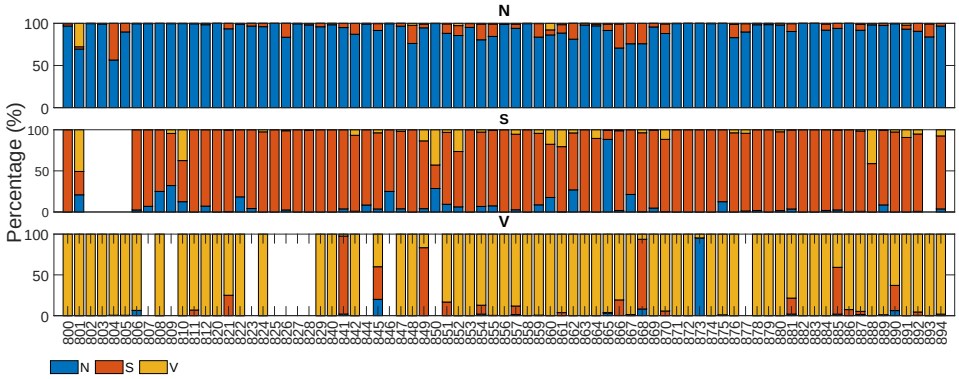

Figure 3.9: Classification of each of the SVDB signals' beats. Each sub-graph represents the classification distribution of the beats of the three classes considered:N, S and V. Each bar in each sub-graph represents the total number of beats of that class of a single patient and how they have been classified (in percentage). The x-axis represent the different patient IDs.

Table 3.7: Patient-based classifier performance, median(IQR range).

	Normal (N)	Supraventricular (S)	Ventricular (V)
#Pat.	139	132	123
Acc. (%)	99.05 (95.40 - 99.74)	99.35 (95.78 - 99.84)	99.87 (99.34 - 99.99)
Se. (%)	99.29 (94.96 - 99.78)	95.83 (87.50 -100)	100 (95.84 - 100)
Sp. (%)	99.54 (96.62 - 100)	99.39 (95.84 - 99.87)	99.90 (99.53 - 100)
PPV (%)	100 (99.78 - 100)	35.68 (9.63 - 69.57)	79.63 (15.71 - 97.39)

Values are expressed as median (25th - 75th) percentile. Acc., Se., Sp. stand for accuracy, sensitivity, specificity, respectively

Table 3.8: Confusion Matrix of total classified beats in percentage. The vertical and horizontal axis represent the true labels and predicted classes, respectively.

	Normal (N)	Supraventricular (S)	Ventricular (V)
Normal (N)	97.52 %	1.94 %	0.54 %
Supraventricular (S)	4.89 %	93.44 %	1.67 %
Ventricular (V)	0.69 %	3.79 %	95.52 %

Poincaré Images is shown for an RR Poincaré Plot (first column), an RR and a dRR Poincaré Image. N Poincaré Plot and Images condensed all samples in a single cluster as R peaks distance follow a regular behaviour. In contrast, as A/V beats disrupt the RR sequence, A/V images contained outlier points independent of the main cluster.

Table 3.9: Poincaré Images computed for each category per database.

Database (DB)	#Patients	N	A/V
SVDB	78	377156	53847
LTSTDB	80	3995	8295
Total	158	381151	62142

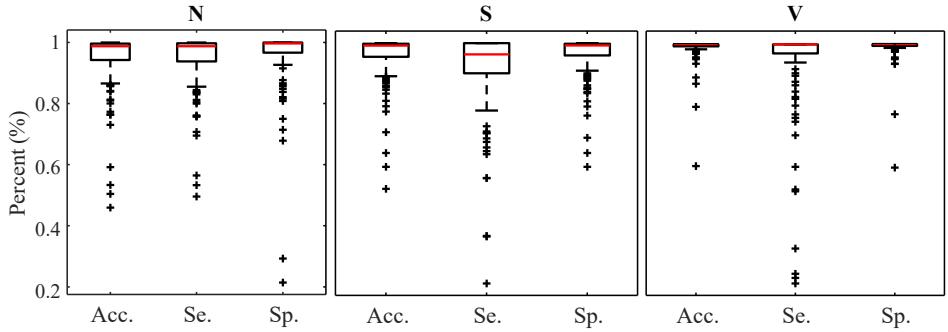

Figure 3.10: Accuracy Sensitivity and Specificity box plot for N, S and V detection calculated for each patient independently.

3.3.5 Segment-wise classifier performance

The classifier performance was evaluated from two different perspectives: element-wise i.e., each Poincaré Image was treated as an individual sample regardless of the patient it belonged to (Table 3.10) and patient-wise i.e., performance was calculated for each specific patient (Table 3.11). Analysis of classification not taking into account patient dependencies as in Table 3.10 disregards the fact that not all patients accounted for the same number of A/V images, and thus offers a biased view of the overall classifier performance. Results from patient-wise perspective in Table 3.11 are expressed as the median performance per patient together with the 75 and 25-percentile. No statistical significant differences between RR, dRR and $RR + dRR$ inputs were appreciated. On Figure 3.11 the patient-specific accuracy, sensitivity, specificity and PPV distribution is displayed. As it can be seen, although most patients exhibited high performance results, outliers were present in all metrics.

Table 3.10: Classification performance on the total of Poincaré Images classified in all K-folds.

	RR	dRR	RR + dRR
Accuracy (%)	95.94	96.16	96.05
Sensitivity (%)	92.00	94.88	93.84
Specificity (%)	96.59	96.36	96.41
PPV (%)	81.46	80.96	80.99

Table 3.11: Patient-wise classification performance. Values are expressed as median (75-percentile, 25-percentile).

	RR	dRR	RR + dRR
Accuracy (%)	97.90 (94.49 - 99.28)	97.29 (92.59 - 99.24)	97.55 (93.65 - 99.26)
Sensitivity (%)	96.03 (89.67 - 98.76)	95.59 (86.84 - 98.87)	96.10 (89.41 - 99.26)
Specificity (%)	98.70 (91.50 - 99.75)	98.12 (89.01 - 99.65)	98.31 (91.05 - 99.73)
PPV (%)	91.91 (70.87 - 99.24)	90.01 (64.22 - 98.86)	90.79 (67.78 - 99.12)

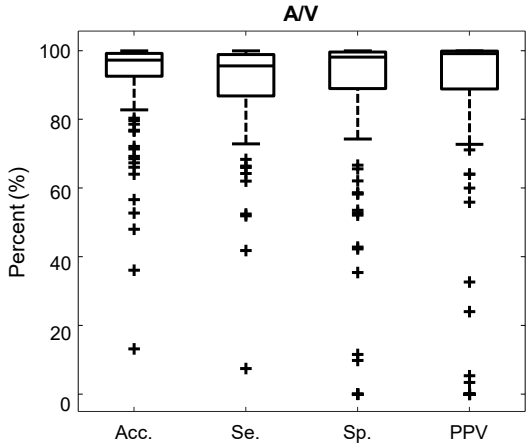

Figure 3.11: Boxplot of classification performance per patient for RR input Poincaré Images. Acc stands for accuracy, Se. for sensitivity and Sp. for specificity

3.4 Discussion

Explicit PAC detection in the ECG has not gained great attention, as it can be noted by the low number of papers solely addressing this problem [110, 111]. Rather, extensive literature can be found regarding a broader beat classification into supraventricular, ventricular and normal categories [66, 112].

PACs in the ECG are characterized by an alteration of the RR interval and P wave morphology. In [117] four different types of PACs are described based on the how they altered the RR interval. Whereas in [118] they explored the different P wave morphologies depending on the site of origin of the ectopic beats. The combination of HRV and morphology features included in the method proposed in this study aimed to take advantage of both characteristics. PAC and PVC share similar characteristics as far as HRV disturbance is referred. The main differences on the ECG trace

between PAC and PVC are the QRS and P wave morphology. However, whereas the QRS distortion can be easily and reliably distinguished, the P wave is susceptible to noise and its morphology may be easily altered by external sources rather than by an electrophysiological disturbance. Indeed, the feature importance selected by the proposed classifier displayed in Table 3.3 and Figures 3.6, 3.7 confirms the relevant role RR interval disturbances and QRS complex morphology have on the PAC and PVC detection capacity. Observing Table 3.3 one could intuitively guess that the model detects an RR alteration and it discriminates between PAC and PVC by checking if the QRS complex morphology is or not altered.

Even if the DWT has been proved useful for extracting relevant ECG information [66,114], according to Table 3.3 morphological features extracted from the DWT seemed to have a lower impact on the overall classifier's performance. While the morphological features obtained using the raw signal accounted for any morphological signal changes, morphological DWT features accounted for changes occurring only at a determined frequency band. Results suggest that the morphological changes induced in the 4th scale of the DWT captured by cross-correlation with respect to the intra-patient templates, were not as representative as those capturing morphological changes in the temporal signal.

It is known that ECG signals acquired from different patients have a considerable inter-patient variability. This dissimilarities hider the definition of universal measures that could serve as descriptors of eletrophysiological events (as PAC or PVCs). As a result, ECG delineators and beat classifiers attaining high performance across different patients and databases are rather challenging. By extracting morphological features that do not depend on precise measures but on the analysis of the evolution of the ECG signal itself, the inter-patient dissimilarity problematic is bypassed. Many classifiers as [67,69] depended on a proper ECG delineation to extract morphological information.

Two approaches were taken to study the proposed model: a proper PAC detector by discriminating between two categories (S - $Other$) and a beat classifier to discriminate among three different classes (N, S and V). Sensitivity values increased slightly for multi-label approach but with a reduction in PPV in comparison to the binary classification. Although PAC detection was the main target of the development of this model, results obtained for the multi-label approach shows that the classifier can be suc-

cessfully adapted to the detection of also ventricular beats without major performance degradation in PAC detection performance.

Furthermore, great care was taken in this work so as to not only maintain a balance among the three beat categories in the train set but also among the amount of beats belonging to different patients, in order to avoid a patient-biased trained model. As it can be seen in the results in Figures 3.8 and 3.9 as well as in Table 3.7, the detector performance varies among patients, evidencing the strong inter-patient influence on discriminating different beat types. One factor contributing to this could be the differences between lead placement on patients for acquiring Holter recordings. Different lead placement for Holter monitoring would influence amplitudes for the ECG segments, specially for regions as the P-wave. In patients with ECG signals attaining a lower P-wave amplitude, morphological distortions would be less evident and thus more difficult to detect.

Feature importance

Permutation importance and SHAP values were computed on the multi-class classifier to increase model explainability. Although impurity-based feature importance is computationally cheap (as it is integrated in the trained random forest model itself), it has some limitations; as it is evaluated on the same data the model was trained with, if the model is overfitted it can lead to misleading results. In addition, correlated features may affect importance estimations.

Permutation importance can be applied to the test set data, solving the problem of using the same training data to explore feature importance. However, it does not overcome the correlation problem. Comparing permutation importance results on the training and test set of the same k-fold can provide some insight about the model's features and training. The differences in feature importances found between both graphs (Figures 3.6 and 3.5) may point to the existence of overfitting on the model or to the presence of correlated features. As SHAP values agree with those of Figure 3.5, it is more possible that differences in permutation importance between train and test are given by the existence of correlated featured. Nevertheless, as a 10-fold crossvalidation was being used to estimate the performance of the proposed methodology, the threshold for correlation filtering was kept high so as to maintain the same features at each k-fold and prevent biased results

towards an specific dataset split. No further feature selection method was applied for the same reason.

SHAP values target both problems but are computationally expensive and their calculation is less intuitive. Values from Figure 3.7 agree with those in Table 3.3 and Figure 3.5 about the relevance of RR and QRS morphological features. Moreover, it also points at the role of P-wave features on S beats detection. This is coherent with the distortion PACs induce on P-wave morphology. In specific, the role of P-wave cross-correlation lag values seemed to provide a higher classification power than the cross-correlation values themselves.

3.4.1 Related work

From the published methods, an initial distinction can be made based on if a proper patient-wise train-test division was made. As demonstrated by [66], there exists a strong bias introduced in algorithms trained and tested with beats belonging to the same patients. A second distinction can be made based on the database used for testing the methodology. The Association for the AAMI guidelines recommend the open-source MIT-BIH Arrhythmia database available at Physionet as a common framework for reporting performance, as it is the only one that contains the five superclasses of arrhythmias. However, as discussed by [112] this database is highly unbalanced and provides misleading results about supraventricular and ventricular beats detection. A standardized train-test division of the MIT-BIH arrhythmia database was proposed by [67], which has been used by many authors as [69–72]. However, most PACs and PVCs occur in single patients in both sets and extrapolation of the performance to other patients is rather doubtful. [66] performed an exhaustive analysis about how the databases used for testing changed significantly the performance reported by the same methodology. Therefore, it is important to understand that comparison between algorithms is not trivial and that it should be interpreted with care.

De Chazal et al. [67] used RR intervals and morphological information of the segmented ECG as features and LDs models as classifier. They used for training and testing the MIT-BIH arrhythmia database divided by the standard DB1 DB2 introduced by themselves. They reported a sensitivity of 75.9%, a PPV of 38.5% and a FPR of 4.7%. In two studies, [65, 66] de-

veloped a classifier including RR interval and morphological features from different scales of the DWT. In a first study [65] used a LDC and tested their method on the DS2-Test set of the MIT-BIH obtaining a SVEB sensitivity of 77% and a PPV of 88%. In addition, they also tested the methodology on the whole MIT-BIH Arrhythmia database reporting a SVEB sensitivity of 61% and a PPV of 73%. In a second study [66] they used up to 8 public databases, among which were the SVDB and the LTSTDB (containing the 82.01% of the total PACs) to train and test their model. They integrated their previously developed classifier together with an unsupervised clustering method. In addition they enabled it to be assisted, semi-assisted or automatic. They obtained a sensitivity and a PPV of 76% and 43% in the full MIT-BIH in automatic mode that increased up to 89% to 88%, respectively, in assisted modality. Similarly in the SVDB they obtained 46% and 50% that increased to 74% and 79% sensitivity and PPV in automatic to assisted, respectively. Finally in the LTSTDB they obtained 50% and 8% sensitivity and PPV in automatic and 51% and 58% in assisted. To the best of our knowledge [66] are the only ones reporting results using the SVDB and LTSTDB as test set.

The classifier presented in this work attained a sensitivity and PPV of 94.12% and 21.21% for the LTSTDB and 91.78% and 66.67% for the SVDB as shown in Table 3.6. Results for both databases were considerably higher in terms of sensitivity in comparison with those reported in [66] for the automatic and assisted classification. PPVs were higher that the ones reported by [66] only in the fully automatic mode. Nevertheless PPV should be interpreted with care as, given the high class imbalance (of more than 2 orders of magnitude) S PPVs would increase if S sensitivity was equal to zero, thus not detecting PACs at all. Therefore, sensitivity values ought to be taken in consideration together with the PPV. The beat classifier presented by [66] attains a higher PPV in the SVDB (76%) in the assisted mode. However, the reported sensitivity in the LTSTDB using that same methodology reaches only a 51% which would not make it suitable for PAC detection. It is evident that using the proposed methodology or the one presented by [66] is a matter of trade-off with regard to the amount of false positives or false negatives as far as PAC detection is concerned.

On the other hand the sensitivity and PPV values obtained for V classification shown in Tables 3.7 and 3.4 (100% and 93.66% for the SVDB and 99.81% and 45.83% for the LTSTDB) were superior to those reported by [66] in the automatic mode (sensitivity and PPV of 82% and 54% per-

cent for the SVDB and 43% and 11% for the LTSTDB). For the assisted mode they obtained a sensitivity and PPV of 88% and 90% for the SVDB and 95% and 99% for the LTSTDB. It should be noted that even if some performance values are higher for S of V detection for the assisted version of the classifier presented in [66], this version requires of the intervention of a user to verify the final classification which could induced human errors as well as variability among users in the final classification. These results suggest that the proposed model could be used not only for PAC, but also for V detection, obtaining performance results for beat classification higher than those present in literature.

The proposed model enabled the detection of almost all PACs included in this study, implying an step ahead in PAC detection, as the available methods' sensitivity was always kept low [65–72]. In addition, PPVs were higher than those reported in literature for the databases included in this study and for fully automatic algorithms. In order to reduce false positives, the exploration of the the integration of an unsupervised learning classifier to the one presented in this work as [66] did, could be considered for future work.

Segment-wise classifier

With the aim of reducing the amount of PAC false positives detected, we worked on a classifier that could serve as a prior filter to a beat classifier. In this work a novel method for the detection of ECG segments containing atrial or ventricular beats was presented. The methodology relies on a 2D representation of RR intervals to capture their non-linear dynamics. The fact that it solely relies on R peak detection, makes this methodology more robust against noise presence than those based on morphological waveform analysis. Performance was analyzed from both a patient-wise and an element-wise point of view (Table 3.10 and 3.11, respectively). Three different Poincaré Image inputs were studied without appreciation of significant differences in patient-wise performance. In all cases, accuracy was superior to 96%, sensitivity higher than 93% and PPV superior to 80%. Given the unbalance nature of the dataset, as shown in Table 3.9, the decrease in PPV accounted for false positives in a greater proportion than false negatives (as sensitivity values were kept high).

This work intended to introduce a methodology to identify regions of interest of the ECG to localize A/V beats. Current beat classifiers focus on single beat classification, most attaining specially low performance on atrial beat detection [66, 119] and thus, not suitable for the study of PAC implications. ECG segment classification could ease the selection of signal regions containing PACs or other types of A/V beats and serve to reduce false positives and identify possible false negatives. Differently, for manual beat annotation, it could reduce considerably the segments of the ECG for the clinician to analyse. As a future work other beat-windows could be explored. In addition, it could be explored how the implementation of this method as a prior classifier to with beat-to-beat classifiers could influence their performance.

3.5 Conclusion

In this work three classification methods for PAC detection are exploited. On one hand a multi-class classifier for N, S and V a binary classifier for S and *Other* beats is presented. In contrast with many methodologies present in literature, the developed methodology does not require of ECG delineation. Although comparison among methodologies presented in different studies is not trivial, the former method outperforms in terms of sensitivity and PPV the state-of-the-art models. Further efforts should be made in order to decrease the inter-patient variability, increase the PPV and reduce false positives so as to be able to use the former method in clinical trials.

On the other hand, an ECG segment classifier is introduced for the detection of segments containing A/V beats. The results obtained suggest this methodology could be used to reduce the ECG beat annotation workload to study PACs and other A/V implications. In addition, it could be used in combination with automatic beat classifiers to find misclassifications and obtain more reliable annotations.

Chapter 4

AI for Cardiac Disorders Automatic Annotation in the 12-Lead ECG

4.1 Motivation

Cardiovascular diseases (CVDss) are the first cause of mortality and morbidity worldwide [120]. The standard 12-lead ECG is an essential tool in clinical practice to diagnose CVDs and to have an initial assessment of a patient's health condition [56]. Typically, manual interpretation of ECGs by expert clinicians is needed. This requires skilled personnel with high degree of specialization and in some cases inter-operators discrepancies are present. An automatic algorithm for detection of cardiac abnormalities and classification of ECG recordings could aid clinical practice, providing clin-

icians with an objective tool to make an early and accurate diagnosis of CVDs [121]. The aim of this work was to put together artificial intelligence techniques and physiological know-how to build and validate such model. Two different approaches were investigated.

Firstly a machine learning ensemble model was developed using 12-lead ECG signals from the 6 annotated datasets available for the 2020 Computing in Cardiology Challenge [16] capable of targeting a limited number of pathologies. Secondly, a deep learning model was developed capable of detecting 26 pathologies ECG signals with different number of leads, signal lengths, acquisition centers and devices using the PhysioNet/Computing in Cardiology Challenge 2021 multi center ECG database [16–22]. This deep learning study aimed to focus on the nature of the detected pathologies: some affect the *morphology* of the ECG, others its *rhythm*, some both. A combination of a modified ResNet with another branch fed with temporal parameters was exploited to leverage both morphological and rhythmic properties of the pathologies, and to improve the classification capabilities of the neural networks. Little can be found in literature about deep+wide integration and training. In this study we presented an analysis of the best training strategy for a model integrating both a deep network and classic machine learning features emphasizing in the integration of both types of information. Moreover, in contrast with [122–124], the presented model further elaborated the integration of the wide features by changing the network structure. Integrating a wide branch into a deep network could increase a model's explainability, provide us with the possibility to force a network to emphasize on determined features or characteristics of a dataset and to increment the generalization capabilities of a deep network.

4.2 Material and Methods

4.2.1 Total data available

Eight different datasets were made available by the PhysioNet CinC Challenge 2020 and 2021 (for further details the reader is referred to [16–22]) with a total of 88253 subjects associated with 133 labels that represent different pathologies. Only 30 out of 133 were of interest. The number of recordings per scored label of each dataset is shown in Table 4.1. Four pathology pairs were considered equivalent, as shown in Table 4.1. Recordings containing

any of these classes were relabeled, and only 26 out of the 30 scored labels were maintained. Signals were acquired with a sampling frequency ranging from 257 Hz up to 1000 Hz, and were of variable duration, with recordings of 5 s, 10 s, 120 s and 30 minutes long.

Table 4.1: Distribution of labels per dataset. Couples that are considered equivalent are highlighted.

Labels	CPSC	CPSC-Extra	St.Pet.	PTB	PTB-XL	Georgia	Chap.-Shao.	Ningbo	Total
AF	1221	153	2	15	1514	570	1780	0	5255
AFL	0	54	0	1	73	186	445	7615	8374
BBB	0	0	1	20	0	116	0	385	522
Brady	0	271	11	0	0	6	0	7	295
CLBBB	0	0	0	0	0	0	0	213	213
CRBBB	0	113	0	0	542	28	0	1096	1779
IAVB	722	106	0	0	797	769	247	893	3534
IRBBB	0	86	0	0	1118	407	0	246	1857
LAD	0	0	0	0	5146	940	382	1163	7631
LAnFB	0	0	0	0	1626	180	0	380	2186
LBBB	236	38	0	0	536	231	205	35	1281
LQRSV	0	0	0	0	182	374	249	794	1599
NSIVCB	0	4	1	0	789	203	235	536	1768
NSR	918	4	0	80	18092	1752	1826	6299	28971
PAC	616	73	3	0	398	639	258	1054	3041
PR	0	3	0	0	296	0	0	1182	1481
PRWP	0	0	0	0	0	0	0	638	638
PVC	0	188	0	0	0	0	0	1091	1279
LPR	0	0	0	0	340	0	12	40	392
LQT	0	4	0	0	118	1391	57	337	1907
QAb	0	1	0	0	548	464	235	828	2076
RAD	0	1	0	0	343	83	215	638	1280
RBBB	1857	1	2	0	0	542	454	195	3051
SA	0	11	2	0	772	455	0	2550	3790
SB	0	45	0	0	637	1677	3889	12670	18918
STach	0	303	11	1	826	1261	1568	5687	9657
SVPB	0	53	4	0	157	1	0	9	224
TAb	0	22	0	0	2345	2306	1876	5167	11716
TInv	0	5	1	0	294	812	157	2720	3989
VPB	0	8	0	0	0	357	294	0	659

Sensitivity (Se.), Specificity (Sp.)

4.2.2 Machine learning approach

Data

Six databases were used: China Physiological Signal Challenge in 2018 (CPSC2018), St Petersburg INCART 12-lead Arrhythmia Database, the PTB Diagnostic ECG Database, the PTB-XL electrocardiography Database and the Georgia 12-Lead ECG Challenge Database. In total, 43,101 labeled 12-lead ECGs with 111 different cardiac abnormalities labelled following the SNOMED-CT coding system. Only 24 disorders were targeted: IAVB, AF, AFL, Brady, IRBBB, LAnFB, LAD, LFBBB, LQRSV, NSIVCB, PR, PAC, PVC, LPR, LQT, QAb, RAD, sinus arrhythmia SA, SB, NSR, STach, TAb and TInv, the rest were considered "Unscored" and ignored. For the first phase, also ST elevation (STE) and depression (STD) were considered.

ECG preprocessing

Two median filters of 200 and 600 ms were applied to obtain the baseline of each ECG signal that was then subtracted to the original ECG to obtain a baseline corrected signal. Power line and high-frequency noise were removed with a finite impulse response low-pass filter with equal ripple in the pass and stop bands. The 3-dB point of the filter was 35 Hz. R peaks were detected using Pan Tompkins' algorithm [125]. To minimize misdetections, R peaks were detected on 5 leads with positive QRS (I, II, III, aVF, V4, V5, V6). An impulse train signal was built based on the R peaks detected for each lead; a Gaussian filter was applied to smooth the R series, then signals were aligned through a cross-correlation procedure, using V5 as reference and finally the median signal among the 5 series was computed. On the obtained signal, the Pan Tompkins algorithm was applied to obtain the R peaks used on the rest of the analysis.

Heart rate variability features

The mean and standard deviation of the RR intervals was computed, along with the standard deviation of the difference of consecutive RR intervals, the percentage of successive interval differences greater than 50 ms (pNN50) and the RMMSD.

Intra and inter-patient models

Inter-patient models that gathered the median behaviour of each electrophysiological disorder in the ECG were built for different signal segments of clinical significance. These models were used to measure the distance of each patient to the ECG pattern typical of each electrophysiological disorder. Firstly, ECG windows of interest were selected based on the expected occurrence of different electrical events (all values referred to the location of R peak): P-wave (from -300 ms to -40 ms) [113], QRS-complex (-70 ms to +60 ms), PQ-ST (PT-segment with the removal of the QRS-segment, from -150 ms to +250 ms, removing the segment from -55 ms to + 55 ms), T-wave (from +100 ms to +350 ms) [126], PR (from -288 ms to the R peak), RT (from the R peak to +258 ms). Secondly, intra-patient templates were computed for each of the above-mentioned segments. For each lead, all windows were aligned and the median waveform, representative of the analyzed subject, was calculated. The mean standard deviation of all windows was stored as a feature. Thirdly, the maximum cross-correlation index was computed between each individual region of interest and the median intra-patient template. Fourthly, for each rhythm, the templates of the correspondent subjects of the training set were aligned and the median rhythm template was obtained. Finally, the maximum cross-correlation between each intra-patient templates and each of the inter-patient models with the corresponding lags were stored as features. In Figure 4.1, an example is visible for the RT segment for NSR, RBBB and LBBB. On the left panels, all RT segments in the recording of a subject are depicted with the obtained intra-patient template in red. On the right, the median intra-patient templates of all the subjects are plotted with the resulting inter-patient template in red.

GEH features

Global Electric Heterogeneity (GEH) features were computed using the three open source toolboxes provided by the challenge: HRV toolbox [127], ECGkit [115] and GEH parameter extraction [128, 129] and vector origin point [130]. These features included: azimuth, elevation, and magnitude of spatial peak QRS, T and spatial ventricular gradient (SVG) vectors, az-

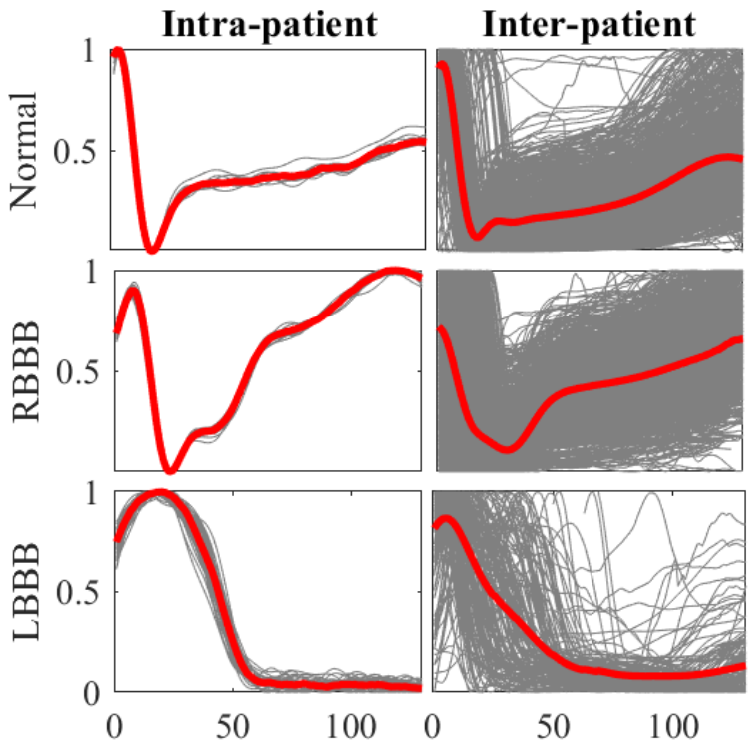

Figure 4.1: Intra and inter-patient templates for the RT segment of the rhythms Normal, RBBB and LBBB. All the segments of a recording are overlapped in the intra-patient plots, whereas in the inter-patients plots, the median segments of all the subjects with the respective condition are depicted. The median segments are in red.

imuth and elevation of QRS, T vectors and Wilson's area SVG, scalar value of the SVG, and peak and are of the spatial QRS-T angle [128, 129].

Classifier ensemble

The classification strategy adopted was to create an ensemble of classifiers, each of them focused on a specific region of the cardiac conductive system. A distinction was made between punctual events (PAC and PVC), and permanent conditions maintained along the whole ECG recording (all the other pathologies). As most signals ranged from 6 s to 60 s, all arrhythmias were considered present on the whole signal.

During a first phase only 9 classes were taken into consideration (AF, AFL, LBBB, PAC, PVC, RBBB, STE, STD, IAVB). The detection of per-

manent conditions was distributed into different classifiers depending on the ECG region where the electrophysiological disorders were manifested (Table 4.3). This yielded one classifier (C1) including the cross-correlation values between inter-patient and intra-patient of the P-wave and PR interval, and another one (C2) including the QRS complex, RT segment and T-wave information. Both classifiers included GEH and HRV features, as well as age and sex. In a second phase, the 24 categories were targeted. Given the increase in the amount of rhythms included, the detection of permanent conditions was distributed differently, according to the cardiac region of occurrence. Furthermore, for each of the sporadic conditions (PAC and PVC) an individual classifier was designed in both phases. This led to a total of 4 classifiers in phase 1 and 7 in phase 2:

C1: Atrial tissue

C[2-4]: Purkinje system

- C2: Sinoatrial node
- C3: Atrioventricular (AV) node
- C4: Bundle branches

C5: Ventricular tissue

C6: PAC

C7: PVC

Age, sex, HRV and GEH features were used by all classifiers. In contrast, the cross-correlation features included varied among the different classifiers: C1 and C3 included features for the P-wave and PR segment; C2 included features for all signal segments; C4 and C5 used QRS complex, PQ-ST segment, T wave, and RT segment features. C6 and C7 included all HRV features and the cross-correlation values between the intra-patient template and the 3 beats with the shortest preceding RR interval for the QSR and P-wave segments.

Signals classified as AF or AFL were not taken into consideration for possible PAC or PVC. In phase 1, C1 and C2 were support vector machines. PAC and PVC classifiers used bagged trees. For the second phase, in which 24 categories were included, all classifiers used boosted trees. All classifiers were implemented in Matlab 2020a and trained using a 10-fold cross-validation.

4.2.3 Deep learning approach

Dataset preparation

A database selection procedure was adopted in order to maximize performances during the training process. The CPSC-Extra, StPetersburg and PTB datasets were excluded due to the poor quality of their recordings. While CPSC-Extra's and PTB's population of acquisition was reduced, the StPetersburg database contained very long signals in comparison to the other databases (30 minutes vs the 10 seconds-mean signals of the rest of the databases). The exclusion of these three sources reduced the number of recordings available to 84210. Unscored signals were discarded and only 10 seconds-long subjects with 500 Hz sampling frequency were considered, reducing the final dataset to 67659 signals. A stratified holdout of 12.5% of training data was performed to create a local test dataset. In addition, given the high unbalance of recordings towards NSR and SB, three different training datasets were created. Each of these datasets contained one third of SB and NSR recordings and all the rest of recordings. This yielded three final training datasets containing 40365 recordings each.

Data Processing

Figure 4.2 summarizes the ECG preprocessing steps prior to the feeding of the classification model. Two preprocessing branches were considered - Deep and Wide branch - whose output was used to feed the Deep and Wide branches of the classification model, respectively. Before feeding the Deep branch, the recordings were resampled to 500 Hz, filtered (Butterworth filter, 3rd order, frequency range [1 47] Hz) and finally each channel was standardized by computing its z-score. Signals shorter than 5000 samples (10 seconds) were zero padded and those with a higher length were clipped at random. Moreover, data augmentation was performed by applying the following distortions:

- Random gaussian noise with standard deviation 1 addition with probability 0.06
- Leads exchange (interchange between any of the available leads) with probability 0.02

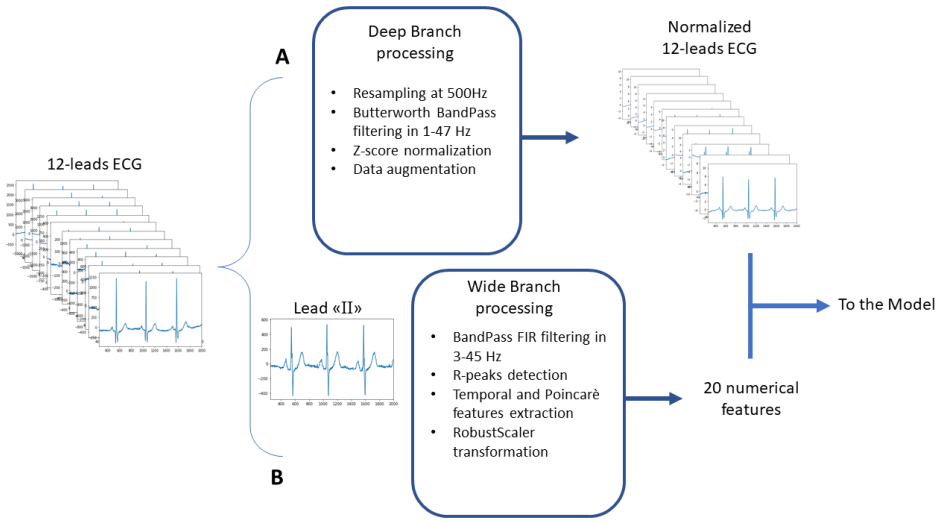

Figure 4.2: The two flows of information. A. Processing of ECG before introducing it into the ResNet; B. Extraction of Machine Learning features for the wide branch. The two branches are then concatenated to produce the output class probabilities.

- Channel inversion (multiplication of the channel by -1) with probability 0.02

The procedure was applied switching at every batch and every epoch the signals modified. This elaboration aimed to improve the model generalization capability towards noisy signals.

To compute the features introduced into the wide branch, R peaks were detected from lead II using Christov's adaptative threshold technique [131] implemented in the bioppsy python library (version 0.6.1). A set of temporal features were computed from the RR sequence and its representation on the Poincaré Plot. All the wide features computed are listed in Table 4.2. In the *Parameter* column the names given to each parameter in the code is listed and described in the column *Description*.

Model Architecture

The model contained two branches: a *Deep* branch formed by a modified ResNet and a *Wide* branch integrating 20 machine learning features embedded into three sequential dense layers. Both branches were concatenated and fed to a final dense sigmoid layer. The full model architecture is represented

Table 4.2: List of the 20 features extracted.

Parameter	Description
mean_nni	*Mean NN interval*
sdnn	*NN interval std* [a]
sdsd	*Std of adjacent NN-intervals differences*
nni_50	*Number of NN exceeding 50ms*
pnni_50	*Percentual of nni50 over total RR intervals*
nni_20	*Number of NN exceeding 20ms*
pnni 20	*Percentual of nni20 over total RR intervals*
rmssd	*RMS* [b] *of NN differences*
median_nni	*Median NN interval*
range_nni	*Max-Min difference*
cvsd	*CV* [c] *of successive differences*
cvnni	*Coefficient of Variation*
mean_hr	*Mean Heart Rate*
max_hr	*Max Heart Rate*
min_hr	*Min Heart Rate*
std_hr	*std of Heart Rate*
sd1	*Std of the major axis*
sd2	*Std of the minor axis*
sd_ratio	*Ratio between sd2 and sd1*
ellipse_area	*Area of fitted ellipse of Poincarè plot*

[a] Standard Deviation, [b] Root Mean Square, [c] Coefficient of Variation

in Figure 4.3. This structure intended to fuse the ResNet's capacity to extract implicit complex ECG morphology patterns together with the explicit cardiac rhythm features.

The *Deep* branch was a modified ResNet [132] with Dilation Convolution layers and Squeeze-and-Excitation Blocks (SEBs). Two different residual Blocks conformed the ResNet structure: the identity and the Convolutional block. The network and specific Blocks composition are shown in Figures 4.3 and 4.4, respectively. The Identity Block was composed by two convolutional layers with variable filter size and dilation rate, followed by an SEB and an additional layer that added the original input of the Block and the output of the dilated convolutions. The dilation was used in the convolutions to expand the distance considered between the values of the kernel to reach a larger field of view without losing resolution. A dropout layer was inserted after the first batch normalization layer for regularization

purposes. The Convolutional Block followed the same structure than the Identity Block but containing an extra convolutional layer as shown in Figure 4.4. In addition, no dilation was performed in the convolutional layers conforming this Block. The SEB performed a feature squeezing through a global average pooling layer, followed by two dense layers with ReLU and sigmoid activation functions that added non-linearity. Finally, a multiply layer reintroduced the original input of the Block and multiplied it by the output of the last dense layer. This module was based on the channel attention intuition: it applies a weight to each channel based on their importance before propagating to the next layer. The described Blocks were integrated as shown in Figure 4.3. The standardized ECGs (output of flow A of Figure 4.2) were fed to an initial convolutional layer with a kernel size of 15 and dilation rate of 4. Identity and Convolutional Blocks were alternated following the typical ResNet structure, maintaining the kernel size and doubling the dilation and filter numbers every two Blocks. A total of 8 residual Blocks were integrated into the network architecture. The last Identity Block was followed by a global max pooling layer.

The wide branch input was the 20 features (flow B of figure 4.2) introduced into a fully connected (FC) network made of 3 dense layers of 20, 15 and 10 neurons each with ReLU activation function. The FC network acted as bottleneck forcing the model to learn an embedding of the input features. The output of the Deep and Wide branches were concatenated and introduced to a final Dense sigmoid layer.

The final model was an ensemble computed by the majority voting of three different models with the architecture described in Figure 4.3. Each of them was trained on one of the three different subsets with a more balanced distribution. A different ensemble model was created for each lead subset. All models maintained the same architecture but with different input sizes, relative to the number of input leads.

Model Training

The deep and wide branches output were 512 and 10 features, respectively. Given the unbalance of trainable parameters and output features from the wide and deep branches, a three-step training was performed, allowing the network to focus on specific phases of the feature extraction.

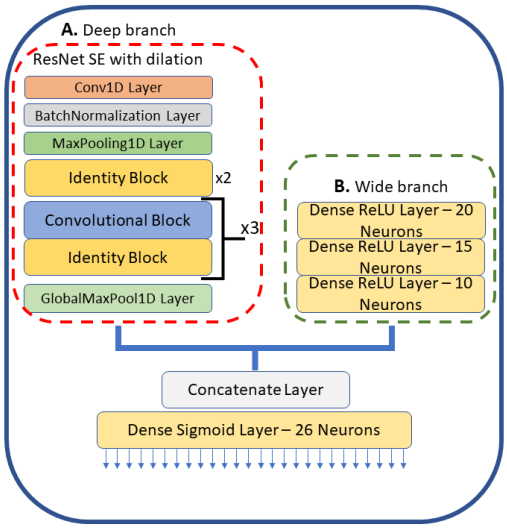

Figure 4.3: Network architecture: block A on the left is the Deep branch, which is the ResNet SEB; block B on the right is the Dense Layer that takes the numerical features and creates an embedding of them. Their outputs are concatenated to perform the final prediction.

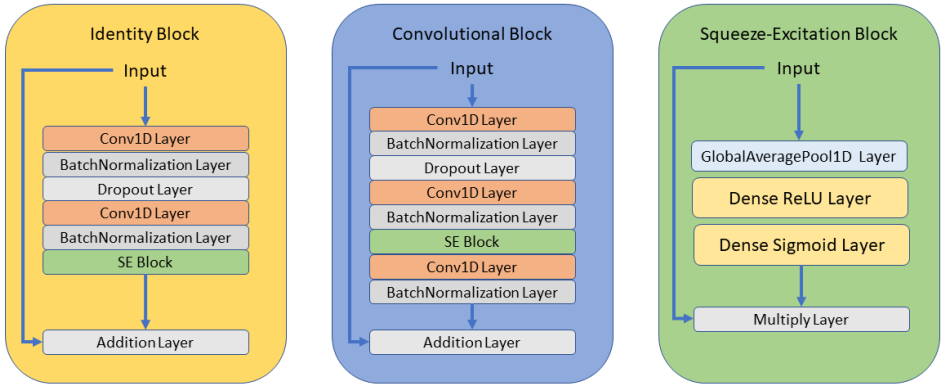

Figure 4.4: The composition of the Residual block and SEB. In all the Blocks the input is reinserted in the pipeline thanks to an Addition or Multiply Layer to improve the performances of the model and better catch the features of the signal.

Firstly, the deep branch was trained alone without the wide branch integration for 30 epochs. Secondly, the deep branch parameters were frozen and only the wide branch was trained for 20 epochs. Finally the deep branch was fine tuned while freezing the wide parameters for 10 epoch. Every training step was performed with early stopping which monitored the validation loss. The learning rate was initially set to 0.003 with a tenfold

decay of every 10 epochs. Binary crossentropy was used as loss function, optimized with Adam optimizer. Batch size was set to 64.

In order to maximize the Challenge Metric (CM) after the model training, a threshold optimization for each sigmoid unit output probabilities was performed. The optimization algorithm followed two steps:

1. A gridsearch of the best threshold vector maximizing the CM score, ranging from 0 to 0.4 with a 0.01 step.
2. The application of Nelder-Mead downhill simplex minimization method over the negative value of the CM score, initialized with the output of step 1.

Model Evaluation

Signals were preprocessed as described in section 4.2.3. As the model was trained for a 5000-sample input, test signals with a lower number of samples were zero padded. Those with a higher number of samples were divided into segments of 5000 samples with a 256 sample overlap and each segment was introduced separately into the model. The output probabilities of each segment were averaged to obtain a final signal classification.

Models were trained following a 5-fold crossvalidation. Thresholds were optimized using the validation set at each k-fold and then performances were assessed on a local test set obtained holding out a stratified set of 12.5% of the training data (9646 recordings). Furthermore, with the aim of evaluating the model's performance dependence on the local holdout test set, a second train-test split was performed to compare performance of models evaluated in different test sets.

4.3 Results

4.3.1 Machine learning

The results obtained after the 10-fold cross-validation of the provided datasets for phase 1 and 2 of the challenge are gathered in Tables 4.3 and 4.4, respectively. Sensitivity and specificity values are displayed along with the number of signals present in the dataset for each class targeted. In addition, the

sub-indexes C1, C2, ..., C7 in the Rhythms column indicate the classifier in which the rhythm was included. In both phases specificity (Sp.) values were higher than sensitivity (Se.) ones. Results in phase 1 were superior than those in phase 2. Table 4.5 gathers the overall results obtained for the classifier ensembles of phases 1 and 2 after a 10-fold cross-validation on the databases provided. The Challenge Metric (CM) used to evaluate the classifiers varied from phase 1 to phase 2. While in phase 1 it corresponded to the geometrical mean between F measure and G measure, on phase 2 a new scoring system was used to reflect the value of the algorithm in a clinical setting: it awarded full credit to correct diagnoses and partial credit to misdiagnoses with similar risks or outcomes similar to those of the actual diagnosis.

Table 4.3: Classification results on the test set after 10-fold cross-validation on the CPSC dataset labelled according to the first phase of the challenge.

Rhythms	Subjects	Se.(%)	Sp.(%)
AF C1	1221	94.6±0.02	95.9±0.01
IAVB C1	722	88.1±0.05	97.1±0.01
LBBB C2	236	85.6±0.08	98.5±0.00
NSR	918	78.8±0.04	93.7±0.01
PAC C3	616	84.5±0.06	90.3±0.01
PVC C4	700	87.8±0.02	93.5±0.01
RBBB C2	1857	89.5±0.02	96.5±0.01
STD C2	869	69.6±0.05	96.8±0.01
STE C2	220	37.2±0.09	99.3±0.00

Sensitivity (Se.), Specificity (Sp.)

4.3.2 Deep learning

Table 4.6 gathers the CM scores obtained in 5-fold crossvalidation training and evaluated in the local test set for the different training steps. As the final model was an ensemble formed by three models, results are presented as the mean and standard deviation of each of the models integrating the ensemble after the 5-fold crossvalidation. In addition, CM scores are presented for each of the different phases of the training: only the deep model

Table 4.4: Classification results on the test set after 10-fold cross-validation on all the datasets labelled according to the second phase of the challenge.

Rhythms	Subjects	Se.(%)	Sp.(%)
AF C1	3475	94.55±1.19	93.97±0.56
AFL C1	314	16.33±6.77	99.70±0.10
IAVB C3	2394	92.32±1.82	81.08±1.76
IRBBB C4	1611	5.78±2.43	99.77±0.14
LAD C4	6086	56.52±5.13	86.84±0.83
LAnFB C4	1806	56.29±14.67	97.60±0.58
LBBB C4	1041	31.38±3.64	99.71±0.09
LQT C5	1513	8.91±2.98	99.08±0.54
PAC C6	1944	65.88±4.20	91.61±0.84
PVC C7	1253	50.22±7.17	85.36±1.45
RBBB C4	3085	89.53±2.39	95.70±0.52
SB C2	2359	43.20±2.69	98.95±0.20
NSR C2	20846	90.48±0.61	70.38±1.55
STach C2	2402	86.78±2.91	99.25±0.20
TAb C5	4673	56.57±7.34	78.75±4.07

Sensitivity (Se.), Specificity (Sp.)

Table 4.5: Classification results on phase 1 and 2 for 10-fold cross-validation of the provided datasets.

Ph.	AUROC	AUPRC	Accuracy	F	Fbeta	Gbeta	CM
1st	0.099±0.009	0.019±0.002	0.944±0.004	0.727±0.019	0.763±0.019	0.525±0.025	0.644±0.031
2nd	0.541±0.002	0.083±0.001	0.148±0.009	0.116±0.004	0.144±0.004	0.062±0.002	0.030±0.009

Ph=phase; AUROC=area under the receiver operating curve;
AUPRC=area under the precision-recall curve.

(D), the deep model integrating the wide branch trained (D+W) and the model integrating both trained branched with a retrain of the deep branch (D+W+D). The local test set results present the CM for the ensemble model integrating the three individual models through major voting for the different training phases.

Table 4.6: Challenge scores for the ensemble of models: metric obtained after 5-fold crossvalidation training and local test of the ensemble on the public training set for each of the models forming the ensemble and their mean. D and W represent the training procedure: D is the model made of only the Deep branch; D+W represent the model composed by both branches, with Wide branch trained after the Deep branch; D+W+D shows results of the model with Wide and Deep branches where the Deep branch has been retrained after the Wide one.

Leads	Model	Training			Local Test		
		D	D+W	D+W+D	D	D+W	D+W+D
12	M1	0.708 ± 0.006	0.704 ± 0.005	0.706 ± 0.006	0.705	0.708	0.709
12	M2	0.704 ± 0.005	0.699 ± 0.007	0.704 ± 0.004			
12	M3	0.702 ± 0.003	0.700 ± 0.003	0.702 ± 0.003			
12	Mean	0.704 ± 0.005	0.701 ± 0.005	0.704 ± 0.004			

Table 4.6 shows an increase in performance of the ensemble model with respect to that of the individual models, as the mean values obtained in the validation set were lower than those obtained in the test set using the ensemble. During crossvalidation, D and D+W+D models showed a similar CM, while D+W attained a slightly lower metric. However, on the model testing the D+W+D obtained the highest CM. In contrast with the validation results both training phases integrating the wide branch performed better on the local test than the deep model alone, pointing at the higher generalization capabilities provided by the integration of a wide branch.

Comparison among models evaluated using a different holdout test set are presented in Table 4.7. Table 4.7 shows the CM obtained for the 2-lead models for D, D+W and D+W+D training phases and evaluated in test set 1 (the same as that presented for the 12 leads in Table 4.6) and a second test set 2. Slight differences can be appreciated between the CM of both test sets. However while in the test 1 the best training strategy is D+W, in test 2 is D+W+D. As in Table 4.6 CM on the test set was always higher for those models integrating the wide branch in their training.

Moreover, 12 and 2-leads model PPV per class are shown in Table 4.8, comparing the three different models D, D+W and D+W+D. In the appendix also sensitivity and f1-score values have been presented In Table 4.8 it can be seen that for the 12-lead model the D, D+W and D+W+D was superior for 6, 6 and 13 classes, respectively. In contrast for the 2-lead model the Deep, D+W and D+W+D was superior for 12, 6 and 8 classes

Table 4.7: Challenge scores for the ensemble models of 2-leads evaluated on two different local test sets for D, D+W and D+W+D.

Leads	Test set	Local Test		
		D	D+W	D+W+D
2	1	0.677	0.680	0.677
2	2	0.667	0.675	0.679

respectively. For wide branch integration the best training strategy for both leads combination was D+W+D.

4.4 Discussion

4.4.1 Machine learning

The aim of this work was to propose an automatic algorithm capable of identifying different cardiovascular diseases using 6 different databases with 43,101 labeled recordings made available by the PhysioNet/Computing in Cardiology Challenge 2020 [130]. Several attempts have been already described in literature [121, 133–135]. Currently, the role of clinicians is still fundamental for the final diagnosis, but a support role from computers could provide a useful tool to aid them for early and correct diagnosis of cardiac abnormalities. The presented method intended to follow a physiologically consistent approach. An ensemble of classifiers was built focusing each of them on specific cardiac regions. The electrophysiological disorders of study were distributed into each of the classifiers based on their cardiac region of incidence. ECG signals were divided into regions of interest and a comparison was performed between intra-patient models and inter-patient rhythm-specific models. In phase 1 the methodology succeeded in detecting all disorders with specificity higher than 90% and sensitivity higher than 84%, except for STE and STD. These results are in line with those already present in literature [15], although as different databases are used, a comparison is not trivial. These results suggest that intra-patient and inter-patient models manage to capture electrophysiological disturbances of different nature and areas of the cardiac tissue. However, results obtained in phase 2 show that the metrics used are not enough if a larger number of conditions with often similar expression in the ECG are targeted, as IRBBB and RBBB, IAVB

Table 4.8: PPV computed for each class of the 12 and 2 leads model with the training steps: Deep branch only, D+W branches and D+W+D.

Labels	12 leads			2 leads			N. Rec.
	D	D+W	D+W+D	D	D+W	D+W+D	
AF	0.365	0.365	0.369	0.362	0.365	0.368	554
AFL	0.673	0.648	0.661	0.652	0.637	0.638	1040
BBB	0.167	0.238	0.206	0.201	0.166	0.168	62
CLBBB | $LBBB$	0.561	0.410	0.469	0.536	0.431	0.438	163
CRBBB | $RBBB$	0.708	0.619	0.605	0.737	0.667	0.666	453
IAVB	0.445	0.483	0.455	0.475	0.515	0.431	374
IRBBB	0.378	0.422	0.408	0.206	0.204	0.291	220
LAD	0.529	0.529	0.539	0.460	0.507	0.499	954
LAnFB	0.427	0.461	0.449	0.377	0.395	0.395	273
LPR	0.161	0.165	0.170	0.185	0.141	0.169	49
LQRSV	0.162	0.186	0.235	0.151	0.171	0.191	199
LQT	0.306	0.315	0.287	0.264	0.25	0.265	237
NSIVCB	0.248	0.270	0.261	0.239	0.262	0.248	220
NSR	0.915	0.897	0.894	0.904	0.892	0.902	3537
PAC | $SVPB$	0.466	0.471	0.482	0.489	0.451	0.481	337
PR	0.791	0.755	0.792	0.739	0.597	0.681	185
PRWP	0.143	0.133	0.152	0.163	0.156	0.112	80
PVC | VPB	0.456	0.508	0.519	0.422	0.454	0.487	217
QAb	0.294	0.321	0.325	0.353	0.288	0.268	259
RAD	0.402	0.392	0.416	0.417	0.382	0.377	159
SA	0.525	0.544	0.594	0.540	0.501	0.494	472
SB	0.934	0.934	0.931	0.948	0.937	0.929	2358
STach	0.897	0.880	0.889	0.853	0.859	0.851	1167
TAb	0.404	0.423	0.428	0.384	0.390	0.390	1461
TInv	0.279	0.278	0.285	0.260	0.255	0.266	497

and LPR, LAnFB and LBBB, among others. In each classifier (C1, C2, . . . , C7) at least a cardiac abnormality was detected with good results i.e. AF in C1, STach in C2, IAVB in C3, RBBB in C4. However, the models built and/or the metric obtained seemed to be unable to distinguish among such an amount of rhythms exhibiting similar morphologies. In order to do so, more complex models as those built using deep learning techniques might be necessary.

4.4.2 Deep learning

The aim of this work was to develop a model that could combine classic machine learning features with deep learning and address the reported poor capability of deep models to effectively capture ECG signal temporal patterns [136]. The presented model encompassed two branches: a deep ResNet SE neural network and a wide 3-layered handcrafted feature classifier. The two branches were concatenated to obtain an embedding of the morphological and temporal features and to provide the recording's probability of belonging to each class. With this approach it was possible to combine implicit complex morphological information obtained by the ResNet model together with the explicit rhythmic data of the RR features. The input to the wide branch were 20 handcrafted features (Table 4.2) fed to a FC network with a final 10-neuron layer acting as bottleneck. This way a reduced feature space input was created for the concatenation with the deep branch, integrating the abstracted useful information from the handcrafted features. This contrasts with the approach used in [124] where the learning of the wide branch was performed only through 16 handcrafted HRV features directly concatenated to the deep branch. The chosen deep branch architecture was a modified ResNet model [132] integrating an SE Block, which had already proven in the past Challenge edition its efficacy for 12-leads ECGs classification due to its skip connections and the interchannel dependencies provided by the SE Block. A fixed window of only 5000 samples for signal classification was set as input, forcing to clip or zero-padd input signals. Although this was a limitation of the presented model, the choice was determined by the computational constrains due to the high dataset dimensions. As most signals contained around 5000 samples, it was set as a common framework to all recordings in order to take advantage of parallel processing power of GPU, reducing the training time.

Data augmentation and elimination from the three datasets was performed to deal with the non-homogeneities present in the database (length, sample frequency and labelling strategy). The database class imbalance shown in Table 4.1 was addressed with the optimization of the threshold used for classification for each class and with the creation of three data subsets with a more balanced class distribution. Each subset was used to train a different model, yielding three models that were integrated into an ensemble. Such ensemble showed a better generalization capacity that individual models, as displayed in Table 4.6, as it integrated the learning from all three

data subsets. A three-phase learning (as described in section 4.2.3) routine was implemented in order to mitigate overfitting and avoid giving excessive importance to the wide branch. After the deep model training without wide implementation, the wide model was introduced and trained. After the wide model training and feature embedding into a reduced space the branch was frozen to enable a final fine tuning of the deep branch considering also the handcrafted features.

From the comparison of the results obtained on the validation and test set gathered in Table 4.6 it could be observed that the integration of a wide branch increased the generalization capacities of the model. This could be hypothesized as test set results of models integrating a wide branch (D+W and D+W+D) were superior to those of just the deep model (D), while in crossvalidation the deep model without wide integration performed equally or better than those integrating the wide branch. Furthermore, test set results were superior in most cases for D+W+D showing that this last step may reduce overfitting and help integrate both types of information, yielding a more robust model. If we were to combine results obtained for individual classes as presented in Table 4.8 and those presented in Table 4.6, we could say that the best model structure would be a model ensemble integrating a D+W branch and trained following a three-step procedure: D+W+D. If Table 4.8 is analyzed, most classes are benefited from the retrain of the network after the wide branch is integrated.

In this study the integration of both types of information was emphasized by exploring both, the structural concatenation of both branches and their training. Little can be found in literature regarding deep+wide structural integration and training. Since both branches present a high difference among the number of trainable parameters, training both branches simultaneously with the same number of epochs and learning rate could lead to rapid overfitting of the wide branch and a very heterogeneous distribution of the importance given to each branch in the final model. To the best of our knowledge this is the first study exploring different training steps to integrate deep and wide branch for ECG classification. Integrating a wide branch into a deep network could provide us with the possibility of forcing a network to emphasize on determined features or characteristics of a dataset. In addition, given the lower variance of classic machine learning features with respect to deep ones, it could induce higher generalization capabilities in deep networks.

Comparison of the presented model with algorithms previous to the Challenge is not simple. The databases made available by the Computing in Cardiology PhysioNet 2021 Challenge have created a change in paradigm of artificial intelligence algorithms for the detection of ECG disorders. Previously, data restrictions did not enable the development of complex models targeting as many disorders as those targeted by the challenge, and those developed limited their activity to AF detection or beat classification [137, 138]. Those algorithms were typically based just in either private databases or in a few public databases [139]. They did not contain the sufficient amount and heterogeneity of annotated signals to target such a wide range of pathologies. The Challenge databases set a common frame for the development and comparison of state of the art artificial intelligence models. The objective of this work was to, not only present a valid solution for automatic ECG annotation, but to explore the behaviour of the integration of deep and wide features for such classification. It should be noted that no official Challenge score on the hidden test set was provided. Therefore, even if the final database used contained a highly heterogeneous set of ECG signals with different pathologies acquired from different centres, the interpretation of the final model's generalizability should be taken with care. Finally, although some compensation strategies were adopted to counteract the data imbalance effect (data subset division, ensemble creation and threshold optimization) as future work other approaches like the use of generative adversarial networks for data augmentation could be explored.

4.5 Conclusion

In this chapter two approaches for the detection of variate electrophysiological pathologies in the ECG are presented. Promising results for the detection of a limited number of abnormalities in short signals were obtained using the machine learning model. However, poor performance was achieved when applied to a higher number of disorders. A more complex model based on deep learning techniques was needed to target them. In the deep learning approach of this work, a model integrating deep and wide features was presented. An analysis was carried out to understand the benefits provided by the integration of explicit rhythm features into a deep model and the strategy that should be adopted for its training. Results pointed at a model's increased generalization power with the implementation of a wide branch. In addition, the influence of the different training steps on the

final outcome were studied. The best performing model was an ensemble of three models for each lead integrating a deep and wide branch and with a three-step training procedure.

Chapter 5

Conclusion and Final Remarks

5.1 Conclusions

This thesis explored different techniques to provide the scientific community with the computational tools for the automatic annotation of ECG signals. Given that ECG automatic annotation can have different scopes and limitations, different approaches were investigated.

5.1.1 Arrhythmia Detection on Single-Lead Constant Monitoring Devices

In a first study, a technique for detection of different tachyarrhythmia based on RR intervals was developed using a different approach with respect to the models present in literature. In contrast with state-of-the-art RR detectors specifically built for AF detection, the presented methodology was able to capture tachyarrhythmia-specific RR patterns and by so, it was applicable to detect episodes other than AF.

The methodology was based on the representation of RR intervals on a discretized version of the Poincaré Plot: the Poincaré Image. Poincaré Atlases were also introduced as the representation of the mean characteristic Poincaré Image of an specific tachyarrhythmia. Different types of Poincaré Images were explored, based on the RR intervals, δRR intervals, and their

combination by juxtaposition in $RR\delta RR$ Poincaré Images. The different bin sizes for the creation of Poincaré Images were explored: 10 ms, 20 ms and 40 ms. In addition, different classification metrics were compared: NMI, 2D correlation and their combination. The best hyperparameter combination was an $RR\delta RR$ Poincaré Image with a bin size of [40x40] ms classified using a combination of NMI and 2D correlation. The optimal combination was evaluated in ECG signal segments of 20, 30, 60 and 120 seconds. Performance was superior in longer signal segments, although it was not highly degraded as signal length was reduced. For a signal segment of 60 s and bin size of [40x40] ms a sensitivity of 94.35±3.68% and 88.86±12.78% and specificity of 85.51±7.46% and 96.10±2.25 was obtained for AF and AB episodes, respectively. Given the lower amount of AB episodes present in the study, a higher standard deviation was obtained. This pointed at the necessity of a higher amount of Poincaré Images to create a Poincaré Atlas representative of AB. Moreover, in order to assess the translability of the method into other types of signals, it was applied to synthetic PPG signals for AF detection. Results suggested that the methodology could be applied also to these kind of signals. However, due to data constrictions, only AF detection could be evaluated. Even if in this study only NSR, AF and AB were targeted, we theorise that it could be generalizable to other rhythms, which could enable the monitoring of other possibly relevant cardiac tachyarrhythmia.

Preliminary results considering only a type of Poincaré Image, bin size and time window as well as a simpler classification metric, were presented in [105]. In [108] instead, the concepts of Poincaré Image and Poincaré Atlas were extensively assessed and the parameters influencing their computation were optimized for obtaining better classification results. The improvements introduced in [108] with respect to [105] include: the exploration of the bin size influence on Poincaré Images and Altases, the study of different types of Poincaré Images configuration, the optimization of the distance metric between Poincaré Images and Atlases and finally the assessment of the methodology's performance for reduced time windows that range from 120 s to 20 s (in [105] only a time window of 120 s was used). The major contribution of [108] was thus, the definition of Poincaré Images and Atlases and the demonstration of their potential to detect and classify different cardiac rhythms with segments as short as 20 s.

5.1.2 Beat Classification for PAC Occurrence Monitoring in Ambulatory 2-lead ECG Holter Recordings

The study of PAC implications in first-time AF occurrence and stroke are hindered by the lack of a proper PAC automatic annotator. Those present in literature have a low PAC sensitivity and thus are unsuitable for any PAC occurrence study. In this thesis, two detectors were developed: one focused only on RR information to classify short ECG segments and another one containing also morphological data to classify individual beats.

Firstly, a beat classifier targeting N, S and V beats was developed. A confrontation was made between a multi-class classifier (including the three classes mentioned) and a binary classifier distinguishing only between S and $Other$ beats. Sensitivity values were slightly higher for multi-label approach but with a reduction in PPV in comparison to the binary classifier. In contrast with state-of-the-art classifiers, the developed method attained a higher S sensitivity preserving the PPV. Differently that other methods in literature, the present model does not depend on ECG delineation methods which may be sensitive to noise interference. Further efforts are still needed to ameliorate the proposed solution and reduce the amount of false positives to a level which allows to apply such model in a clinical setting.

For that aim, a segment-wise detector was developed. The detector exploited the RR representation method developed in the previous chapter by transforming the RR sequence into Poincaré Images. A CNN was trained to distinguish 30-beat segments containing one or more PAC or PVC. The network managed to detect Poincaré Images containing A/V beats with a patient-wise sensitivity and specificity of 96.03 (89.67 - 98.76)% and 98.70 (91.50 - 99.75)%, respectively. These results suggest that the proposed classifier could be used to detect ECG regions containing S and V beats with high precision, reducing the number of beats to be classified. If a manual annotation is sought, a segment-wise classification could drastically reduce the workload from the trained professional. On the other hand, if an automatic beat classification is wanted, selecting only those regions that could potentially contain A/V could aid in the reduction of false positives.

The multi-class and binary classifiers were initially published in [140] as a full paper. The segment-wise classifier was subsequently developed and published in [141] as a conference proceedings.

5.1.3 AI for Cardiac Disorders Automatic Annotation in the 12-Lead ECG

Automatic classification of a wide variety of electrophysiological disorders on short 12-lead ECG signals could be used to annotate large retrospective databases and carry out different large-scale analysis. In this work two approaches were taken to develop such a model.

Firstly a machine learning ensemble was developed. Each of the sub-models targeted a specific cardiac region and the pathologies that could affect it. Results were promising when 9 pathologies were included. However, when a higher and more complex amount of cardiac disorders were targeted, the machine learning system failed to provide accurate predictions. To solve this, a deep learning model based on a modified ResNet and combined with classic machine learning was developed. The reasoning behind the combination of deep and wide features was that deep features extracted with the modified ResNet model contain complex implicit specific morphological information that could be completed by more generalizable and simpler rhythm features. In addition, by introducing rhythm features relevant rhythm indexes could be forced to be integrated into the classification decision process. However, training this hybrid model is not trivial, as the two branches contain a very unbalanced number of parameters. A three-step training approach was explored in this work to balance the effect of both branches on the final classification. The final best performing model was an ensemble of three modified ResNets for each lead integrating both wide and deep branches and trained in a three-step fashion.

The higher explainability of machine learning makes it more suitable for its application on a medical context. Therefore, they ought to be a previous choice to deep learning. However, deep learning its proving its potential in undergoing complex tasks. In this case, its superiority with respect to a machine learning ensemble in detecting a wide range of cardiac disorders was made evident. Still work needs to be done to provide a better model explainability so as to make deep networks gain a higher trust on the medical field.

The machine learning classifier ensemble was originally developed as a contribution to the Computers in Cardiology (CinC) PhysioNet Challenge 2020, and published in the proceedings of the conference [142]. The first deep learning approach was instead presented to the CinC PhysioNet

Challenge 2021 [122] in a first attempt of combining a ResNet with machine learning features. Finally, the hybrid model training and combination strategy was further explored in a journal publication in [143].

5.2 Final Remarks

Automatic ECG annotation plays a fundamental role in cardiac health understanding and monitoring. There exists a wide and diverse range of devices for the acquisition of cardiac eletrophysiological activity and an enormous amount of information is being gathered every day in both a clinical and an ambulatory setting. Manual revision of all these data is unrealistic, and to extract and take advantage of this information, automatic annotation ought to be applied.

Classic atrial arrhythmia detection included mainly AF identification and was based on the computation of determined parameters and the setting of thresholds and relationships to detect the arrhythmia. With the development of machine learning and deep learning models, more complex algorithms have been introduced. Nevertheless, most of these algorithms still mainly focus on AF monitoring rather than on other types of atrial tachyarrhythmia, which are being proved to be more relevant than suspected in the past. The introduction of techniques able to detect other types of atrial tachyarrhthymia in different types of signals may enable to fulfill a two-fold scope: on one hand they could be used to exploit the information from extensive databases to perform clinical studies and on the other hand they could be used to monitor the occurrence of relevant events on ambulatory recordings to assess a patient's cardiac activity.

This thesis presents a contribution for the automatic annotation of ECG signals for different scopes. By presenting an RR interval-based classifier able to target tachyarrhythmia other than AF, a method for monitoring cardiac activity on low quality signals suitable for long-term constant monitoring devices is presented. On the other hand, the beat classifier able to detect PACs presented was developed to enable off-line automatic classification of ambulatory medium-term ECG recordings. This would allow the study of PAC implications as well as the detection of these events on patients' ECG registrations. Finally, a more complex deep learning model was developed using short ECG signals acquired in a clinical setting. This model could help alleviate clinicians' workload on the interpretation of ECG

recordings. More importantly, it could be used to analyze large databases whose manual and homogeneous annotation are unfeasible to carry out and by so, enable the development of future clinical studies.

Appendix

Chapter 3: Beat Classification for PAC Detection in Ambulatory 2-lead ECG Holter Recordings

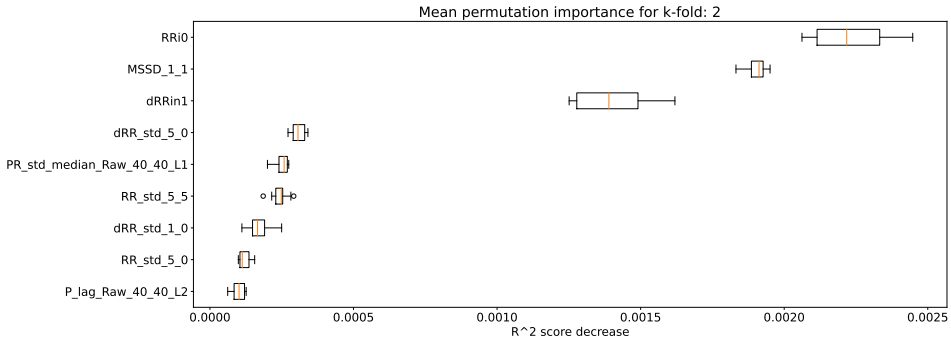

Figure 5.1: Mean permutation importance calculated using the test set for the random forest classifier model of k-fold 2. Only the 10 most relevant features according to their decrease in the model's score are presented.

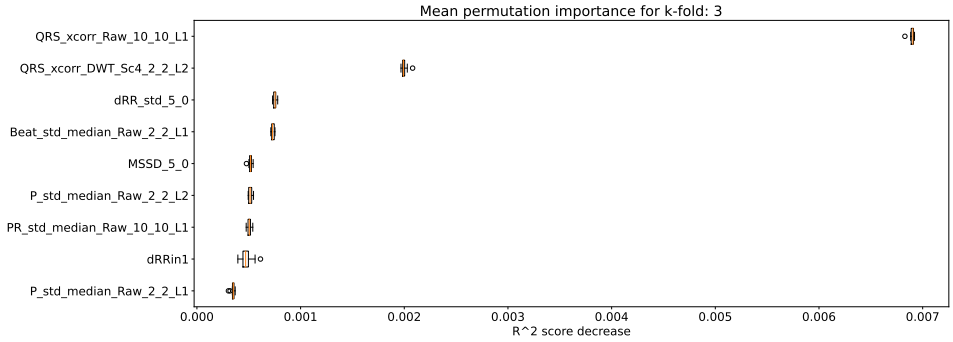

Figure 5.2: Mean permutation importance calculated using the test set for the random forest classifier model of k-fold 3. Only the 10 most relevant features according to their decrease in the model's score are presented.

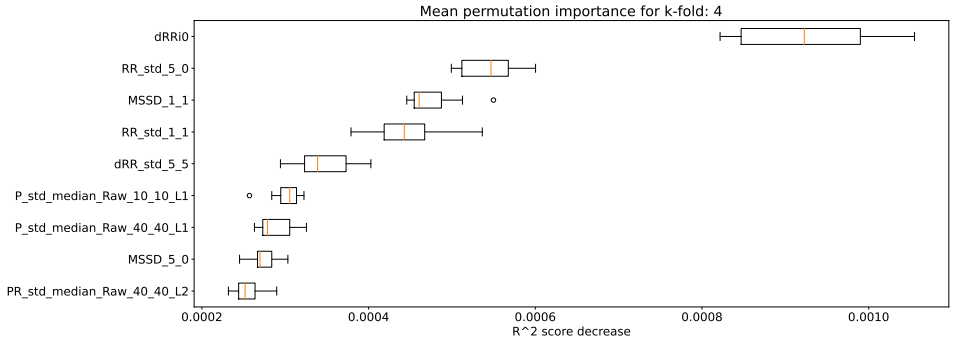

Figure 5.3: Mean permutation importance calculated using the test set for the random forest classifier model of k-fold 4. Only the 10 most relevant features according to their decrease in the model's score are presented.

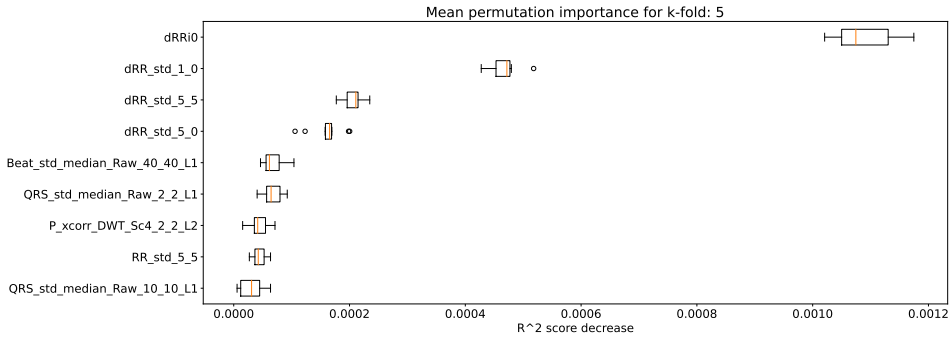

Figure 5.4: Mean permutation importance calculated using the test set for the random forest classifier model of k-fold 5. Only the 10 most relevant features according to their decrease in the model's score are presented.

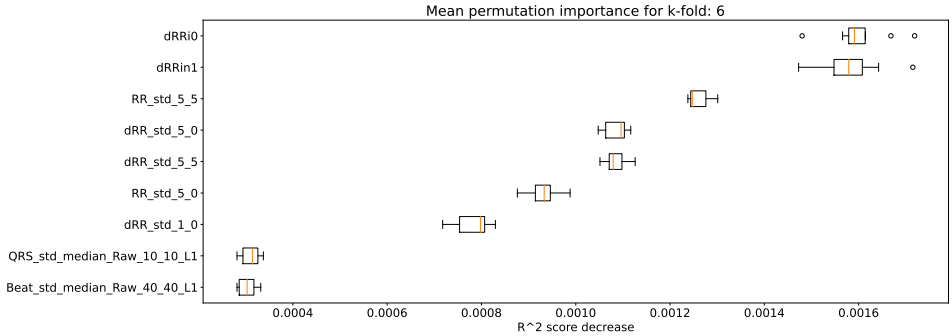

Figure 5.5: Mean permutation importance calculated using the test set for the random forest classifier model of k-fold 6. Only the 10 most relevant features according to their decrease in the model's score are presented.

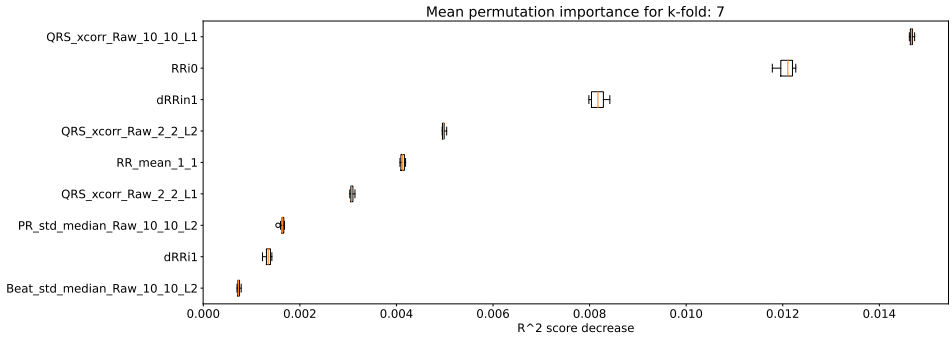

Figure 5.6: Mean permutation importance calculated using the test set for the random forest classifier model of k-fold 7. Only the 10 most relevant features according to their decrease in the model's score are presented.

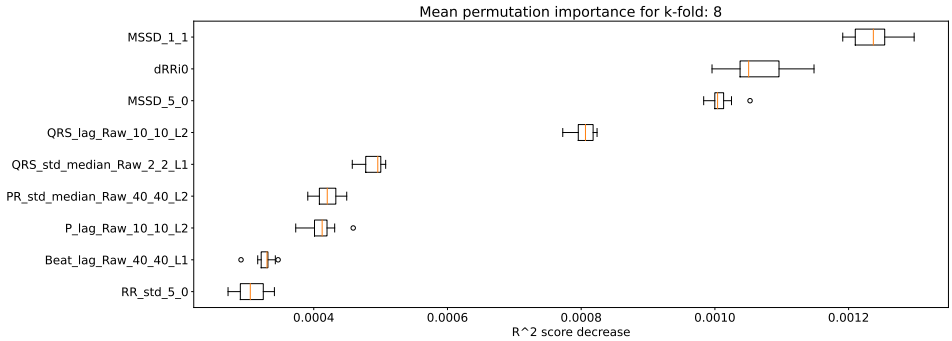

Figure 5.7: Mean permutation importance calculated using the test set for the random forest classifier model of k-fold 8. Only the 10 most relevant features according to their decrease in the model's score are presented.

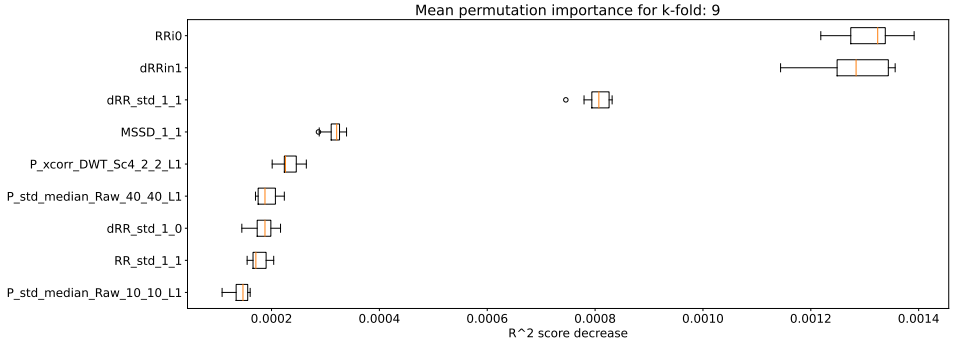

Figure 5.8: Mean permutation importance calculated using the test set for the random forest classifier model of k-fold 9. Only the 10 most relevant features according to their decrease in the model's score are presented.

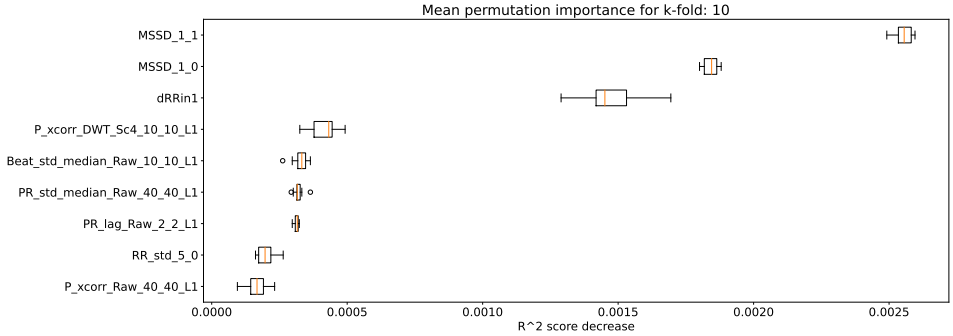

Figure 5.9: Mean permutation importance calculated using the test set for the random forest classifier model of k-fold 10. Only the 10 most relevant features according to their decrease in the model's score are presented.

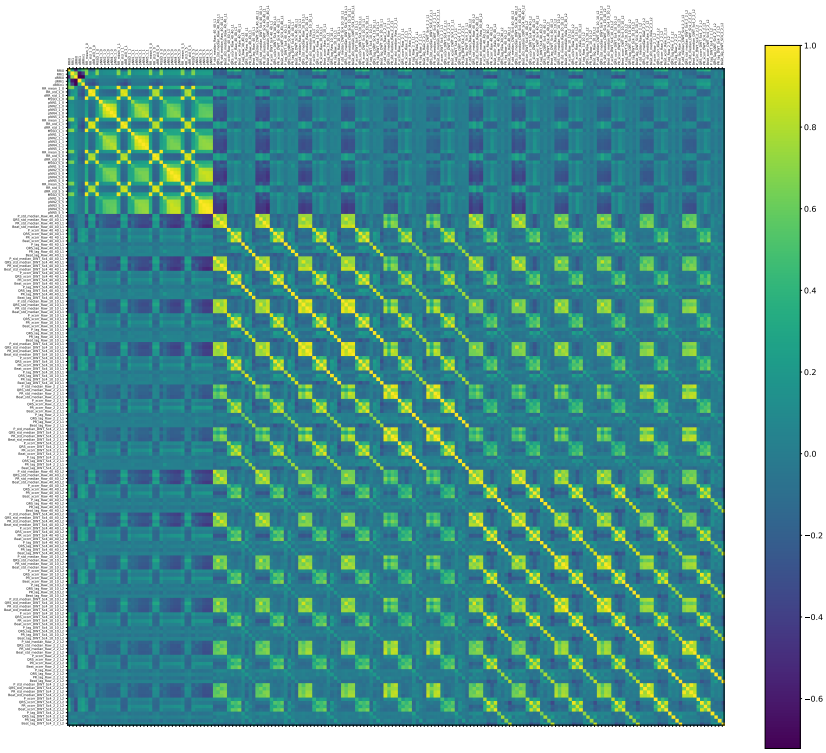

Figure 5.10: Correlation matrix of the training set of k-fold 1 for all the features computed.

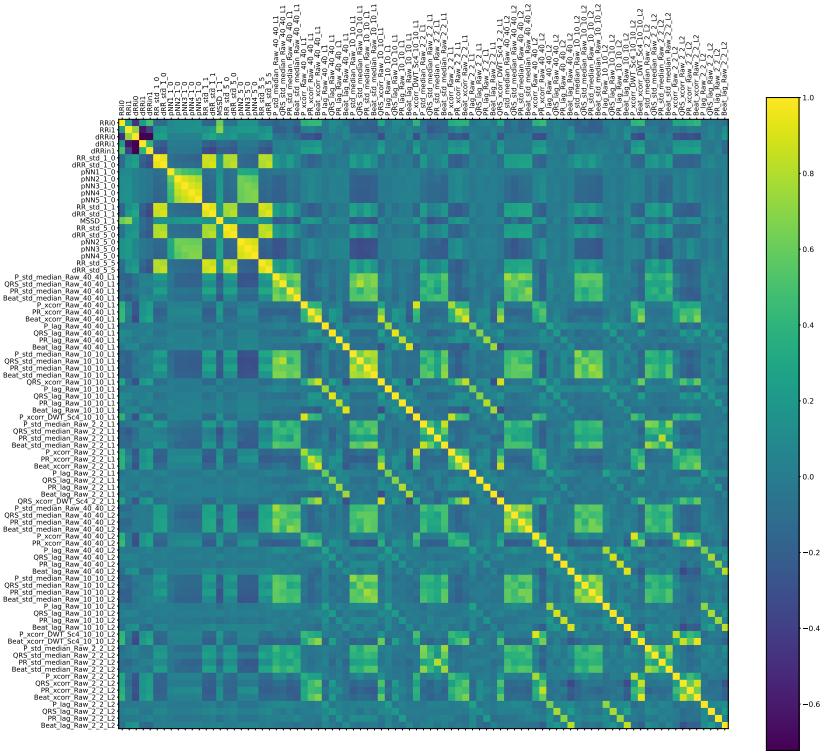

Figure 5.11: Correlation matrix of the features selected after variance and correlation filter on k-fold 1 (elimination of features with >0.90 correlation values) application.

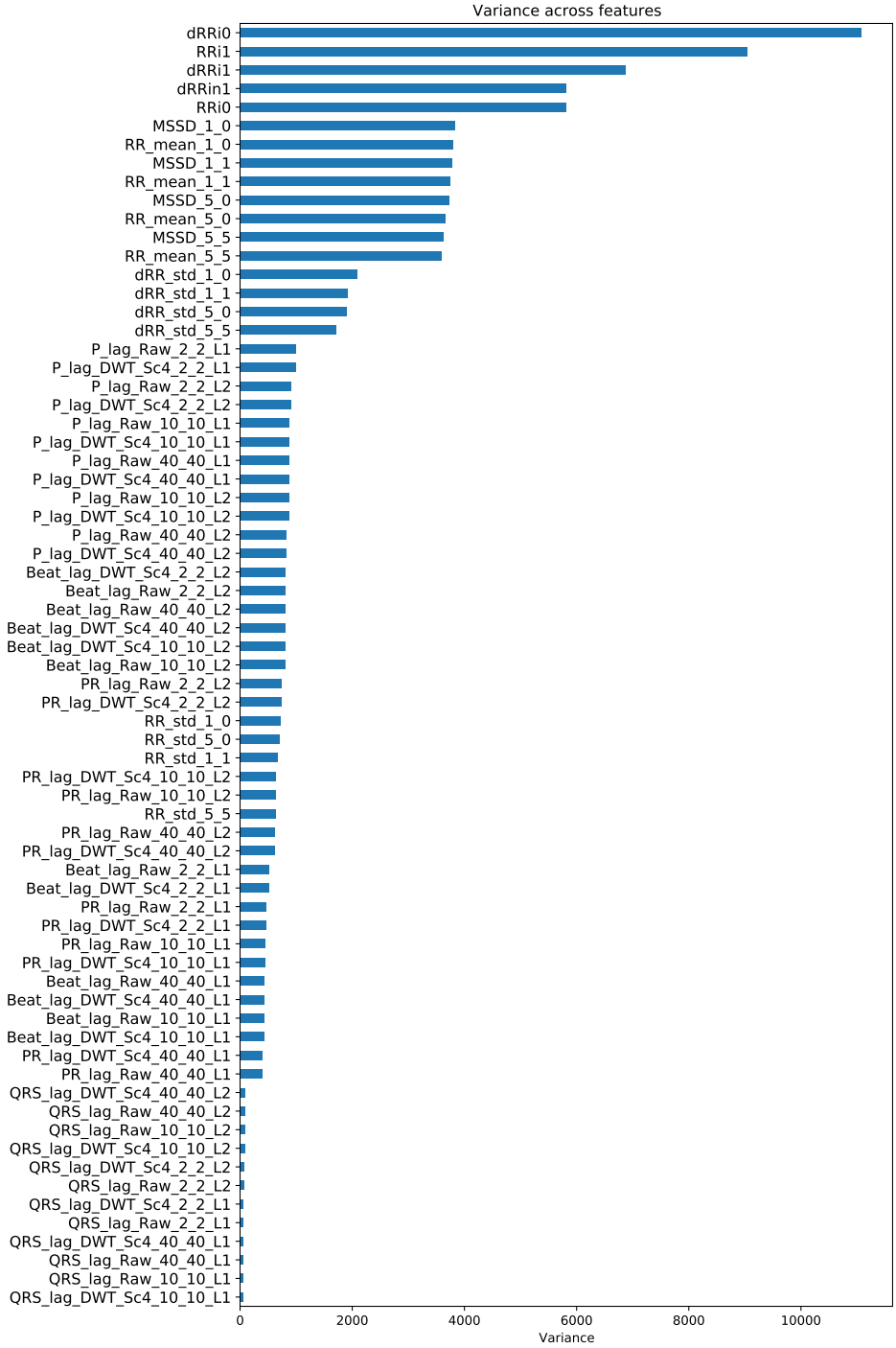

Figure 5.12: Variance value for each feature. Features have been sorted by decreasing value of their variance. For a matter of space only those features with variance > 0.5 are displayed.

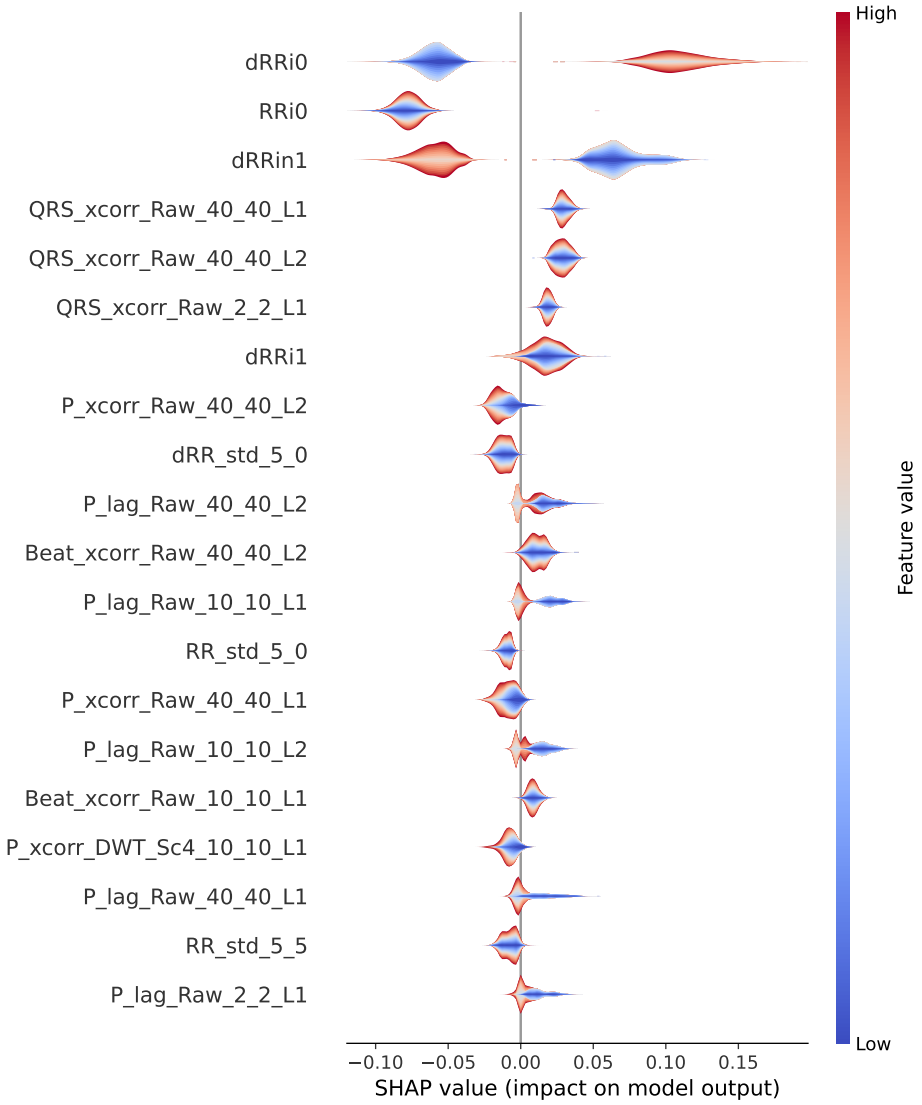

Figure 5.13: SHAP values for S predictions for k-fold 2 model.

Chapter 4: AI for Cardiac Disorders Automatic Annotation in the 12-Lead ECG

Table 5.1: F1-score computed for each class of the 12 and 2 leads model with the training steps: Deep branch only, D+W branches and D+W+D.

Labels	12 leads		
	D	D+W	D+W+D
AF	0.527	0.525	0.529
AFL	0.796	0.780	0.789
BBB	0.245	0.299	0.283
LBBB	0.684	0.569	0.619
RBBB	0.805	0.752	0.739
IAVB	0.598	0.629	0.606
IRBBB	0.484	0.512	0.507
LAD	0.676	0.675	0.683
LAnFB	0.586	0.616	0.605
LPR	0.249	0.245	0.250
LQRSV	0.200	0.202	0.176
LQT	0.414	0.415	0.397
NSIVCB	0.334	0.347	0.344
NSR	0.926	0.925	0.924
PAC	0.611	0.611	0.618
PR	0.834	0.806	0.827
PRWP	0.232	0.215	0.229
VPB	0.598	0.637	0.644
QAb	0.282	0.272	0.263
RAD	0.519	0.519	0.538
SA	0.669	0.682	0.713
SB	0.959	0.958	0.957
STach	0.941	0.933	0.938
TAb	0.553	0.564	0.567
TInv	0.406	0.404	0.411

Table 5.2: Sensitivity computed for each class of the 12 and 2 leads model with the training steps: Deep branch only, D+W branches and D+W+D.

Labels	12 leads		
	D	D+W	D+W+D
AF	0.902	0.903	0.905
AFL	0.946	0.940	0.943
BBB	0.982	0.988	0.985
LBBB	0.986	0.976	0.981
RBBB	0.979	0.971	0.969
IAVB	0.952	0.959	0.954
IRBBB	0.967	0.972	0.970
LAD	0.911	0.911	0.914
LAnFB	0.963	0.967	0.966
LPR	0.983	0.985	0.986
LQRSV	0.957	0.964	0.973
LQT	0.956	0.958	0.952
NSIVCB	0.954	0.958	0.956
NSR	0.945	0.943	0.942
PAC	0.961	0.961	0.963
PR	0.993	0.992	0.993
PRWP	0.967	0.967	0.974
VPB	0.974	0.978	0.979
QAb	0.963	0.966	0.967
RAD	0.978	0.977	0.978
SA	0.955	0.958	0.965
SB	0.979	0.979	0.979
STach	0.985	0.983	0.984
TAb	0.786	0.803	0.806
TInv	0.888	0.888	0.892

References

[1] Giuseppe Lippi, Fabian Sanchis-Gomar, and Gianfranco Cervellin. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. *International Journal of Stroke*, 16(2):217–221, 2021.

[2] Massimo Zoni-Berisso, Fabrizio Lercari, Tiziana Carazza, and Stefano Domenicucci. Epidemiology of atrial fbrillation: European perspective. *Clinical Epidemiology*, 6(1):213–220, 2014.

[3] Enrico Vizzardi, Antonio Curnis, Maria G. Latini, Francesca Salghetti, Elena Rocco, Laura Lupi, Riccardo Rovetta, Filippo Quinzani, Ivano Bonadei, Luca Bontempi, Antonio D'Aloia, and Livio Dei Cas. Risk factors for atrial fibrillation recurrence: A literature review. *Journal of Cardiovascular Medicine*, 15(3):235–253, 2014.

[4] Luciano A. Sposato, Lauren E. Cipriano, Patricia M. Riccio, Vladimir Hachinski, and Gustavo Saposnik. Very short paroxysms account for more than half of the cases of atrial fibrillation detected after stroke and TIA: A systematic review and meta-analysis. *International Journal of Stroke*, 10(6):801–807, 2015.

[5] David J. Gladstone, Paul Dorian, Melanie Spring, Val Panzov, Muhammad Mamdani, Jeff S. Healey, Kevin E. Thorpe, R. Aviv, K. Boyle, J. Blakely, R. Cote, J. Hall, M. K. Kapral, N. Kozlowski, A. Laupacis, M. O'Donnell, K. Sabihuddin, M. Sharma, A. Shuaib, H. Vaid, A. Pinter, S. Abootalebi, R. Chan, S. Crann, L. Fleming, C. Frank, V. Hachinski, K. Hesser, B. S. Kumar, P. Soros, M. Wright, V. Basile, K. Boyle, J. Hopyan, Y. Rajmohan, R. Swartz, H. Vaid, G. Valencia, J. Ween, H. Aram, P. A. Barber, S. Coutts, A. M. Demchuk, K. Fischer, M. D. Hill, G. Klein, C. Kenney, B. Menon,

M. McClelland, A. Russell, K. Ryckborst, P. Stys, E. E. Smith, T. W. Watson, S. Chacko, D. Sahlas, J. Sancan, R. Côté, L. Durcan, E. Ehrensperger, J. Minuk, T. Wein, L. Wadup, N. Asdaghi, J. Beckman, N. Esplana, P. Masigan, C. Murphy, E. Tang, P. Teal, K. Villaluna, A. Woolfenden, S. Yip, M. Bussière, D. Dowlatshahi, M. Sharma, G. Stotts, S. Robert, K. Ford, D. Hackam, L. Miners, T. Mabb, J. D. Spence, B. Buck, T. Griffin-Stead, R. Jassal, M. Siddiqui, A. Hache, C. Lessard, F. Lebel, A. Mackey, S. Verreault, C. Astorga, L. K. Casaubon, M. Del Campo, C. Jaigobin, L. Kalman, F. L. Silver, L. Atkins, K. Coles, A. Penn, R. Sargent, C. Walter, Y. Gable, N. Kadribasic, B. Schwindt, A. Shuaib, P. Kostyrko, D. Selchen, G. Saposnik, P. Christie, A. Jin, D. Hicklin, D. Howse, E. Edwards, S. Jaspers, F. Sher, S. Stoger, D. Crisp, A. Dhanani, V. John, M. Levitan, M. Mehdiratta, and D. Wong. Atrial Premature Beats Predict Atrial Fibrillation in Cryptogenic Stroke: Results from the EMBRACE Trial. *Stroke*, 46(4):936–941, 2015.

[6] Tran Thong, James McNames, Mateo Aboy, and Brahm Goldstein. Prediction of Paroxysmal Atrial Fibrillation by Analysis of Atrial Premature Complexes. *IEEE Transactions on Biomedical Engineering*, 51(4):561–569, 2004.

[7] Zeynep Binici, Theodoros Intzilakis, Olav Wendelboe Nielsen, Lars Køber, and Ahmad Sajadieh. Excessive supraventricular ectopic activity and increased risk of atrial fibrillation and stroke. *Circulation*, 121(17):1904–1911, 2010.

[8] Boon Hor Chong, Vincent Pong, Kwok Fai Lam, Shasha Liu, Ming Liang Zuo, Yuk Fai Lau, Chu Pak Lau, Hung Fat Tse, and Chung Wah Siu. Frequent premature atrial complexes predict new occurrence of atrial fibrillation and adverse cardiovascular events. *Europace*, 14(7):942–947, 2012.

[9] Shinya Suzuki, Koichi Sagara, Takayuki Otsuka, Hiroto Kano, Shunsuke Matsuno, Hideaki Takai, Tokuhisa Uejima, Yuji Oikawa, Akira Koike, Kazuyuki Nagashima, Hajime Kirigaya, Junji Yajima, Hiroaki Tanabe, Hitoshi Sawada, Tadanori Aizawa, and Takeshi Yamashita. Usefulness of frequent supraventricular extrasystoles and a high CHADS 2 score to predict first-time appearance of atrial fibrillation. *American Journal of Cardiology*, 111(11):1602–1607, 2013.

[10] Bao Tao Huang, Fang Yang Huang, Yong Peng, Yan Biao Liao, Fei Chen, Tian Li Xia, Xiao Bo Pu, and Mao Chen. Relation of premature atrial complexes with stroke and death: Systematic review and meta-analysis. *Clinical Cardiology*, 40(11):962–969, 2017.

[11] Boon-hor Chong, Vincent Pong, Kwok-fai Lam, Shasha Liu, Mingliang Zuo, Yuk-fai Lau, Chu-pak Lau, Hung-fat Tse, and Chung-wah Siu. Frequent premature atrial complexes predict new occurence of atrial fibrillation and adverse cardiovascular events. *Europace*, 14:942–947, 2011.

[12] Bjørn Strøier Larsen, Preman Kumarathurai, Julie Falkenberg, Olav W. Nielsen, and Ahmad Sajadieh. Excessive Atrial Ectopy and Short Atrial Runs Increase the Risk of Stroke beyond Incident Atrial Fibrillation. *Journal of the American College of Cardiology*, 66(3):232–241, 2015.

[13] Christina F. Pacchia, Nazem W. Akoum, Stephen Wasmund, and Mohamed H. Hamdan. Atrial bigeminy results in decreased left ventricular function: An insight into the mechanism of PVC-induced cardiomyopathy. *PACE - Pacing and Clinical Electrophysiology*, 35(10):1232–1235, 2012.

[14] Derick M. Todd, Simon P. Fynn, Andrew P. Walden, W. Julian Hobbs, Sanjay Arya, and Clifford J. Garratt. Repetitive 4-Week Periods of Atrial Electrical Remodeling Promote Stability of Atrial Fibrillation: Time Course of a Second Factor Involved in the Self-Perpetuation of Atrial Fibrillation. *Circulation*, 109(11):1434–1439, 2004.

[15] Ary L. Goldberger and Luis A. N. Amaral. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, 101(23):215–220, 2012.

[16] Amit Shah Chad Robichaux An-Kwok Ian Wong Chengyu Liu Feifei Liu Ali Bahrami Rad Andoni Elola Salman Seyedi Qiao Li Ashish Sharma Gari D. Clifford Matthew A. Reyna Erick A. Perez Alday, Annie Gu. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. *Physiol. Meas.*

[17] MA Reyna, N Sadr, EA Perez Alday Computing in ..., and Undefined 2021. Will two do? varying dimensions in electrocardiography: The physionet/computing in cardiology challenge 2021. *Computing in Cardiology*, 48:1–4, 2021.

[18] Feifei Liu, Chengyu Liu, Lina Zhao, Xiangyu Zhang, Xiaoling Wu, Xiaoyan Xu, Yulin Liu, Caiyun Ma, Shoushui Wei, Zhiqiang He, Jianqing Li, and Eddie Ng Yin Kwee. An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. *Journal of Medical Imaging and Health Informatics*, 8(7):1368–1373, 2018.

[19] Jianwei Zheng, Huimin Chu, Daniele Struppa, Jianming Zhang, Sir Magdi Yacoub, Hesham El-Askary, Anthony Chang, Louis Ehwerhemuepha, Islam Abudayyeh, Alexander Barrett, Guohua Fu, Hai Yao, Dongbo Li, Hangyuan Guo, and Cyril Rakovski. Optimal Multi-Stage Arrhythmia Classification Approach. *Scientific Reports*, 10(1):1–17, 2020.

[20] Patrick Wagner, Nils Strodthoff, Ralf Dieter Bousseljot, Dieter Kreiseler, Fatima I. Lunze, Wojciech Samek, and Tobias Schaeffter. PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, 7(1):1–15, 2020.

[21] R. Bousseljot, D. Kreiseler, and A. Schnabel. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik*, 40(s1):317–318, 1995.

[22] Jianwei Zheng, Jianming Zhang, Sidy Danioko, Hai Yao, Hangyuan Guo, and Cyril Rakovski. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. *Scientific Data*, 7(1):1–8, 2020.

[23] K. Tateno and L. Glass. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals. *Medical and Biological Engineering and Computing*, 39(6):664–671, 2001.

[24] S. Dash, K. H. Chon, S. Lu, and E. A. Raeder. Automatic real time detection of atrial fibrillation. *Annals of Biomedical Engineering*, 37(9):1701–1709, 2009.

[25] Chao Huang, Shuming Ye, Hang Chen, Dingli Li, Fangtian He, and Yuewen Tu. A novel method for detection of the transition between atrial fibrillation and sinus rhythm. *IEEE Transactions on Biomedical Engineering*, 58(4):1113–1119, 2011.

[26] P. Langley, M. Dewhurst, L. Y. Di Marco, P. Adams, F. Dewhurst, J. C. Mwita, R. Walker, and A. Murray. Accuracy of algorithms for detection of atrial fibrillation from short duration beat interval recordings. *Medical Engineering and Physics*, 34(10):1441–1447, 2012.

[27] D. E. Lake and J. R. Moorman. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *AJP: Heart and Circulatory Physiology*, 300(1):H319–H325, 2011.

[28] Jinseok Lee, Yunyoung Nam, David D. McManus, and Ki H. Chon. Time-varying coherence function for atrial fibrillation detection. *IEEE Transactions on Biomedical Engineering*, 60(10):2783–2793, 2013.

[29] Xiaolin Zhou, Hongxia Ding, Benjamin Ung, Emma Pickwell-MacPherson, and Yuanting Zhang. Automatic online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy. *BioMedical Engineering Online*, 13(1), 2014.

[30] Andrius Petrenas, Vaidotas Marozas, and Leif Sörnmo. Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Computers in Biology and Medicine*, 65:184–191, 2015.

[31] Jinseok Lee, Bersain A. Reyes, David D. McManus, Oscar Mathias, and Ki H. Chon. Atrial fibrillation detection using an iphone 4S. *IEEE Transactions on Biomedical Engineering*, 60(1):203–206, 2013.

[32] Martin Stridh, Andreas Bollmann, S. Olsson, and Leif S?rnmo. Detection and feature extraction of atrial tachyarrhythmias. *IEEE Engineering in Medicine and Biology Magazine*, 25(6):31–39, 2006.

[33] Jinseok Lee, David D. McManus, Peter Bourrell, Leif Sörnmo, and Ki H. Chon. Atrial flutter and atrial tachycardia detection using Bayesian approach with high resolution time-frequency spectrum from ECG recordings. *Biomedical Signal Processing and Control*, 8(6):992–999, 2013.

[34] Arthur C. Guyton and John E. Hall. *Textbook of Medical Physiology*. Saunders Elservier, 2006.

[35] Lars Kaestner, Qinghai Tian, and Peter Lipp. Action Potentials in Heart Cells. (May):163–182, 2011.

[36] Leif Sörnmo and Pablo Laguna. *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. 2005.

[37] Lara Ortiz-Martin, Pablo Picazo-Sanchez, Pedro Peris-Lopez, and Juan Tapiador. Heartbeats do not make good pseudo-random number generators: An analysis of the randomness of inter-pulse intervals. *Entropy*, 20(2):1–19, 2018.

[38] JAMES L. WELLS, ROBERT B. KARP, NICHOLAS T. KOUCHOUKOS, WILLIAM A.H. MACLEAN, THOMAS N. JAMES, and ALBERT L. WALDO. Characterization of Atrial Fibrillation in Man: Studies Following Open Heart Surgery. *Pacing and Clinical Electrophysiology*, 1(4):426–438, 1978.

[39] Rania Gaspo, Ralph F. Bosch, Elias Bou-Abboud, and Stanley Nattel. Tachycardia-induced changes in Na+ current in a chronic dog model of atrial fibrillation. *Circulation Research*, 81(6):1045–1052, 1997.

[40] Ralph F. Bosch, Xiaorong Zeng, Joachim B. Grammer, Katarina Popovic, Christian Mewis, and Volker Kühlkamp. Ionic mechanisms of electrical remodeling in human atrial fibrillation. *Cardiovascular Research*, 44(1):121–131, 1999.

[41] David R. Van Wagoner, Amber L. Pond, Michelle Lamorgese, Sandra S. Rossie, Patrick M. McCarthy, and Jeanne M. Nerbonne. Atrial L-type Ca2+ currents and human atrial fibrillation. *Circulation Research*, 85(5):428–436, 1999.

[42] Maurits Allessie, Jannie Ausma, and Ulrich Schotten. Electrical , contractile and structural remodeling during atrial fibrillation. 54:230–246, 2002.

[43] M. C.E.F. Wijffels, C. J.H.J. Kirchhof, L. V.A. Boersma, R. Dorland, and M. A. Allessie. Atrial fibrillation begets atrial fibrillation. *New Trends in Arrhythmias*, 9(2):147–152, 1993.

[44] Gerrit D. Dispersyn, Jannie Ausma, Fred Thoné, Willem Flameng, Jean Louis J. Vanoverschelde, Maurits A. Allessie, Frans C.S. Ramaekers, and Marcel Borgers. Cardiomyocyte remodelling during myocardial hibernation and atrial fibrillation: Prelude to apoptosis. *Cardiovascular Research*, 43(4):947–957, 1999.

[45] Jayson R. Baman, Daniel T. Mathew, Michael Jiang, and Rod S. Passman. Mobile Health for Arrhythmia Diagnosis and Management. *Journal of General Internal Medicine*, 37(1):188–197, 2022.

[46] US Food and Drug Administration. Mobile medical applications.

[47] Przemyslaw Guzik and Marek Malik. ECG by mobile technologies. *Journal of Electrocardiology*, 49(6):894–901, 2016.

[48] E. Jonathan and Martin Leahy. Investigating a smartphone imaging unit for photoplethysmography. *Physiological Measurement*, 31(11), 2010.

[49] Nicole Lowres, Lis Neubeck, Glenn Salkeld, Ines Krass, Andrew J. McLachlan, Julie Redfern, Alexandra A. Bennett, Tom Briffa, Adrian Bauman, Carlos Martinez, Christopher Wallenhorst, Jerrett K. Lau, David B. Brieger, Raymond W. Sy, and S. Ben Freedman. Feasibility and cost-effectiveness of stroke prevention through community screening for atrial fibrillation using iPhone ECG in pharmacies: The SEARCH-AF study. *Thrombosis and Haemostasis*, 111(6):1167–1176, 2014.

[50] Nino Isakadze and Seth S. Martin. How useful is the smartwatch ECG? *Trends in Cardiovascular Medicine*, 30(7):442–448, 2020.

[51] S. Suave Lobodzinski. ECG patch monitors for assessment of cardiac rhythm abnormalities. *Progress in Cardiovascular Diseases*, 56(2):224–229, 9 2013.

[52] Paddy M. Barrett, Ravi Komatireddy, Sharon Haaser, Sarah Topol, Judith Sheard, Jackie Encinas, Angela J. Fought, and Eric J. Topol. Comparison of 24-hour Holter monitoring with 14-day novel adhesive patch electrocardiographic monitoring. *American Journal of Medicine*, 127(1):11–95, 2014.

[53] Axel Müller, Wilfried Scharner, Tilo Borchardt, Wolfgang Och, and Harald Korb. Reliability of an external loop recorder for automatic recognition and transtelephonic ECG transmission of atrial fibrillation. *Journal of Telemedicine and Telecare*, 15(8):391–396, 2009.

[54] Shantanu Sarkar, David Ritscher, and Rahul Mehra. Communications A Detector for a Chronic Implantable Atrial. *IEEE Transactions on Biomedical Engineering*, 55(3):1219–1224, 2008.

[55] Charlotte Eitel, Daniela Husser, Gerhard Hindricks, Manuela Frühauf, Sebastian Hilbert, Arash Arya, Thomas Gaspar, Ulrike Wetzel, Andreas Bollmann, and Christopher Piorkowski. Performance of an implantable automatic atrial fibrillation detection device: Impact of software adjustments and relevance of manual episode analysis. *Europace*, 13(4):480–485, 2011.

[56] Leif Sörnmo and Pablo Laguna. Chapter 6: The electrocardiogram - A Brief Background. In *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, pages 411–452. 2005.

[57] Xiaolin Zhou, Hongxia Ding, Wanqing Wu, and Yuanting Zhang. A real-time Atrial fibrillation detection algorithm based on the instantaneous state of heart rate. *PLoS ONE*, 10(9), 2015.

[58] Jie Lian, Lian Wang, and Dirk Muessig. A simple method to detect atrial fibrillation using RR intervals. *American Journal of Cardiology*, 107(10):1494–1497, 2011.

[59] Jinho Park, Sangwook Lee, and Moongu Jeon. Atrial fibrillation detection by heart rate variability in Poincare plot. *BioMedical Engineering Online*, 8:1–12, 2009.

[60] Shantanu Sarkar, David Ritscher, and Rahul Mehra. A detector for a chronic implantable atrial tachyarrhythmia monitor. *IEEE Transactions on Biomedical Engineering*, 55(3):1219–1224, 2008.

[61] Hyun Jun Cho, Cheol Hyun Lee, Jongmin Hwang, Hyoung-Seob Park, Sang-Woong Choi, In-Cheol Kim, Yun-Kyeong Cho, Hyuck-Jun Yoon, Hyungseop Kim, Chang-Wook Nam, Seung-Ho Hur, Byung Chun Jung, Yoon Nyun Kim, and Seongwook Han. Accuracy of implantable loop recorders for detecting atrial tachyarrhythmias after atrial fibrillation catheter ablation. *International Journal of Arrhythmia*, 21(1), 2020.

[62] Sami Pakarinen, Mika Lehto, Jaap Ruiter, and Willem G. de Voogt. Enhanced detection of atrial tachyarrhythmias with pacing devices by using more accurate atrial sensing. *Journal of Interventional Cardiac Electrophysiology*, (0123456789), 2021.

[63] Rod S. Passman, Kenneth M. Weinberg, Mark Freher, Pablo Denes, Andi Schaechter, Jeffrey J. Goldberger, and Alan H. Kadish. Accuracy

of mode switch algorithms for detection of atrial tachyarrhythmias. *Journal of Cardiovascular Electrophysiology*, 15(7):773–777, 2004.

[64] Hirofumi Tomita, Shingo Sasaki, Joji Hagii, and Norifumi Metoki. Covert atrial fibrillation and atrial high-rate episodes as a potential cause of embolic strokes of undetermined source: Their detection and possible management strategy. *Journal of Cardiology*, 72(1):1–9, 2018.

[65] Mariano Llamedo and Juan Pablo Martinez. Driven by Database Generalization Criteria. *IEEE Transactions on Biomedical Engineering*, 58(3):616–625, 2011.

[66] Mariano Llamedo and Juan Pablo Martinez. An automatic patient-adapted ECG heartbeat classifier allowing expert assistance. *IEEE Transactions on Biomedical Engineering*, 59(8):2312–2320, 2012.

[67] Philip De Chazal, Maria O'Dwyer, and Richard B. Reilly. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 51(7):1196–1206, 2004.

[68] Eduardo José Da S. Luz, Thiago M. Nunes, Victor Hugo C. De Albuquerque, João P. Papa, and David Menotti. ECG arrhythmia classification based on optimum-path forest. *Expert Systems with Applications*, 40(9):3561–3573, 2013.

[69] Zhancheng Zhang, Jun Dong, Xiaoqing Luo, Kup Sze Choi, and Xiaojun Wu. Heartbeat classification using disease-specific feature selection. *Computers in Biology and Medicine*, 46(1):79–89, 2014.

[70] Tanis Mar, Sebastian Zaunseder, Juan Pablo Martínez, Mariano Llamedo, and Rdiger Poll. Optimization of ECG classification by means of feature selection. *IEEE Transactions on Biomedical Engineering*, 58(8):2168–2177, 2011.

[71] Sung Nien Yu and Ying Hsiang Chen. Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. *Pattern Recognition Letters*, 28(10):1142–1150, 2007.

[72] Sung Nien Yu and Kuan To Chou. Integration of independent component analysis and neural networks for ECG beat classification. *Expert Systems with Applications*, 34(4):2841–2846, 2008.

[73] Michael A. Lee, Richard Weachter, Scott Pollak, Mark S. Kremers, Ajay M. Naik, Russell Silverman, Joann Tuzi, Wayne Wang, Linda J. Johnson, and David E. Euler. The effect of atrial pacing therapies on atrial tachyarrhythmia burden and frequency: Results of a randomized trial in patients with bradycardia and atrial tachyarrhythmias. *Journal of the American College of Cardiology*, 41(11):1926–1932, 2003.

[74] Saeed Babaeizadeh, Richard E. Gregg, Eric D. Helfenbein, James M. Lindauer, and Sophia H. Zhou. Improvements in atrial fibrillation detection for real-time monitoring. *Journal of Electrocardiology*, 42(6):522–526, 2009.

[75] Steven Ladavich and Behnaz Ghoraani. Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity. *Biomedical Signal Processing and Control*, 18:274–281, 2015.

[76] Yong Xia, Naren Wulan, Kuanquan Wang, and Henggui Zhang. Detecting atrial fibrillation by deep convolutional neural networks. *Computers in Biology and Medicine*, 93(July 2017):84–92, 2018.

[77] Juan Ródenas, Manuel García, Raúl Alcaraz, and José J. Rieta. Wavelet entropy automatically detects episodes of atrial fibrillation from single-lead electrocardiograms. *Entropy*, 17(9):6179–6199, 2015.

[78] Petr Nejedly, Adam Ivora, Radovan Smisek, Ivo Viscor, Zuzana Koscova, Pavel Jurak, and Filip Plesinger. Classification of ECG Using Ensemble of Residual CNNs with Attention Mechanism. *Computing in Cardiology*, 2021-Septe:1–4, 2021.

[79] Hyeongrok Han, Seongjae Park, Seonwoo Min, Hyun Soo Choi, Eunji Kim, Hyunki Kim, Sangha Park, Jinkook Kim, Junsang Park, Junho An, Kwanglo Lee, Wonsun Jeong, Sangil Chon, Kwonwoo Ha, Myungkyu Han, and Sungroh Yoon. Towards High Generalization Performance on Electrocardiogram Classification. *Computing in Cardiology*, 2021-Septe:4–7, 2021.

[80] Nima L. Wickramasinghe and Mohamed Athif. Multi-Label Cardiac Abnormality Classification from Electrocardiogram Using Deep Convolutional Neural Networks. *Computing in Cardiology*, 2021-Septe:1–4, 2021.

[81] Taya V. Glotzer, Emile G. Daoud, D. George Wyse, Daniel E. Singer, Michael D. Ezekowitz, Christopher Hilker, Clayton Miller, Dongfeng Qi, and Paul D. Ziegler. The Relationship between daily atrial tachyarrhythmia burden from implantable device diagnostics and stroke risk the trends study. *Circulation: Arrhythmia and Electrophysiology*, 2(5):474–480, 2009.

[82] Craig T. January, L. Samuel Wann, Joseph S. Alpert, Hugh Calkins, Joaquin E. Cigarroa, Joseph C. Cleveland, Jamie B. Conti, Patrick T. Ellinor, Michael D. Ezekowitz, Michael E. Field, Katherine T. Murray, Ralph L. Sacco, William G. Stevenson, Patrick J. Tchou, Cynthia M. Tracy, and Clyde W. Yancy. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: A report of the American college of Cardiology/American heart association task force on practice guidelines and the heart rhythm society. *Journal of the American College of Cardiology*, 64(21):e1–e76, 2014.

[83] Leif Sörnmo, Andrius Petr, Vaidotas Marozas, and Public E C G Databases. Atrial Fibrillation from an Engineering Perspective. pages 49–71, 2018.

[84] Kirk H. Shelley. Photoplethysmography: Beyond the calculation of arterial oxygen saturation and heart rate. *Anesthesia and Analgesia*, 105(SUPPL. 6):31–36, 2007.

[85] Andrius Sološenko, Andrius Petrėnas, Vaidotas Marozas, and Leif Sörnmo. Modeling of the photoplethysmogram during atrial fibrillation. *Computers in Biology and Medicine*, 81(October 2016):130–138, 2017.

[86] John Allen. Photoplethysmography and its application in clinical physiological measurement, 3 2007.

[87] Axel Schäfer and Jan Vagedes. How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram, 6 2013.

[88] Andrius Petrėnas, Leif Sörnmo, Arūnas Lukoševičius, and Vaidotas Marozas. Detection of occult paroxysmal atrial fibrillation, 2015.

[89] John A. Cairns, Stuart Connolly, Sean McMurtry, Michael Stephenson, and Mario Talajic. Canadian Cardiovascular Society atrial fib-

rillation guidelines 2010: Prevention of stroke and systemic thromboembolism in atrial fibrillation and flutter. *Canadian Journal of Cardiology*, 27(1):74–90, 2011.

[90] Navinder S. Sawhney and Gregory K. Feld. Diagnosis and Management of Typical Atrial Flutter. *Medical Clinics of North America*, 92(1):65–85, 2008.

[91] P. Langley, M. Dewhurst, L. Y. Di Marco, P. Adams, F. Dewhurst, J. C. Mwita, R. Walker, and A. Murray. Accuracy of algorithms for detection of atrial fibrillation from short duration beat interval recordings. *Medical Engineering and Physics*, 34(10):1441–1447, 2012.

[92] Hans D. Esperer, Chris Esperer, and Richard J. Cohen. Cardiac arrhythmias imprint specific signatures on Lorenz plots. *Annals of Noninvasive Electrocardiology*, 13(1):44–60, 2008.

[93] Tsuyoshi Anan, Kenji Sunagawa, Haruo Araki, and Motoomi Nakamura. Arrhythmia analysis by successive RR plotting. *Journal of Electrocardiology*, 23(3):243–248, 1990.

[94] Chandan K. Karmakar, Ahsan H. Khandoker, Jayavardhana Gubbi, and Marimuthu Palaniswami. Complex correlation measure: a novel descriptor for Poincaré plot. *Biomedical engineering online*, 8:17, 2009.

[95] Lijuan Zhang, Tianci Guo, Bin Xi, Yang Fan, Kun Wang, Jiacheng Bi, and Ying Wang. Automatic recognition of cardiac arrhythmias based on the geometric patterns of poincaré plots. *Physiological Measurement*, 36(2):283–301, 2015.

[96] Michael Brennan, Marimuthu Palaniswami, and Peter Kamen. Distortion properties of the interval spectrum of IPFM generated heartbeats for heart rate variability analysis. *IEEE Transactions on Biomedical Engineering*, 48(11):1251–1264, 2001.

[97] Michael Brennan, Marimuthu Palaniswami, and Peter Kamen. Poincaré plot interpretation using a physiological model of HRV based on a network of oscillators. *American Journal of Physiology - Heart and Circulatory Physiology*, 283(5):H1873–H1886, 2002.

[98] Paul Viola and Wells III William M. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[99] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multi-modality image registration by maximization of mutual information. *Proceedings of the Workship on Mathematical Methods in Biomedical Image Analysis*, 16(2):14–22, 1996.

[100] Jpw Pluim. Mutual information based registration of medical images. *Ponsen & Looijen*, XX:1–21, 2000.

[101] C. Studholme, D. L.G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[102] Alexander Strehl and Joydeep Ghosh. Cluster ensembles - A knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3(3):583–617, 2003.

[103] Thomas R Dawber, H Emerson Thomas, and Patricia M Mcnamara&dagger. CHARACTERISTICS OF THE DICROTIC NOTCH OF THE ARTERIAL PULSE WAVE IN CORONARY HEART DISEASE. Technical report.

[104] Valentina D.A. Corino, Rita Laureanti, Lorenzo Ferranti, Giorgio Scarpini, Federico Lombardi, and Luca T. Mainardi. Detection of atrial fibrillation episodes using a wristband device. *Physiological Measurement*, 38(5):787–799, 2017.

[105] Guadalupe García-Isla, Valentina D.A. Corino, and Luca T. Mainardi. Cardiac Tachyarrhythmia Detection by Poincaré Plot-Based Image Analysis. In *Computing in Cardiology*, Singapore, 2019.

[106] P. Kirchhof et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *European Heart Journal*, 37(38), 2016.

[107] Gregory W. Botteron and Joseph M. Smith. Quantitative assessment of the spatial organization of atrial fibrillation in the intact human heart. *Circulation*, 93(3):513–518, 1996.

[108] Guadalupe Garcia-Isla, Valentina D.A. Corino, and Luca Mainardi. Poincaré Plot image and rhythm-specific atlas for atrial bigeminy and atrial fibrillation detection. *IEEE Journal of Biomedical and Health Informatics*, 25(4):1–1, 2020.

[109] Pak Hei Chan, Chun Ka Wong, Yukkee C. Poh, Louise Pun, Wangie Wan Chiu Leung, Yu Fai Wong, Michelle Man Ying Wong, Ming Zher Poh, Daniel Wai Sing Chu, and Chung Wah Siu. Diagnostic Performance of a Smartphone-Based Photoplethysmographic Application for Atrial Fibrillation Screening in a Primary Care Setting. *Journal of the American Heart Association*, 5(7):1–7, 2016.

[110] V. Krasteva, I. Jekova, and I. Christov. Automatic detection of premature atrial contractions in the electrocardiogram. *Clbme.Bas.Bg*, (April 2014):9–10, 2006.

[111] M. Visinescu, C. A. Bashour, O. Wazni, and B. Gopakumaran. Automatic detection of conducted premature atrial contractions to predict atrial fibrillation in patients after cardiac surgery. *Computers in Cardiology*, 31:429–432, 2004.

[112] Eduardo José da S. Luz, William Robson Schwartz, Guillermo Cámara-Chávez, and David Menotti. ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer Methods and Programs in Biomedicine*, 127:144–164, 2016.

[113] Federica Censi, G Calcagnini, C Ricci, R P Ricci, M Santini, A Grammatico, and P Bartolini. P-Wave Morphology Assessment by a Gaussian Functions-Based Model in Atrial Fibrillation Patients. *IEEE Transactions on Biomedical Engineering*, 54(4):663–672, 2007.

[114] Juan Pablo Martínez, Rute Almeida, Salvador Olmos, Ana Paula Rocha, and Pablo Laguna. A Wavelet-Based ECG Delineator Evaluation on Standard Databases. *IEEE Transactions on Biomedical Engineering*, 51(4):570–581, 2004.

[115] Andrés Julio Demski and Mariano Llamedo Soria. ecg-kit a Matlab Toolbox for Cardiovascular Signal Processing. *Journal of Open Research Software*, 4:2–5, 2016.

[116] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su In Lee. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67, 1 2020.

[117] Andrius Petrenas, Vaidotas Marozas, Andrius Sološenko, Raimondas Kubilius, Jurgita Skibarkiene, Julien Oster, and Leif Sörnmo. Electrocardiogram modeling during paroxysmal atrial fibrillation: Application to the detection of brief episodes. *Physiological Measurement*, 38(11):2058–2080, 2017.

[118] Peter M. Kistler, Kurt C. Roberts-Thomson, Haris M. Haqqani, Simon P. Fynn, Suresh Singarayar, Jitendra K. Vohra, Joseph B. Morton, Paul B. Sparks, and Jonathan M. Kalman. P-Wave Morphology in Focal Atrial Tachycardia. Development of an Algorithm to Predict the Anatomic Site of Origin. *Journal of the American College of Cardiology*, 48(5):1010–1017, 2006.

[119] Eduardo José da S. Luz, William Robson Schwartz, Guillermo Cámara-Chávez, and David Menotti. ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer Methods and Programs in Biomedicine*, 127, 2016.

[120] Massimo F. Piepoli, Arno W. Hoes, Stefan Agewall, Christian Albus, Carlos Brotons, Alberico L. Catapano, Marie Therese Cooney, Ugo Corrà, Bernard Cosyns, Christi Deaton, Ian Graham, Michael Stephen Hall, F. D.Richard Hobbs, Maja Lisa Løchen, Herbert Löllgen, Pedro Marques-Vidal, Joep Perk, Eva Prescott, Josep Redon, Dimitrios J. Richter, Naveed Sattar, Yvo Smulders, Monica Tiberi, H. Bart Van Der Worp, Ineke Van Dis, W. M.Monique Verschuren, Simone Binno, Guy De Backer, Marco Roffi, Victor Aboyans, Norbert Bachl, Scipione Carerj, Leslie Cho, John Cox, Johan De Sutter, Günther Egidi, Miles Fisher, Donna Fitzsimons, Oscar H. Franco, Maxime Guenoun, Catriona Jennings, Borut Jug, Paulus Kirchhof, Kornelia Kotseva, Gregory Y.H. Lip, François Mach, Giuseppe Mancia, Franz Martin Bermudo, Alessandro Mezzani, Alexander Niessner, Piotr Ponikowski, Bernhard Rauch, Adrienne Stauder, Guillaume Turc, Olov Wiklund, Stephan Windecker, Jose Luis Zamorano, Stephan Achenbach, Lina Badimon, Gonzalo Barón-Esquivias, Helmut Baumgartner, Jeroen J. Bax, Veronica Dean, Çetin Erol, Oliver Gaemperli, Philippe Kolh, Patrizio Lancellotti, Petros Nihoyannopoulos, Adam Torbicki, António Vaz Carneiro, Bernhard Metzler, Ruslan Najafov, Valeriy Stelmashok, Catherine De Maeyer, Mirza Dilić, Ivan Gruev, Davor Miličić, Helena Vaverkova, Ida Gustafsson, Ihab Attia, Davit Duishvili, Jean Ferrières, Nela Kostova, Zurab Klimi-

ashvili, Rainer Hambrecht, Konstantinos Tsioufis, Eszter Szabados, Karl Andersen, Carl Vaughan, Barak Zafrir, Salvatore Novo, Kairat Davletov, Fisnik Jashari, Alina Kerimkulova, Iveta Mintale, Georges Saade, Zaneta Petrulioniene, Charles Delagardelle, Caroline J. Magri, Victor Rudi, Latifa Oukerraj, B. Ersen Çölkesen, Henrik Schirmer, Roberto Palma Dos Reis, Daniel Gherasim, Sergey Nedogoda, Marco Zavatta, Vojislav Giga, Slavomira Filipova, Luis Rodríguez Padial, Anna Kiessling, Abdallah Mahdhaoui, Dilek Ural, Elena Nesukay, and Chris Gale. 2016 European Guidelines on cardiovascular disease prevention in clinical practice. *European Heart Journal*, 37(29):2315–2381, 2016.

[121] Harold Smulyan. The Computerized ECG: Friend and Foe. *American Journal of Medicine*, 132:153–160, 2019.

[122] Stefano Magni, Andrea Sansonetti, Chiara Salvi, Tiziana Tabiadon, Guadalupe Garc, and Politecnico Milano. Combining ResNet Model with Handcrafted Temporal Features for ECG Classification with Varying Number of Leads. 2021.

[123] Federico M. Muscato, Valentina D.A. Corino, and Luca T. Mainardi. Ensemble Learning of Modified Residual Networks for Classifying ECG with Different Set of Leads. *Computing in Cardiology*, 2021-Septe:3–6, 2021.

[124] Annamalai Natarajan, Yale Chang, Sara Mariani, Asif Rahman, Gregory Boverman, Shruti Vij, and Jonathan Rubin. A Wide and Deep Transformer Neural Network for 12-Lead ECG Classification. *Computing in Cardiology*, 2020-Septe:12–15, 2020.

[125] Jiapu Pan and Willis J. Tompkins. A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236, 2007.

[126] G. Goovaerts, C. Varon, B. Vandenberk, R. Willems, and S. Van Huffel. Tensor-based detection of T wave alternans in multilead ECG signals. In *Computing in Cardiology*, 2014.

[127] Adriana N. Vest, Giulia Da Poian, Qiao Li, Chengyu Liu, Shamim Nemati, Amit J. Shah, and Gari D. Clifford. An open source benchmarked toolbox for cardiovascular waveform and interval analysis. *Physiological Measurement*, 39(10), 2018.

[128] Jonathan W. Waks, Colleen M. Sitlani, Elsayed Z. Soliman, Muammar Kabir, Elyar Ghafoori, Mary L. Biggs, Charles A. Henrikson, Nona Sotoodehnia, Tor Biering-Sørensen, Sunil K. Agarwal, David S. Siscovick, Wendy S. Post, Scott D. Solomon, Alfred E. Buxton, Mark E. Josephson, and Larisa G. Tereshchenko. Global Electric Heterogeneity Risk Score for Prediction of Sudden Cardiac Death in the General Population: The Atherosclerosis Risk in Communities (ARIC) and Cardiovascular Health (CHS) Studies. *Circulation*, 133:2222–2234, 2016.

[129] Jason A. Thomas, Erick A. Perez-Alday, Allison Junell, Kelley Newton, Christopher Hamilton, Yin Li-Pershing, David German, Aron Bender, and Larisa G. Tereshchenko. Vectorcardiogram in athletes: The Sun Valley Ski Study. *Annals of Noninvasive Electrocardiology*, page e12614, 2018.

[130] Erick Andres Perez-Alday, Yin Li-Pershing, Aron Bender, Christopher Hamilton, Jason A. Thomas, Kyle Johnson, Tiffany L. Lee, Ryan Gonzales, Aaron Li, Kelley Newton, and Larisa G. Tereshchenko. Importance of the heart vector origin point definition for an ECG analysis: The Atherosclerosis Risk in Communities (ARIC) study. *Computers in Biology and Medicine*, 104(September 2018):127–138, 2019.

[131] Ivaylo I. Christov. Real time electrocardiogram QRS detection using combined adaptive threshold. *BioMedical Engineering Online*, 3:1–9, 2004.

[132] Zhibin Zhao, Hui Fang, Samuel D. Relton, Ruqiang Yan, Yuhong Liu, Zhijing Li, Jing Qin, and David C. Wong. Adaptive Lead Weighted ResNet Trained with Different Duration Signals for Classifying 12-lead ECGs. *Computing in Cardiology*, 2020-Septe:1–4, 2020.

[133] Miquel Alfaras, Miguel C. Soriano, and Silvia Ortín. A Fast Machine Learning Model for ECG-Based Heartbeat Classification and Arrhythmia Detection. *Frontiers in Physics*, 7(July):1–11, 2019.

[134] Jürg Schläpfer and Hein J. Wellens. Computer-Interpreted Electrocardiograms: Benefits and Limitations. *Journal of the American College of Cardiology*, 70(9):1183–1192, 2017.

[135] Xin Zhang, Kai Gu, Shumei Miao, Xiaoliang Zhang, Yuechuchu Yin, Cheng Wan, Yun Yu, Jie Hu, Zhongmin Wang, Tao Shan, Shenqi

Jing, Wenming Wang, Yun Ge, Yin Chen, Jianjun Guo, and Yun Liu. Automated detection of cardiovascular disease by electrocardiogram signal analysis: A deep learning system. *Cardiovascular Diagnosis and Therapy*, 10(2):227–235, 2020.

[136] Genshen Yan, Shen Liang, Yanchun Zhang, and Fan Liu. Fusing Transformer Model with Temporal Features for ECG Heartbeat Classification. *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, pages 898–905, 2019.

[137] Xinwen Liu, Huan Wang, Zongjin Li, and Lang Qin. Deep learning in ECG diagnosis: A review. *Knowledge-Based Systems*, 227:107187, 2021.

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

[139] G. B. Moody and R. G. Mark. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001.

[140] Guadalupe García-Isla, Luca Mainardi, and Valentina D.A. Corino. A Detector for Premature Atrial and Ventricular Complexes. *Frontiers in Physiology*, 12(June):1–11, 2021.

[141] Guadalupe García-Isla, Luca Mainardi, and Valentina D A Corino. A Poincare Image-Based Detector of ECG Segments Containing Atrial and Ventricular Beats. volume 48, pages 1–4. IEEE, 2021.

[142] Guadalupe Garcia-Isla, Rita Laureanti, Valentina D. Corino, and Luca T. Mainardi. ECG Morphological Decomposition for Automatic Rhythm Identification. In *Computing in Cardiology*, volume 2020-September. IEEE Computer Society, 9 2020.

[143] Guadalupe Garcia-Isla, Federico M. Muscato, Andrea Sansonetti, Stefano Magni, Valentina D.A. Corino, and Luca T. Mainardi. Ensemble classification combining ResNet and handcrafted features with three-steps training. *Physiological measurement*, 43(9), 9 2022.

List of Publications

Journal Publications

- **García-Isla, G.**, Corino, V., Mainardi, L. (2020). Poincaré plot image and rhythm-specific atlas for atrial bigeminy and atrial fibrillation detection. *IEEE Journal of Biomedical and Health Informatics, 25(4)*, 1093-1100.
- **García-Isla, G.**, Mainardi, L., Corino, V. D. (2021). A Detector for Premature Atrial and Ventricular Complexes. *Frontiers in Physiology*, 820.
- **García-Isla, G.***, Muscato, F.*, Sansonetti, A., Magni, S., Corino, V., Mainardi, L. (2022). Ensemble Classification Combining ResNet and Handcrafted Features with Three-Steps Training. *Physiological Measurement, 43(9)*, 094003.

Conference Proceedings

- **Garcia-Isla, G.**, Corino, V., Mainardi, L. (2019, September). Cardiac tachyarrhythmia detection by Poincaré plot-based image analysis. In *2019 Computing in Cardiology (CinC)* (pp. Page-1). IEEE.
- **García-Isla, G.**, Laureanti, R., Corino, V. D., Mainardi, L. T. (2020, September). ECG Morphological Decomposition for Automatic Rhythm Identification. In *2020 Computing in Cardiology* (pp. 1-4). IEEE.
- **García-Isla, G.**, Corino, V. D., Mainardi, L. T., "Poincaré Image-Based Atrial Fibrillation Detection for Photoplethysmography Signals," in *2020 Gruppo Nazionale di Bioingegneria (GNB)* proceedings, 2021.
- **García-Isla, G.**, Mainardi, L., Corino, V. D. (2021, September). A Poincaré Image-Based Detector of ECG Segments Containing Atrial and

Ventricular Beats. In 2021 Computing in Cardiology (CinC) (Vol. 48, pp. 1-4). IEEE.

- Magni, S.*, Sansonetti, A.*, Salvi, C., Tabiadon, T., **García-Isla, G.** (2021, September). Combining ResNet Model with Handcrafted Temporal Features for ECG Classification with varying number of leads. In 2021 Computing in Cardiology (CinC) (Vol. 48, pp. 1-4). IEEE.