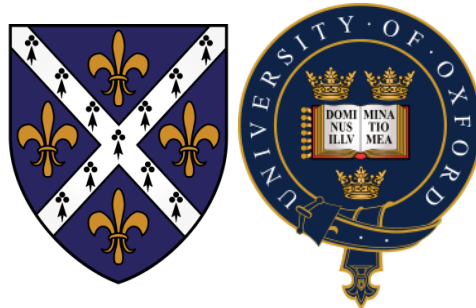


*COMPUTATIONAL DETECTION OF
ELECTROPHYSIOLOGICAL ABNORMALITIES
IN HYPERTROPHIC CARDIOMYOPATHY FOR
RISK STRATIFICATION*



Aurore Lyon

St Hugh's College

Computational Cardiovascular Science Group

Department of Computer Science

University of Oxford

Michaelmas 2017

**This thesis is submitted to the Department of Computer Science, University of Oxford,
for the degree of Doctor of Philosophy. This thesis is entirely my own work, and,
except where otherwise indicated, describes my own research**

To my parents

Avec une pensée pour Claude et Philippe

COMPUTATIONAL DETECTION OF ELECTROPHYSIOLOGICAL ABNORMALITIES IN HYPERTROPHIC CARDIOMYOPATHY FOR RISK STRATIFICATION

ABSTRACT

Hypertrophic cardiomyopathy (HCM) is a common cardiac genetic disease and a major cause of sudden cardiac death (SCD) in young adults. While most patients remain asymptomatic, others suffer from SCD triggered by ventricular arrhythmias. An accurate detection of these high-risk patients in order to provide them with appropriate treatment is therefore essential, and it remains a challenge as current electrocardiogram (ECG) biomarkers are not specific. In this thesis, we develop computational methods for the analysis and interpretation of electrophysiological clinical data in order to investigate the diversity of HCM phenotypes with the ultimate aim to improve risk stratification in HCM. First, by combining computational clustering and mathematical modelling, we identify four distinct ECG phenotypes exhibiting differences in hypertrophy distribution and risk of arrhythmia. The group with primary repolarization abnormalities and coexistence of apical and septal hypertrophy shows a higher HCM Risk-SCD score compared to other groups. Second, we explore the influence of structural and electrophysiological mechanisms on the ECG to explain the four HCM phenotypes identified, by using a whole-body personalized 3D computer simulation framework. We show that apico-basal repolarization heterogeneities explain the T wave inversions in the high risk group, and that an abnormal Purkinje system may explain the QRS abnormalities in another group. Finally, we further investigate the presence of repolarization biomarkers in HCM, specifically action potential alternans, and show that HCM cardiomyocytes do not overall exhibit more action potential alternans compared to controls, which may suggest other whole-organ mechanisms responsible for T wave alternans. Overall, this thesis contributes towards a better understanding of HCM phenotypic heterogeneity and the improvement of individual patient management.

Publications

The publications resulting from the work described in this thesis are listed below.

- **A. Lyon**, A. Mincholé, J P. Martínez, P. Laguna and B. Rodriguez. Computational techniques for ECG analysis in light of their contribution to medical advances. *J. R. Soc. Interface*. 2018 15 20170821; DOI: 10.1098/rsif.2017.0821.
- **A. Lyon**, R. Ariga, A. Mincholé, M. Mahmood, E. Ormondroyd, P. Laguna, N. de Freitas, S. Neubauer, H. Watkins and B. Rodriguez. Distinct ECG phenotypes identified in hypertrophic cardiomyopathy using machine learning associate with arrhythmic risk markers. *Front Physiol*. 2018; 9:213. DOI: 10.3389/fphys.2018.00213
- **A. Lyon**, A. Mincholé. E. Passini and B. Rodriguez. Investigation of the Presence and Mechanisms of Action Potential Alternans in Hypertrophic Cardiomyopathy. *Computing in Cardiology*. 2017.
- **A. Lyon** et al. Investigation of the effect of structural abnormalities on the ECG in hypertrophic cardiomyopathy patients using MRI-based computer models of the human heart and torso. *Submitted to Europace*.
- **A. Lyon**, R. Ariga, A. Mincholé, P. Laguna, N. de Freitas, S. Neubauer, H. Watkins, B. Rodriguez. Risk stratification in hypertrophic cardiomyopathy based on QRS and T wave morphological biomarkers identifies three phenotypic subgroups. *European Heart Journal*, 37 (Abstract Supplement), 412-413. 2016
- A. Ghetti, V. Grau, S. Harmer, I. Kopljar, P. Lambiase, H. Rong Lu, **A. Lyon**, A. Mincholé, A. Muszkiewicz, J. Oster, M. Paci, E. Passini, S. Severi, P. Taggart, A. Tinker, J.-P. Valentin, A. Varro, M. Wallman, X. Zhou, Human-based approaches to pharmacology and cardiology: an interdisciplinary and intersectorial workshop. *Europace* 18(9). 2016.
- **A. Lyon**, A. Mincholé, R. Ariga, P. Laguna, N. de Freitas, S. Neubauer, H. Watkins, B. Rodriguez. Extraction of morphological QRS-based biomarkers in hypertrophic cardiomyopathy for risk stratification using L1 regularized logistic regression. *Computing in Cardiology*. DOI: 10.1109/CIC.2015.7408573. 2015

Conference presentations

The work described in this thesis was presented at the following conferences:

- **CMR2018**, Barcelona, February 2018 (invited talk)
- **Computing in Cardiology**, Rennes, September 2017 (talk)
- **BHF Centre for Research Excellence Symposium** – Oxford, September 2017 (invited talk)
- **European Medical and Biological Engineering Conference (EMBEC)**, Tampere, June 2017 (poster)
- **BHF Centre for Research Excellence Symposium** – Oxford, September 2016 (poster)
- **European Society of Cardiology (ESC) Congress**, Rome, August 2016 (talk)
- **Young Investigator Award Competition at Computing in Cardiology**, Nice, 2015 (talk)
- **BHF Centre for Research Excellence Symposium** – Oxford, September 2014 (poster)

Prizes

The work described in this thesis was awarded the following prizes:

- **Best talk**, Oxford and Cambridge Women in Computer Science Conference, Oxford, March 2017
- **Winner of the Young Investigator Award**, Computing in Cardiology, Nice, September 2015
- **Honorary mention for Best Poster**, BHF Centre for Research Excellence Symposium, Oxford, September 2014

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors, Professor Blanca Rodríguez and Dr Ana Mincholé, for all their support over the last three years. Ana, thank you for being available, supportive, reassuring and optimistic every single day. Most of this work would never have been done without your help! Blanca, you have been an incredible model to look up to and I have learnt so much by your side. Thank you for supporting me through failure and doubt, for reminding me of the big picture when I needed it, and for teaching me life lessons I will not forget!

The work presented here would not exist without the help and support of many collaborators. Special thanks to Dr Rina Ariga, for providing us with the rich dataset used in this work, and for all her hard work and patience. I would also like to thank Professor Hugh Watkins for his implication in this research, Dr Ernesto Zacur, for all the work and guidance on the imaging work, Professor Nando de Freitas, for his fruitful ideas, and Professor Pablo Laguna. Pablo, muchas gracias por todas las discusiones tan productivas que hemos tenido, tus geniales ideas y sobre todo por tu apoyo y generosidad. Thank you to the Department of Computer Science, for a fantastic work environment, and the British Heart Foundation for funding my DPhil.

A big thank you to all my friends who made these three years amazing: Laure, for being such a good friend, and the best food and sport mate ever, Héctor, for all your inappropriate jokes and your continuous patience with me, Cristian, for your never-ending Italian energy, Anna, my DPhil mum, Nicolas, my best tennis partner, and everyone in the CCS group for making everyday life in the office so nice. Thanks to my friends back in France: Emilie, Adrien, Tatiana, Nils, for staying in touch despite my rare trips back to Paris. Thank you to St Hugh's College and its MCR for making me feel part of the family, and to all the sporting teams I was lucky to join, especially OULTC, and OULGC.

Separately, thank you to Kayla Friedman and Malcolm Morgan of the Centre for Sustainable Development, University of Cambridge, for the Microsoft Word thesis template used to produce this document.

Finally, I would like to thank all my family, especially my mum, dad and sister for supporting me every step of the way!

CONTENTS

PUBLICATIONS	IV
CONFERENCE PRESENTATIONS	V
PRIZES	V
1 INTRODUCTION	1
1.1 INTRODUCTION	1
1.2 THESIS AIMS.....	4
1.3 THESIS OUTLINE	5
2 CARDIAC ELECTROPHYSIOLOGY AND HYPERTROPHIC	
CARDIOMYOPATHY	8
2.1 INTRODUCTION	8
2.2 ELECTROPHYSIOLOGY OF THE HEALTHY HEART	9
2.2.1 <i>Cardiac anatomy and function</i>	9
2.2.2 <i>Cardiac electrical propagation: from single cell to whole organ</i>	11
2.2.3 <i>Electrocardiogram recording</i>	15
2.3 HYPERTROPHIC CARDIOMYOPATHY (HCM).....	19
2.3.1 <i>HCM structural and electrophysiological abnormalities</i>	20
2.3.2 <i>Risk stratification in HCM</i>	23
2.3.3 <i>ECG abnormalities and standard biomarkers in HCM</i>	25
2.3.4 <i>Imaging techniques for the identification of HCM structural abnormalities</i>	26
3 COMPUTATIONAL METHODS FOR ELECTROCARDIOGRAM ANALYSIS	
AND INTERPRETATION	28
3.1 INTRODUCTION	28

3.2 SIGNAL PROCESSING TECHNIQUES FOR ECG SIGNALS	29
3.2.1 Noise removal and filtering of Holter recordings.....	29
3.2.2 Delineation of ECG waveforms	31
3.3 MACHINE LEARNING TECHNIQUES FOR CLASSIFICATION OF ECG SIGNALS	38
3.3.1 Objectives of state of the art methods and available databases	38
3.3.2 Feature extraction and dimensionality reduction.....	40
3.3.3 Main machine learning techniques for ECG classification	43
3.3.4 Discussion	54
3.4 PERSONALIZED WHOLE-BODY COMPUTER SIMULATIONS	56
3.4.1 Modelling of cardiac anatomy	57
3.4.2 Modelling of the cardiac electrical propagation	59
3.4.3 Modelling of cardiac heterogeneities	62
4 EXTRACTION OF ECG PHENOTYPES IN HCM BASED ON QRS MORPHOLOGICAL BIOMARKERS	64
4.1 INTRODUCTION.....	64
4.2 MOTIVATION	65
4.3 METHODS.....	66
4.3.1 Clinical database	66
4.3.2 ECG signal processing	68
4.3.3 Extraction of standard and Hermite based biomarkers.....	71
4.3.4 Binary classification between HCM and controls	74
4.3.5 Dimensionality reduction and clustering.....	75
4.3.6 Statistical analysis	77

4.4 RESULTS	77
4.4.1 Binary classification between HCM and Controls	77
4.4.2 Identification of three ECG phenotypes based in QRS biomarkers.....	83
4.4.3 Association with MRI and arrhythmic risk factors.....	87
4.5 DISCUSSION	90
4.6 CONCLUSIONS	91
5 EXTRACTION OF HCM PHENOTYPES WITH QRS AND T WAVE MORPHOLOGICAL BIOMARKERS	92
5.1 INTRODUCTION	92
5.2 METHODS.....	93
5.2.1 Extraction of standard and KLT-based repolarization biomarkers	93
5.2.2 Computational clustering	96
5.3 RESULTS	98
5.3.1 Identification of four ECG phenotypes based on QRS and T wave biomarkers.	98
5.3.2 Association with MRI and arrhythmic risk factors.....	101
5.4 DISCUSSION	106
5.5 CONCLUSIONS	109
6 INTERPRETATION OF ECG PHENOTYPES USING PERSONALIZED WHOLE-BODY CMR-BASED WHOLE-BODY 3D SIMULATIONS	111
6.1 INTRODUCTION	111
6.2 MOTIVATION.....	112
6.3 METHODS.....	113
6.3.1 3D ventricular mesh generation from cardiac magnetic resonance images	113
6.3.2 Cardiac electrophysiology modelling and ECG simulation.....	117

6.3.3 <i>Modelling of structural and electrophysiological abnormalities</i>	120
6.4 RESULTS.....	123
6.4.1 <i>Anatomy alone does not explain ECG abnormalities identified in HCM phenotypes</i>	123
6.4.2 <i>Conduction abnormalities can explain deep and wide S waves in Group 3 patients</i>	128
6.4.3 <i>Apex-base repolarization heterogeneities can explain T wave inversions as primary repolarization abnormalities</i>	137
6.5 DISCUSSION.....	141
6.6 CONCLUSIONS	143
7 INVESTIGATION OF THE PRESENCE OF ACTION POTENTIAL ALTERNANS IN HCM	145
7.1 INTRODUCTION.....	145
7.2 MOTIVATION	146
7.3 METHODS.....	147
7.3.1 <i>Populations of control and HCM cardiomyocytes models</i>	147
7.3.2 <i>Cardiac action potentials simulation</i>	149
7.3.3 <i>Computation of action potential biomarkers</i>	150
7.4 RESULTS.....	152
7.4.1 <i>Comparison of action potential alternans models in HCM and control populations</i>	152
7.5 DISCUSSION.....	158
7.6 CONCLUSIONS	159
8 CONCLUSIONS AND FUTURE DIRECTIONS	161

8.1 SUMMARY OF THE THESIS.....	161
8.1.1 <i>Key contributions and main findings</i>	164
8.2 FUTURE DIRECTIONS	166
8.2.1 <i>Implementation of the novel ECG criteria in a larger database with cardiovascular end-points</i>	166
8.2.2 <i>Modelling of complex T wave shapes</i>	166
8.2.3 <i>Investigation of the presence of T wave alternans in HCM</i>	167
8.2.4 <i>Simulation of ischaemia in HCM patients</i>	167
8.3 CONCLUSIONS	168
9 BIBLIOGRAPHY	170
10 APPENDIX.....	190
APPENDIX A: SUMMARY TABLE OF MACHINE LEARNING METHODS USED FOR ECG SIGNAL CLASSIFICATION.....	191

LIST OF TABLES

TABLE 3.1: SUMMARY TABLE OF THE MAJOR DATABASES USED FOR CLASSIFICATION OF ECG SIGNALS.	39
TABLE 4.1: CHARACTERISTICS OF NORMAL CONTROLS AND HCM PATIENTS.....	78
TABLE 4.2: CLINICAL AND GENOTYPE CHARACTERISTICS OF HCM PATIENTS.....	79
TABLE 4.3: DISCRIMINATIVE POWER OF THE STANDARD QRS BIOMARKERS.....	82
TABLE 4.4: RANKING OF BIOMARKERS ACCORDING TO THEIR INFLUENCE ON THE CLASSIFICATION PROCESS AS WELL AS THE CUMULATIVE PERFORMANCE EVALUATION	82
TABLE 4.5: STANDARD ECG FEATURES FOR QRS-BASED HCM GROUPS	86
TABLE 4.6: CLINICAL FEATURES FOR QRS-BASED HCM GROUPS.....	87
TABLE 5.1: CHARACTERISTICS OF THE FOUR HCM PHENOTYPES FROM COMBINED QRS AND T WAVE CLUSTERING	99
TABLE 5.2: CHARACTERISTICS OF THE FOUR HCM PHENOTYPES FROM COMBINED QRS AND T WAVE CLUSTERING	103
TABLE 6.1: DIFFERENT SETS OF INTRACELLULAR AND EXTRACELLULAR CONDUCTIVITIES EXPLORED	121
TABLE 6.2: IONIC ELECTROPHYSIOLOGICAL REMODELLING IN HCM AS DESCRIBED IN (PASSINI ET AL. 2016).....	122
TABLE 7.1: IONIC ELECTROPHYSIOLOGICAL REMODELLING IN HCM AS DESCRIBED IN (PASSINI ET AL. 2016).....	149
TABLE 7.2: NUMBER OF CONTROLS AND HCM MODELS DISPLAYING ALTERNANS	154

LIST OF FIGURES

FIGURE 2.1: HEALTHY CARDIAC ANATOMY.	10
FIGURE 2.2: ORGANIZATION OF THE CARDIOMYOCYTES..	11
FIGURE 2.3: SEQUENCE OF CARDIAC ELECTRICAL ACTIVATION.....	12
FIGURE 2.4: SCHEMATIC REPRESENTATION OF A VENTRICULAR CARDIAC ACTION POTENTIAL	14
FIGURE 2.5: PLACEMENT OF THE ELECTRODES FOR A 12-LEAD ECG RECORDING.	16
FIGURE 2.6: RELATION BETWEEN THE CELLULAR ACTION POTENTIALS IN VARIOUS CARDIAC REGIONS AND THE BODY SURFACE ECG FOLLOWING THE ELECTRICAL PROPAGATION PATH.	17
FIGURE 2.7: ECG WAVEFORMS, EXAMPLES OF NORMAL AND ABNORMAL ECG RECORDINGS..	19
FIGURE 2.8: STRUCTURAL ABNORMALITIES AFFECTING HCM HEARTS.....	22
FIGURE 2.9: IMAGING TECHNIQUES FOR STRUCTURAL ABNORMALITIES DETECTION..	27
FIGURE 3.1: EXAMPLE OF DATA AVAILABLE FOR THE ANALYSIS OF ECG SIGNALS.	42
FIGURE 3.2: MAIN MACHINE LEARNING METHODS USED FOR ECG CLASSIFICATION.....	44
FIGURE 3.3: COMPUTATIONAL PIPELINE FOR ECG SIMULATION FROM MAGNETIC RESONANCE IMAGES	59
FIGURE 4.1: ECG SIGNAL PROCESSING PIPELINE.....	71

FIGURE 4.2: STANDARD BIOMARKERS EXTRACTED FROM THE QRS WAVEFORM.....	72
FIGURE 4.3: FIRST FOUR HERMITE BASIS USED IN THE QRS RECONSTRUCTION	73
FIGURE 4.4: COMPARISON OF STANDARD (QRS WIDTH, AMPLITUDE, ASCENDING AND DESCENDING SLOPES) AND MORPHOLOGICAL (FOUR HERMITE COEFFICIENTS) QRS BIOMARKERS.....	81
FIGURE 4.5: CLUSTERING WITH QRS MORPHOLOGY ALONE IDENTIFIES THREE HCM GROUPS SHOWING GROUP DIFFERENCES IN 3 FEATURES IN LEAD II AND THE LATERAL PRECARDIAL LEADS.....	84
FIGURE 4.6: COMPARISON OF THE FOUR HERMITE COEFFICIENTS IN LEADS I-II, V1-6 FOR THE THREE QRS-BASED HCM GROUPS.	84
FIGURE 4.7: THREE DISTINCT QRS MORPHOLOGIES OF THE THREE SUBGROUPS IDENTIFIED..	86
FIGURE 4.8: LOCATION OF MAXIMUM LEFT VENTRICULAR WALL THICKNESS IN THE THREE GROUPS.	89
FIGURE 5.1: STANDARD BIOMARKERS EXTRACTED FROM THE STT SEGMENT WAVEFORM....	94
FIGURE 5.2: FOUR KARHUNEN-LOÈVE TRANSFORM BASIS FUNCTIONS FOR THE ST-T SEGMENT.....	95
FIGURE 5.3: SUMMARY OF THE METHODOLOGIES APPLIED IN THIS WORK FOR THE ANALYSIS AND CLASSIFICATION OF 86 HCM PATIENT ECGS.	97
FIGURE 5.4: FOUR GROUPS WITH DISTINCT ECG MORPHOLOGY ARE IDENTIFIED BY COMBINED CLUSTERING WITH QRS MORPHOLOGY AND T WAVE BIOMARKERS.	98

FIGURE 5.5: FOUR COEFFICIENTS OF THE KL TRANSFORM FOR GROUPS 1A, 1B, 2 AND 3 IN LEADS I,II, V1-V6.....	101
FIGURE 5.6: FOUR DISTINCT ECG PHENOTYPES IN HYPERTROPHIC CARDIOMYOPATHY EXHIBIT DIFFERENCES IN HYPERTROPHY MORPHOLOGY AND ARRHYTHMIC RISK..	105
FIGURE 6.1: VOLUMETRIC MESH GENERATION.....	115
FIGURE 6.2:HISTOGRAMS OF EDGE RATIO, ASPECT GAMMA, MINIMUM DIHEDRAL ANGLE AND RADIUS RATIO FOR THE TETGEN-GENERATED MESH OF EDGE LENGTH 4MM.....	116
FIGURE 6.3: SENSITIVITY ANALYSIS FOR CONDUCTIVITY VALUES PERFORMED ON A 1D FIBRE	118
FIGURE 6.4: SIMULATION CONFIGURATION.....	119
FIGURE 6.5: SUMMARY OF THE SIMULATION PIPELINE.	120
FIGURE 6.6: ACTION POTENTIAL DURATION MAP SHOWING LONGER ACTION POTENTIAL DURATION IN THE APICAL REGION WITH HYPERTROPHY.....	122
FIGURE 6.7: TOTAL MESH AND HEART MESH FOR EACH REPRESENTATIVE PATIENT OF GROUPS 1A, 1B AND 3.....	124
FIGURE 6.8: REAL ECG (BLACK) AND SIMULATED ECG (BLUE) OF PATIENTS IN GROUP 1A (PANEL A), GROUP 1B (PANEL B) AND GROUP 3 (PANEL C) IN THE NORMAL INITIAL CONFIGURATION, FOR LEADS V1-V6.....	126
FIGURE 6.9: ACTIVATION MAPS (MS) FOR PATIENTS IN GROUP 1A, GROUP 2 AND GROUP 3	127
FIGURE 6.10: ANALYSIS OF CONDUCTIVITY CHANGES ON THE ECG..	128

FIGURE 6.11: FRACTIONAL ANISOTROPY MAPS FOR GROUP 3 PATIENT, AT DIASTOLE, FOR BASE, MID AND APEX SLICES..	129
FIGURE 6.12: SIMULATED ECG WITH ISOTROPIC ORIENTATION OF THE MYOCYTE FIBRES ..	130
FIGURE 6.13: MODELLING OF SLOW CONDUCTION IN VARIOUS LOCATIONS OF THE MYOCARDIUM.	132
FIGURE 6.14: MODELLING OF FAST CONDUCTION IN VARIOUS LOCATIONS OF THE MYOCARDIUM.	133
FIGURE 6.15: CHANGES IN THE ACTIVATION SEQUENCE OF THE LEFT VENTRICLE WITH DIFFERENT BLOCKING COMBINATIONS	135
FIGURE 6.16: ALTERATION OF THE FAST ENDOCARDIAL PROPAGATION.	136
FIGURE 6.17: MODELLING OF ALTERED FAST PROPAGATION SYSTEM	137
FIGURE 6.18: SIMULATED ECG WITH LONGER APD IN THE APEX.....	138
FIGURE 6.19: REPOLARIZATION MAPS SHOWING THE TIME OF REPOLARIZATION (MS) OF EACH NODE, FOR A NORMAL SIMULATION AND A SIMULATION WITH THE I_{Kr} CONDUCTANCE DIVIDED BY 2 IN THE APEX	139
FIGURE 6.20: SIMULATION WITH HCM IONIC REMODELLING IN THE APEX.	140
FIGURE 7.1: (A) BIOMARKERS EXTRACTED FROM THE CARDIAC ACTION POTENTIAL: ACTION POTENTIAL DURATION AND AMPLITUDE. (B) BIOMARKERS EXTRACTED FROM THE CARDIAC CALCIUM TRANSIENT TRACES: CAT DURATION AND AMPLITUDE.	151
FIGURE 7.2: TWO CONSECUTIVE BEATS SHOWING ACTION POTENTIAL DURATION ALTERNANS	152

FIGURE 7.3: BOXPLOT OF APD90 (MS) AND CATD90 (MS) FOR HCM AND CONTROL POPULATIONS AT A CYCLE LENGTH OF 1000MS	153
FIGURE 7.4: PERCENTAGE OF MODELS DISPLAYING ALTERNANS FOR CONTROL AND HCM.	154
FIGURE 7.5: BOXPLOT COMPARING THE MAGNITUDE OF AP ALTERNANS (MV) IN CONTROL AND HCM POPULATIONS.	157

LIST OF ABBREVIATIONS AND ACRONYMS

AF – atrial fibrillation

AHA – American Heart Association

AP – action potential

APA – action potential amplitude

APDX – action potential duration at X% repolarization

AV - atrioventricular

BP – blood pressure

BPR – blood pressure response

CaT – calcium transient

CaTA – calcium transient amplitude

CaTDX – calcium transient duration at X% repolarization

CMR – cardiovascular magnetic resonance

CVD – cardiovascular disease

DTI – diffusion tensor imaging

ECG – electrocardiogram

FA – fractional anisotropy

G+LVH- – genotype positive HCM with normal left ventricular wall thickness

HCM – hypertrophic cardiomyopathy

HMM – Hidden Markov model

HRV – heart rate variability

ICD – implantable cardioverter-defibrillator

LA – left atrium

LAD – left axis deviation

LGE – late gadolinium enhancement

LV – left ventricle

LVH – left ventricular hypertrophy

LVOT – left ventricular outflow tract

MRI – magnetic resonance images

MYBPC3 – myosin-binding protein C

MYH7 – beta-myosin heavy chain

NSVT – non-sustained ventricular tachycardia

NYHA – New York Heart Association

PVC – premature ventricular beat

RA – right atrium

RMS – root mean square

RV – right ventricle

SA – sinoatrial

SCD – sudden cardiac death

SNR – signal to noise ratio

SVM – support vector machine

TNNT2 – troponin T

TWA – T wave alternans

TWI – T wave inversion

VEB – ventricular ectopic beat

VT – ventricular tachycardia

VF – ventricular fibrillation

1 INTRODUCTION

1.1 Introduction

The heart is a muscular organ, divided into four chambers, which constantly provides oxygenated blood and nutrients to the rest of the organism. It works as a pump contracting periodically to eject blood from the lungs to the body. This pumping behaviour is ruled by electrical signals initiated by pacemaker cells and propagating through the conduction system in a specific sequence to generate cardiac contraction. The atria (upper chambers of the heart) first receive the contraction signal and push the blood to the ventricles (lower chambers). The ventricles then receive the electrical signal to contract and send the blood to the rest of the body. This cycle must remain steady and regular to avoid life-threatening episodes such as arrhythmias that may lead to sudden cardiac arrest. Currently, cardiovascular diseases (CVD) are a leading cause of mortality worldwide and the World Health Organization reported that 17.7 million people died from CVD in 2015, representing 31% of all global deaths. Cardiovascular research, aiming at better understanding heart conditions and improving diagnosis and treatment, is therefore an active field of research to reduce the burden of CVDs.

Hypertrophic cardiomyopathy (HCM) is a common yet challenging genetic heart disease, with a prevalence of 1 in 500 people (Semsarian et al. 2015), and characterized by the

thickening of the muscle between the two cardiac ventricles. It is a major cause of sudden cardiac death (SCD) in young adults, yet most patients remain asymptomatic with a normal life expectancy (Maron 2002). Because of this heterogeneous clinical course, the identification of patients at risk is a clinical priority as implantable cardioverter-defibrillators (ICD) could successfully treat ventricular arrhythmias triggering SCD. The use of ICD has confirmed that ventricular arrhythmias are the precipitating mechanism in sudden cardiac death (Maron et al. 2000). HCM hearts suffer from structural changes such as hypertrophy (Spirito et al. 2000), myocyte disarray (Varnava et al. 2001) or fibrosis (Adabag et al. 2008), as well as ion-channel dysfunction (Coppini et al. 2013) that may create an unstable electrical milieu predisposing to arrhythmia. These abnormalities may affect the electrical activation (contraction of the heart) and repolarization (relaxation), and, in turn, the electrocardiogram (ECG), which records the cardiac electrical activity on the body surface. ECG abnormalities have been reported in HCM such as pathological Q waves, wide QRS, prolonged QT interval and inverted T waves (Lakdawala et al. 2011) (Savage et al. 1978), but these biomarkers are not specific to identify patients at risk (Sherrid et al. 2009) (Maron, Roberts, and Epstein 1982). These discrete clinically interpreted ECG features, visually extracted from paper ECGs, do not objectively capture the shape diversity of ECG complexes, and despite this biophysical underpinning, the ECG does not yet feature in current risk assessments for HCM (Christiaans et al. 2010). Indeed, initial risk stratification is comprised of six clinical factors: family history of HCM-related sudden death, unexplained collapse, massive thickening of the heart, short-lived ventricular tachycardia (potentially fatal heart rhythm), left atrial size and blood pressure drop on exercise (O'Mahony et al. 2014). These however may overestimate risk and lead to inappropriate ICD implantation in some patients. Risk may also be underestimated in other patients as SCD may occur in low risk patients (Maron et al. 2007). The validation of new methods for risk stratification remains a challenge because of the low numbers of adverse

events in HCM. Indeed, in spite of being a major cause of cardiac mortality in young adults, the incidence of SCD in HCM is less than 1% per year (O'Mahony, Elliott, and McKenna 2013), which therefore prevents its use as an endpoint in most studies. Clinical risk factors and risk scores such as the one developed by the European Society of Cardiology guidelines (Elliott et al. 2014) are thus more used. HCM risk stratification may therefore be improved through further characterisation of the ECG by using novel computational analysis techniques such as the ones proposed here, as well as additional biomarkers such as the occurrence of repolarization alternans (Kuroda et al. 2002).

Computational techniques have recently emerged as powerful tools to analyse biological and medical data thanks to their ability to handle large and heterogeneous datasets. Machine learning techniques have for instance been widely developed for classification and clustering of ECG signals, from heartbeat classification (Llamedo and Martínez 2011) (de Chazal and Reilly 2006) to automated patient diagnosis (Pourbabae and Lucas 2008) (Rahman et al. 2015). Similarly, computer modelling and simulation frameworks have been developed in the last decades and have provided useful insight in cardiac mechanisms from cell to whole organ. They allow the personalized simulation of the cardiac activity based on magnetic resonance images (MRI) (Vigmond et al. 2009) (Dutta et al. 2016). The combination of detailed imaging modalities such as diffusion tensor imaging, or T1 mapping, with computer modelling has also enabled to explain the relationships between structure of the heart and cardiac function, which are challenging to investigate *in vivo* (Sermesant et al. 2006). Recent computational techniques such as machine learning and computer modelling therefore provide a set of methods able to make sense of multivariate complex datasets and detect differences that might be challenging for the human eye (Obermeyer and Emanuel 2016) (Deo 2015).

The ECG signal provides a non-invasive representation of the cardiac activity and its analysis can lead to diagnosis of cardiac disorders. To this end, signal processing and computational techniques allow the extraction of biomarkers, features describing the ECG, such as time-interval biomarkers, characterizing the temporal dynamics of the electrical propagation during a cardiac cycle, morphological biomarkers, quantifying the shape of the ECG waveforms such as amplitude, width or overall morphology, or the occurrence of T wave alternans, illustrating an alternation in the beat-to-beat behaviour of the signal at different frequencies that has been linked to increased arrhythmic risk.

Therefore, by developing computational approaches to analyse and simulate medical data, one can help answer key questions of arrhythmia research and, in the context of HCM, understand its phenotypic heterogeneity and accelerate improvements in individual patient management.

1.2 Thesis aims

The overall goal of this thesis is to develop computational methods for the analysis and interpretation of electrophysiological clinical data in order to investigate the diversity of HCM phenotypes with the ultimate aim to improve risk stratification in HCM. As mentioned before, HCM is a major cause of sudden cardiac death in young adults but most patients remain asymptomatic and reliable methods to distinguish who will remain well with HCM and who will be devastated by disease complications are lacking. We propose a novel approach of combining clustering, mathematical modelling and computer simulations to address this challenge. Our work is based on the hypothesis that both HCM structural and electrophysiological abnormalities affect the activation and repolarization phase of the cardiac cycle resulting in electrophysiological abnormalities in the ECG of different degrees of severity in different patient groups.

This overall goal can be divided into three specific research questions:

1. Can we identify distinct phenotypes in hypertrophic cardiomyopathy associated with different arrhythmic risk based on the ECG alone?
2. What are the HCM structural and electrophysiological abnormalities that can explain the distinct ECG phenotypes identified through aim 1?
3. To what extent does HCM affect rate-dependant biomarkers such as repolarization alternans?

The first question is addressed using mathematical modelling and computational clustering based on morphological QRS and T wave biomarkers to identify distinct ECG phenotypes and compare them in terms of clinical markers and markers of arrhythmic risk. Whole-body simulations help address the second question by modelling the impact of structural and electrophysiological changes on the electrical propagation, in order to investigate the mechanisms that may explain the ECG abnormalities observed in the different phenotypes identified in the first research question. To answer the third question, we hypothesize that the HCM abnormalities affecting the repolarization dynamics such as dysregulation of the calcium cycling may promote the appearance of pro-arrhythmic alternans. Computational modelling and simulation using a population of models approach is implemented to account for inter-subject variability and investigate the presence of action potential alternans in HCM.

1.3 Thesis outline

Chapter 2 begins this thesis by presenting the background knowledge on cardiac electrophysiology and hypertrophic cardiomyopathy. We first describe the healthy cardiac anatomy and the mechanisms regulating the cardiac electrical propagation from the exchange of ion channels to the creation of electrical waveforms at the body surface level

recorded by the electrocardiogram. We then introduce hypertrophic cardiomyopathy and the structural and electrophysiological abnormalities that characterize it. The main challenges of HCM risk stratification are discussed and we then review the current knowledge for detection of ECG and imaging abnormalities in HCM.

Chapter 3 describes the computational methods currently in use for ECG analysis and interpretation. We first review the signal processing methods and machine learning techniques developed for analysis and classification of ECG signals, along with their applications, performances and challenges. We then introduce the field of 3D computer simulations and how they have recently become powerful tools for the understanding of electrophysiological mechanisms from cell to whole-organ.

Chapter 4 constitutes the first result chapter of this thesis, presenting the identification of distinct ECG phenotypes in HCM based only on the QRS waveform of the ECG, using mathematical modelling and blinded clustering. These groups are compared in terms of arrhythmic risk and clinical markers from cardiac magnetic resonance images. We conclude that QRS morphology allows the extraction of three groups with significantly distinct ECG features, but which do not associate with differences in clinical features and markers of arrhythmic risk, suggesting that cardiac repolarization may play a key role in arrhythmogenesis.

Chapter 5 naturally follows by introducing T wave biomarkers and evaluating their influence on the identification of HCM phenotypes when added to the QRS biomarkers in the clustering methodology. Four groups with distinct ECG morphology are identified and exhibit differences in terms of arrhythmic risk and hypertrophy distribution. In particular, the group displaying normal QRS complexes and inverted T waves shows higher HCM SCD-Risk score (O'Mahony et al. 2014) compared to other groups.

Chapter 6 presents the personalized multiscale whole-body 3D simulation framework used to investigate the influence of HCM structural and electrophysiological abnormalities on the ECG and explain the mechanisms behind the ECG abnormalities observed in the four phenotypes identified in Chapter 5. In particular, we highlight that abnormalities in the simplified Purkinje-based activation model may be responsible for the QRS abnormalities in lateral leads in one of the groups. We also show that apex-base repolarization heterogeneities potentially due to ionic remodelling can explain primary repolarization abnormalities in the group with apical hypertrophy.

Repolarization abnormalities proving to be of importance in HCM risk stratification, Chapter 7 extends the investigation of repolarization-based biomarkers to the cellular and ionic basis of repolarization alternans in HCM. Computational modelling and simulation using a population of models approach accounting for inter-subject variability are used to simulate Control and HCM action potentials at various frequencies to investigate the presence of alternans. We observe that HCM cardiomyocytes do not overall exhibit more action potential alternans compared to Controls, which may suggest that T wave alternans on the ECG may be due to other dynamics such as tissue heterogeneities at the whole-organ level.

Finally, Chapter 8 summarizes the main contributions and findings of the thesis, discusses the limitations in the methodology, and suggests future lines of research, beyond the scope of this thesis, to build on the work presented here.

2 CARDIAC ELECTROPHYSIOLOGY AND HYPERTROPHIC CARDIOMYOPATHY

2.1 Introduction

This chapter introduces the necessary background in cardiac electrophysiology and describes hypertrophic cardiomyopathy and its challenges. We first present the electrophysiology of the healthy heart by detailing cardiac anatomy and function. We describe the structure of the heart and the organization of cardiac cells. We then detail the mechanisms of electrical propagation in the heart from cellular to organ level and how this is recorded on the electrocardiogram from the body surface. Finally, we introduce the hypertrophic cardiomyopathy disease and review the state of the art knowledge for its risk stratification and the current techniques allowing the identification of structural and electrophysiological abnormalities.

2.2 Electrophysiology of the healthy heart

2.2.1 Cardiac anatomy and function

The heart is a muscular organ responsible for constantly pumping blood between the lungs and the rest of the body. It ensures an uninterrupted flow of oxygen and nutrients to the organism. The heart is composed of four chambers (Figure 2.1, Panel A): the right and left atria (RA and LA), the smaller upper chambers, and the right and left ventricles (RV and LV), the bigger lower chambers. The upper border of the heart, including the left atrium and part of the right atrium is referred to as the base, and the lowest part of the heart, opposite the base and including the inferolateral part of the left ventricle, is called the apex. Atria and ventricles contract in sequence, creating a cycle of pressure and volume changes called cardiac cycle (Figure 2.1, Panel B). During ventricular diastole, blood flows from the atria to the ventricles. During ventricular systole, blood is then ejected to the pulmonary artery and the lungs, in the case of the right ventricle, and in the aorta and the rest of the body for the left ventricle. The left ventricle pushes the blood out to the body and its wall is usually thicker because generating a higher pressure due to a higher systemic cardiovascular resistance. The right ventricle ejects the blood in the pulmonary circulation to the lungs. The two sides of the heart are separated by a muscular wall called the septum.

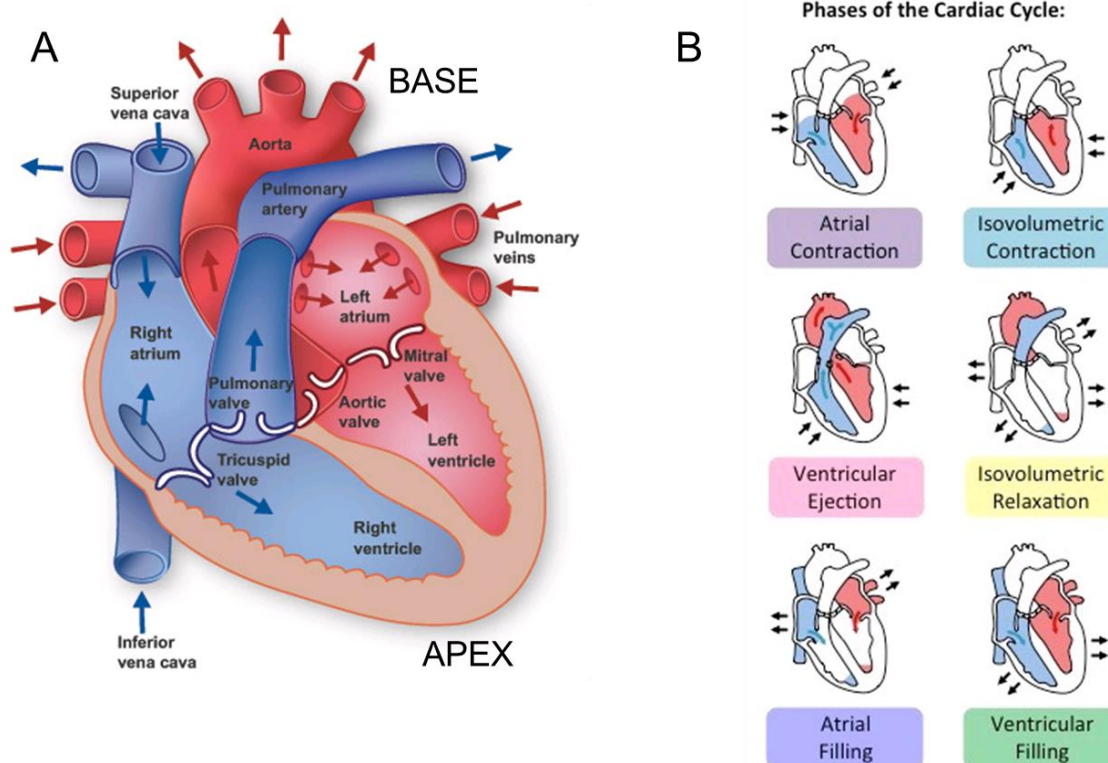


Figure 2.1: Healthy cardiac anatomy. Panel A: the blue part represents the right side of the heart, dealing with deoxygenated blood, and the right part is the left side of the heart, dealing with oxygenated blood. From the Texas Heart Institute (www.texasheart.org). Panel B: diagram representing the different phases of the cardiac cycle (<http://ib.bioninja.com.au>).

The tissue surrounding the chambers is called the myocardium and contracts to trigger the ejection of the blood. The myocardium is composed of muscular cells, the cardiomyocytes, arranged in adjacent layers and connected by gap junctions, responsible for the transition of ions between neighbouring cells (Figure 2.2, Panel A). Cardiomyocytes have a cylindrical shape and measure around $100\mu\text{m}$, with a diameter of $10\text{-}25\mu\text{m}$. They allow the beating of the heart by shortening of the sarcomeres (the basic contractile units of the myocytes) (Bers 2002). Their orientation sets the preferred direction of contraction of the heart and defines the propagation of the electrical signal (Figure 2.2, Panel B). Any alteration to this organization because of structural changes such as fibre disarray or scarring will affect the proper beating of the heart.

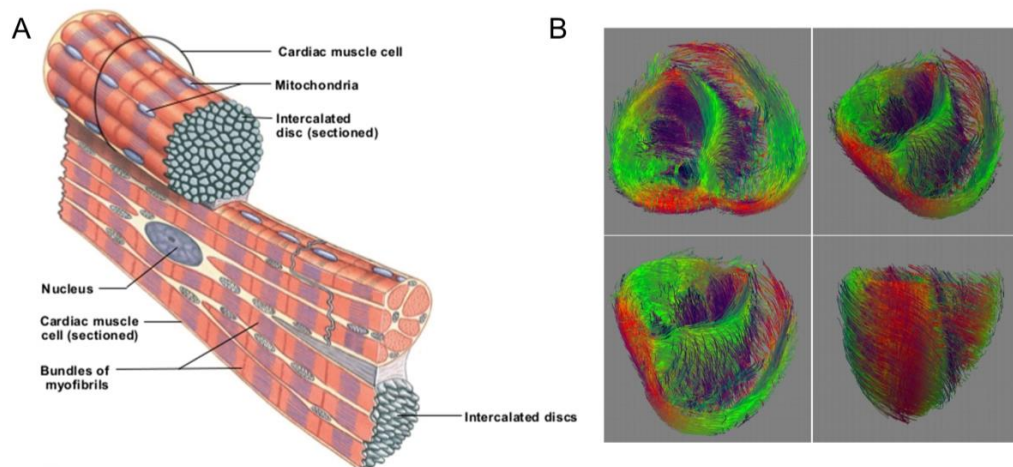


Figure 2.2: Organization of the cardiomyocytes. Panel A: structure of the cardiac muscular cells in sheets, connected with intercalated discs (gap junctions). From: © 2012 Pearson Education, Inc. Figure 20-5b. Panel B: reconstruction of the heart muscle fibres displaying their arrangement in a spiral direction (Zhukov and Barr 2003).

2.2.2 Cardiac electrical propagation: from single cell to whole organ

The mechanical contraction of the heart is generated by an electrical impulse that propagates through the cells, in a specific pattern, triggering their contraction. Most of the cardiomyocytes are not able to self-excite and rely on the reception of an external electrical stimulus to produce an electrical signal, known as an action potential. The electrical signal originates from the sinoatrial (SA) node, composed of pacemaker cells able to control the heart rate, which stimulates the atria. It then reaches the atrioventricular (AV) node, from which it propagates to the ventricles through the fast conduction system: the bundle of His and the Purkinje fibres. This is known as the activation sequence of the heart and is illustrated on Figure 2.3.

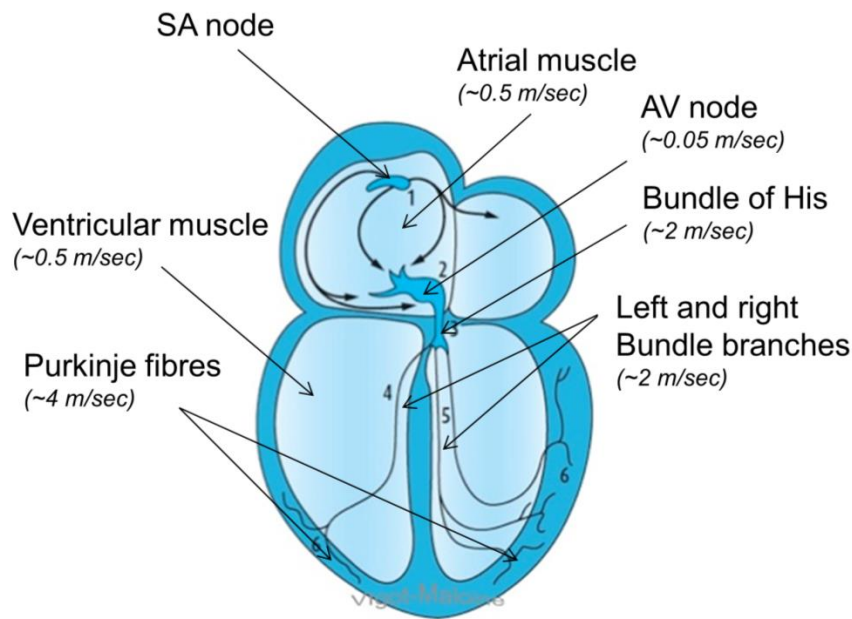


Figure 2.3: Sequence of cardiac electrical activation: from the sinoatrial (SA) node to the atrioventricular (AV) node, followed by the bundle of His and the Purkinje fibres. Speeds of propagation are also mentioned. The bundle of His and Purkinje fibres constitute the fast conduction system (from www.cvphysiology.com)

When the non-self excitable cardiac cells are not stimulated, their membrane remains at a negative potential because of the balance of the concentrations of ion channels around it. When excited by the activation stimulus, the cardiomyocytes produce electrical activity, known as an action potential (AP), composed of changes in voltage, initiated by a rapid depolarization and followed by a slower repolarization phase back to the resting state (Figure 2.4, Panel A). These changes in the membrane voltage are due to changes in specific ionic currents as explained below.

The shape of the ventricular cardiomyocyte action potential may vary across the different cells of the ventricle, but it usually has a duration between 150 and 350ms and is composed of the following four different phases (Figure 2.4, Panel A). Phase 0 corresponds to the initial upstroke, the depolarization phase, due to an increase in sodium ions (Na^+) in the

cell. This is followed by Phase 1, which initiates the repolarization phase when the maximum depolarization has been reached. I_{Na} becomes inactive, and the transient outward potassium current (I_{to}) decreases, resulting in the first deflection (notch) observed on the AP trace and the ejection of potassium ions (K^+) out of the cell. The next phase is called the plateau phase, in which the L-type calcium current (I_{CaL}), slow delayed rectifying potassium current (I_{Ks}), inward rectifying potassium current (I_{K1}) and sodium calcium exchanger (I_{NaCa}) and sodium potassium pump (I_{NaK}) play a role. In this phase, Ca^{2+} ions enter the cell, and K^+ ions leave the cell. In Phase 3, the cell repolarizes completely and the potassium ions leave the cells using the inward rectifying potassium current (I_{K1}) and the fast and slow delayed rectifying potassium currents (I_{Kr} and I_{Ks}). Finally, in Phase 4, the cell comes back to the resting phase in which the voltage is around -85mV and K^+ ions steadily enter the cell. A more detailed description of the ionic channel properties during a normal cardiac cycle can be found in (Grant 2009). Figure 2.4, Panel B models a cardiac cell and recapitulates the main actors in ionic currents interactions.

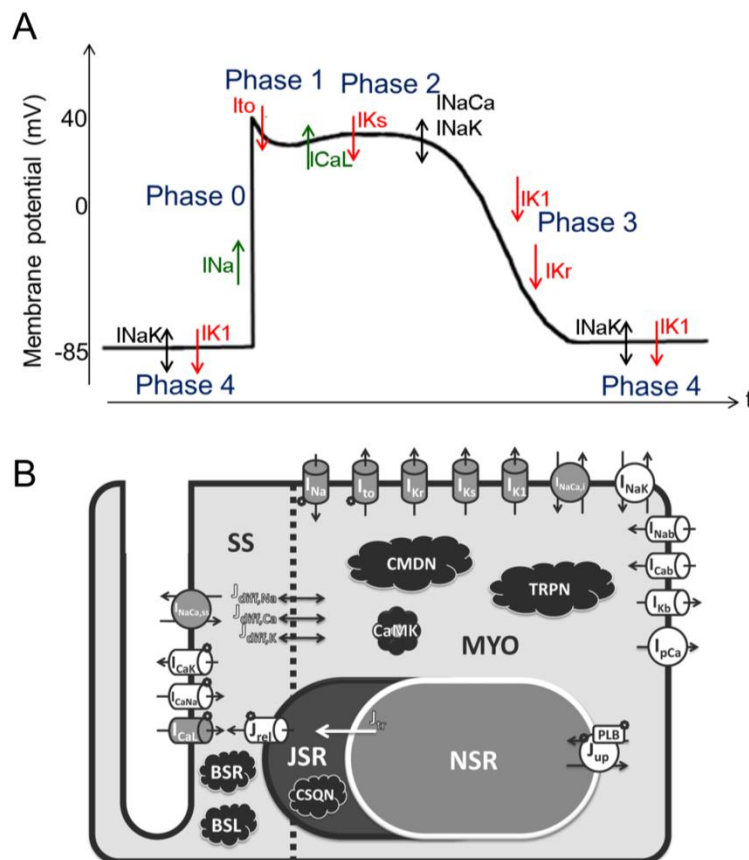


Figure 2.4: Panel A: Schematic representation of a ventricular cardiac action potential and its four main phases: upstroke (phase 0), initial repolarization (phase 1), plateau (phase 2), complete repolarization (phase 3) and resting phase (4). The ionic currents i involved in each phase are also represented. Panel B: schematic representation of a cardiac cell, along with the main ionic currents and pumps involved.

This exchange of ions in the cardiomyocytes is made possible by the presence of ion channel gating, mechanism which regulates the flux of ion channels through the membrane. When the cardiomyocyte is at resting state, most of the ion channels are closed, maintaining a negative membrane voltage. Under the trigger of the electrical stimulus, the membrane potential voltage changes and initiates the opening of the ion channels which, in turn, allow more ions going through the membrane.

2.2.3 Electrocardiogram recording

The propagation of the electrical activity throughout the heart can be recorded on the body surface by placing electrodes on the skin. The signal acquired is the electrocardiogram (ECG). Firstly introduced by Waller in the 1880s, the ECG was really developed by Einthoven in 1902-1903, who defined the location of the electrodes for the recording of the three limb leads I, II and III, and measured the first ECG signals (Herring and Paterson 2006). The ECG is cheap, non-invasive, easy to acquire and therefore widely used in clinical practice, allowing the diagnosis of heart rhythm disorders and cardiac diseases. In order to record the variations in electrical potential during the cardiac cycle, electrodes are placed on the skin at specific positions. In a standard 12-lead ECG, ten electrodes are positioned on the patient. Six electrodes are placed over the chest (V1 to V6), and four others on the patient's arms (RA: right arm, LA: left arm) and legs (RL: right leg, LL: left leg) (Figure 2.5, Panel A). With this electrodes' configuration, 12 leads are defined as follows:

- Limb leads, formed by electrodes located on the limbs:
 - Lead I = LA – RA, measuring the voltage between LA and RA.
 - Lead II = LL-RA.
 - Lead III = LL-LA.
- Augmented limb leads, using the Wilson's terminal (V_w) that averages the measurements from RA, LA and LL electrodes:
 - Lead aVR= $3/2 (RA - V_w)$
 - Lead aVL= $3/2 (LA - V_w)$
 - Lead aVF= $3/2 (LL - V_w)$
- Precordial leads, perpendicular to the six other leads and measuring the voltage between each of the electrodes (V1 to V6) (Figure 2.5, Panel A) and the V_w .

Each of the 12 leads represents the electrical activity from 12 different angles providing different views of the cardiac electrical conduction as shown in Figure 2.5, Panel B. Leads II, III and aVF are referred to as the inferior leads, providing information on the inferior wall of the left ventricle. Leads V1 and V2 are referred to as septal leads, providing information on the septal wall of the left ventricle. They are often grouped together with the anterior leads (V3 and V4), which view the anterior wall of the left ventricle. ST-segment elevation in leads V1-V4 is referred to as “LAD occlusion”. Finally, the lateral leads (I, aVL, V5 and V6) give information on the lateral wall of the left ventricle.

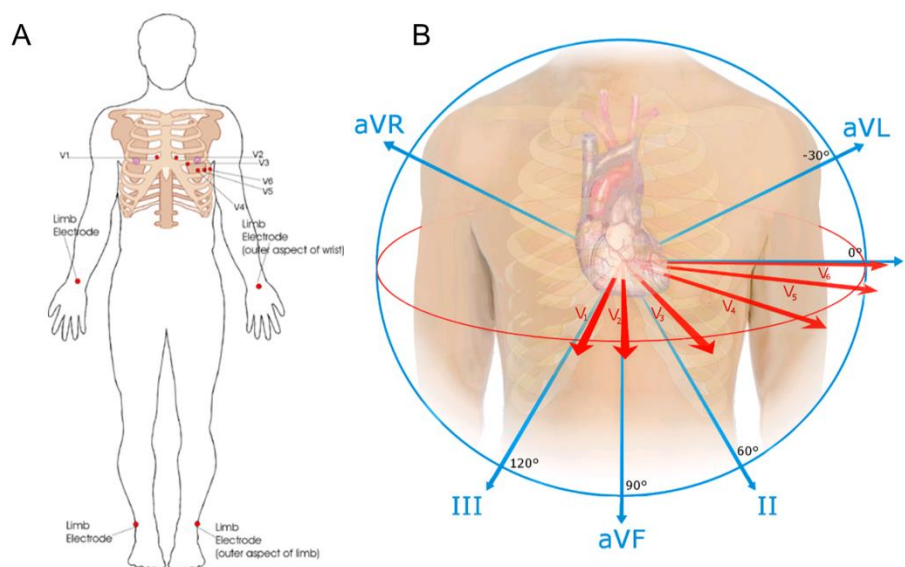


Figure 2.5: Placement of the electrodes for a 12-lead ECG recording. Six electrodes are placed on the chest (V1-V6), and four electrodes are placed on the arms and legs (limb electrodes). From: <http://www.gla.ac.uk/>

ECG is a measure of the global cardiac electrical activity, and it is therefore derived from the action potentials in the cardiac cells described in the previous section. Figure 2.6 nicely describes the succession of action potentials from the sinoatrial node to the ventricles according to the propagation path of the electrical impulse producing the ECG waveform.

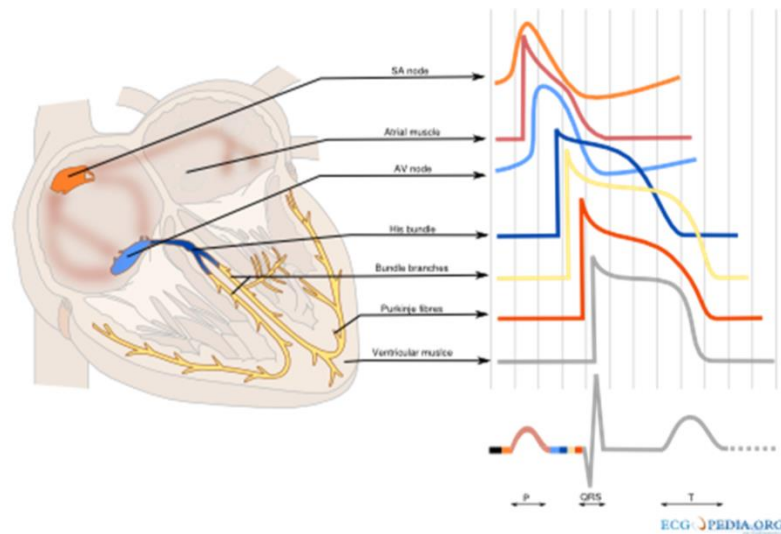


Figure 2.6: Relation between the cellular action potentials in various cardiac regions and the body surface ECG following the electrical propagation path.

Durations of recordings may vary, from around 10 seconds in a standard ECG routine, to hours or days thanks to the use of portable ECG recording devices (such as Holter recordings).

The ECG signal is a succession of heartbeats, each heartbeat representing a complete cardiac cycle from contraction to relaxation. A heartbeat is composed of three main waveforms (represented on Figure 2.7, Panel A): the P wave, representing the depolarization of the atria, the QRS complex, representing the depolarization of the ventricles, and the T wave, representing the ventricular repolarization (a U wave is also sometimes observed, and has been associated to repolarization of the papillary muscles and also a delayed repolarization of the Purkinje fibres). The first detected wave of an ECG beat is the P wave. Of upright polarity across all the leads except in aVR, the P wave amplitude should not be more than 0.3mV. Tall (respectively wide) P waves may illustrate right (respectively left) atrial enlargement. The next waveform is the QRS complex, representing the depolarization (activation) of the ventricles. Its usual duration for normal

beats is around 70-110ms and its amplitude is the largest of any ECG waveform (around 2-3mV). It is itself composed of different waves. The Q wave, sometimes not visible, is the first downward deflection after the P wave, and corresponds to the depolarization in the septum. The R wave is the first upward deflection after the P wave and corresponds to the early depolarization of the ventricles. Finally, the S wave is the first negative deflection after the R wave, corresponding to the late depolarization of the ventricles. Across the different leads, the QRS morphology follows an evolution, especially in the chest leads, starting from a short R wave and deep S wave morphology in lead V1 until a large R wave morphology in V6 (Figure 2.7, Panel B). Finally, the T wave is the final waveform of an ECG beat, and represents the repolarization of the ventricles. It is normally round and upright across the leads (and can be inverted in leads V1 and III with respiration), although it can exhibit a wide range of morphological variability. Any alteration of this normal ECG morphology (some diseased cases are illustrated in Figure 2.7, Panel C) may indicate abnormalities in the heart electrical conduction and therefore detect a cardiac disease.

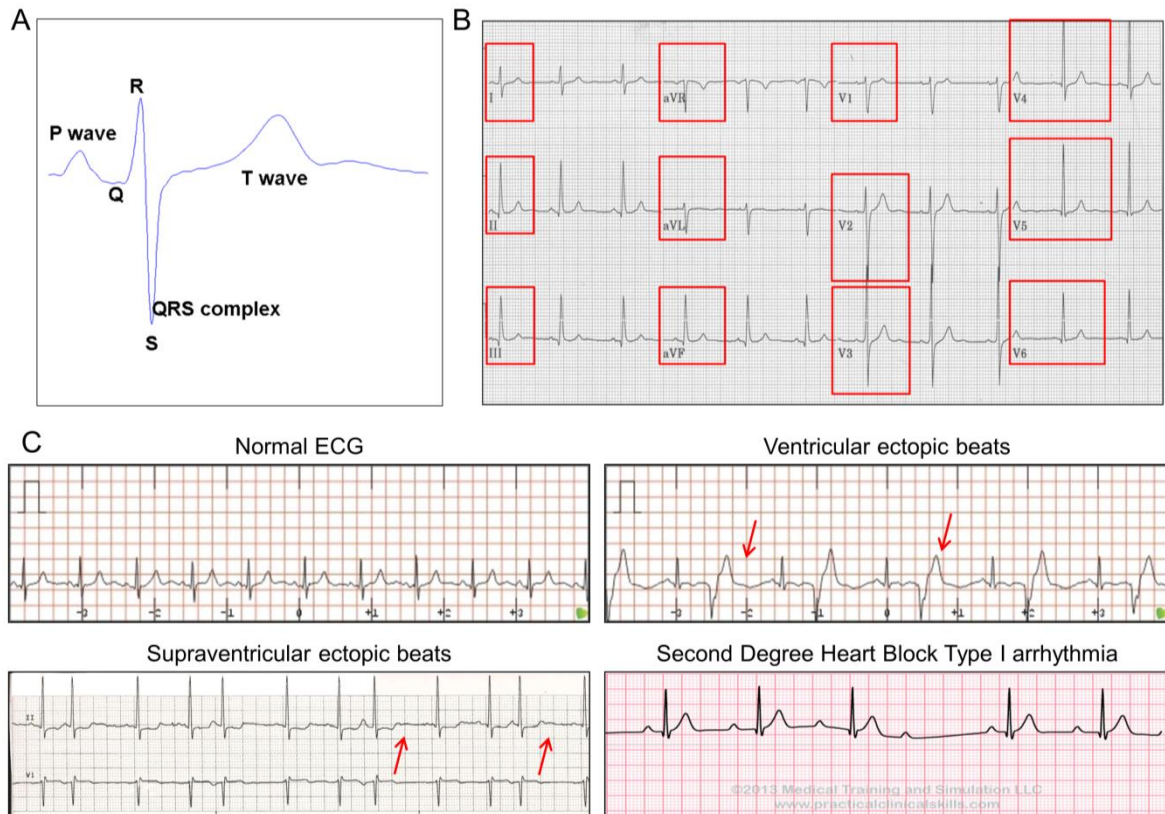


Figure 2.7: ECG waveforms, examples of normal and abnormal ECG recordings. Panel A: The main ECG waveforms constituting a heartbeat: the P wave, representing the depolarization of the atria, the QRS complex, representing the depolarization of the ventricles, and the T wave, representing the repolarization of the ventricles. Panel B: example of a normal ECG morphology in the 12 leads I, II, III, aVR, aVL, aVF, V1 to V6. Panel C: example of various ECG abnormalities (ventricular ectopic beats, supraventricular ectopic beats, and second degree heart block).

2.3 Hypertrophic cardiomyopathy (HCM)

Hypertrophic cardiomyopathy (HCM) is a common cardiac genetic disease with a prevalence of 1 in 500 people (Semsarian et al. 2015), due to the presence of mutations in the genes coding for the proteins of the cardiac sarcomere. It is a major cause of death in young adults (and athletes) and may lead to severe outcomes such as stroke or heart failure. However, the majority of HCM patients remains asymptomatic and does not develop disease complications. Diagnosis is performed either based on the presence of a genetic mutation, and/or left ventricular hypertrophy not related to any other cause (usually

observed on echocardiography or magnetic resonance imaging) (Maron and Maron 2013). Implantable cardioverter-defibrillators allow the efficient treatment of arrhythmias triggering sudden cardiac death (SCD) in HCM, but the rate of inappropriate shocks remains a concern (Lin et al. 2009, Magnusson et al. 2015) and may lead to complications when they occur (Maron et al. 2000). This is why the main challenge in management of HCM is the accurate identification of high-risk patients in order to provide them with appropriate treatment, and to limit the number of unnecessary defibrillators implantation.

2.3.1 HCM structural and electrophysiological abnormalities

In HCM, both ionic remodelling (Coppini et al. 2013) and structural abnormalities (hypertrophy (Spirito et al. 2000), myocyte disarray (Varnava et al. 2001) and fibrosis (Adabag et al. 2008)) create a pro-arrhythmic substrate to different extents in specific patients.

First, HCM hearts are strongly affected by structural abnormalities that may lead to a disruption of the normal electrical propagation. They exhibit left ventricular (LV) hypertrophy (Figure 2.8, Panel A). Its clinical diagnosis is made by the observation of a hypertrophied non-dilated left ventricle (defined as a wall thickness ≥ 15 mm), in the absence of another cardiac disease that could cause hypertrophy (Maron et al. 2003).

In addition to LV hypertrophy, fibre disarray is a hallmark of HCM (Figure 2.8, Panel C). Indeed, HCM mutations affect 11 or more genes coding for myofilament contractile proteins (Watkins 2001). 70% of gene positive patients exhibit mutations in genes coding for two main proteins: the β -myosin heavy chain (MYH7) and myosin-binding protein C (MYBPC3) (Watkins et al. 1992). Troponin T (TNNT2) and other genes may also be affected and account for 5% or less of cases (Watkins et al. 1995) (Alcalai, Seidman, and Seidman 2008). These mutations lead to a thickening of the cardiac sarcomere, responsible

for the contraction of the cells, and, in turn, disrupt the regular alignment of the cardiomyocytes. Fibre disarray therefore alters cardiac contraction and may lead to life-threatening disruption of the heart rhythm and sudden cardiac death. However, the extent and progression of fibre disarray in HCM patients remain unknown. Indeed, disarray may be patchy, and therefore not clearly identifiable by biopsy, and may also be present to various degrees in normal and diseased hearts. Until recently, disarray could only be identified post mortem, making its diagnostic use very poor. With the recent progress in imaging techniques, diffusion tensor imaging has allowed the assessment of fibre disarray non-invasively (more details are provided in section 2.3.4).

Finally, HCM hearts may show signs of myocardial fibrosis (Figure 2.8, Panel B). Different extent of replacement fibrosis (scarring) and interstitial fibrosis are observed in HCM patients, with more severity in areas of larger hypertrophy. Myocardial fibrosis is the result of a large increase in the collagen volume fraction of myocardial tissue. Collagen is a protein that ensures a normal shape and function of the ventricle. When the collagen is present in large quantity, the myocytes are replaced by tissue unable to contract, affecting the diastole and systole behaviours of the ventricle. Late gadolinium enhancement (LGE) and T1-mapping techniques allow the identification of fibrosis on CMR images. More details are provided in section 2.3.4.

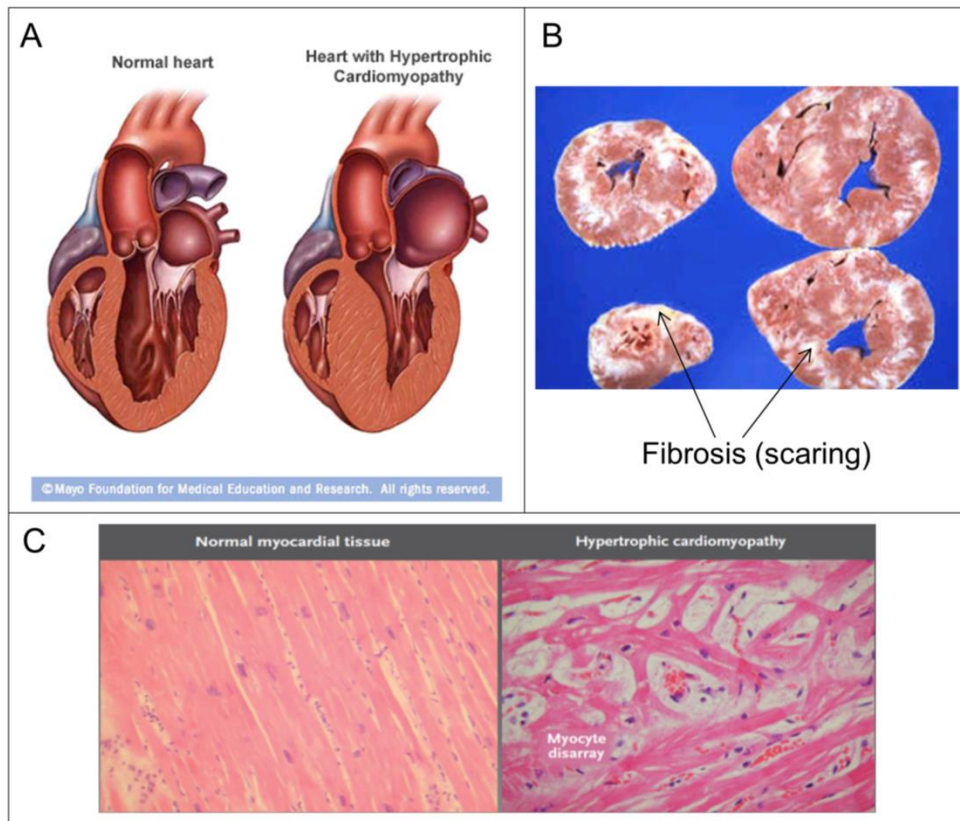


Figure 2.8: Structural abnormalities affecting HCM hearts. Panel A: hypertrophy – in the right heart, the left ventricular septum is thicker than in the normal left heart. Panel B: fibrosis – white patches on the heart slices indicate fibrosis (or scarring). Panel C: fibre disarray – on the left tissue, myocytes are properly aligned along the same direction, while on the right tissue, the fibres are completely disorganised (Watkins et al, 2011).

In addition to these structural abnormalities, HCM hearts also suffer from electrophysiological abnormalities and ionic remodelling. Abnormalities were reported in the cardiomyocytes action potential, calcium subsystem, sarcomere ionic currents and mRNA expression (Coppini et al. 2013). More precisely, HCM cardiomyocytes show an increase in L-type calcium current (I_{CaL}), an increase in Late sodium current (I_{NaL}) and a reduction in potassium (K^+) currents: transient outward current (I_{to}) and inward rectifier potassium current (I_{K1}), which lead to increased action potential and calcium transient duration and reduced mRNA expression of genes coding for K^+ channels. HCM cardiomyocytes also exhibit more repolarization abnormalities than control

cardiomyocytes such as early afterdepolarization (spontaneous depolarizations during the plateau phase of the action potential), due to a reopening of Na^+ and Ca^{2+} channels, and delayed afterdepolarization, due to spontaneous Ca^{2+} release.

These electrophysiological changes may also explain some of the phenotypic signature of HCM. Indeed, longer action potentials durations and increased occurrence of repolarization abnormalities may explain an increased risk of arrhythmias, and prolonged intracellular transients and higher diastolic concentration Ca^{2+} may explain abnormal muscle contraction as well as diastolic dysfunction in HCM patients (Coppini et al. 2013).

2.3.2 Risk stratification in HCM

The identification of patients at high risk of arrhythmias and sudden cardiac death (SCD) is crucial in HCM, as implantable cardioverter defibrillator (ICD) can effectively prevent SCD (Maron et al. 2000). However, the heterogeneous outcomes of the disease and the unpredictability of arrhythmia triggers make HCM risk stratification a challenge. In addition, the threshold for ICD implementation cannot be too low, as it is a costly and invasive therapy that may lead to inappropriate shocks and the development of disease complications. Therefore, the accurate identification of patients who would benefit for ICD treatment is essential in HCM risk assessment. Current risk stratification in HCM is based on five conventional risk factors providing clinical utility (Christiaans et al. 2010):

- non-sustained ventricular tachycardia (NSVT), defined as three or more consecutive ventricular beats at a rate of ≥ 120 bpm lasting less than 30s on a clinical 3-lead 24- to 48-hour Holter monitoring
- unexplained syncope, defined as one or more episodes of unexplained syncope

- family history of sudden cardiac death (SCD), defined as the presence of history of SCD in one or more first degree relative younger than 40 years old, or SCD in a first degree relative with confirmed HCM at any age
- massive left ventricular hypertrophy (LVH), defined as left ventricular wall thickness in any myocardial segment greater than 30mm on short axis CMR images
- abnormal exercise blood pressure (BP) response, defined as a rise in systolic BP inferior to 20 mmHg or a fall of more than 10 mmHg from baseline to peak exercise in patients ≤ 40 years old

Despite providing guidance for the measurement of risk in HCM patients, these risk factors still exhibit low specificity, as shown by the occurrence of SCD in patients labelled as low risk (Maron 2010). In addition, it is notable that patients displaying none of these risk factors may still suffer from sudden death, as described in (Spirito et al. 2014), highlighting the need for a more accurate assessment of SCD risk in HCM. In addition, these risk factors can sometimes not be easily applied to children (Spirito et al. 2000).

In 2014, a new prospectively validated prediction model was introduced by the ESC guidelines (Elliott et al. 2014). It showed to perform better than conventional risk factors, although also with limitations (Maron et al. 2015). This HCM Risk-SCD score is calculated for each patient using seven disease variables in the following formula:

$$probability_{SCD \text{ at } 5 \text{ years}} = 1 - 0.998 * \exp(Prognostic \text{ index}),$$

where

$$Prognostic \text{ index} = [0.15939858 * maximal \text{ wall thickness (mm)}] - [0.00294271 \times maximal \text{ wall thickness}^2 (mm^2)] + [0.0259082 \times left \text{ atrial diameter (mm)}] + [0.00446131 \times maximal \text{ (rest/Valsalva) LVOT gradient (mmHg)}] + [0.4583082 \times$$

$$\begin{aligned} & \text{family history of SCD}] + [0.82639195 \times \text{NSVT}] + [0.71650361 \times \\ & \text{unexplained syncope}] - [0.01799934 \times \text{age at clinical evaluation (years)}] \end{aligned}$$

A risk score of $\geq 6\%$ is classified as high risk and ICD implantation is recommended, 4–6% is intermediate risk and ICD may be considered, and $< 4\%$ is low risk.

2.3.3 ECG abnormalities and standard biomarkers in HCM

The resting 12-lead ECG is not currently used for risk stratification, due to a lack of reliable prediction with existing analysis methods and biomarkers, based on visually inspection of the paper ECG as described in (Lakdawala et al. 2011) (Savage et al. 1978). Q waves and repolarization abnormalities were reported as the most distinguishable abnormalities in HCM patients. Indeed, the Q-wave, representing the left to right depolarization of the septum, may be deeper and/or narrower, because of an asymmetrical septum due to hypertrophy. Abnormalities in the ST segment and T-waves may also be found, due to the electrophysiological abnormalities affecting HCM patients. T-wave inversions were also reported in some type of HCM (such as apical HCM). Wider QRS complexes may also be observed because of the wall-thickening affecting the conduction properties of the tissue and leading to a slower electrical propagation. However, despite the importance of left ventricular wall thickness in HCM diagnosis, ECG voltage criteria for left ventricular hypertrophy (such as Cornell voltage, Sokolow and Lyon index, or Romhilt-Estes point score), measures based on waveform amplitudes, were not found to be specific to HCM patients, with some patients not meeting any ECG criteria for LV hypertrophy (Dumont et al. 2006). In addition, it must be noted that normal ECG patterns may be found in HCM patients.

Despite providing a better description of the electrocardiographic signature of HCM, this approach failed to produce reliable stratification (Sherrid et al. 2009) (Elliott et al. 2000),

as ECG biomarkers did not prove to be specific. This may be due to the high heterogeneity of HCM and the limitations of visually assessing electrocardiographic features from paper ECGs. More sophisticated approaches and new knowledge are therefore required to uncover the potential of the ECG in HCM phenotyping.

2.3.4 Imaging techniques for the identification of HCM structural abnormalities

As mentioned in the sections above, HCM hearts suffer from structural abnormalities that can be detected using imaging techniques. For example, left ventricular hypertrophy can be detected by echocardiography (Figure 2.9, Panel A) as well as cardiac magnetic resonance (CMR) images, which provides a more precise and complete measure of left ventricular wall thickness thanks to a better spatial resolution and tomographic imaging (Figure 2.9, Panel B).

Fibre disarray and fibrosis are more complicated to detect on standard CMR images, but new techniques have recently been developed to help reveal these abnormalities. Diffusion tensor imaging (DTI), used in Chapter 6, is a very useful non-invasive tool to investigate the presence of myofibre disarray in living patients. This method measures the diffusion of water in the myocardium and enables to compute the magnitude, degree of anisotropy of the fibres and the orientation of diffusion anisotropy. The tensor of fibres orientation obtained provides information on the structure of the myocardium. For example, fractional anisotropy (FA) measures the directionality of water diffusion in the tissue. A low FA in a specific region could therefore be due to disarray in the fibres of this region, or increased spacing between myocytes because of the presence of fat cells in the tissue (Figure 2.9, Panel C).

The late gadolinium enhancement (LGE) technique is another CMR-based method, which allows the *in vivo* detection of myocardial fibrosis. The method is based on the observation of the diffusion of a contrast agent, the gadolinium, in the tissue. Normal tissue washes out the gadolinium after some time, while abnormal myocardium, exhibiting a larger volume and slower kinetics, still shows some remaining gadolinium after this time. Detecting the region with remaining gadolinium helps highlighting the regions with abnormal fibrosis tissue. T1-mapping is also used to detect diffuse fibrosis in HCM patients (Figure 2.9, Panel D). T1 is the longitudinal relaxation time constant of the myocardium. Comparing T1 and T1 after injection of gadolinium helps investigate the presence of fibrosis.

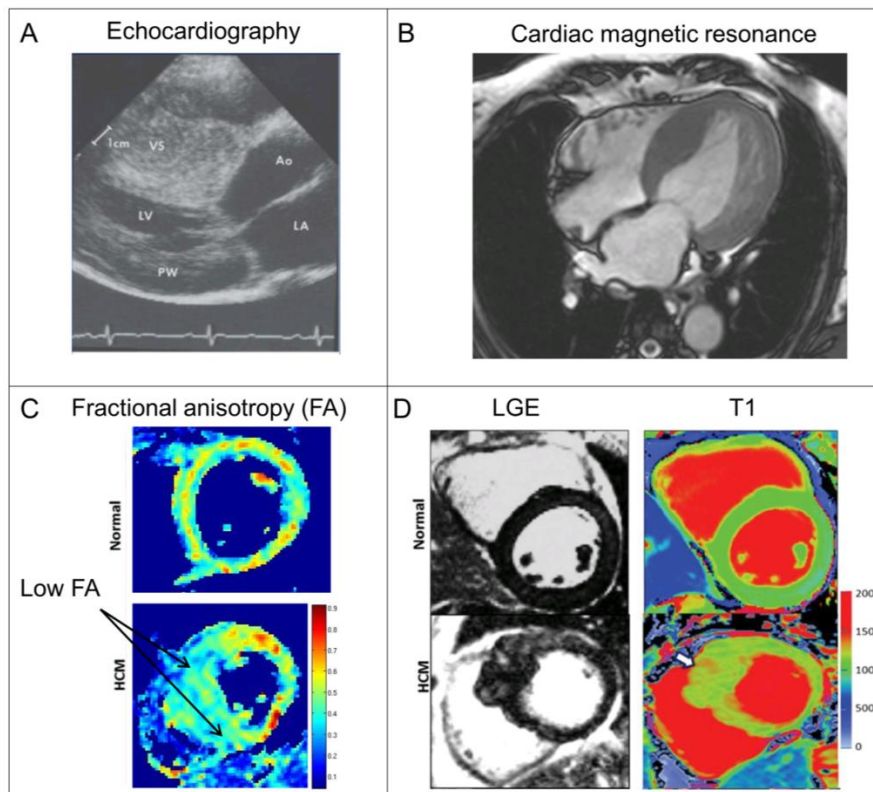


Figure 2.9: Imaging techniques for structural abnormalities detection. Panel A: left ventricular (LV) hypertrophy on echocardiography. Panel B: LV hypertrophy on cardiac magnetic resonance (CMR) images. Panel C: Fractional anisotropy (FA) from diffusion tensor imaging in normal and HCM patients. Fractional anisotropy (FA) is lower in the septum of the HCM patient. Panel D: late gadolinium enhancement (LGE) and T1 mapping for normal and HCM patient. Patches of high T1 value and white late gadolinium are present in the HCM patient.

3 COMPUTATIONAL METHODS FOR ELECTROCARDIOGRAM ANALYSIS AND INTERPRETATION

Sections of this chapter are part of: “Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances” by Lyon et al., 2018, Journal of the Royal Society: Interface, DOI: 10.1098/rsif.2017.0821

3.1 Introduction

Computational methods such as signal processing, pattern recognition, classification and 3D simulations have recently emerged as powerful tools for the analysis and interpretation of medical data, due to the heterogeneity and complexity of the multi-modality datasets available. Analysis of the electrocardiogram has benefited from these recent methodologies as visual inspection of ECG recordings is limited by the struggle of the human eye in identifying complex patterns in larger datasets. In this chapter, we introduce the main techniques developed to analyse and interpret ECG signals. We begin with a description of the signal processing methods used to denoise and segmentate ECG recordings. We then

focus on machine learning methods for classification and clustering, large field of research allowing the automated classification of large datasets of ECG signals. A wide range of classification methods have been proposed to analyse ECG datasets, finding various applications in heartbeat classification, disease diagnosis and long-term monitoring. Because of the variety of techniques available, reviewing these different classes enables to highlight their specific strengths, such as performance, interpretability or automaticity, as well as applications. Chapters 4 and 5 of this thesis are implementing ECG-based classification techniques and face similar challenges as the ones described here. Finally, we introduce the field of personalized 3D computer simulations, implemented in Chapter 6 of this thesis, and how they help better understand the cardiac mechanisms underlying cardiac electrical activity, as well as improve the interpretation of ECG abnormalities.

3.2 Signal processing techniques for ECG signals

3.2.1 Noise removal and filtering of Holter recordings

Analysing ECG signals, and more precisely Holter recordings, presents many challenges. In this section, we will highlight these difficulties and describe some of the methods developed to address them. These techniques are described in the book by (Sörnmo and Laguna 2005). Digital ECGs obtained from hospital recordings may be noisy because of various factors such as stress of the patient resulting in high frequency noise on the ECG signal, breathing or movement, or electrical recording interferences. Signal-to-noise ratio (SNR) estimation and filtering is an initial step to obtain a cleaner signal, by removing the beats with low SNR, in which the noise is more dominant than the actual signal. High-frequency noise and artefacts may also be removed by low-pass filtering, while keeping the main waveforms of interest exhibiting frequencies below 45Hz. Some ECG excerpts may also present an abnormally high baseline, disrupting the continuity of the ECG line. This baseline wander is not due to cardiac activity, but to experimental conditions such as

movement of the arms during a stress test, or breathing. Two different methods are commonly in use for baseline wander removal: time invariant filtering and cubic spline technique. Time invariant filtering assumes that the baseline wander frequency spectrum is below 0.5Hz and consists in a high-pass filtering of the ECG in order to remove the baseline while keeping the important cardiac information. The cubic spline method is based on the approximation of the baseline using knots, positions indicating the isoelectric level between the P wave and the QRS complex. A first order polynomial fitting will not represent the isoelectric line accurately as broken lines will connect the knots. However, increasing the order of the polynomial may lead to a risk of overfitting and increased computational complexity. For cubic spline estimation, a third order polynomial is used. Note that this technique performs better when the heart rate becomes higher, as the number of knots available for the fitting increases. Electromagnetic fields from the recording devices may also interfere with the signal by creating a powerline at 50-60Hz (50Hz in the UK). In order to reject this powerline, a notch filter, rejecting a specific band of the spectrum, may be used in order to remove this narrow frequency band.

These artificial noises are straightforward to remove as they mostly lie in different frequencies than the ECG waveforms. Muscular noise, which may appear during exercise recordings, is, on the contrary, challenging because of its overlap with the ECG waveforms. Time-varying low pass filtering with variable frequency response offers a suitable solution, as it allows a local filtering of the signal, more severe in smooth regions and almost null in QRS regions. However, it may introduce artificial wave because of signal distortion. Other methods such as single value decomposition filters have been proposed but none of them have been found suitable for clinical purposes and this still remains an open field of research.

3.2.2 Delineation of ECG waveforms

Wave delineation is a crucial step in ECG preprocessing. Indeed, ECG recordings capture the cardiac electric activity as a succession of heartbeats. In order to analyse the morphology of each heartbeat and different waves, the heart beats must be segmented and the beginning, end and main points of each waveforms identified. The accuracy of delineation is essential in ECG analysis, as any error or approximation at this stage will affect average signals, or any derived features. Therefore, it has become a large research interest in the signal processing field and many methods have been developed to tackle this challenge. The section below reviews the main techniques developed for ECG delineation: morphology-based methods, filtering techniques, Bayesian inference techniques, classification and pattern recognition methods, and transformative methods.

- Morphology-based methods

A direct approach to locate the boundaries of heartbeats waveforms is to mimic the manual segmentation of experts. Based on this idea, several techniques exploit the changes in the signal morphology such as changes in amplitude, slopes, or voltage. One major method for QRS detection was proposed by (Pan and Tompkins 1985), based on a dual-threshold method taking into account QRS slopes, amplitude and width. Two thresholds were computed over the filtered signal, and the signal obtained from moving window integration to improve the detection sensitivity and reduce false negatives. These thresholds were inferred during a learning phase, based on the MIT-BIH database annotations. This method reached 99.3% of accuracy for QRS detection on the MIT-BIH database (Moody and Mark 2001). As mentioned in this study, the slopes contain most of the information on the waves morphology. Therefore, the analysis of the first derivative of the signal should identify the boundaries. Based on this assumption, (Kemmelings et al. 1994) computed two signals: the sum of the absolute values of the first derivatives of the ECG leads and the sum of the

absolute values of the ECG leads. They used changes in the magnitude of the first derivatives signal as an indicator of the presence of the QRS complex, higher than for other waves such as the T-wave. The detection of QRS onsets and offsets were then performed by minimizing the sum of the absolute values of the signal. This method reached 95% of positive results, compared to human expert annotations, in the delineation of ventricular tachycardia signals, challenging because of the overlap of depolarization and repolarization waves. In another study, (de Chazal and Celler 1996) estimated global values over all the leads for QRS onsets and offsets after QRS detection. Onsets and offsets for each individual lead were then identified using a threshold method over the slopes. This method was validated over the Common Standards for Electrocardiography database tolerances for QRS onsets and offsets. Other biomarkers of ECG morphology were considered in different studies. For example, (Nygårds and Sörnmo 1983) used the envelope of the ECG, obtained from the Hilbert transform of the signal, to delineate QRS complexes. The envelope is a representation of the QRS in which it contains only one single positive pulse, easy to identify by derivatives. Morphology-based delineation is dependent on the quality of the preprocessed signal to avoid approximations due to noise. Thus, (Daskalov and Christov 1999) proposed the implementation of a strong preprocessing method, detecting QRS onset and offset on the exact signal, without distorting the shape of the ECG. The detection was performed by comparing the angle between two consecutive samples and a defined threshold. Morphology-based methods provide straightforward implementations at low computational costs, but require the definition of arbitrary thresholds, lowering the generalization performances to other databases.

- Filtering techniques

Filtering techniques alter the ECG signal in order to make its delineation easier. (Poli, Cagnoni, and Valli 1995) presented a method based on polynomial filters, optimized by

genetic algorithm for the choice of parameters. A linear filter with continuous samples provided the best predictivity and sensitivity of 99.6%. Later, (Soria-Olivas et al. 1998) applied adaptive filtering to the delineation of P and T waves, providing robustness to noise and baseline drift, and adaptable to clinicians criteria. Early, (Laguna et al. 1990) presented a robust method for QT interval analysis based on low-pass filtering of the signal and derivative filtering to extract waves. This method was resistant to variability in wave morphologies, but sensitive to noise and based on arbitrary thresholds. (Vullings, Verhaegen, and Verbruggen 1998) compared Laguna et al's method to a dynamic time wrapping-based technique. After piecewise linear approximation of the signal, the fiducial points were chosen by dynamic time wrapping. This led to a better error measure but suffered from higher standard deviation. Similarly to morphology-based methods, filtering techniques are intuitive and simple to implement. However, they require the definition of many arbitrary parameters, which values will vary from a dataset to another. In addition, the threshold criteria used are not always accurate and robust to noise.

- Bayesian inference techniques

From prior distributions and likelihood of the data, a Bayesian model estimates the posterior distribution of the unknown parameters. This is known as Bayesian inference. (Sayadi and Shamsollahi 2009) extended a Bayesian framework based on methods such as the extended Kalman filter. Each heartbeat was modelled as a combination of five Gaussian kernels, each of them corresponding to an ECG waveform. This method led to predictivity and sensitivity of 99.9%. A main advantage of this method is the robustness to generalization thanks to the adaptive nature of the extended Kalman filter and the absence of arbitrary thresholds. Later, (Lin et al 2010) combined a Bayesian algorithm with a Markov chain Monte Carlo method in order to perform both waveforms delineation and estimation at the same time. Based on posterior probabilities, a partially collapsed Gibbs

sampler was used to segmentate the ECG. This method provided reliable results compared to the benchmark, particularly for P and T wave detection and estimation, but suffered from high computational costs

- Classification and pattern recognition techniques

(Saini, Singh, and Khosla 2013) used a K-nearest neighbour algorithm for classification of ECG waveforms. The gradient of the signal was defined as feature vector and the training data was labelled as 1 if belonging to the wave and -1 if not. QRS complexes were first detected, then replaced by a baseline signal to repeat the same procedure with the T wave, and then the P waves. Simple, this method could adapt to different ECG morphologies. Its accuracy was computed from the wave durations obtained from the delineation marks. They obtained 100% accuracy for durations computed over 30 recordings. Other machine learning classifiers have been explored, such as support vector machines (SVM), studied by (Mehta and Lingayat 2008) for QRS detection. After preprocessing of the ECG signal, a feature vector formed of ten slope values extracted from a defined time window was defined. Several feature vectors were created by moving the time window along the signal. Training data was labelled as 1 if the feature vector lied in a complete QRS region, -1 if in a complete non-QRS region. This led to 99.7% of QRS detection accuracy. Later, (Mehta, Chaturvedi, and Lingayat 2009) proposed detection and delineation of the QRS via fuzzy C-means algorithm, an unsupervised clustering method. The gradient of the signal and the probability of belonging to the QRS were computed. From these probabilities, QRS entropies were computed and used as input vector to the clustering algorithm. Another approach to ECG delineation is the use of syntactic methods. For example, (Trahanias and Skordalakis 1990) used syntactic pattern grammar for ECG analysis. Later, (Koski 1996) mentioned their first approach to ECG delineation based on syntactic recognition of the waveforms. The principle was to decompose the signal with simpler pieces (called

primitives) and define relations between them to create more complex patterns. They showed that attribute grammars were efficient for ECG analysis, but required the choice of many parameters and struggled to introduce expert knowledge in the representation, to describe good sets of primitives.

Among this class of pattern recognition techniques, neural networks and hidden Markov models (HMM) seem to be popular as well. (Koski 1996) proposed the segmentation of the ECG by using broken line approximation to segmentate the beats and HMM to model the signal. The ECG signal was first approximated by continuous lines. The duration of the line segment and the amplitude of the starting point were then computed. These features were given to a HMM, composed of 15 to 30 states. Later, (De Lannoy et al. 2008) modelled the ECG by a 5-hidden state HMM with each state representing the waves of the ECG: P wave, baseline, QRS complex, T wave and baseline. The input signal was the original signal preprocessed by wavelet transform with stepwise optimization. They reached an accuracy of 94.7% over 10 recordings. The optimization of the wavelet transform increased the accuracy of 1.3%. The use of HMM enables to adapt to the specificities of different ECG configurations and learn statistical properties from the model. One limitation is the lack of duration constraints that may lead to delineation errors. Following the same approach, (Clavier et al. 2002) combined a basis functions transform preprocessing (using the Haar transform) for QRS detection and an HMM modelling for delineation of the P wave. Each activation time of the heart was represented by a hidden state of the HMM, resulting in 10 states. The change of states could be considered as the positions of the fiducial points. (Bystricky and Safer 2002) focused on the detection of the T wave end using a 2-layer perceptron with logisitic activation function. The network was fed with the values of the signal during different time windows. This method presented a good handling of the variability in T wave morphologies.

One of the main interests of pattern recognition techniques for ECG delineation is their ability to fit the data structure without arbitrary parameters or prior knowledge. However, they require a large amount of data to train the model and generalize properly, which can be an issue in case of small dataset of a specific disease and can explain why the accuracy results are not the highest with these techniques.

- Transformative methods

Exploiting the potential of frequency transformations, (Murthy and Niranjan 1992) early presented a method based on the discrete Fourier transform (DFT). In the DFT domain, the components of each waveforms do not overlap, allowing a simple filtering method to separate them. Many other studies then focused on the analysis of the signal in a different domain in which the ECG waves are clearly distinct, such as the discrete cosine transform. For example, following (Li 1995)'s work, a major paper presented the wavelet transform as an ECG delineator and was published by (Martínez et al. 2004). The wavelet transform used in this paper decomposed the signal according to different scales of energy. This allows the identification of ECG waveforms at different scales. For example, the QRS complex exhibits very little energy at scale 2^5 while P and T waves still have components at that scale. Peaks and boundaries are then detected by the analysis of zeros and maxima of the signal at these different scales. This method achieved 99.8% of predictivity and sensitivity for delineation of QRS, P and T waves, and introduced the multiscale approach, providing robustness to noise. Since this publication, several works have focused on implementing an online version of the delineator. (Boichat et al. 2009)'s implementation reached similar results compared to the offline algorithm and proposed its embedding on a wearable sensor. Combining the wavelet transform work and the constraints of online implementation, (Marco and Chiari 2011) proposed a method focusing on scales 2^2 and 2^3 of the wavelet transform in addition with locally defined thresholds. They removed the use

of non linear operators such as root mean square (RMS) and floating point algebra, reducing the computational costs to implement an online version of the algorithm. Wavelet-based delineators require the tuning of many parameters, usually defined empirically according to a particular database (Martínez et al. 2004). Some techniques have been proposed to optimize the choice of these parameters and improve the generalization properties of the delineator. For example, (Dumont, Hernández, and Carrault 2010) presented an evolutionary optimization process defining optimal values for the parameters in (Martínez et al. 2004) by stochastic method.

In summary, different approaches to ECG waveforms delineation have been investigated. Morphological and filtering methods mimic the behaviour of human clinicians, are simple and intuitive, but rely on the definition of arbitrary thresholds. Others use pattern recognition techniques and machine learning methods, providing good results by learning the statistical properties of the data because and avoiding the need of expert knowledge. Other approaches process the signal with basis functions transforms such as the wavelet transform, changing the domain of analysis of the signal and reducing the noise. The introduction of the multiscale also provided more accurate results. The accuracy results presented in these studies reach almost 99.9%. However, major challenges, such as generalization to other databases, remain. The lack of testing annotated databases limits the improvement of delineation techniques. As mentioned earlier, many ECG analysis methods rely on the accuracy of the delineation process, which currently acts as a bottleneck in these works. Novel approaches to ECG analysis, using more complex techniques such as convolutional neural network, may try and avoid the need of delineation by analysing the signal as a time series, without focusing on beat segmentation (Pourbabae, Roshtkhari and Khorasani 2017).

3.3 Machine learning techniques for classification of ECG signals

3.3.1 Objectives of state of the art methods and available databases

ECG classification studies may have different objectives, and investigate different types of heart beats or diseases. Some studies focus on heartbeat binary classification to distinguish between normal and abnormal beats (Koski 1996) (Gao et al. 2005), normal and premature ventricular beats (Inan, Giovangrandi, and Kovacs 2006a) (de Oliveira, Andreão, and Sarcinelli-Filho 2010) or normal and diseased beats (Ouyang, Yamauchi, and Ikeda 1998) (illustrated on Figure 3.1, Panel B). Other works follow the classification recommendations of labelling rules such as the AAMI guidelines (normal, ventricular, supraventricular, fusion of normal and ventricular, and unknown beats) (de Chazal and Reilly 2006). The MIT-BIH Arrhythmia database considers fifteen heartbeat classes, which have been also used in other studies (de Chazal, O'Dwyer, and Reilly 2004). Other works may focus on recordings classification rather than analysing single beats (Rahman et al. 2015) (Kannathal et al. 2003). Due to this variety of ECG label sets (summarized in Table 3.1), the classification objectives of the different studies may be different, making their performances harder to compare.

ECG classification can also be performed in recordings of different length, such as standard 12-lead ECGs (only 10 seconds) or longer recordings such as Holter ECGs (24 hours). Longer recordings allow the analysis of the ECG over time and the identification of time-dependent abnormalities, such as changes in the beat morphologies with time or changes in heart rate. In addition to the length of the recording, the number of ECG leads may differ and various methods are proposed to handle multi-channel data. Some studies focus only on single lead data (Mehta and Lingayat 2008). Others combine features computed over several different leads in a single feature vector (de Chazal, O'Dwyer, and Reilly 2004). Another approach is to combine the output of classifiers when applied over

the different leads following a voting approach. For example, in (Zhang and Luo 2014), the outputs of several support vector machine classifiers were merged based on the decision of each classifier. Other works such as (Llamedo, Khawaja, and Martínez 2012) combine the signal's multiple channels before performing feature extraction. They investigated several strategies such as combining wavelet transform features from all the leads, computing the wavelet transform from the vectocardiogram leads, or computing the features from the two first principal component of the ECG leads' principal component analysis.

The field of ECG classification is therefore very broad and studies may differ in terms of goals and data, but the general objective of this field remains the development of automatic pipelines to classify heartbeats of different nature and detect abnormal ECG patterns. The following paragraphs introduce examples of the technical solutions currently proposed in the literature to address this question.

Table 3.1: Summary table of the major databases used for classification of ECG signals.

Database	Type of recordings	Number of recordings	Annotations
MIT/BIH Arrhythmia * (Moody and Mark 2001)	- 30-minute excerpts - 2-channel ambulatory ECG - 360 Hz	48	Beat-by-beat annotations for each beat in each recording (~110,000 annotations)
QT database * (Laguna et al. 1997)	- 15-minute excerpts - 2-channel ECGs - 250 Hz	105	- reference beat annotations - segmentation of waveforms (for 30 to 100 normal beats per recording)
AHA ventricular arrhythmia *	- 2-channel excerpts of analog ambulatory ECG recordings - 250 Hz	80 for training – 75 for testing	- 8 classes of recordings (level of ventricular ectopy) - final 30min annotated beat-by-beat

INCART*	- 30 minutes recordings - 12 leads - 275 Hz	75	- 175,000 beat annotations - 10 classes pathological diagnosis
UCI Machine Learning: Arrhythmia dataset (Lichman 2013)	- 279 attributes (age, sex, height, waveforms description over 12 leads such as duration, amplitudes, areas)	452	16 arrhythmia classes labelled

* Datasets from PhysioBank (Goldberger et al. 2000) (www.physiobank.com), freely available, which gathers 60 databases (4TB) of physiological signals (cardiopulmonary, neural, and other biomedical signals) for healthy subjects and patients (sudden cardiac death, congestive heart failure, epilepsy, gait disorders, sleep apnea, aging)

3.3.2 Feature extraction and dimensionality reduction

Most of these machine learning classification techniques require the definition of a feature vector to describe the ECG beat and train a classifier. Each heartbeat is composed of multiple waves describing different events of the cardiac cycle (P-wave, QRS complex, T-wave). Morphological features, such as slopes, peaks, amplitudes (Bortolan, Degani, and Willems 1990) (de Chazal, O'Dwyer, and Reilly 2004), describe the shape of the ECG waveforms. They may be able to capture changes in the heart rhythm, such as sinus rhythm versus fibrillation, in which the complexes exhibit different morphologies. Some works focus on time interval features (as illustrated on Figure 3.1, Panel A) to characterize the dynamics of ECG phenomena such as QRS duration, QT interval or RR interval, also known as heart rate, defined as the number of regular complexes per unit of time (Nasiri et al. 2009) (Gharaviri et al. 2008) (Gao et al. 2005) (Ganeshkumar and YSKumaraswamy 2012) (Lagerholm et al. 2000). Other significant features include the coefficients of the Hermite transform or the discrete cosine transform (Inan, Giovangrandi, and Kovacs 2006) (Linh, Osowski, and Stodolski 2003) that aim at modelling the ECG beat instead of

extracting features from the raw data. In most studies, a combination of these features is used to characterize the ECG signal.

One of the challenges is that the ratio between the amount of available training data and the number of extracted features is too small, which may lead to overfitting. The number of features must therefore be reduced for good generalization and performance and two main techniques are usually used: dimensionality reduction and feature selection. Dimensionality reduction aims at reducing the size of the space in which the data is represented by computing a reduced number of dimensions that contain most of the information of the dataset. As many features can be extracted from ECG signals, dimensionality reduction algorithms are often performed before running the classifier. Examples of dimensionality reduction techniques include principal component analysis (linear or non-linear) (Nasiri et al. 2009) or linear discriminant analysis (de Chazal, O'Dwyer, and Reilly 2004). Feature selection selects only a small subset of the most significant features in the classification. For this purpose, some studies include an optimization step testing different feature combinations (via optimisation algorithms such as genetic algorithm or particle swarm optimization, or statistical distribution analysis such as Gini's index) and retrieving only the relevant features for further analysis (Doughty and Goldfeder 2013) (Nasiri et al. 2009) (Melgani and Bazi 2008). For example, (Mar et al. 2011) performed classification between normal, ventricular, premature ventricular and fusion beats based on the idea of (Llamedo and Martínez 2011) to use the sequential forward floating search feature selection procedure. This improved the classification accuracy of the multi-layer perceptron classifier from 79% using 71 features to 90% with only 9 features.

Feature quality and robustness is a challenge as poor quality features resulting from low quality delineators, filtering or approximations may lead to low performance and generalization properties despite powerful classification algorithms. Solutions to tackle this issue have been proposed such as in (Llamedo and Martínez 2011). In their study they introduced the use of robust surrogates of typical features, using for example directly the wavelet transform signal instead of the QRS width to reduce errors due to delineation.

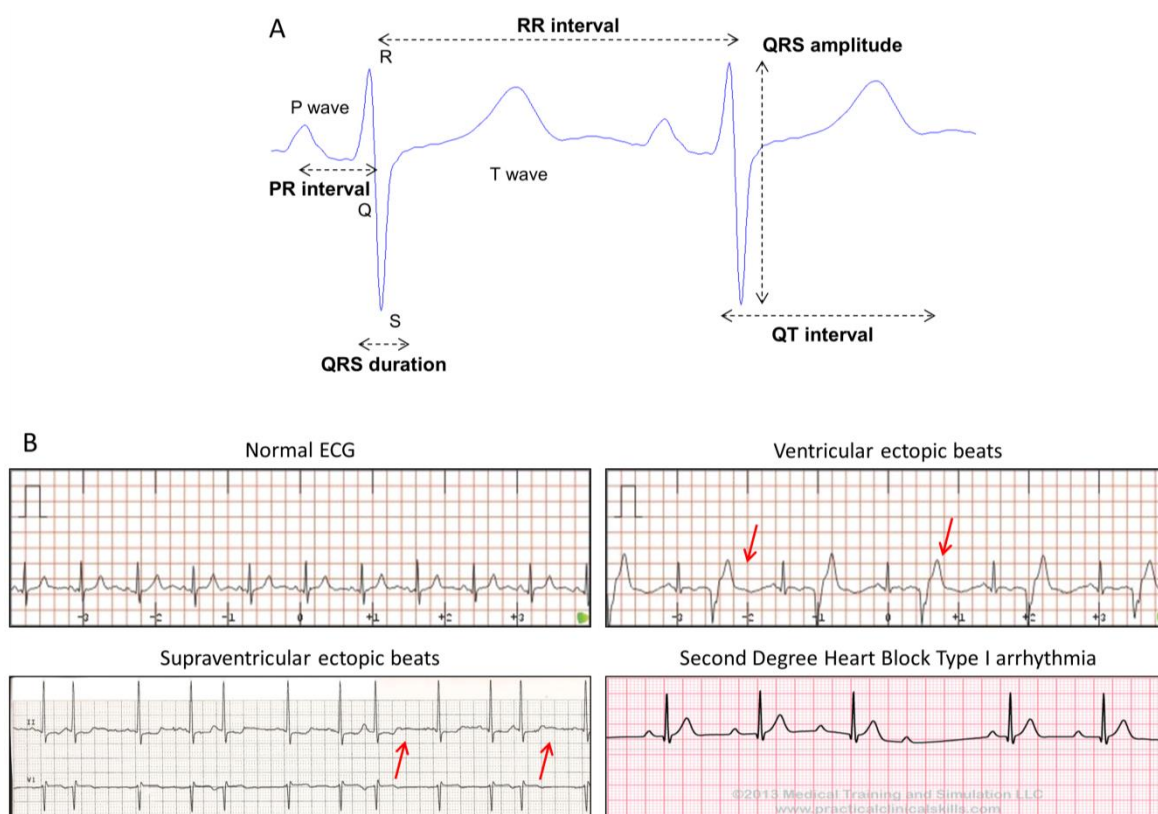


Figure 3.1: Example of data available for the analysis of ECG signals. Panel A: ECG waveforms (P, QRS and T waves) and standard features extracted from an ECG beat. Panel B: examples of different ECG waveforms: normal ECG, ventricular ectopic beats, supraventricular ectopic beats (from www.cardionetics.com) and second degree heart block arrhythmia (from www.practicalclinicalskills.com).

3.3.3 Main machine learning techniques for ECG classification

The main machine learning methods developed in the field of heartbeat classification and their performances are presented here and illustrated on Figure 3.2. (Appendix A provides a summary table). Many classification approaches follow the common strategy of training a classifier on a set of training data and validating this trained classifier on a set of testing data to evaluate their performances. Some studies propose more elaborate ways of using the classifiers that may lead to patient-specific techniques and allow expert assistance. For example, (de Chazal and Reilly 2006) proposed a patient adaptable system in which a classifier is trained to annotate the first beats of a recording, then checked by a clinical expert. This patient specific adaptation was then added to a global classifier. This technique showed to increase the performance of the classifier by 10% (accuracy of 94%) for the classification of heartbeats according to the AAMI classes. Another work by (Llamedo and Martínez 2012) also combined a global classifier (linear discriminant analysis) and a patient specific step (an expectation-maximization clustering algorithm). This led to an automatic patient-adaptable technique which can also incorporate the input of a cardiologist (semi-automatic) if clusters are difficult to classify.

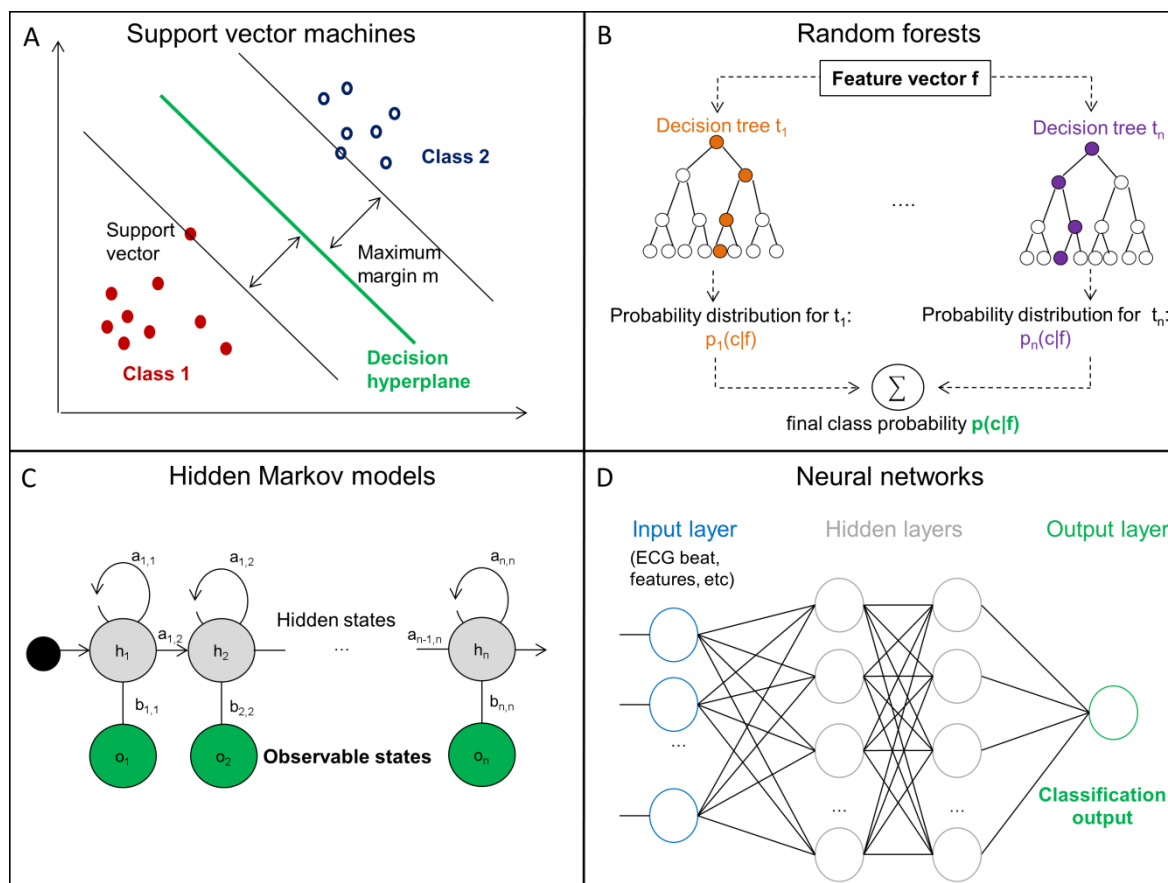


Figure 3.2: Main machine learning methods used for ECG classification. Panel A: Support vector machine binary classification by maximization of the margin m . Panel B: Random forest classification using n decision trees. Panel C: Hidden Markov model with n states, transition matrix $(a_{i,j})$ and emission matrix $(b_{i,j})$. Panel D: Neural network with two hidden layers.

3.3.3.1 Heartbeat classification

- Linear and quadratic discriminant

In their approach, (de Chazal, O'Dwyer, and Reilly 2004) implemented linear discriminants with weighted likelihood to classify 50,000 heartbeats in 5 classes (normal, ventricular ectopic, supraventricular ectopic, fusion of a normal and ventricular ectopic, or unknown beat type) from 22 recordings of the MIT-BIH arrhythmia database using QRS-based and time intervals features. The LD technique computes mean and covariance for the training data in order to maximise the likelihood. Posterior probabilities are then computed

to output the final classification labelling. Their study compared 12 different classifier configurations and reached a sensitivity of 75.9% (positive predictivity 38.5%, false positive rate 4.7%) for supraventricular ectopic beats, and a sensitivity of 77.7% (positive predictivity 81.9%, false positive rate 1.2%) for ventricular ectopic beats. Another work by (de Chazal and Reilly 2000) used both quadratic and linear discriminants with discrete wavelet transform and time domain features to classify 12-lead ECGs in a cohort of 500 cases with 7 classes. The range of accuracies was 60% to 75%. Similarly, another work by (Llamedo and Martínez 2011) used LD and quadratic discriminants to perform the classification of normal, ventricular and supraventricular beats. They used ~100,000 beats from the MIT-BIH Arrhythmia database split in two dataset for the evaluation of their model. Their method was validated over the MIT-BIH Supraventricular Arrhythmia (185,000 beats), and INCART (180,000 beats) databases. They extracted features from the RR interval and characterized the morphology of the ECG by discrete wavelet transform. They reached a global accuracy of 93% (84% sensitivity and 75% positive predictivity) on the MIT-BIH arrhythmia database and similar results on the other databases. Similar work was conducted by (Yeh, Wang, and Chiou 2009), using LD for classification between normal and 5 different classes of abnormal beats for arrhythmia diagnosis. Evaluated on the recordings from the MIT-BIH database, their method reported correct detection between 84.68% and 98.97% for the five classes studied.

- Support Vector Machines

Support vector machines (SVM) are a very popular class of machine learning algorithms because of their good classification and generalization properties. It is a supervised learning method which learns the best hyperplane to maximize the margin between two classes (Figure 3.2, Panel A). Usually, a kernel is applied to the inner product, which makes the problem equivalent to finding the best decision boundary separating classes in

an equivalent high-dimensionality space resulting from non-linear transformation of the feature vector. This decision boundary is then used to classify unknown testing data (Cortes and Vapnik 1995). SVM application to ECG classification and its optimization have been largely studied. For example, (Übeyli 2007) applied SVM with error output correction code to classify heartbeats from 4 classes (normal, congestive heart failure, ventricular tachyarrhythmia and atrial fibrillation) from the Physionet database. Discrete wavelet transform was used to preprocess the data and extract features. They reached an accuracy of 98.61% (sensitivity of 98.89%) on a testing set of 360 beats. SVM optimization techniques have been investigated to improve the choice of features and parameters, reduce overfitting and speed up the classification. For example, (Melgani and Bazi 2008) performed arrhythmia classification with SVM optimized by particle swarm optimization (PSO). PSO reduced the number of features from 303 to 46 and reached an overall accuracy of 89.72% in detecting 5 heartbeat classes (normal, atrial premature beat, ventricular premature beat, right bundle branch block, left bundle branch block and paced beat) from the MIT-BIH database. Another work by (Asl, Setarehdan, and Mohebbi 2008) focused on the classification of heartbeats in 6 classes for arrhythmia detection (normal, premature ventricular contraction, atrial fibrillation, sick sinus syndrome, ventricular fibrillation and 2° heart block). They used 1367 ECG segments from the MIT-BIH database and designed feature vectors based on heart rate variability. They reached 99.16% of accuracy with discriminant analysis feature reduction (15 to 5 features). In another study, (Nasiri et al. 2009) used principal component analysis and genetic algorithm to determine the best parameters to tune the SVM algorithm and to perform feature reduction to remove the features that may lower the accuracy. They reached an accuracy of 93.46% for classifying between normal, right and left bundle branch block, and paced beat on the MIT-BIH arrhythmia database with genetic algorithm. Methods based on SVM and optimized SVM therefore reach high accuracies ranging from 90% to 99% in multi-class

classification. A drawback of SVM classification for clinical applications is its lack of interpretability in order to evaluate each feature's influence and extract relevant discriminant biomarkers. Similar applications of SVM algorithms to ECG beat classification were implemented by (Li, Rajagopalan, and Clifford 2014), (Rabee and Barhumi 2012), or (Mehta and Lingayat 2008)

- Random forests

The random forest method (Breiman 2001) is an ensemble learning technique combining the classification outputs of the decision trees that compose the forest (Figure 2, Panel B). A useful property of random forests is their ability to rank the variables according to their importance in the classification and therefore allow feature selection to avoid overfitting. (Ganeshkumar and YSKumaraswamy 2012) investigated arrhythmia detection by identifying 6 heartbeat classes (normal, premature ventricular contraction, paced, atrial premature beat, and left and right bundle branch block) using random forest. They reached an accuracy of 92.16% on 150 beats extracted from the MIT-BIH database with 30 trees but their method was not validated over an independent testing set. In a recent work, (Rahman et al. 2015b) compared the heartbeats of 12-lead ECG from 1000 HCM patients to a group of ischemic and non ischemic cardiomyopathy patients. Their implementation used a random forest algorithm with 500 trees and 264 features (after feature selection) obtained from ECG waveforms morphology and duration. They reached an accuracy of 89% validated with 5-fold cross validation. Similar random forest classifiers for ECG classification were developed by (Sathish and Vimal 2010), or (Emanet 2009).

- Bayesian networks

Bayesian networks are probabilistic graphical models representing variables and their probabilistic relationships. Following the successful application of Bayesian frameworks

techniques to ECG segmentation (Sayadi and Shamsollahi 2009) and denoising (Sameni et al. 2007), (de Oliveira, Andreão, and Sarcinelli-Filho 2010) introduced the first approach of ECG beat classification using Bayesian networks. Their work focused on premature ventricular beat (PVC) detection by the implementation of a Bayesian network framework using channel fusion. They reached a sensitivity of 99.69% (positive predictivity 98.79%). Validation of their method was performed by splitting the 82,509 labeled heart beats from the QT database with 75% as training set (58,744 beats) and 25% (23,765 beats) as test set. Similarly, (Oster et al. 2015) developed a model-based filtering approach using switching Kalman filters to classify ventricular beats. In addition to the classification performance, their approach presented the ability to detect unknown beat morphologies. They reached an accuracy of 97.3% (positive predictivity 99.96%) for ventricular beats classification on the validation dataset composed of 50,000 beats from the MIT-BIH database and 175,000 beats from the INCART database.

A simple type of Bayesian networks are Hidden Markov models (HMM) which represent the system of study as a Markov process, composed of hidden states and transition probabilities where the sequence of hidden states is estimated from the observations (Rabiner 1989) (Figure 3.2, Panel C). HMM do not require prior knowledge and the estimation of the parameters is automated. This technique is popular for sequence modelling and temporal pattern analysis. Hence, several studies focus on the application of HMM to ECG modelling and prediction. Early, (Coast et al. 1990) investigated the use of HMM for classification and modelling of normal and ventricular ectopic beats (VEB) from the American Heart Association database (six 30-min ECG recordings). Each ECG waveform and duration was represented by a HMM state. Their method reached a sensitivity of 97.25% (specificity of 85.67% over 799 VEB beats). Similarly, (Koski 1996) performed ECG segmentation and classification between normal and premature ventricular

beats (PVB) signals using HMM. They estimated the ECG signal by broken line approximation. The classifier was composed of 20 to 30 states. They reported an accuracy of 100% on the very small testing set used (4 testing beats). Later, (Andreão, Dorizzi, and Boudy 2006) performed modelling and classification of premature ventricular contraction (PVC) beats using 59 recordings from the QT database. They combined an HMM and a rule based system for beats identification (based on the beat prematurity and enlarged QRS morphology criteria). They reached a sensitivity of 99.79% (specificity of 99.96%) for PVB detection over the testing set. Most of these techniques based on HMM also integrate and estimate the parameters of the delineation of ECG waveforms in the model, avoiding the need for a separate segmentation step.

- Neural networks

In the last fifteen years, neural networks (Bishop 1995) have been very popular in ECG classification. Inspired by biological processes, the interconnected neurons of the system learn the structure of the data from training examples (Figure 3.2, Panel D). Neural networks are powerful for their ability to detect patterns and extract data structure without expert knowledge. An example of neural network application to ECG heartbeat classification was presented by (Issac Niwas, Shantha Selva Kumari, and Sadasivam 2005). They implemented an artificial neural network trained on feature vectors composed of heartbeat intervals and spectral entropy. They reached an accuracy of 99.02% in classification of testing heartbeats in 10 classes from the MIT-BIH database. Similarly, (Inan, Giovangrandi, and Kovacs 2006) studied the detection of premature ventricular contractions (PVC) beats. They used morphological features derived from the wavelet transform and time intervals features as input to a feed-forward multi-layer perceptron neural network with a single hidden layer. They obtained 96.82% of accuracy for PVC beat classification on 22 ECG recordings from the MIT BIH database.

In another study, (Übeyli 2009) used a more complex neural network architecture, recurrent neural networks (RNN), which differ from standard ANN by the presence of direct cycles in the neurons architecture. This allows the representation of dynamic temporal processes. They trained the RNN with Levenberg–Marquardt training algorithm and eigenvector based features. They tackled the classification of heartbeats in 4 classes on 720 heart beats from the Physionet database and reached an accuracy of 98.06%. Another work by (Lagerholm et al. 2000) presented a clustering method for arrhythmia over 48 recordings of the MIT-BIH database. They preprocessed the signal with the Hermite transform which provided a better robustness to noise. The Hermite transform and the RR interval defined the feature vector. They implemented self organizing networks (SON) to perform clustering in 25 groups and reach 1.5% of misclassification. SON also conserves the neighbouring structure of the data, representing similar clusters physically close on the map, and therefore facilitating the interpretation of the results by the clinicians. Similarly, (Linh, Osowski, and Stodolski 2003) used the Hermite transform of QRS complexes coupled to a TSK fuzzy network to classify 6 types of heart rhythms (premature ventricular ectopic, left bundle branch block, right bundle branch block, atrial premature beat, ventricular flutter wave, and ventricular escape beat) for arrhythmia. They reached 96% of accuracy in detecting these heartbeat classes, although no validation dataset is described in the paper. In addition, their method showed to be less sensitive to morphological variations of the ECG and handled heartbeat variability. Another method implementing fuzzy techniques was developed by (Ozbay, Ceylan, and Karlik 2006). Their fuzzy clustering neural network architecture achieved 99.9% of accuracy in arrhythmia detection (classifying normal beat, sinus bradycardia, ventricular tachycardia, sinus arrhythmia, atrial premature contraction, paced beat, right and left bundle branch block, atrial fibrillation and atrial flutter) on 5342 segments from the MIT/BIH database. Similar works

on neural networks were conducted by (Jadhav, Nalbalwar, and Ghatol 2010), (Meau et al. 2006), (Dokur and Ölmez 2001) or (Das et al. 2014).

Most of the studies based on neural networks perform well, reaching 95-99% of accuracy in classifying heartbeats in five or six different classes (mostly for arrhythmia), but these results remain similar to the ones obtained with other techniques such as SVM algorithm. Neural networks have the ability to take raw ECG recordings as input and learn directly from the signal.

3.3.3.2 Patient classification and disease diagnosis

The methods presented above can also be used in a different context such as patients' classification based on features of a whole segment of ECG signal. Indeed, instead of classifying different classes of heartbeats, some studies focus on diagnosing patients from their ECG and aim at performing recording classification. These works find useful clinical application in risk stratification of diseases where the identification of high risk patients is crucial, or in the monitoring of diseases where abnormal events may strike at unpredicted times. Several works approach patient diagnosis by identifying abnormal excerpts or events in the recordings. For example, (Faganeli and Jager 2010) developed a decision-tree based classification to distinguish between ST-segment changes due to transient ischaemia episodes and non-ischemic ST-segment deviation, due to changes in heart rate, which make automatic identification of the true ischemic episodes more challenging. They evaluated their performance by bootstrap method and reached a sensitivity of 98.1% and specificity of 85.2% with features from heart rate and Legendre orthonormal polynomial coefficients. Another work by (Bailón et al. 2003) aimed at diagnosing coronary heart disease based on exercise ECG indexes. They developed an automated method to extract repolarization, depolarization and heart rate variability (HRV) indexes from noisy exercise recordings. A multivariate discriminant analysis then classified patients into two classes:

ischemic and low risk patients based on independent set of indexes. HRV indexes provided the best results with a sensitivity of 94% and specificity of 92% to classify 65 ischemic and 40 low risk cases. Another disease condition investigated is Brugada syndrome, which leads to a high risk of sudden cardiac death because of episodes of ventricular fibrillation (VF) in patients with no structural heart disease (Brugada and Brugada 1992). (Kawazoe et al. 2016) investigated risk of VF in patients with Brugada syndrome thanks to a logistic regression model, cross-validated over a database of 143 patients (35 with VF, 108 without). Syncope episodes, r–J interval in lead V1, QRS duration in lead V6, and Tpeak–Tend dispersion were the best discriminating features identified by logistic regression. They led to a sensitivity of 97.1% and a specificity of 63.0% by leave-one-out cross validation.

The study by (Rahman et al. 2015), mentioned earlier, presented a patient classifier to identify patients affected by hypertrophic cardiomyopathy (HCM) from controls, based on standard 12-lead ECG. They classified a patient as HCM if the majority of the heartbeats in the recording were labelled as HCM by the classifier. Their dataset was composed of 221 HCM patients and 541 control (non HCM patients, but with implemented cardiac defibrillator). 264 standard ECG features, such as time intervals and waveforms amplitude, were extracted by feature selection (information gain criterion), and used to perform SVM and random forest classifications with 5-fold cross validation. They reached similar performance results with the two classifiers, with a precision of 0.84 (0.89 sensitivity, 0.93 specificity).

Atrial fibrillation (AF) is a condition investigated by many studies as its irregular and repetitive episodes can lead to heart failure, stroke and double the risk of mortality (Alonso et al. 2009). In (Pourbabae and Lucas 2008), time interval and morphological features were extracted from the three main ECG waveforms (P wave, QRS and T wave) of 25

patients with paroxysmal AF episodes and 25 healthy subjects from the 2001 Computers in Cardiology Challenge database. The multi-layer perceptron implemented classified accurately 87% of the testing dataset, and suggested the importance of QRS-based features in the classification after testing the influence of each set of features. Similarly, (Colloca et al. 2013) performed detection of atrial fibrillation episodes via support vector machines, evaluated on the MIT-BIH database with a specificity of 99.72% on the testing set. In another study by (Asgari, Mehrnia, and Moussavi 2015), stationary wavelet transform and support vector machine were designed to identify AF events with sensitivity and specificity of 97.0% and 97.1% respectively.

(Acharya et al. 2003) were interested in diagnosing several groups of cardiac disorders based on ECG signals: ischemic or dilated cardiomyopathy, complete heart block, sick sinus syndrome, atrial fibrillation (AF) or ectopics, and normal beats. They developed an artificial neural network (ANN) coupled with a fuzzy equivalence relation and reached an accuracy of 85-95% using four heart rate variability features. In another work, (Zheng et al. 2014) applied convolutional neural network (CNN) to time-series classification using a multi-channel technique. They aimed at classifying congestive heart failure from 2-lead ECG of 15 subjects by extracting sub-sequences of the signal. The model learned features from the ECG throughout the layers of the CNN and reached an accuracy of 94.7% for congestive heart failure detection on this dataset.

ECG analysis can also be used for long term monitoring of patients in clinical care, to detect abnormal rhythmic events that may occur suddenly. For example in (Kannathal et al. 2003), three degrees of disease severity (normal, abnormal, life threatening) were predicted from the ECG of patients in intensive care unit (ICU) using artificial neural networks (600 training patients, 200 testing patients). This reached an accuracy of 99% with RBF (radial basis function) networks. Similarly in (Zhang et al. 2016), a combination of genetic

algorithm and SVM was used to detect false critical arrhythmia alarms in ICUs with a true positive rate of 93% (true negative rate of 94%).

Patient classification and diagnosis differs from heartbeat classification by automatically analysing whole recordings. These studies may be able to provide insight in the structure of diseased population thanks to clustering techniques, highlight which biomarkers are significant to distinguish two disorders, and provide automatic diagnosis (or semi-automatic, taking into account expert input). They also allow the analysis of ECG changes over time, which may be a tedious task to perform visually.

3.3.4 Discussion

The electrocardiogram is a cheap, non-invasive and simple modality, widely used in clinical practice. As a recording of the body surface electrical activity, it provides information about heart rhythm abnormalities and helps detect diseases. However, visual inspection of the ECG provides discrete clinically interpreted features which cannot objectively capture the diversity of ECG abnormalities and morphologies. This is why computational methods are required, as they can make sense of multivariate complex datasets and detect differences that might be challenging for the human eye. Indeed, as mentioned in the sections above, machine learning techniques have been used to classify specific types of heartbeats for arrhythmia detection, such as supraventricular and ventricular ectopic beats, or second degree heart blocks. They also helped in patient classification and diagnosis, used for atrial fibrillation, cardiomyopathies or intensive care unit monitoring. These studies show the ability of machine learning and classification techniques to analyse the ECG and uncover hidden patterns in the data, to highlights groups of similar ECG morphology. This is key to achieve the first goal of this thesis, presented in Chapters 4 and 5, to identify different ECG phenotypes in HCM.

However, analysing ECG data presents many challenges. First, most large clinical studies still record ECG in paper print-outs, requiring manual digitization before computational techniques can be applied. The habit of digital ECG acquisition is still to be implanted in clinical practice in some countries. In addition, many ECG databases are not publicly available, gather low numbers of patients, and require extensive signal preprocessing techniques to denoise the recordings for computational analysis. As a consequence, most of the methods presented here focus on large, publicly available databases of analog ECG recordings and many methods are trained and tested on the same databases, despite ensuring separated training and testing sets. This may limit the generalization of these techniques to different clinical databases or other cardiac disorders. To address these challenges of ECG analysis, alternative datasets to the standard ECG recordings emerge. For example, the body surface mapping (BSM) technique samples multiple points around the chest to provide a more detailed mapping of the body surface electrical activity than the standard 12-lead ECG. It has been shown to be more accurate than the ECG in detecting transient episodes like myocardial ischaemia (Robinson and Curzen 2009). However, it is less commonly used and accessible than a standard ECG. Intracardiac recordings also present an alternative to measure the electrical activity in the heart, but they are invasive, and could therefore not be obtained for large numbers of patients, which would show little interest for computational investigations.

The gold standard for electrocardiographic abnormality detection remains the expert annotations, and the development of larger, more diverse and annotated datasets would help improve the performances for clinical use. However, these techniques provide automaticity and consistency in the analysis of large datasets, which are more and more needed when manual classification cannot be performed. More importantly, the benefit of computer techniques goes beyond these classification tasks in which computational methods try to reproduce the expert judgement. As implemented in this thesis, they can

help uncover new knowledge and discover new biomarkers by unravelling structures in the data that were unknown before.

The recent interest in computational ECG analysis is twofold. For the computational field, the challenge and diversity offered by ECG recordings provide a rich environment to develop new methods. From the clinical perspective, these methods provide a new horizon on how ECGs can be analysed, biomarkers, discovered and diseases, predicted. Progress in this field is recent, which explains why several studies investigate the behaviours of different machine learning methods on ECG data. The fact that the ECG signal can be analysed by so many techniques actually provides a wide range of options depending on the goal of the study. In Chapters 4 and 5, interpretable techniques are favoured because of the requirement of the research question to describe and understand the phenotypes identified. These computational methods were introduced in the clinical environment recently, and even if growing interest has been shown from the clinical field (Obermeyer and Emanuel 2016) (Ambale-Venkatesh et al. 2017), changing clinical practice takes time and the actual clinical impact of these computational methods cannot be yet entirely measured. However, they definitely provide interesting lines of research and future directions for fruitful discussions.

3.4 Personalized whole-body computer simulations

Complex multiscale computer models can be used in combination with ECG analysis techniques to unravel the influence of ionic, tissue and whole-organ mechanisms on the ECG. Such approach is implemented in Chapter 6, exploring the structural and electrophysiological mechanisms that may explain the ECG abnormalities identified in previous chapters with classification techniques. The aim of this section is to review the

different computational and mathematical methods developed for modelling of the ECG and motivate their use in this thesis.

3.4.1 Modelling of cardiac anatomy

Geometrical approaches to ventricular models have been developed, based on a simple definition of the heart anatomy with geometrical shapes. In 2013, (Sovilj et al. 2013) presented a simplified 3D model of a torso-embedded heart obtained by combining different geometries, associated with a bidomain model. They simulated myocardial infarctions at various locations and were able to recover ECG changes in the ST segment. Another work by (Bacharova et al. 2012) used computer simulations to untangle the mechanisms leading to different ECG patterns in the context of left ventricular (LV) hypertrophy, especially increased QRS amplitude, and to show the influence of LV mass on QRS changes. However, the simple geometry of the model cannot handle complex physiological systems with changes in anatomy. In a more recent work, (Bacharova et al. 2016) investigated the effect of slow ventricular activation on the QRS complex and showed how alteration of electrical properties may mimic ECG morphologies associated with anatomical abnormalities. These simplified approaches have the benefit of speed and low memory requirements, but they do not take into account the realistic anatomy of the heart which may limit their use in understanding physiological cardiac mechanisms.

Recent developments in cardiac computer simulations therefore introduced the use of realistic geometries, obtained from clinical cardiac images to build the mesh anatomy (Figure 3.3). These approaches improve the clinical accuracy of the model and allow the personalization of the simulations in order to address the challenges of human variability. For example, (Vadakkumpadan et al. 2009) published a methodology to construct detailed computer models from cardiac magnetic resonance images and diffusion tensor imaging data. This methodology was then applied to investigate arrhythmia and defibrillation in

rabbit hearts and the effect of shocks and re-entries. Another work by (Cardone-Noott et al. 2016) investigated the effect of changes in conductivities on the ECG morphology and how the variability in the activation sequence relates to changes in QRS biomarkers, by designing a 3D simulation pipeline with a human volumetric mesh, based on the cellular (ten Tusscher and Panfilov 2006) model. In (Chen et al. 2011), they used an MRI-based computer simulation model from patients with acute myocardial infarction to show that instabilities in QT interval may predict ventricular tachycardia onset. Similarly, (Wilhelms, Dössel, and Seemann 2011) applied realistic detailed 3D computer simulations to better understand the underlying mechanisms leading to shifts in the ECG ST segment in acute cardiac ischaemia.

The simulation of the ECG through 3D computer simulations also finds applications in cardiotoxicity and understanding the effect of drugs on the ECG (Benson, Al-Owais, and Holden 2011) (Zemzemi et al. 2013). (Zemzemi et al. 2013) investigated the influence of drugs on the ECG signal thanks to 3D computer simulations, allowing drug testing in silico rather than in vivo. They simulated channel conductance block for the hERG and fast sodium channels and observed QT prolongation for 50% hERG block (6%) as well as QRS and QT prolongation for 50% fast sodium current block (12% and 5% respectively). In addition to body surface, their pipeline also allowed the investigation of changes at the cellular level through the APD, showing that the mechanisms behind these ECG changes are different for potassium and sodium block. Similarly, (Sebastian et al. 2010) investigated the effect of dofetilide, a drug affecting the IKr current on a patient-specific model simulating the pseudo-ECG. They observed a prolongation of the QT interval of 100ms with a total IKr block, in line with a prolongation of the APD90 in cardiac cells.

In this thesis, computer simulations are used to explore the effect of structure and electrophysiology, such as anatomy, ionic remodelling or fibre disarray, on specific HCM

patients. Realistic personalized models are therefore needed to model the uniqueness of each patient.

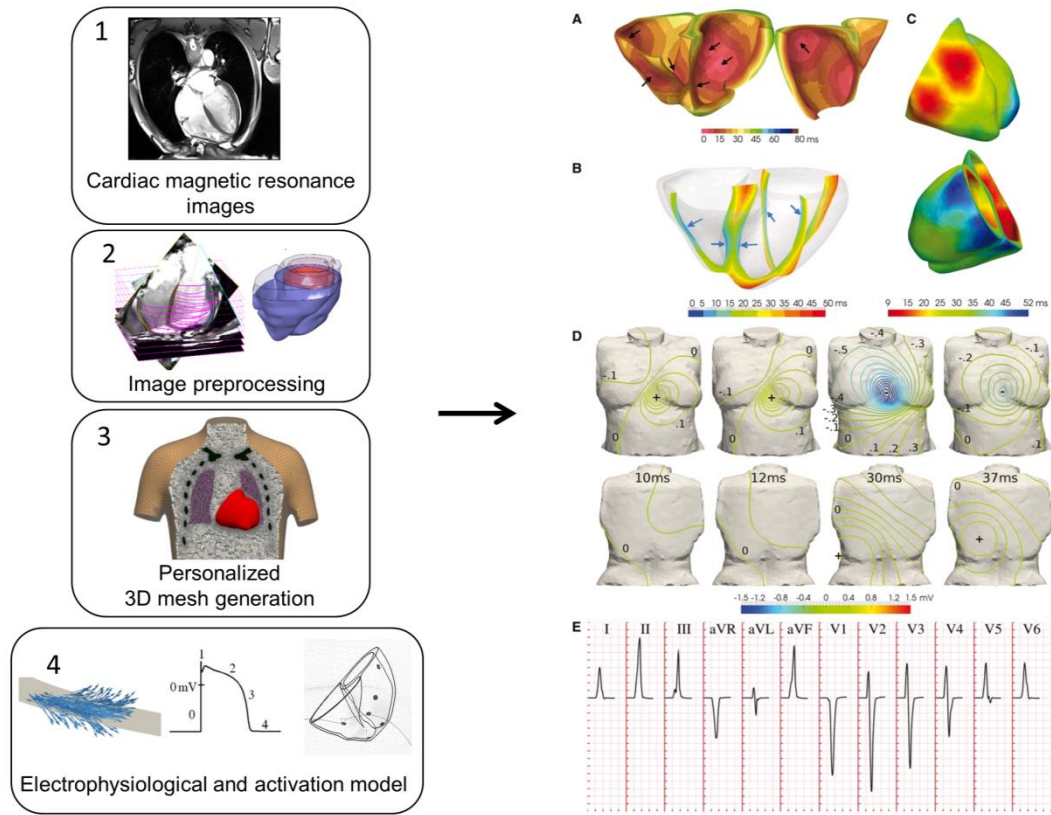


Figure 3.3: Computational pipeline for ECG simulation from magnetic resonance images through 3D meshes and cellular electrophysiological models (left) and the obtained simulation of the cardiac electrical activity (right – from (Cardone-Noott et al. 2016)). Personalized cardiac magnetic resonance images (1) are segmented and preprocessed (2). From this information, surface and volumetric meshes are generated (3) and an electrophysiological model defining the electrical activity in the cells is implemented (4). The obtained simulated electrical conduction (right panel) can then be investigated.

3.4.2 Modelling of the cardiac electrical propagation

Two different physiological approaches have been proposed to model the propagation of the electrical signal in cardiac tissue: discrete and continuous representations. Discrete models define the structure of each cardiac cell individually, as well as the connections between them. They are straightforward to implement and allow the explicit definition of specific elements of cardiac architecture such as gap junctions. This allows more realistic

behaviour in the case of cell decoupling (Leon and Roberge 1991) (Diaz, Rudy, and Plonsey 1983). However, for large scale simulations and whole-organ modelling, as the ones implemented in Chapter 6 of this thesis, they are computationally too expensive and continuous approximations, considering larger pieces of tissue rather than individual cells, are required. When modelling the propagation of the cardiac excitation through the cardiomyocytes, two distinct domains, connected via the cell membrane, may be defined: intracellular and extracellular space, exhibiting different properties, especially unequal anisotropy ratios (ratio of conductivities parallel and perpendicular to the fibres). Under the bidomain model, the two domains are defined, and different electrical potentials (φ_i and φ_e) are defined for each of them, following partial differential equations.

The transmembrane potential V_m is defined as:

$$V_m = \varphi_i - \varphi_e$$

In each domain, the electrical potential is uniform and continuous and defined as:

$$-\nabla \cdot (\sigma_e \nabla \varphi_e) = \beta (C_m \frac{dV_m}{dt} + I_{ion})$$

$$\nabla \cdot (\sigma_i \nabla \varphi_i) = \beta (C_m \frac{dV_m}{dt} + I_{ion})$$

with σ_i and σ_e the intracellular and extracellular conductivities, β the membrane to cell volume ratio (in m^{-1}), I_{ion} the ionic current over the membrane (in Am^{-2}), V_m the transmembrane potential, and C_m the cell membrane capacitance (in Fm^{-2}).

When equal anisotropy ratios in both domains are considered (i.e $\sigma_i = a\sigma_e, \forall a$), or when the extracellular potentials are considered as negligible, the bidomain equations can be simplified to the monodomain equations:

$$\nabla \cdot (\sigma \nabla V_m) = \beta (C_m \frac{dV_m}{dt} + I_{ion})$$

with σ the intracellular conductivity.

In addition to large scale reasons, the bidomain equations are particularly important when the extracellular space matters, for example when investigation defibrillation shocks (Trayanova, Plank, and Rodríguez 2006).

In this thesis, we use the finite element method (FEM) to solve these bidomain equations as implemented on Chaste (Pitt-Francis et al. 2009). This numerical technique approximates the solution of the differential equations by dividing the entire space in smaller elements. The system composed of the simple equations modelling the smaller finite elements can then be solved with methods minimizing a defined error function (Bernabeu et al. 2014). This method is particularly useful for cardiac simulations as it is able to handle complicated unstructured meshes as the realistic ones obtained from cardiac images.

Phenomenological approaches have also been proposed to model cardiac propagation. These models focus on characterizing the relationships between variables with simple mathematical equations, rather than describing the variables themselves, reducing the complexity of the problem. Studies have for example investigated the modelling and simulation of cardiac electrical propagation using fast algorithms based on graph based theory (Dam, Oostendorp, and Oosterom 2009) (Wallman, Smith, and Rodriguez 2011). These techniques may facilitate both simulations of the ECG and parameter inference from the ECG, from clinical datasets, by taking into account parameter uncertainty such as in (Wallman, Smith, and Rodriguez 2014) and (Konukoglu et al. 2011).

3.4.3 Modelling of cardiac heterogeneities

Cardiac tissue is not homogeneous, and three main types of electrophysiological heterogeneities have been reported through experimental measurements. First, transmural heterogeneities describe the variation in cellular properties across the membrane. Indeed, cardiomyocytes near the epicardial surface of the heart have been shown to exhibit different ion channel properties, conductivities and behaviours than endocardial cells. More specifically, in human, the endocardial tissue shows longer action potential duration (APD) than epicardial tissue (Szabó et al. 2005) (Näbauer et al. 1996), and the transient outward current I_{to} exhibits higher density in epicardial cells (Wettwer et al. 1994). In addition, based on recent experimental measurements (Gaborit et al. 2007) (Soltysinska et al. 2009), the (O'Hara et al. 2011) human ventricular model also modifies conductances of potassium (I_{Kr} , I_{K1} , I_{Kb}), late sodium (I_{NaL}), and calcium (I_{CaL}) channels, as well as the sodium-potassium pump, and calcium transport, to model the transmural heterogeneities and the various cell types. A third group of cells, called M cells, has also been reported (Antzelevitch et al. 1991), but there is substantial debate on these, as studies show they are observed only in wedge preparations rather than human (Janse et al. 2012). The second type of reported heterogeneity is interventricular, as, action potential duration has been shown by in vivo recordings to be longer in the human left ventricle than in the right ventricle (Bueno-Orovio et al. 2012). Finally, apico-basal heterogeneities have also been found, although with contradictory results. Indeed, most studies report longer APD in the base than the apex in human (Szentadrassy et al. 2005), but shorter APD in the base have been measured in other species such as dogs (Burgess et al. 1972) or rabbit (Cheng et al. 1999).

Realistic computer simulations therefore aim at taking these electrophysiological heterogeneities into account when modelling the activation and repolarization dynamics in

the heart. The use of these heterogeneities in this thesis is discussed in Chapter 6, when the 3D model is defined. Several studies carried out experiments to investigate the influence of modelling these heterogeneities on the simulated ECG, especially the T wave. For example, (Weiss et al. 2007) explored the role of these heterogeneities on the ECG with biventricular-torso simulations and concluded that without interventricular heterogeneities, transmural gradients were enough to generate a realistic positive T wave. In the case of interventricular heterogeneities, both apico-basal and transmural heterogeneities were needed for a positive T wave. The activation, and in turn the QRS complex, were not affected by these heterogeneity of the ionic currents. (Keller et al. 2012) similarly reported that modelling the apico-basal gradient with an increase of I_{Ks} by 50% in the apex was the most successful in simulating realistic, positive T wave morphologies.

In addition, structural heterogeneities are also present in the heart. Modelling the arrangement of cardiomyocytes in sheets, local conductivities in the heart can be defined as orthogonal vectors describing the conductivity along the fibre, sheet, and intersheet. (Caldwell et al. 2009) suggest that orthotropic conductivities are important to model accurately the electrical propagation in the heart. Fibre direction is usually defined following rule-based methods, such as (Legrice, Hunter, and Smaill 1997) (Pravdin et al. 2013) (Streeter et al. 1969), or according to the results of diffusion tensor imaging (DTI) data. The initial framework of the simulations presented in Chapter 6 considers the fibre orientations defined by (Streeter et al. 1969)'s rule based method. Following a more personalized approach, the information contained in the DTI data is then analysed to model structural abnormalities related to each patient, such as fibre disarray.

4 EXTRACTION OF ECG PHENOTYPES IN HCM BASED ON QRS MORPHOLOGICAL BIOMARKERS

The work in this chapter is part of: “Distinct ECG phenotypes identified in hypertrophic cardiomyopathy using machine learning associate with arrhythmic risk markers” by Lyon et al., 2018, Frontiers in Physiology, DOI: 10.3389/fphys.2018.00213 and “Extraction of morphological QRS-based biomarkers in hypertrophic cardiomyopathy for risk stratification using L1 regularized logistic regression” by Lyon et al., 2015, Computing in Cardiology.

4.1 Introduction

Chapter 2 introduced hypertrophic cardiomyopathy (HCM) and reviewed the challenges that remain for HCM risk stratification and the identification of high risk patients. In this chapter, we present the computational methods developed to identify different patients' subgroups in HCM based on biomarkers extracted from the QRS complex of Holter ECG

recordings. We then compare the identified subgroups in terms of clinical markers obtained from cardiac magnetic resonance images (CMR) as well as traditional risk factors. The database collection and the clinical work described in this chapter was performed by Dr Rina Ariga (Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, UK).

4.2 Motivation

As mentioned in Chapter 2, hypertrophic cardiomyopathy (HCM) remains a common yet challenging genetic heart muscle disease due to its heterogeneous clinical course. Ventricular arrhythmias in HCM are a major cause of sudden cardiac death (SCD) in young people (Maron 2002) (Maron et al. 2000), and the accurate identification of high risk patients is a clinical priority since implantable cardioverter-defibrillators (ICD) can successfully treat ventricular arrhythmias triggering SCD. Previous studies have attempted to assess the electrophysiological signature of HCM by visually inspecting paper ECGs (Lakdawala et al. 2011) (Savage et al. 1978), but the approach failed to produce reliable stratification (Sherrid et al. 2009) (Maron, Roberts, and Epstein 1982). This may be due to the limitations in the methodology and therefore, more sophisticated approaches and new knowledge are required. In HCM, five conventional risk factors for SCD are currently in use: non-sustained ventricular tachycardia (NSVT), unexplained syncope, family history of SCD, massive left ventricular hypertrophy (LVH) and abnormal exercise blood pressure response, and provide clinical utility (Christiaans et al. 2010) (Maron et al. 2007). More recently, the prospectively validated HCM Risk-SCD prediction model (O'Mahony et al. 2014) has been recommended by the 2014 ESC guidelines (Elliott et al. 2014), and shown to perform better than conventional risk factors, although also with limitations (Maron et al. 2015). The resting 12-lead ECG is not used for risk stratification due to the lack of insight mentioned above.

In this chapter, we hypothesize that HCM structural abnormalities, namely hypertrophy (Spirito et al. 2000), myocyte fibre disarray (Varnava et al. 2001) and fibrosis (Adabag et al. 2008), may create a pro-arrhythmic substrate to different extents in specific patients and affect in turn the QRS waveform in the ECG, which represent the depolarization (activation) of the cardiac electrical propagation. We therefore aim at identifying discrete subgroups of HCM patients with differences in electrophysiological and structural phenotype using a novel computational analysis of high fidelity 12-lead Holter ECGs, by combining computational clustering and mathematical modelling on the QRS. CMR imaging and clinical markers are evaluated to further characterize the patients' subgroups.

4.3 Methods

4.3.1 Clinical database

Eighty-six HCM patients were recruited from the University of Oxford Inherited Cardiac Conditions clinic, John Radcliffe Hospital, Oxford. Diagnosis of HCM was made by the presence of a pathogenic genetic mutation or, in the absence of an identified mutation, HCM was defined as left ventricular hypertrophy (LVH) ($\geq 15\text{mm}$) not originating from another cause. Patients were excluded if they had a diagnosis of coronary artery disease, diabetes, hypertension, atrial fibrillation or an ICD with paced rhythm. Thirty-eight normal controls, non-smokers, without cardiovascular disease, hypertension or diabetes, and no family history of cardiomyopathy or SCD were also recruited. All participants underwent 12-lead ambulatory Holter monitoring (at a sampling frequency of 1kHz, H12+, Mortara Instrument) for 24 hours. Standard 12-lead resting ECGs were also acquired from a Burdick 8500 digital electrocardiograph with the Glasgow algorithm (Macfarlane et al. 1990). Left ventricular volume, mass and function were obtained by CMR imaging, performed on a 3.0 Tesla Trio MR system (Siemens Healthcare, Erlangen, Germany). The

extent and morphology of hypertrophy were identified on short axis images and categorized into 4 subtypes:

- no hypertrophy: genotype positive patients with wall thickness $\leq 12\text{mm}$ (G+LVH-)
- septal hypertrophy: basal and/or mid septal wall thickness $>12\text{mm}$ in genotype positive patients or $\geq 15\text{mm}$ in genotype negative patients
- apical hypertrophy: apical wall thickness $\geq 15\text{mm}$ below papillary muscle level
- mixed hypertrophy: coexisting hypertrophy in septal and apical segments.

Clinical assessments of HCM patients included history, physical examination, family pedigree, ambulatory 3-lead Holter ECG and echocardiography. Five conventional risk factors (Elliott et al. 2000) were considered: non-sustained ventricular tachycardia (NSVT) (three or more consecutive ventricular beats at a rate of ≥ 120 bpm lasting $<30\text{s}$ on clinical 3-lead 24- to 48-hour Holter monitoring), unexplained syncope (≥ 1 episode of unexplained syncope), family history of SCD (history of SCD in ≥ 1 first degree relative ≤ 40 years old or SCD in a first degree relative with confirmed HCM at any age), massive LVH (LV wall thickness in any myocardial segment of $\geq 30\text{mm}$ on short axis CMR images) and abnormal exercise blood pressure (BP) response (rise in systolic BP <20 mmHg or a fall of >10 mmHg from baseline to peak exercise in patients ≤ 40 years old). The HCM Risk-SCD score (2014 ESC guidelines) was calculated for each patient using 7 disease variables as in (O'Mahony et al. 2014). A risk score of $\geq 6\%$ is classified as high risk and ICD implantation is recommended, 4–6% is intermediate risk and ICD may be considered, and $<4\%$ is low risk.

4.3.2 ECG signal processing

The raw digitalized Holter ECGs acquired required pre-processing to remove artefacts due to muscular noise, respiration and normal postural changes carried out during normal activities among others, as described in Chapter 3, section 3.2.1. The first 30-minute excerpts (~2000 beats) from the Holter ECGs were used to analyze the 8 linearly independent leads (I, II, V1-6) with custom-built software using MATLAB (Mathworks, MA, USA). A sensitivity analysis showed that analysing different 30-minute excerpts in the recording yielded the same results. 30 minutes also provided enough data to screen the excerpts and avoid changes in beat morphology due to changes in heart rate, and the segments were manually checked to ensure an average heart rate of 70bpm. The methodological pipeline is presented in Figure 4.1.

A wavelet based delineator (Martínez et al. 2004) extracted the peaks and boundaries of the ECG waveforms. The wavelet transform of a signal $x(t)$ is defined as its decomposition according to a base of functions derived from an initial wavelet:

$$WT_{a,b}(x) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \varphi\left(\frac{t-b}{a}\right) dt$$

with φ the initial wavelet, a a scale coefficient controlling the width of the basis, and b a translation coefficient. The coefficients a and b can be chosen to be discrete: $a = 2^k$, $k \in \mathbb{N}$. The idea of this algorithm is that the frequency bands of the different ECG waveforms lie at different scales. For instance, the energy of the ECG is located between the scales 21 and 25. The energy of the QRS is very low for scales higher than 24, while T and P waves still have consequent energies at scale 25. By representing these waveforms at different scales, the algorithm can perform the wavelet transform of the signal and reduce the delineation problem to simpler tasks such as zero crossing or, local optima detection. This is made possible using the derivative nature of the wavelet basis. Indeed, a zero

crossing in the wavelet transform corresponds to a maximum or minimum in the initial signal, highlighting the waveform evaluated. The algorithm first detects the QRS complex, then identifies the waves of the QRS, then determines the QRS beginning and end points, and finally detects the T and P-waves.

High frequency noise was removed using a low-pass Butterworth filter with cut-off frequency of 45Hz for QRS complexes and 25 Hz for T waves to allow the waveforms of interest. The magnitude of the frequency response (gain G) of a low pass Butterworth filter is given as:

$$G^2(w) = G_0^2 / (1 + (\frac{w}{w_c})^{2n}),$$

where w is the angular frequency, G_0 is the DC gain (gain at zero frequency), w_c the cut-off frequency (45Hz), and n the order of the filter ($n = 11$). This step enabled to smooth the signal in order to remove notches or peaks due to noise that would lead to a misinterpretation of the morphology of the signal

Baseline drift was removed by a cubic spline method to avoid morphology changes of non cardiac origin. The cubic spline method is based on the approximation of the baseline using knots indicating the isoelectric level between the P wave and the QRS complex. Here, the knots are chosen 80ms before the fiducial point of the QRS, on the PR segment. A first order polynomial fitting will not represent the isoelectric line accurately as lines will connect the knots. However, increasing the order of the polynomial leads to a risk of overfitting and increase of computational complexity. For cubic spline estimation, a third order polynomial is used. The polynomial fitting takes its root in Taylor series expansion. Indeed, by defining the knots as $\{t_i, i = 0, 1, \dots\}$, the baseline is computed by a Taylor serie:

$$y_{\infty}(t) = \sum_{n=0}^{\infty} \frac{(t-t_i)^n}{n!} y_{\infty}^{(l)}(t_i).$$

This formula can be truncated at the order of the polynomial (3 in our case) and after different approximations over the derivatives, the baseline is estimated, and then removed from the initial ECG. In addition, a notch filter rejected the 50Hz mains power artefact, coming from the recording devices.

The first twenty beats with maximal ST segment–T wave (ST-T) signal-to-noise ratio, defined as the ratio between the ST-T peak to peak amplitude and the root mean square value of the background noise >20Hz, were considered for analysis and were aligned with respect to the QRS complex by Woody's method (Sörnmo and Laguna 2005). This alignment technique is based on the computation of the correlation between each beat and a template beat. The template beat is computed once, at the beginning of the segmentation. It is an average of the first twenty beats mentioned before. No particular filtering or denoising method is applied, as the only thing that matters is the beats time position. The computation of the correlation between each beat and the template leads to a lag matrix composed of the possible time differences and the correlation they achieved. A lag window of [-90,90] is defined to explore all the possible shifting of the curves and the corresponding correlations. The lag number that achieves the maximum correlation between the two curves, computed by maximum likelihood estimation, is stored and the beat is shifted to this position to ensure alignment. This technique is repeated over all the beats until the maximum correlation is achieved for all. A window of 180ms centred on the mean energy of the QRS and a 500ms window aligned to the STT onset were computed in all 8 leads per participant. ECG excerpts were checked to ensure same heart rate in the segment and avoid misalignment from the STT onset due to changes in heart rate. Average QRS and STT waveforms were then computed.

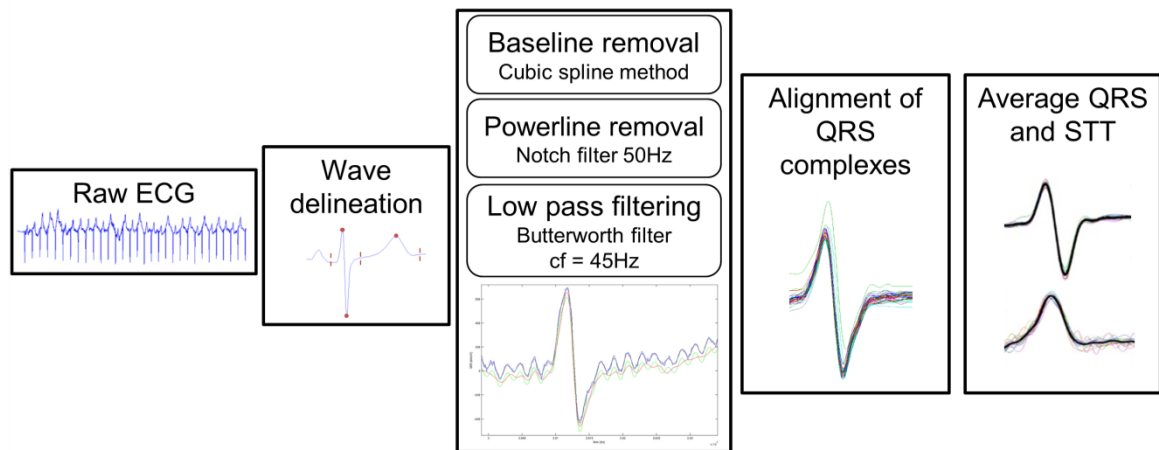


Figure 4.1: ECG signal processing pipeline: from the raw ECG to an average QRS and STT waveforms per lead per subject.

4.3.3 Extraction of standard and Hermite based biomarkers

All biomarkers were calculated per lead from the average QRS waveform. Five standard QRS biomarkers were defined as below, based on standard QRS measures and visual inspection of the ECGs. They are also represented on Figure 4.2:

- QRS amplitude, defined as the absolute value of the difference between maximum and minimum of the QRS complex values
- QRS width, defined as the time interval between the QRS onset and the QRS offset
- the maximum ascending/descending or descending/ascending slopes of the QRS defined as the first ascending and descending slopes depending on the QRS complex pattern
- the percentage of negative time samples in the QRS with respect to the isoelectric level (mostly representing the percentage of negative S wave with respect to QRS duration)

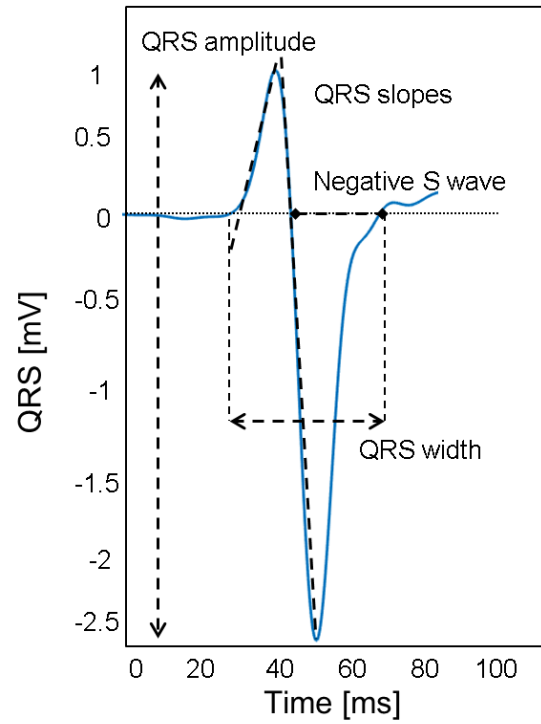


Figure 4.2: Standard biomarkers (amplitude, slopes, negative S wave, width) extracted from the QRS waveform

The QRS morphology (shape) was also quantified using computational modelling based on a combination of orthogonal Hermite functions (Figure 4.3), with well-established ability to provide a compact representation of the QRS complex and denoise the signal (Sandryhaila, Kovacevic, and Puschel 2011). Indeed, three Hermite functions enable to recover 98% of the QRS energy in control subjects. (Laguna et al. 1996). In our implementation, four Hermite functions were used in HCM because of the greater QRS heterogeneity found in HCM patients. They were defined as:

$$\forall x \in \mathbb{R}, \forall n \in \mathbb{N}, \forall a \in \mathbb{N}^*, \Psi_n(x, a) = (a2^n n! \sqrt{\pi})^{-\frac{1}{2}} e^{-\frac{x^2}{2a^2}} H_n\left(\frac{x}{a}\right)$$

where n is the order of the Hermite function Ψ and of the associated Hermite polynomial H , and a the width parameter.

The QRS was therefore expressed as a linear combination of the first four Hermite bases and a reconstruction error:

$$\forall t \in [1, 2, \dots, T], \hat{x}(t) = \sum_{n=1}^4 w_n b_n(t) + v(t)$$

where $\hat{x}(t)$ is the reconstructed QRS signal, $b_n(t)$ the Hermite function of order n , t the time samples from $T = 1$ to 180ms, w_n the Hermite coefficients in the linear combination and $v(t)$ the reconstruction error.

The width of the Hermite functions, allowing the consideration of a larger variability in QRS duration, was computed for each subject by minimizing the fitting error defined as the root mean square of the difference between the reconstructed and original QRS.

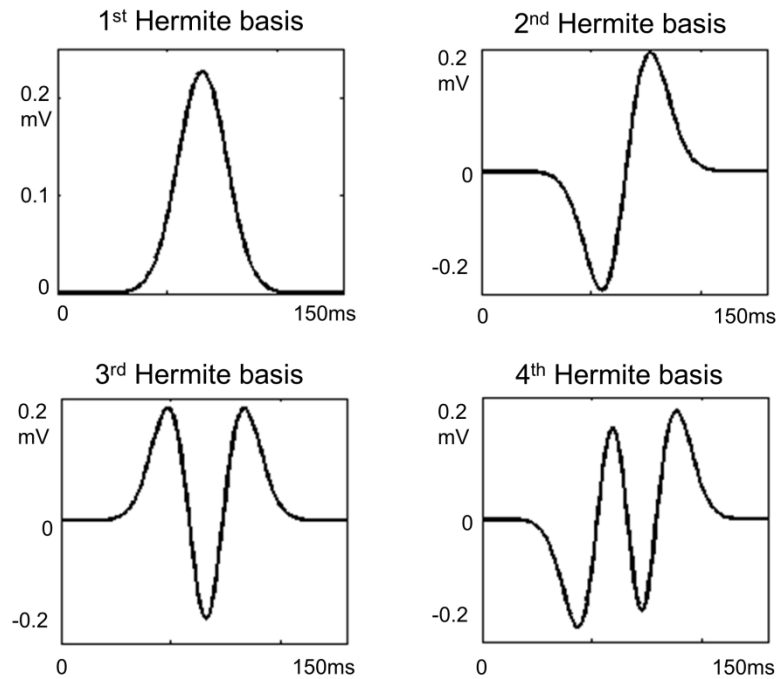


Figure 4.3: First four Hermite basis used in the QRS reconstruction

From this Hermite transform characterization, five biomarkers, now referred to as morphological QRS biomarkers, were computed in each lead for each subject:

- the four coefficients of the Hermite basis required to reconstruct the QRS, providing a compact representation of the morphology of the QRS
- the fitting error, defined as the root mean square of the difference between the reconstructed QRS and the original signal, providing a measure of the difficulty of reconstructing the QRS from four Hermite basis.

The width parameter a of the Hermite basis was not included as a feature because it did not show any significant difference between subjects (14.9 ± 1.2 for controls, 15.1 ± 1.6 for HCM, $p = 0.53$).

4.3.4 Binary classification between HCM and controls

A first step to characterize the HCM population was to investigate how distinct or similar their ECG features were compared to control. Binary classification between HCM patients and control subjects was therefore performed, using a logistic regression L1-regularized classifier, a supervised learning technique. Logistic regression is a probabilistic classification model based on the estimation of the probabilities of one point $\mathbf{P}$ belonging to one class or the other: $(C_1|\mathbf{P}) = \sigma(\mathbf{z}^T \mathbf{P})$, with σ the logistic sigmoid function. As a probabilistic model, the logistic regression aims at maximizing the conditional likelihood of the data: $\mathcal{L}(\mathbf{z}) = \sum_n \log p(t_n|\mathbf{P}_n)$, where t_n is the output label for the point $\mathbf{P}_n$. Regularized logistic regression introduces in addition a penalization to the objective function via a prior on the parameters. The prior is here defined as a L1 prior in order to push the weights $\mathbf{z}$ to zero and ensure sparsity:

$$p(\mathbf{z}) \propto \exp \frac{-\sum_i |z_i|}{2\gamma} \text{ with } \gamma \text{ being a constant parameter and } i \in [1, \dots, nb_features]$$

The objective function therefore becomes:

$$\mathcal{L}(\mathbf{z}) = \sum \log p(\mathbf{t}|\mathbf{P}) - \frac{\sum_i |z_i|}{2\gamma}$$

$\alpha = \frac{1}{2\gamma}$ is the regularization parameter and allows the investigation of the weight of each feature in the classification by varying its value (α was varied from 15, value that ensured all the features weights were set to 0, to 1).

For each subject, a feature vector was designed, composed of the standard QRS biomarkers and the morphological biomarkers described in 4.3.3, for each lead, resulting in an 80-dimensional feature vector. Considering this large number of features with respect to the amount of data, the Lasso technique (Tibshirani 1994) was implemented to remove the features with redundant information and identify the combination of the first four most discriminant features in the classification.

In our implementation, the dataset was randomly split into a training set for learning and a testing set for performance evaluation using 5-fold cross validation. The classifications were run 100 times to assess the performance of the classifier in terms of accuracy, sensitivity with respect to HCM and specificity.

4.3.5 Dimensionality reduction and clustering

The identification of subgroups in the HCM population was performed using blinded, unsupervised clustering. Each patient was represented by a vector of QRS biomarkers as described in 4.3.4. Seven significant features were selected, using the unsupervised multi-cluster feature selection algorithm (MCFS) published in (Cai, Zhang, and He 2010) to discriminate the dataset. To ensure proper comparison, Hermite coefficients were normalised by QRS amplitude for each patient. MCFS aims at preserving the multi-cluster structure of the data and relies on spectral analysis technique combined to optimization

solving methods like L1-regularized least square. The selected features were then reduced to two dimensions using Laplacian eigenmaps dimensionality reduction. This data-driven dimensionality reduction technique aims at extracting the low-dimensionality structure of the data from the high-dimension dataset, hypothesizing that the data lie in a low dimension manifold. It preserves the data local geometrical properties by building an adjacency matrix, representing the neighbouring connections between the data points $d_1 \dots d_k$, in our case, the morphological QRS biomarkers for the $k = 86$ patients. This preservation of the locality structure makes the method resistant to outliers and noise. Each data point is then connected to the others based on their proximity (following the heat kernel), creating a graph W such as:

$$W_{ij} = \begin{cases} \exp\left[-\frac{\|d_i - d_j\|^2}{t}\right], & \text{if } d_i \in N_j \text{ or } d_j \in N_i \\ 0, & \text{otherwise} \end{cases}$$

with N_i and N_j the n -nearest neighbours ($n \in \mathbb{N}$) of d_i and d_j and $t \in \mathbb{R}$, parameter of the heat kernel.

The Laplacian matrix is defined as: $L = D - W$ with $D_{ii} = \sum_j W_{ji}$.

The aim is to construct a set of points $y_1 \dots y_k$, living in the low dimensional manifold (in our case, two dimensions, as we are looking to reduce the dataset to 2D for visualisation purposes) that can represent $d_1 \dots d_k$. This means minimizing the objective function $\sum_{ij} (y_i - y_j)^2 W_{ij}$ so that points in the low dimension space are close if close in the initial space. The embedding map is therefore computed by calculating the eigenvectors of the eigenvalue problem: $L\mathbf{f} = \lambda D\mathbf{f}$.

The embedding on the low m -dimensional space then obtained is: $d_i \rightarrow (\mathbf{f}_1(i), \dots, \mathbf{f}_m(i))$.

A density-based clustering algorithm, DBSCAN (Ester et al. 1996), was applied on this

low-dimensional representation of the dataset to investigate the existence of distinct subgroups. This algorithm identified clusters by maximizing the density in each of the clusters. The minimum number of individuals in a cluster was set to $\lceil m/25 \rceil = 3$, with $m = 86$ HCM patients, as recommended in (Ester et al. 1996). The distance between points in the clusters was evaluated using the Euclidean distance.

4.3.6 Statistical analysis

In the tables presented in this chapter, data are expressed as mean $\pm$ standard deviation or median and range. Data were tested for normality using the Kolmogorov-Smirnov test. Normally-distributed data were compared using t-tests or analysis of variance. Non-normally distributed data were compared using the Mann-Whitney U-test or Kruskal-Wallis test. Categorical data were compared with Chi-square or Fisher's exact tests. Statistical significance was assumed when $p < 0.05$ (after Bonferroni adjustment for multiple comparisons, where appropriate).

4.4 Results

4.4.1 Binary classification between HCM and Controls

The CMR and QRS characteristics of HCM and control patients are summarized in Table 4.1. In terms of electrocardiogram features, HCM patients presented a more leftward QRS axis, larger QRS duration and amplitude, steeper QRS slopes and abnormal Q waves compared with controls (all $p < 0.03$). As expected, HCM patients also exhibited larger ejection fraction, mass index and maximum ventricular wall thickness than controls (all $p < 0.001$).

Table 4.1: Characteristics of normal controls and HCM patients [mean $\pm$ standard deviation, or number of participants (%)]

	Normal controls (n=38)	HCM patients (n=86)	p value
Age, years	47 $\pm$ 15	45 $\pm$ 14	0.54
Male	25 (66)	59 (69)	0.76
Body mass index, kg/m ²	24 $\pm$ 4	26 $\pm$ 4	0.06
Systolic BP, mmHg	117 $\pm$ 14	118 $\pm$ 14	0.76
Diastolic BP, mmHg	70 $\pm$ 10	72 $\pm$ 12	0.33
CMR dimensions and function			
LV end-diastolic volume, ml	155 $\pm$ 35	154 $\pm$ 33	0.88
LV end-systolic volume, ml	50 $\pm$ 17	41 $\pm$ 16	0.007
LV ejection fraction, %	68 $\pm$ 5	74 $\pm$ 7	0.00003
LV mass index, g/m ²	54 $\pm$ 11	72 $\pm$ 25	0.00002
Maximal LV wall thickness, mm	11 $\pm$ 1	20 $\pm$ 6	2 $\times$ 10 ⁻¹⁴
Left atrial diameter, mm	33 $\pm$ 7	39 $\pm$ 7	0.00004
Standard QRS features			
Heart rate, bpm	58 $\pm$ 11	58 $\pm$ 10	0.81
QRS axis, °	40 $\pm$ 34	13 $\pm$ 43	0.0007
QRS duration, ms	91 $\pm$ 9	98 $\pm$ 15	0.02
QRS amplitude, mV	1441 $\pm$ 414	1825 $\pm$ 652	0.002
QRS ascending slope	74 $\pm$ 22	89 $\pm$ 32	0.025
QRS descending slope	-118 $\pm$ 51	-150 $\pm$ 57	0.001
Pathological Q wave	0	20 (23)	0.0004

BP, blood pressure; CMR, cardiovascular magnetic resonance; LV, left ventricular.

Table 4.2 describes the clinical characteristics of the HCM cohort. Patients were mainly asymptomatic (median NYHA functional class = 1) and were low risk for SCD (median HCM Risk-SCD score = 2.4%; median total risk factor = 1). 19 patients had an ICD implanted for primary prevention with a median follow-up of 3 years. Only 1 patient (5%) received an appropriate shock, which is in keeping with primary prevention discharge rates seen in a previous study (Maron et al. 2000). 10% of patients were genotype positive with normal wall thickness (G+LVH-) and the majority had septal hypertrophy (69%).

Table 4.2: Clinical and genotype characteristics of HCM patients [number of patients (%)]

NYHA class, [median (range)]	1 (1-3)
NYHA III/IV	5 (6)
LVOT obstruction (gradient ≥ 30 mmHg)	11 (13)
Implantable cardioverter-defibrillator	19 (22)
Appropriate ICD shocks	1
ICD follow-up, years [median (range)]	3 (0-12)
Hypertrophy morphology	
No LVH (G+LVH-)	9 (10)
Septal LVH	59 (69)
Apical LVH	4 (5)
Mixed septal & apical LVH	14 (16)
HCM Risk-SCD score, % [median (range)]	2.4 (0.8-11.0)
Conventional risk factors	
NSVT	23 (27)
Syncope	10 (12)
Family History SCD	17 (20)
Abnormal exercise BP response	7 (8)
Massive LVH ≥ 30 mm	1 (1)

<i>Total number of risk factors [median (range)]</i>	1 (0-3)
<i>0 Risk factors</i>	35 (41)
<i>1 Risk factor</i>	45 (52)
<i>≥ 2 Risk factors</i>	6 (7)
Genotype	
Gene negative	27(32)
MYBPC3	33 (38)
MYH7	24 (28)
Troponin I	1 (1)
Mitochondrial	1 (1)

NYHA, New York Heart Association; LVOT, left ventricular outflow tract; ICD, implantable cardioverter-defibrillator; LVH, left ventricular hypertrophy; G+LVH-, genotype positive HCM with normal wall thickness; SCD, sudden cardiac death; NSVT, non-sustained ventricular tachycardia; BP, blood pressure; ESC, European Society of Cardiology; HCM, hypertrophic cardiomyopathy; MYBPC3, myosin-binding protein C; MYH7, beta-myosin heavy chain

QRS biomarkers were evaluated to identify abnormalities in HCM patients with respect to controls. Figure 4.4 highlights the existing overlap between standard QRS biomarkers in controls and HCM populations. Only QRS amplitude and ascending slope in the septal leads (V1-V2) were significantly larger in HCM patients, due to septal hypertrophy. One QRS morphological biomarker (the first Hermite coefficient) also showed significant differences in the lateral leads V4-V5 (Figure 4.4) between HCM patients and controls.

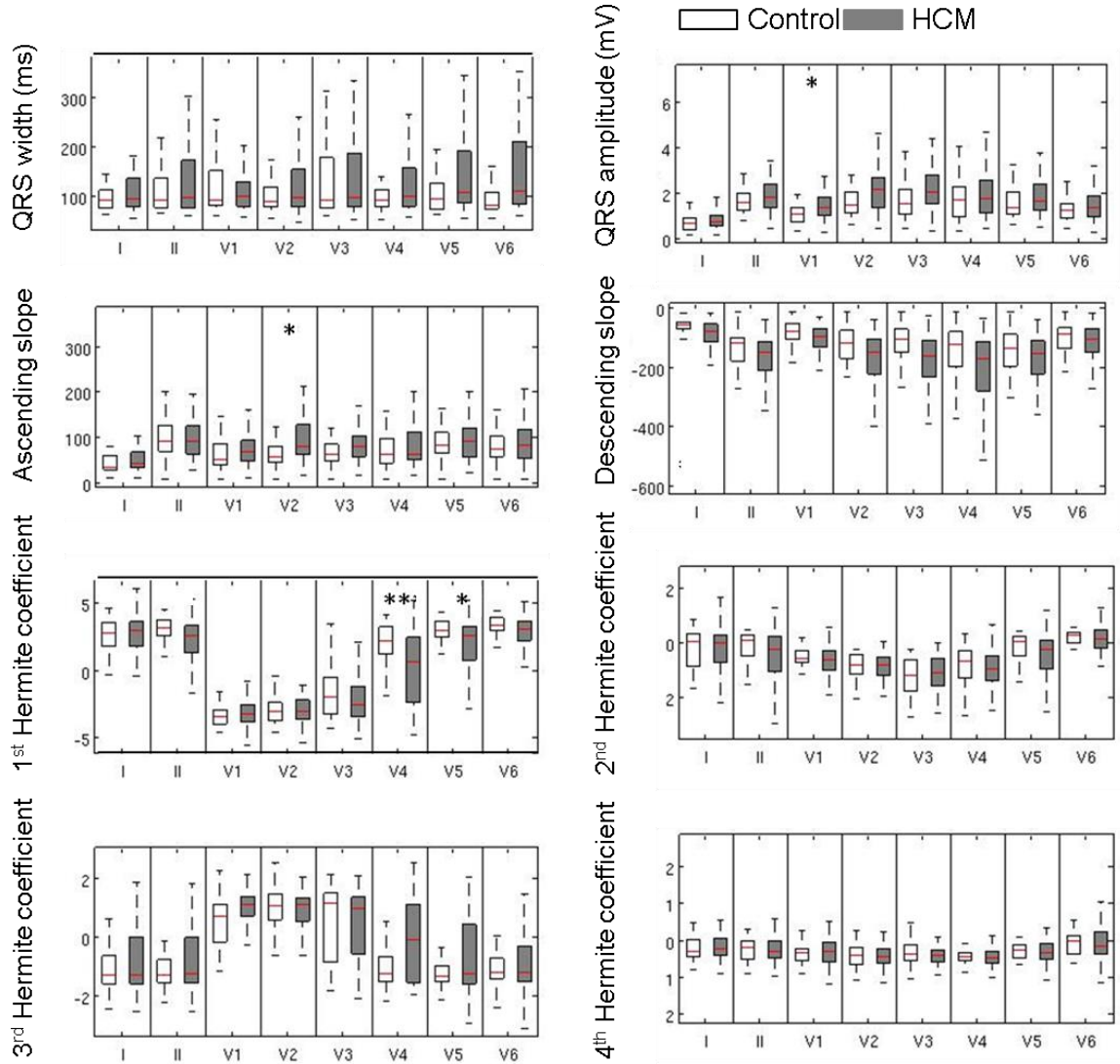


Figure 4.4: Comparison of standard (QRS width, amplitude, ascending and descending slopes) and morphological (four Hermite coefficients) QRS biomarkers in leads I-II and V1-6 between HCM and controls. Mean, standard deviation and interquartile range; p-values are Bonferroni-corrected (2x8) to account for multiple comparisons (*p<0.05, **p<0.01, ***p<0.001).

Therefore, when performing classification with logistic regression L1-regularized, only 66% of HCM patients were correctly classified with respect to controls using standard QRS biomarkers (Table 4.3). This may be explained by the large heterogeneity of the HCM population. With standard and morphological QRS biomarkers, we performed feature selection to reduce the number of features in the classifier and rank the biomarkers

according to their influence in the classification process. The top 4 ranking results are presented in Table 4.4, as well as the cumulative accuracy reached for combining these first biomarkers. The addition of QRS morphological biomarkers increased the classification accuracy to 75%. The QRS biomarkers that yielded better discrimination between HCM and control were the first Hermite coefficient in V4 and V6, the third coefficient in V5, and the percentage of S wave duration in V6, which were obtained by feature selection.

Table 4.3: Discriminative power of the standard QRS biomarkers: accuracy of the logistic regression classifier

Biomarker	Number of features	Accuracy (%)	Sensitivity to HCM (%)	Specificity (%)
QRS amplitude	8	63.5	69.1	55.2
QRS width	8	63.4	66.2	57.1
QRS slopes (ascending and descending)	8	61.7	71.5	34.7
QRS amplitude + width + slopes	24	66.6	76.1	50.2

Table 4.4: Ranking of biomarkers according to their influence on the classification process as well as the cumulative performance evaluation

Biomarker	Cumulative accuracy (%)	Sensitivity to HCM (%)	Specificity (%)
(1) 1 st Hermite coefficient lead V4	64.7	69.5	61.6
+ (2) 1 st Hermite coefficient lead V6	65.7	72.4	60.1
+ (3) % negative regions lead V6	70.9	73.7	65
+ (4) 3 rd Hermite coefficient lead V5	75.1	77.6	66.2

4.4.2 Identification of three ECG phenotypes based in QRS biomarkers

Based on QRS morphology alone, three HCM groups were obtained by clustering (Figure 4.5, Panel A). The main differences were seen in first, second and third Hermite basis in lead II and lateral precordial leads (V4-V6) (Figure 4.5, Panel B-D; Figure 4.6).

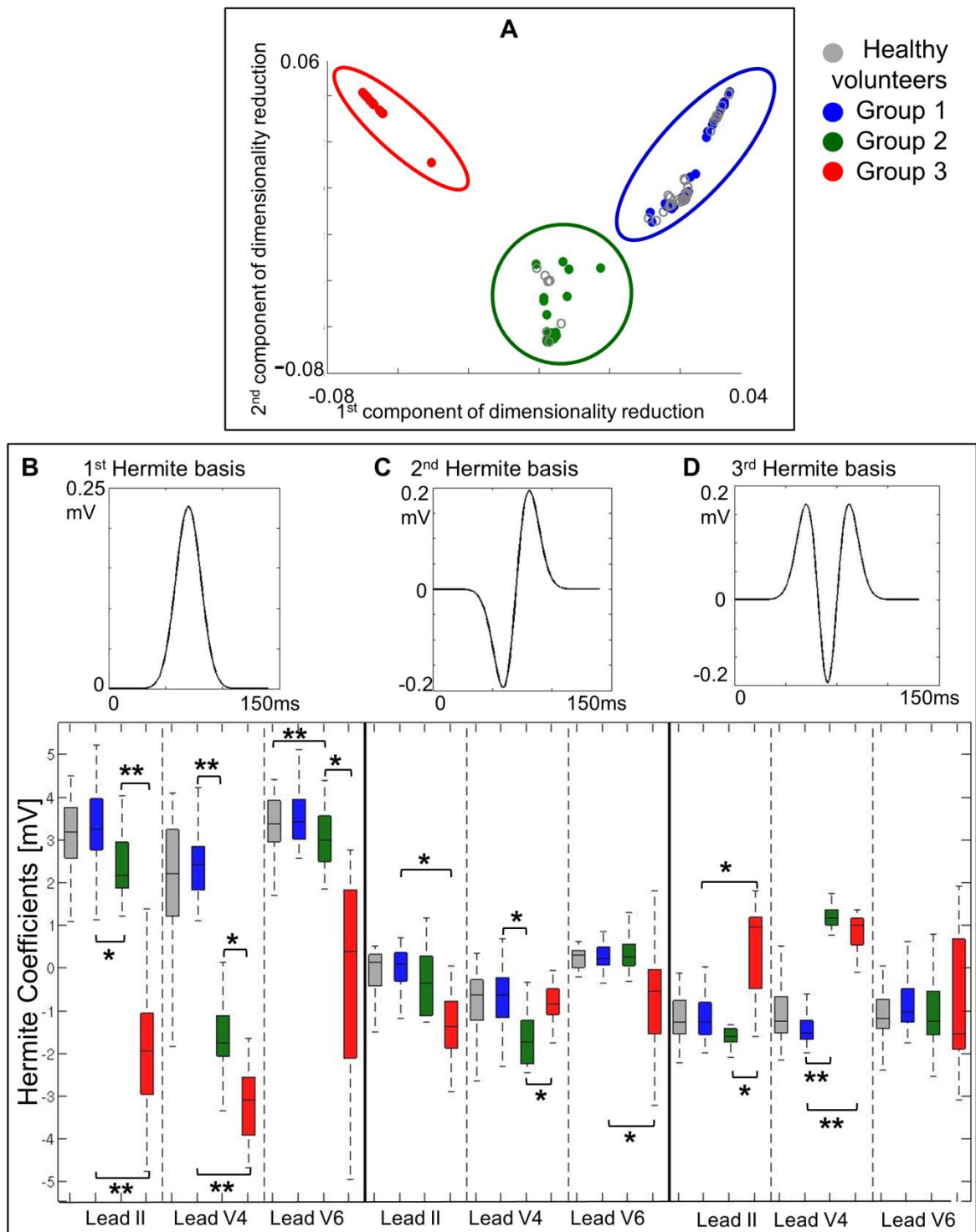


Figure 4.5: Clustering with QRS morphology alone identifies three HCM groups showing group differences in 3 features in lead II and the lateral precordial leads. (A) The three QRS-based HCM groups identified by cluster analysis using QRS morphological biomarkers alone are shown on the 2-dimensional space obtained by dimensionality reduction. (B-D) These QRS-based HCM groups show differences in the 1st, 2nd and 3rd Hermite coefficients (mathematical functions representing the QRS shape: QRS morphological biomarkers) in leads II, V4 and V6. Controls are shown for visual comparison but were not included in Kruskal-Wallis ANOVA as the focus of this section is to characterize subgroups within the HCM population (** $p < 1 \times 10^{-6}$, * $p < 0.001$). Comparison with Controls can be seen in section 4.4.1.

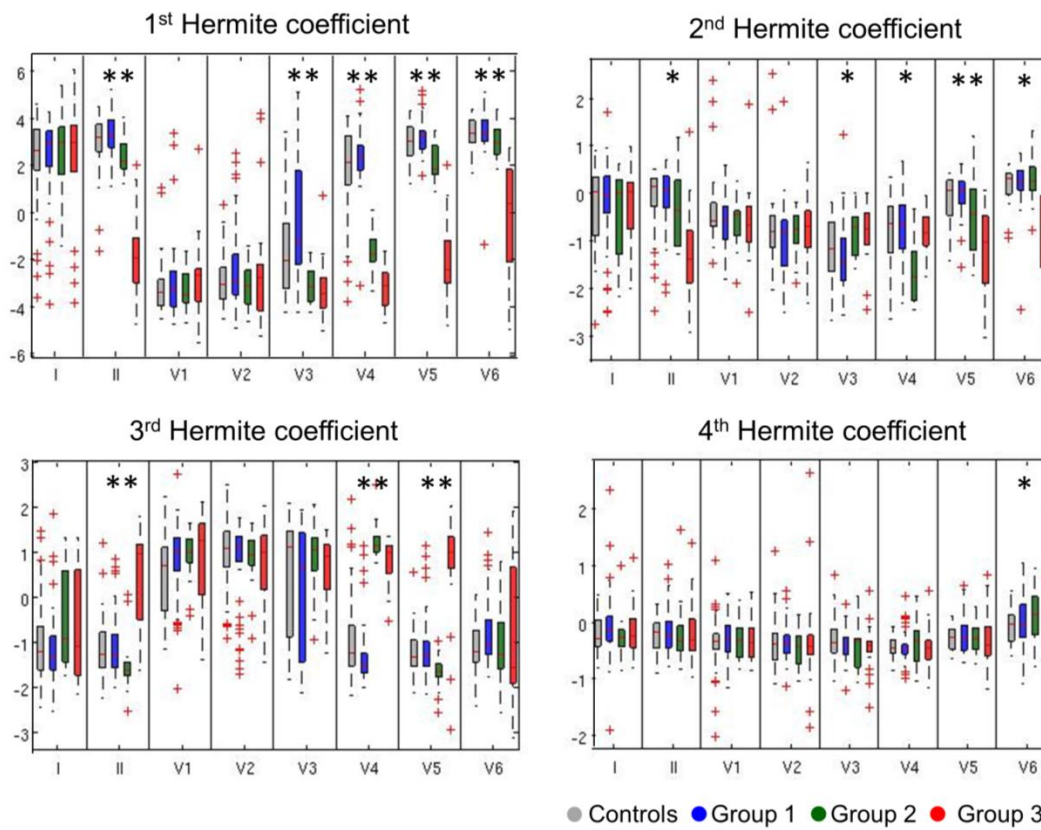


Figure 4.6: Comparison of the four Hermite coefficients in leads I-II, V1-6 for the three QRS-based HCM groups. Controls are shown for visual comparison but were not included in Kruskal-Wallis one-way analysis of variance. Mean, standard deviation and interquartile range; p-values are Bonferroni-corrected for multiple comparisons (* p -value <0.001 , ** p -value $<1 \times 10^{-6}$).

Group 1 was the largest group with 52% of the patients ($n = 45$), and displayed normal QRS morphology, suggested by Hermite coefficients not significantly different compared to controls (Figure 4.5), and QRS axis in normal ranges (Table 4.5). Group 2 (22% of patients, $n = 19$) showed differences in V4 in the first three Hermite bases compared to controls and Group 1, but no difference in V6 (Figure 4.5, all post-hoc $p < 0.001$). Lead V4 in Group 2 displayed shorter R wave duration (2: 38ms, 1: 47ms; 2 vs 1, $p = 0.003$) and deeper S waves (2: $-1170\mu\text{v}$, 1: $-568\mu\text{v}$; 2 vs 1, $p = 0.0003$) compared to Group 1 (Table 4.5). Group 3 (26% of patients, $n = 22$) exhibited vast differences in lead II and V4-6 in the first three Hermite bases compared to the other HCM groups and controls (Figure 4.5, all post hoc $p < 6 \times 10^{-6}$). Differences in lead II were manifested as abnormally shifted QRS axis towards left axis deviation (LAD) (3: -37° , 1: 30° , 2: 29° ; 3 vs 1 and 3 vs 2, all post hoc $p < 2 \times 10^{-7}$). In V4, the R wave duration was shorter than Group 1 and amplitude was shorter than Groups 1 and 2, S wave duration was longer than Group 1, and S amplitude was deeper than Group 1 (3 vs 1 and 3 vs 2, all post hoc $p < 0.003$). In V6, R wave duration and amplitude were shorter, and S wave duration and amplitude larger than both the other groups (3 vs 1 and 3 vs 2, all post hoc $p < 0.01$, Table 4.5). The three QRS morphologies are summarized on Figure 4.7, and Table 4.5 summarizes the standard ECG features for the QRS-based groups.

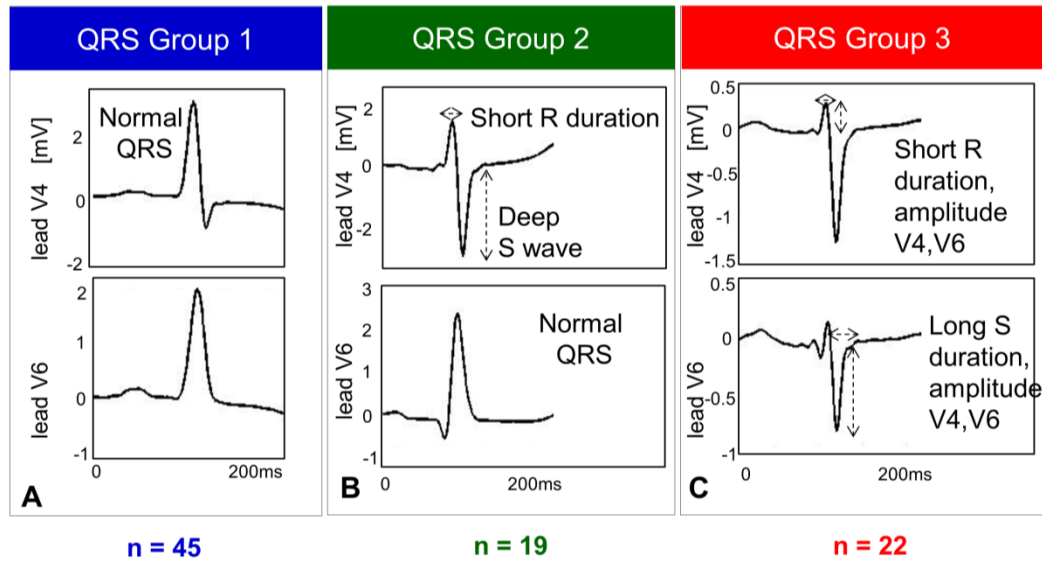


Figure 4.7: Three distinct QRS morphologies of the three subgroups identified. (A) Group 1 shows normal QRS. (B) Group 2 shows short R wave duration and deep S wave in V4. (C) Group 3 shows short R wave duration and amplitude and long S wave duration and amplitude in V4 and V6.

Table 4.5: Standard ECG features for QRS-based HCM groups [mean $\pm$ standard deviation, or number of patients (%)]

	Group 1 (n=45)	Group 2 (n=19)	Group 3 (n=22)	p value (group comparison)
Heart rate, bpm	58 $\pm$ 9	59 $\pm$ 13	55 $\pm$ 10	0.46
QRS axis, $^{\circ}$	30 $\pm$ 33 ^b	29 $\pm$ 34 ^c	-37 $\pm$ 28	1 $\times$ 10 ⁻¹¹
QRS duration, ms	97 $\pm$ 17	96 $\pm$ 16	102 $\pm$ 13	0.13
QRS amplitude, mV	1876 $\pm$ 702	1807 $\pm$ 588	1737 $\pm$ 613	0.88
QRS ascending slope	94 $\pm$ 34	92 $\pm$ 34	76 $\pm$ 23	0.10
QRS descending slope	-149 $\pm$ 58	-155 $\pm$ 60	-149 $\pm$ 56	0.95
Pathological Q waves	3 (7) *	8 (42)	9 (41)	0.0003
Lead V4				
R duration V4, ms	47 $\pm$ 11 ^{a,b}	38 $\pm$ 10	41 $\pm$ 17	0.0001

R amplitude V4, μv	1310 ± 706^b	1307 ± 648^c	551 ± 433	0.00002
S duration V4, ms	36 ± 17^b	41 ± 13	47 ± 19	0.01
S amplitude V4, μv	$-568 \pm 413^{a,b}$	-1170 ± 570	-1270 ± 863	0.00002
Lead V6				
R duration V6, ms	54 ± 13^b	51 ± 14^c	31 ± 18	0.000006
R amplitude V6, μv	1290 ± 696^b	1227 ± 451^c	$635 \pm 455^*$	0.00008
S duration V6, ms	17 ± 18^b	15 ± 17^c	35 ± 19	0.0004
S amplitude V6, μv	-118 ± 148^b	-122 ± 190^c	$-501 \pm 460^*$	0.00006

^aGroup 1 vs 2, ^b1 vs 3, ^c2 vs 3, $p < 0.05$ on post hoc pairwise comparisons (p-values multiplied by 3 for Bonferroni adjustment of 3 tests). * $p < 0.05$ on post hoc contingency table analysis (p-values multiplied by 6 for Bonferroni adjustment of 3×2 combinations).

4.4.3 Association with MRI and arrhythmic risk factors

Although ECG features were significantly different between the QRS-based groups, clinical features from CMR and markers of arrhythmic risk, presented in Table 4.6 below, did not differ significantly, suggesting that QRS biomarkers alone may not be useful for risk stratification in HCM.

Table 4.6: Clinical features for QRS-based HCM groups [mean $\pm$ standard deviation, median (range) or number of patients (%)]

	Group 1 (n=45)	Group 2 (n=19)	Group 3 (n=22)	p value (group comparison)
Age, years	47 ± 15	44 ± 12	43 ± 13	0.34
Male	31 (69)	12 (63)	16 (73)	0.80
Body mass index, kg/m^2	27 ± 5	25 ± 4	25 ± 4	0.16
Systolic BP, mmHg	120 ± 13	117 ± 13	114 ± 15	0.28
Diastolic BP, mmHg	74 ± 11	73 ± 13	68 ± 12	0.13

HCM Risk-SCD score, %	2.5 (1-11)	2.1 (1-6)	2.5 (1-9)	1.00
Conventional risk factors				
NSVT	14 (31)	4 (21)	5 (23)	0.69
Syncope	6 (13)	2 (11)	2 (9)	1.00
Family History SCD	7 (16)	5 (26)	5 (23)	0.52
Abnormal exercise BPR	5 (11)	0	2 (9)	0.44
Massive LVH ≥ 30 mm	0	0	1 (5)	-
Hypertrophy				
LV mass index, g/m ²	74 $\pm$ 28	63 $\pm$ 17	73 $\pm$ 26	0.30
Max LV wall, mm	19 $\pm$ 6	19 $\pm$ 5	21 $\pm$ 6	0.22
Hypertrophy morphology				0.008
No LVH (G+LVH-)	8 (18)	1 (5)	0	
Septal LVH	23 (51)	16 (85)	20 (90)	
Apical LVH	2 (4)	1 (5)	1 (5)	
Mixed septal & apical LVH	12 (27)	1 (5)	1 (5)	
Other clinical features				
LV end-diastolic volume, ml	152 $\pm$ 28	155 $\pm$ 33	156 $\pm$ 41	0.87
LV end-systolic volume ml	40 $\pm$ 14	43 $\pm$ 14	41 $\pm$ 21	0.50
LV ejection fraction, %	74 $\pm$ 8	72 $\pm$ 8	74 $\pm$ 7	0.61
Left atrial diameter, mm	40 $\pm$ 7	37 $\pm$ 6	38 $\pm$ 8	0.25
LVOT gradient, mmHg	7.0 (4-111)	6.7 (2-110)	6.8 (3-92)	0.92
Genotype				
Gene positive	29 (64)	13 (68)	17 (77)	0.65

HCM, hypertrophic cardiomyopathy; SCD, sudden cardiac death; NSVT, non-sustained ventricular tachycardia; BP, blood pressure; BPR, blood pressure response; LVH, left ventricular hypertrophy; LV, left ventricular; G+LVH-, genotype positive HCM with normal wall thickness; LVOT, left ventricular outflow tract.

In addition, the location of the maximum left ventricular wall thickness was similar in the three groups, as presented in Figure 4.8. The majority of patients in all of the groups displayed a maximum LV wall thickness located in the basal anteroseptal, basal inferoseptal and mid inferoseptal segments. The location of the maximum hypertrophy therefore did not explain the differences in QRS abnormalities identified in the three subgroups. Moreover, Group 1 gathered the majority of patients exhibiting mixed septal and apical hypertrophy and genotype positive patients with no hypertrophy (Figure 4.8).

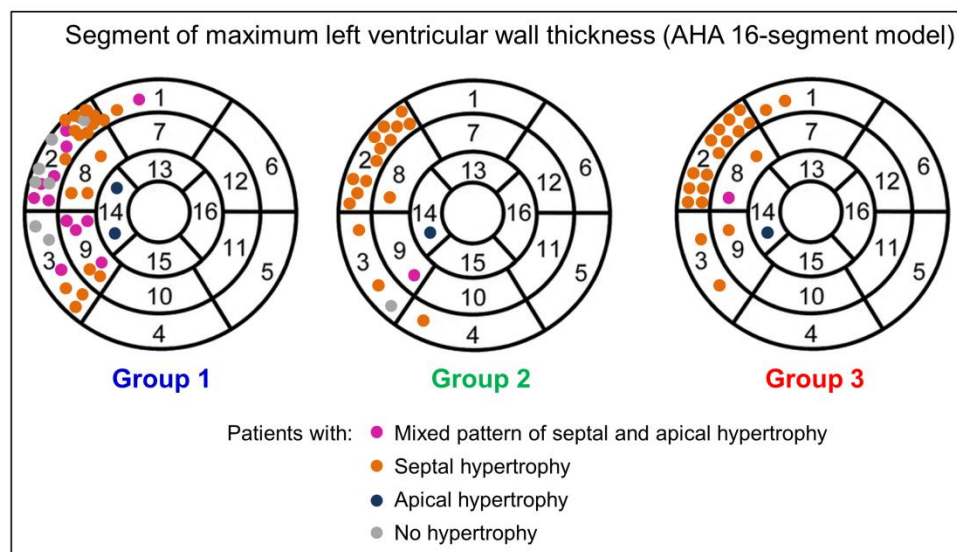


Figure 4.8: Location of maximum left ventricular wall thickness in the three groups: Group 1 had a predominance of patients with mixed septal and apical hypertrophy (pink), and genotype positive patients with no hypertrophy (grey). Group 2 and 3 patients predominantly exhibited septal hypertrophy (orange). Segments of maximum wall thickness are represented on the American Heart Association (AHA) 16-segment model of the left ventricle for each group.

4.5 Discussion

In this chapter, we investigated the role of QRS biomarkers in unfolding the phenotypic heterogeneity of HCM. We hypothesized that, as QRS morphology reflects the cardiac depolarization sequence, it may therefore be affected by HCM structural abnormalities such as hypertrophy, fibre disarray or fibrosis. We quantitatively assessed the QRS morphology as a whole using mathematical modelling rather than describing it with discrete standard features. Computational clustering with QRS morphology alone identified three groups with unique and distinct QRS features (Figure 4.7). Group 1 showed normal QRS morphology, similar to Control, explaining the poor accuracy results of the binary classification based on QRS markers. Group 2 presented QRS abnormalities in lead V4 with short R wave duration and deep S wave. Group 3 exhibited severe QRS abnormalities in both leads V4 and V6 with short R wave duration and amplitude, and long S wave duration and amplitude and left axis deviation (LAD). LAD is well-known to be associated with LVH but other factors such as a degree of left anterior fascicular block could also account for this leftward axis. Group 3 patients mainly had isolated septal hypertrophy, yet QRS abnormalities were seen in the lateral leads suggesting that remodeling may occur distal from the septum causing less uniform electrical propagation in the lateral leads.

These QRS variations could not be accounted for by differences in left ventricular hypertrophy. Right ventricular hypertrophy (RVH) may explain Group 3 ECG phenotype, but RV wall thickness was not performed in the clinical assessment of our dataset so cannot be compared between the groups. Other structural changes such as fibrosis (Dumont et al. 2006) and fibre disarray may also affect depolarization and QRS particularly in patients in Groups 2 and 3. This analysis suggests that while QRS features may be informative for diagnosis in HCM (Konno et al. 2004), categorizing HCM groups

on this basis showed no differences in association with known markers of risk. No genotype association was seen across any group but our sample size may not have had adequate power to assess genotype-phenotype correlations.

4.6 Conclusions

This chapter focused on the clustering of the HCM population using morphological QRS biomarkers only. The results of the analysis suggest that the QRS alone may not be sufficient for risk stratification and that repolarization may play a key role in arrhythmogenesis and help identify patients at risk. Chapter 5 will therefore investigate the effect of including T wave biomarkers in the clustering analysis, accounting for the repolarization phase. Similarly, the subgroups obtained will be compared in terms of clinical markers and markers of arrhythmic risk.

5 EXTRACTION OF HCM PHENOTYPES WITH QRS AND T WAVE MORPHOLOGICAL BIOMARKERS

The work in this chapter is part of: “Distinct ECG phenotypes identified in hypertrophic cardiomyopathy using machine learning associate with arrhythmic risk markers” by Lyon et al., 2018, Frontiers in Physiology, DOI: 10.3389/fphys.2018.00213

5.1 Introduction

In Chapter 4, we developed signal processing and computational clustering techniques to blindly identify and characterize three phenotypes in HCM based on QRS biomarkers, hypothesizing that HCM structural abnormalities would affect the electrical activation and lead to changes in the QRS complex on the ECG. These three subgroups exhibited distinct QRS morphologies, but they did not correlate with markers of arrhythmic risk or clinical markers, suggesting that the QRS alone is not sufficient for risk stratification in HCM and that repolarization may play a key role in arrhythmogenesis. In addition to the structural

abnormalities affecting HCM hearts presented in the previous chapter, ionic remodelling (Coppini et al. 2013) plays a major role in creating a pro-arrhythmic substrate in HCM. Indeed, HCM patients suffer from abnormalities in the calcium subsystem, especially an increase of L-type calcium, as well as an increase of Late Sodium current and a reduction of potassium currents. These may affect ventricular repolarization and lead to abnormalities in the T wave in the ECG. Electrical instability in HCM therefore may be reflected as abnormalities in both the QRS complex (depolarization) and the T wave (repolarization). It is known that T wave changes can occur when the QRS is normal – a primary repolarization abnormality or, as a direct result of altered depolarization – a secondary repolarization abnormality (where both the QRS and T wave are altered) (Fisch 1992). In this chapter, we investigate the influence of repolarization in HCM risk stratification and, more generally, the discriminative power of the combination of QRS and T wave morphological biomarkers, using clustering based on T wave biomarkers in conjunction with the QRS biomarkers used in Chapter 4.

5.2 Methods

5.2.1 Extraction of standard and KLT-based repolarization biomarkers

Similarly to the computation of QRS biomarkers presented in 4.3.2, all ST-T biomarkers were calculated per lead from the average ST-T waveform obtained after preprocessing the ECG signal as described in 4.3.1. Six standard ST-T biomarkers were defined as below and are represented on Figure 5.1.

- T wave amplitude, defined as the absolute value of the difference between maximum and minimum of the T wave values
- ST segment elevation, measured at J point (defined as the junction between the end of the QRS complex and the beginning of the ST-segment) + 80ms (as heart rate is less than 120 bpm) (Jager et al. 2003)

- QTc interval, defined as the duration between the beginning of the QRS and the end of the T wave, corrected with the Bazett's formula (most clinically used correction) in each lead
- JTc interval, defined as the duration between J point and the end of the T wave, corrected with the Bazett's formula in each lead
- T peak to T end interval, defined as the duration between the peak of the T wave and the end of the T wave in each lead
- T wave inversion, defined as a negative inflexion of at least 0.1mV in at least two contiguous leads in V3-6.

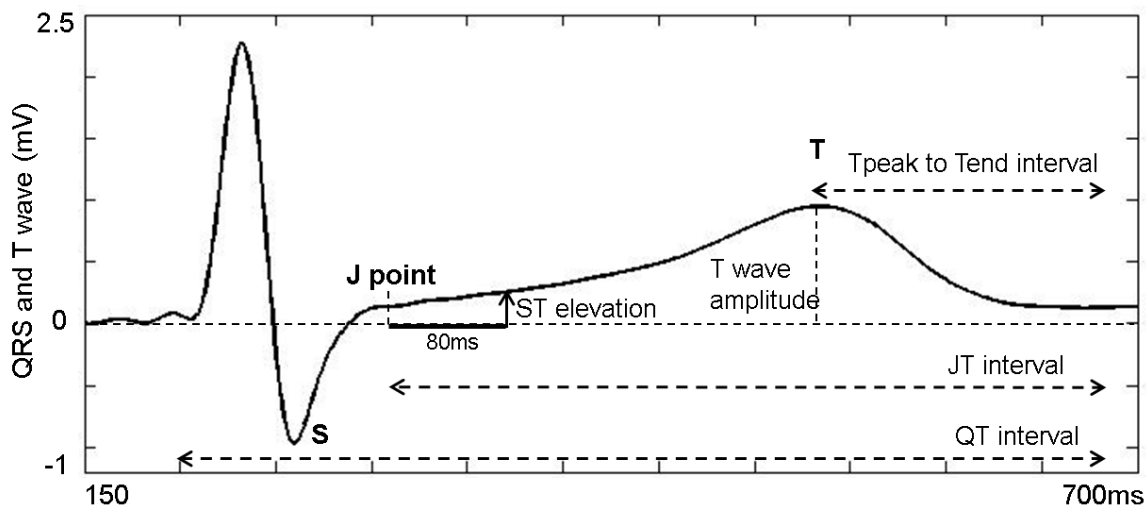


Figure 5.1: Standard biomarkers extracted from the STT segment waveform

For the computation of these biomarkers, ECG waveforms onset and offset (such as QRS beginning and end, T wave beginning and end) were obtained from the delineation algorithm detailed in section 4.3.2.

In addition, the ST-T segment was characterised using a set of mathematical functions known as the Karhunen-Loève transform (KLT) (Zaharov et al. 2014). This transform is well-established in reconstructing the ST-T segment and providing a meaningful representation of the shape of the ST-T waves by maximizing the energy of the ST-T

population and concentrating the signal information in a minimum number of parameters (Laguna, Moody, and Mark 1994) (Olcayto 1997). The KLT is an orthogonal linear transform which can be used to reconstruct ECG signals with a specific number of bases, similarly to the Hermite transform characterization introduced in Chapter 4. However, at the difference of the Hermite transform, the KLT is signal-dependent. Indeed, it must be derived from the statistics of the signal, in this case, ST-T segments. A training dataset is therefore required in order to compute the KL bases. The KLT bases are then computed as the orthogonal eigenvectors of the covariance matrix of the training dataset. In this work, we used a set of bases already computed in a previous study over a large independent dataset of 200,000 preprocessed ST-T waveforms (García et al. 1998) (Figure 5.2). 90% of the signal energy can be reconstructed using four KLT coefficients, motivating the use of four KL coefficients in this work for the reconstruction of the ST-T waveforms (Laguna, Moody, and Mark 1994). The choice of these specific bases is also discussed in section 5.4.

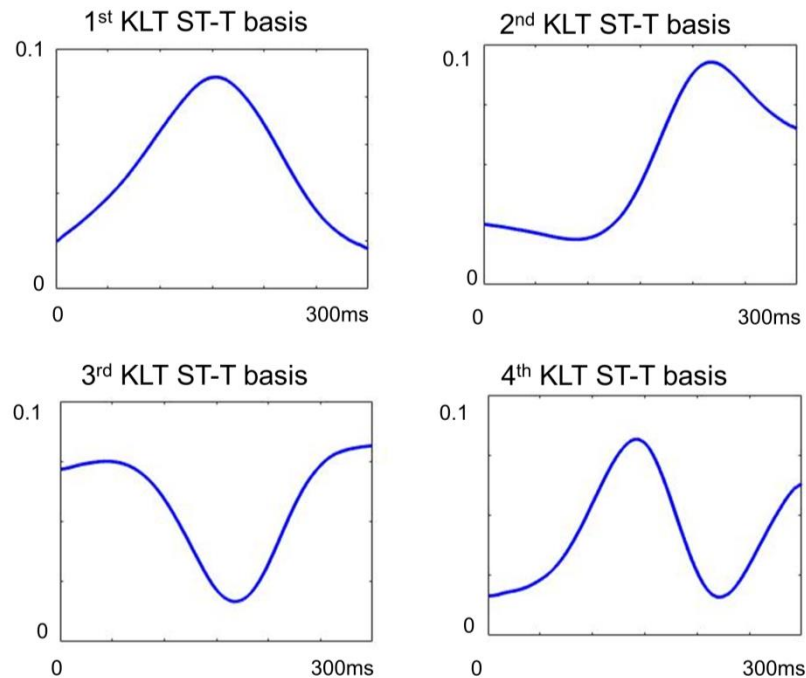


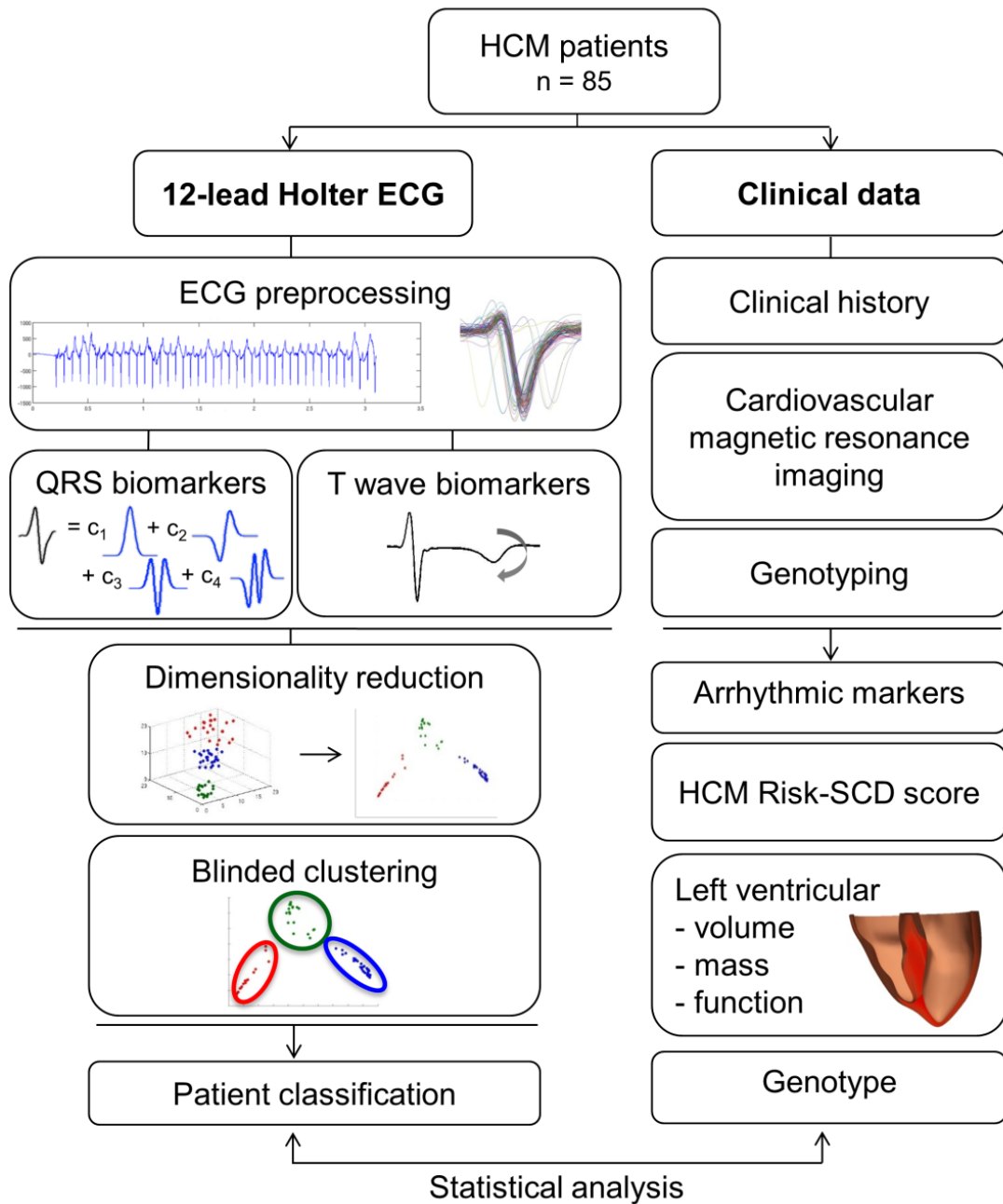
Figure 5.2: Four Karhunen-Loève transform basis functions for the ST-T segment

Five KL-based ST-T biomarkers were therefore computed:

- the four coefficients of the KL basis required to reconstruct the ST-T
- the fitting error, defined as the root mean square of the difference between the reconstructed and the original ST-T, providing a measure of the difficulty of reconstructing the STT from only four KL basis functions.

5.2.2 Computational clustering

The identification of subgroups in the HCM population was performed using the methodology introduced in Chapter 4, with the addition of the T wave biomarkers presented in section 5.2.1. Dimensionality reduction and clustering were similarly performed using Laplacian dimensionality reduction and DBSCAN clustering algorithm. Figure 5.3 recapitulates the methodological pipeline implemented for the identification of HCM subgroups in this thesis.



$C_1 \dots C_4$: first to fourth Hermite transform coefficients (Methods section)

Figure 5.3: Summary of the methodologies applied in this work for the analysis and classification of 86 HCM patient ECGs using mathematical modelling and machine learning, and the analysis of their clinical and cardiovascular magnetic resonance features.

5.3 Results

5.3.1 Identification of four ECG phenotypes based on QRS and T wave biomarkers

In order to get a complete electrocardiographic signature and account for both the activation and repolarization phases, QRS morphological biomarkers obtained by feature selection in Chapter 4 were combined with the selected repolarization biomarkers just described, and the clustering analysis was recomputed.

In the feature selection process, T wave inversion was the principal repolarization biomarker selected and led to the identification of four groups. Groups 2 and 3 from the QRS clustering remained as before, and the two remaining groups were the result of the separation of Group 1, from the QRS-based clustering, based on T wave inversion. Patients in Group 1A exhibited T wave inversion (n = 21; Figure 5.5) while patients in Group 1B showed no T wave inversion (n = 24; Figure 5.5).

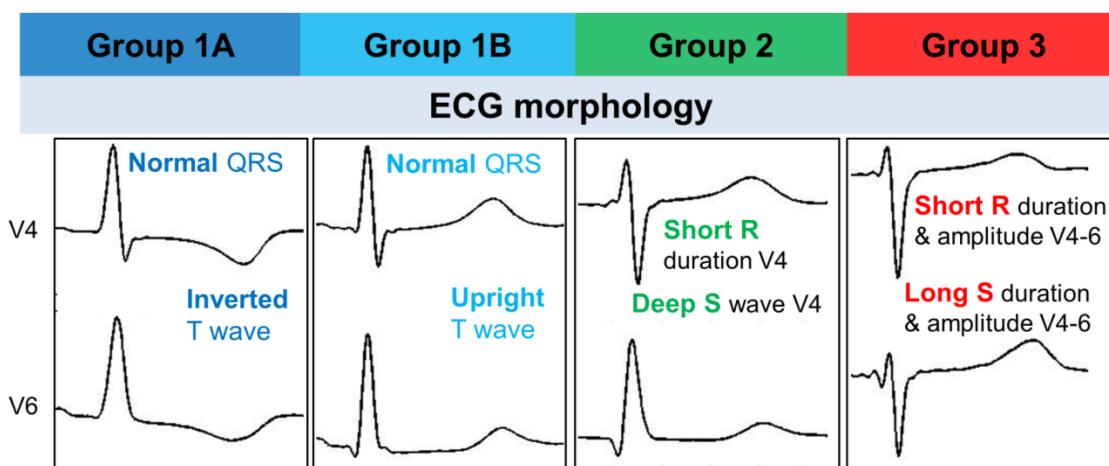


Figure 5.4: Four groups with distinct ECG morphology are identified by combined clustering with QRS morphology and T wave biomarkers. Group 1A – normal QRS with inverted T wave and abnormal T wave axis deviation (primary repolarization abnormality), Group 1B – normal QRS with upright T wave, Group 2 – short R wave duration and deep S wave in V4, Group 3 – left axis deviation, short R wave duration and amplitude, and long S wave duration and amplitude in V4 and V6.

Table 5.1 summarizes the standard ECG features for the four HCM groups. Group 1A had a more rightward T wave axis (1A: 156°, 1B 42°, 2: 49°, 3: 68°; 1A vs 1B, 1A vs 2, 1A vs 3, all post hoc $p < 2 \times 10^{-6}$) and the greatest number of patients with an abnormal T wave axis (17 patients; post hoc $p = 2 \times 10^{-6}$; Table 5.1). Group 1B had an absence of patients with pathological Q waves and abnormal T wave axes (all post hoc $p < 0.01$).

Table 5.1: Characteristics of the four HCM phenotypes from combined QRS and T wave clustering [mean $\pm$ standard deviation, median (range) or number of patients (%)]

	Group 1A (n=21)	Group 1B (n=24)	Group 2 (n=19)	Group 3 (n=22)	p value (group comparison)
Standard ECG features					
Heart rate, bpm	58 $\pm$ 10	58 $\pm$ 8	59 $\pm$ 13	58 $\pm$ 10	0.66
QRS axis, °	14 $\pm$ 30 ^c	44 $\pm$ 29 ^e	29 $\pm$ 34 ^f	-37 $\pm$ 28	8 $\times 10^{-13}$
QRS duration, ms	94 $\pm$ 8	100 $\pm$ 21	96 $\pm$ 16	102 $\pm$ 13	0.23
QRS amplitude, mV	2006 $\pm$ 780	1762 $\pm$ 621	1807 $\pm$ 588	1737 $\pm$ 613	0.70
QRS ascending slope	93 $\pm$ 36	95 $\pm$ 33	92 $\pm$ 34	76 $\pm$ 23	0.19
QRS descending slope	-160 $\pm$ 60	-139 $\pm$ 56	-155 $\pm$ 60	-149 $\pm$ 56	0.47
Pathological Q waves	3 (14)	0 [†]	8 (38)	9 (43)	0.0003
T wave axis, °	156 $\pm$ 45 _{a,b,c}	42 $\pm$ 27	49 $\pm$ 47	68 $\pm$ 41	2 $\times 10^{-11}$
Abnormal T axis	17 (81) [†]	0 [†]	6 (33)	5 (25)	7 $\times 10^{-8}$
T amplitude, mV	-303 $\pm$ 272 _{a,b,c}	429 $\pm$ 198	325 $\pm$ 298	401 $\pm$ 250	1 $\times 10^{-9}$
T wave inversion	21 (100) [†]	0	4 (21)	1 (5)	8 $\times 10^{-17}$
Giant T wave inversion	5 (24) [†]	0	1 (5)	0	0.003
T peak to T end, ms	91 $\pm$ 17	80 $\pm$ 20	86 $\pm$ 16	84 $\pm$ 19	0.19

ST segment displacement, mV	-13 ± 51 <i>a,b,c</i>	32 ± 33	38 ± 51	48 ± 38	0.0001
QTc interval, ms	452 ± 22	435 ± 26	429 ± 26	443 ± 29	0.05
JTc interval, ms	359 ± 47	349 ± 87	342 ± 97	366 ± 145	0.42

HCM, hypertrophic cardiomyopathy; SCD, sudden cardiac death; NSVT, non-sustained ventricular tachycardia; BPR, blood pressure response; LVH, left ventricular hypertrophy; LV, left ventricular; G+LVH-, genotype positive HCM with normal wall thickness; EDV, end-diastolic volume; ESV, end-systolic volume; LVOT, left ventricular outflow tract. ^a Group 1A vs 1B, ^b 1A vs 2, ^c 1A vs 3, ^d 1B vs 2, ^e 1B vs 3, ^f 2 vs 3, *p*<0.05 on post hoc pairwise comparisons (*p*-values multiplied by 6 for Bonferroni adjustment of 6 tests). * † *p*<0.05 on post hoc contingency table analysis (*p*-values multiplied by 16 and 8 for Bonferroni adjustment of 4×4 and 4×2 combinations respectively).

The KLT coefficients were also analysed in light of the four groups identified. Figure 5.6 shows the distribution of the coefficients in the four groups. Interestingly, Group 1A showed significant differences compared to the other groups, exhibiting similar distribution, in all leads especially in the 1st and 3rd coefficients, mostly accounting for the polarity of the ST-T segment in the KL reconstruction. This was in agreement with the specificity of Group 1A to exhibit T wave inversions compared to the other groups.

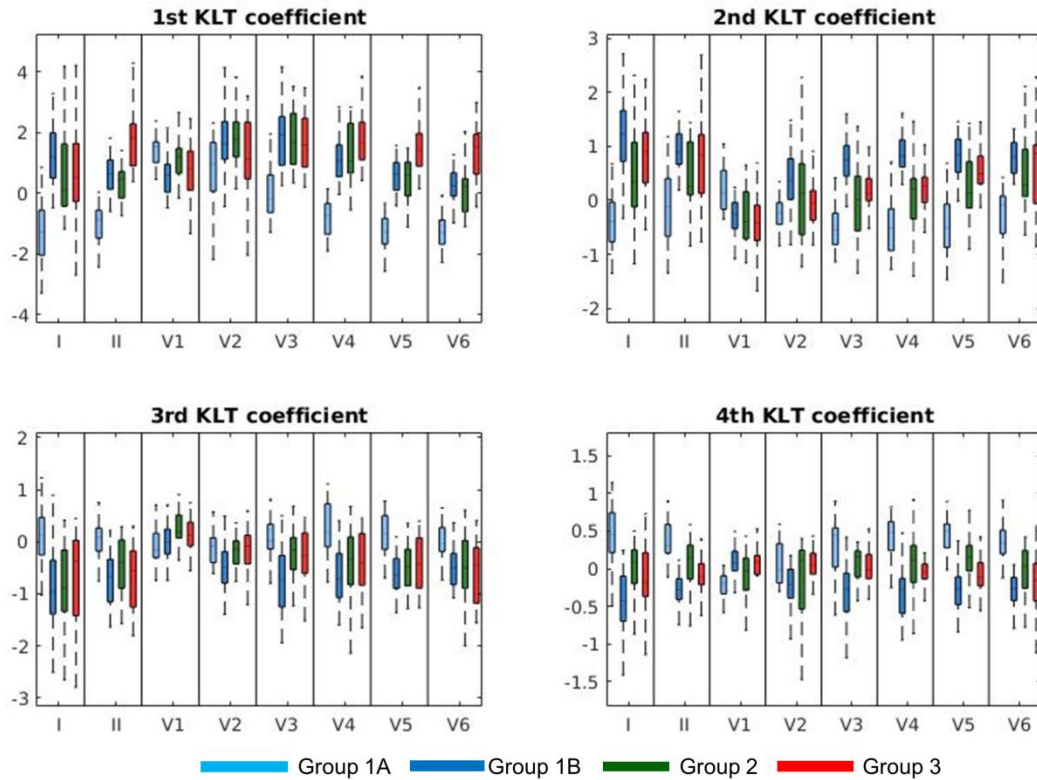


Figure 5.5: Four coefficients of the KL transform for Groups 1A, 1B, 2 and 3 in leads I,II, V1-V6.

5.3.2 Association with MRI and arrhythmic risk factors

Given the low incidence of SCD in HCM, with an annual SCD rate less than 1% (O'Mahony, Elliott, and McKenna 2013), the four HCM subgroups were further evaluated using arrhythmic risk markers and the HCM Risk-SCD score established clinically as well as features derived from cardiac resonance imaging. Clinical markers are reported in Table 5.2 and the main findings related to the four groups are illustrated on Figure 5.6.

Group 1A had the highest median HCM Risk-SCD score (1A: 3.8%, 1B: 1.8%, 2: 2.1%, 3: 2.5%; 1A vs 1B, 1A vs 2, 1A vs 3, all post hoc $p \leq 0.02$; Figure 5.7, Panel C). There was a trend to increased NSVT in Group 1A (NSVT differed between the 4 phenotypes, $p=0.025$ but this was not significant when corrected for the three other risk factors also tested). When considering Group 1A relative to the rest of the HCM cohort, the odds ratio of

NSVT in Group 1A was 4.9 (95% confidence interval, 1.7 to 14.0; $p=0.004$, Figure 5.7, Panel C). This group also contained the only ICD patient with an appropriate shock. Hypertrophy was increased in this group as demonstrated by greater wall thickness than Group 1B (1A: 22mm, 1B: 16mm; 1A vs 1B, post hoc $p=0.002$), increased mass index than Groups 1B and 2 (1A: 90g/m², 1B: 62g/m²; 2: 63g/m²; 1A vs 1B and 1A vs 2, all post hoc $p<0.004$), and the absence of G+LVH- patients. Interestingly, Group 1A exhibited a predominance of patients with mixed septal and apical hypertrophy (1A: 11 patients, 1B: 1 patient, 2: 1 patient, 3: 1 patient; post hoc $p=4\times10^{-6}$; Figure 5.7, Panel B).

Group 1B had less hypertrophy and contained the most G+LVH- patients (88%, Table 5.2, post hoc $p=0.0003$). This was consistent with their normal QRS and T wave morphologies.

Group 3 patients had a similar magnitude of hypertrophy to Group 1A (wall thickness, 3: 21mm, 1A: 22mm; mass index, 3: 73g/m², 1A: 90 g/m²; 3 vs 1A, all post hoc $p=\text{non-significant}$). In spite of their more abnormal QRS morphology, Group 3 patients had a lower HCM Risk-SCD score than Group 1A patients (3: 2.5%, 1A: 3.8%; 3 vs 1A; post hoc $p=0.02$). Group 3 patients mainly had septal hypertrophy (90%).

There were no associations seen with other clinical features (LV volumes, ejection fraction, left atrial diameter or LVOT gradient) or genotype across the groups ($p=\text{non-significant}$).

Table 5.2: Characteristics of the four HCM phenotypes from combined QRS and T wave clustering [mean $\pm$ standard deviation, median (range) or number of patients (%)]

	Group 1A (n=21)	Group 1B (n=24)	Group 2 (n=19)	Group 3 (n=22)	p value (group comparison)
Age, years	47 $\pm$ 12	48 $\pm$ 18	44 $\pm$ 12	43 $\pm$ 13	0.53
Male	16 (76)	15 (63)	12 (63)	16 (73)	0.70
Body mass index, kg/m ²	28 $\pm$ 6	26 $\pm$ 4	25 $\pm$ 4	25 $\pm$ 4	0.06
HCM Risk-SCD score, %	3.8 (1-11) ^{a,b,c}	1.8 (1-4)	2.1 (1-6)	2.5 (1-9)	0.0004
Conventional risk factors					
NSVT	11 (52) [*]	3 (13)	4 (21)	5 (23)	0.025
Syncope	2 (10)	4 (17)	2 (10)	2 (9)	0.88
Family History SCD	5 (24)	2 (8)	5 (26)	5 (23)	0.38
Abnormal exercise BPR	2 (10)	3 (12)	0	2 (9)	0.51
Massive LVH ≥ 30 mm	0	0	0	1 (5)	-
Appropriate ICD shock	1	0	0	0	-
Hypertrophy					
LV mass index, g/m ²	90 $\pm$ 27 ^{a,b}	62 $\pm$ 23	63 $\pm$ 17	73 $\pm$ 26	0.0003
Max LV wall, mm	22 $\pm$ 4 ^a	16 $\pm$ 5	19 $\pm$ 5	21 $\pm$ 6	0.002
Hypertrophy morphology					3 $\times 10^{-7}$
No LVH (G+LVH-)	0	8 (33) [*]	1 (5)	0	
Septal LVH	8 (38) [*]	15 (63)	16 (85)	20 (90)	
Apical LVH	2 (10)	0	1 (5)	1 (5)	

Mixed septal & apical LVH	11 (52)*	1 (4)	1 (5)	1 (5)	
Other clinical features					
LV EDV, ml	150 ± 26	154 ± 31	155 ± 33	156 ± 41	0.94
LV ESV, ml	35 ± 10	43 ± 15	43 ± 14	41 ± 21	0.26
LV ejection fraction, %	76 ± 7	72 ± 8	72 ± 8	74 ± 7	0.28
Left atrial diameter, mm	41 ± 5	39 ± 8	37 ± 6	38 ± 8	0.25
LVOT gradient, mmHg	6.5 (5-110)	7.2 (4-111)	6.7 (2-110)	6.8 (3-92)	0.81
Genotype					
Gene positive	10 (48)	19 (79)	13 (68)	17 (77)	0.11

HCM, hypertrophic cardiomyopathy; SCD, sudden cardiac death; NSVT, non-sustained ventricular tachycardia; BPR, blood pressure response; LVH, left ventricular hypertrophy; LV, left ventricular; G+LVH-, genotype positive HCM with normal wall thickness; EDV, end-diastolic volume; ESV, end-systolic volume; LVOT, left ventricular outflow tract. ^a Group 1A vs 1B, ^b 1A vs 2, ^c 1A vs 3, ^d 1B vs 2, ^e 1B vs 3, ^f 2 vs 3, $p < 0.05$ on post hoc pairwise comparisons (p-values multiplied by 6 for Bonferroni adjustment of 6 tests). * † $p < 0.05$ on post hoc contingency table analysis (p-values multiplied by 16 and 8 for Bonferroni adjustment of 4×4 and 4×2 combinations respectively).

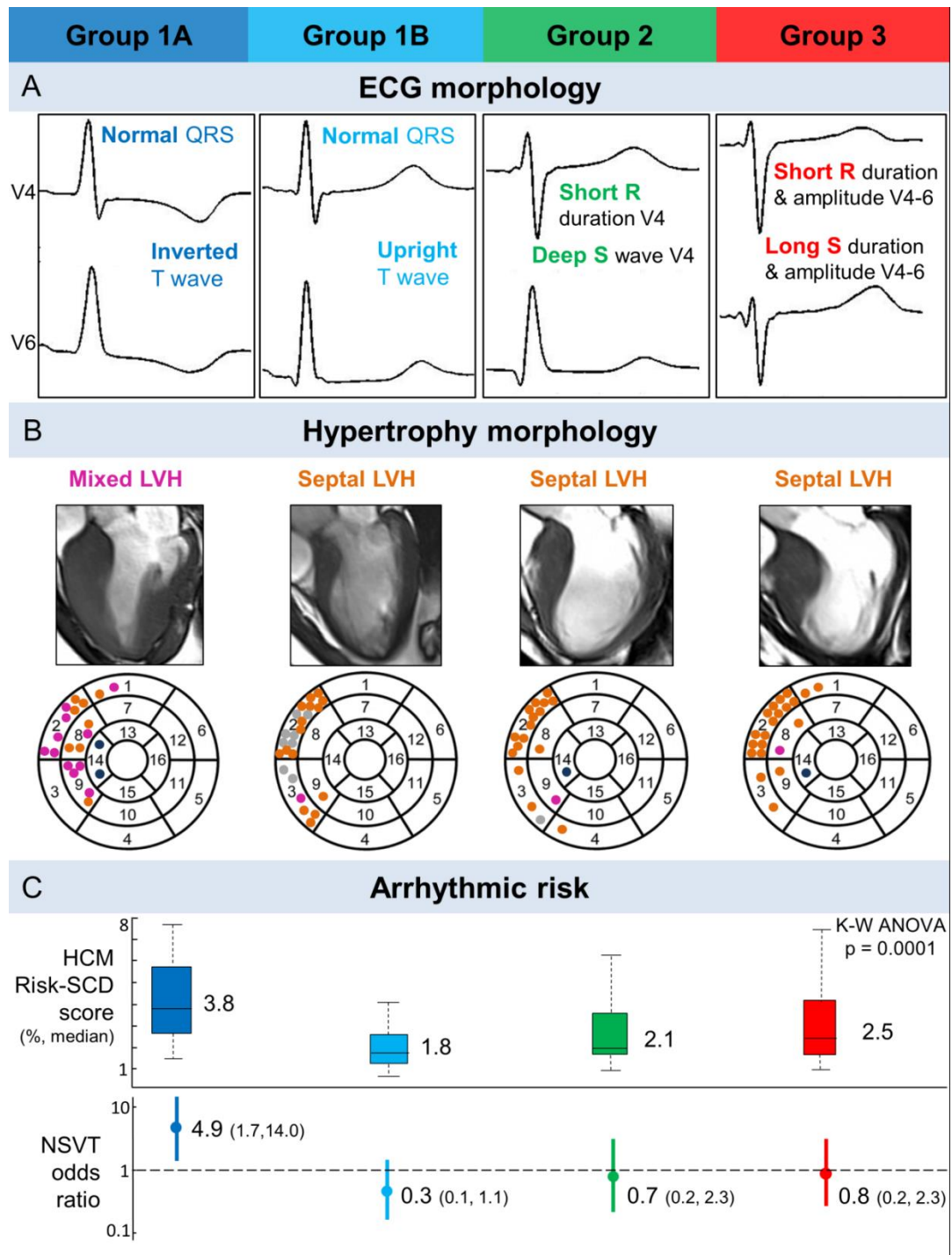


Figure 5.6: Four distinct ECG phenotypes in hypertrophic cardiomyopathy exhibit differences in hypertrophy morphology and arrhythmic risk. (A) Four groups with distinct ECG morphology are identified by combined clustering with QRS morphology and T wave biomarkers. Group 1A – normal QRS with inverted T wave and abnormal T wave axis deviation (primary repolarization abnormality), Group 1B – normal QRS with upright T wave, Group 2 – short R wave duration and deep S wave in V4, Group 3 – left axis deviation, short R wave duration and amplitude, and long S wave duration and amplitude in V4 and V6. (B) Group 1A had a predominance of patients with mixed septal and apical left ventricular hypertrophy (pink). Group 1B had the most genotype

positive patients with no hypertrophy (grey). Group 2 and 3 patients mainly had septal hypertrophy (orange). Four patients had apical hypertrophy (navy). (C) Patients with a primary repolarization abnormality – normal QRS with T wave inversion and abnormal T wave axis deviation (Group 1A), had (top) an increased HCM Risk-SCD score (3.8%) and (bottom) an increased odds ratio of non-sustained ventricular tachycardia relative to the rest of the HCM cohort (odds ratio 4.9, 95% confidence interval, 1.7 to 14.0, $p=0.004$).

5.4 Discussion

In this chapter, we identified four different HCM phenotypes based on a combination of QRS and T wave biomarkers analyzed computationally. The groups showed differences in HCM Risk-SCD score, non-sustained ventricular tachycardia and LV hypertrophy from CMR.

Lateral T wave inversion, a common repolarization abnormality in HCM (Papadakis et al. 2009), was the critical T wave biomarker which when added to QRS morphology in the clustering analysis, identified four HCM groups with differences in clinical features. Group 1A patients, even though with a normal QRS, had a higher HCM Risk-SCD score, a trend to increased NSVT, an appropriate ICD shock and a predominance of mixed septal and apical pattern of hypertrophy. In the presence of normal depolarization and normal QRS axis, the T wave inversion and abnormal direction of T wave axis in Group 1A patients constitute a primary repolarization abnormality (Rautaharju et al. 2001) (Kors et al. 1998). Patients in Groups 2 and 3 had T wave inversion in much smaller numbers (4 and 1 patients respectively) than Group 1A. These represent secondary or combined primary and secondary repolarization abnormalities given that these patients had abnormal depolarization and QRS complexes (Rautaharju, Surawicz, and Gettes 2009).

T wave inversion per se in HCM has been shown to increase SCD risk in some studies (Ostman-Smith et al. 2010) (Kuroda et al. 2002) but not in others (Sherrid et al. 2009) (Maron, Roberts, and Epstein 1982). Our results suggest that it is a primary repolarization abnormality that increases SCD risk in HCM, rather than T wave inversion alone which can be secondary to depolarization abnormalities. A larger cohort is needed to confirm whether risk differs between patients with primary and secondary repolarization changes. In addition, simple descriptors of the T wave morphology such as polarity were spread across different KL basis. We should note that the KL transform characterization proposed is based on the ST-T segment (as a continuation from the QRS complex analysed in Chapter 4). This waveform is very sensitive to changes in heart rate, which may lose the morphological details of the T wave when averaged. Future work could therefore focus on characterizing the T wave morphology in more details, by extracting a window around the T wave and aligning the waveforms on the T wave peak.

In this work, we used Hermite bases to characterize the QRS complex because of their good performances to model its morphology as discussed in section 4.3.3, and the KL transform to model ST-T segments. Hermite bases could be used to similarly characterise the T wave, and future work could investigate the potential of the 1st and 2nd bases in modelling inverted and biphasic T wave morphologies.

Clinically, T wave inversion may be caused by repolarization abnormalities due to structural and ionic remodeling. In HCM, the overexpression of the L-type Ca^{2+} current, increased late sodium current and reduction of repolarization currents lead to prolongation and heterogeneity in repolarization (Coppini et al. 2013) (Passini et al. 2016). T wave inversion may also be the result of ischaemia from microvascular dysfunction commonly seen in HCM (Petersen et al. 2007). Petersen et al. demonstrated that reduced endocardial perfusion was proportional to the magnitude of hypertrophy. Ischaemia may explain the T

wave inversion seen in Group 1A patients as they had greater hypertrophy, which is also consistent with an increased sensitivity of T wave inversion in diagnosing HCM as LVH increases (Konno et al. 2007). However, Group 3 patients had a comparable magnitude of hypertrophy to Group 1A, yet there was no increased SCD risk, suggesting that additional factors may be implicated with increasing risk. A further study has shown that patients with apical HCM and cavity obliteration had increased perfusion defects and NSVT rates as a result of ischaemia from extravascular compression of the coronary artery due to myocardial pressure during cavity obliteration (Matsubara et al. 2003). This may also be in keeping with Group 1A having the greatest number of patients with a mixed pattern of coexisting septal and apical hypertrophy which tend to show cavity obliteration and a worse prognosis (Yan et al. 2012); while Group 3 with 90% septal hypertrophy, were at lower risk despite having similar wall thickness and mass index to 1A. Further imaging studies with novel perfusion assessment may improve our understanding of the mechanisms of ischaemia and in turn, the repolarization abnormalities in HCM.

Interestingly, patients in Group 1B (28%) with normal QRS and upright T waves were indistinguishable from controls based on ECG features and had less LVH. These patients are likely to have less severe disease and better prognosis (McLeod et al. 2009). We may also speculate that they will have minimal fibrosis, disarray and ischaemia giving rise to relatively normal depolarization and repolarization. Group 3 on the other hand, had marked QRS abnormalities: LAD with QRS differences in V4 to V6. LAD is well-known to be associated with LVH but other factors such as a degree of left anterior fascicular block could also account for this leftward axis. Group 3 patients mainly had isolated septal hypertrophy, yet QRS abnormalities were seen in the lateral leads suggesting that remodeling may occur distal from the septum causing less uniform electrical propagation

in the lateral leads. No genotype association was seen across any group but our sample size may not have had adequate power to assess genotype-phenotype correlations.

This study provides evidence that ECG phenotyping with advanced computational QRS morphology and T wave analysis is a powerful method of characterizing HCM heterogeneity. Data suggest that HCM patients with a primary repolarization abnormality as manifested by T wave inversion and abnormal T wave axis deviation but normal QRS are at greater risk of arrhythmia and SCD in our low to moderate risk cohort. This risk was associated with the distribution of hypertrophy (greater number of segments involved in mixed septal and apical pattern of hypertrophy) rather than magnitude of hypertrophy (as measured by maximum wall thickness or mass index). A large scale longitudinal study with cardiovascular end-point data will allow robust assessment of ECG phenotyping as an independent tool for accurate risk stratification. Computational image-based modelling and simulation, implemented in Chapter 6, may also help disentangling the relative contribution of structural, ischemic and ionic factors which are likely to determine the heterogeneity in ECG biomarkers (Dutta et al. 2016). This better understanding of HCM will hopefully contribute to the improvement of current therapies.

5.5 Conclusions

In this chapter, four HCM phenotypes were identified based on QRS morphology and T wave biomarkers using a combination of clustering and mathematical modelling. Patients with a primary repolarization abnormality as manifest by T wave inversion and abnormal T wave axis deviation not secondary to QRS abnormalities had an increased HCM Risk-SCD score, a trend to increased NSVT, and coexistence of septal and apical hypertrophy. As a next step in understanding HCM heterogeneity, Chapter 6 will explore the structural and electrophysiological mechanisms that may explain the distinct ECG phenotypes identified in this chapter. These results, and the nature of the underlying processes captured by the

ECG, suggest that computational ECG phenotyping has the potential to be a novel and independent risk factor for clinical stratification.

6 INTERPRETATION OF ECG PHENOTYPES USING PERSONALIZED CMR-BASED WHOLE-BODY 3D SIMULATIONS

6.1 Introduction

Chapter 4 and 5 presented the identification of different phenotypes in HCM based on electrographic features from QRS and T wave morphologies. This chapter follows up on this result by exploring the connection between these ECG abnormalities and the presence of structural and electrophysiological changes in these patients. The image processing work and the generation of surface meshes presented in this chapter was performed by Dr Ernesto Zacur (Institute of Biomedical Engineering, University of Oxford).

6.2 Motivation

In Chapter 5, we identified four different HCM phenotypes exhibiting differences in ECG morphologies and associated with different hypertrophy distribution and risk of arrhythmia. The structural and electrophysiological mechanisms leading to such electrocardiographic abnormalities in HCM remain unclear. The aim of this chapter is therefore to investigate the influence of structural changes (such as hypertrophy, or fibre disarray) and electrophysiological abnormalities (ionic remodelling) on the ECG to understand the specific ECG changes observed in these HCM phenotypes. We hypothesise that structural changes such as hypertrophy and fibre disarray may explain the QRS abnormalities identified in Group 3, by altering the electrical activation in these patients' hearts, and that the dysregulation of ionic currents observed in HCM (Coppini et al. 2013) could explain the repolarization abnormalities (inverted T waves) highlighted in Group 1A. This work uses a computational modelling framework allowing personalized whole-body simulations of the electrocardiogram from cardiac magnetic resonance (CMR) images. As described in Chapter 3, CMR-based models allow the use of realistic geometries, constructed from clinical cardiac magnetic resonance images to build the mesh anatomy. This approach improves the clinical accuracy of the model and allows the personalization of the simulations in order to address the challenges of human variability. In this thesis, the consideration of CMR data allows a personalized analysis of the distinct phenotypes identified in HCM. Whole-organ cardiac computer simulations have already been implemented in previous studies (Dutta et al. 2016) (Cardone-Noott et al. 2016) (Kayvanpour et al. 2015) to understand multiscale mechanisms in the heart. This motivates their use in this project, to explore the mechanisms linking structural and electrophysiological abnormalities with altered electrical propagation in hypertrophic cardiomyopathy, based on CMR imaging data.

6.3 Methods

The whole-body simulation pipeline implemented in this work takes as input cardiac magnetic resonance (CMR) images of specific HCM patients and allows the simulation of the electrical propagation in these patients, as well as the recording of the electrocardiogram (ECG). Modifying structural and electrophysiological parameters in the initial pipeline configuration then enables to investigate the influence of these parameters on the cardiac electrical activity and the ECG.

6.3.1 3D ventricular mesh generation from cardiac magnetic resonance images

The section below describes the computation of 3D ventricular meshes representing the patients' anatomy from CMR images.

6.3.1.1 Mesh generation

First, four representative patients were chosen, based on their ECG and CMR images, to represent each of the HCM phenotypes identified in Chapter 5. The choice of these patients was firstly based on the presence of the ECG abnormalities characteristic of the phenotypes identified (Group 1A: normal QRS, inverted T waves in V4-V6, Group 1B: normal QRS, upright T waves in V4-V6, Group 2: short R duration, deep S wave in V4, Group 3: short R duration and amplitude, long S duration and amplitude in V4-V6). In addition, we considered the quality of the acquired CMR images to allow an accurate 3D representation. For example, male torsos were easier to model, and a large field of view, with sufficient contrast, and accurate contouring of the ventricle were required. The study focused on the interpretation of the ECG of patients from Group 1A and 3, exhibiting the most extreme and interesting ECG abnormalities.

The first step of the pipeline consisted in constructing 3D surfaces of the heart, torso, lungs and ribs from CMR images. The bi-ventricular heart geometries were derived directly from

CMR by image processing. MRI slices of the heart (long and short axis) were recorded in the clinic, and pre-processing steps such as alignment, registration, and manual contour delineation were applied in order to generate 3D surfaces of the left and right ventricles corresponding to the patient's CMR images and anatomy. Torso and lungs surface geometries were derived from combining segmentation localizers with a statistical model of the torso. Ribs were adapted from a template in order to fit the patient's anatomy. This part of the work was carried out by Dr Zacur.

The next step was to mesh these surfaces in order to obtain fully connected 3D volumetric meshes. From the 3D surfaces obtained in the previous step, 3D volumetric meshes were constructed using the meshing software Tetgen (Si 2015). 3D meshing provided a volumetric mesh composed of tetrahedral elements. Each tetrahedron was defined by four nodes (Figure 6.1, Panel A) and allowed the propagation of the electrical signal throughout the mesh. The meshing process required the definition of a tetrahedral edge length and multiple optimization parameters. In our case, a fine and detailed mesh was required in the heart in order to solve the complex propagation equations (described in Chapter 3 section 3.4.3) with good accuracy. Torso, lungs and ribs were designed as passive tissues with different fixed conductivities, allowing a coarser mesh to be defined in these regions (following (Cardone-Noott et al. 2016)'s work). The edge length of each element was defined as 0.4mm for the heart, 3mm for the lungs, 2mm for the ribs and 10mm for the torso based on (Cardone-Noott et al. 2016) and (Niederer et al. 2011)'s benchmarking work. An adaptive meshing, providing smooth transitions between different edge lengths, was implemented by controlling the volume of the tetrahedra in order to ensure a smooth propagation of the electrical signal in the whole domain.

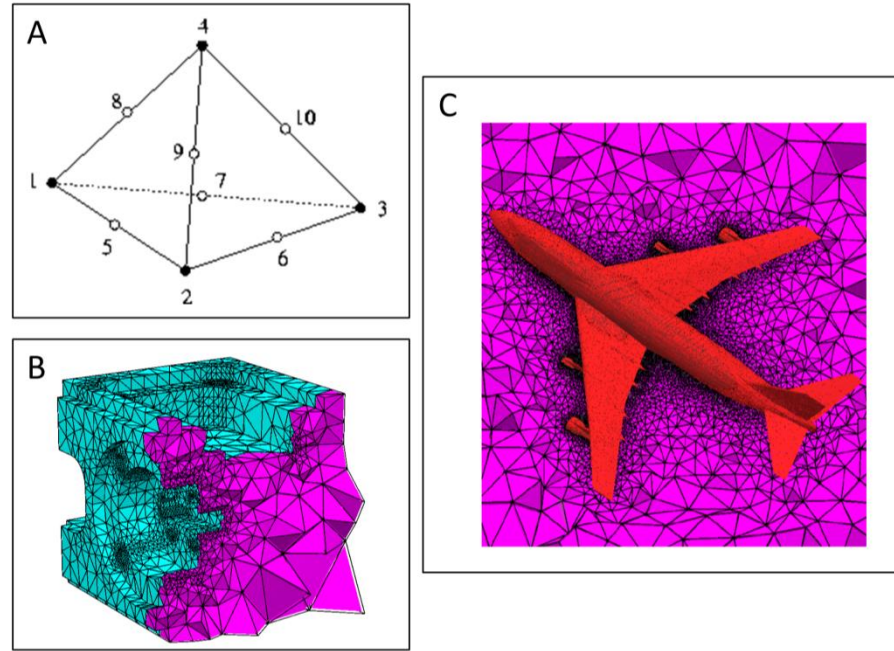


Figure 6.1: Volumetric mesh generation. Panel A: 3D mesh tetrahedron element (4 nodes and 6 edges). Panel B: example of a 3D volumetric mesh generated using Tetgen. Panel C: adaptive meshing in Tetgen to adapt to smaller edge length (From Weierstrass Institute, Berlin)

6.3.1.2 Mesh quality assessment

The quality of the Tetgen-generated meshes was assessed using the following different quality metrics for each tetrahedral element. Well-behaved elements usually have metrics in the acceptable ranges described (Knupp et al. 2006).

- Edge ratio, defined as the ratio between the length of the longest edge and the shortest edge of the element L_{max}/L_{min} . The acceptable range is between $[1,3]$.
- Aspect gamma, comparing the root-mean-square edge length $R = \sqrt{\frac{\sum_{i=0}^5 \|L_i\|^2}{6}}$ to the volume V of the element. The acceptable range is between $[1,3]$.
- Minimum dihedral angle, defined as the (nonoriented) dihedral angle of two neighbouring faces of the element that are adjacent along edge i :

$\alpha_i = \frac{180^\circ}{\pi} \arccos(n_{i1}, n_{i2})$ with n_{i1}, n_{i2} normal unit vectors to the two faces. The acceptable range is between $[40^\circ, 180/\pi * \arccos(1/3) \approx 70.5^\circ]$.

- Radius ratio, defined as the normalized ratio of the radius of the inscribed sphere to the radius of the circumsphere. The acceptable range is between $[1,3]$.

Figure 6.2 below summarizes the different quality metrics for a volumetric mesh of edge length 4mm generated with Tetgen.

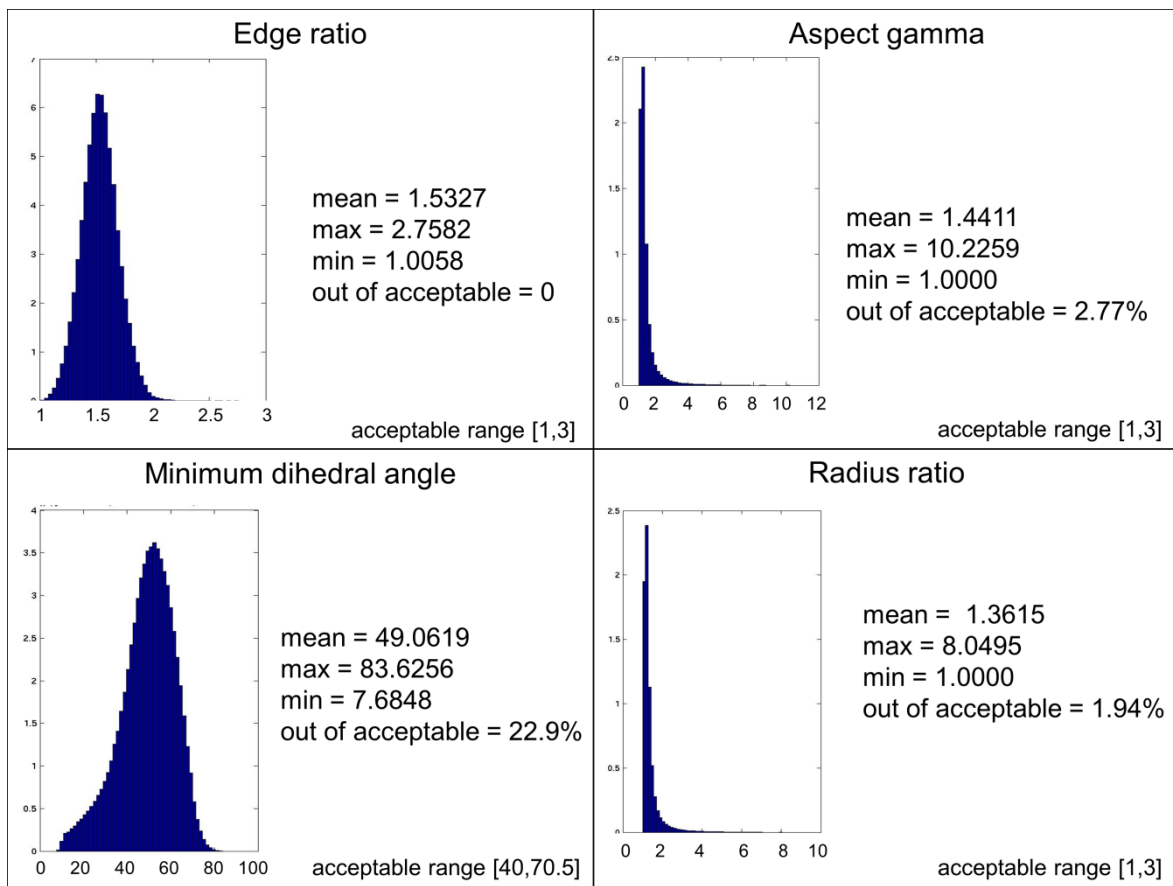


Figure 6.2: Histograms of edge ratio, aspect gamma, minimum dihedral angle and radius ratio for the Tetgen-generated mesh of edge length 4mm. Mean metric, maximum and minimum values, as well as the number (%) of elements out of the acceptable range are also indicated.

6.3.2 Cardiac electrophysiology modelling and ECG simulation

The 3D biventricular volumetric meshes were therefore composed of a set of nodes (vertices of the tetrahedra, defined by their coordinates) and elements (tetrahedra, defined by four nodes). The electrophysiological simulation described below followed the framework published in (Cardone-Noott et al. 2016). The electrophysiological (O’Hara et al. 2011) model of ventricular cells was used to define cellular activity. Intracellular conductivities in the active tissues were chosen in order to generate conduction velocities close to 67, 30 and 17 cm/s in the longitudinal, transversal and transmural directions (Caldwell et al. 2009). Extracellular conductivities were defined according to resistivity ratios (Clerc 1976). Extracellular and intracellular conductivities were adapted from (Cardone-Noott et al. 2016) based on a convergence analysis performed by Héctor Martínez-Navarro. As described on Figure 6.3, a scaling factor of 1.8 led to longitudinal and transverse conduction velocities within experimental ranges (67 and 30 cm/s). Initial conductivities were therefore set to (2.7, 0.81, 0.405 mS/cm) for intracellular conductivities and (9.81, 3.65 and 3.65 mS/cm) for extracellular conductivities. In passive tissues, isotropic conductivities were chosen to be 0.389 mS/cm in the lungs, 0.2 mS/cm in the bones, and 2.16 mS/cm in the rest of the torso (Gabriel, Lau, and Gabriel 1996) (Rush, Abildskov, and Mcfee 1963) (Cardone-Noott et al. 2016).

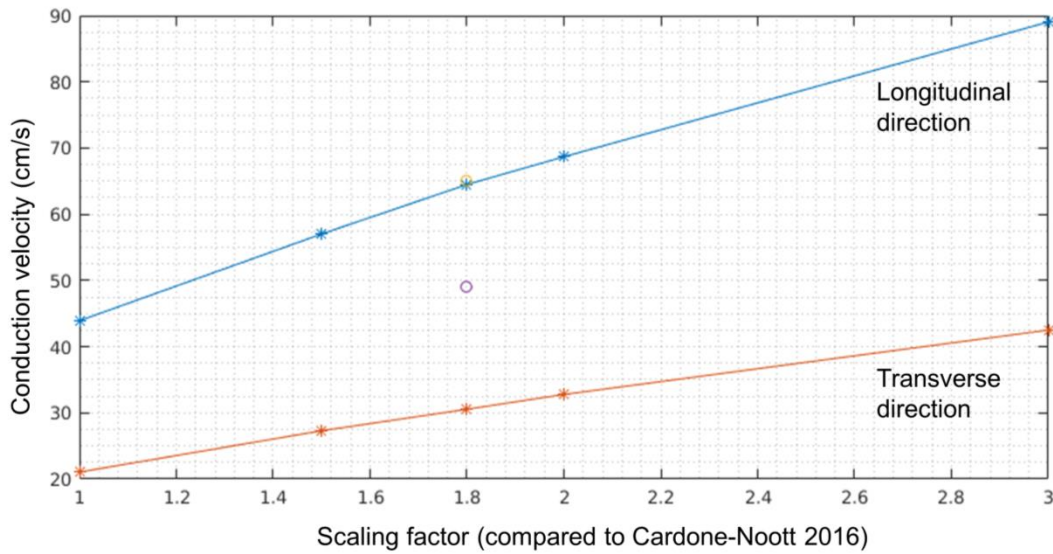


Figure 6.3: Sensitivity analysis for conductivity values performed on a 1D fibre. A scaling factor of 1.8 from Cardone-Noott 2016's values provided physiological conduction velocity close to 67, 30 and 17 cm/s. This work was carried out by Héctor Martínez-Navarro.

The activation sequence was then modelled by defining seven stimuli points (root nodes) in the endocardium (four in the left ventricle and three in the right ventricle) to initiate the activation. These endocardial sites of earliest activation modelled the role played by Purkinje branches in real physiology (Figure 6.4, Panel A). They were adapted from (Cardone-Noott et al. 2016) by Dr. Zacur, by computing a mapping to each patient's anatomy. The stimulus in these points was initially defined by a duration of 1ms, a magnitude of $120,000 \text{ mA/cm}^3$, and a speed of 0.179 cm/s. The shortest distances between each node were computed using the (Dijkstra 1959) algorithm in the left and right ventricles, in order to model the Purkinje network and define the propagation of the stimuli from the root nodes to the rest of the endocardial surface nodes. In order to simulate the ECG on the torso surface, ten nodes were defined to model electrodes placed on the torso, adapting the location defined in (Cardone-Noott et al. 2016) to each personalized torso by geometrical deformation (Figure 6.4, Panel B). Fibre orientations were computed for each geometry following a rule-based method replicating findings in (Streeter et al. 1969).

Apex-to-base electrophysiological heterogeneities were introduced to represent the dispersion of the action potential duration (APD) (as described in Chapter 3 section 3.4.4). These were achieved by modifying the delayed rectifier potassium current (IKs) conductance (Burton and Cobbe 2001). Apico-basal dispersion of activation was considered by defining a shorter action potential duration in the apex (Ramanathan et al. 2006). This gradient was modelled by increasing g_{Ks} by 50% in the apex. Apico-basal heterogeneities were sufficient to reproduce realistic T waves in all ECG leads from normal subjects, therefore controversial transmural heterogeneities, not observed in vivo but only on wedge preparation (Boukens et al. 2017) (Taggart et al. 2001) (Chapter 3 section 3.4.4) were not used in the models. The initial configuration is summarised in Figure 6.4, Panel C.

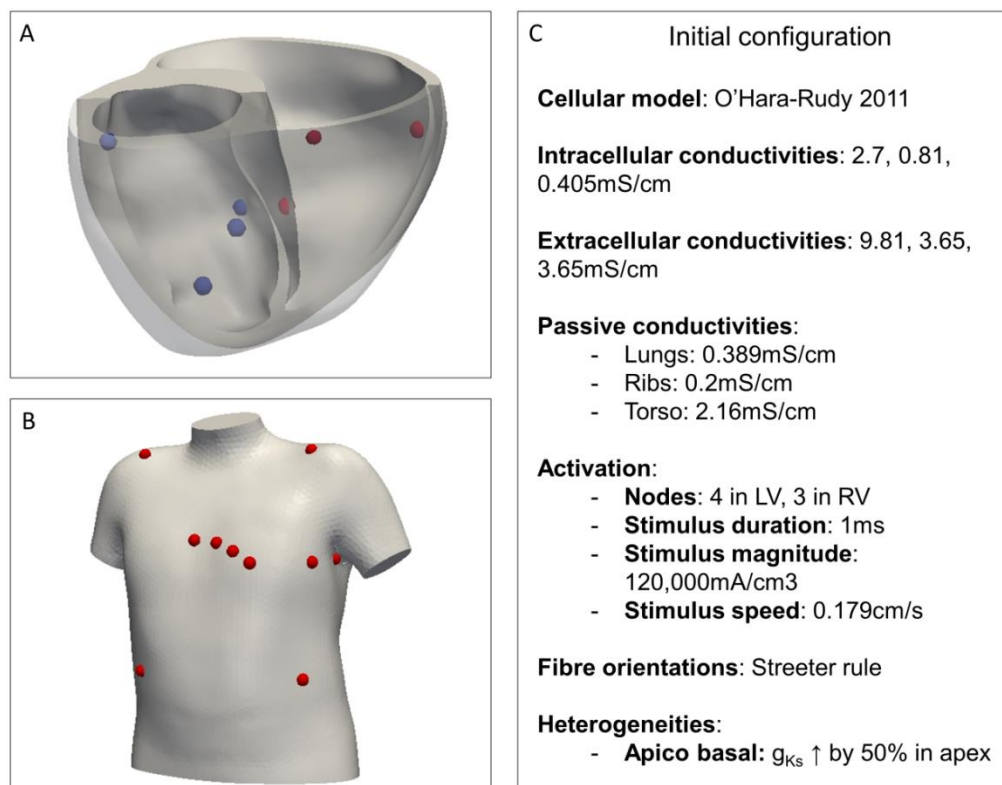


Figure 6.4: Simulation configuration. Panel A: four (resp. three) activation points are defined in the left - blue (resp. right – red) ventricle. Panel B: ten electrodes are defined to measure the ECG on the torso. Panel C: summary of the initial normal configuration of the simulations.

Simulations were run using Chaste (Pitt-Francis et al. 2009), a simulation package aimed at multi-scale and computationally expensive biomedical problems, on the Archer UK national supercomputer (<http://www.archer.ac.uk/>) using parallelization on 960 cores. Simulating 500ms of cardiac activity (duration of one heartbeat) took around 2 hours and 30 minutes, depending on the patient and more specifically, the size of the volumetric mesh. The whole running pipeline is summarized in Figure 6.5.

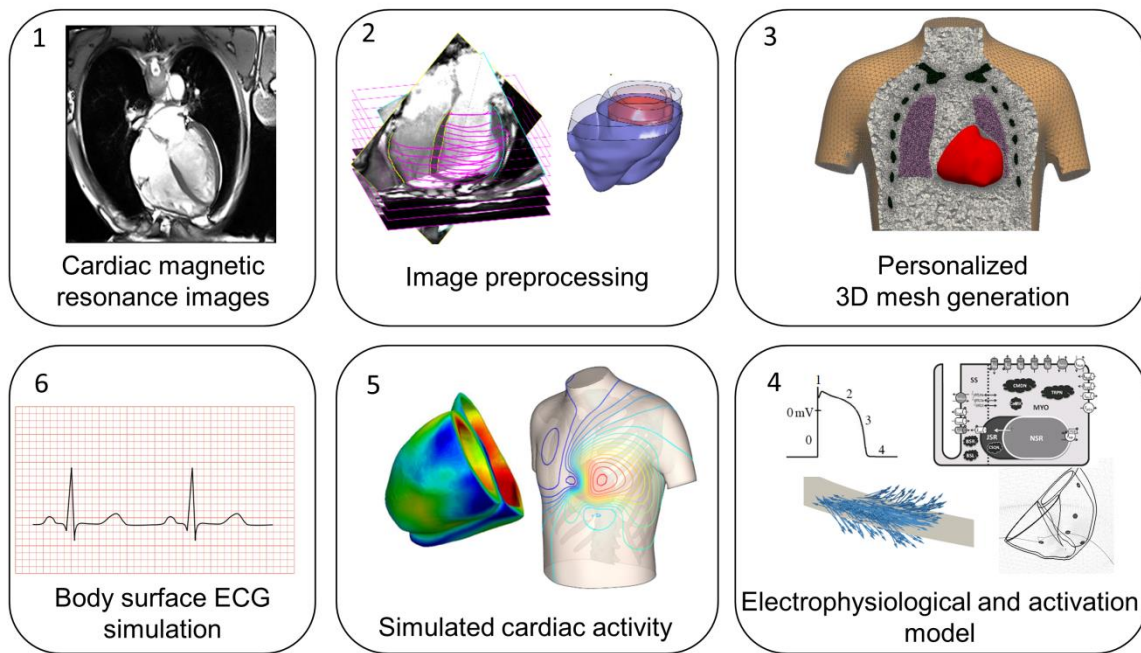


Figure 6.5: Summary of the simulation pipeline: cardiac magnetic resonance images (1) are preprocessed (2) in order to generate a personalized 3D volumetric mesh (3). An electrophysiological model is then defined (4) and the cardiac electrical activity can be simulated (5) and the ECG recorded (6).

6.3.3 Modelling of structural and electrophysiological abnormalities

In order to investigate the influence of the anatomy on the ECG in HCM patients, we simulated 500ms of cardiac electrical activity in the normal initial settings defined earlier (Figure 6.4, Panel C). In order to account for variability, tissue conductivities in the intracellular and extracellular domains were varied (as described in Table 6.1) to explore the sensitivity of the ECG to changes in conductivities.

Table 6.1: Different sets of intracellular and extracellular conductivities explored

	Intracellular conductivities (mS/cm)	Extracellular conductivities (mS/cm)
Initial configuration	(2.7, 0.81, 0.41)	(9.81, 3.65, 3.65)
(IC) and (EC) x 2	(5.40, 1.62, 0.82)	(19.62, 7.3, 7.3)
(IC) and (EC) x 0.5	(1.35, 0.41, 0.21)	(4.91, 1.83, 1.83)

IC: intracellular conductivities, EC: extracellular conductivities

Electrophysiological heterogeneities were investigated by introducing changes in the ion channels conductances. We hypothesised that inverted T waves in the lateral leads in Group 1A patients may be due to an increased action potential duration (APD) in apical cells compared to the base, leading to an inversion of the repolarization currents. Indeed, upright T waves in a normal ECG are explained by opposite directions of depolarization and repolarization (Franz et al. 1987) (Gima and Rudy 2002). Forcing the repolarization direction to be the same as the depolarization one, using electrophysiological changes, may therefore lead to inversions of the T waves. This hypothesis was supported by the HCM cardiomyocytes experimental phenotype, exhibiting longer APD due to changes in calcium, potassium and late sodium currents (Coppini et al. 2013) (Passini et al. 2016). This also associated with the hypertrophy distribution of Group 1A patients, the only group exhibiting coexistence of septal and apical hypertrophy (Chapter 5). We therefore modelled apico-basal heterogeneities by modifying the rapid potassium current I_{Kr} conductance between apex and base, a smaller g_{Kr} in the apex leading to a longer APD (Burton and Cobbe 2001) (Figure 6.6). We also explored the effect of modelling HCM ionic remodelling, described in (Coppini et al. 2013) in apical cells. Specific changes in ionic properties are defined as in Table 6.2 below.

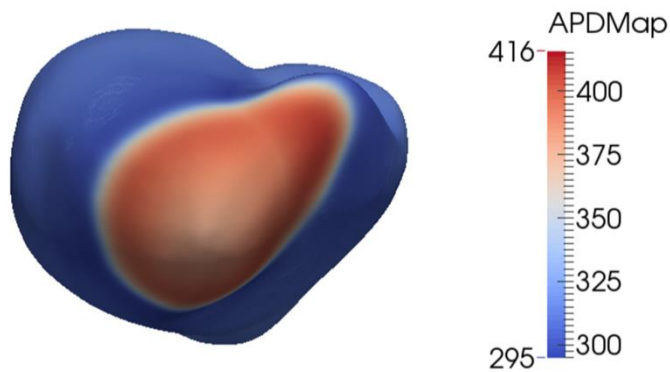


Figure 6.6: Action potential duration map showing longer action potential duration in the apical region with hypertrophy.

Table 6.2: Ionic electrophysiological remodelling in HCM as described in (Passini et al. 2016)

Ionic Current	HCM	Ionic Current	HCM
I_{Na}	-	I_{to}	-70%
I_{NaL}	+165%	I_{K1}	-30%
I_{Nab}	+165%	I_{Kr}	-45%
I_{CaL}	+40%	I_{Ks}	-45%
τI_{CaL} fast inactivation	+35%	I_{NaK}	-30%
τI_{CaL} slow inactivation	+20%	J_{Up}	-25%
I_{NCX}	+30%	J_{Rel}	-20%

Finally, fibre disarray was modelled using information from diffusion tensor imaging. Cardiac diffusion tensor imaging (DTI), presented in Chapter 2 section 2.3.4, is an imaging method that measures the diffusion of water in the myocardium and enables to compute the magnitude, degree of anisotropy and the orientation of diffusion anisotropy (Zhukov and Barr 2003). The tensor of fibres orientation obtained provides information on the structure of the myocardium. For example, a low fractional anisotropy (FA), resulting in a more isotropic diffusion, may illustrate disarray in the fibres of these regions, fibrosis, myocytes enlargement, or increased spacing between the myocytes. DTI maps and measurements for the Group 3 patient exhibiting QRS abnormalities and a reference control were compared to identify the areas of low FA and potential abnormally slow conduction. The conduction properties of the identified abnormal areas were then modified to explore the influence of the speed of the conductivities and conduction in these regions.

6.4 Results

6.4.1 Anatomy alone does not explain ECG abnormalities identified in HCM phenotypes

The whole-body simulation pipeline presented in the Methods section was implemented for the representative patients of Group 1A, Group 1B and Group 3, which presented the most interesting and extreme ECG phenotypes among those identified. 3D volumetric meshes were computed and contained 52 million elements for Group 1A patient, 35 million for Group 1B patient, and 48 million for Group 3 patient. More details on the meshes computed and the simulation time required for each patient are represented on Figure 6.7 below.

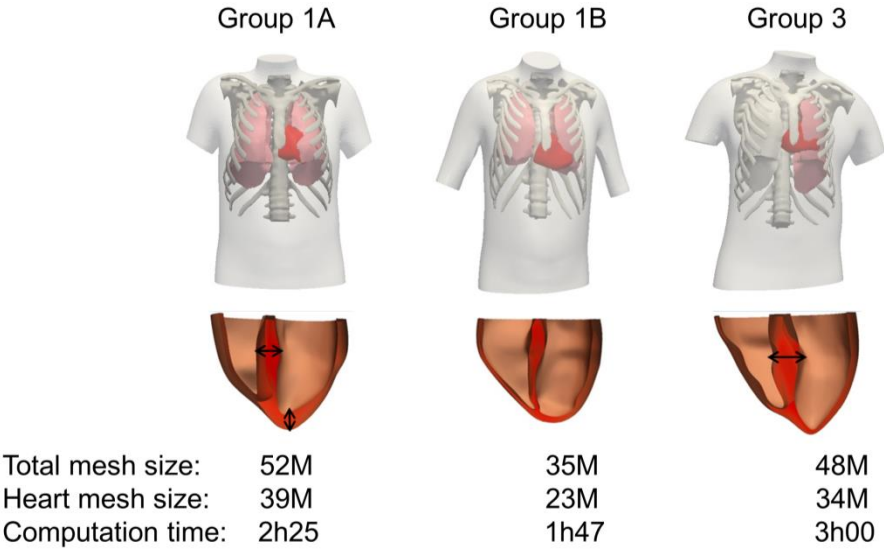


Figure 6.7: Total mesh and heart mesh for each representative patient of Groups 1A, 1B and 3. Total mesh size (number of elements), heart mesh size (number of elements) and computation time for simulating 500ms on 960 cores are also detailed.

Simulations in the normal initial configuration (as described in Figure 6.4, Panel C) were run to investigate the role of the anatomy alone on the ECG. Figure 6.8 shows the real and simulated ECG for Group 1A, Group 1B and Group 3 patient in leads V1 to V6.

The simulated QRS of Group 1A patient exhibited a similar morphology to the real QRS observed in this patient, with a normal looking evolution from leads V1 to V6. However, the simulated T waves presented a normal upright morphology in leads V4-6, while the real ECG of this patient showed inverted T waves in V4-6. The anatomy alone did not explain the repolarization abnormalities in Group 1A patient.

In Group 1B patient, the QRS complexes showed similar normal patterns compared to the real QRS in leads V1-6. The T waves also exhibited upright normal morphologies as observed the real ECG. The control simulation settings therefore reproduced the normal looking ECG morphology of Group 1B patients.

In Group 3 patient, the simulated QRS complexes in the septal leads V1-V3 exhibited a similar morphology to the real QRS. However, the simulated QRS in lateral leads V4-V6 showed a normal morphology with tall R wave and small S wave, instead of exhibiting small R duration and amplitude and abnormal deep and wide S waves as observed on the real ECG. Simulated T waves were upright, as observed on the real ECG. The anatomy alone therefore did not explain the abnormal QRS morphologies observed in this patient.

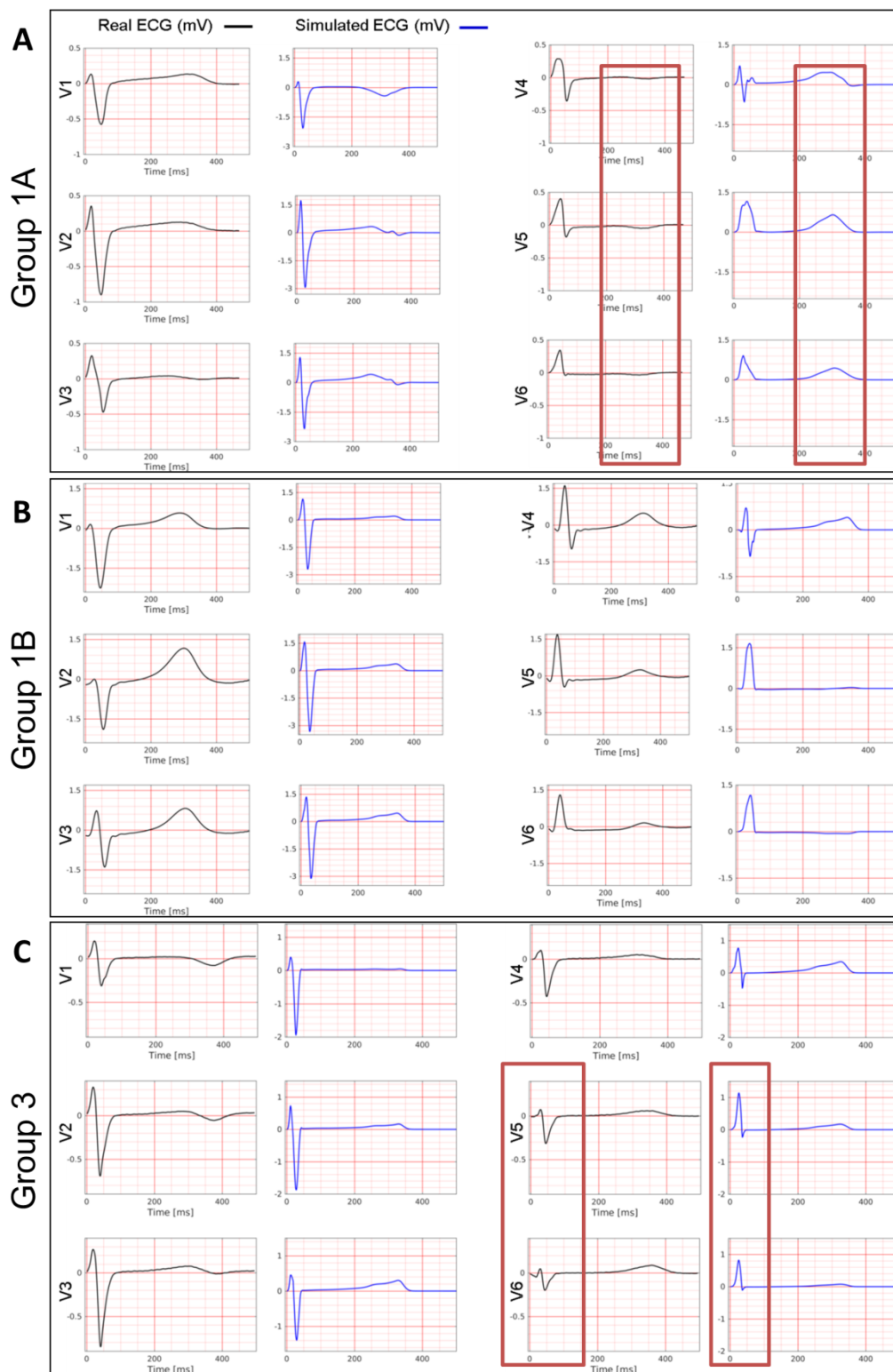


Figure 6.8: Real ECG (black) and simulated ECG (blue) of patients in Group 1A (Panel A), Group 1B (Panel B) and Group 3 (Panel C) in the normal initial configuration, for leads V1-V6. Red boxes highlight characteristic ECG abnormalities (inverted T waves in Group 1A in V4-6, deep S waves in Group 3 in V4-6) not reproduced by the simulation with normal parameters.

Figure 6.9 below compares the activation maps for patients in Group 1A, Group 1B and Group 3 in the normal initial configuration. The activation sequences were the same in the three hearts: the activation was triggered in the seven activation points defined, propagated through the ventricles and finished in the apex. These similar activation patterns confirmed that the various types of abnormalities observed on the ECG phenotypes may be due to other factors than the anatomy alone.

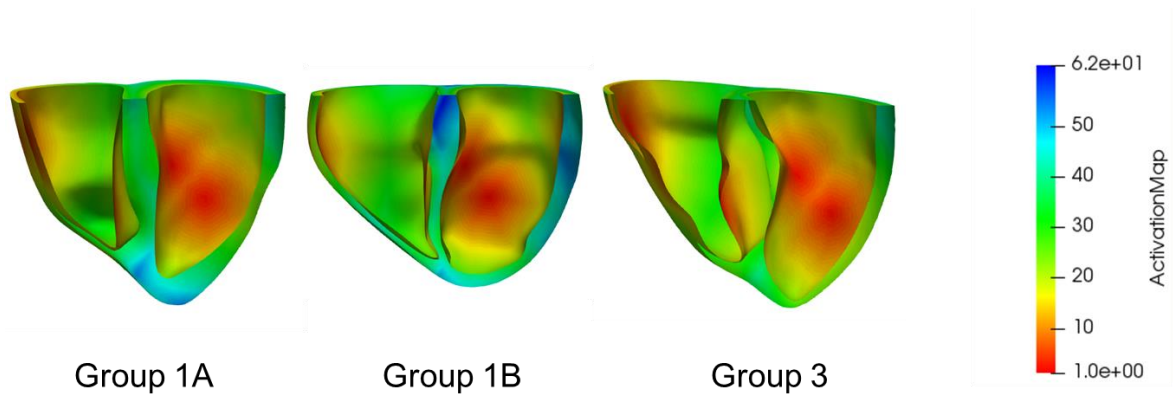


Figure 6.9: Activation maps (ms) for patients in Group 1A, Group 1B and Group 3. Activation is initiated in the seven activation nodes and propagates in the ventricles to the apex.

As a sensitivity analysis, we analysed the influence of varying both the intracellular and extracellular conductivities on the shape of the QRS, to explore whether these changes could explain the QRS abnormalities observed in Group 3 patients. Results are shown on Figure 6.10 below. Reducing the conductivities increased the amplitude and width of the QRS complexes, especially the R wave in V1-3, but did not change the morphology and polarity of the QRS waveforms, especially in leads V5-6. Doubling the conductivities did not yield severe changes in QRS morphology. It slightly reduced the width of the QRS in all leads, overall decreased the amplitude of the QRS (especially the R wave in leads V1-4), and slightly increased the amplitude of the S wave in V5-6 while still exhibiting a tall R wave in V5-6. In all cases, the QRS in V4-6 still exhibited a large R amplitude with small S wave, different from the real ECG showing a small R amplitude R with deep and wide S

wave. These results suggest that simple homogeneous conductivity changes do not lead to severe morphological abnormalities in the lateral leads of Group 3 patients.

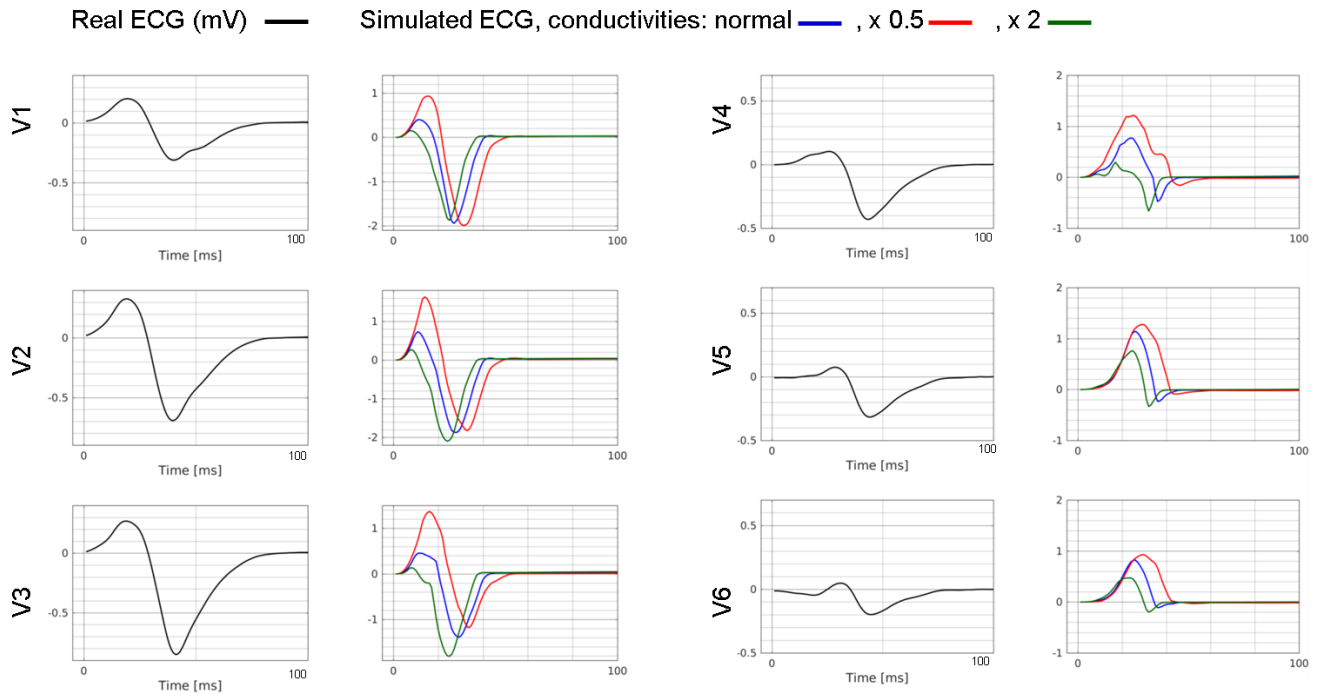


Figure 6.10: Analysis of conductivity changes on the ECG. Real ECG (black) is represented along with a normal simulation (blue), a simulation with intracellular and extracellular conductivities divided by 2 (red), and a simulation with intracellular and extracellular conductivities multiplied by 2 (green).

6.4.2 Conduction abnormalities can explain deep and wide S waves in Group 3 patients

6.4.2.1 Isotropic myocardial propagation does not affect the QRS polarities of the lateral leads

HCM hearts exhibit structural abnormalities such as fibre disarray that may affect the ECG, in particular the activation, and in turn the QRS complex as observed in Group 3 patients. As mentioned in Chapter 2 section 2.2.1, the orientation of the fibres in the healthy heart is anisotropic, and sets a preferred orientation for the cardiomyocytes defining myocardial propagation. Fibre disarray corresponds to a disarrangement of these fibres and a loss of the anisotropy property. Diffusion tensor imaging provides information

on the regions that may be affected by fibre disarray. Indeed, a low fractional anisotropy (FA) may be the indicator of disarray, fibrosis, cell enlargement, or a combination of all. The inspection of these images (Figure 6.11) for the representative Group 3 patient led to various possible altered areas. Indeed, in the basal and mid slices, low FA is observed in the anterior RV/LV insertion point at the epicardial border. Patchy low FA is also noticeable all around the septum. In the apical slice, low FA is seen in the endocardium of the anterior and posterior LV/RV insertion point.

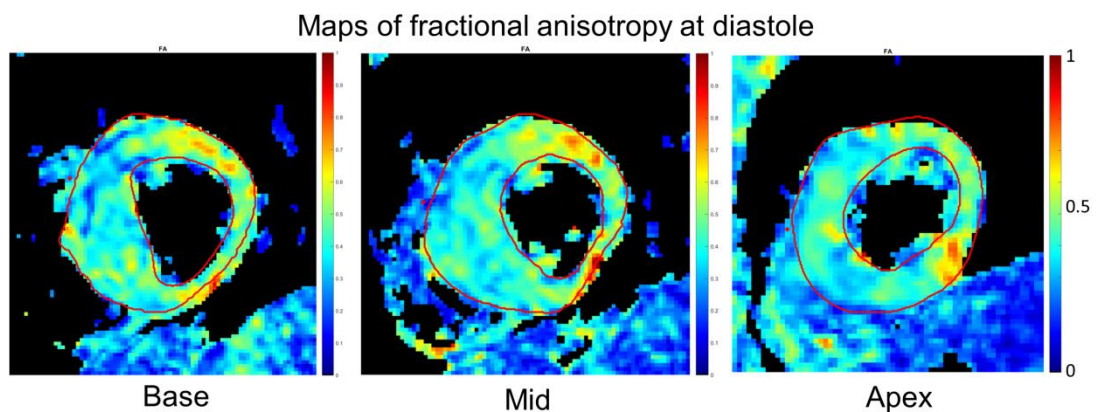


Figure 6.11: Fractional anisotropy maps for Group 3 patient, at diastole, for base, mid and apex slices. Red contours delimit endocardium and epicardium and were manually drawn by Dr. Ariga. Blue regions in the myocardium correspond to areas of low fractional anisotropy.

We first modelled fibre disarray by defining isotropic fibre orientations in the regions identified by DTI as the ones with low FA. This was modelled by defining the same diffusion in all directions, resulting in a velocity of 67 cm/s. The effect of slower conduction is investigated in the following paragraph (Figure 6.13). The resulting ECG is presented on Figure 6.12 below. We concluded that the simple modelling of isotropic propagation in regions of low fractional anisotropy did not affect the ECG. Other abnormalities may therefore alter the morphology of the QRS complex in Group 3 patients.

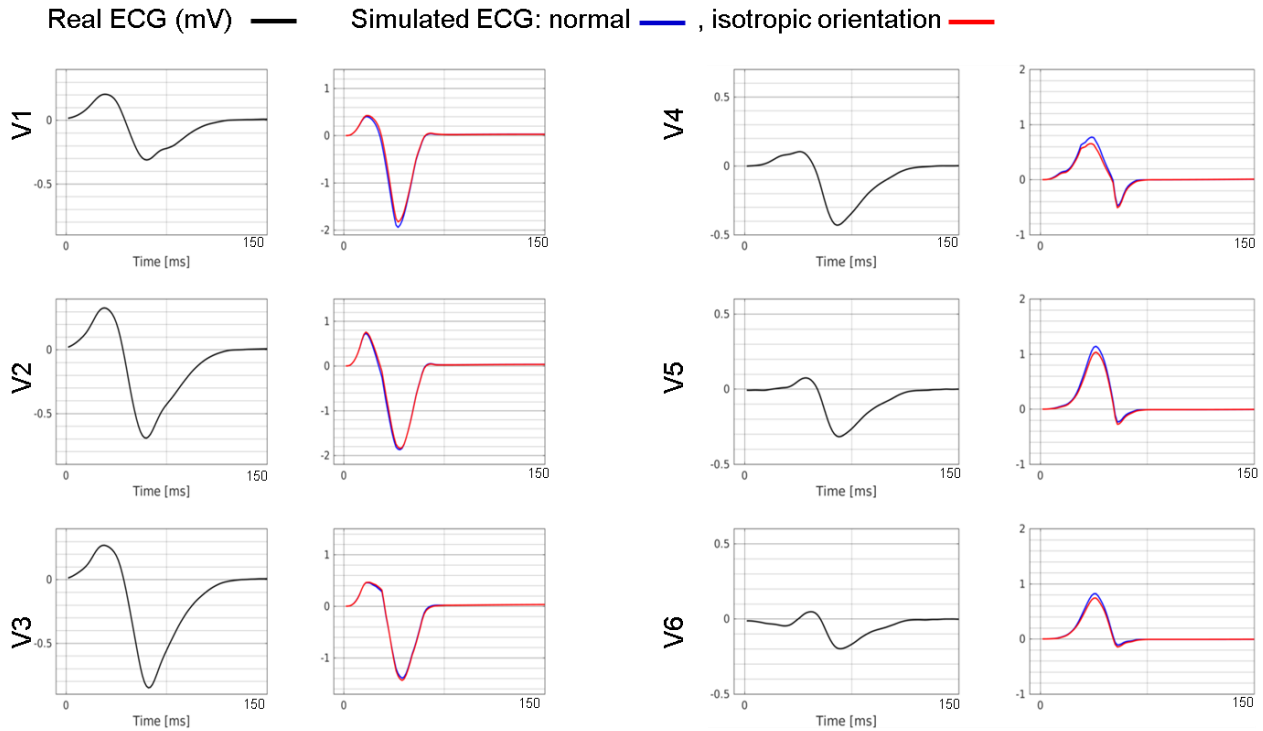


Figure 6.12: Simulated ECG with isotropic orientation of the myocyte fibres (red) along with normal simulation (blue) and real ECG (black) in leads V1 to V6.

Hypertrophy may lead to slower conduction in the myocardium due to fibre disarray, or faster conduction because of hypertrophied cells as described in (Wiegerinck et al. 2006). We therefore investigated the two possible hypotheses. We first assumed that disarrangement in the myocyte fibre could lead to an impaired and slower electrical propagation. We then considered that the increased volume of the myocytes due to hypertrophy could lead to an increase in electrical propagation speed

In order to slow the conduction, local conductivities in the affected regions were reduced by four, and the endocardial propagation speed was also reduced by increasing the node distance in the Dijkstra algorithm. Simulated ECG for four different regions (mid-septum, left ventricular basal free wall, apex and septum from base to apex) are represented on Figure 6.13.

Slowing the conduction in the mid-septum region affected mostly the septal leads V1-V4 with very abnormal QRS morphology in these leads. The QRS in the lateral leads exhibited a similar morphology to the normal simulation, with a slightly larger width and smaller amplitude. Slowing the conduction in the left ventricular basal free wall did not affect the QRS in the septal leads V1-V3, despite a smoother upright S slope to the baseline. Leads V5 and V6 showed similar amplitude than in the normal simulation, and did not exhibit any S waves. V6 showed an abnormally long activation with a very wide QRS. Slowing the conduction in the apex affected mostly leads V4-V6 with severely abnormal patterns of wide QRS. This probably illustrated an abnormally delayed activation of the left ventricle, due to the slow conduction in the apex. Finally, slowing down the conduction in the entire septum from base to apex led once again to severe QRS abnormalities in septal leads V2-V4 with reduced QRS amplitude but similar morphologies in leads V5-V6.

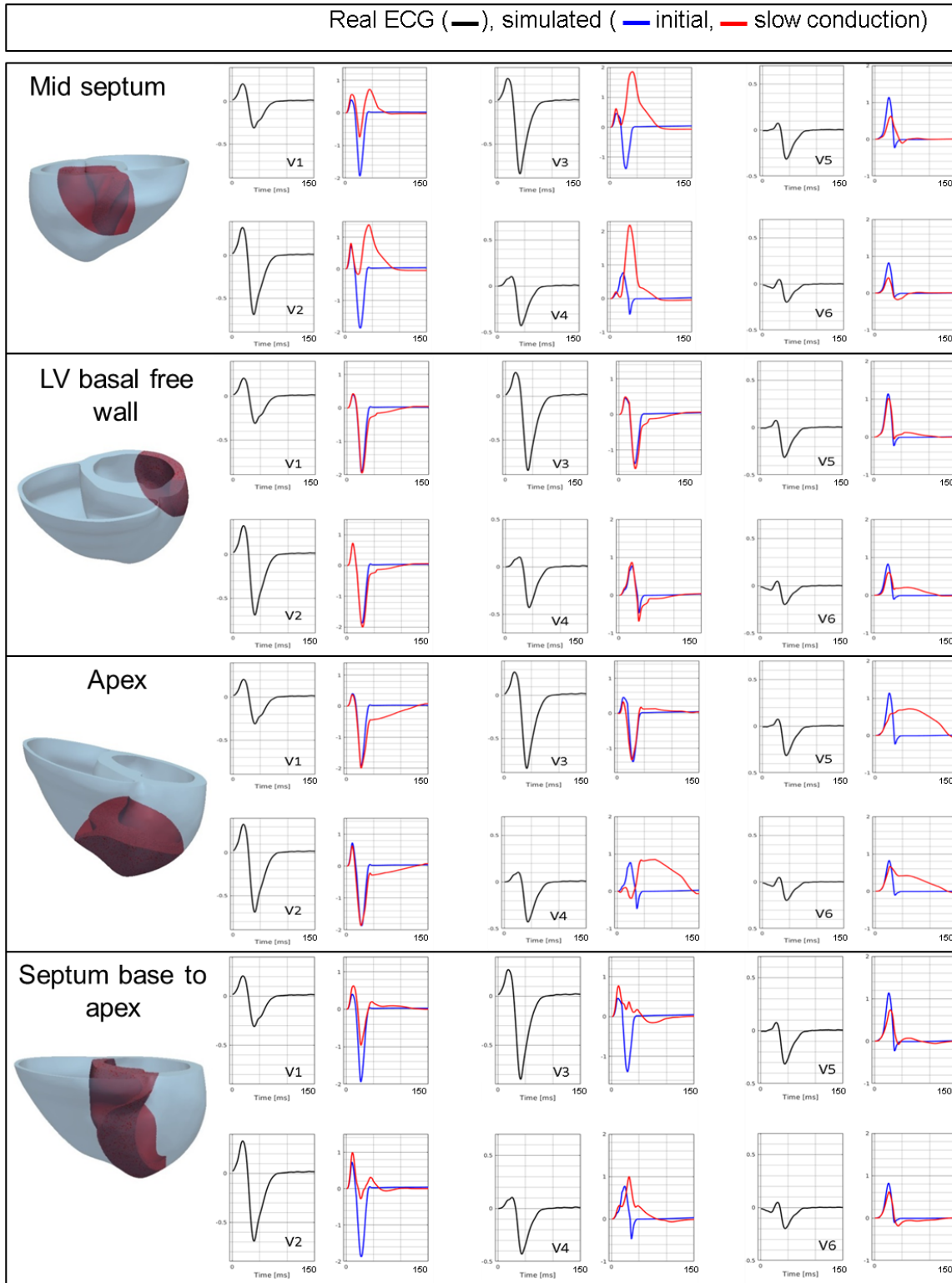


Figure 6.13: Modelling of slow conduction in various locations of the myocardium (mid-septum, left ventricular basal free wall, apex and septum from base to apex), highlighted in red. Real ECG (black), simulated ECG in normal conditions (blue) and simulated ECG with slow conduction (red) are represented.

We then similarly modelled the effect of speeding up the conduction in localized regions of the myocardium. Simulated ECGs are presented on Figure 6.14 below. Faster conduction in both apical and septal regions increased the amplitude and reduced the width of the QRS in leads V3-V4, but did not modify the polarity of the S wave in leads V5 and V6.

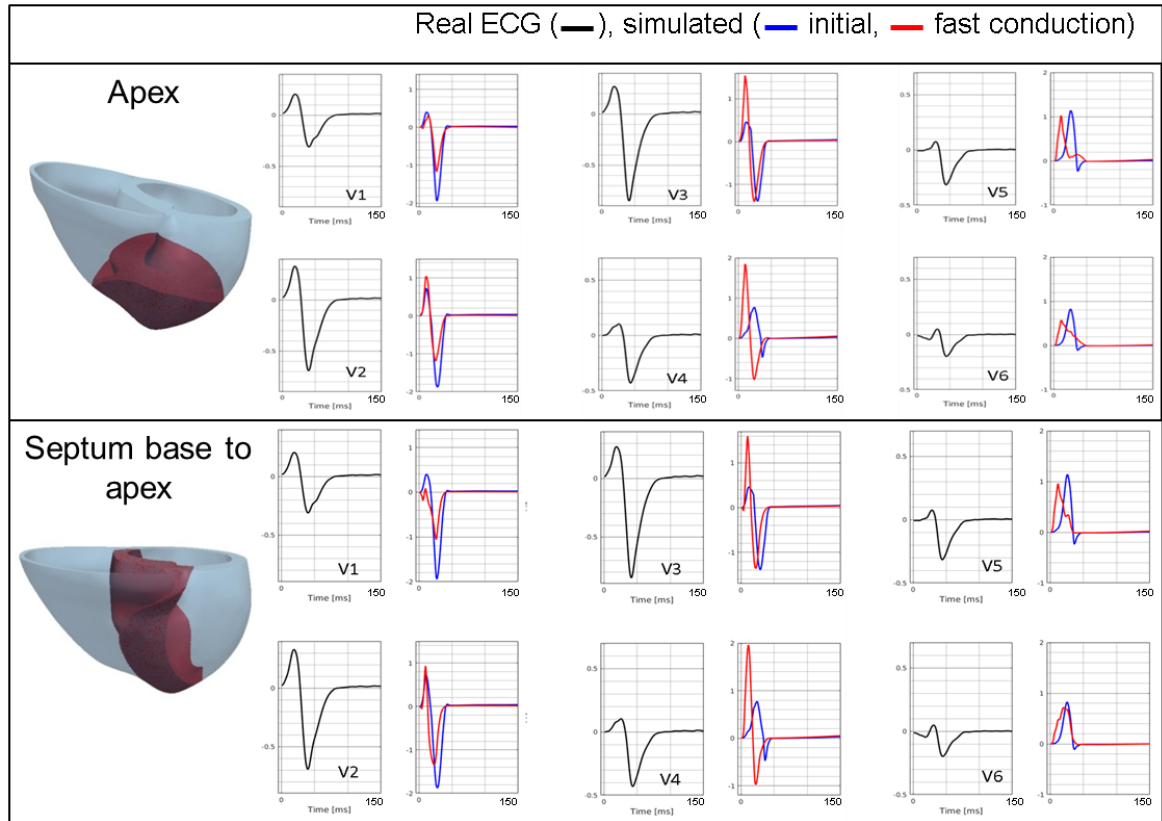


Figure 6.14: Modelling of fast conduction in various locations of the myocardium (apex and septum from base to apex), highlighted in red. Real ECG (black), simulated ECG in normal conditions (blue) and simulated ECG with fast conduction (red) are represented.

These results suggested that homogeneous changes, either a decrease or an increase, in the conduction speed in local regions of the myocardium could not explain the abnormalities observed in Group 3 patients. Indeed, these changes affected the width and amplitude of the QRS in the septal leads, but did not inverse the polarity of the S wave in leads V4-6 which would allow to reproduce the deep and wide S waves observed in these patients.

6.4.2.2 Alteration of the left ventricular early activation sites does not affect the morphology of the QRS waveforms

As described in the previous section, homogeneous changes in the conduction speed, both in specific regions and in the entire myocardium, did not reproduce the abnormal morphology of the QRS observed in Group 3 patients. We therefore hypothesized that early activation sites could be affected by the hypertrophic areas and lead to morphological changes in the lateral leads, as described in (Cardone-Noott et al. 2016). We investigated the effect of changes in the activation by exploring the ECG resulting from various combinations of activation blocks, defined as knock-outs of early activation sites in various regions (Figure 6.15). Simulation results showed that various combinations of activation blocks in the LV affected the duration of the activation, and therefore the width of QRS complexes, but did not change the overall morphology of the QRS in lateral leads. Blocking the entire activation in the left ventricle (left bundle branch block LBBB) led to very wide QRS complexes due to the slow activation of the left ventricle from the right ventricle stimulation. Knocking out the posterior branch of the left ventricle did not affect the morphology of the QRS in all leads, slightly increasing the amplitude of the QRS in V2-4. Knocking out the left ventricular activation except in the septum did not change the QRS morphology in any leads, but increased the QRS amplitude in leads V2-4 and width in V5-V6. Finally, we investigated the effect of blocking the entire activation in the right ventricle (right bundle branch block RBBB). Interestingly, this led to deep and wide S waves in lateral leads, as observed on Group 3 patient. However, the precordial leads were strongly affected, with upright R waves in leads V1-V3, not observed on the real ECG of Group 3 patient, which showed normal QRS morphology in these leads.

We concluded that altering the activation sequence of the left and right ventricles did not lead to the morphological abnormalities observed in the lateral leads V5-6 combined with normal QRS morphology in precordial leads.

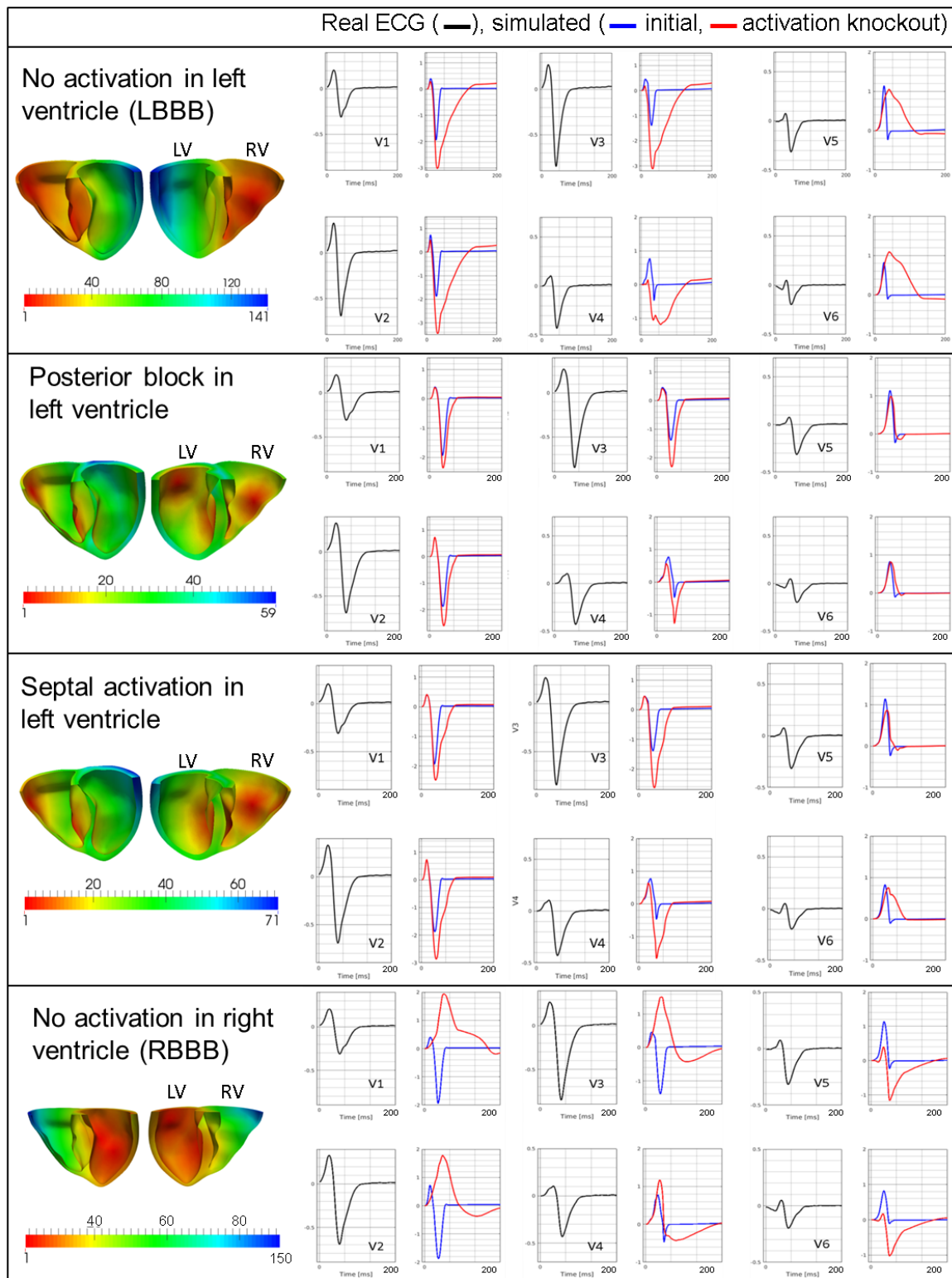


Figure 6.15: Changes in the activation sequence of the left ventricle with different blocking combinations (no activation in the left ventricle, posterior block, free wall activation only, septal activation only). Activation maps in both ventricles are shown along with the specific simulated ECG (red), normal simulation (blue) and real ECG (black) in leads V1 to V6.

6.4.2.3 Alteration of the fast endocardial conduction system can reproduce deep and wide S waves in lateral leads

Considering the inability of homogeneous conduction changes and alteration of the activations sequence to reproduce Group 3 QRS abnormalities, we hypothesized that the fast endocardial conduction system may be affected in these patients. We modelled this abnormality by reducing the coupling in the branches of the Purkinje system coupled with the endocardium. This was explicitly introduced in the model by creating an imbalance between conductivities speed (increased by four in the entire myocardium) and stimulus strength (maintained at 120,000 mA/cm³). This led to a patchy activation and diffuse activation sequence (as illustrated on Figure 6.16), which translated into deep and wide S waves in lateral leads V4-V6 of the ECG, without affecting the morphology of the septal leads (Figure 6.17).

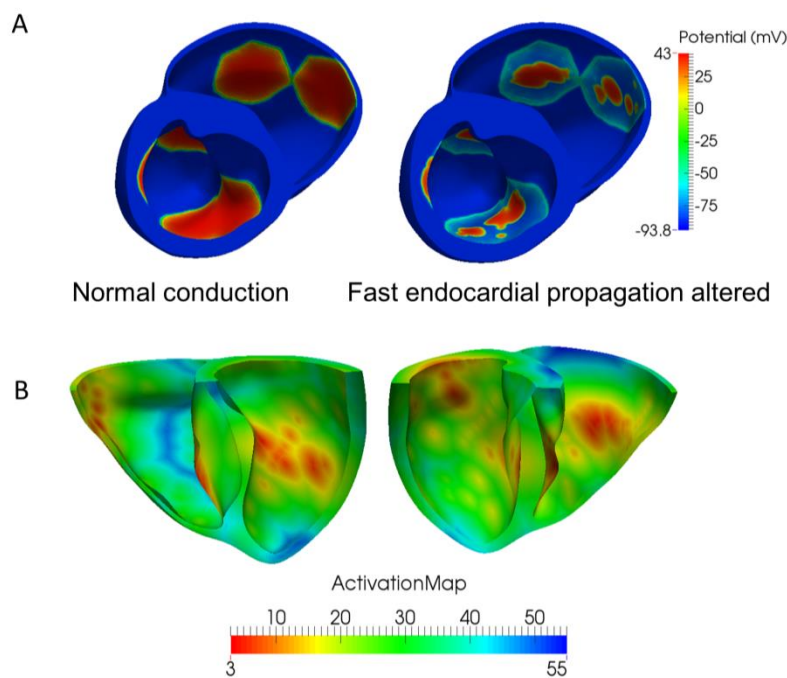


Figure 6.16: Alteration of the fast endocardial propagation. Panel A: Patchy activation in the simulation with altered fast endocardial propagation compared to the normal conduction. Panel B: Activation maps for the simulation with altered fast endocardial propagation show diffuse activation.

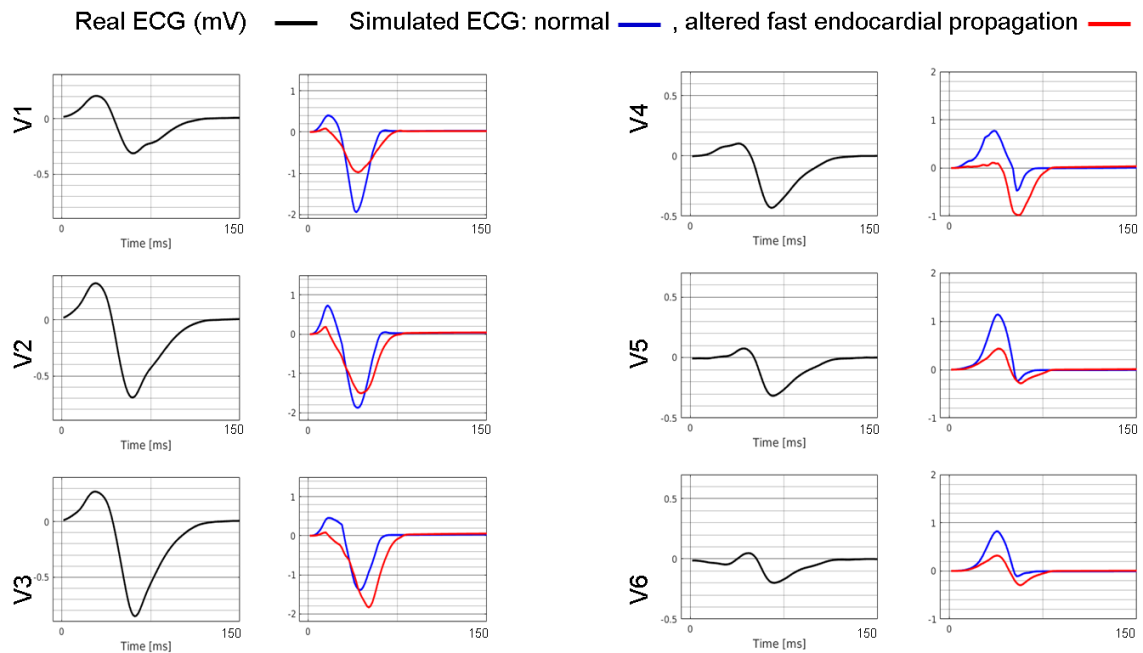


Figure 6.17: Modelling of altered fast propagation system (red) with normal simulation (blue) and real ECG (black) in leads V1 to V6.

Among the various hypothesis explored to explain the presence of abnormal deep and wide S wave in the lateral leads of Group 3 patients, an abnormal fast endocardial conduction system was the only one that reproduced these abnormalities.

6.4.3 Apex-base repolarization heterogeneities can explain T wave inversions as primary repolarization abnormalities

In order to investigate the presence of inverted T waves in leads V4-6 in patients of Group 1A, we hypothesized that longer action potential duration in the apex would lead to an inversion of the normal repolarization currents, and, in turn, lead to the inversion of the T wave. As explained in the Methods section, we reduced g_{Kr} by two in the apical region. The resulting ECG is showed on Figure 6.18.

The QRS complex, representing the activation phase of the cardiac cycle, was not affected by the apico-basal heterogeneity, and remained the same than in the normal simulation. This was to be expected as I_{Kr} is involved in the repolarization phase of the cardiac cycle and does not affect the activation. Interestingly, the T wave morphology was altered and

showed inversions in the lateral leads V4-6, as observed on the real ECG, confirming the hypothesis that apico-basal heterogeneities can be responsible for abnormalities in the repolarization currents expressed as inverted T waves.

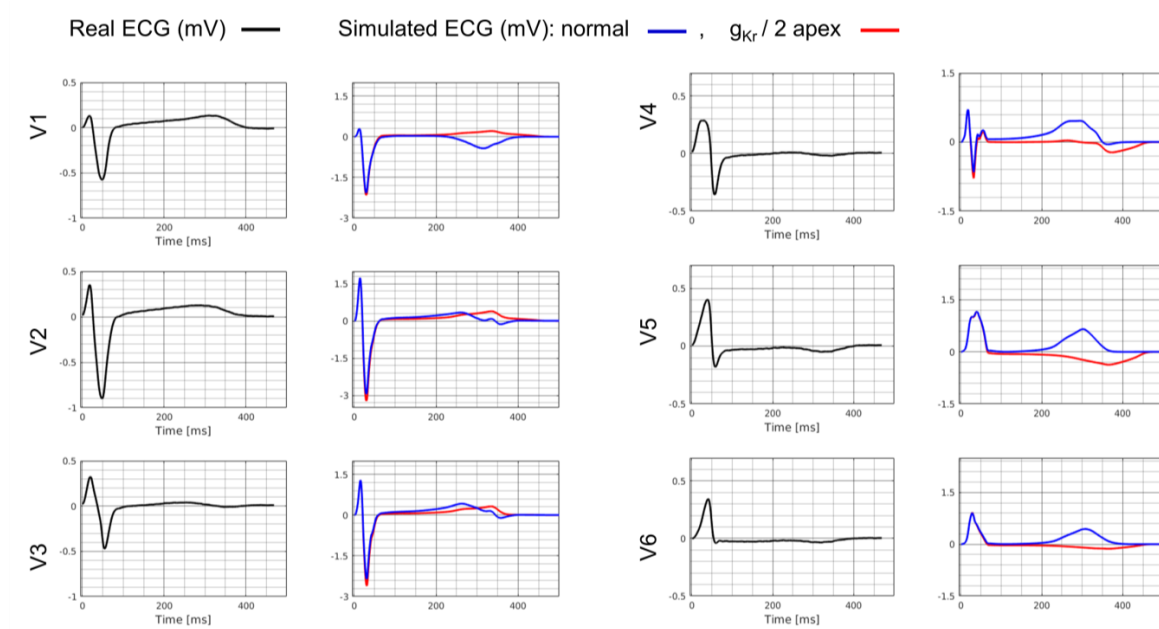


Figure 6.18: Simulated ECG with longer APD in the apex (I_{Kr} conductance divided by 2) (red), along with the real ECG (black) and simulated ECG (blue) in the normal configuration, for Group 1A patient for leads V1-V6.

The repolarization maps presented in Figure 6.19 illustrate the alteration of the repolarization currents in the simulation with reduced apical g_{Kr} . In the normal simulation, the apex was the earliest region of the heart to repolarize (around 225ms), but with a reduced conductance of I_{Kr} , the apex became the latest region to repolarize (around 325ms), inverting the normal pattern of repolarization, and translating into inverted T waves on the ECG.

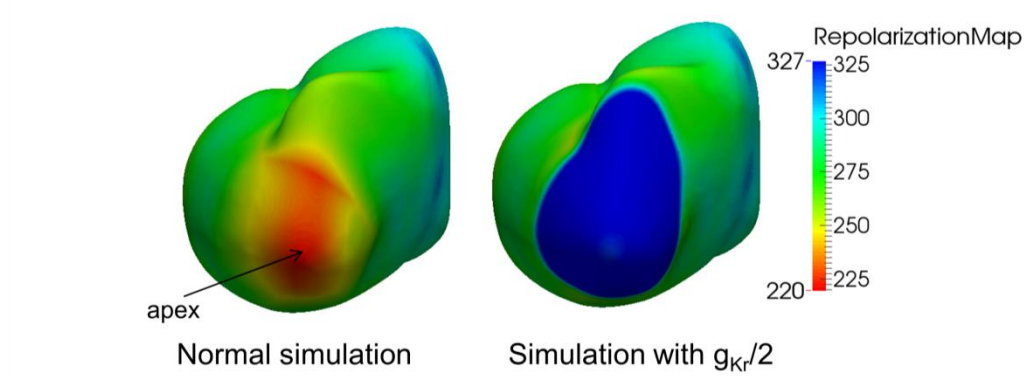


Figure 6.19: Repolarization maps showing the time of repolarization (ms) of each node, for a normal simulation (the apex repolarizes first at 220ms), and a simulation with the IKr conductance divided by 2 in the apex (the apex repolarized the latest, at 325 ms)

We then introduced the HCM ionic remodelling described in (Passini et al. 2016) (Coppini et al. 2013) in the apex. This led to an increased APD in the apical region (555ms vs 314ms in control, Figure 6.20, Panel B) and, in turn, an inversion of the T wave in leads V4-6 (Figure 6.20, Panel A). The width of the T wave in the simulation with HCM ionic remodelling was longer than in the normal simulation or in the simulation with reduced g_{Kr} . This is explained by the difference in APDs modelled (555ms with HCM remodelling, 423ms with reduced g_{Kr} , and 314ms in control).

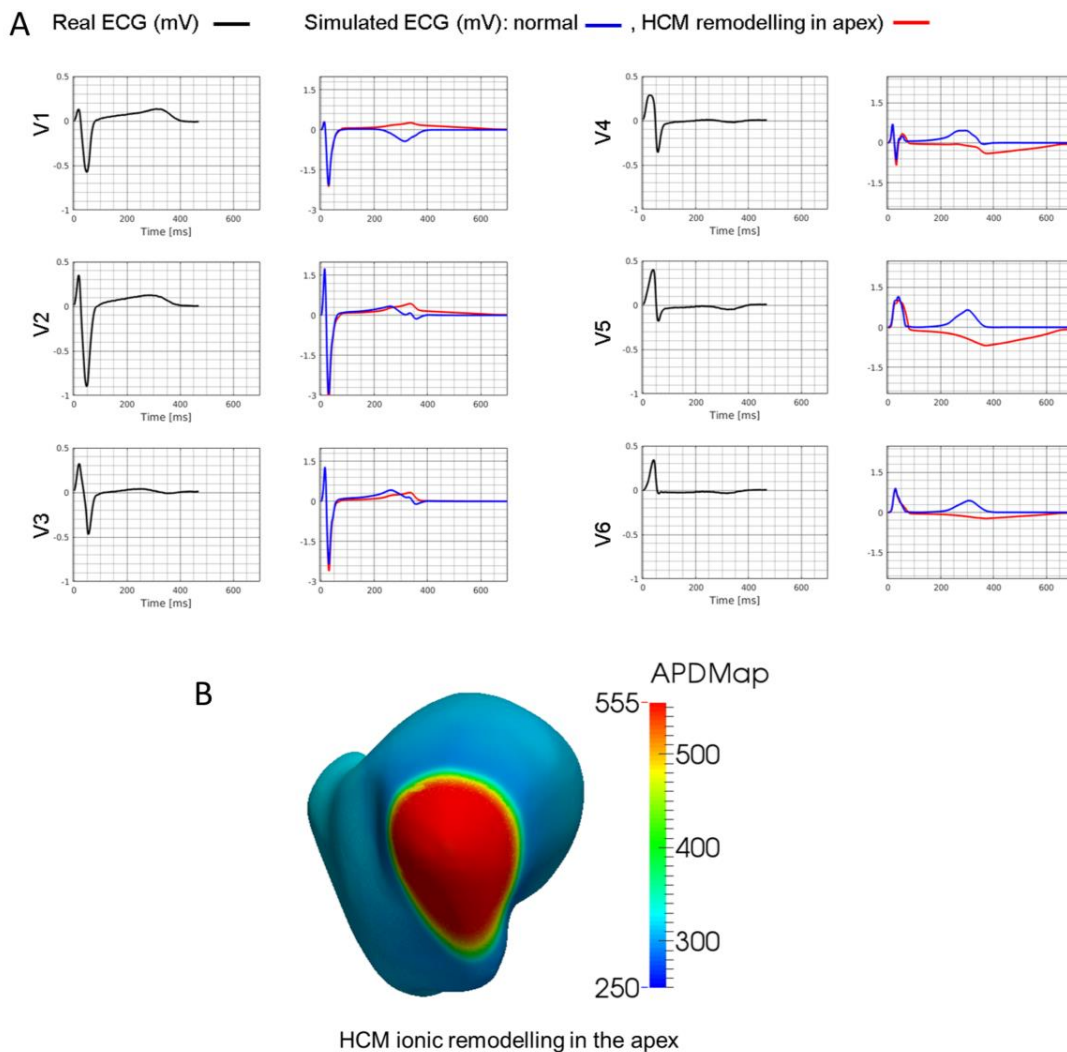


Figure 6.20: Simulation with HCM ionic remodelling in the apex. Panel A: Real ECG (black), simulated ECG (blue) in the initial configuration, simulated ECG (red) with HCM ionic remodelling in the apex for leads V1-V6. Panel B: Action potential duration map showing the APD (ms) in the myocardium, for the simulation with the HCM ionic remodelling in the apex. The apex shows longer APD (555ms) compared to the rest of the ventricle (313ms).

These results therefore suggested that apico-basal repolarization heterogeneities, illustrated by differences in APD between base and apex, potentially due to HCM ionic remodelling, may explain the presence of inverted T waves in Group 1A patients, exhibiting coexistence of septal and apical hypertrophy.

6.5 Discussion

In this project, we investigated the influence of structural and electrophysiological abnormalities on the ECG by using a whole-body 3D simulation pipeline. Our goal was to uncover the cardiac mechanisms responsible for the distinct ECG phenotypes identified in HCM and make the connection between structure and function in HCM patients.

Three main findings were presented in this chapter. First, we proved that cardiac anatomy alone did not account for the distinct ECG abnormalities identified in Chapters 4 and 5. Indeed, the simulated ECG obtained in the normal configuration, without modelling any changes, showed normal QRS and T wave morphology in Groups 1A, 1B and 3, and did not reproduce the specific abnormalities observed in Groups 1A and 3. Changes in amplitude may be noted between the real and simulated ECG, and may be due to differences in body conductivities. Indeed, variability in body conductivities has been shown to strongly affect QRS amplitude (Cardone-Noott et al. 2016). The ECG waveforms were therefore compared in terms of overall morphology and polarity of the waves rather than amplitude. This result suggested that left ventricular hypertrophy alone was not responsible for the severe ECG abnormalities identified in these HCM patients and was therefore not key to distinguish HCM patients and predict higher risk of arrhythmia.

Secondly, we showed that an abnormal Purkinje conduction system may explain the QRS abnormalities (deep and wide S waves) observed in lateral leads of Group 3 patient. The effects of several structural abnormalities were tested such as isotropic propagation and changes in conduction speed due to fibre disarray. None of these homogeneous changes in conduction reproduced the Group 3 abnormalities. Interestingly, altering the activation sequence by modelling various conduction blocks, in both left and right ventricles, did not yield the abnormalities either. Affecting the fast endocardial system related to the Purkinje fibres was the only way to simulate them. We modelled a reduced coupling in the Purkinje

branches coupled to the endocardium which led to a patchy activation and deep and wide S waves. These findings may illustrate an abnormal Purkinje conduction in these patients and were in agreement with the study by (Fananapazir et al. 1989) reporting 23% of HCM patients with abnormal His-Purkinje conduction system. Cases of HCM patients with abnormal His-Purkinje system were also reported in (Doven et al. 2004), with the interesting mention of a patient with blockage of the His potential and third degree AV block exhibiting a deep S wave morphology in V6.

Finally, we showed that apico-basal repolarization heterogeneities affected the T wave in the ECG and may be a potential mechanism to explain the presence of inverted T waves in Group 1A patients, exhibiting primary repolarization abnormalities. More precisely, longer action potential duration in the apex, potentially due to HCM ionic remodelling (Coppini et al. 2013) (Passini et al. 2016), led to T wave inversions in leads V4-6 without affecting the normal morphology of the QRS complex. The location of these changes was, in addition, in agreement with the specific hypertrophy distribution of these patients. Group 1A was indeed the only group in which patients exhibited coexistence of apical and septal hypertrophy, suggesting that ionic remodelling abnormalities may occur in the hypertrophied apical region of these patients.

As a limitation, Group 2 was not investigated as it presented only moderate QRS abnormalities in lead V4 and lower or similar SCD-Risk score compared to Group 3. Future work could investigate the mechanisms at stake in this particular group of patients. In addition, only three representative patients were simulated, because of time and computational costs. Follow-up studies could perform the same investigation on additional patients from the same groups to confirm the role of the mechanisms described here.

To summarise, these results highlighted two different possible mechanisms for ECG abnormalities in HCM. Abnormal Purkinje conduction may explain the severe QRS abnormalities in the lateral leads in Group 3 patients with deep and wide S waves, while ionic remodelling may be the mechanisms responsible for inverted T waves in Group 1A patients. Interestingly, Group 1A patients presented a higher risk of arrhythmia and showed the highest HCM Risk-SCD score compared to other groups, especially Group 3. This could suggest that repolarization abnormalities due to ionic remodelling may play a key role in arrhythmogenesis, as suggested in (Monitillo et al. 2016) (Iacoviello et al. 2007). Conduction abnormalities in the Purkinje system may be less pro-arrhythmic as they may not perturb the overall synchronicity of the cardiac cycle. The very different nature of these two mechanisms shows potential implications in the treatment of HCM patients and reinforces the importance of ECG phenotyping in risk assessment and patient management.

Finally, this work generated hypotheses to explain the mechanisms responsible for the specific ECG abnormalities identified in HCM patients. Invasive techniques such as endocardial mapping could help verify these hypotheses and improve the understanding of the connection between structure and function in hypertrophic cardiomyopathy patients.

6.6 Conclusions

This chapter explored the relationships between structural and electrophysiological abnormalities and the morphology of the ECG, in order to provide insight in the mechanisms that may lead to the HCM phenotypes identified in Chapter 5. A personalized whole-body simulation framework allowed the modelling of the effect of various abnormalities such as hypertrophy, fibre disarray, changes in conduction speed or ionic changes on the ECG. We showed that an abnormal Purkinje conduction may explain the QRS abnormalities in one of the HCM phenotypes. We also concluded that apico-basal repolarization heterogeneities may explain the primary repolarization abnormalities

observed in the higher risk group, exhibiting inverted T wave and normal QRS with coexistence of septal and apical hypertrophy. Overall, this study provided novel hypotheses for structure-function relationships in HCM that may be tested by further invasive studies. This work also confirmed the importance of HCM phenotyping for patient management, and the potential of computer modelling and simulation in describing and understanding cardiac heterogeneity.

7 INVESTIGATION OF THE PRESENCE OF ACTION POTENTIAL ALTERNANS IN HCM

The work in this chapter was published in: “Investigation of the Presence and Mechanisms of Action Potential Alternans in Hypertrophic Cardiomyopathy” by Lyon et al., 2017, Computing in Cardiology.

7.1 Introduction

In Chapters 4 and 5, we investigated the influence of morphological biomarkers extracted from the QRS and ST-T segment waveforms for risk stratification in HCM. Combined with blinded computational clustering, they allowed the novel identification of four distinct phenotypes with differences in ECG morphology, hypertrophy distribution and arrhythmic risk, highlighting the importance of ventricular repolarization in arrhythmogenesis. The results highlight the likely importance of repolarization abnormalities in the risk of lethal arrhythmias. The work presented in this chapter aims at further investigating abnormalities in repolarization in cardiomyocytes with HCM phenotype using computer modelling.

Specifically, the study aims to investigate whether the electrophysiological remodelling in HCM promotes the occurrence of alternans in cardiac repolarization, which have been linked to increased arrhythmic risk (Tse et al 2016). Our hypothesis is that abnormalities in the calcium cycling in HCM may promote the occurrence of alternans. This study focuses on APD alternans in single cell to evaluate the potential relevance of repolarization alternans in HCM. This may motivate future work to characterize T wave alternans in Holter ECGs.

7.2 Motivation

Several studies (Kuroda et al. 2002) (Momiya et al. 1997) (Antunes et al. 2017) have suggested that T wave alternans (TWA), defined as a beat-to-beat variation in the shape of the T wave on the ECG, may be a potential biomarker to identify HCM patients at risk, but the mechanisms behind TWA are still unclear. Indeed, two main possible mechanisms responsible for TWA have been reported: a steep action potential duration restitution, responsible for changes in the membrane voltage that can be caused, for example, by anatomical abnormalities such as myocardial fibre disarray and scarring, and abnormalities in the calcium cycling dynamics, leading to a failure of the myocyte to balance release and reuptake of calcium (Verrier, Kumar, and Nearing 2009) (Cutler and Rosenbaum 2009).

As mentioned earlier, HCM hearts suffer from electrophysiological abnormalities affecting the repolarization phase. In addition, the results from Chapter 5, identifying the phenotype with primary repolarization abnormalities as at higher risk, highlight the importance of ventricular repolarization in arrhythmogenesis. Repolarization abnormalities stem from ionic remodelling and in particular an increase of the L-type calcium, an increase of the Late Sodium current and a reduction of the potassium currents (Coppini et al. 2013) (Passini et al. 2016). It was shown that abnormalities in calcium handling may lead to a

beat-to-beat alternation of the repolarization process illustrated by action potential amplitude (voltage) alternans at the cellular level (Bayer et al. 2010) (Narayan et al. 2008). Indeed, the cardiac action potential amplitude has been shown to be sensitive to conductances of I_{CaL} , I_{NaK} and I_{to} (Zhou et al. 2015) and will therefore be affected by such electrophysiological HCM abnormalities.

In this chapter, we hypothesize that alterations in repolarization currents and calcium dynamics in HCM may promote the occurrence of repolarization alternans in HCM. We investigate the presence of action potential alternans in HCM cardiomyocytes by using a population of model approach described in (Passini et al. 2016) to account for the human variability present in HCM and control populations. We analyse two consecutive heart beats at steady state, record their action potentials and compare the presence of alternans in action potential biomarkers at different pacing lengths in both groups.

7.3 Methods

7.3.1 Populations of control and HCM cardiomyocytes models

The inter-subject variability between human beings is large and observed at all biological levels from cell to organ. It is illustrated by differences in physiological behaviours in various individuals of a same species such as a different response to drugs, or outcome of pathologies. It is crucial to account for this variability when investigating the behaviour of a population, and several standard methods are based on averaging or using a single model, which loses the range of possible behaviours in a population. The population of models method proposes a different approach than the use of a single model in order to take into account variability in populations. It is based on the assumption that cardiac inter-subject variability will be expressed primarily by differences in ion channels properties rather than structure (Britton et al. 2013). Therefore, a cardiac action potential model, defining the structure of the ion channels, is chosen and the parameters of the model such as

conductances, defining the properties of the ion channels, are varied to obtain a group of different models representing the range of possible behaviours.

Therefore, in order to account for inter-subject variability in HCM and control populations, we used the population of models approach described in (Passini et al. 2016). The baseline model used is the human endocardial (O'Hara et al. 2011) model, the state-of-the-art model for human cardiac ventricular cells. An initial population composed of 30,000 cell models is built by varying 11 conductances of the main ion channels in the baseline model, namely g_{Kr} , g_{Ks} , g_{K1} , g_{to} , g_{CaL} , g_{NaL} , g_{Na} , g_{NCX} , g_{NaK} (Na^+/K^+ pump), g_{Rel} (RyR) and g_{Up} (SERCA). The conductances were varied in a range from 50% to 150% using Latin Hypercube Sampling (McKay, Beckman, and Conover 1979). From this set of models, only the ones with maximum and minimum action potential and calcium transient biomarkers values in accordance with experimental data (Coppini et al. 2013) were accepted in the final population. This resulted in a control population of 9,118 models.

The HCM population was then derived from the control population by scaling the control ionic current conductances to recreate the electrical remodelling observed in HCM based on experimental data voltage-clamp experiment (Coppini et al. 2013). As described in (Passini et al. 2016), I_{NaL} was increased by 165%, and I_{CaL} by 40%. I_{to} was decreased by 70% and I_{K1} by 30%. Fast and slow time constants of voltage and Ca^{2+} dependent I_{CaL} inactivation were increased by 35% and 20% respectively. In addition, I_{Kr} and I_{Ks} were decreased by 45%, the SERCA pump by 25%, RyRs release by 20% and Na^+/Ca^{2+} exchanger was increased by 30%. The affinity of Troponin for Ca^{2+} was increased by 50%, the Na^+/K^+ pump was reduced by 30% and the background Na^+ current was increased by 165% similarly as I_{NaL} . These changes are summarized in Table 7.1 below.

Table 7.1: Ionic electrophysiological remodelling in HCM as described in (Passini et al. 2016)

Ionic Current	HCM	Ionic Current	HCM
I_{Na}	-	I_{to}	-70%
I_{NaL}	+165%	I_{K1}	-30%
I_{Nab}	+165%	I_{Kr}	-45%
I_{CaL}	+40%	I_{Ks}	-45%
τI_{CaL} <i>fast inactivation</i>	+35%	I_{NaK}	-30%
τI_{CaL} <i>slow inactivation</i>	+20%	J_{Up}	-25%
I_{NCX}	+30%	J_{Rel}	-20%

7.3.2 Cardiac action potentials simulation

In order to simulate the electrical activity in the cardiomyocytes, the models defined in 7.3.1 were paced at 1000ms until steady state (500 beats), when the cells are stabilized and the amount of calcium entering the cell is the same as the one leaving it, in order to ensure stable ionic concentrations of Na^+ and K^+ . Simulations were run at various pacing cycle lengths to investigate the occurrence of alternans at different frequencies: 700ms, 600ms, 500ms, 450ms, 400ms, 350ms, 300ms and 250ms. Once in steady state, we recorded action potential and calcium transient (CaT) traces for two consecutive beats at the different pacing cycle lengths. Simulations were performed using Chaste, an open source cardiac simulation software (Pitt-Francis et al. 2009) and the analysis of action potential and CaT traces was performed using custom Matlab software.

7.3.3 Computation of action potential biomarkers

In order to compute action potential and CaT biomarkers, the models exhibiting repolarization abnormalities, namely failure at the end of diastole (defined as $V_m > -65$ mV), or a positive derivative of transmembrane voltage after 150ms from the AP peak, were discarded in accordance with experiments.

The following biomarkers were computed for the two beats of each model at each cycle length (illustrated in Figure 7.1):

- Action potential amplitude (APA): computed as the absolute value of the difference between the maximum of the AP and the resting membrane potential
- Action potential duration (APD): computed between the upstroke of the AP and the time of X% (APDX, with $X = 90, 50, 20$) of repolarization.
- Calcium transient (CaT) duration: computed as the maximum of the CaT trace between the upstroke of the CaT trace and the time of X% (CaTDX, with $X = 90, 50, 20$) of repolarization.
- CaT amplitude (CaTA): computed as the absolute value of the difference between the maximum of the CaT trace and the resting membrane potential

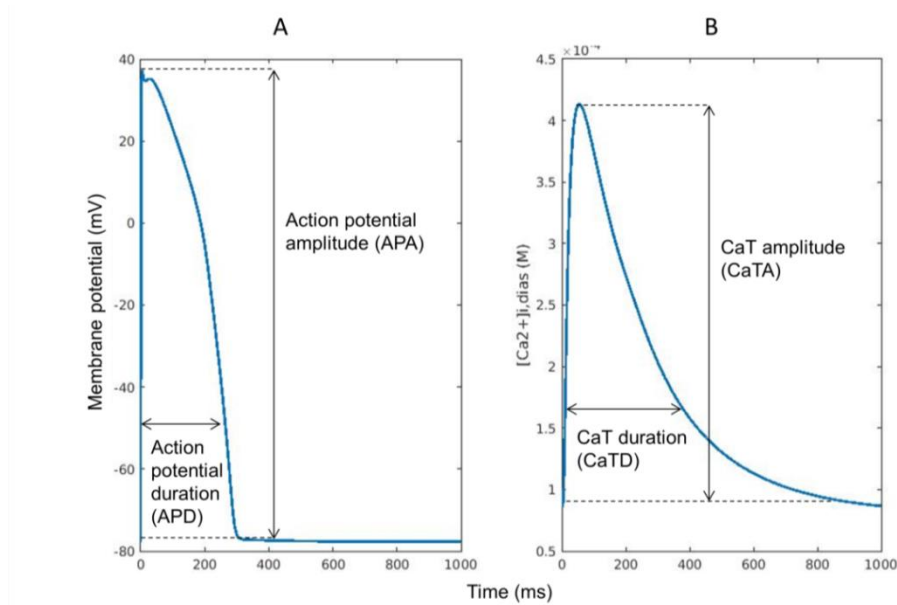


Figure 7.1: (A) Biomarkers extracted from the cardiac action potential: action potential duration and amplitude. (B) Biomarkers extracted from the cardiac calcium transient traces: CaT duration and amplitude.

The models were considered to present amplitude alternans if the difference in amplitude between the two consecutive beats was more than 1% of the magnitude of the amplitude of the first beat. This threshold, discussed in section 7.5, was set to this low value to observe enough alternans for analysis. Duration alternans were defined by a difference in duration between the two consecutive beats larger than 5ms as in (Zhou et al. 2015) (Figure 7.2).

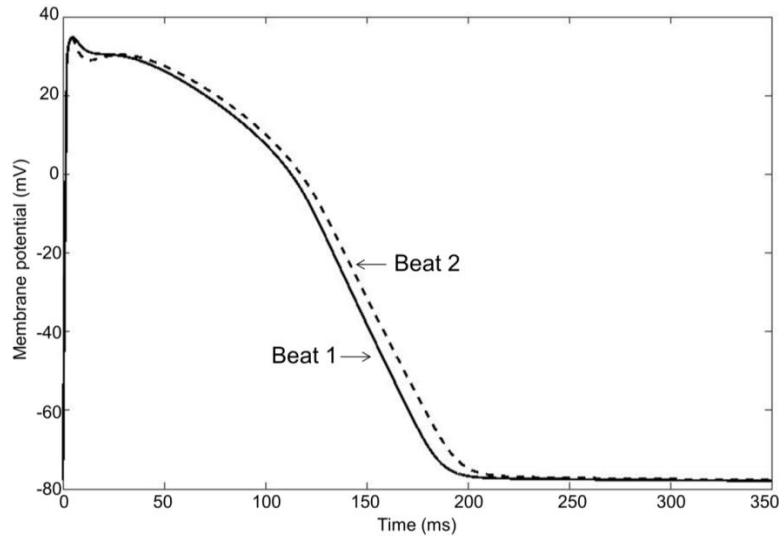


Figure 7.2: Two consecutive beats showing action potential duration alternans: the first beat presents a shorter action potential duration than the second beat.

7.4 Results

7.4.1 Comparison of action potential alternans models in HCM and control populations

In agreement with the HCM experimental phenotype (Passini et al. 2016), the HCM models showed a longer APD compared to controls, with greater variability (486 ± 84 ms vs 286 ± 52 ms, $p < 0.001$ for cycle length of 1000ms) and a longer CaTD compared to controls, with greater variability (691 ± 82 ms vs 522 ± 65 ms, $p < 0.001$ for cycle length of 1000ms) (Figure 7.3).

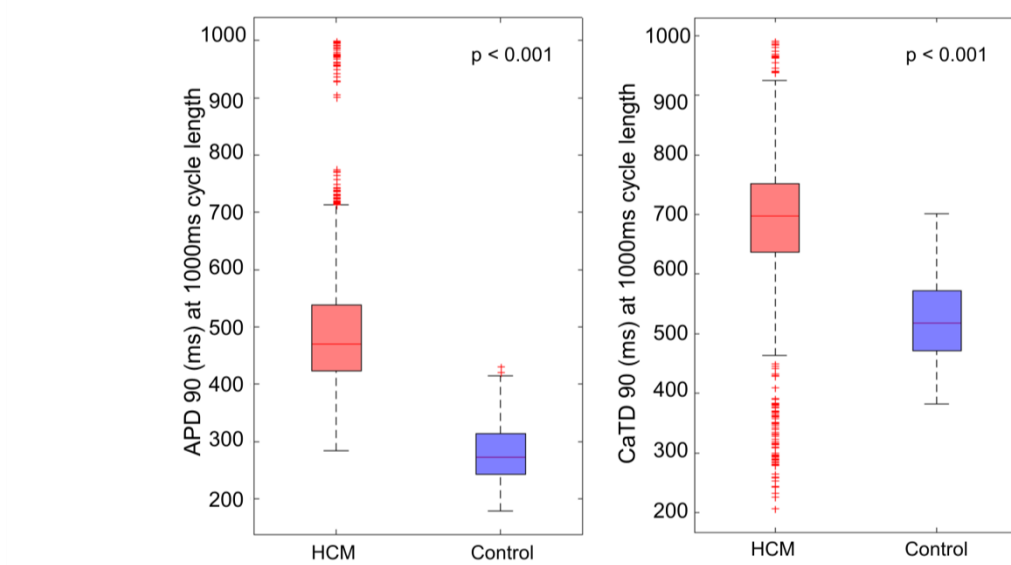


Figure 7.3: Boxplot of APD90 (ms) and CaTD90 (ms) for HCM and Control populations at a cycle length of 1000ms

The percentage of models displaying alternans in control and HCM populations for different pacing lengths is represented on Figure 7.4 and reported in more details in Table 7.1.

For high cycle lengths (400ms to 700ms), both groups show little numbers of action potential amplitude (APA) alternans, with controls showing strictly no APA alternans at these frequencies. For low cycle lengths (350ms to 250ms), HCM models show significantly more APA alternans than controls (250ms: 33% for HCM, 19% for Control, 300ms: 13.4% for HCM, 2.6% for Control, 350ms: 3% for HCM, 0.1% for Control, all $p < 0.001$).

For APD90, CaTA and CaTD90 alternans, the percentages of models displaying alternans in the HCM and control groups do not differ significantly.

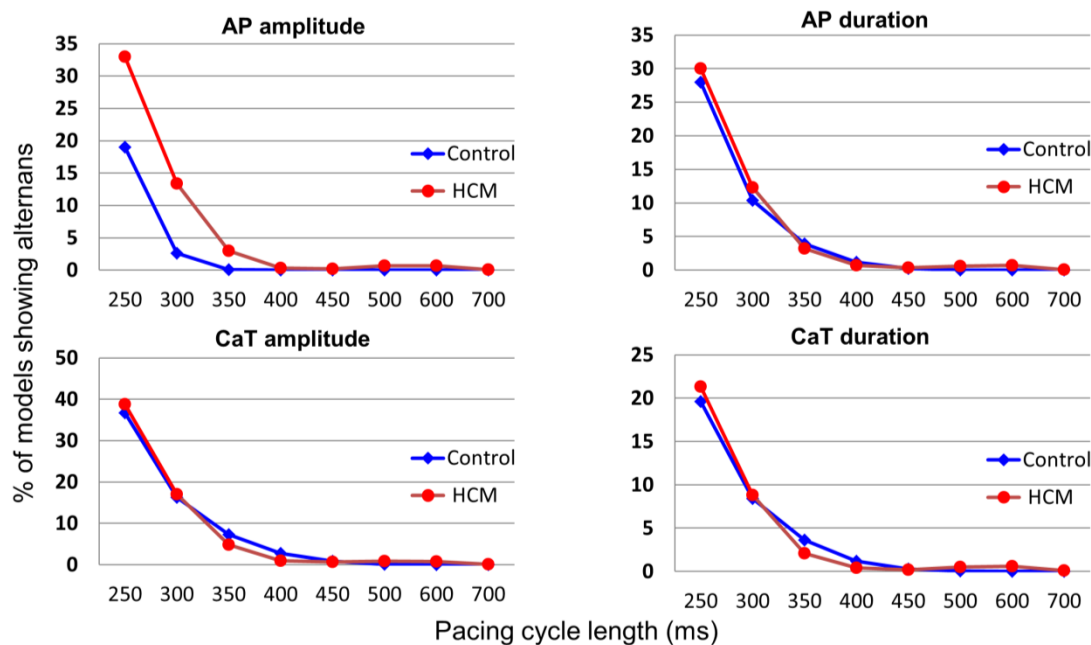


Figure 7.4: Percentage of models displaying alternans for Control and HCM, for action potential amplitude (top left), duration (top right), and Ca²⁺ transient amplitude (bottom left) and duration (bottom right), at pacing cycle length from 250ms to 700ms.

Table 7.2: Number of controls and HCM models displaying alternans for APD90, APD50, APD20, APA, APA and APD90 simultaneously, CATD90 and CATA, at 700ms, 600ms, 500ms, 450ms, 400ms, 350ms, 300ms and 250ms pacing cycle lengths [number of models (%)]

Biomarker	Number of models displaying alternans	
	Control	HCM
700ms (Control: 9,118 accepted models, HCM: 8,447)		
APD90	0	11 (0.1)
APD50	0	9 (0.1)
APD20	0	6 (0.07)
APA	0	7 (0.08)
APA & APD90	0	7 (0.08)
CaTD90	0	12 (0.1)

CaTA	0	11 (0.1)
600ms (Control: 9,118 accepted models, HCM: 8,607)		
APD90	0	57 (0.7)
APD50	0	51 (0.7)
APD20	0	37 (0.4)
APA	0	58 (0.7)
APA & APD90	0	46 (0.5)
CaTD90	0	49 (0.6)
CaTA	0	68 (0.8)
500ms (Control: 9,118 accepted models, HCM: 8,749)		
APD90	1 (0.01)	56 (0.6)
APD50	0	51 (0.6)
APD20	0	37 (0.4)
APA	0	59 (0.7)
APA & APD90	0	48 (0.5)
CaTD90	4 (0.04)	44 (0.5)
CaTA	12 (0.1)	79 (0.9)
450ms (Control: 9,118 accepted models, HCM: 8,704)		
APD90	23 (0.25)	32 (0.37)
APD50	0	20 (0.2)
APD20	2 (0.002)	10 (0.1)
APA	0	19 (0.2)
APA & APD90	0	13 (0.1)
CaTD90	28 (0.3)	18 (0.2)
CaTA	80 (0.9)	58 (0.7)
400ms (Control: 9,118 accepted models, HCM: 8,651)		
APD90	111 (1.2)	67 (0.7)

APD50	28 (0.3)	55 (0.6)
APD20	33 (0.4)	38 (0.4)
APA	0	32 (0.3)
APA & APD90	0	22 (0.3)
CaT90	110 (1.2)	34 (0.4)
CaTA	257 (2.8)	90 (1)
350ms (Control: 9,118 accepted models, HCM: 8,538)		
APD90	359 (3.9)	277 (3.2)
APD50	205 (2.2)	243 (2.8)
APD20	159 (1.7)	214 (2.5)
APA	10 (0.1)	261 (3)
APA & APD90	1 (0.01)	180 (2.1)
CaT90	326 (3.6)	177 (2.1)
CaTA	668 (7.3)	410 (4.8)
300ms (Control: 9,116 accepted models, HCM: 6,386)		
APD90	950 (10.4)	784 (12.3)
APD50	739 (8.1)	737 (11.5)
APD20	598 (6.5)	627 (9.8)
APA	246 (2.6)	858 (13.4)
APA & APD90	102 (1.1)	646 (10)
CaT90	764 (8.4)	562 (8.8)
CaTA	1488 (16.3)	1090 (17)
250ms (Control: 8,669 accepted models, HCM: 1,934)		
APD90	2411 (28)	585 (30)
APD50	2136 (25)	567 (29)
APD20	1700 (19)	501 (26)
APA	1653 (19)	635 (33)

APA & APD90	953 (10.9)	551 (28.5)
CaT90	1702 (19.6)	412 (21.3)
CaTA	3179 (36.7)	751 (38.8)

Accepted models = models that did not exhibit repolarization abnormalities as defined in 7.3.3. APD = action potential duration, APA = action potential amplitude, CaT = calcium transient duration, CaTA = calcium transient amplitude.

In addition to the number of models exhibiting alternans, we computed the magnitude of these alternans for both groups, measured as the absolute difference between the two consecutive beats. HCM models showed significantly higher magnitude of APA alternans compared to controls (250ms: $7.58 \pm 3.5\text{mV}$ vs $4.01 \pm 3.38\text{mV}$, 300ms: $6.32 \pm 3.52\text{mV}$ vs $2.61 \pm 2.26\text{mV}$, 350ms: $4.59 \pm 2.63\text{mV}$ vs $1.27 \pm 0.26\text{mV}$, all $p < 0.001$, Figure 7.5, Panel A). In addition, HCM showed higher magnitude of APD alternans compared to controls (250ms: $21.61 \pm 8.29\text{ms}$ vs $11.89 \pm 5.72\text{ms}$, 300ms: $21.01 \pm 11.29\text{ms}$ vs $9.81 \pm 3.79\text{ms}$, 350ms: $18.69 \pm 9.90\text{ms}$ vs $8.37 \pm 2.41\text{ms}$, 400ms: $15.68 \pm 8.18\text{ms}$ vs $7.66 \pm 2.01\text{ms}$, 450ms: $13.56 \pm 8.89\text{ms}$ vs $6.61 \pm 1.32\text{ms}$ all $p < 0.001$, all $p < 0.001$, Figure 7.5, Panel B).

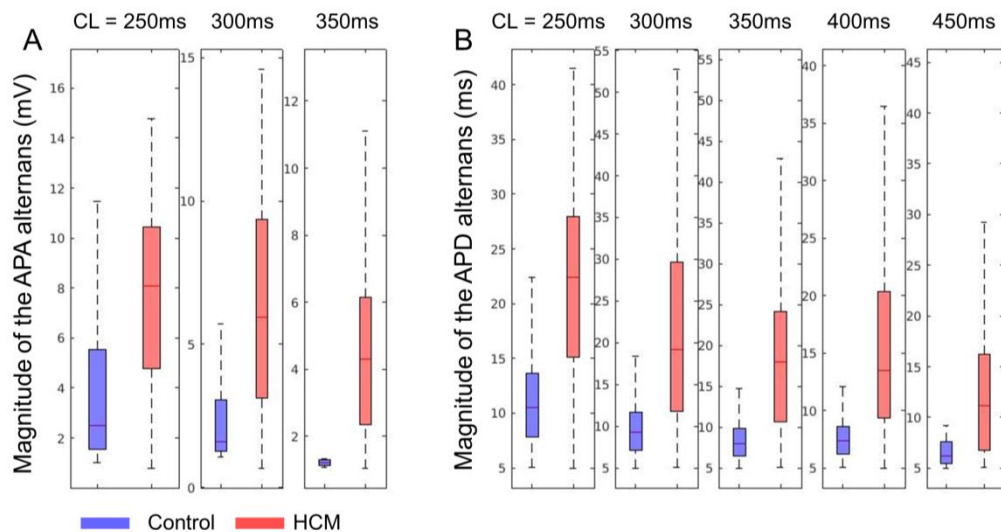


Figure 7.5: Boxplot comparing (A) the magnitude of AP amplitude alternans (mV) in Control and HCM populations at 250, 300 and 350ms pacing length, (B) the magnitude of AP duration alternans (ms) in Control and HCM populations at 250, 300, 350, 400 and 450ms pacing length.

7.5 Discussion

In this work, we investigated the presence of alternans at the cellular level in HCM compared to controls using a population of models approach. HCM models displayed significantly more AP amplitude alternans than controls at low cycle lengths (250-350ms), but no significant differences were observed at pacing cycle lengths between 400ms and 1000ms. Physiologically, the presence of these AP alternans at low frequencies may be relevant in the case of young people in exercise, who may reach these fast heart rates (171-240bpm). HCM being a major cause of SCD in young adults, this may be relevant for this group of patients. HCM and Control models also exhibited similar number of alternans in AP duration, CaT duration or CaT amplitude alternans at all frequencies. Despite low numbers, the magnitude of alternans in HCM models was significantly higher than in Controls in terms of action potential amplitude and duration.

The study by (Narayan et al. 2008) addressed a similar question in heart failure patients and they concluded that magnitude of action potential amplitude alternans explained T wave alternans (TWA) at $< \text{ or } = 109$ beats/min while magnitude of action potential duration did not. In their study, computer models linked these findings with reduced sarcoplasmic reticulum calcium uptake in these heart failure patients, concluding that AP alternans may be caused by abnormalities in the cellular calcium dynamics, linked with risk of VT/VF. This resonates with the absence, in our study, of an increased number of action potential duration alternans in HCM. However, despite observing a higher number of AP amplitude alternans at high frequencies, the threshold for positive alternans (1% difference) captures very subtle changes between two consecutive beats, suggesting that the alternans observed are not very marked.

Therefore, this study does not strongly suggest a larger presence of cellular action potential alternans in HCM cardiomyocytes compared to Controls. This may suggest that TWA on

the ECG, observed in HCM and proposed as potential tool for risk stratification (Antunes et al. 2017), may be due to other dynamics than single cell abnormalities. Alteration of the tissue structure leading to heterogeneities and conduction delays may rather be the cause of TWA. Indeed, increased dispersion-of-refractoriness in heterogeneous regions of the myocardium may lead to areas where recovery times are prolonged. Increased recovery may lead to partial conduction blocks and facilitate re-entry (Kuo et al. 1983) (Armoundas, Tomaselli, and Esperer 2002). Computer simulations have shown that refractoriness heterogeneities could lead to an alternating partial depolarization and result in repolarization alternans on the ECG (Smith and Cohen 1984). More specifically in HCM, structural abnormalities like fibre disarray or fibrosis, which may create regional conduction abnormalities could in turn lead to an enhanced dispersion of refractoriness and promote the occurrence of pro-arrhythmic TWA as suggested in (Kon-No et al. 2001). Further simulation studies could potentially investigate the influence of these abnormalities on the occurrence of alternating behaviours on the ECG.

7.6 Conclusions

This chapter focused on the exploration of repolarization abnormalities at the single cell level, investigating the presence of action potential alternans in HCM to explore whether they could suggest the presence of T wave alternans on the ECG. To this end, a population of models approach for both HCM and Control cardiomyocytes was used, and action potentials simulated at different pacing cycle lengths. HCM showed more action potential amplitude alternans than control at low (250-300ms) cycle lengths, but there were no differences at higher cycle lengths (400-700ms). In addition, no differences were found in terms of the probability of APD90, CATA and CAT90 alternans between the two groups. HCM alternans also showed higher magnitude than controls. These results therefore suggest that single cell action potential alternans do not motivate the occurrence of TWA

on the ECG. These may be due to other dynamics at the whole-organ level such as tissue heterogeneities leading to conduction blocks and promoting the occurrence of pro-arrhythmic TWA in HCM.

8 CONCLUSIONS AND FUTURE DIRECTIONS

8.1 Summary of the thesis

HCM is a common cardiac genetic disease characterized by the thickening of the wall of the ventricles (usually the left ventricle), and a major cause of sudden cardiac death in young adults. Most patients remain asymptomatic but others suffer from sudden cardiac death triggered by ventricular arrhythmias. Due the heterogeneous course of this disease, there is a crucial need to identify high-risk patients to provide them with treatment.

Current HCM risk stratification is based on clinical markers (Christiaans et al. 2010) (Maron et al. 2007), and the resting 12-lead ECG is not included in risk stratification assessment. This may be due to the lack of insight of current techniques used for ECG analysis, which failed to produce reliable stratification (Sherrid et al. 2009) (Maron, Roberts, and Epstein 1982). Based on visual inspections of ECG recordings, they may not manage to handle the phenotypic heterogeneity of HCM and fail to identify specific ECG morphologies in large cohort. Therefore, more complex ECG analysis techniques, able to make sense of the heterogeneity of the data, are needed to improve HCM risk assessment.

In this thesis, we have investigated electrophysiological abnormalities in HCM by developing novel computational methods in order to improve risk stratification in HCM. First, we identified distinct ECG phenotypes using blinded clustering and mathematical modelling on QRS biomarkers alone, and we then added T wave biomarkers to account for the repolarization phase. We then explained the possible structural and electrophysiological mechanisms responsible for the distinct ECG phenotypes previously identified by running personalized 3D computer simulations from cell to whole organ. We finally investigated abnormalities in the repolarization dynamics by exploring the presence of pro-arrhythmic alternans at single cell level in HCM and controls using a population of models approach.

We began in Chapter 2, by introducing important background on cardiac electrophysiology for the normal heart, describing the propagation of electrical signals at the ionic channel level and the whole organ level. We then reviewed the literature on hypertrophic cardiomyopathy, describing the disease progression, and the abnormalities reported on the ECG and cardiac images. We then discussed the challenges that remain for risk stratification.

In Chapter 3, we reviewed the computational methods developed for the analysis and interpretation of ECG signals, with a specific focus on signal processing techniques, machine learning methods for automated classification and clustering, and computer simulations methods, allowing the interpretation of the ECG signal through a better understanding of how the cardiac mechanisms may influence the waveforms.

In Chapter 4, we explored the presence of distinct phenotypes in HCM based solely on QRS morphology accounting for abnormalities in the electrical activation. We implemented signal processing, mathematical modelling and clustering techniques to

analyse a database of HCM Holter ECG recordings, and we identified three groups with different ECG morphologies. However, these subgroups did not exhibit any statistical differences in terms of risk factors and clinical markers from imaging.

In Chapter 5, we included T wave biomarkers in the clustering to account for the repolarization phase. This led to the identification of four distinct subgroups with different ECG morphologies, including one group with normal QRS and inverted T waves (primary repolarization abnormalities), one group with normal QRS and T wave morphology, one group with mild QRS abnormalities in anterior leads, and one group with marked QRS abnormalities in lateral leads but normal T wave. These groups showed significant differences in arrhythmic risk and hypertrophy morphology. Indeed, the group with primary repolarization abnormalities exhibited a higher HCM-Risk-SCD score (O'Mahony et al. 2014), and coexistence of septal and apical hypertrophy

Chapter 6 aimed at interpreting the different ECG phenotypes identified in Chapter 4 and 5 by using a multiscale personalized computer simulation framework from cardiac magnetic resonance images to ECG, in order to explore the role of anatomy and structural and electrophysiological abnormalities on the ECG. We showed that anatomy alone did not account for the various ECG abnormalities identified, but that an abnormal Purkinje conduction system may explain the QRS abnormalities seen in lateral leads. We also concluded that apex-to-base repolarization heterogeneities, potentially due to ionic remodelling, could explain the presence of inverted T waves and normal QRS in the group exhibiting primary repolarization abnormalities with coexisting apical and septal hypertrophy.

Chapter 7, motivated by the importance of repolarization abnormalities highlighted in the previous chapters, finally explored the presence of action potential alternans in HCM, driven by the hypothesis that they may motivate the presence of T wave alternans at the

ECG level. We used a population of cardiomyocytes models for HCM and controls, and compared both groups in terms of action potential alternans. HCM models did not show a significantly larger number of action potential alternans compared to Controls, suggesting that T wave alternans on the ECG may be due to tissue heterogeneities at the whole-organ level.

Overall, this thesis has therefore investigated electrophysiological abnormalities in HCM from cell to whole-organ for risk stratification.

8.1.1 Key contributions and main findings

The first main contribution of this thesis is the development of a novel methodology to analyse and model ECG signals from high-fidelity Holter recordings. We proposed a combination of mathematical modelling and clustering, to capture the morphology of the ECG waveforms and extract subgroups from the HCM population. The power of this methodology is to use complex techniques able to handle large and heterogeneous data to extract simple criteria that would not have been uncovered by a simple visual inspection of the ECG, but can now be applied in the clinic for the analysis of ECG recording. This work shows the benefits of using computational methods to effectively dissect HCM heterogeneity, and presents a crucial step forward to gain improvements in individual patient management and risk stratification.

The second main contribution of this work is the identification of four distinct ECG phenotypes, using the methodology mentioned above, exhibiting differences HCM Risk-SCD score, ventricular arrhythmic events and extent of left ventricular hypertrophy. Patients with normal QRS complexes but with T wave inversion have a near two fold increase in the HCM risk-SCD score compared to other groups, (including the ones with T wave inversion secondary to QRS abnormalities), and exhibit mixed septal and apical

hypertrophy (rather than only septal hypertrophy as in other groups). T wave abnormalities in the presence of normal QRS may reflect underlying myocardial processes (such as remodelling of repolarization currents) not captured by current risk predictors in HCM. The results of this work therefore suggest that computational ECG phenotyping has the potential to be a novel and independent risk factor for clinical stratification.

The third major contribution of this work is the unravelling of the possible structural and electrophysiological mechanisms at the cellular level that may be responsible for the ECG abnormalities identified in the four phenotypes of HCM patients. Using personalized 3D simulations of the electrical propagation in the heart, we showed that the anatomy alone could not explain the ECG abnormalities observed in the representative patients. Abnormal Purkinje conduction may explain deep and wide S waves in lateral leads in Group 3, and apex-to-base heterogeneities due to ionic remodelling may explain the presence of inverted T waves in the Group 1A patients. This work explores the connection between changes at the cellular and tissue levels on the whole-organ level to provide a better interpretation and understanding of ECG.

The final contribution of this thesis is the finding that HCM cardiomyocytes do not display significantly more action potential alternans, either in amplitude or duration, than Controls, which may therefore suggest that T wave alternans at the ECG level may be due to abnormalities at the whole organ level such as tissue heterogeneities.

8.2 Future directions

8.2.1 Implementation of the novel ECG criteria in a larger database with cardiovascular end-points

Our study used digital ECG data from 12-lead Holter recordings. This enabled the identification of four distinct phenotypes using novel computational methods, which is not possible with standard paper ECGs collected in large studies. Given the large information content gathered for each patient in our study, the database was necessarily limited in the number of patients assembled. However, the power of this complex methodology is to output simple criteria that can be used in the clinic for visual inspection of ECG and therefore tested on larger database of paper ECGs. Over the next 5-10 years, a large scale longitudinal study with cardiovascular end-point data, as well as a larger number of high-risk patients, such as the multicenter Hypertrophic Cardiomyopathy Registry (HCMR) (2750 patients) (Kramer et al. 2015) may provide the data to determine whether our findings allow improving current risk stratification for SCD using scanned paper 12 lead ECGs. Our current work provides the detailed analysis based on high fidelity ECG recordings that would enable such validation.

8.2.2 Modelling of complex T wave shapes

In our study, the main biomarker describing the T wave morphology that showed discriminative power in the HCM database was T wave inversion. This may be explained by the fact that other simple descriptors of the T wave morphology such as polarity were spread across different KL basis. We should note that the KL transform characterization proposed is based on the ST-T segment (as a continuation from the QRS complex). This waveform is sensitive to changes in heart rate, which may lead to a loss of the morphological details of the T wave when averaging. Future work could therefore focus on characterizing the T wave morphology in more depth, by extracting a window centred on

the T wave peak by alignment of the waveforms. The discriminative power of this more detailed morphology could then be investigated using a similar methodology than the one presented in Chapter 5.

8.2.3 Investigation of the presence of T wave alternans in HCM

The results of Chapter 7 do not suggest that HCM alternans at the cellular level are responsible for T wave alternans in HCM. However, TWA have been reported as a potential tool for risk stratification in HCM (Kuroda et al. 2002) (Momiya et al. 1997) (Antunes et al. 2017) and future work could therefore investigate the presence of TWA in HCM on the database of Holter recordings following for example the methodology implemented in (Monasterio et al. 2012). If TWA are indeed observed on the ECG, this would suggest that they are driven by whole-organ abnormalities such as fibre disarray or fibrosis inducing tissue heterogeneities in these patients. This could be explored via computer modelling and simulations. A next step would be to investigate the discriminative power of TWA in the HCM population using a similar methodology than the one proposed in Chapter 5 to assess their importance for risk stratification.

8.2.4 Simulation of ischaemia in HCM patients

Myocardial ischaemia has also been reported in HCM and suggested as a cause of sudden death in young adults but is not included in risk assessment (Maron et al. 2009) (Ismail et al. 2014) (Basso et al. 2000). Myocardial ischaemia occurs when an artery in the heart is blocked and the blood flow to the heart is in turn reduced. Ischaemia can lead to cardiac arrest and in time, lead to scarring and damage cardiac tissue. An extension of this thesis could explore ECG markers of ischemia such as ST-T segment deviation (Mincholé, Jager, and Laguna 2010) in HCM, and use it as a biomarker in the clustering methodology presented in Chapter 5. The relation between the extent or location of ischemia in particular groups of HCM patients, and risk factors or other clinical markers could lead to

interesting results. Similarly to what was conducted in this work, personalized computer simulations studies could also help investigate the influence of HCM ischemic regions on the ECG of these patients. For example, Petersen et al. demonstrated that reduced endocardial perfusion was proportional to the magnitude of hypertrophy (Petersen et al. 2007). Ischemia could therefore account for the T wave inversion seen in Group 1A patients as they had greater hypertrophy.

8.3 Conclusions

HCM is a major cause of sudden cardiac death in young adults and a very heterogeneous cardiac genetic disease. Risk stratification remains a challenge and the current ECG biomarkers are not specific to identify high-risk patients. In this thesis, we propose a novel computational approach to HCM risk stratification by exploring electrophysiological abnormalities. We develop mathematical modelling and clustering methods to investigate the presence of ECG subgroups and we identify four distinct phenotypes displaying differences in ECG morphology, arrhythmic risk and hypertrophy morphology. We then introduce a personalized 3D computer simulation framework allowing the simulation of the ECG from cardiac magnetic resonance images. With this methodology, we explore the influence of structural and electrophysiological abnormalities in HCM on the ECG and explain how they may lead to the four phenotypes we identified. We show that anatomy alone does not account for these ECG changes. Abnormal Purkinje conduction may explain the abnormalities in the lateral leads in the group with marked QRS abnormalities, and apex-to-base heterogeneities potentially due to HCM ionic remodelling can explain the inverted T waves in the group with primary repolarization abnormalities and apical hypertrophy. Finally, we investigate the presence of action potential alternans in HCM,

using a population of models approach to account for inter-subject variability. We show that HCM do not display more action potential alternans than Controls, which may suggest that T wave alternans in HCM are due to whole heart tissue abnormalities. Further research is needed to investigate the implication of these findings. This thesis presents a novel approach to ECG analysis, modelling, and interpretation, and shows that computational ECG phenotyping has the potential to become a novel and independent risk factor for clinical stratification.

9 BIBLIOGRAPHY

Adabag A., Maron B., Appelbaum E., Harrigan C., Buross J., Gibson C., Lesser J., et al. 2008. “Occurrence and Frequency of Arrhythmias in Hypertrophic Cardiomyopathy in Relation to Delayed Enhancement on Cardiovascular Magnetic Resonance.” *Journal of the American College of Cardiology* 51 (14): 1369–74. doi:10.1016/j.jacc.2007.11.071.

Alcalai R., Seidman J., and Seidman C. 2008. “Genetic Basis of Hypertrophic Cardiomyopathy: From Bench to the Clinics.” *Journal of Cardiovascular Electrophysiology* 19 (1): 104–10. doi:10.1111/j.1540-8167.2007.00965.x.

Alonso A., Agarwal S., Soliman E., Ambrose M., Chamberlain A., Prineas R., and Folsom A. 2009. “Incidence of Atrial Fibrillation in Whites and African-Americans: The Atherosclerosis Risk in Communities (ARIC) Study.” *American Heart Journal* 158 (1): 111–17. doi:10.1016/j.ahj.2009.05.010.

Ambale-Venkatesh B., Yang X., Wu C., Liu K., Hundley W., McClelland R., Gomes A., et al. 2017. “Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis.” *Circulation Research*, August, CIRCRESAHA.117.311312. doi:10.1161/CIRCRESAHA.117.311312.

Andreão R., Dorizzi B., and Boudy J. 2006. “ECG Signal Analysis through Hidden Markov Models.” *IEEE Transactions on Bio-Medical Engineering* 53 (8): 1541–49. doi:10.1109/TBME.2006.877103.

Antunes M., Samesima N., Filho H., Matsumoto A., Verrier R., Pastore C., Arteaga-Fernández E., and Mady C. 2017. “Exercise-Induced Quantitative Microvolt T-Wave Alternans in Hypertrophic Cardiomyopathy.” *Journal of Electrocardiology* 50 (2): 184–90. doi:10.1016/j.jelectrocard.2016.10.010.

Antzelevitch C., Sicouri S., Litovsky S., Lukas A., Krishnan S., Di Diego J., Gintant G., and Liu D. 1991. “Heterogeneity within the Ventricular Wall. Electrophysiology and Pharmacology of Epicardial, Endocardial, and M Cells.” *Circulation Research* 69 (6): 1427. doi:10.1161/01.RES.69.6.1427.

- Armoundas A., Tomaselli G., and Esperer H. 2002. "Pathophysiological Basis and Clinical Application of T-Wave Alternans." *Journal of the American College of Cardiology* 40 (2): 207–17.
- Asgari S., Mehrnia A., and Moussavi M. 2015. "Automatic Detection of Atrial Fibrillation Using Stationary Wavelet Transform and Support Vector Machine." *Computers in Biology and Medicine* 60 (May): 132–42. doi:10.1016/j.combiomed.2015.03.005.
- Asl B., Setarehdan S., and Mohebbi M. 2008. "Support Vector Machine-Based Arrhythmia Classification Using Reduced Features of Heart Rate Variability Signal." *Artificial Intelligence in Medicine* 44 (1): 51–64. doi:10.1016/j.artmed.2008.04.007.
- Bacharova L., Szathmary V., Potse M., and Mateasik A. 2012. "Computer Simulation of ECG Manifestations of Left Ventricular Electrical Remodeling." *Journal of Electrocardiology* 45 (6): 630–34. doi:10.1016/j.jelectrocard.2012.07.009.
- Bacharova L., Szathmary V., Svehlikova J., Mateasik A., and Tysler M. 2016. "QRS Complex Waveform Indicators of Ventricular Activation Slowing: Simulation Studies." *Journal of Electrocardiology* 49 (6): 790–93. doi:10.1016/j.jelectrocard.2016.07.032.
- Bailón R., Mateo J., Olmos S., Serrano P., García J., del Río A., Ferreira I., and Laguna P. 2003. "Coronary Artery Disease Diagnosis Based on Exercise Electrocardiogram Indexes from Repolarisation, Depolarisation and Heart Rate Variability." *Medical and Biological Engineering and Computing* 41 (5): 561–71. doi:10.1007/BF02345319.
- Basso C., Thiene G., Corrado D., Buja G., Melacini P., and Nava A. 2000. "Hypertrophic Cardiomyopathy and Sudden Death in the Young: Pathologic Evidence of Myocardial Ischemia." *Human Pathology* 31 (8): 988–98. doi:10.1053/hupa.2000.16659.
- Bayer J., Narayan S., Lalani G., and Trayanova N. 2010. "Rate-Dependent Action Potential Alternans in Human Heart Failure Implicates Abnormal Intracellular Calcium Handling." *Heart Rhythm: The Official Journal of the Heart Rhythm Society* 7 (8): 1093–1101. doi:10.1016/j.hrthm.2010.04.008.
- Benson A., Al-Owais M., and Holden A. 2011. "Quantitative Prediction of the Arrhythmogenic Effects of de Novo hERG Mutations in Computational Models of Human Ventricular Tissues." *European Biophysics Journal* 40 (5): 627–39. doi:10.1007/s00249-010-0663-2.
- Bernabeu M., Southern J., Wilson N., Strazdins P., Cooper J., and Pitt-Francis J. 2014. "Chaste: A Case Study of Parallelisation of an Open Source Finite-Element Solver with Applications to Computational Cardiac Electrophysiology Simulation." *International Journal of High Performance Computing Applications* 28 (1): 13–32.
- Bers D. 2002. "Cardiac excitation–contraction coupling". *Nature* 415, 198–205. doi:10.1038/415198a
- Bishop C. 1995. *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc.
- Boichat N., Boichat N., Atienza D., and Khaled N. 2009. "Wavelet-Based ECG Delineation on a Wearable Embedded Sensor Platform." In *Sixth International Workshop*

on Wearable and Implantable Body Sensor Networks, 2009. BSN 2009, 256–61. doi:10.1109/BSN.2009.30.

Bortolan G., Degani R., and Willems J.L. 1990. “Neural Networks for ECG Classification.” In *Computers in Cardiology 1990, Proceedings.*, 269–72. doi:10.1109/CIC.1990.144212.

Boukens B., Meijborg V., Belterman C., Opthof T., Janse M., Schuessler R., Coronel R., and Efimov I. 2017. “Local Transmural Action Potential Gradients Are Absent in the Isolated, Intact Dog Heart but Present in the Corresponding Coronary-perfused Wedge.” *Physiological Reports* 5 (10). doi:10.14814/phy2.13251.

Breiman L. 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Britton O., Bueno-Orovio A., Van Ammel K., Lu H., Towart R., Gallacher D., and Rodriguez B. 2013. “Experimentally Calibrated Population of Models Predicts and Explains Intersubject Variability in Cardiac Cellular Electrophysiology.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (23): E2098-2105. doi:10.1073/pnas.1304382110.

Brugada P., and Brugada J. 1992. “Right Bundle Branch Block, Persistent ST Segment Elevation and Sudden Cardiac Death: A Distinct Clinical and Electrocardiographic Syndrome. A Multicenter Report.” *Journal of the American College of Cardiology* 20 (6): 1391–96.

Bueno-Orovio A., Hanson B., Gill J., Taggart P., and Rodriguez B. 2012. “In Vivo Human Left-to-Right Ventricular Differences in Rate Adaptation Transiently Increase Pro-Arrhythmic Risk Following Rate Acceleration.” Edited by Rajesh Gopalrao Katare. *PLoS ONE* 7 (12): e52234. doi:10.1371/journal.pone.0052234.

Burgess .M, Green L., Millar K., Wyatt R., and Abildskov J. 1972. “The Sequence of Normal Ventricular Recovery.” *American Heart Journal* 84 (5): 660–69. doi:10.1016/0002-8703(72)90181-0.

Burton F., and Cobbe S. 2001. “Dispersion of Ventricular Repolarization and Refractory Period.” *Cardiovascular Research* 50 (1): 10–23. doi:10.1016/S0008-6363(01)00197-3.

Bystricky W., and Safer A. 2002. “Modelling T-End in Holter ECGs by 2-Layer Perceptrons.” In *Computers in Cardiology, 2002*, 105–8. doi:10.1109/CIC.2002.1166718.

Zheng L. 1995. “Detection of ECG Characteristic Points Using Wavelet Transforms.” *IEEE Transactions on Bio-Medical Engineering* 42 (1): 21–28. doi:10.1109/10.362922.

Cai D., Zhang C., and He X. 2010. “Unsupervised Feature Selection for Multi-Cluster Data.” In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 333–342. KDD '10. New York, NY, USA: ACM. doi:10.1145/1835804.1835848.

Caldwell B., Trew M., Sands G., Hooks D., LeGrice I., and Smaill B. 2009. “Three Distinct Directions of Intramural Activation Reveal Nonuniform Side-to-Side Electrical

- Coupling of Ventricular Myocytes.” *Circulation. Arrhythmia and Electrophysiology* 2 (4): 433–40. doi:10.1161/CIRCEP.108.830133.
- Cardone-Noott L., Bueno-Orovio A., Mincholé A., Zemzemi N., and Rodriguez B. 2016. “Human Ventricular Activation Sequence and the Simulation of the Electrocardiographic QRS Complex and Its Variability in Healthy and Intraventricular Block Conditions.” *Europace* 18 (Suppl 4): iv4-iv15. doi:10.1093/europace/euw346.
- Lin C, Mailhes C. 2010. “P- and T-Wave Delineation in ECG Signals Using a Bayesian Approach and a Partially Collapsed Gibbs Sampler.” *IEEE Transactions on Bio-Medical Engineering* 57 (12): 2840–49. doi:10.1109/TBME.2010.2076809.
- de Chazal P., and Celler B. 1996. “Automatic Measurement of the QRS Onset and Offset in Individual ECG Leads.” In *Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine*, 4:1399–1400 vol.4. doi:10.1109/IEMBS.1996.647474.
- de Chazal P., O’Dwyer M., and Reilly R. 2004. “Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features.” *IEEE Transactions on Biomedical Engineering* 51 (7): 1196–1206. doi:10.1109/TBME.2004.827359.
- de Chazal P. and Reilly R. 2006. “A Patient-Adapting Heartbeat Classifier Using ECG Morphology and Heartbeat Interval Features.” *IEEE Transactions on Bio-Medical Engineering* 53 (12 Pt 1): 2535–43. doi:10.1109/TBME.2006.883802.
- de Chazal P. and Reilly R. 2000. “A comparison of the ECG classification performance of different feature sets”. *Computers in Cardiology*. (27): 327-330. doi: 10.1109/CIC.2000.898523
- Chen X., Hu Y., Fetics B., Berger R., and Trayanova N. 2011. “Unstable QT Interval Dynamics Precedes VT Onset in Patients with Acute Myocardial Infarction: A Novel Approach to Detect Instability in QT Interval Dynamics from Clinical ECG.” *Circulation. Arrhythmia and Electrophysiology* 4 (6): 858–66. doi:10.1161/CIRCEP.110.961763.
- Cheng J., Kamiya K., Liu W., Tsuji Y., Toyama J., and Kodama I. 1999. “Heterogeneous Distribution of the Two Components of Delayed Rectifier K⁺ Current: A Potential Mechanism of the Proarrhythmic Effects of Methanesulfonanilideclass III Agents.” *Cardiovascular Research* 43 (1): 135–47. doi:10.1016/S0008-6363(99)00061-9.
- Christiaans I., van Engelen K., van Langen I., Birnie E., Bonzel G., Elliott P., and Wilde A. 2010. “Risk Stratification for Sudden Cardiac Death in Hypertrophic Cardiomyopathy: Systematic Review of Clinical Risk Markers.” *Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology: Journal of the Working Groups on Cardiac Pacing, Arrhythmias, and Cardiac Cellular Electrophysiology of the European Society of Cardiology* 12 (3): 313–21. doi:10.1093/europace/eup431.
- Clavier L., Boucher J-M., Lepage R., Blanc J.-J., and Cornily J.-C.. 2002. “Automatic P-Wave Analysis of Patients Prone to Atrial Fibrillation.” *Medical and Biological Engineering and Computing* 40 (1): 63–71. doi:10.1007/BF02347697.
- Clerc L. 1976. “Directional Differences of Impulse Spread in Trabecular Muscle from Mammalian Heart.” *The Journal of Physiology* 255 (2): 335–46.

Coast D., Stern R., Cano G., and Briller S. 1990. "An Approach to Cardiac Arrhythmia Analysis Using Hidden Markov Models." *IEEE Transactions on Biomedical Engineering* 37 (9): 826–36. doi:10.1109/10.58593.

Colloca R., Johnson A., Mainardi L., and Clifford G. 2013. "A Support Vector Machine Approach for Reliable Detection of Atrial Fibrillation Events." In *Computing in Cardiology* 2013, 1047–50.

Coppini R., Ferrantini C., Yao L., Fan P., Del Lungo M., Stillitano F., Sartiani L., et al. 2013. "Late Sodium Current Inhibition Reverses Electromechanical Dysfunction in Human Hypertrophic Cardiomyopathy." *Circulation* 127 (5): 575–84. doi:10.1161/CIRCULATIONAHA.112.134932.

Cortes C., and Vapnik V. 1995. "Support-Vector Networks." *Machine Learning* 20 (3): 273–97. doi:10.1007/BF00994018.

Cutler M., and Rosenbaum D. 2009. "Risk Stratification for Sudden Cardiac Death: Is There a Clinical Role for T Wave Alternans?" *Heart Rhythm, Proceedings of the 2008 California Heart Rhythm Symposium* October 2-4, 2008, 6 (8, Supplement): S56–61. doi:10.1016/j.hrthm.2009.05.025.

Weiss D., Seemann G., Keller D., Farina D., Sachse F., and Dossel O. 2007. "Modeling of Heterogeneous Electrophysiology in the Human Heart with Respect to ECG Genesis." In *2007 Computers in Cardiology*, 49–52. doi:10.1109/CIC.2007.4745418.

Dam, P., Oostendorp T., and van Oosterom A. 2009. "Application of the Fastest Route Algorithm in the Interactive Simulation of the Effect of Local Ischemia on the ECG." *Medical & Biological Engineering & Computing* 47 (1): 11–20. doi:10.1007/s11517-008-0391-2.

Das M., and Ari S. 2014. "ECG Beats Classification Using Mixture of Features, ECG Beats Classification Using Mixture of Features." *International Scholarly Research Notices*, *International Scholarly Research Notices* 2014, 2014 (September): e178436. doi:10.1155/2014/178436, 10.1155/2014/178436.

Daskalov I., and Christov I. 1999. "Electrocardiogram Signal Preprocessing for Automatic Detection of QRS Boundaries." *Medical Engineering & Physics* 21 (1): 37–44. doi:10.1016/S1350-4533(99)00016-8.

De Lannoy G., Frenay B., Verleysen M., and Delbeke J.. 2008. "Supervised ECG Delineation Using the Wavelet Transform and Hidden Markov Models." In , 22:22–25. doi:10.1007/978-3-540-89208-3_7.

Deo R. 2015. "Machine Learning in Medicine." *Circulation* 132 (20): 1920–30. doi:10.1161/CIRCULATIONAHA.115.001593.

Diaz P., Rudy Y., and Plonsey R.. 1983. "Intercalated Discs as a Cause for Discontinuous Propagation in Cardiac Muscle: A Theoretical Simulation." *Annals of Biomedical Engineering* 11 (3–4): 177–89.

Dijkstra E. 1959. "A Note on Two Problems in Connexion with Graphs." *Numerische Mathematik* 1 (1): 269–271.

- Dokur Z., and Ölmez T. 2001. "ECG Beat Classification by a Novel Hybrid Neural Network." *Computer Methods and Programs in Biomedicine* 66 (2–3): 167–81. doi:10.1016/S0169-2607(00)00133-4.
- Doughty E., and Goldfeder R. 2013. "Characterizing And Diagnosing Hypertrophic Cardiomyopathy From ECG Data." CS 229 Final Report. Stanford University. <http://cs229.stanford.edu/proj2013/DoughtyGoldfeder-CharacterizingAndDiagnosingHypertrophicCardiomyopathyFromECGData.pdf>.
- Doven O., Cicek D., Pekdemir H., Camsari A., Parmaksiz T., Cin G., and Akkus N. 2004. "Abnormal His-Purkinje System Conduction Leading to Complete Atrioventricular Block in Patients with Hypertrophic Cardiomyopathy: A Report of 3 Cases." *Japanese Heart Journal* 45 (2): 347–52.
- Dumont C., Monserrat L., Soler R., Rodríguez E., Fernandez X., Peteiro J., Bouzas A., Bouzas B., and Castro-Beiras A. 2006. "Interpretation of Electrocardiographic Abnormalities in Hypertrophic Cardiomyopathy with Cardiac Magnetic Resonance." *European Heart Journal* 27 (14): 1725–31. doi:10.1093/eurheartj/ehl101.
- Dumont J., Hernández A., and Carrault G. 2010. "Improving ECG Beats Delineation with an Evolutionary Optimization Process." *IEEE Transactions on Bio-Medical Engineering* 57 (3): 607–15. doi:10.1109/TBME.2008.2002157.
- Dutta S., Mincholé A., Zacur E., Quinn T., Taggart P., and Rodriguez B. 2016. "Early Afterdepolarizations Promote Transmural Reentry in Ischemic Human Ventricles with Reduced Repolarization Reserve." *Progress in Biophysics and Molecular Biology* 120 (1–3): 236–48. doi:10.1016/j.pbiomolbio.2016.01.008.
- Elliott P., Poloniecki J., Dickie S., Sharma S., Monserrat L., Varnava A., Mahon N., and McKenna W. 2000. "Sudden Death in Hypertrophic Cardiomyopathy: Identification of High Risk Patients." *Journal of the American College of Cardiology* 36 (7): 2212–18.
- Elliott P., Borger M., Borggrefe M., Cecchi F., Charron P., Hagege A., Lafont A., et al. 2014. "2014 ESC Guidelines on Diagnosis and Management of Hypertrophic Cardiomyopathy: The Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC)." *European Heart Journal* 35 (39): 2733–79. doi:10.1093/eurheartj/ehu284.
- Emanet N. 2009. "ECG Beat Classification by Using Discrete Wavelet Transform and Random Forest Algorithm." In *Fifth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control, 2009. ICSCCW 2009*, 1–4. doi:10.1109/ICSCCW.2009.5379457.
- Ester M., Kriegel H-P., Jörg S., and Xu X. 1996. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." In , 226–231. AAAI Press.
- Faganeli J., and Jager F.. 2010. "Automatic Classification of Transient Ischaemic and Transient Non-Ischaemic Heart-Rate Related ST Segment Deviation Episodes in Ambulatory ECG Records." *Physiological Measurement* 31 (3): 323. doi:10.1088/0967-3334/31/3/004.

- Fananapazir L., Tracy C., Leon M., Winkler J., Cannon R., Bonow R., Maron B., and Epstein S. 1989. "Electrophysiologic Abnormalities in Patients with Hypertrophic Cardiomyopathy. A Consecutive Analysis in 155 Patients." *Circulation* 80 (5): 1259–68.
- Fisch C. 1992. "Electrocardiography and Vectorcardiography." Braunwald E, *Heart Disease. A Textbook of Cardiovascular Medicine* 4th Ed, 116–60.
- Franz M., Bargheer K., Rafflenbeul W., Haverich A., and Lichtlen P. 1987. "Monophasic Action Potential Mapping in Human Subjects with Normal Electrocardiograms: Direct Evidence for the Genesis of the T Wave." *Circulation* 75 (2): 379–86. doi:10.1161/01.CIR.75.2.379.
- Gaborit N., Le Bouter S., Szuts V., Varro A., Escande D., Nattel S., and Demolombe S. 2007. "Regional and Tissue Specific Transcript Signatures of Ion Channel Genes in the Non-Diseased Human Heart." *The Journal of Physiology* 582 (Pt 2): 675–93. doi:10.1113/jphysiol.2006.126714.
- Gabriel S., Lau R., and Gabriel C.. 1996. "The Dielectric Properties of Biological Tissues: III. Parametric Models for the Dielectric Spectrum of Tissues." *Physics in Medicine and Biology* 41 (11): 2271–93.
- Ganeshkumar R., and Kumaraswamy Y. 2012. "Investigating Cardiac Arrhythmia in ECG Using Random Forest Classification." *International Journal of Computer Applications* 37 (4): 31–34. doi:10.5120/4599-6557.
- Gao D., Madden M., Chambers D., and Lyons G.. 2005. "Bayesian ANN Classifier for ECG Arrhythmia Diagnostic System: A Comparison Study." In 2005 IEEE International Joint Conference on Neural Networks, 2005. IJCNN '05. Proceedings, 4:2383–88 vol. 4. doi:10.1109/IJCNN.2005.1556275.
- García J., Lander P., Sörnmo L., Olmos S., Wagner G., and Laguna P. 1998. "Comparative Study of Local and Karhunen–Loève-Based ST–T Indexes in Recordings from Human Subjects with Induced Myocardial Ischemia." *Computers and Biomedical Research* 31 (4): 271–92. doi:10.1006/cbmr.1998.1481.
- Gharaviri A., Dehghan F., Teshnelab M., and Moghaddam H. 2008. "Comparison of Neural Network, ANFIS, and SVM Classifiers for PVC Arrhythmia Detection." In 2008 International Conference on Machine Learning and Cybernetics, 2:750–55. doi:10.1109/ICMLC.2008.4620504.
- Gima K., and Rudy Y. 2002. "Ionic Current Basis of Electrocardiographic Waveforms." *Circulation Research* 90 (8): 889–96. doi:10.1161/01.RES.0000016960.61087.86.
- Goldberger A., Amaral L., Glass L., Hausdorff J., Ivanov P., Mark R., Mietus J., Moody G., Peng C-K., and Stanley H. 2000. "PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals." *Circulation* 101 (23): e215–20. doi:10.1161/01.CIR.101.23.e215.
- Grant A. 2009. "Cardiac Ion Channels". *Circulation: Arrhythmia and Electrophysiology*. 2009;2:185-194. <https://doi.org/10.1161/CIRCEP.108.789081>

- Herring N., Paterson D. 2006. "ECG diagnosis of acute ischaemia and infarction: past, present and future". *QJM: An International Journal of Medicine* 99(4): 219–230. <https://doi.org/10.1093/qjmed/hcl025>
- Iacoviello M., Forleo C., Guida P., Romito R., Sorgente A., Sorrentino S., Catucci S., Mastropasqua F., and Pitzalis M. 2007. "Ventricular Repolarization Dynamicity Provides Independent Prognostic Information Toward Major Arrhythmic Events in Patients With Idiopathic Dilated Cardiomyopathy." *Journal of the American College of Cardiology* 50 (3): 225–31. doi:10.1016/j.jacc.2007.02.071.
- Inan O., Giovangrandi L., and Kovacs G. 2006. "Robust Neural-Network-Based Classification of Premature Ventricular Contractions Using Wavelet Transform and Timing Interval Features." *IEEE Transactions on Biomedical Engineering* 53 (12): 2507–15. doi:10.1109/TBME.2006.880879.
- Ismail T., Hsu L-Y, Greve A., Gonçalves C., Jabbour A., Gulati A., Hewins B., et al. 2014. "Coronary Microvascular Ischemia in Hypertrophic Cardiomyopathy - a Pixel-Wise Quantitative Cardiovascular Magnetic Resonance Perfusion Study." *Journal of Cardiovascular Magnetic Resonance* 16 (August): 49. doi:10.1186/s12968-014-0049-1.
- Niwas I., Kumari R., and Sadasivam V. 2005. "Artificial Neural Network Based Automatic Cardiac Abnormalities Classification." In *Sixth International Conference on Computational Intelligence and Multimedia Applications*, 2005, 41–46. doi:10.1109/ICCIMA.2005.13.
- Jadhav S., Nalbalwar S., and Ghatol A.. 2010. "Artificial Neural Network Based Cardiac Arrhythmia Classification Using ECG Signal Data." In *Electronics and Information Engineering (ICEIE)*, 2010 International Conference On, 1:V1-228-V1-231. doi:10.1109/ICEIE.2010.5559887.
- Janse M., Coronel R., Opthof T., Sosunov E., Anyukhovskiy E., and Rosen M. 2012. "Repolarization Gradients in the Intact Heart: Transmural or Apico-Basal?" *Progress in Biophysics and Molecular Biology* 109 (1): 6–15. doi:10.1016/j.pbiomolbio.2012.03.001.
- Jager F., Taddei A., Moody G., Emdin M., Antolič G., Dorn R., Smrdel A., Marchesi C., Mark R. 2003. "Long-Term ST database: a reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia". *Med. Biol. Eng. Comput.*, 41:172-182
- Kannathal N., Acharya U., Lim C., Sadasivan P., and Krishnan S.. 2003. "Classification of Cardiac Patient States Using Artificial Neural Networks." *Experimental & Clinical Cardiology* 8 (4): 206–11.
- Kawazoe H., Nakano Y., Ochi H., Takagi M., Hayashi Y., Uchimura Y., Tokuyama T., et al. 2016. "Risk Stratification of Ventricular Fibrillation in Brugada Syndrome Using Noninvasive Scoring Methods." *Heart Rhythm* 13 (10): 1947–54. doi:10.1016/j.hrthm.2016.07.009.
- Kayvanpour E., Mansi T., Sedaghat-Hamedani F., Amr A., Neumann D., Georgescu B., Seegerer P., et al. 2015. "Towards Personalized Cardiology: Multi-Scale Modeling of the Failing Heart." *PLOS ONE* 10 (7): e0134869. doi:10.1371/journal.pone.0134869.

Keller D., Weiss D., Dossel O., and Seemann G. 2012. "Influence of I(Ks) Heterogeneities on the Genesis of the T-Wave: A Computational Evaluation." *IEEE Transactions on Bio-Medical Engineering* 59 (2): 311–22. doi:10.1109/TBME.2011.2168397.

Kemmelings J., Linnenbank A., Muilwijk S., SippensGroenewegen A., Peper A., and Grimbergen C. 1994. "Automatic QRS Onset and Offset Detection for Body Surface QRS Integral Mapping of Ventricular Tachycardia." *IEEE Transactions on Biomedical Engineering* 41 (9): 830–36. doi:10.1109/10.312090.

Knupp P., Ernst C., Thompson D., Stimpson C., and Pebay P. 2006. "The Verdict Geometric Quality Library." SAND2007-1751. Sandia National Laboratories. doi:10.2172/901967.

Konno T., Fujino N., Hayashi K., Uchiyama K., Masuta E., Katoh H., Sakamoto Y., Tsubokawa T., Ino H., and Yamagishi M. 2007. "Differences in the Diagnostic Value of Various Criteria of Negative T Waves for Hypertrophic Cardiomyopathy Based on a Molecular Genetic Diagnosis." *Clinical Science (London, England: 1979)* 112 (11): 577–82. doi:10.1042/CS20060293.

Konno T., Shimizu M., Ino H., Yamaguchi M., Terai H., Uchiyama K., Oe K., Mabuchi T., Kaneda T., and Mabuchi H. 2004. "Diagnostic Value of Abnormal Q Waves for Identification of Preclinical Carriers of Hypertrophic Cardiomyopathy Based on a Molecular Genetic Diagnosis." *European Heart Journal* 25 (3): 246–51. doi:10.1016/j.ehj.2003.10.031.

Kon-No Y., Watanabe J., Koseki Y., Koyama J., Yamada A., Toda S., Shinozaki T., et al. 2001. "Microvolt T Wave Alternans in Human Cardiac Hypertrophy: Electrical Instability and Abnormal Myocardial Arrangement." *Journal of Cardiovascular Electrophysiology* 12 (7): 759–63.

Konukoglu E., Relan J., Cilingir U., Menze B., Chinchapatnam P., Jadidi A., Cochet H., et al. 2011. "Efficient Probabilistic Model Personalization Integrating Uncertainty on Data and Parameters: Application to Eikonal-Diffusion Models in Cardiac Electrophysiology." *Progress in Biophysics and Molecular Biology* 107 (1): 134–46. doi:10.1016/j.pbiomolbio.2011.07.002.

Kors J., de Bruyne M., Hoes A., van Herpen G., Hofman A., van Bommel J., and Grobbee D. 1998. "T Axis as an Indicator of Risk of Cardiac Events in Elderly People." *The Lancet* 352 (9128): 601–5. doi:10.1016/S0140-6736(97)10190-8.

Koski A. 1996. "Modelling ECG Signals with Hidden Markov Models." *Artificial Intelligence in Medicine* 8 (5): 453–71. doi:10.1016/S0933-3657(96)00352-1.

Kramer C., Appelbaum E., Desai M., Desvigne-Nickens P., DiMarco J., Friedrich M., Geller N., et al. 2015. "Hypertrophic Cardiomyopathy Registry: The Rationale and Design of an International, Observational Study of Hypertrophic Cardiomyopathy." *American Heart Journal* 170 (2): 223–30. doi:10.1016/j.ahj.2015.05.013.

Kuo C., Amlie J., Munakata K., Reddy C., and Surawicz B. 1983. "Dispersion of Monophasic Action Potential Durations and Activation Times during Atrial Pacing, Ventricular Pacing, and Ventricular Premature Stimulation in Canine Ventricles." *Cardiovascular Research* 17 (3): 152–61.

- Kuroda N., Ohnishi Y., Yoshida A., Kimura A., and Yokoyama M. 2002. "Clinical Significance of T-Wave Alternans in Hypertrophic Cardiomyopathy." *Circulation Journal* 66 (5): 457–62. doi:10.1253/circj.66.457.
- Lagerholm M., Peterson C., Braccini G., Edenbrandt L., and Sornmo L. 2000. "Clustering ECG Complexes Using Hermite Functions and Self-Organizing Maps." *IEEE Transactions on Biomedical Engineering* 47 (7): 838–48. doi:10.1109/10.846677.
- Laguna P., Jane R., Olmos S., Thakor N., Rix H., and Caminal P. 1996. "Adaptive Estimation of QRS Complex Wave Features of ECG Signal by the Hermite Model." *Medical & Biological Engineering & Computing* 34 (1): 58–68.
- Laguna P., Mark R., Goldberg A., and Moody G. 1997. "A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG." In *Computers in Cardiology 1997*, 673–76. doi:10.1109/CIC.1997.648140.
- Laguna P., Moody G., and Mark R. 1994. "Analysis of the Cardiac Repolarization Period Using the KL Transform: Applications on the ST-T Database." In *Computers in Cardiology 1994*, 233–36. doi:10.1109/CIC.1994.470207.
- Laguna P., Thakor N., Caminal P., Jané R., Yoon H., Bayés de Luna A., Martí V., and Guindo J. 1990. "New Algorithm for QT Interval Analysis in 24-Hour Holter ECG: Performance and Applications." *Medical & Biological Engineering & Computing* 28 (1): 67–73.
- Lakdawala N., Thune J., Maron B., Cirino A., Havndrup O., Bundgaard H., Christiansen M., et al. 2011. "Electrocardiographic Features of Sarcomere Mutation Carriers with and without Clinically Overt Hypertrophic Cardiomyopathy." *The American Journal of Cardiology* 108 (11): 1606–13. doi:10.1016/j.amjcard.2011.07.019.
- Legrice I., Hunter P., and Smaill B. 1997. "Laminar Structure of the Heart: A Mathematical Model." *The American Journal of Physiology* 272 (5 Pt 2): H2466–2476.
- Leon L., and Roberge F. 1991. "Directional Characteristics of Action Potential Propagation in Cardiac Muscle. A Model Study." *Circulation Research* 69 (2): 378–95.
- Li Q., Rajagopalan C., and Clifford G. 2014. "A Machine Learning Approach to Multi-Level ECG Signal Quality Classification." *Computer Methods and Programs in Biomedicine* 117 (3): 435–47. doi:10.1016/j.cmpb.2014.09.002.
- Lichman M. 2013. "UCI Machine Learning Repository." University of California, School of Information and Computer Science. <http://archive.ics.uci.edu/ml>.
- Lin G., Nishimura R., Gersh B., Phil D., Ommen S., Ackerman M., Brady P. 2009. "Device complications and inappropriate implantable cardioverter defibrillator shocks in patients with hypertrophic cardiomyopathy." *Heart* 2009;95:709–714.
- Linh T., Osowski S., and Stodolski M. 2003. "On-Line Heart Beat Recognition Using Hermite Polynomials and Neuro-Fuzzy Network." *IEEE Transactions on Instrumentation and Measurement* 52 (4): 1224–31. doi:10.1109/TIM.2003.816841.

Llamedo, M., Khawaja A., and Martínez J.P. 2012. "Cross-Database Evaluation of a Multilead Heartbeat Classifier." *IEEE Transactions on Information Technology in Biomedicine* 16 (4): 658–64. doi:10.1109/TITB.2012.2193408.

Llamedo M., and Martínez J.P. 2011. "Heartbeat Classification Using Feature Selection Driven by Database Generalization Criteria." *IEEE Transactions on Bio-Medical Engineering* 58 (3): 616–25. doi:10.1109/TBME.2010.2068048.

Llamedo M., and Martínez J.P. 2012. "An Automatic Patient-Adapted ECG Heartbeat Classifier Allowing Expert Assistance." *IEEE Transactions on Biomedical Engineering* 59 (8): 2312–20. doi:10.1109/TBME.2012.2202662.

Macfarlane P., Devine B., Latif S., McLaughlin S., Shoat D., and Watts M. 1990. "Methodology of ECG Interpretation in the Glasgow Program." *Methods of Information in Medicine* 29 (4): 354–61.

Magnusson P., Gadler F., Liv P., and Mörner S. 2015. "Hypertrophic Cardiomyopathy and Implantable Defibrillators in Sweden: Inappropriate Shocks and Complications Requiring Surgery". *J Cardiovasc Electrophysiol*, 26: 1088–1094.

Mar T., Zaunseder S., Martínez J.P., Llamedo M., and Poll R. 2011. "Optimization of ECG Classification by Means of Feature Selection." *IEEE Transactions on Biomedical Engineering* 58 (8): 2168–77. doi:10.1109/TBME.2011.2113395.

Marco L., and Chiari L. 2011. "A Wavelet-Based ECG Delineation Algorithm for 32-Bit Integer Online Processing." *BioMedical Engineering OnLine* 10 (1): 23. doi:10.1186/1475-925X-10-23.

Maron B., Roberts W., and Epstein S. 1982. "Sudden Death in Hypertrophic Cardiomyopathy: A Profile of 78 Patients." *Circulation* 65 (7): 1388–94.

Maron B., Shen W., Link M., Epstein A., Almquist A., Daubert J., Bardy G., et al. 2000. "Efficacy of Implantable Cardioverter-Defibrillators for the Prevention of Sudden Death in Patients with Hypertrophic Cardiomyopathy." *The New England Journal of Medicine* 342 (6): 365–73. doi:10.1056/NEJM200002103420601.

Maron B. 2002. "Hypertrophic Cardiomyopathy: A Systematic Review." *JAMA* 287 (10): 1308–20.

Maron B., McKenna W., Danielson G., et al. 2003. "American College of Cardiology/European Society of Cardiology clinical expert consensus document on hypertrophic cardiomyopathy". *J Am Coll Cardiol*;42:1687–713

Maron B. 2010. "Contemporary Insights and Strategies for Risk Stratification and Prevention of Sudden Death in Hypertrophic Cardiomyopathy." *Circulation* 121 (3): 445–56. doi:10.1161/CIRCULATIONAHA.109.878579.

Maron B., Casey S., Chan R., Garberich R., Rowin E., and Maron M. 2015. "Independent Assessment of the European Society of Cardiology Sudden Death Risk Model for Hypertrophic Cardiomyopathy." *The American Journal of Cardiology* 116 (5): 757–64. doi:10.1016/j.amjcard.2015.05.047.

- Maron B., and Maron M. 2013. "Hypertrophic Cardiomyopathy." *The Lancet* 381 (9862): 242–55. doi:10.1016/S0140-6736(12)60397-3.
- Maron B., Spirito P., Shen W-K., Haas T., Formisano F., Link M., Epstein A., et al. 2007. "Implantable Cardioverter-Defibrillators and Prevention of Sudden Cardiac Death in Hypertrophic Cardiomyopathy." *JAMA* 298 (4): 405–12. doi:10.1001/jama.298.4.405.
- Maron M., Olivotto I., Maron B., Prasad S., Cecchi F., Udelson J., and Camici P. 2009. "The Case for Myocardial Ischemia in Hypertrophic Cardiomyopathy." *Journal of the American College of Cardiology* 54 (9): 866–75. doi:10.1016/j.jacc.2009.04.072.
- Martínez J.P., Almeida R., Olmos S., Rocha A., and Laguna P. 2004. "A Wavelet-Based ECG Delineator: Evaluation on Standard Databases." *IEEE Transactions on Bio-Medical Engineering* 51 (4): 570–81. doi:10.1109/TBME.2003.821031.
- Matsubara K., Nakamura T., Kuribayashi T., Azuma A., and Nakagawa M. 2003. "Sustained Cavity Obliteration and Apical Aneurysm Formation in Apical Hypertrophic Cardiomyopathy." *Journal of the American College of Cardiology* 42 (2): 288–95.
- McKay M., Beckman R., and Conover W. 1979. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." *Technometrics* 21 (2): 239–45. doi:10.2307/1268522.
- McLeod C., Ackerman M., Nishimura R., Tajik A., Gersh B., and Ommen S. 2009. "Outcome of Patients with Hypertrophic Cardiomyopathy and a Normal Electrocardiogram." *Journal of the American College of Cardiology* 54 (3): 229–33. doi:10.1016/j.jacc.2009.02.071.
- Meau Y.P., Ibrahim F., Narainasamy S., and Omar R. 2006. "Intelligent Classification of Electrocardiogram (ECG) Signal Using Extended Kalman Filter (EKF) Based Neuro Fuzzy System." *Computer Methods and Programs in Biomedicine* 82 (2): 157–68. doi:10.1016/j.cmpb.2006.03.003.
- Mehta S., Chaturvedi C., and Lingayat N. 2009. "Identification and Delineation of QRS Complex in Electrocardiogram Using Fuzzy C-Means Algorithm." *Journal of Theoretical and Applied Science*.
- Mehta S., and Lingayat N. 2008. "Detection of QRS Complexes in Electrocardiogram Using Support Vector Machine." *Journal of Medical Engineering & Technology* 32 (3): 206–15. doi:10.1080/03091900701507183.
- Melgani F., and Bazi Y. 2008. "Classification of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization." *IEEE Transactions on Information Technology in Biomedicine* 12 (5): 667–77. doi:10.1109/TITB.2008.923147.
- Mincholé A., Jager F., and Laguna P. 2010. "Discrimination between Ischemic and Artifactual ST Segment Events in Holter Recordings." *Biomedical Signal Processing and Control* 5 (1): 21–31. doi:10.1016/j.bspc.2009.09.001.
- Momiyama Y., Hartikainen J., Nagayoshi H., Albrecht P., Kautzner J., Saumarez R., McKenna W., and Camm A. 1997. "Exercise-Induced T-Wave Alternans as a Marker of High Risk in Patients with Hypertrophic Cardiomyopathy." *Japanese Circulation Journal* 61 (8): 650–56.

Monasterio V., Laguna P., Cygankiewicz I., Vázquez R., Bayés-Genís A., Bayés de Luna A., and Martínez J.P. 2012. “Average T-Wave Alternans Activity in Ambulatory ECG Records Predicts Sudden Cardiac Death in Patients with Chronic Heart Failure.” *Heart Rhythm* 9 (3): 383–89. doi:10.1016/j.hrthm.2011.10.027.

Monitillo F., Leone M., Rizzo C., Passantino A., and Iacoviello M. 2016. “Ventricular Repolarization Measures for Arrhythmic Risk Stratification.” *World Journal of Cardiology* 8 (1): 57–73. doi:10.4330/wjc.v8.i1.57.

Moody G., and Mark R. 2001. “The Impact of the MIT-BIH Arrhythmia Database.” *IEEE Engineering in Medicine and Biology Magazine: The Quarterly Magazine of the Engineering in Medicine & Biology Society* 20 (3): 45–50.

Murthy I., and Niranjan U. 1992. “Component Wave Delineation of ECG by Filtering in the Fourier Domain.” *Medical and Biological Engineering and Computing* 30 (2): 169–76. doi:10.1007/BF02446127.

Näbauer M., Beuckelmann D., Überfuhr P., and Steinbeck G. 1996. “Regional Differences in Current Density and Rate-Dependent Properties of the Transient Outward Current in Subepicardial and Subendocardial Myocytes of Human Left Ventricle.” *Circulation* 93 (1): 168–77.

Narayan S., Bayer J., Lalani G., and Trayanova N. 2008. “Action Potential Dynamics Explain Arrhythmic Vulnerability in Human Heart Failure: A Clinical and Modeling Study Implicating Abnormal Calcium Handling.” *Journal of the American College of Cardiology* 52 (22): 1782–92. doi:10.1016/j.jacc.2008.08.037.

Nasiri J., Naghibzadeh M., Yazdi H., and Naghibzadeh B. 2009. “ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm.” In *Third UKSim European Symposium on Computer Modeling and Simulation, 2009. EMS '09*, 187–92. doi:10.1109/EMS.2009.39.

Niederer S., Kerfoot E., Benson A., Bernabeu M., Bernus O., Bradley C., Cherry E., et al. 2011. “Verification of Cardiac Tissue Electrophysiology Simulators Using an N-Version Benchmark.” *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 369 (1954): 4331–51. doi:10.1098/rsta.2011.0139.

Nygårds M.-E., and Sörnmo L. 1983. “Delineation of the QRS Complex Using the Envelope of the E.c.g.” *Medical and Biological Engineering and Computing* 21 (5): 538–47. doi:10.1007/BF02442378.

Obermeyer Z., and Emanuel E. 2016. “Predicting the Future — Big Data, Machine Learning, and Clinical Medicine.” *New England Journal of Medicine* 375 (13): 1216–19. doi:10.1056/NEJMp1606181.

O’Hara T., Virág L., Varró A., and Rudy Y. 2011. “Simulation of the Undiseased Human Cardiac Ventricular Action Potential: Model Formulation and Experimental Validation.” *PLoS Computational Biology* 7 (5): e1002061. doi:10.1371/journal.pcbi.1002061.

Olcayto E. 1997. “DISCRETE RANDOM SIGNALS AND STATISTICAL SIGNAL PROCESSING, C. W. Therrien, Prentice-Hall International, Hemel Hempstead, 1992, ISBN 0-13-217-985-7, £25.39 (Softback, Disc and VAT).” *International Journal of*

- Adaptive Control and Signal Processing 11 (3): 256–256. doi:10.1002/(SICI)1099-1115(199705)11:3<256::AID-ACS425>3.0.CO;2-T.
- Oliveira L., Andreão R., and Sarcinelli-Filho M. 2010. “The Use of Bayesian Networks for Heart Beat Classification.” *Advances in Experimental Medicine and Biology* 657: 217–31. doi:10.1007/978-0-387-79100-5_12.
- O’Mahony C., Elliott P., and McKenna W. 2013. “Sudden Cardiac Death in Hypertrophic Cardiomyopathy.” *Circulation: Arrhythmia and Electrophysiology* 6 (2): 443–51. doi:10.1161/CIRCEP.111.962043.
- O’Mahony C., Jichi F., Pavlou M., Monserrat L., Anastasakis A., Rapezzi C., Biagini E., et al. 2014. “A Novel Clinical Risk Prediction Model for Sudden Cardiac Death in Hypertrophic Cardiomyopathy (HCM Risk-SCD).” *European Heart Journal* 35 (30): 2010–20. doi:10.1093/eurheartj/eh439.
- Oster J., Behar J., Sayadi O., Nemati S., Johnson A., and Clifford G. 2015. “Semisupervised ECG Ventricular Beat Classification With Novelty Detection Based on Switching Kalman Filters.” *IEEE Transactions on Biomedical Engineering* 62 (9): 2125–34. doi:10.1109/TBME.2015.2402236.
- Ostman-Smith I., Wisten A., Nylander E., Bratt E-L., de-Wahl Granelli A., Oulhaj A., and Ljungström E. 2010. “Electrocardiographic Amplitudes: A New Risk Factor for Sudden Death in Hypertrophic Cardiomyopathy.” *European Heart Journal* 31 (4): 439–49. doi:10.1093/eurheartj/ehp443.
- Ouyang N., Yamauchi K., and Ikeda M. 1998. “Training a NN with ECG to Diagnose the Hypertrophic Portions of HCM.” In *The 1998 IEEE International Joint Conference on Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence*, 1:306–9 vol.1. doi:10.1109/IJCNN.1998.682282.
- Ozbay Y., Ceylan R., and Karlik B. 2006. “A Fuzzy Clustering Neural Network Architecture for Classification of ECG Arrhythmias.” *Computers in Biology and Medicine* 36 (4): 376–88. doi:10.1016/j.combiomed.2005.01.006.
- Pan J., and Tompkins W. 1985. “A Real-Time QRS Detection Algorithm.” *IEEE Transactions on Biomedical Engineering* BME-32 (3): 230–36. doi:10.1109/TBME.1985.325532.
- Papadakis M., Basavarajaiah S., Rawlins J., Edwards C., Makan J., Firoozi S., Carby L., and Sharma S. 2009. “Prevalence and Significance of T-Wave Inversions in Predominantly Caucasian Adolescent Athletes.” *European Heart Journal* 30 (14): 1728–35. doi:10.1093/eurheartj/ehp164.
- Passini E., Mincholé A., Coppini R., Cerbai E., Rodriguez B., Severi S., and Bueno-Orovio A. 2016. “Mechanisms of pro-Arrhythmic Abnormalities in Ventricular Repolarisation and Anti-Arrhythmic Therapies in Human Hypertrophic Cardiomyopathy.” *Journal of Molecular and Cellular Cardiology, Special Issue: Computational Modelling of the Heart*, 96 (July): 72–81. doi:10.1016/j.yjmcc.2015.09.003.
- Petersen S., Jerosch-Herold M., Hudsmith L., Robson M., Francis J., Doll H., Selvanayagam J., Neubauer S., and Watkins H. 2007. “Evidence for Microvascular Dysfunction in Hypertrophic Cardiomyopathy: New Insights from Multiparametric

Magnetic Resonance Imaging.” *Circulation* 115 (18): 2418–25. doi:10.1161/CIRCULATIONAHA.106.657023.

Pitt-Francis J., Pathmanathan P., Bernabeu M., Bordas R., Cooper J., Fletcher A., Mirams G., et al. 2009. “Chaste: A Test-Driven Approach to Software Development for Biological Modelling.” *Computer Physics Communications*, 40 YEARS OF CPC: A celebratory issue focused on quality software for high performance, grid and novel computing architectures, 180 (12): 2452–71. doi:10.1016/j.cpc.2009.07.019.

Poli R., Cagnoni S., and Valli G. 1995. “Genetic Design of Optimum Linear and Nonlinear QRS Detectors.” *IEEE Transactions on Biomedical Engineering* 42 (11): 1137–41. doi:10.1109/10.469381.

Pourbabaee B., and Lucas C. 2008. “Automatic Detection and Prediction of Paroxysmal Atrial Fibrillation Based on Analyzing ECG Signal Feature Classification Methods.” In 2008 Cairo International Biomedical Engineering Conference, 1–4. doi:10.1109/CIBEC.2008.4786068.

Pourbabaee B., Roshtkhari M. and Khorasani K. 2017. “Deep Convolutional Neural Networks and Learning ECG Features for Screening Paroxysmal Atrial Fibrillation Patients”. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 99:1-10. doi: 10.1109/TSMC.2017.2705582

Pravdin S., Berdyshev V., Panfilov A., Katsnelson L., Solovyova O., and Markhasin V. 2013. “Mathematical Model of the Anatomy and Fibre Orientation Field of the Left Ventricle of the Heart.” *BioMedical Engineering OnLine* 12 (June): 54. doi:10.1186/1475-925X-12-54.

Rabee A., and Barhumi I. 2012. “ECG Signal Classification Using Support Vector Machine Based on Wavelet Multiresolution Analysis.” In 2012 11th International Conference on Information Science, Signal Processing and Their Applications (ISSPA), 1319–23. doi:10.1109/ISSPA.2012.6310497.

Rabiner L. 1989. “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.” *Proceedings of the IEEE* 77 (2): 257–86. doi:10.1109/5.18626.

Rahman Q., Tereshchenko L., Kongkatong M., Abraham T., Abraham M., and Shatkay H. 2015. “Utilizing ECG-Based Heartbeat Classification for Hypertrophic Cardiomyopathy Identification.” *IEEE Transactions on NanoBioscience* 14 (5): 1–1. doi:10.1109/TNB.2015.2426213.

Rajendra A., Bhat P., Iyengar S., Rao A., and Dua S. 2003. “Classification of Heart Rate Data Using Artificial Neural Network and Fuzzy Equivalence Relation.” *Pattern Recognition* 36 (1): 61–68. doi:10.1016/S0031-3203(02)00063-8.

Ramanathan C., Jia P., Ghanem R., Ryu K., and Rudy Y. 2006. “Activation and Repolarization of the Normal Human Heart under Complete Physiological Conditions.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (16): 6309–14. doi:10.1073/pnas.0601533103.

Rautaharju, P., Nelson J., Kronmal R., Zhang Z., Robbins J., Gottdiener J., Furberg C., Manolio T., and Fried L. 2001. “Usefulness of T-Axis Deviation as an Independent Risk

Indicator for Incident Cardiac Events in Older Men and Women Free from Coronary Heart Disease (the Cardiovascular Health Study).” *The American Journal of Cardiology* 88 (2): 118–23.

Rautaharju P., Surawicz B., and Gettes L. 2009. “AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram Part IV: The ST Segment, T and U Waves, and the QT Interval A Scientific Statement From the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society Endorsed by the International Society for Computerized Electrocardiology.” *Journal of the American College of Cardiology* 53 (11): 982–91. doi:10.1016/j.jacc.2008.12.014.

Robinson M., and Curzen N. 2009. “Electrocardiographic Body Surface Mapping: Potential Tool for the Detection of Transient Myocardial Ischemia in the 21st Century?” *Annals of Noninvasive Electrocardiology: The Official Journal of the International Society for Holter and Noninvasive Electrocardiology, Inc* 14 (2): 201–10. doi:10.1111/j.1542-474X.2009.00284.x.

Rush S., Abildskov J., and Mcfee R. 1963. “Resistivity of Body Tissues at Low Frequencies.” *Circulation Research* 12 (1): 40–50. doi:10.1161/01.RES.12.1.40.

Saini I., Singh D., and Khosla A. 2013. “Delineation of ECG Wave Components Using K-Nearest Neighbor (KNN) Algorithm: ECG Wave Delineation Using KNN.” In *2013 Tenth International Conference on Information Technology: New Generations (ITNG)*, 712–17. doi:10.1109/ITNG.2013.76.

Sameni R., Shamsollahi M., Jutten C., and Clifford G. 2007. “A Nonlinear Bayesian Filtering Framework for ECG Denoising.” *IEEE Transactions on Biomedical Engineering* 54 (12): 2172–85. doi:10.1109/TBME.2007.897817.

Sandryhaila A., Kovacevic J., and Puschel M. 2011. “Compression of QRS Complexes Using Hermite Expansion.” In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 581–84. doi:10.1109/ICASSP.2011.5946470.

Sathish B., and Vimal C. 2010. “Random Forest Classifier Based ECG Arrhythmia Classification.” *Int. J. Healthc. Inf. Syst. Inform.* 5 (2): 1–10. doi:10.4018/jhisi.2010040101.

Savage D., Seides S., Clark C., Henry W., Maron B., Robinson F., and Epstein S. 1978. “Electrocardiographic Findings in Patients with Obstructive and Nonobstructive Hypertrophic Cardiomyopathy.” *Circulation* 58 (3 Pt 1): 402–8.

Sayadi O., and Shamsollahi M. 2009. “A Model-Based Bayesian Framework for ECG Beat Segmentation.” *Physiological Measurement* 30 (3): 335. doi:10.1088/0967-3334/30/3/008.

Sebastian R., Heidenreich E., Dux-Santoy L., Rodriguez J., Ferrero J.M., and Saiz J. 2010. “Modeling Drug Effects on Personalized 3D Models of the Heart: A Simulation Study.” In *Statistical Atlases and Computational Models of the Heart*, 222–31. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-15835-3_23.

- Semsarian C., Ingles J., Maron M., Maron B. 2015. “New Perspectives on the Prevalence of Hypertrophic Cardiomyopathy”. *Journal of the American College of Cardiology* Mar 2015, 65 (12) 1249-1254; DOI: 10.1016/j.jacc.2015.01.019
- Sermesant M., Moireau P., Camara O., Sainte-Marie J., Andriantsimiavona R., Cimirman R., Hill D., Chapelle D., and Razavi R. 2006. “Cardiac Function Estimation from MRI Using a Heart Model and Data Assimilation: Advances and Difficulties.” *Medical Image Analysis* 10 (4): 642–56. doi:10.1016/j.media.2006.04.002.
- Sherrid M., Cotiga D., Hart D., Ehlert F., Haas T., Shen W-K., Link M., et al. 2009. “Relation of 12-Lead Electrocardiogram Patterns to Implanted Defibrillator-Terminated Ventricular Tachyarrhythmias in Hypertrophic Cardiomyopathy.” *The American Journal of Cardiology* 104 (12): 1722–26. doi:10.1016/j.amjcard.2009.07.056.
- Si H. 2015. “TetGen, a Delaunay-Based Quality Tetrahedral Mesh Generator.” *ACM Trans. Math. Softw.* 41 (2): 11:1–11:36. doi:10.1145/2629697.
- Smith J., and Cohen R. 1984. “Simple Finite-Element Model Accounts for Wide Range of Cardiac Dysrhythmias.” *Proceedings of the National Academy of Sciences of the United States of America* 81 (1): 233–37.
- Soltysinska E., Olesen S-P., Christ T., Wettwer E., Varro A., Grunnet M., and Jespersen T. 2009. “Transmural Expression of Ion Channels and Transporters in Human Nondiseased and End-Stage Failing Hearts.” *Pflügers Archiv : European Journal of Physiology* 459 (1): 11–23. doi:10.1007/s00424-009-0718-3.
- Soria-Olivas E., Martinez-Sober M., Calpe-Maravilla J., Guerrero-Martinez J.F., Chorro-Gasco J., and Espi-Lopez J. 1998. “Application of Adaptive Signal Processing for Determining the Limits of P and T Waves in an ECG.” *IEEE Transactions on Biomedical Engineering* 45 (8): 1077–80. doi:10.1109/10.704877.
- Sörnmo L., and Laguna P. 2005. *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Edited by Leif Sörnmo Pablo Laguna. Biomedical Engineering. Burlington: Academic Press. <http://www.sciencedirect.com/science/article/pii/B9780124375529500015>.
- Sovilj S., Magjarević R., Lovell N., and Dokos S. 2013. “A Simplified 3D Model of Whole Heart Electrical Activity and 12-Lead ECG Generation.” *Computational and Mathematical Methods in Medicine* 2013. doi:10.1155/2013/134208.
- Spirito P., Bellone P., Harris K., Bernabo P., Bruzzi P., and Maron B. 2000. “Magnitude of Left Ventricular Hypertrophy and Risk of Sudden Death in Hypertrophic Cardiomyopathy.” *The New England Journal of Medicine* 342 (24): 1778–85. doi:10.1056/NEJM200006153422403.
- Spirito P., Autore C., Formisano F., Assenza G., Biagini E., Haas T., Bongioanni, S. et al. 2014. “Risk of Sudden Death and Outcome in Patients with Hypertrophic Cardiomyopathy with Benign Presentation and without Risk Factors.” *The American Journal of Cardiology* 113 (9): 1550–55. doi:10.1016/j.amjcard.2014.01.435.

- Streeter D., Spotnitz H., Patel D., Ross J., and Sonnenblick E. 1969. "Fiber Orientation in the Canine Left Ventricle during Diastole and Systole." *Circulation Research* 24 (3): 339–47.
- Szabó G., Szentandrassy N., Bíró T., Tóth B., Czifra G., Magyar J., Bányász T., Varró A., Kovács L., and Nánási P. 2005. "Asymmetrical Distribution of Ion Channels in Canine and Human Left-Ventricular Wall: Epicardium versus Midmyocardium." *Pflügers Archiv: European Journal of Physiology* 450 (5): 307–16. doi:10.1007/s00424-005-1445-z.
- Szentadrassy N., Banyasz T., Biro T., Szabo G., Toth B., Magyar J., Lazar J., Varro A., Kovacs L., and Nanasi P. 2005. "Apico-Basal Inhomogeneity in Distribution of Ion Channels in Canine and Human Ventricular Myocardium." *Cardiovascular Research* 65 (4): 851–60. doi:10.1016/j.cardiores.2004.11.022.
- Taggart P., Sutton P., Opthof T., Coronel R., Trimlett R., Pugsley W., and Kallis P. 2001. "Transmural Repolarisation in the Left Ventricle in Humans during Normoxia and Ischaemia." *Cardiovascular Research* 50 (3): 454–62. doi:10.1016/S0008-6363(01)00223-1.
- Tibshirani R. 1994. "Regression Shrinkage and Selection Via the Lasso." *Journal of the Royal Statistical Society, Series B* 58: 267–288.
- Trahanias P., and Skordalakis E. 1990. "Syntactic Pattern Recognition of the ECG." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (7): 648–57. doi:10.1109/34.56207.
- Trayanova N., Plank G., and Rodríguez B. 2006. "What Have We Learned from Mathematical Models of Defibrillation and Postshock Arrhythmogenesis? Application of Bidomain Simulations." *Heart Rhythm* 3 (10): 1232–35. doi:10.1016/j.hrthm.2006.04.015.
- Tse G., Wong S., Tse V., Lee Y., Lin H., and Yeo J. 2016. "Cardiac dynamics: Alternans and arrhythmogenesis." *Journal of Arrhythmia*, 32(5), 411–417. <http://doi.org/10.1016/j.joa.2016.02.009>
- Tusscher, ten K., and Panfilov A. 2006. "Alternans and Spiral Breakup in a Human Ventricular Tissue Model." *American Journal of Physiology. Heart and Circulatory Physiology* 291 (3): H1088–1100. doi:10.1152/ajpheart.00109.2006.
- Übeyli E. 2007. "ECG Beats Classification Using Multiclass Support Vector Machines with Error Correcting Output Codes." *Digital Signal Processing* 17 (3): 675–84. doi:10.1016/j.dsp.2006.11.009.
- Übeyli E. 2009. "Combining Recurrent Neural Networks with Eigenvector Methods for Classification of ECG Beats." *Digital Signal Processing* 19 (2): 320–29. doi:10.1016/j.dsp.2008.09.002.
- Vadakkumpadan F., Rantner L., Tice B., Boyle P., Prassl A., Vigmond E., Plank G., and Trayanova N. 2009. "Image-Based Models of Cardiac Structure with Applications in Arrhythmia and Defibrillation Studies." *Journal of Electrocardiology* 42 (2): 157.e1–10. doi:10.1016/j.jelectrocard.2008.12.003.

- Varnava A., Elliott P., Mahon N., Davies M., and McKenna W. 2001. "Relation between Myocyte Disarray and Outcome in Hypertrophic Cardiomyopathy." *The American Journal of Cardiology* 88 (3): 275–79.
- Verrier R., Kumar K., and Nearing B. 2009. "Basis for Sudden Cardiac Death Prediction by T-Wave Alternans from an Integrative Physiology Perspective." *Heart Rhythm : The Official Journal of the Heart Rhythm Society* 6 (3): 416–22. doi:10.1016/j.hrthm.2008.11.019.
- Vigmond E., Vadakkumpadan F., Gurev V., Arevalo H., Deo M., Plank G., and Trayanova N. 2009. "Towards Predictive Modelling of the Electrophysiology of the Heart." *Experimental Physiology* 94 (5): 563. doi:10.1113/expphysiol.2008.044073.
- Vullings H., Verhaegen M., and Verbruggen H. 1998. "Automated ECG Segmentation with Dynamic Time Warping." In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1998, 163–66 vol.1. doi:10.1109/IEMBS.1998.745863.
- Wallman M., Smith N., and Rodriguez B. 2011. "Estimation of Activation Times in Cardiac Tissue Using Graph Based Methods." In *Functional Imaging and Modeling of the Heart*, 71–79. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-21028-0_9.
- Wallman M., Smith N., and Rodriguez B. 2014. "Computational Methods to Reduce Uncertainty in the Estimation of Cardiac Conduction Properties from Electroanatomical Recordings." *Medical Image Analysis* 18 (1): 228–40. doi:10.1016/j.media.2013.10.006.
- Watkins H. 2001. "Hypertrophic cardiomyopathy: from molecular and genetic mechanisms to clinical management". *European Heart Journal Supplements* (3):L43–L50. https://doi.org/10.1016/S1520-765X(01)90064-1
- Watkins H., Rosenzweig A., Hwang D-S., Levi T., McKenna W., Seidman C., and Seidman J. 1992. "Characteristics and Prognostic Implications of Myosin Missense Mutations in Familial Hypertrophic Cardiomyopathy". *N Engl J Med.* 326:1108-1114. DOI: 10.1056/NEJM199204233261703
- Watkins H., McKenna W., Thierfelder L., Suk H., Anan R., O'Donoghue A., Spirito P., Matsumori A., Moravec C., Seidman J., and Seidman C. 1995. "Mutations in the Genes for Cardiac Troponin T and α -Tropomyosin in Hypertrophic Cardiomyopathy". *N Engl J Med*; 332:1058-1065. DOI: 10.1056/NEJM199504203321603
- Wettwer E., Amos G., Posival H., and Ravens U. 1994. "Transient Outward Current in Human Ventricular Myocytes of Subepicardial and Subendocardial Origin." *Circulation Research* 75 (3): 473–82.
- Wiegerinck R., Verkerk A., Belterman C., van Veen T., Baartscheer A., Opthof T., Wilders R., de Bakker J. and Coronel R. 2006. "Larger Cell Size in Rabbits With Heart Failure Increases Myocardial Conduction Velocity and QRS Duration". *Circulation*;113:806-813. https://doi.org/10.1161/CIRCULATIONAHA.105.565804
- Wilhelms M., Dössel O., and Seemann G. 2011. "In Silico Investigation of Electrically Silent Acute Cardiac Ischemia in the Human Ventricles." *IEEE Transactions on Bio-Medical Engineering* 58 (10): 2961–64. doi:10.1109/TBME.2011.2159381.

- Yan L., Wang Z., Xu Z., Li Y., Tao Y., and Fan C. 2012. "Two Hundred Eight Patients with Apical Hypertrophic Cardiomyopathy in China: Clinical Feature, Prognosis, and Comparison of Pure and Mixed Forms." *Clinical Cardiology* 35 (2): 101–6. doi:10.1002/clc.20995.
- Yeh Y-C., Wang W-J, and Chiou C.W. 2009. "Cardiac Arrhythmia Diagnosis Method Using Linear Discriminant Analysis on ECG Signals." *Measurement* 42 (5): 778–89. doi:10.1016/j.measurement.2009.01.004.
- Zaharov V., Farahi R., Snyder P., Davison B., and Passian A. 2014. "Karhunen–Loève Treatment to Remove Noise and Facilitate Data Analysis in Sensing, Spectroscopy and Other Applications." *Analyst* 139 (22): 5927–35. doi:10.1039/C4AN01300J.
- Zemzemi N., Bernabeu M., Saiz J., Cooper J., Pathmanathan P., Mirams G., Pitt-Francis J., and Rodriguez B. 2013. "Computational Assessment of Drug-Induced Effects on the Electrocardiogram: From Ion Channel to Body Surface Potentials." *British Journal of Pharmacology* 168 (3): 718–33. doi:10.1111/j.1476-5381.2012.02200.x.
- Zhang Q., Chen X., Fang Z., and Xia S. 2016. "False Arrhythmia Alarm Reduction in the Intensive Care Unit Using Data Fusion and Machine Learning." In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 232–35. doi:10.1109/BHI.2016.7455877.
- Zhang Z., and Luo X. 2014. "Heartbeat Classification Using Decision Level Fusion." *Biomedical Engineering Letters* 4 (4): 388–95. doi:10.1007/s13534-014-0158-7.
- Zheng Y., Liu Q., Chen E., Ge Y., and Zhao J. 2014. "Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks." In *Web-Age Information Management*, edited by Feifei Li, Guoliang Li, Seung-won Hwang, Bin Yao, and Zhenjie Zhang, 298–310. Springer International Publishing. http://link.springer.com/chapter/10.1007/978-3-319-08010-9_33.
- Zhou X., Bueno-Orovio A., Orini M., Hanson B., Hayward M., Taggart P., Lambiase P., Burrage K., and Rodriguez B. 2015. "In Vivo and In Silico Investigation into Mechanisms of Frequency Dependence of Repolarization Alternans in Human Ventricular Cardiomyocytes." *Circulation Research*, November, CIRCRESAHA.115.307836. doi:10.1161/CIRCRESAHA.115.307836.
- Zhukov L., and Barr A. 2003. "Heart-Muscle Fiber Reconstruction from Diffusion Tensor MRI." In *IEEE Visualization*, 2003. VIS 2003., 597–602. doi:10.1109/VISUAL.2003.1250425.

10 APPENDIX

APPENDIX A: SUMMARY TABLE OF MACHINE LEARNING METHODS USED FOR ECG SIGNAL	
CLASSIFICATION	191

APPENDIX A: SUMMARY TABLE OF MACHINE LEARNING METHODS USED FOR ECG SIGNAL CLASSIFICATION

Summary table of reviewed machine learning classification methods, along with their objective, dataset, performance and validation.

Study	Classes and goal	Methods	Performance	Validation
Heartbeat classification				
Chazal et al (2004)	Normal, ventricular ectopic beat (VEB), supraventricular ectopic beat (SVEB), fusion of normal and VEB, unknown	Linear discriminants (LD) with QRS-based and time intervals features	SVEB: 75.9% (sensitivity) VEB: 77.7% (sensitivity)	On 50,000 independent beats (MIT-BIH)
Llamedo and Martínez (2011)	Normal, VEB and SVEB classification	LD (RR-intervals and wavelet transform features) with floating feature selection	93% global accuracy	On independent MIT-BIH beats and INCART
Yeh et al (2009)	Normal, left bundle branch block (LBBB), right bundle branch block (RBBB), premature ventricular contraction (PVC), atrial premature contractions	Linear discriminants	96.23% (global accuracy)	On 14 30-min excerpts (MIT-BIH)
Ubeyli (2007)	Normal, congestive heart failure, ventricular tachyarrhythmia, atrial fibrillation (AF)	Support Vector Machine (SVM) with error output correction code and	98.61% accuracy	On 360 independent beats (Physionet)

		discrete wavelet transform		
Melgani and Bazi (2008)	Normal, atrial premature beat, PVC, RBBB, LBBB, and paced beat	SVM optimized by particle swarm optimization	89.72% accuracy	On 40,438 independent beats (from 20 patient records of MIT-BIH)
Asl et al (2008)	Normal, PVC, AF, sick sinus syndrome, ventricular fibrillation, 2° heart block	SVM with heart rate variability features and discriminant analysis feature reduction	99.16% accuracy	On 463 testing segments of MIT-BIH (average over 100 different runs)
Nasiri et al (2009)	Normal, RBBB, LBBB, and paced beat	SVM with principal component analysis and genetic algorithm	93.46% accuracy	50% of MIT-BIH for testing
Ganeshkumar and Kumaraswamy (2012)	Normal, PVC, paced, atrial premature beat, LBBB and RBBB	Random Forest (30 trees) on 150 beats from MIT-BIH	92.16% accuracy	Not validated on independent dataset
de Oliveira et al (2010)	PVC detection	Bayesian network framework using channel fusion	99.69% sensitivity	On QT database (25% - 23,765 beats for testing)
Coast and al (1990)	VEB detection (over American Heart association database)	Hidden Markov Model (HMM) with states corresponding to ECG waveforms or intervals	97.25% sensitivity	On 799 independent beats
Koski (1996)	PVC detection	HMM and broken line	100% accuracy	On only 4

	approximation (30 states)			beats
Andreao et al (2006)	PVC detection	HMM and rule based system	99.79% sensitivity	On 61,543 test beats from QT database
Niwas et al (2005)	Normal, LBBB, RBBB, Atrial Premature beat , SVEB, PVC, AF, Ventricular Fibrillation, Sick Sinus Syndrome, Fusion of VEB and normal	Artificial neural network with heartbeat intervals and spectral entropy features	99.02% accuracy	On 180 (unspecified) independent datasets
Inan et al (2006)	PVC detection	Feed-forward multi-layer perceptron with wavelet transform and time intervals features	96.82% accuracy	On 22 ECG recordings from MIT-BIH
Ubeyli (2009)	Normal, congestive heart failure, ventricular tachyarrhythmia, AF	Recurrent neural network with Levenberg–Marquardt training algorithm and eigenvectors	98.06% accuracy	On 360 beats from Physionet
Lagerholm and al (2000)	Normal, LBBB, RBBB, atrial premature, abberated atrial premature, nodal premature, SVEB, VEB, fusion normal and VEB, ventricular flutter, atrial escape, nodal escape, ventricular escape, unknown	Self organizing networks with Hermite transform and RR intervals features	1.5% of misclassification	Not validated on independent database
Linh et al (2003)	Normal, LBBB, RBBB, atrial premature, VEB, ventricular flutter	TSK fuzzy network with Hermite	96% accuracy	On 3,668 beats from MIT-BIH

	wave, ventricular escape	transform		
Ozbay et al (2006)	Normal, sinus bradycardia, ventricular tachycardia, sinus arrhythmia, atrial premature contraction, paced, RBBB, LBBB, AF and atrial flutter	Multi-layer perceptron with fuzzy clustering neural network architecture	99.9% accuracy	On 5342 segments from 92 patients (MIT-BIH)
Patient diagnosis, monitoring and risk stratification				
Faganeli and Jager (2010)	Patients with transient ischemia episodes	Decision trees with heart rate and Legendre polynomial coefficients features	98.1% sensitivity, 85.2% specificity	Bootstrap method
Rahman et al (2015)	Hypertrophic cardiomyopathy detection on 12-lead ECG	SVM and random forest with 264 time intervals and waveforms amplitude features	precision of 0.84	5-fold cross validation over 10,930 beats
Bailon et al (2003)	Diagnosis of coronary artery disease	Multivariate discriminant analysis with repolarization, depolarization and heart rate variability features	94% sensitivity, 92% specificity	Cross validated estimation (leave one out)
Pourbabaee and Lucas (2008)	Paroxysmal AF episodes detection	Multi-layer perceptron with time interval and morphological waveform features	87.5% accuracy	Over 16 recordings from 2001 Computing in Cardiology challenge
Colloca et al (2013)	AF episodes detection	SVM optimized with grid-search	85.45% accuracy with 100% sensitivity to AF	Series 200 of the MIT-BIH Arrhythmia

				(with 8 AF subjects)
Asgari et al (2015)	AF episodes detection	SVM with stationary wavelet transform	97.0% sensitivity	2-fold stratified cross-validation on MIT-BIH
Acharya et al (2003)	Ischemic/dilated cardiomyopathy, complete heart block, sick sinus syndrome, AF, ectopics, normal	Artificial neural network with fuzzy equivalence	85-95% accuracy	On 66 testing samples
Zheng et al (2014)	Congestive heart failure (from 2-lead ECGs)	Convolutional neural network (CNN)	94.7% accuracy	10-fold cross validation over 15 subjects
Kawazoe et al	Risk of ventricular fibrillation in Brugada syndrome patients	Logistic regression with syncope, r-J interval, QRS duration, and Tpeak-Tend dispersion as features	97.1% sensitivity, 63.0% specificity	Leave-one-out cross validation over 143 patients
Kannathal et al (2003)	Normal, abnormal (PVC, RBBB, LBBB, paced), life threatening (sick sinus syndrome, ischemic heart disease, ventricular fibrillation)	Artificial neural networks with radial basis function	99% accuracy	On 200 independent testing patients
Zhang et al (2016)	Asystole, extreme bradycardia, extreme tachycardia, ventricular tachycardia or ventricular flutter/fibrillation arrhythmia	SVM with genetic algorithm	93% true positive rate	5-fold cross validation over 750 recordings (2015 Physionet Challenge)

Real-time episodes' detection and wearable devices				
Kiranyaz et al (2016)	VEB and SVEB detection	1-D patient specific CNN	98.6% accuracy	On 41,766 testing beats from MIT-BIH
Chauhan and Vig (2015)	PVC, atrial premature contraction, paced, ventricular couplet	Deep long short term memory (LSTM) network	0.975 precision, 0.9645 F-score	Testing set of unknown size
Jeon et al (2014)	Normal beats, AF, myocardial ischemia classification	SVM on ARM processor	95.1% sensitivity	10-fold cross validation over MIT-BIH AF, 2001 and 2004 CinC challenge and STT database
Leutheuser et al (2014)	Normal and abnormal (all MIT-BIH labels that are not normal), on Android devices	Decision tree classifier with heartbeat features	89.6% accuracy	Not validated on independent dataset
Oresko et al (2010)	Normal, RBBB, PVC, paced or fusion of paced and normal beat detection on smartphone	Feedforward multilayer perceptron with QRS morphological beat pattern	99% accuracy for normal – 81% accuracy for fusion	3-fold cross validation over 5421 beats (MIT-BIH)
Oster et al (2015)	Normal and SVEB, fusion and VEB, and unknown	Switching Kalman filters with X-factor mode	99.2% F1-score	Independent validation on INCART