Applications of Machine Learning and Computational Methods in the Prediction of Cardiovascular Remodeling

A DISSERTATION PRESENTED

BY

Eleni E. Angelaki

TO

THE DEPARTMENT OF PHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE SUBJECT OF

PHYSICS

UNIVERSITY OF CRETE

CRETE, GREECE

December 2023

Thesis advisor: Professor G. P. Tsironis Eleni E. Angelaki

Applications of Machine Learning and Computational Methods in the Prediction of Cardiovascular Remodeling

Abstract

Machine learning (ML) is a growing field poised to change the way we practice cardiovascular medicine, providing new tools for interpreting data and making decisions. Cardiac digital images or biosignals defy traditional statistical methods and require the deployment of ML. In this work we show that ML classifiers trained using anthropometric features and ECG-derived features, can: (a) detect abnormal left ventricular geometry, even before the onset of left ventricular hypertrophy; (b) detect hypertension using 12-lead-ECG-derived features; and (c) detect hypertension in populations without cardiovascular disease from single-lead-ECGs. The latter classifiers can be useful in raising awareness in people with undiagnosed hypertension. We then present a computational method to simulate the dynamics of action potential propagation using the three-variable Fenton-Karma model and account for both normal and damaged cells through spatially inhomogeneous voltage diffusion coefficient.

Επιβλέπων: Καθ. Γ. Τσιρωνης Ελενη Ε. Αγγελακη

Εφαρμογες Μηχανικης Μαθησης και Υπολογιστικων Μεθοδων στην προβλεψη της Καρδιοαγγειακης Αναδιαμορφωσης

Περιληψη

Η μηχανική μάθηση (ΜΜ) είναι ένας γοργά αναπτυσσόμενος τομέας που πιθανόν να αλλάξει τον τρόπο με τον οποίο ασκούμε την καρδιολογική ιατρική, προσφέροντας νέα εργαλεία για την ερμηνεία των δεδομένων και τη λήψη αποφάσεων. Οι ψηφιακές εικόνες και τα βιοσήματα της καρδιάς δεν μπορουν να αναλυθούν με τις παραδοσιακές στατιστικές μεθόδους και χρειαζονται εφαρμογές της ΜΜ. Σε αυτήν την εργασία δείχνουμε ότι οι ταξινομητές ΜΜ που εκπαιδεύονται χρησιμοποιώντας ανθρωπομετρικά χαρακτηριστικά και χαρακτηριστικά που προέρχονται από το ηλεκτροκαρδιογράφημα (ΗΚΓ), μπορούν: (α) να ανιχνεύσουν την ανώμαλη γεωμετρία της αριστερής κοιλίας, ακόμη και πριν την εμφάνιση υπερτροφίας της αριστερής κοιλίας, (β) να ανιχνεύσουν την υπέρταση σε πληθυσμούς χωρίς καρδιαγγειακή νόσο, χρησιμοποιώντας χαρακτηριστικά που προέρχονται από το ΗΚΓ 12 απαγωγών, και (γ) να ανιχνεύσουν την υπέρταση σε πληθυσμούς χωρίς καρδιαγγειακή νόσο, από το ΗΚΓ της μίας απαγωγής. Στην τελευταία περίπτωση, οι ταξινομητές θα μπορουσαν να είναι χρήσιμοι για την ευαισθητοποίηση άτομων με αδιάγνωστη υπέρταση. Στη συνέχεια, παρουσιάζουμε μια υπολογιστική μέθοδο για την προσομοίωση της διάδοσης του δυναμικού ενέργειας χρησιμοποιώντας το μοντέλο Fenton-Karma σε καρδιακά κύτταρα με χωρικά ανισότροπο συντελεστή διάχυσης.

ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ ΜΟΥ

Contents

Listing of figures

List of Tables

Originality

Statement

The writing of this thesis is my original work, either with or without collaborators. Some of it is relevant prior or concurrent work included for reference here:

Published papers

[1] Eleni Angelaki, Georgios D. Barmparis, George Kochiadakis, Spyros Maragk- oudakis, Eirini Savva, Emmanuel Kampanieris, Spyros Kassotakis, Petros Kalo- moirakis, Panos Vardas, Giorgos P. Tsironis, and Maria E. Marketou. Artificial intelligence-based opportunistic screening for the detection of arterial hypertension through ecg signals. *Journal of hypertension*, 40(12):2494–2501, 2022.

[2] Eleni Angelaki, Maria E. Marketou, Georgios D. Barmparis, Alexandros Patrianakos, Panos E. Vardas, Fragiskos Parthenakis, and Giorgos P. Tsironis. Detection of ab- normal left ventricular geometry in patients without cardiovascular disease through machine

learning: An ecg-based approach. *The journal of clinical hypertension (Greenwich, Conn.)*, 23(5):935–945, 2021.

Submitted papers

[3] Eleni Angelaki, Maria E. Marketou, Georgios D. Barmparis, Alexandros Patrianakos, Panos E. Vardas, Fragiskos Parthenakis, and Giorgos P. Tsironis. Diagnostic per- formance of single-lead electrocardiograms for arterial hypertension diagnosis: A machine learning approach. 2023.

[4] Eleni Angelaki, Nikos Lazarides, Georgios D. Barmparis, Ioannis Kourakis, Maria E. Marketou, and Giorgos P. Tsironis. T-wave inversion through inhomogeneous voltage diffusion within the FK3V cardiac model. arXiv:submit/5234360, 2023.

Acknowledgments

I HOPE TO EXPRESS MY DEEP APPRECIATION to everyone who supported and accompanied me throughout this incredible journey. Thank you!

I would like to thank my main advisor Prof. George Tsironis; without his initial encouragement I probably would not have sailed on this voyage. His mentorship and valuable guidance on Physics, research, and life as a graduate student, was instrumental in the success of my studies. He always knew when to leave room for growth and when to be firm.

My research involved delving into a topic I initially knew little about, the cardiovascular system, and for this, my other advisor, Prof. Maria Marketou, has offered valuable guidance. Her unwithering support, mentoring, and encouragement saved the day many times. I am also very appreciative to her for paying so much attention to good data collection despite running a busy hospital clinic, among her other numerous duties. Both advisors played an crucial role in helping me find my footing at the intersection between Physics and Medicine.

I would like to thank the rest of the members of my dissertation committee: Prof. George Kochiadakis, Prof. Andreas Zezas, and Prof. Ioannis Kominis, for their support and thought-provoking comments during my thesis defense; Prof. Kostas Makris, for an inspiring course on Statistical Physics; Prof. Issac Lagaris, who, throughout my graduate years, provided encouragement as well as valuable advice on methods of optimization.

I had the privilege to teach data science and machine learning for years in Dr. Pavlos Protopapas' courses CS109A&B. He introduced me to the field of AI and has been a valuable teacher and mentor ever since.

I am honored to have had the opportunity to do research with the following collaborators: George Barmparis, whose encouragement and thoughtful comments helped improve this work, Nikos Lazarides, and Marios Mattheakis, who advised me on all things graduate. I am thankful to David Sondak, at Harvard, for insightful conversations, and to my former co-instructor Chris Tanner, who showed me how joyful teaching can be. A cheerful thank you to my collaborator and friend George Neofotistos whose practical advice and psychological help accompanied me throughout the whole process. I hope to express my special thanks to Javier Zazo, formerly at Harvard, for very interesting discussions about data science and deep learning, and to Prof. Hanspeter Pfister of Harvard, for welcoming me in his group as an Associate. I would like to thank Stefanos Trachanas, who, although not directly involved in my graduate studies, has been an inspirational Physics teacher.

I owe a large part of my research to the medical practitioners who collected the data used in this work. Sincere thanks to Konstantinos Fragkiadakis, Spyros Maragkoudakis, Evangelos Zacharis, Anthi Plevritaki, Emmanouil Kampanieris, Petros Kalomoirakis, and Spyros

Kassotakis. I never lost sight of the fact that each data point represented a patient, and I am grateful to all of them as well.

I also hope to thank Maria Matalliotaki for administrating the physics graduate program professionally and with kindness.

The computations in this dissertation were run on the FAS RC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University.

I am thankful to my Rethymno family and friends: my brother Kostas, and sister-in-law Manolia, who kept my spirits high when I needed it; my friends Katerina, Argiro, and Chrysanthi, for their encouragement.

A huge thank you to my husband, Efthimios Kaxiras, for his support, patience, and illuminating discussions. Warm thanks to my daughter Daphne and son Vassilis, who were always there with love, encouragement, and inspiring conversations.

List of Abbreviations

BMI	body mass index
BSA	body surface area
CVD	cardiovascular diseases
soko	Sokolow-Lyon voltage (RV_5+SV_1)
BMI-soko	BMI-modified Sokolow-Lyon voltage (BMI divided by SV_1+RV_5)
BMI-adj-soko	BMI-adjusted Sokolow-Lyon criteria
BMI-adj-cornell	BMI-adjusted Cornell criteria (BMI multiplied by RaVL+SV_3)
RaVL-QRS	Cornell product (RaVL+SV_3 multiplied by QRS complex duration)

1

Introduction

Humans have always revelled at explaining the physical world through Physics and Math. Initially with just our brains, then using computers programmed to follow a list of formal, mathematical rules, called algorithms, designed to solve straight-forward problems. In the last decades, these algorithms have evolved to solving difficult, even intuitive tasks, such as recognizing faces in a photo, making conversation, discovering patterns in hetero-

geneous data, or detecting diseases in medical imaging modalities. Algorithms aiming at creating systems that mimic cognitive functions akin to humans, constitute the broad term of *artificial intelligence* (AI). Today, AI is a growing field with many practical applications and active research areas including in cardiology.

Cardiovascular diseases (CVDs) are a group of disorders of the heart and blood vessels which may lead to deaths due to heart attacks and strokes – according to the World Health Organization, CVDs are the leading cause of death globally, taking an estimated 17.9 million lives each year.[1] Many patients are asymptomatic until the late stages when they present with acute, life-threatening disease. Early detection, risk assessment, and behavior/health-factor intervention are essential to lower CVD morbidity and mortality. Cardiac remodeling is considered an important aspect of CVD progression and is therefore emerging as a significant therapeutic target.[2,3,4,5] More specifically, arterial hypertension is associated with a spectrum of cardiac geometric adaptation matched to systemic hemodynamics and ventricular load, which has important prognostic implications.[2,3,4,5]

The heart, this incredibly complicated physical system that keeps us alive, produces a synchronized mechanical contraction, or heartbeat, initiated by a self-generated electrical stimulus that propagates through the cardiac muscle creating what is called an *action potential* (AP). APs are macroscopically traced via the body surface *electrocardiogram* (ECG). One of cardiology's most inexpensive, easy to use, and noninvasive modalities, the EKG, enables physicians to look for patterns that correlate with disease manifestation, providing unique information that cannot be obtained by any other technique. An ECG can be rapidly recorded with computer-enabled portable equipment in a so-called 12 lead, or

single-lead configuration, the latter supported also by wearable devices.

The motivation for this dissertation is first to understand the interplay between experimental observations in ECG signals and disease manifestation, and second, to gain insight into the connection of cardiac elements on a cell level and the macroscopically observed ECG signals. The content is arranged as follows: Chapter 1 discusses the purpose of this work and gives its outline, and Chapter 2 provides a brief description of the function of the heart with some details on ECG acquisition. Chapter 3 gives an overview of AI, its subcategories, and methods of implementation. Examples of AI in medicine and specifically in cardiology are given.

Chapter 4 discusses the methods common to all the experimental studies of the next three Chapters. Chapter 5 describes cardiac remodeling and the role of ECG in predicting abnormal left ventricular geometry, and Chapter 6 presents the results of opportunistically detecting the existence of arterial hypertension, using a 12-lead. Wrapping the machine learning part is Chapter 7, where we detect hypertension from a single-lead ECG.

Chapter 8 provides an introduction to the mathematical concepts involved in AP propagation. Chapter 9 employs a low-dimensional mathematical model for the AP, and constructs the so-called *pseudo-ECG*, reproducing some features of the body surface ECG. We then alter the parameters of the model to mimic defected regions in the cardiac tissue. In this Chapter we also present our findings in T-wave inversion when changing the diffusivity of the cell medium. Last, in Chapter 10 we make our conclusions and discuss about future work.

2

Cardiac anatomy and physiology

The heart is located between the lungs in the middle of the chest, enclosed in a double layered membrane called the pericardium, whose inner layer is attached to the heart muscle. A coating of fluid separates the two layers of the membrane, letting the heart move as it beats. Internally, the heart is composed of four chambers: the left and right atrium, located in the upper part, and the left and right ventricle located in the lower part, as shown

in figure 2.1.

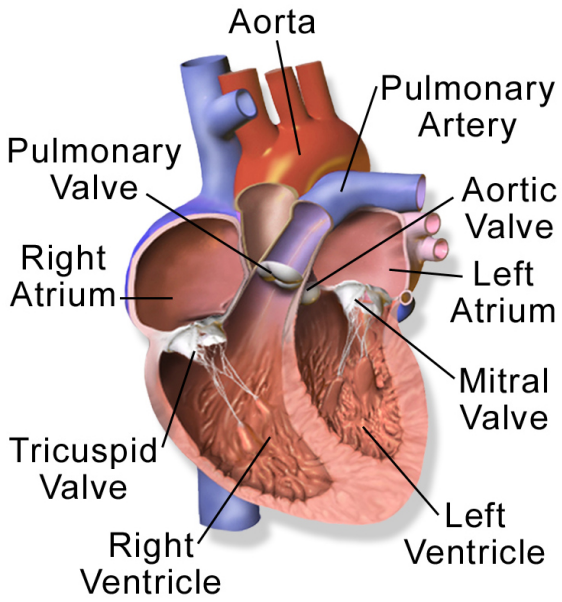

Fig. 2.1: Anatomy of the heart as a double pump. Image from [6], CC.

When relaxed, the right atrium receives deoxygenated blood that has circulated throughout the body including the heart itself after delivering oxygen and nutrients, while the left fills with newly oxygenated blood from the lungs. When the atria contract they push the blood through valves into the relaxed ventricles, which then contract to push blood out. The right ventricle pumps blood to the lungs while the left one pumps blood to the body. This continuous cycle of synchronized contractions repeats approximately 60-100 times per minute and circulates close to six liters of blood per minute through the body.

2.1 The electrocardiogram

British physiologist Augustus D. Waller of St Mary's Medical School, London, recorded the first human electrocardiogram in 1887, using a capillary electrometer from Thomas Goswell, a technician in the laboratory[7], thus inventing a cornerstone in cardiovascular diagnostics. In early 20th century, Dutch physician Willem Einthoven, considered the father of electrocardiography as we know it today, coined the term "electrocardiogram". Einthoven, using an improved electrometer and a correction formula, distinguished five deflections which he named P, Q, R, S and T[8]. His revolutionary invention transformed cardiac diagnostics, earning him the Nobel Prize in Physiology or Medicine in 1924.

Electrodes placed in specific locations on the extremities and torso detect the currents reaching the skin and are configured in *leads* to record the voltage differences between those electrodes. Today, recordings of 12 leads are produced, and the standardization of the ECG waves (P, QRS, and T) allows for consistent readings. ECG recording devices, called *electrocardiographs*, evolved from cumbersome, large machines, to small and portable ones. Their recordings also progressed to include not just paper outputs but also digital files that can be interpreted, stored, and analysed by computer. Digital interpretation of ECGs via computational methods and ML further enhance diagnostic accuracy. The addition of wearable technology, mobile apps, and telemedicine platforms, make ECG readings accessible even beyond the clinical setting.

The use of multiple leads helps capture the activity of the heart through multiple views, the outputs of which are amplified, filtered, digitized, and displayed to produce an ECG

recording. Leads are divided in two groups:

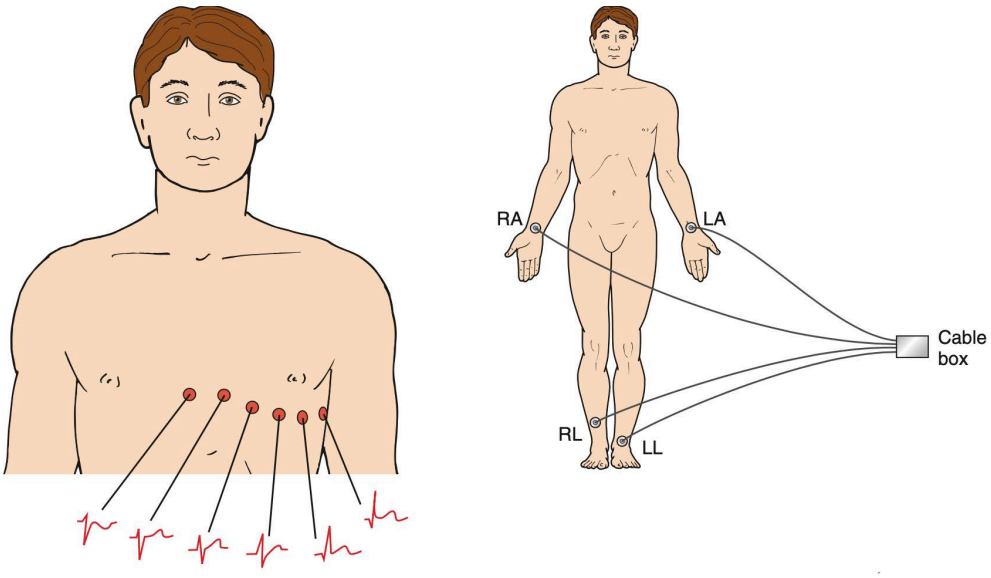

Fig. 2.2: Electrodes attached to the body surface to take an ECG. Chest leads give a close multidimensional view of cardiac electrical activity (left). In the limp leads (right), the right leg (RL) electrode functions solely as a ground to prevent alternating current interference. LA stands for left arm; LL for left leg; and RA for right arm. Image from[9]. Used with permission from Elsevier.

(a) The six *limb* leads consisting of the standard *bipolar* (I, II, III) and *augmented* (aVR, aVL, and aVF) leads. The bipolar leads record the voltage difference between two extremities as follows: lead I records the difference between the left arm (LA) and the right arm (RA) electrodes, lead II records the difference between the right arm (RA) and the left leg (LL) electrodes, and lead III records the difference between the left leg (LL) and the left arm (LA) electrodes. Dr. Emanuel Golberger invented the three augmented leads which are unipolar in the sense that those electrodes record the voltage in one location rather than two. To be exact, they record the difference between that location and the middle of the chest considered to have *zero* potential. Fig. 2.2 (right) shows the placement of the limb

electrodes on the body. Leads I, II, and III can be represented schematically as a triangle, called *Einthoven's triangle*, showing the spatial orientation of those three leads. The augmented leads can also be placed on Einthoven's triangle as shown in Fig. 2.3 (left).

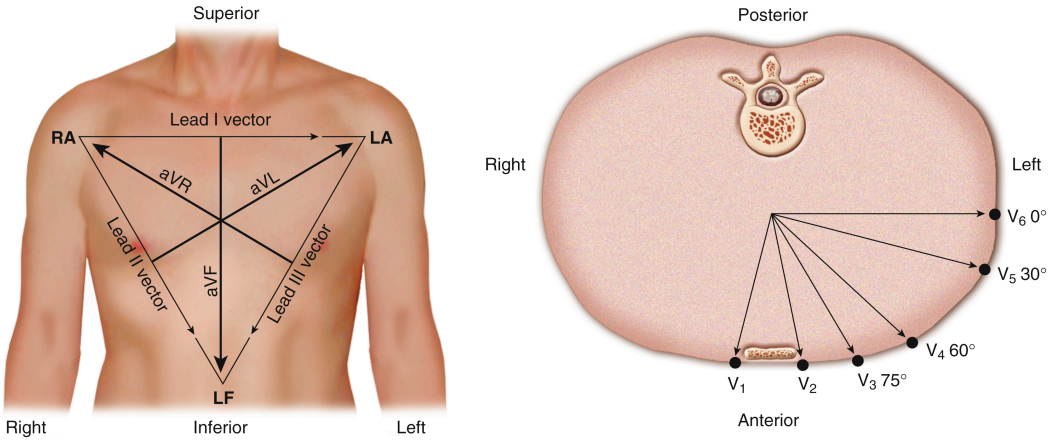

Fig. 2.3: Einthoven's triangle extended to include not only the three standard limb leads, but also the three augmented leads (left). Vectors for the six unipolar precordial leads are shown on the right. LA, Left arm; LF, left foot; RA, right arm. Image from [10]. Used with permission from Elsevier.

(b) The six *precordial* leads (V1 to V6), also called the *chest* leads, record the electrical currents of the heart via electrodes placed directly on the chest wall. Shown in Fig. 2.2 (left), and Fig. 2.3 (right), they are unipolar and, being closer to the heart, better suited for the detection of certain conditions such as left ventricular hypertrophy (LVH).

The ECG connects basic science and life-saving decisions made by physicians, in the most fascinating way. The criteria for interpreting an ECG in a clinical setting have evolved for over a century and are based on correlations between physiologic findings and feature measurements in large populations. Several different criteria with varying accuracy have been proposed for clinical conditions, sometimes, based purely on empirical knowledge.

Consequently, ECGs can be viewed using in different frames of reference. One, is formulated from the statistical probability that a physiologic abnormality exists based on the phenomenological findings on the ECG, and the other, uses knowledge accumulated from studies in cardiac electrophysiology. A typical output ECG reading is shown in Fig. 2.4.

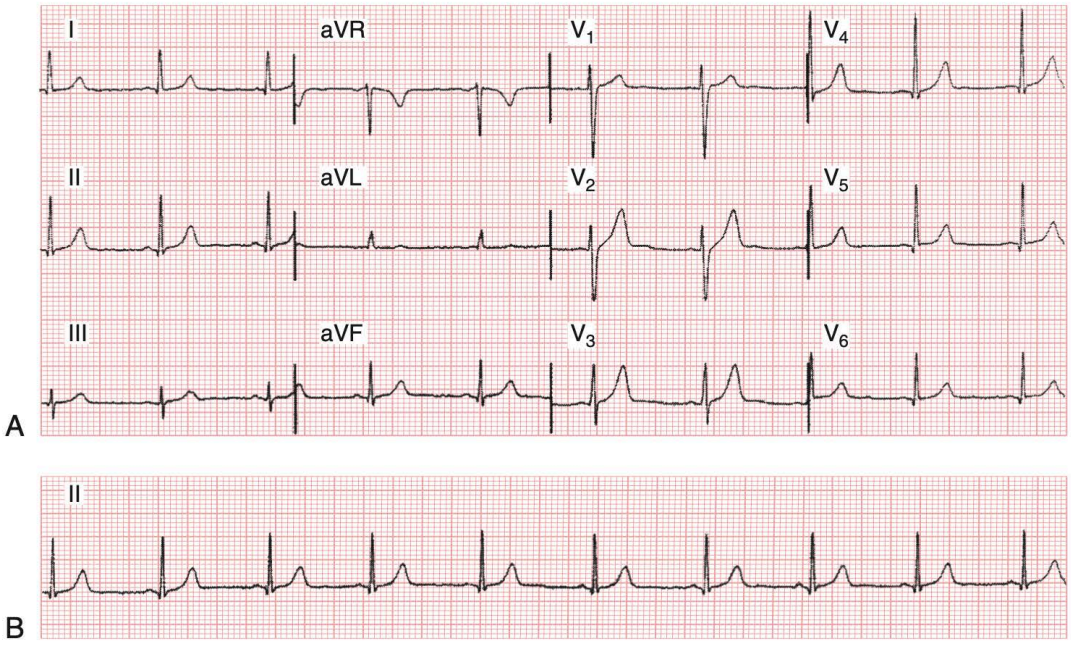

Fig. 2.4: A. Sample ECG showing all 12 standard leads. Each lead is displayed for a fourth of the total time which is usually 10 sec. **B**. A closer look at a whole 10 sec tracing for lead II. The average heart rate for this person is $50 - 60$ beats per minute. Image from [9]. Used with permission from Elsevier.

2.2 The action potential

The study of the heart's electrical system is about excitable media, a class of nonlinear complex systems, whose behavior is described in terms of the *potentials* and *currents* they produce, observed in the medium's interior, across their membranes, and in their surrounding

space. Charge-carrying ions within electrolytes – mainly ions of sodium and potassium – constitute currents in living tissue. Except where it crosses the membrane, current is mostly one directional, moving along the fiber's axis. Excitable media are composed of elementary segments each of which has three basic characteristics: a resting state, an excitation threshold, and a diffusive-type coupling among neighboring segments.[11] Myocardial cells exhibit all three characteristics: (1) They have a resting potential of about -80 to -90 mV with respect to surrounding extracellular fluid. (2) Their threshold for excitation of about -70 mV ensures that only external stimuli above a certain value induce the cell to change its state from resting to excited. When in an excited state, the cell produces a pulse in time which propagates without damping and whose shape and nature are determined by the nonlinear properties of the medium and not by the form of the external excitation. This change happens due to charged cations moving in and out of the cell, a movement controlled by specialized proteins embedded in the cell membranes that make up ion channels, transporters, and pumps, collectively referred to as *gates*. (3) propagation of the AP involves the diffusion of ions via the gap junctions, as well as their transmembrane transfer through the gates.

At the top of Fig. 2.5 we see the sodium (Na^+) entry via the fast sodium channel which is responsible for the rapid upstroke of the action potential in ventricular cells. Calcium (Ca^{++}) enters the cell via the calcium channel, depicted next to the sodium channel, and is responsible for the depolarization of the muscle cells. Potassium (K^+) repolarizes the cell by exiting via potassium channels. Potassium channels are the only ones that are always open in order to maintain the resting potential. All channels are considered *gated* because

they open and close at different rates depending on the transmembrane voltage and on the length of time during which the cell has maintained certain voltages.

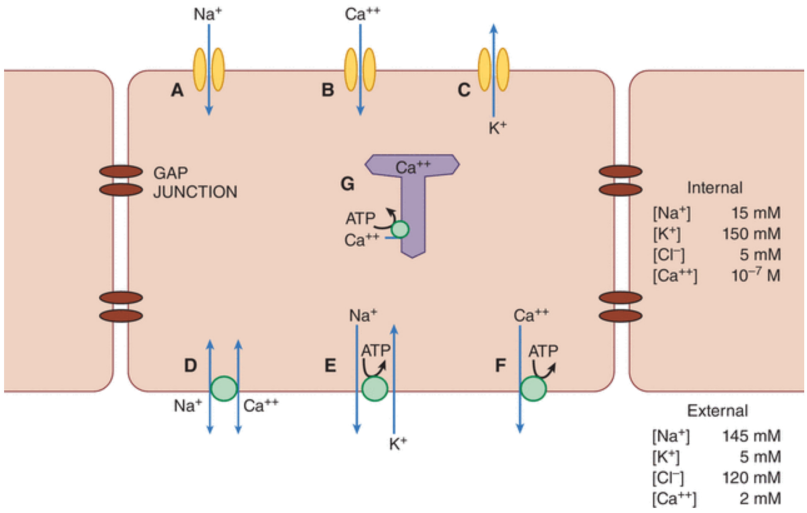

Fig. 2.5: Ion channels, gap junctions, and transporters in the myocyte. At the top of the Figure we can see the sodium (Na^+) entry via the fast sodium channel which is responsible for the rapid upstroke of the action potential. Image from [12] used with permission from Wolters Kluwer.

Effectively, gap junctions slow down propagation by having a larger resistance than the cytoplasm. They allow heart cells to function in a coordinated, synchronized manner, ensuring they're electrically connected as a single unit. These junctions are predominantly found at the ends of cells. As a result, the anatomic characteristics of groups of cardiac muscle differ based on the orientation they're studied from, a trait known as "anisotropy". Conduction velocity is about two to three times faster along the length of the fiber compared to across its width.

There are 5 types of cardiac muscle which differ anatomically and functionally: sinoatrial (SA) node, atrioventricular (AV) node, His-Purkinje system, atrial muscle, and ven-

tricular muscle.[13] The myocardium, the muscle in the walls of the heart chambers, is the thickest part of the heart muscle – thicker in the ventricles and thinner in the atria; its cells, called myocytes, are roughly shaped as cylinders and in the ventricles are 100 μm long and 25 μm wide[14]. Each cell contains myofibrils, long chains of individual sarcomeres which are the basic contractile element of the cell, the part that makes the heart contract. Sarcolemma, as the cell's membrane, named defines the limits of the cell and separates its exterior from its interior, while also acting as an electrical parallel plate capacitor. Its asymmetric lipid molecule bilayer forms an interior insulating core. Myocytes contain regions called intercalated discs (IDs), present only in cardiac cells. Shown along with sarcomeres in Fig. 2.6, IDs appear as dark transverse lines that cross chains of cardiac cells at various intervals and include gap junctions, complexes that metabolically and electrically connect adjacent cardiomyocytes.

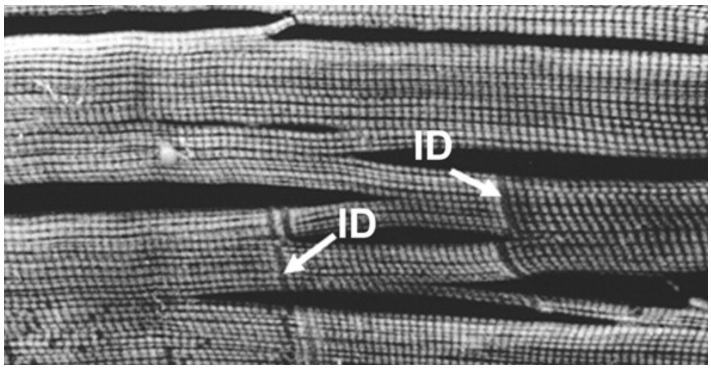

Fig. 2.6: Electron imaging of normal human left ventricular cardiomyocytes. Intercalated discs (arrows) and sarcomere striations are clearly seen; magnification ×600. ID indicates intercalated disc. Image from[15].

Normally, the AP originates from a group of specialized pacemaker cells, found in high proportion of the right atrium, and constituting the sinoatrial node (SAN), as shown in

Fig. 2.7. From the SAN, the electrical signal spreads through the right atrium into the left

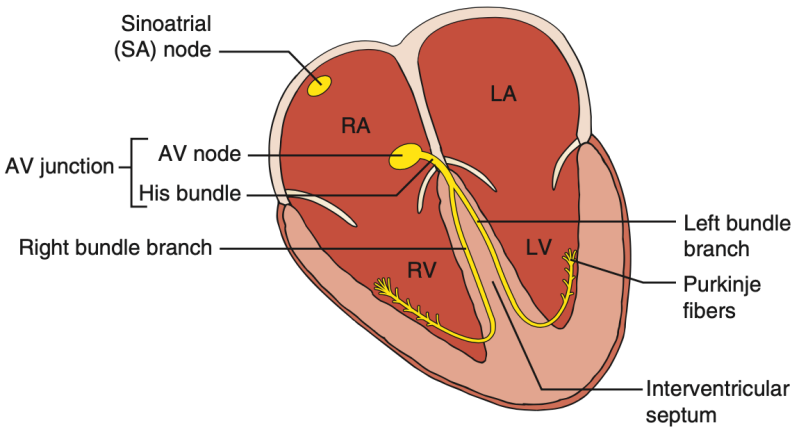

Fig. 2.7: Anatomy of the heart as an electrically timed pump. Image from [16] used with permission from Elsevier.

atrium and then to the atrioventricular (AV) node which, under normal circumstances, is the sole electrical conduction point between the atria and the ventricles. The AP proceeds to the bundle of His, and from there to the left ventricle via the left bundle branch, as well as the right ventricle via the right bundle branch. The conduction of the electrical signal in the ventricles is rapid due to specialized cells called Purkinjie fibers that ensure simultaneous depolarization and therefore contraction of the ventricles.

Life is sustained due to the reliable propagation of APs across the myocardium, which ensures its coordinated excitation and contraction, the heartbeat, as mentioned. The AP is essentially an electrical disturbance, which propagates over long distances, and once initiated by excitation from a stimulus current, preserves its propagation independently of the triggering stimulus – achieving an "autopreserving" status. The resting potential for the membrane is around -85 mV and is depicted by phase 4 in Fig. 2.8. To initiate the

AP, the stimulus current must assume a threshold value of certain amplitude, between -60 to -70 mV, after which a dramatic change in the potential is shown by the rapid upstroke of phase 0 in Fig. 2.8. A fast influx of Na^+ ions through the membrane depolarizes the membrane by over 100 mV in only milliseconds (phase 1 in in Fig. 2.8). Then the fast Na^+ channels close and the slow Ca^++ ion channels open to let the influx of Ca^++ as well as some Na^+. These slow currents are accompanied by K^+ leaning the cell in the plateau phase (phase 2 in Fig. 2.8). Then the Ca^++ stops and rest state is restored by the outward flux of K^+ (phase 3 in Fig. 2.8).

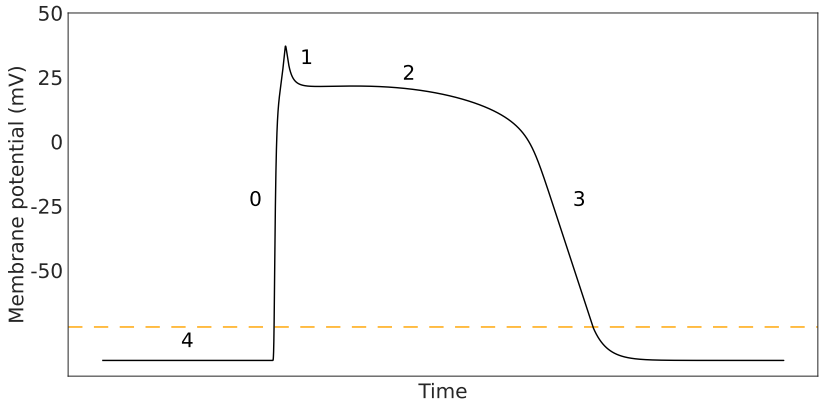

Fig. 2.8: Schematic representation of AP. The resting potential is represented by phase 4. Phase 0 is the rapid upstroke of depolarization which happens because of Na^+ influx; a transient outward potassium current is responsible for partial repolarization during phase 1; slow Ca^{++} influx balanced by K^+ efflux results in the plateau of phase 2; and final rapid repolarization results largely from further K+ efflux during phase 3. The orange dotted line denotes the threshold potential for cell firing

Alterations in the electrical properties of ventricular tissue form the basis of ischemic arrhythmogenesis producing changes in the AP pulses and the ECG.[17,18] Such alterations, involving, for example, the remodeling of ionic currents due to changes in intracellular and

extracellular ionic concentrations, has been studied in the literature.[19] Spatial heterogeneity such as cell-to-cell decoupling, occurring usually in later stages of ischemia, has also been shown experimentally to lead to propagation disruptions and a reduction in conduction velocity.[20]

Fig. 2.9 shows the cardiac conduction system as a network of specialized cells comprising of the sinoatrial (SA) node, the artial muscle, the atrioventricular (AV) node, the His bundle and its bundle branches, the Purkinjie fibers, and finally the ventricular myocytes.

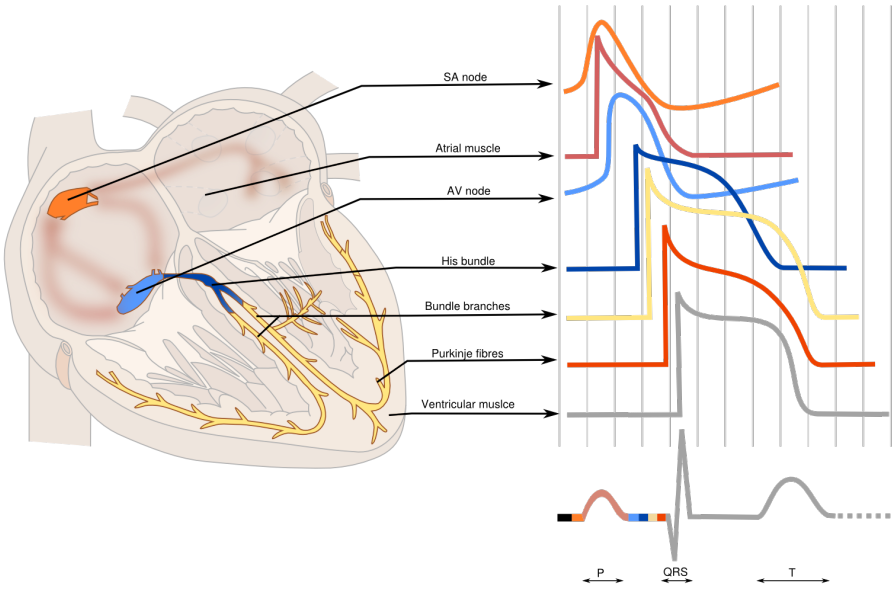

Fig. 2.9: Cardiac action potential (AP) shown for different types of cardiomyocytes, and how they relate to the electrocardiogram (ECG). Depicted is the membrane potential w.r.t. time, for the duration of a single heartbeat. P waves relate to the depolarization of the atrial myocytes, the QRS complex relates to the depolarization of the ventricles, and T waves relate to the repolarization of the ventricles. We notice that, the AP of the ventricular cells, depicted by the grey curve at the bottom of the series of curves, has a longer duration than that of the SA node, drawn as the top curve; also, the Purkinje cell AP is similar to the ventricular action potential except for a sharper initial peak.These relations are color-coded in the small realistic electrocardiogram (ECG) at the bottom right of the image. Image courtesy of Dr. De Voogt and ECGpedia.org.

We can see that AP curve shapes are different for each type of cardiac cell. In the small realistic electrocardiogram (ECG) at the bottom right of the image, we see how these AP curves are associated with parts of an actual ECG (introduced in section 2.1).

Disruptions in AP propagation are the manifestations of underlying cardiac abnormalities, e.g. in myocardial ischemia, one such abnormality, the blood supply to the heart's

coronary arteries cannot meet the demand. The basis of ischemic arrhythmogenesis is alteration in the electrical properties of ventricular tissue, producing changes in the action potential and the body surface ECG[17,18]. One such alteration, the remodeling of ionic currents due to changes in intracellular and extracellular ionic concentrations, has been studied in the literature[19]. In addition to ionic remodeling, spatial heterogeneity such as cell-to-cell decoupling, occurring usually in later stages of ischemia, has been shown experimentally to lead to propagation disruptions and a reduction in conduction velocity[20].

3

Artificial Intelligence in cardiovascular medicine

The term AI defines an interdisciplinary area that includes the techniques we call machine learning (ML). In today's literature, those two terms are often interchangeable, justified by the absence of practical applications in the area outside of their intersection; an area reserved for Artificial General Intelligence (AGI), a theoretical form of AI where machines would have an intelligence similar to humans, and Artificial Super Intelli-

gence (ASI) where machines would surpass the intelligence of the human brain.

3.1 Machine learning

ML is defined as a set of mathematical algorithms that enable computers to detect complex patterns in data and then use those patterns to predict future behavior, thus assisting humans in decision-making, while automating repetitive human labor. ML focuses more on making predictions, whereas statistics is geared towards making inferences about data. ML algorithms, usually called *models*, improve at tasks with experience; the more you train them, the better they become. Models such as *logistic regression* (LR), use hand-crafted features to determine outcomes, e.g., whether to recommend cesarean delivery[21]. The performance of these simple models depends on the representation of the data, as in our previous example, the *features* that the doctor will choose to input, such as the absence of a uterine scar. LR learns how each of these features of the patient correlates with various outcomes. More complicated algorithms such as *Random Forests* (RFs) still use predefined features but can handle more complicated and nonlinear relationships between them.

When models discover not only the mapping from representation to outcome, but also the representation itself, we have what is called *representation learning*. For some tasks, such as interpreting images, learned representations often result in much better performance than can be obtained with hand-designed features. When asking a model to discriminate cats from dogs, we do not describe explicitly how a cat looks, we just show the model many images of cats with the label "this is a cat". By acquiring knowledge from experience and learning complicated concepts out of simpler ones, we avoid having to formally specify

all the knowledge that the problem requires. Model architectures in this category include "shallow", meaning with 1-2 *layers*, *feed forward neural networks* (FFNN), and *shallow autoencoders*. For problems with a lot of existing human expert knowledge, e.g., data in a tabular form, adding that data to the algorithm might increase its accuracy.

Last in our tour of the main types of AI, is *deep learning* which is expressed by "deep", meaning with many layers, FFNNs, *convolutional neural networks* (CNN), *Recurrent neural networks* (RNN), and a zoo of other architectures.[22] See Fig. 3.1 for a drawing of a recap of the basic AI types.

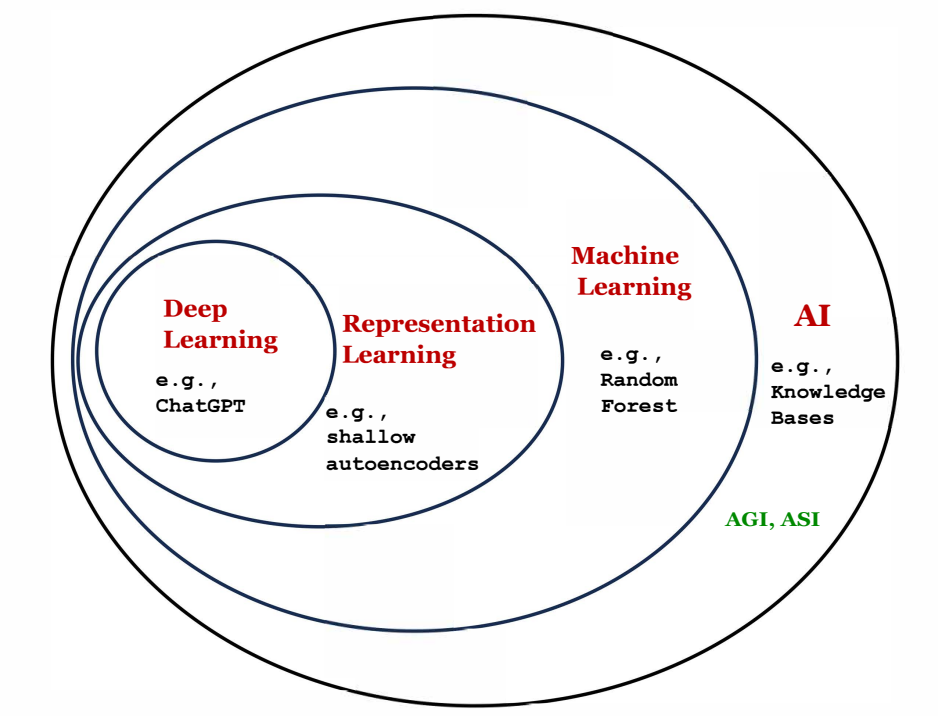

Fig. 3.1: A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning, which is used for many but not all approaches to AI. Each section of the Venn diagram includes an example of an AI technology. Image adapted from the book "Deep Learning".[23] Used with permission from MIT Press.

In terms of the relationship between the type of training and the kind of data used, ML is usually divided into three main categories, supervised learning, unsupervised learning, and reinforcement learning. Fig. 3.2 shows a schematic representation of these three categories, the details of which are explained in the following paragraphs.

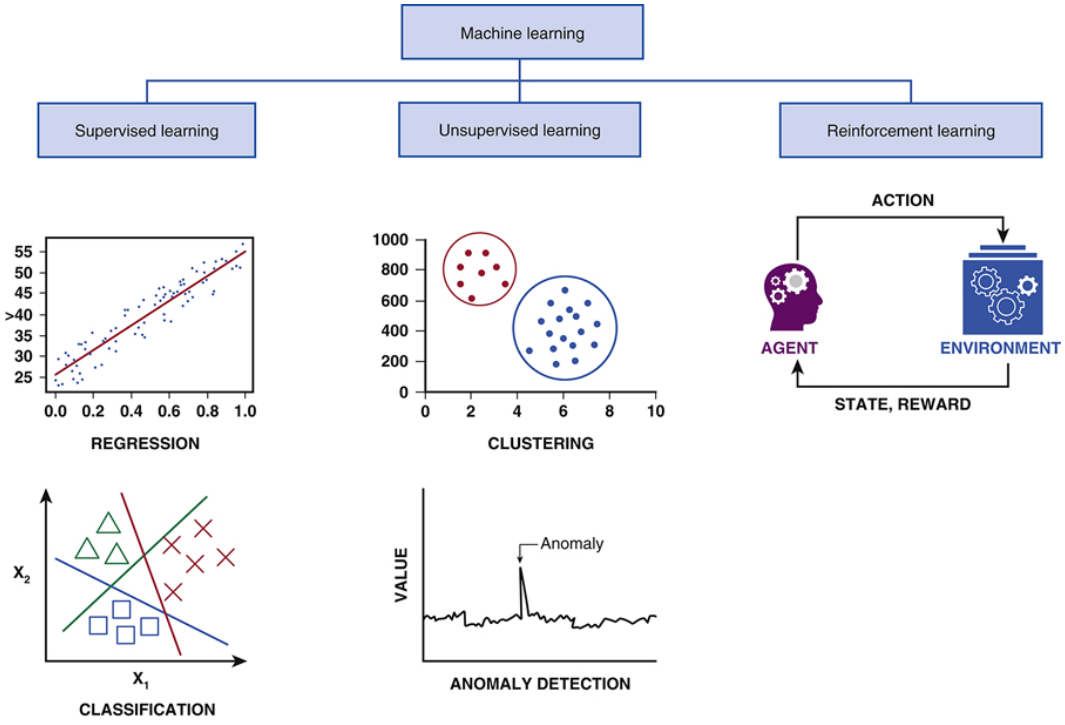

Fig. 3.2: Graphic chart depicting the three major subtypes of ML: supervised learning, unsupervised learning, and reinforcement learning. Image from[9]. Used with permission from Elsevier.

In *supervised learning* the algorithm is given a set of input–output pairs

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N},$$

where N is the number of samples in the dataset, also called the training set, and the goal

is to learn a mapping from $\mathbf{x}$ inputs to y outputs. The input $\mathbf{x} = \{x_1, x_2, x_3\}$ is a vector of values called *predictors* or *features*, and y_i is a single value called the *response variable*. For example, x_1 could be the `age` of a patient, x_2 their `BMI`, x_3 their `blood pressure`, and the response variable could be $y = 1$, code for a `hypertensive` person, or $y = 0$ for a non-hypertensive (`normotensive`). Most algorithms assume that the response variable is either *categorical*, having values in one of $\mathcal{K}$ different classes or categories, or *continuous*, taking on numerical values. We refer to problems with a continuous response as *regression* problems, and those involving a categorical response as *classification* problems. In our previous example, y was categorical, assuming values from the finite set $y_i \in \{0, 1\}$. The algorithm is trained on this set of data, the metric of good performance being how close to the real output are the outputs of the algorithm when applied to an unseen test set, i.e., the error rate:

$$\frac{1}{N}\sum_{i=1}^{N} \mathtt{I}(y_i \neq \hat{y}_i),$$

or otherwise, the fraction of incorrect classifications, where $\hat{y}_i$ is the class label predicted for the ith observation by the model, and y_i is the true class for the same observation. The indicator variable `I` is equal to 1 if $y_i \neq \hat{y}_i$, and 0 if $y_i = \hat{y}_i$. In other words, if $y_i = \hat{y}_i$, then the ith observation was classified correctly by our model; if $y_i \neq \hat{y}_i$, it was *misclassified*. For more on metrics see Section 3.1.2.

The success of the supervised technique relies heavily on the availability of large annotated datasets.

In *unsupervised learning* we only have inputs

$$\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^{N},$$

but no outputs, and the main goal is to find *patterns* in the data. This, of course, is not well defined, as we might not know what kind of patterns we are looking for. Observing patterns in the data allows a deep learning model to cluster inputs. Fig. 3.3 shows a dataset with two features, $\mathbf{x} = \{x_1, x_2\}$, with a clear, non-overlapping (left), and a not-so-clear (right) clustering tendency for three groups.

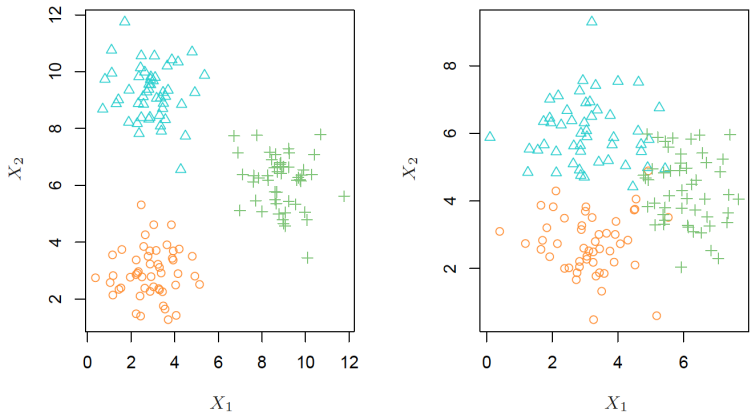

Fig. 3.3: Clustering tendency of dataset with two features. Each group is shown using a different colored symbol. Left: The three groups are well-separated. In this setting, a clustering approach should successfully identify the three groups. Right: There is some overlap among the groups. Now the clustering task is more difficult. Image from[24]. Used with permission from Springer.

One obvious advantage of unsupervised learning is that it can leverage large amount of unlabeled data that would otherwise require a large effort to be labeled.

Other ML techniques include: (a) self-supervised learning, where a model is trained on a task using the data itself to generate supervisory signals, rather than relying on external

labels, e.g., when we hide a portion of a sentence and the model is asked to generate it; (b) semi-supervised, where the data consists of a small number of labeled examples and a large number of unlabeled examples, requiring a combination of supervised and unsupervised techniques.

Although large sets are abundant in tasks such as common object identification (labeled photos of cats and dogs), in medicine, there is a plethora of data being produced with little or no labels, while the opposite, annotated medical datasets, are scarce.[25] Labeled medical datasets are generally in the hundreds of samples and rarely in the thousands or millions. *Transfer learning* offers a way to overcome this obstacle. With this method, model parameters trained on other tasks, are transferred to a medical task with some extra training. For example, we could use a model architecture useful for reading text or music and tweak it to read ECG voltage traces. Or, image models trained on the `ImageNet`, a large visual database designed for use in visual object recognition software, can be used to read clinical imaging photos. However, the characteristics that the model learns might not be relevant to the medical task at hand. Another technique to compensate for the lack of large datasets, is *data augmentation* where you supplement the data with "alternative views" of them, e.g., one could apply transformations to an ECG to obtain these alternative views, as shown in Fig. 3.4. What types of augmentation are suitable for the particularities of ECG nature, is debatable, though.

Unlabeled datasets can be leveraged to build *self-supervised* models,[25] models that learn from data without labels. For example, one could occlude a portion of an ECG and train the model to "generate" it. This way, a good representation of the ECG can be learned

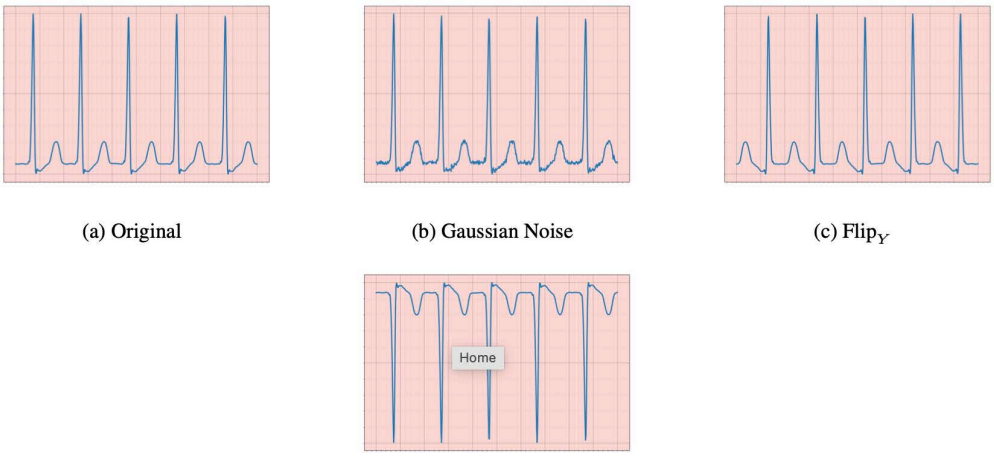

Fig. 3.4: Transformation applied to an ECG segment to augment training a) original, b) with additive Gaussian noise, c) after being flipped temporally, Flip$_Y$, and d) after being flipped along the x-axis, Flip$_X$. Image from[26].

by the model and later used for supervised learning with small amount of labels. Another self-supervised technique, is *contrastive learning*, where the primary objective is to make the model associate similar samples and disassociate dissimilar ones. Similar and dissimilar samples can be generated using augmentation, as previously mentioned, and as shown in Fig. 3.5.

The third large category of ML is *reinforcement learning*. Large language models (LLMs) in their final training stage, have humans rate their responses in a feedback loop (reward or penalty), thus *reinforcing* correct behavior. ChatGPT's[27] impressive behavior in making human-like conversation is based on fine-tuning by lots of people rating its responses. ChatGPT can, for example, write computer programs for processing and visualizing data, translate foreign languages, decipher explanation-of-benefits notices and laboratory tests

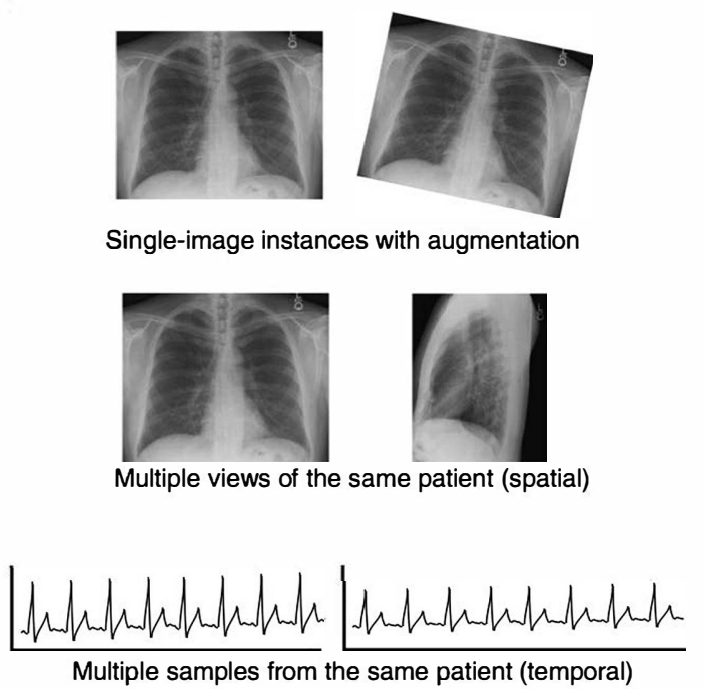

Fig. 3.5: Examples of positive data pairs used in contrastive learning. From top to bottom: two rotations of the same X-ray image; two spatial views of the same patient; two time points of an ECG trace. Image from [25].

for readers unfamiliar with the language used in each, and, perhaps controversially, write emotionally supportive notes to patients.[28] At its present form, it is difficult to know whether ChatGPT's answers are grounded in appropriate facts.

Given the inefficiency of the human involvement, scientists are looking for other ways to train LLMs, such as having AI systems rate each other's responses, taking the human out of the loop.

3.1.1 Data science

Data science is a broader set of tools for making sense of complex datasets to which machine learning is an important, but not the sole, component. In recent years, we have seen a staggering increase in the scale and scope of data collection across virtually all areas of science and industry.[24] Data science uses statistical methods, so, since there is terminology overlap, we will talk about the three types of datasets we use in this work. The training set, the validation set, and the test set. First, we divide the dataset into training and test sets. We put the test set away and use it only once to evaluate the performance of our final model(s). The reported metrics are always on the test set. The training set can be further divided into training set and validation set either explicitly or via cross-validation. Explicitly is when we have enough data to "spare" some for the validation set. The model is fit on the training set, and the fitted model is used to predict the responses for the (known) observations in the validation set. Depending on the resulting error rate, typically assessed using mean squared error (MSE) for quantitative responses, the model is trained again optimizing for this error. If we do not have enough data, we use what is called *k-fold cross-validation*. This method randomly divides the dataset into k groups of equal size, each of which is successively treated as the validation set, while the model is trained on the remaining $k - 1$ groups. The MSE is then computed on the observations in the held-out group. This procedure is repeated k times resulting in k estimates of the test error, MSE_1, MSE_2, etc. The total error

is computed by averaging these values,

$$\frac{1}{k}\sum_{i=1}^{k}\mathrm{MSE}_i.$$

Fig. 3.6 illustrates this approach.

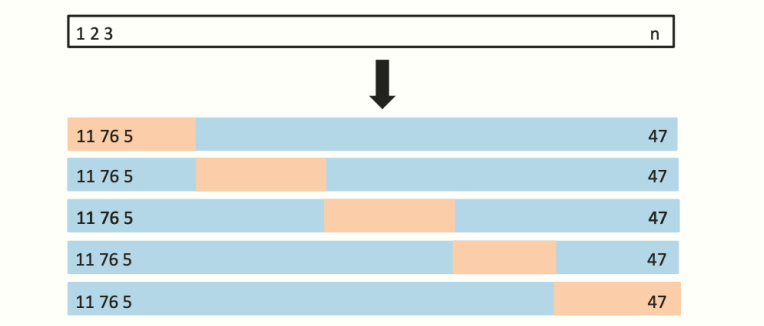

Fig. 3.6: A schematic display of 5-fold CV. A set of n observations is randomly split into five non-overlapping groups. Each of these fifths acts as a validation set (shown in beige), and the remainder as a training set (shown in blue). The test error is estimated by averaging the five resulting MSE estimates. Image from.[29] Used with permission from Springer.

3.1.2 Classification metrics

Classification metrics are based on the number of correct predictions, namely, true positive (TP), true negative (TN), false positive (FP), and false negative (FN) outcomes. They

include

$$\text{sensitivity or true positive rate} = \frac{\text{TP}}{\text{TP+FN}}$$

$$\text{specificity or true negative rate} = \frac{\text{TN}}{\text{TN+FP}}$$

$$\text{false positive rate} = 1 - \text{specificity}$$

$$\text{accuracy} = \frac{\text{TP+TN}}{\text{TN+FN+FP+TP}}$$

Sensitivity and specificity are measures of a model's ability to correctly classify a person as having a disease or not having a disease. A highly sensitive model means that there are few false negative results, and thus fewer cases of disease are missed. The specificity of a model is its ability to designate an individual who does not have a disease as negative. A highly specific model means that there are few false positive results, thus avoiding unnecessary further diagnostic procedures. It is desirable to have a test that is both highly sensitive and highly specific, but unfortunately, this is rarely feasible.

The accuracy is a good metric only when the dataset is *balanced*, meaning the number of positive outcomes is roughly the same as the number of negative outcomes. In cases of *imbalanced* datasets, the area under the receiver operating characteristic curve (AUC/ROC), measured using the trapezoidal rule, is a more accurate metric. The ROC curve is a popular graphic for simultaneously displaying the two types of errors for all possible thresholds[29]. A threshold is the percentage after which an outcome is considered positive. If our threshold is set to 0.5 (50%), and our model classifies a person as being hypertensive with a

probability of 55%, then the outcome is positive. The name "ROC" is historic, and comes from communications theory; an acronym for receiver operating characteristics. The overall performance of a classifier, summarized over all possible thresholds, is given by the area under the (ROC) curve (AUC). The closest to 1 is the AUC the better the classifier. We expect a classifier that performs no better than a coin flip to have an AUC of 0.5. For an example of an ROC see Fig. 7.2.

3.1.3 Tools

The programming language `Python`[30] and the `Scikit-learn`[31] library was used for all ML algorithms and numerical calculations. Custom `Python` code produced the additional ECG waveform measurements. We used the `scipy.stats.spearmanr` function for measuring correlation. All figures and plots, unless otherwise noted, were created directly from the data, using the plotting library `matplotlib`[32]. For implementing the Random Forest classifier (introduced below) we used `RandomForest` classifier. All data analysis was done with the `Pandas`[33] library.

3.1.4 Random Forest classification

A RF is a method of machine learning, an ensemble of Decision Trees[34]. Decision trees are formal structures that have been successfully applied to making decisions in medicine; they resemble decision processes, a method familiar to many doctors. Each decision tree in a RF (see Fig. 3.7) performs a series of binary decisions (splits) by selecting a subgroup of the input features (such as age, body mass index, and body surface area), effectively trying

out different feature order and feature combinations.

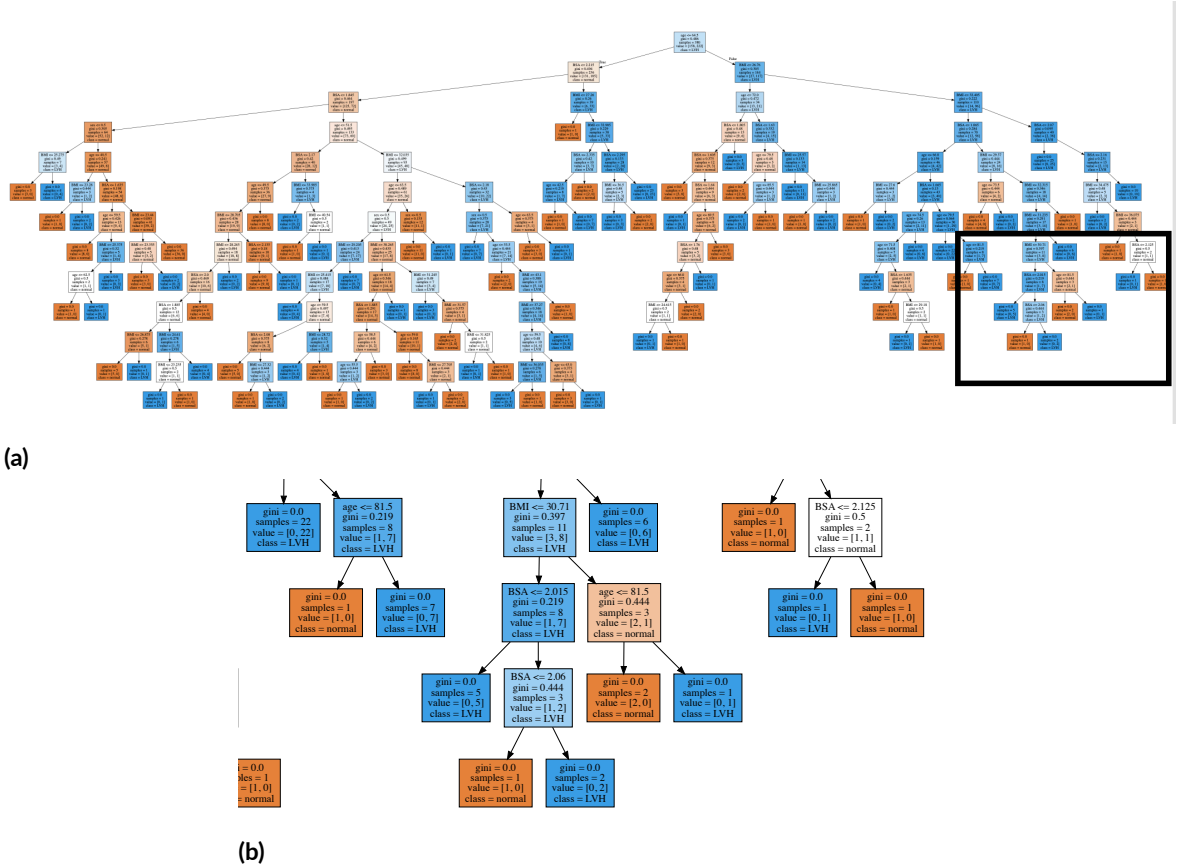

Fig. 3.7: (a) Depiction of a single decision tree out of many in a Random Forest. (b) The area inside the black box enlarged for visibility. The end nodes are called leaf nodes and their color denotes the class they represent; orange for LVH and blue for no-LVH. LVH: left ventricular hypertrophy, see Chapter 5.

RFs build a large collection of de-correlated trees, and then average their votes for the predicted class [29]. RFs are good predictors even with smaller datasets due to a technique called *bagging* (*b*ootstrap *agg*regat*ing*). Bootstrap creates overlapping, randomly selected, distinct data sets by repeatedly sampling observations from the original data set with replacement. Let's assume our original dataset is $\mathbf{x} = \{x_1, x_2, x_3, x_4\}$, one bootstraped dataset

would be $\mathbf{b}_1 = \{x_1, x_1, x_3, x_2\}$, another one would be $\mathbf{b}_2 = \{x_1, x_3, x_3, x_4\}$, etc. Bagging trains multiple trees on bootstrapped subsets of the data, and makes the final decision based on the majority vote of the different trees.

RFs use what is left out in a bootstrapped subset as a validation dataset. This is called *Out-Of-Bag* (OOB) estimate[35] of the training error. This estimate shows the ability of the model to predict on unobserved data. The OOB procedure can be divided in 4 steps: (a) entries left out from the bootstrapped subsets are considered part of a "test set"; (b) the model is evaluated on that test set; (c) the classification error is calculated; (d) the two previous steps are repeated until the error is minimized; The model optimizes its hyperparameters by minimizing this built-in OOB error estimate, which, as it turns out, is almost identical to that obtained by the usual k-fold cross-validation [29]. Therefore this technique enables RFs to be trained and cross-validated in one pass. A RF is also capable of handling non-linear interactions as well as cross-correlations among features.

Ensemble methods such as RFs perform better on tabular data that NNs. For a benchmark comparison of NN vs RF erformance on tabular data see.[36]

3.1.5 Feature importance via Shapley Additive explanations

Explaining predictions from tree models is always desired and is particularly important in medical applications, where the patterns uncovered by a model are often more important than the model's prediction performance [37]. The tree ensemble method in the *scikit-learn* package allows for the computing of measures of feature importance. These measures aspire to provide insight into which features drive the model's prediction. Mean Decrease

in Impurity (MDI), an approach popular among medical researchers, calculates each feature importance as the sum over the number of splits (across all trees). It was shown that the impurity-based feature importance can inflate the role of numerical features and bias the contribution of categorical, low cardinality ones [38]. Furthermore, these significances are computed on training set statistics and therefore do not reflect the usefulness of the feature in predictions that generalize to the test set. A better method is Permutation Importance which randomly shuffles a feature and calculates the error after running the model; if the error increased, then that feature is deemed important. We go one step further and calculate a recent feature importance metric called Shapley Additive explanations (SHAP), a game theoretic approach to explain the output of any machine learning model[39,37]. SHAP connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions. Visualizing feature importance using SHAP values is thought to be more accurate for global and local feature importance (importance calculated on *each* feature instead of all of them). SHAP values have already been used in medical papers [40].

3.1.6 t-distributed Stochastic Neighbor Embedding (t-SNE)

Feature visualization through dimensionality reduction using t-distributed Stochastic Neighbor Embedding (t-SNE)[41] is a popular method in the medical community; a variation of Stochastic Neighbor Embedding, t-SNE maps points from high-dimensional space to low-dimensional space by minimizing the difference in all pairwise similarities between points in high- and low-dimensional spaces. Embedding high dimensional data

into 2 dimensions allows us to visualize them in a way that gives useful insights on what differentiates study participants with a condition from those without. We performed visualization as part of our feature selection, before running any models, to inspect the features that seemed to best discriminate hypertensive from normotensive participants. The axes of the low-dimensional space are given in arbitrary units. The algorithm proceeds as follows: first, the pairwise distance matrix is calculated in high-dimensional space; next, the distance matrix is transformed to a similarity matrix using a varying Gaussian kernel, so that the similarity between points $\mathbf{X}_i$ and $\mathbf{X}_j$ represents the joint probability that $\mathbf{X}_i$ will choose $\mathbf{X}_j$ as its neighbor or vice versa (based on their Euclidean distance and local density); then, a random low-dimensional mapping is rendered and pairwise similarities are computed for points in the low-dimensional map. However, the low-dimensional similarities are computed using Student's t-distribution rather than a Gaussian distribution. Finally, gradient descent is used to minimize the Kullback-Leibler divergence between the two probability distributions, leading to the final low-dimensional map.

3.2 Role of artificial intelligence in cardiovascular medicine

ML has already been shown to outperform human experts in fields as diverse as speech recognition,[42] image classification,[43] and game playing.[44] ML is also gaining ground in medicine, performing comparably with physicians at tasks such as interpreting medical images.[45] ML is reducing the cost of development and time consumption in drug design.[46] Digital health technologies, such as AI-enabled tools and remote monitoring devices, have the potential to allow health systems to provide better, more accessible care while easing

the overburdening of clinicians. The use of telemedicine and AI may be able to address the global shortage of health care providers;[47] it has lately sparked interest in areas such as the care of retinopathy of premature infants, addressing the relative shortage of ophthalmologists.[48]

Cardiology is no exception.[49] ML is impacting both clinical diagnostics[50,51] and research,[52,53] and will certainly change the way we practice cardiovascular medicine, providing new tools for leveraging data and making decisions.[54,55] Several studies apply ML for the prediction of heart failure.[56] Cardiovascular imaging is an area where ML may reduce cost and improve value by facilitating image acquisition, measurement, and subsequent evaluation.[57,58] In a recent paper, the authors implement a deep learning–based tool to estimate right ventricular ejection fraction (RVEF) from 2D echocardiographic videos.[59] Cardiovascular image segmentation is a another field in which AI-based methods have recently shown substantial improvements in performance[60] – segmentation is the process of content extraction that takes as an input a medical image, volume, or sequence of images or volumes, to produce associated shapes, for example, defining the chambers of the heart or heart vasculature.

3.3 Applications of machine learning in the electrocardiogram

Although cardiovascular imaging with various modalities such as calcium scoring computed tomography, coronary CT angiography, and cardiac MRI, have advanced rapidly over recent years, the ECG remains the primary evaluation method for a person suspected of having a cardiovascular pathology.[61] Applications of ML on the ECG are evolving

rapidly with tremendous future implications on cardiovascular medicine.[62,63,64,65] Analysis of the ECG with ML is already helping clinicians detect cardiovascular diseases such as coronary artery disease.[66]

Modalities such as ECGs and echocardiogram images, defy traditional statistical methods and require the deployment of ML. ECG signals and patterns largely unrecognizable to human eye interpretation can be detected by machine learning algorithms. In a study, an ML model detected arrhythmias equally well as cardiologists.[67]. In another, a model trained on sinus rhythm ECGs identified patients with paroxysmal atrial fibrillation (AF).[68] New combinations of ECG parameters were used to predict Echo-LVH.[69] Left ventricular diastolic function was assessed with ML and ECG Features.[70]

3.3.1 Wearables

Cardiovascular medicine can benefit from smart analysis of massive amounts of data produced by wearable sensors that have the capacity to incorporate health data with lifestyle parameters. Assuming privacy and security issues are addressed, ambulatory devices could become a source for continuous monitoring and early warning, especially for vulnerable populations. The Apple Watch has moved beyond photoplethysmographic pulse detection to generate an ECG that is similar to a single-lead (or Lead I) ECG,[71] and the relevant app can detect atrial fibrillation (AF).[72] Chest patches can record single-lead ECGs.[73]

3.3.2 The need for interpretability

Medicine is a high-stakes field where justifying decisions is important. Given the complexity of ML models, especially deep learning models, it is not always possible to explain their predictions, which is a barrier to the adoption of ML in medicine. Better performance, though, of some of these models compared to traditional methods, necessitates their use, and makes the quest for interpretability all the more necessary. Saliency maps[74], a visualization technique that identifies pixels whose relevance to the network's prediction is high, and Grad-CAM[75], which highlights the regions of the image that the model considered most relevant for predicting the given class, are steps towards interpretability. Other such techniques include LIME,[76] and the aforementioned SHAP.

Data consist of numbers, of course. But these numbers are fed into the computer, not produced by it. These are numbers to be treated with considerable respect, never to be tampered with, nor subjected to a numerical process whose character you do not completely understand.

Bill Press, computational biologist and computer scientist, from his book[77]

4

Research methods

Data acquisition for this thesis lasted three years, from November 2019 to November 2022 , and involved two Cardiology centers, the Heraklion University General Hospital, and the Chania St George General Hospital, both located in Crete, Greece. The population was, by assumption, local to the island of Crete, although we did not explicitly collect such information. This is the only AI ECG-based study that we are aware of, that uses

data from the population of the island of Crete. While there is value in using standardized and highly tested databases to train and evaluate an ML algorithm, such as the ones in Physionet,[78] there are also caveats; the population from which the data were collected may not be the right one for a novel hypothesis. Or, the necessary features might be missing, as was the case with our research; data on whether human subjects were hypertensive did not exist in public databases. By gathering our own data, we addressed these issues. Also, participants in our study were carefully selected to eliminate clinical parameters that could mislead the model. This way, we improve the quality of input data and avoid pitfalls due to the large diversity of pathological conditions that formed the basis for the training data. Last, our data were collected prospectively across real-world clinical settings and by many operators.

The study was conducted in accordance with the Declaration of Helsinki, the protocol was approved by the Hospital Ethics Committee, and patients gave written informed consent to their participation in the study.

4.1 Sample/cohort selection and exclusion criteria

Participants to all three experimental studies, had no indications of CVD, underwent a physical examination, an ECG recording, a routine echocardiography examination, and laboratory tests, before inclusion. Height and weight were measured during the same visit as the ECG acquisition, and the individuals were classified using the World Health Organization (WHO) classification of body mass index (BMI).

Hypertensive patients were recruited from the outpatient clinics of the two centers. Nor-

motensive healthy individuals were referred either for the investigation of atypical chest pain or for the modification of risk factors for cardiovascular disease such as hyperlipidemia. The diagnosis of hypertension was based on the recommendations of the European Society of Hypertension/European Society of Cardiology[79]; essential hypertension was defined as office BP $>$ 140/90 mmHg, measured in three consecutive visits, or in one visit when the diagnosis was confirmed by out of office measurements. In addition, out of office measurements were performed to exclude two medical conditions: a) what is known as white-coat hypertension, referred to as an untreated condition in which office blood pressure measurements are elevated, but 24-h ambulatory blood pressure measurement, home blood pressure measurement, or both show no abnormalities; and b) what is known as masked hypertension, characterized by a normal office blood pressure and hypertensive levels on 24-h ABPM .

Emphasizing the importance of data quality over quantity, we had every ECG carefully reviewed by cardiologists to exclude subjects with certain conditions that could confuse the model. Subjects with any of the following characteristics were excluded: tachy- or bradyarrhythmia; permanent atrial fibrillation, RBBB, LBBB or other conduction abnormalities on ECG, coronary artery disease; moderate or severe valvular heart disease, cardiomyopathy, cerebrovascular, liver or renal disease; history of acute coronary syndrome or myocarditis; ejection fraction $<$ 55%; history of drug or alcohol abuse; any chronic inflammatory or other infectious disease during the last 6 months; thyroid gland disease; pregnant or lactating women. Vascular or neoplastic conditions were also ruled out vascular or neoplastic conditions were ruled out by a careful examination of the history and routine laboratory

tests. Functional tests for myocardial ischemia, coronary computed tomography angiography or invasive coronary angiography were performed according to physician's judgement, in order to exclude coronary artery disease. The study was conducted in accordance with the Declaration of Helsinki, the protocol was approved by the Hospital Ethics Committee, and patients gave written informed consent to their participation in the study.

4.2 Electrocardiogram acquisition

A 12-lead ECG of 10 seconds duration in resting position was performed on each subject using a digital 6-Channel machine (Biocare iE 6, Shenzhen, P.R. China) and was stored in a digital file with sampling rate 1000 Hz using eXtensive Markup Language format (XML). We used custom code to automatically read the XML and extract the information we needed. We used all 12 leads for the hypertrophy study and the 12-lead hypertension experiment, and in order to simulate devices that record only a single lead for our last study, we isolated the tracings of lead I from our digital files for further processing.

Automated measurements of wave/complex duration and wave amplitude, calculated by Biocare's software, were then extracted from the digital files. These measurements were based on representative complexes, corresponding to individual heart beats, which, according to the manufacturer, were calculated as follows: the multiple cardiac cycles of 10-sec duration that are recorded for each lead, are overlaid digitally to form a single 1-sec duration representative beat for each lead; this reduces the effects of beat-to-beat variation in the waveforms and removes potential artifacts. Some features extracted from the ECG, such as heights of waveforms, are lead-specific, while others such as time intervals, are measured

in the averaged waveforms of all leads overlaid on each other. We moved all representative beats vertically until the ST segment was on the zero horizontal line, and that became our *isoelectric* line, or baseline. The process is schematically shown in Fig. 4.1.

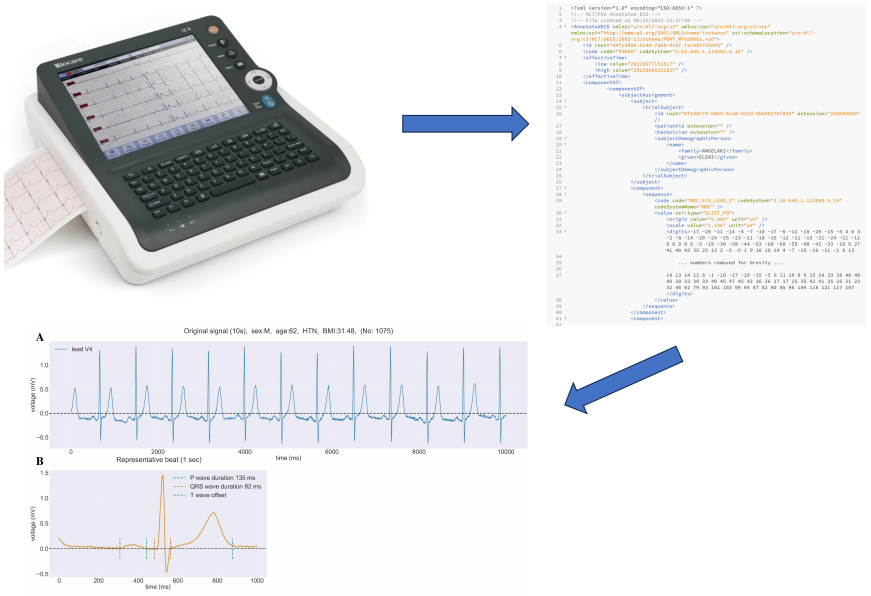

Fig. 4.1: The data collection process at a glance. The electrocardiograph records an ECG from the patient with the help of a healthcare professional. The machine then saves an XML file, out of which we extract a 10 sec ECG signal with some measurements, and then a 1 sec representative signal.

4.3 Echocardiography

Echocardiography provides detailed data on cardiac structure, including the size and shape of cardiac chambers, as well as the morphology and function of cardiac valves. It is often the second most common cardiac imaging modality performed, after the ECG. Echocardiography is based on the standard principles of ultrasound imaging in which high-frequency sound waves emitted from piezoelectric crystals in a transducer, traverse through internal

body structures, interact with tissues, and after reflecting back to the transducer, are processed by computer to generate an image and detailed measurements. The time required for the waves to be reflected back is translated in scan lines that depict both location (depth of reflection) and amplitude (intensity of reflection)

A full echocardiographic study was performed in all subjects using a Vivid 7 (General Electric, Horten, Norway) ultrasound device with a 1.5 - 3.6 MHz wide angle phased-array transducer (M4S) according to the recommendations of the European Association of Cardiovascular Imaging and American Society of Echocardiography[80] by two experienced echocardiographists blindly. All examinations were performed the same day with the ECG and digitally stored for off-line analysis as the mean of three cardiac cycles.

4.4 Feature engineering and selection

Choosing good predictive features/markers from data is arguably the most challenging part in an ML project even more important than the choice of the specific learning algorithm. Finding the right markers to extract from the raw data, those that could reveal psychological signals of interest and better predict target outcomes, is especially hard with rich sources such as ECGs. In addition to known metrics, i.e., what physicians look for in an ECG, we constructed our own features with the suggestions of the cardiologists in the team. Feature selection was part of pre-processing, before any models were trained, so as not to bias the model.

Below is a list of the extracted and calculated features, in categories. In subsequent Chapters we describe how they were selectively used for each experiment.

4.4.1 Anthropometric features

(1) `age`, years.

(2) `height`, cm.

(3) `weight`, kgr.

(4) `sex`, binary, 0: male, 1: female.

(5) `BMI`, kg/m^2. Body mass index. Calculated by taking a person's weight, divided by their height squared, or BMI = weight (in kg)/ height^2 (in m^2). BMI is a statistical index to provide an estimate of body fat in males and females of any age[81,82]. The National Institute of Health (NIH) now uses BMI to define a person as underweight, normal weight, overweight, or obese instead of traditional height vs. weight charts. However, individual variations do exist, and BMI is insufficient as the sole means of classifying a person as obese or malnourished. In certain populations, like elite athletes and body-builders, an elevated BMI does not directly correlate to their health status due to their increased muscle mass and weight falsely increasing their BMI. Weight and height were measured during the same visit as the ECG acquisition.

(6) `BMI_bins`, categorical variable assigning a person to one of four categories according to their BMI: underweight (BMI $<$ 18.5 kg/m^2, category 0), normal weight (18.5 $\leq$ BMI $\geq$ 24.9 kg/m^2, category 1), overweight (24.9 $<$ BMI $\geq$ 29.9 kg/m^2, category 2), and obesity (BMI $>$ 29.9 kg/m^2, category 3). We did not have underweight people in our study, so, effectively we only have categories 1, 2, and 3.

(7) `BSA`, in m^2, body surface area, calculated with Mosteller's simplified formula[83]:

$$\text{BSA}\ (m^2) = \sqrt{\frac{\text{height (cm)} \cdot \text{weight (kgr)}}{3600\ (\text{cm kgr/m}^4)}}$$

(8) `BF`, %. Percentage of body fat. Defined by the formulae: [adult males] BF(%) = 1.20 · BMI + 0.23 · age - 10.8 - 5.4, and [adult females] BF(%) = 1.20 · BMI + 0.23 · age - 5.4.[84]

4.4.2 ECG-derived features

Electrocardiographic terms are consistent with AMA Manual of Style (2019, 11th edition). We chose to include ECG measurements adjusted for BMI based on studies showing that larger body mass decreases the amplitude of the R and S waves in specific leads due to the electrical currents traveling longer distances.[85] For reference, an ECG tracing with its major components is shown in Fig. 4.2.

Based on time markers on the 1 sec representative beat derived by the electrocardiograph, ECG-based features were calculated (see Fig. A.1 for an example of an XML excerpt). Those are, (a) the `P_onset`, time of start of the P wave, (b) the `P_offset`, end of the P wave, (c) the `QRS_onset`, start of the QRS complex, and (d) `QRS_offset`, end of the QRS complex. And last, (e) `T_offset`, the end of the T wave. All time points are measured from the beginning of the ECG waveform, as shown in Fig. 4.3.

Durations

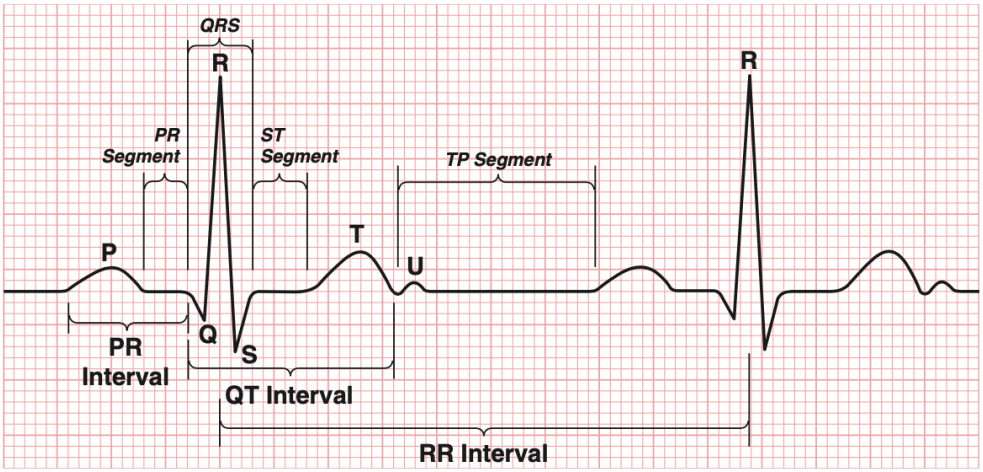

Fig. 4.2: Basic components of the ECG signal, including the P, QRS, ST, T, and U waveforms, the RR, PR, QRS, and QT intervals, and the PR, ST, TP segments. Image from[9]. Used with permission from Elsevier.

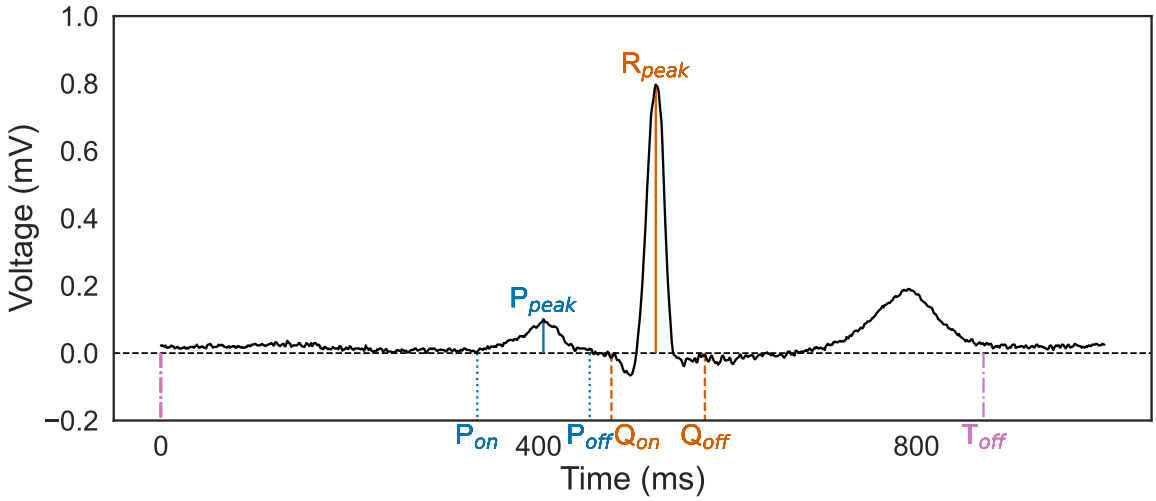

Fig. 4.3: Time markers derived by the machine. P_{on} is the onset of the P wave, P_{off} denotes the offset, Q_{on} is the onset of the QRS complex, and Q_{off}, its offset. For the T interval only the offset is derived by the machine.

(1) `hr`, beats per minute. Heart rate. The average RR interval, or the distance between two consecutive R peaks (see Fig. 4.2 in Chapter 2). The measurement was performed by the machine and included in the XML file, as shown in Fig. A.2.

(2) `P_duration`, ms. Time from onset of the P wave to its offset.

(3) `PQ_duration`, ms. Duration of P wave. Measured from the onset of the P wave to the onset of the QRS. It is also referred to as the PR interval as a Q wave is not always present.

(4) `QRS_duration`, ms. QRS interval duration. Measured from the beginning to the end of the QRS complex.

(5) `QT_interval`, ms. The QT interval comprises the QRS complex, the ST segment, and the T wave. It is calculated from the start of the QRS complex until the end of the T wave.

(6) $\texttt{QT}_\texttt{c}$, ms. QT corrected. One difficultly of QT interpretation is that the QT interval gets shorter as the heart rate increases. This problem can be solved by correcting the QT time for heart rate using the Bazett formula:

$$\mathrm{QT_c} = \frac{\mathrm{QT}}{\sqrt{\mathrm{RR\ interval\ (sec)}}}$$

(7) `ST_duration`, ms. Duration of the ST segment. Time distance from the offset of the QRS complex until the onset of the T wave. Because of the complexity of the ECG in the area where the T wave starts, the machine did not make measurements for this variable. We used `Python` code to infer the start of the T wave from the slope of the

rising wave. As seen in Fig. 4.4, the start of the T wave was marked where the slope hits the baseline horizontal line.

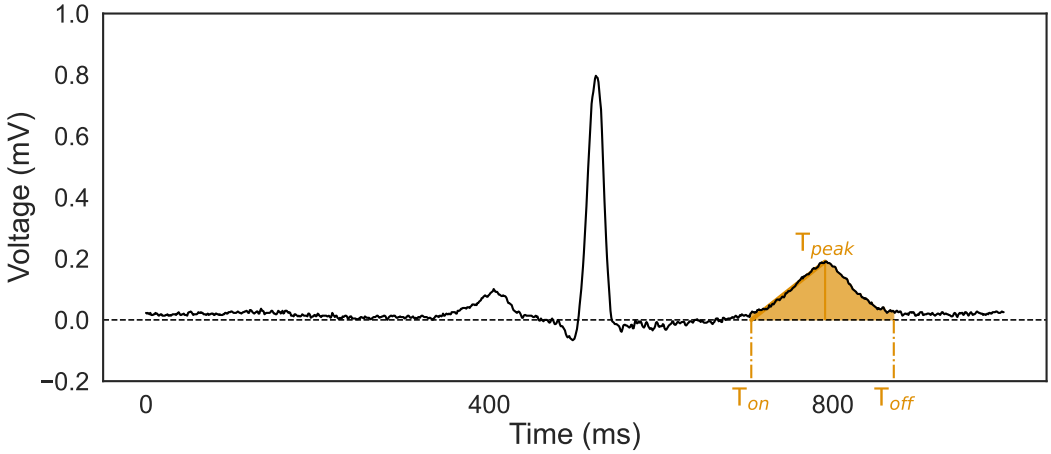

Fig. 4.4: T wave markers, T_{on} is T_onset, T_{off} denotes the offset. For the T interval only the offset is derived by the machine, the onset is calculated by our software. The shaded area is the area under the T wave, namely the `TOI`.

(8) `ID_I`, ms. Intrinsicoid deflection, or R-wave peak time. Represents the early phase of ventricular depolarization and is defined as the time period from the onset of the QRS complex to the peak of the R wave. The lead number comes after the dash. In previous studies, delayed intrinsicoid deflection (DID) $\leq$ 50 ms in lateral precordial leads V5 or V6 has been associated with left ventricular hypertrophy (LVH) and is included in the Romhilt-Estes criteria for ECG diagnosis of LVH. The ID can be seen in Fig. 4.5.

(9) `T_duration`, ms. T wave duration. Time from onset of the T wave to its offset (see Fig. 4.4).

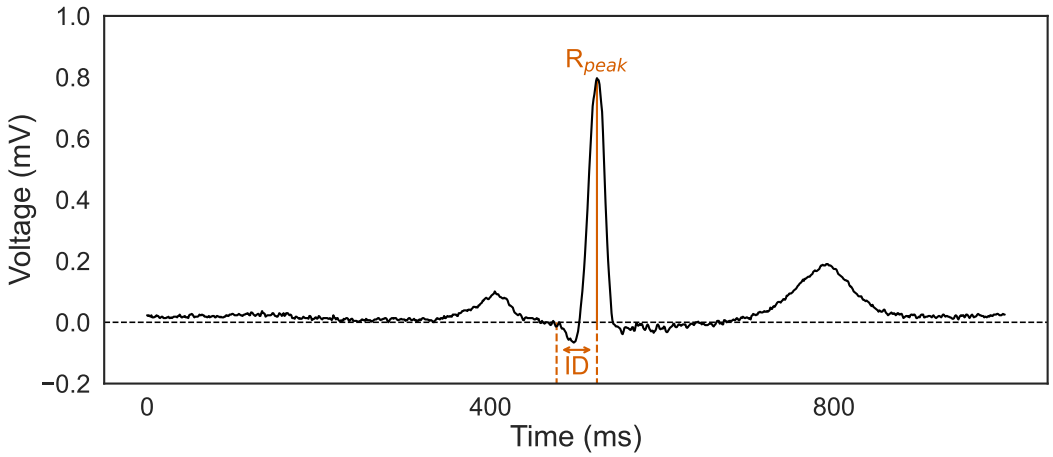

Fig. 4.5: Arrow indicates the intrinsicoid deflection (ID), the time duration between the onset of the QRS complex and the peak of the R wave.

Amplitudes

1. `R_I`, mV. R wave voltage amplitude in lead I (or other lead indicated by the roman numeral). Height of the peak of the R wave (see Fig. 4.3).

2. `T_V2`, mV. T wave voltage amplitude in lead V_2 (or other lead indicated by the roman numeral). Height of the peak of the T wave (see Fig. 4.4).

3. `P_V5`, mV. P wave voltage amplitude (see P peak in Fig. 4.4).

4. `S_V5`, mV. S wave voltage amplitude, defined as the height of the S wave, in absolute value since S is usually below the isoelectric.

5. `R_max`, mV. Maximum voltage R height. The largest R wave voltage amplitude of all the limb leads.

Axes

Axes values are produced by the electrocardiograph's software, while the feature `QRS_T_angle` is calculated by our software, as explained below.

1. `P_axis_fr`, degrees. Angle of the electrical axis of the P wave in frontal plane.
2. `QRS_axis_fr`, degrees. Angle of the electrical axis of the QRS complex in frontal plane. The QRS axis signifies the sum of all individual vectors generated by the depolarization waves of ventricular myocytes, and is considered an important ECG parameter that reflects the diastolic function of the left ventricle.[86] For a depiction of this vector see Fig. 4.6.

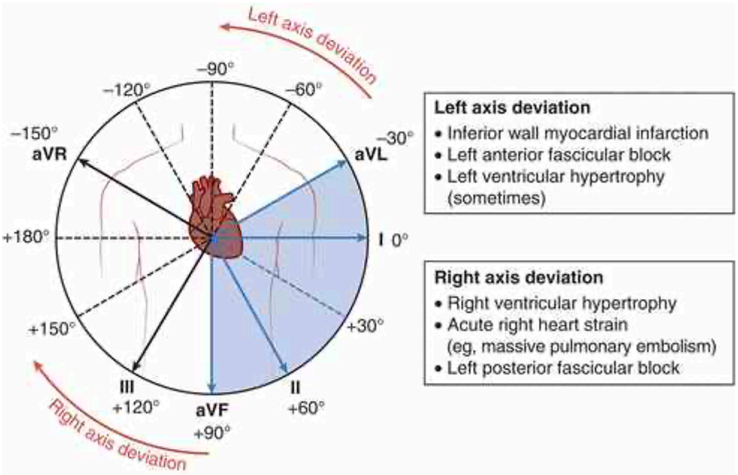

Fig. 4.6: Position of the mean QRS vector which defines the feature `QRS_axis_fr`. Normal values are considered those from -30° to +90° in the frontal plane. Image from[12] used with permission from Wolters Kluwer.

3. `T_axis_fr`, degrees. Angle of the electrical axis of the T wave in frontal plane.

4. QRS_T_angle, degrees. Planar frontal QRS-T angle, defined as the angle between the frontal QRS_axis_ fr and T_axis_fr vectors; it is the angle between the directions of ventricular depolarization and repolarization, therefore, a wide QRS-T angle reflects either structural abnormalities affecting the depolarization or regional pathophysiological changes in ionic channels altering the sequence of repolarization.[87] It can be calculated from a standard 12-lead ECG as the absolute value of the difference between the frontal plane QRS axis and T axis. If such a difference exceeds 180 degrees, then frontal QRS-T angle is calculated as 360° minus the absolute value of the difference between the frontal plane QRS axis and T axis.[88]

Areas

1. ROI_aVF, mVms, calculated as 1/2 of the area/integral below the R wave in the aV_L lead. The reason we take 1/2 of the area and not whole is purely by convention. The above definition applies to other leads, e.g., ROI_V2, would be the area/integral below the R wave in the V_2 lead.

2. TOI_I, in mVms, the whole area under the T wave, as seen in Fig. 4.4.

We removed features that exhibited correlation $> 85\%$ in Spearman's rank correlation test; highly correlated features contribute the same amount of information and including both of them in a RF model might not affect performance, but it will divide, thus lessen, each feature's predictive significance. We choose the Spearman test because of the possibility of non-linear relationships among the data and then calculated the rank-sum statistic

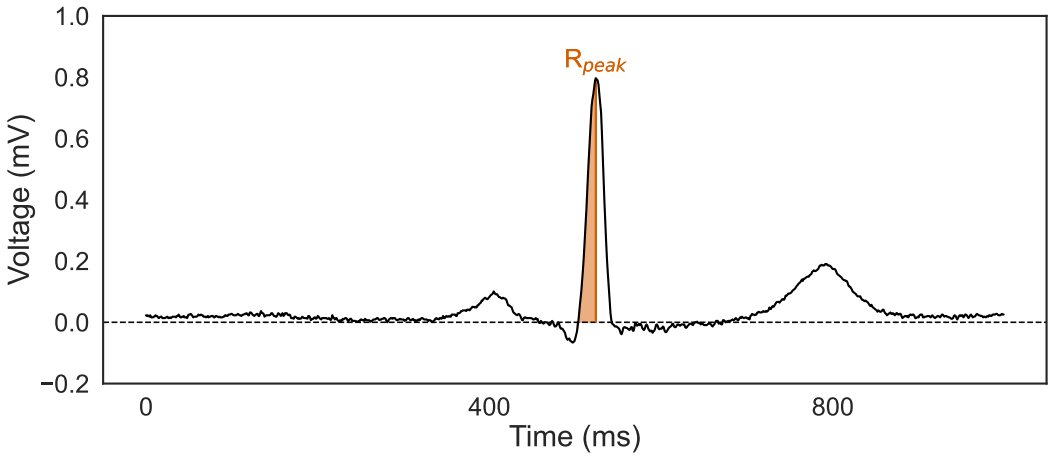

Fig. 4.7: Shaded area shows is variable `ROI_I`, defined as 1/2 of the area under the R curve, by convention.

and ranked the features according to their p-value. This feature selection is part of preprocessing, before the model was trained.

4.5 Dataset preparation

Unless otherwise indicated, the dataset for each study was split into a train set (80%), used directly to learn the parameters of the model, and a test set (20%), consisting of data the model had not seen during training and was used exclusively for performance evaluation of the final model. Reported metrics are on the test set. Stratification for sex and history of hypertension, during the partition, ensured the two sets contained the same proportions of these two features. For validation while training the Random Forest (RF) we used the model's internal out-of-bag (OOB) set. Feature importance graphs are also on the test set, as, using the train set inflates the importance of some features which might not be as important in predicting the outcome. We also made sure that data from the same patient

was not included in both the train and test set.

5

Detection of abnormal left ventricular geometry

The left ventricle (LV) is the largest and strongest chamber in the heart, strong enough to generate pressure and displace a volume of blood through the aortic valve and into the systemic circulation. LV performance is determined by the preload (venous return, end-diastolic volume), myocardial contractility (the force generated at any given end-diastolic volume), and afterload (aortic impedance and wall stress)[10]. The LV responds

to systemic hemodynamic and ventricular load under declining cardiac function, by triggering a hypertrophic response, leading to an increase in myocyte size, left ventricular wall thickness and mass. When cardiac enlargement occurs, the heart muscle cells become either longer, deemed eccentric hypertrophy, or wider, in concentric hypertrophy. Although this remodeling response initially restores wall stress, it ultimately leads to a prognosis of increased risk for major cardiovascular complications.[89] Cardiac remodeling is considered an important aspect of CVD progression and is therefore emerging as a significant therapeutic target.[2,3,4,5]

The detection of hypertension-mediated organ damage, such as abnormal LVG, is a useful approach toward risk stratification of a hypertensive population.[79] The evaluation of cardiac structure and function is encouraged since it might influence treatment decisions.[79] Transthoracic echocardiography has received a strong indication for the initial evaluation of suspected hypertensive heart disease. Abnormal LVG is the early marker of LV remodeling that precedes hypertrophy and is frequently associated with LV diastolic dysfunction.[80]

Left ventricular hypertrophy (LVH), is usually defined as an increase in LV mass, although such an increase is not the only attribute consistent with LVH. More broadly we can define hypertrophy as an increase in size of the tissue due to an increase in size of the cells in that tissue without increasing their number, thus differentiating hypertrophy from hyperplasia. Verma et al.[90] define 3 patterns of LV remodeling based on measurements of the LV mass index (LVMi) and relative wall thickness (RWT): (a) concentric remodeling (CR) defined as normal LVMi and increased RWT; (b) eccentric hypertrophy (EH) having

increased LVMi and normal RWT; and (c) concentric hypertrophy (CH) characterized by both increased LVMi and RWT. For completeness, (d) normal geometry (NG) is defined as both normal LVMi and RWT. CH and EH are collectively denoted as LVH. LV remodeling is defined as either CR or EH or CH. Values are slightly different for males and females, as indicated in the graphic depiction of these categories shown in Fig. 5.1.

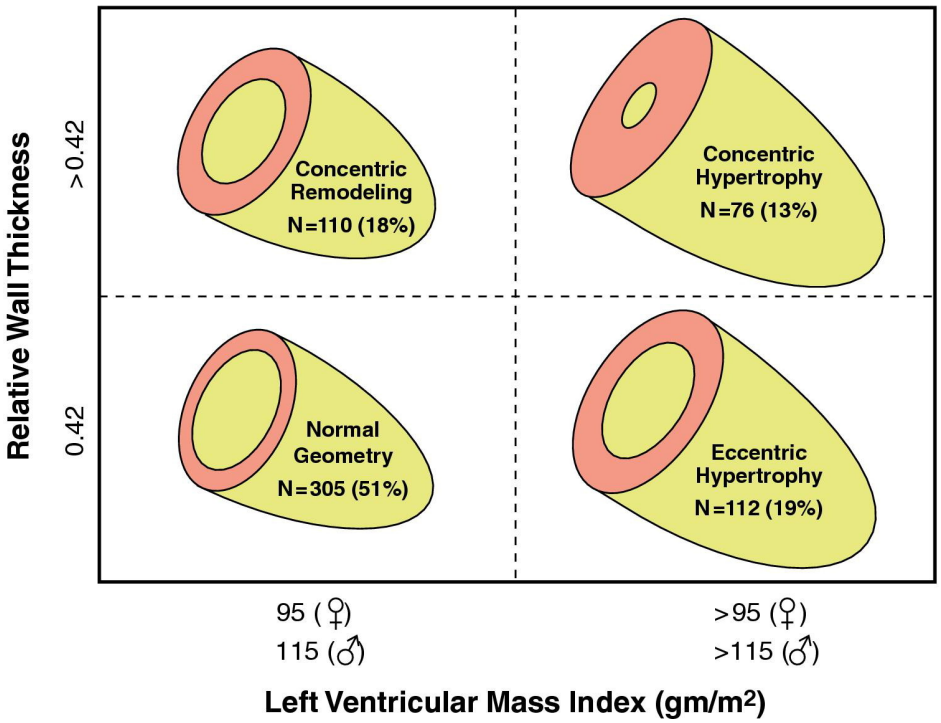

Fig. 5.1: Patterns of cardiac remodeling based on sex and measurements of the left ventricular mass index and relative wall thickness. N is the number of samples/patients in each category. Sex is defined as binary (male/female) with ♀ being the symbol for females. Image from [90].

The three abnormal patterns were associated with a higher risk of subsequent cardiovascular events, carrying progressively worse prognosis. In hypertensive patients, LV remodeling„ is predictive of the incidence of cardiovascular events.[79] CR not only precedes

LVH, but also most other cardiac dysfunctions while it progresses asymptomatically. The early detection of abnormal LVG can result in early detection of subclinical hypertension-mediated organ damage and may help clinical decision on follow up.

An ECG is the primary modality for detecting LVH or other abnormalities in hypertensive individuals. An effect of advanced LVH on the ECG is a prominent negative S wave on the right chest leads and tall R waves on the left chest leads. However, the ECG is not a sensitive method of detecting left ventricular hypertrophy, and, with existing knowledge, cannot detect signs of LVG at early stages, before LVH is present. Notably, although ECG criteria demonstrate relatively high specificity, the sensitivity for the detection of LVH is low,[91] approximately 30%, and in some studies it is as low as 6.9%.[92] The echocardiogram is always suggested as an additional diagnostic evaluation for this reason.

We propose a digital interpretation of the ECG with ML methods, to extract information that is not easily and directly detected by the human eye, especially within a busy clinical setting, that would expand the diagnostic capabilities of ECG in detecting patients with LV remodeling and LVH, and refer them for further echocardiographic evaluation. This study was designed to test the hypothesis that a 12 lead ECG, being a routine and inexpensive screening procedure, can, through ML methods, provide further accuracy in detecting abnormal LVG, even at the early stages before the onset of LVH, in a population without established CVD. We also seek to understand which features contribute to the ML model's decisions by calculating global feature importance and feature interactions.

5.1 Feature engineering and selection

Besides the features described in Chapter 4, additional echocardiographic features and criteria for LVH used in daily practice were calculated:

(1) `LVM`, gr. Left ventricular mass. Equal to the internal dimension measured at end-diastole, according to the American Society of Echocardiography guidelines.[93]

(2) `LVMi`, gr·m^2. Left ventricular mass indexed for BSA. Some say that this tends to underestimate the prevalence of LV hypertrophy in overweight patients.

(3) `RWT`. Relative wall thickness. Defined as the ratio of twice left ventricular diastolic posterior wall thickness (PWT) to left ventricular end-diastolic diameter (LVEDD), RWT is a measure of LV geometry and has been associated with LVH and myocardial ischemia. RWT was derived by $2\times$ PWT/LVIDd (LVIDd–left ventricular internal diastolic dimension, PWT–posterior wall thickness). Normal RWT was defined as $<$ 0.43. Septal and posterior wall thicknesses were measured at end-diastole, according to the American Society of Echocardiography guidelines.[93,79]

(4) `hyper`, binary. 1: existence, 0: absence of hypertension.

(5) `soko`, mV. Sokolow-Lyon (SL) voltage. Sum $SV_1 + RV_5 > 3.5$ mV, where SV_1 is the S wave voltage in V_1 lead, and RV_5 is the R wave voltage in the V_5 lead, as defined by European guidelines on hypertension and the detection of LVH.[79]

(6) `BMI_adj_soko`, mV. BMI adjusted SL index. Calculated by adding 0.4mV for being overweight, and 0.8mV for being obese, to the regular SL index, based on Hesham et.

al.[94]

$$\texttt{BMI\_adj\_soko} = \begin{cases} \text{SL} + 0.4, & \text{if overweight} \\ \text{SL} + 0.8, & \text{if obese.} \end{cases}$$

(7) `Cornell_sum`, or `RaVL+SV3`, mV. Cornell voltage index. Calculated as the sum of the R wave height for lead aVL, `R_aVL`, and the S wave height for V_3 in absolute value, `S_V3`.

(8) `Cornell_product`, or `RaVL_QRS`, ms·mV. Cornell product. The product of the Cornell index with the QRS complex duration.

(9) `BMI_adj_cornell`, mV·kgr/m^2. BMI adjusted Cornell product. Calculated by multiplying the continous value of `BMI` with the Cornell product `RaVL+SV3` based on Hesham et. al.[94]

In addition to the above standard LVH criteria, we have our own criterion:

(10) `BMI_soko`, or `BMI/SL`, in kgr/m^{2}mV, a custom-made criterion, defined as continuous BMI divided by the SL index. We introduce the feature because we hypothesize that body mass affects the amplitude of the R and S waves, as the electrical currents travel different distances before reaching the lead electrodes.

We performed feature selection to reduce the dimensionality of our space by eliminating irrelevant features, and correcting for high correlation among some of the features, as assessed by Pearson's correlation test. Keeping only one of two correlated features retains all the information while providing a clearer picture of the remaining feature's contribution

(see Chapter4); Pearson's coefficient > 0.90 was our threshold for removal. Our model was trained on 32 features shown in Table 5.1.

5.2 Results

After careful screening of 903 hypertensive and normotensive healthy individuals, we enrolled 528 consecutive subjects older than 30 years of age, with and without essential hypertension and no indications of CVD. Of the chosen subjects, 56 % were female and 71% were hypertensive. The mean age was 62.3±11.9 years for women, and 60.5±12.4 years for men. Based on `BMI`, 45.7 % of them were obese, 40.2 % were overweight, and 13.9 % were within normal range; there were no underweight individuals. `CR` was present in 37.2% of the individuals, `LVH` was present in 17 %, while 45.7 % of them had normal geometry. Figure 5.2 shows the box plots for four features and their distributions among the three categories, `NG`, `CR`, and `LVH`. Specifically, Figure 5.2D reveals a tendency of the product of the `BMI` with the SL criteria, (`BMI_soko`) to discriminate the `CR` class from the other two.

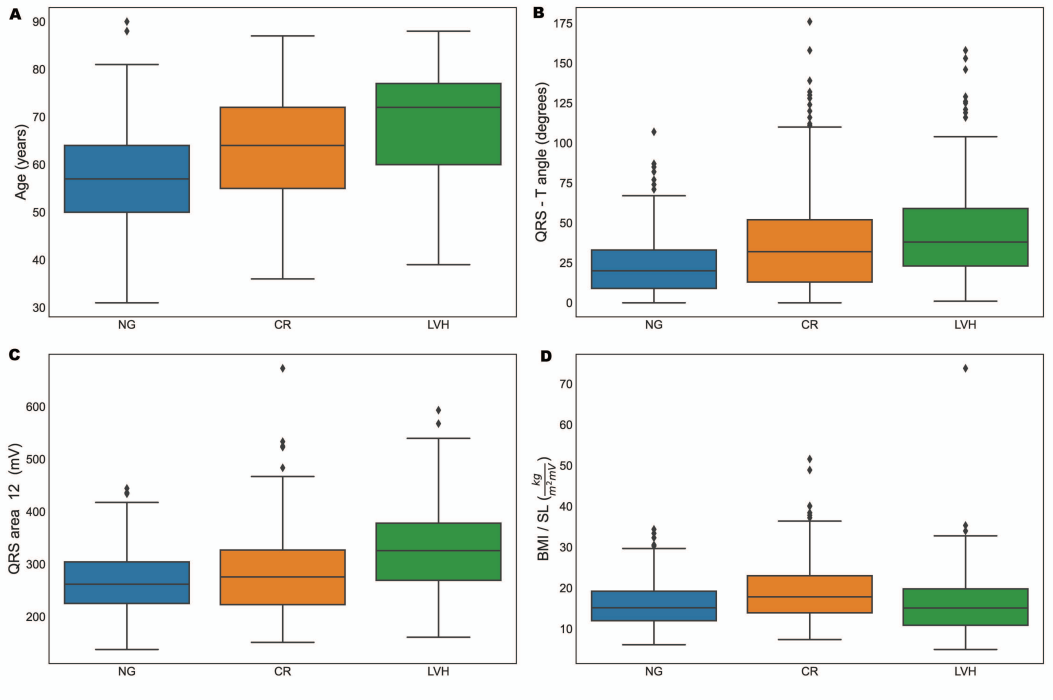

Fig. 5.2: (**A-D**) Box plots of feature distributions in patients for class NG, CR, and LVH. In (**D**) we notice that BMI/SL shows a small tendency to discriminate the CR class. The individual dots in the box plots depict outliers, patients with predictor values very different from other patients. BMI, body mass index; CR, concentric remodeling; LVH, left ventricular hypertrophy; NG, normal geometry; BMI/SL, BMI_soko.

Trained on a set of features, an ML classifier's goal is to assign each individual (observation) to one of various classes (response variable). We tried using just the anthropometric features as predictors and then adding the ECG-derived features. For evaluating the performance of our model we calculated accuracy, sensitivity, specificity, and AUC/ROC.

When trained using only clinical variables (sex, age, BMI_bins, BSA, hyper, and height), our model and got an accuracy of 79 % in the test set, with a sensitivity of 84 % and a specificity of 73 %. The addition of the ECG-derived features seen in table 5.1, improved our model's accuracy to 87 % (in the same test set), with a sensitivity of 97 % and a specificity of

75 % for the default threshold of 0.5. AUC/ROC was 0.91.

Concentrating on the LVH class, we trained the RF to classify NG+CR vs. textttLVH. We achieved an accuracy of 89, specificity 93, sensitivity 67, and AUC/ROC 0.89. Due to the imbalance in the dataset (ratio of NG+CR to LVH was 5/1) when divided into these categories, we performed *oversampling* for imbalance correction using RandomOverSampler.[95] The model results are summarized in Table 5.2.

We then visualized the global feature importance and local explanations for the binary classification using SHAP. An interesting finding is the effect of specific features on each individual subject separately, as well as the interaction effects between pairs of features. In Fig 5.3 we notice that hypertension has a strong positive effect on being classified as CR+LVH, while normotensive patients have a different risk for being classified as CR+LVH. Age plays an important role in the risk of being classified as CR+LVH with a cutoff around 65 years, and the risk is higher for men under 65, while over 65 the risk appears higher for women, as shown in Fig. 5.3 (C).

Feature	Description	Range	Mean or Mode
sex	0, M; 1,F	2	F
age	age, years	31.0 - 90.0	61.5
hyper	history of hypertension, 0, NT; 1, HTN	2	1
BMI_bins	BMI class, 0, normal; 1, overweight; 2, obese	3	3
height_cm	height, cm	137.0 - 194.0	165
BSA	BSA, m^2	1.3 - 2.8	1.93
P_dur	P wave duration, ms	0.0 - 154.0	113
QRS_dur	QRS interval duration, ms	68.0 - 138.0	92.7
QTc_dur	QT-interval corrected for heart rate, ms	368.0 - 510.0	422
ST_dur	ST segment duration, ms	0.0 - 277.0	112
S_V1	S wave amplitude in V_1, mV	0.0 - 1.9	0.732
soko	SL voltage, mV	0.5 - 4.5	1.93
R_max	Tallest R wave in limb leads (RE criteria), mV	0.3 - 1.9	0.903
BMI_soko	BMI divided by SL voltage, kgr/m^2 mV	6.0 - 73.8	17.4
ID_V5	Intrinsicoid deflection in V_5, ms	24.0 - 57.0	39.5
ROI_aVF	Area under R wave in aVF, ms mV	0.0 - 21.3	3.46
SOI_V1	Area under S wave in V_1, ms mV	0.0 - 62.2	17.3
QRSOI_V5	Area under QRS interval in V_5, ms mV	11.0 - 80.2	33.7
QRSOI_12	Total area in all leads (sum), ms mV	137.5 - 672.9	285
R_V2	R wave amplitude in V_2, mV	0.0 - 2.0	0.41
R_aVL_QRS	Cornell product, ms mV	3.6 - 243.8	52.6
S_V2	S wave amplitude in V_2, mV	0.1 - 2.3	0.784
S_V5	S wave amplitude in V_5, mV	0.0 - 1.3	0.381
QRS_T_angle	Planar Frontal QRS-T angle, degrees	0.0 - 176.0	33.3
QRS_axis_fr	QRS axis front, degrees, degrees	-77.0 - 97.0	15.6
T_axis_fr	T axis front, degrees	-59.0 - 258.0	39.8
T_aVL	T wave amplitude in aVL, mV	-0.3 - 0.3	0.0815
T_V2	T wave amplitude in V_2, mV	-0.3 - 1.2	0.253
S_V3	S wave amplitude in V_3, mV	0.1 - 2.7	0.828
RaVL_SV3	Cornell sum, mV	0.2 - 3.8	1.39

Table 5.1: Anthropometric and ECG feature inputs to the machine learning model for the detection of abnormal left ventricular geometry. NT, normotensives; HTN, hypertensives.

Features	Categories	Acc (%)	Spec (%)	Sens (%)	AUC/ROC
6 clinical	NG vs. CR+LVH	79	73	84	0.85
32 (6 clinical+ 26 ECG)	**NG vs. CR+LVH**	**87**	**75**	**97**	**0.91**
32 (6 clinical+ 26 ECG+oversampling)	NG+CR vs. LVH	89	93	67	0.89

Table 5.2: Performance metrics for the machine learning classifier in discriminating between various categories. Acc, accuracy; Spec, specificity; Sens, sensitivity.

5.3 Discussion

We have shown the promising potential of ML modeling for the diagnostic screening of abnormal LVG and cardiac remodeling through ECG. We found specific clinical and ECG features that can predict early pathological changes of LVG in patients without established CVD and detect the population who will benefit from a detailed echocardiographic evaluation. We used not only the traditional ECG criteria for LVH but also a novel ECG marker that increased the accuracy of our ML model.

Our results show that age plays an important role in the risk of someone having `CR` or `LVH` with a cut-off around 64.5 years. The risk appears higher for men younger than 64.5 while after that age the risk seems higher for women. Our results indicate that BMI adjusted SL criteria seem to differentiate for the `CR` class. `Hypertension`, textttage, and `BMI` were most significant, as expected; the area under the QRS complex summed over all 12 leads, the Planar Frontal QRS-T angle, and QTc duration, among others, were important in predicting risk.

There are limited data in the literature that attempt to predict cardiac structural or functional abnormalities with ECG data interpreted through ML algorithms.[70,96] However, the existing knowledge has focused only on patients who have already shown LVH. There are no data for patients in earlier stages of cardiac geometry change prior to hypertrophy. The present prospective ML study also differs from previous ones in that it involves patients who were very carefully selected, thereby excluding those with CVD.[70] This may explain the fact that in our study, analysis of patients with LVH achieved a higher AUC in

comparison to recently published work,[70] despite the fact that the number of our patients is smaller. ML is susceptible to major errors in interpretation, and generalizability. The fact that participants in our study did not have CVD is a major strength since in effect it largely eliminates other clinical parameters that could mislead our model. In this way, we improve the quality of input data and avoid various pitfalls that could arise due to the large diversity of pathological conditions that formed the basis for the training process.

We have discovered new relationships and showed that a quantitative assessment of abnormal LVG can be performed by using easily obtained clinical data and ECG features. This novel approach has the potential to serve as a cost-effective screening tool for early detection of LV remodeling, to dramatically optimize treatment and patient-management. ECGs are more easily obtainable and cost-effective than echocardiography or cardiac magnetic resonance imaging (MRI), and for those reasons more often used in current clinical practice. The goal of hypertension treatment is to prevent pathological changes in LVG or to reverse CR and LVH. Deep learning could potentially detect patients with hypertension-mediated organ damage at an early stage and with simple and widely used clinical tools.

There are some limitations to our work. The number of subjects we have included is not large since this is a single center study over a specific time period. Nonetheless, our results are clear and mainly due to the fact that our subject population is carefully chosen and does not have other CVD that could influence ECG features.

Although cardiac MRI has been suggested as the most accurate diagnostic test for cardiac remodeling and LVH, it is limited by cost and lack of availability. Most importantly,

it is not recommended for routine clinical use for this reason. 2D echocardiography is the imaging test of choice for assessing those patients and is also the only guideline-approved modality for monitoring volumes and mass,[79,80] which also has well studied prognostic and clinical implications.[79,97]

We did not perform coronary angiogram in all patients, and this may bias the outcomes. However, we believe that this bias is small since patients underwent a meticulous work out to exclude coronary artery disease while performing coronary angiogram in low probability patients would be unethical. RWT is not always reflective of true LVG in patients with asymmetric hypertrophy. However, it is the most widely used index for this purpose in routine clinical practice for hypertensive patients.[79,80]

Fig. 5.3 ***(following page)*****:** Global and local importance for the 20 most important features in the RF binary classifier for detecting CR+LVH, and feature interactions for four of them. All plots are on the test set. **(a)**: Bar chart of mean feature importance for the classification. **(b)**: SHAP summary plot showing the effect of each feature on individual patients. The long tails indicate rare but high magnitude risk factors. The red part on the Hypertension row shows that hypertension has a strong positive effect on someone being classified as CR+LVH, and the spread of the blue line shows that non-hypertensive people have different risk for being classified as CR+LVH. **(c)**: Effect of age on detecting CR+LVH. The plot shows that age plays an important role in the risk of being classified as CR+LVH with a cut-off around 65 years of age. The risk is higher for men under 65, while over 65 the risk appears higher for women, as the blue and red dots indicate. **(d)**: Effect of the BMI/SL on detecting CR+LVH with a visible cut-off of around 18 kg/(m^2 mV). **(e)**: Effect of QRS-T angle on the same risk. Risk appears higher after a value of 27 degrees . **(f)**: Effect of QTc duration on the same risk, with a visible cut-off point at 420 ms.
RF: random forest, CR: concentric remodeling, ECG: electrocardiogram, NG: normal geometry, LVH: left ventricular hypertrophy, SHAP: (SHapley Additive exPlanations), BSA: body surface area, BMI: body mass index.

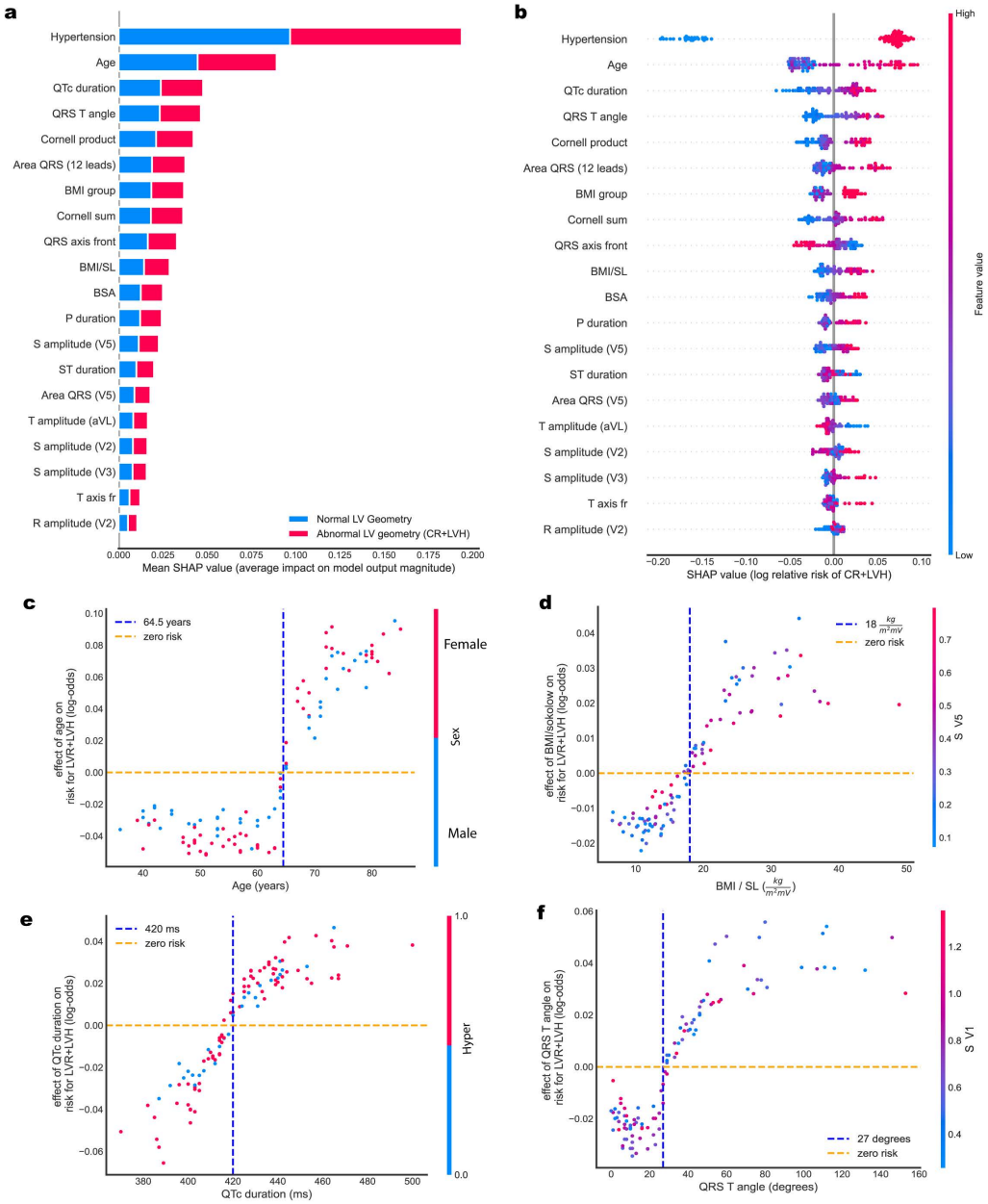

Fig. 5.3: (continued)

Everything should be made as simple as possible, but not simpler.

Quote attributed to Albert Einstein

6

Opportunistic screening for the detection of arterial hypertension through 12-lead ECG

Hypertension is defined as blood pressure (BP) of greater than 140 and/or 90 mmHg according to the ESC/ESH guidelines for hypertension [79] or 130 and/or 80 mmHg according to the ACC/AHA guidelines. [98] Essential hypertension is the type not attributed to under-

lying, identifiable cause. Hypertension is one of the most significant risk factors for cardiovascular disease (CVD) and a major cause of premature mortality and rising health care costs.[99] It is a leading modifiable cause in 54% of stroke cases and 47% of ischemic heart disease incidences worldwide.[100] The global prevalence of hypertensive heart disease, having risen steadily over the last decades, is expected to continue to rise due to population growth and aging.[101] Unfortunately, control rates among people with hypertension are very poor, approximately 23% for women and 18% for men, with a large number of hypertensives not properly identified.[102]

Unawareness of hypertension is an important contributing factor to the inadequate control of the disease and absence of appropriate antihypertension treatments. Population screening programs have shown that more than 50% of hypertensives were unaware they had hypertension.[103,104] Despite the progress in blood pressure (BP) measurement techniques, a substantial proportion of hypertensive patients is not identified as such, and are thus incorrectly diagnosed and managed.[105] Although current practices, with the use of ambulatory and home BP measurements, have become more powerful in detecting the 'real' hypertensive population by discarding the white-coat effect and discovering masked hypertension, still a large proportion of patients escape diagnosis. Considering the importance of this disease for public health, exploring novel tools that potentially minimize the unawareness and increase the diagnostic performance of hypertension in daily clinical practice seems of vital importance.

In this Chapter we show our work in detecting whether a person is hypertensive using features derived from a 12-lead ECG, as well as basic anthropometric features.

6.1 Feature selection

Initially we had 60 features in our dataset (see Table B.1). Some pairs of features exhibited high correlation, as calculated by `Spearman's` rank correlation test; highly correlated features contribute the same amount of information and including both of them in a RF model might not affect performance, but it will divide, thus lessen, each feature's predictive significance. We chose a cut-off of 0.90 % correlation for removal. In our `python` code we used the `spearmanr` function from the `scipy.stats` package. We choose the particular test because of the possibility of non-linear relationships among the data and then calculated the Rank-Sum statistic and ranked the features according to their p-value.

6.2 Results

After carefully screening 2156 healthy individuals, we enrolled 1091 consecutive subjects aged 30 to 80 years (Fig. 6.1). Of our participants, 617 (56.5 %) were female, 474 (43.5 %) were male, and 712 (65.2 %) were hypertensive based on the definition given above.

At the time of the study, 613 (86.22 %) of those hypertensives were receiving RAAS blockers, 185 (26 %) CCBs, 102 (14.35 %) diuretics, and 93 (5.49 %) some other form of medication. Of the non hypertensives, 11 (2.91 %) were receiving RAAS blockers, 93 (24 %) CCBs, 52 (13.75 %) diuretics, and 3 (0.79 %) some other form of medication. Overall, the mean age was 59.3± 11.2 years; 60.3±10.8 years for females, and 58±11.6 years for males. According to their BMI, 505 (46.3 %) of them were obese, 417 (38.0 %) were overweight, and 169 (15.0 %) were within normal range. Based on blood pressure cate-

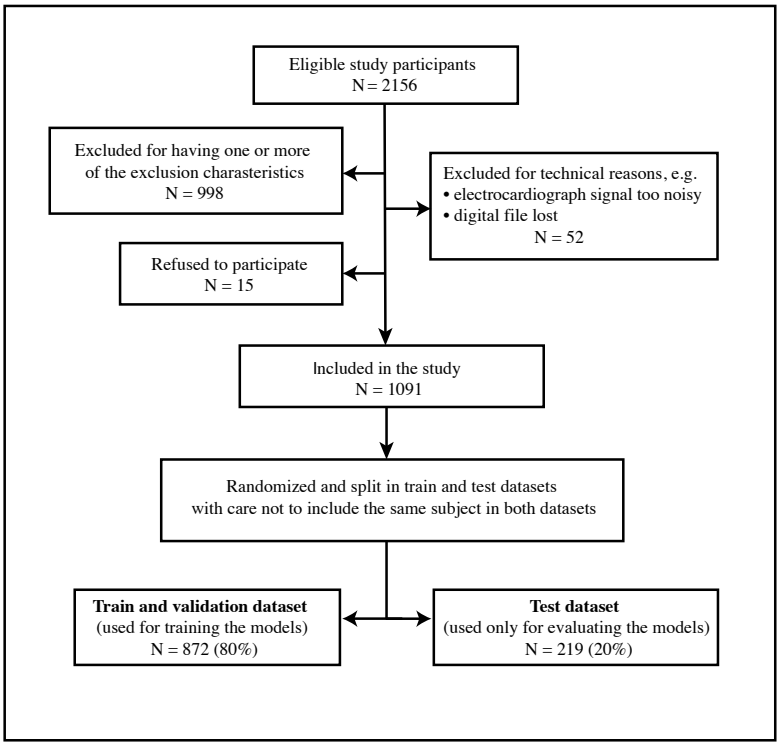

Fig. 6.1: Study flowchart of participant selection from initial hospital evaluation until data inclusion in the ML models.

gory, 6.7 % were optimal, 10.47 % were normal, 49.49 % were in the upper normal range, 27.55 % were grade 1 hypertensives, 5.14 % were grade 2 hypertensives, and 0.64 % were grade 3 hypertensives. Compared with the normotensive group, the hypertensive group was older, had higher BMI, and tended to have slightly more female participation. The comparative statistics for a range of anthropometric and ECG features between hypertensive and normotensive population are shown in Table 6.1.

Feature	HTN			NT			P-value*
	Mean	Std	Range	Mean	SD	Range	
Age, years	62.5	10.5	30.0 - 80.00	53.3	10.2	30.0 - 80.00	<0.001
Systolic blood presure, mmHg	139.8	13.8	100.0 - 185.00	127.5	9.1	91.0 - 175.00	<0.001
Diastolic blood pressure, mmHg	84.6	8.9	42.0 - 127.00	80.7	7.6	52.0 - 133.00	<0.001
Mean blood pressure, mmHg	103.0	9.2	68.0 - 146.30	96.3	7.2	70.7 - 147.00	<0.001
Heart rate, bpm	69.9	11.6	40.0 - 129.00	71.1	12.0	48.0 - 109.00	0.108
Body mass index, kgr/m^2	31.4	5.4	18.8 - 56.64	28.1	5.2	17.6 - 48.87	<0.001
Body fat, kgr/m^2	41.6	9.0	18.9 - 75.45	36.6	8.1	17.3 - 67.04	<0.001
BMI-adjusted Cornell, mV·kgr/m^2	45.1	18.5	4.3 - 128.13	34.3	16.2	3.1 - 118.90	<0.001
R wave amplitude in aVL, mV	0.6	0.3	0.0 - 1.91	0.5	0.3	0.0 - 1.61	<0.001
Cornell criteria, mV	1.4	0.5	0.1 - 3.77	1.2	0.5	0.1 - 3.37	<0.001
Area under R wave in I, ms·mV	9.3	3.9	0.8 - 33.36	7.8	3.3	0.8 - 21.10	<0.001
QRS axis front, degrees°	13.6	32.4	-77.0 - 188.00	26.0	30.6	-82.0 - 88.00	<0.001
Corrected QT interval, ms	423.7	24.6	337.0 - 500.00	414.8	22.1	364.0 - 506.00	<0.001
P wave duration, ms	114.6	15.5	0.0 - 196.00	111.7	10.6	75.0 - 149.00	<0.001
PQ interval duration, ms	167.0	27.4	0.0 - 277.00	159.0	20.8	112.0 - 226.00	<0.001
QT interval duration, ms	397.9	31.7	290.0 - 501.00	387.5	28.7	312.0 - 496.00	<0.001
R wave amplitude in III, mV	0.2	0.2	0.0 - 1.37	0.3	0.3	0.0 - 1.74	<0.001
Planar frontal QRS-T angle, degrees°	37.0	35.3	0.0 - 178.00	26.2	25.3	0.0 - 168.00	<0.001
Area under R wave in aVF, ms·mV	3.9	3.6	0.0 - 25.39	4.7	3.7	0.0 - 23.51	<0.001
Area under T wave divided by QRS complex area	1.0	0.5	0.0 - 3.46	1.1	0.5	0.1 - 3.12	<0.001
Area under R wave in III, ms·mV	1.9	2.4	0.0 - 21.05	2.6	3.0	0.0 - 19.12	<0.001
BMI-modified Sokolow-Lyon voltage, kgr/m^2· mV	17.5	7.7	4.9 - 96.88	15.4	5.9	5.8 - 41.95	<0.001
BMI-adjusted Sokolow-Lyon voltage, mV	2.6	0.6	0.2 - 5.66	2.4	0.6	1.0 - 4.73	<0.001
Total QRS area in all leads, ms·mV	291.2	78.0	118.4 - 734.18	272.2	78.4	133.9 - 942.09	<0.001
S wave amplitude in V_5, mV	0.4	0.3	0.0 - 1.46	0.3	0.2	0.0 - 1.60	<0.001
T wave amplitude in V_5, mV	0.3	0.2	-0.7 - 1.02	0.3	0.2	-0.5 - 0.82	<0.001
S wave amplitude in V_3, mV	0.9	0.4	0.0 - 2.84	0.8	0.4	0.0 - 3.15	<0.001
QRS complex duration, ms	92.6	11.0	62.0 - 153.00	90.6	11.2	55.0 - 147.00	0.002
P wave amplitude in II, mV	0.1	0.0	0.0 - 0.28	0.1	0.0	0.0 - 0.41	0.003
Area under QRS interval in V_5, ms·mV	34.7	12.3	10.5 - 93.14	32.5	11.1	11.7 - 97.14	0.004
Q vs. S vector	0.9	0.6	-0.8 - 2.91	0.9	0.6	-0.8 - 2.65	0.008
J point deflection, mV	-0.0	0.0	-0.1 - 0.14	-0.0	0.0	-0.1 - 0.13	0.01
Q wave duration, ms	10.5	8.1	0.0 - 36.00	11.9	8.6	0.0 - 48.00	0.01
P axis in frontal plane, degrees°	46.6	22.5	-59.0 - 116.00	50.4	26.4	-61.0 - 268.00	0.011
Intrincicoid deflection in II, ms	41.9	6.7	5.0 - 95.00	41.1	6.7	4.0 - 64.00	0.042
Area under S wave in V_1, ms·mV	18.6	11.5	0.0 - 77.63	17.2	9.9	0.0 - 51.79	0.054
T wave duration, ms, ms	204.0	42.3	43.0 - 345.00	200.6	32.4	77.0 - 312.00	0.071
T wave amplitude in III, mV	0.0	0.1	-0.4 - 0.37	0.0	0.1	-0.3 - 0.57	0.077

Table 6.1: Characteristics and Comparative Statistics for Hypertensive and Normotensive Study Participants. BMI: body mass index; SD: standard deviation; HTN: participants with hypertension, NT: participants not diagnosed with hypertension.

*The alternative hypothesis for this P-value is that there exists a difference between HTN and NT participants for each feature/variable

Embedding high-dimensional data into 2 dimensions allows us to visualize them in a way that gives useful insights on what differentiates study participants with a condition from those without. We performed visualization as part of our feature selection, before running any models, to inspect the features that seemed to best discriminate hypertensive from normotensive participants. We employed the t-SNE algorithm. We visualized various subsets of the anthropometric and ECG features using t-SNE. In Fig. 6.2, each point is a participant characterized by the following set of features: `age`, `BF`, `BMI_adj_cornell`, `R_aVL`, and `BMI_soko`. This particular subset of features seems to visually separate hypertensive patients, who are represented by dots mostly on the upper left corner, from normotensive patients which are the ones in the rest of the plot. This feature selection was part of preprocessing.

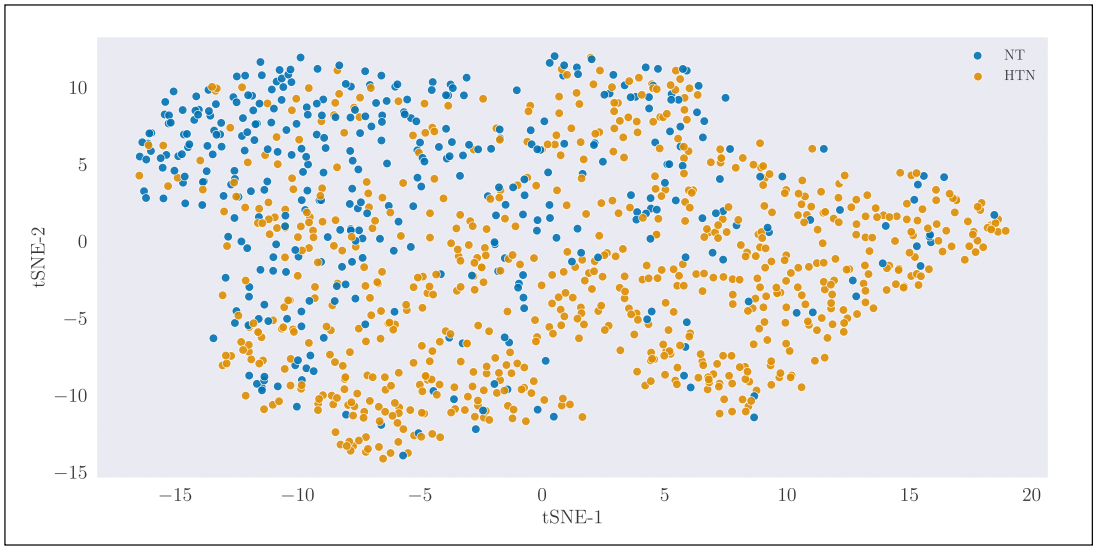

Fig. 6.2: Study subject clustering using t-distributed Stochastic Neighbor Embedding (t-SNE). NT signifies the normotensive participants and HTN the hypertensive. The axes of the 2-dimensional space are given in arbitrary units.

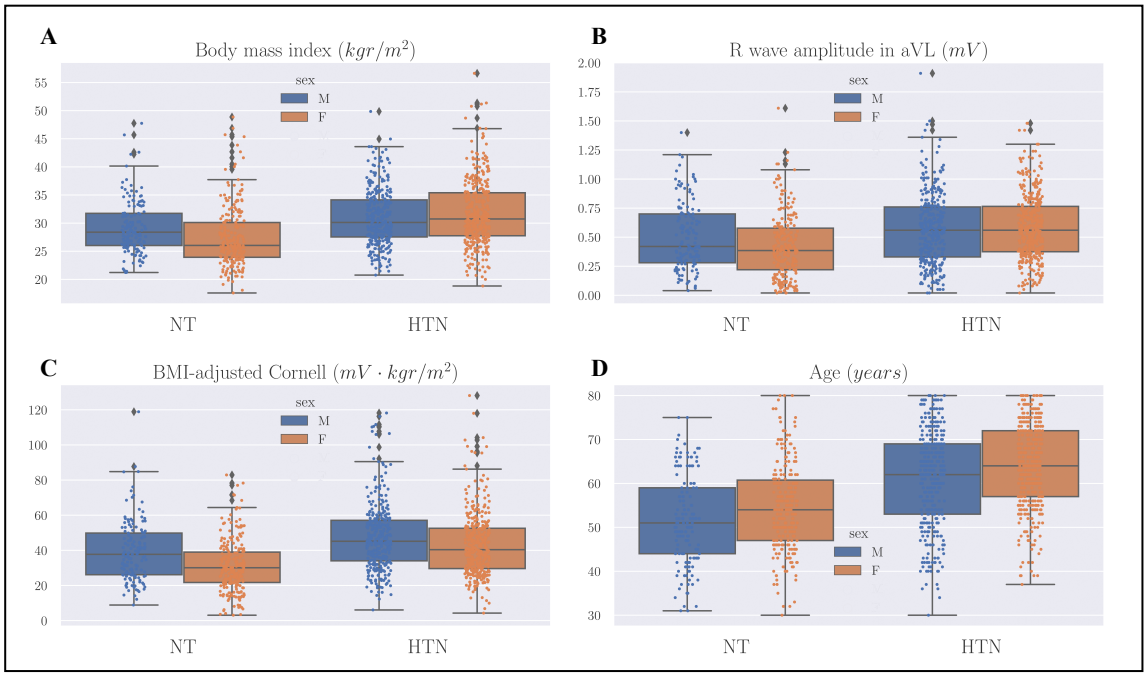

Fig. 6.3: Feature distribution comparison. These box plots show the distributions of body mass index (BMI) (**A**), R wave amplitude in aVL (**B**) , BMI-adjusted Cornell criteria (BMI multiplied by RaVL+SV_3) (**C**) , and age (**D**) , between normotensive (NT) and hypertensive (HTN) individuals. Scatterplots (dots) of the data were superimposed for a more detailed visualization of the distributions. Each plot is also subdivided in male and female participants.

Based on the discriminatory ability of these features, we depict the distributions of `BMI`, `R_aVL`, `BMI_adj_ cornell`, and `age`, between normotensive and hypertensive individuals in four separate plots, shown in Fig. 6.3). In each plot, we separate the distributions for male and female individuals. We notice that the distributions for `BMI`, `age`, and `BMI_adj_ornell` index, are shifted towards larger values for hypertensive than normotensive participants.

Our RF model's accuracy on detecting hypertension using was 84.2 %, its specificity was 66.7 %, and its sensitivity was 91.4 %; AUC/ROC was 0.86, for the standard decision

threshold of 0.5. By moving the threshold to 0.6, we increased our specificity to 78.0 %, without sacrificing the sensitivity too much (new value was 84.0 % vs. 91.4 %). The results for all our models are in Table 6.2

Model	Features	Accuracy (%)	AUC(ROC)	Sensitivity (%)	Specificity (%)
Random Forest 1 (thresh.= 0.6)	age, sex, BF, BMI-adj-Cornell, R in aVL, BMI-modified SL	**84.2**	**0.89**	**84.0**	**78.0**
Random Forest 1	age, sex, BF, BMI-adj Cornell, R in aVL, BMI-modified SL	84.2	0.86	91.4	66.7
Random Forest 2	age, sex, BF, BMI-adj-Cornell, R in aVL	82.0	0.86	90.0	66.0
Random Forest 3	age, sex, BF, BMI-adj Cornell, R in aVL, BMI-modified SL, BMI	83.2	0.867	91.4	63.2
Logistic Regression	age, sex, BF, BMI-adj Cornell, R in aVL, BMI-modified SL	77.8	0.77	93.6	39.7
K Nearest Neighbors	age, sex, BF, BMI-adj-Cornell, R in aVL, BMI-modified SL	78.8	0.87	97.6	40.2

Table 6.2: Classification performance metrics for random forest, logistic regression, and k-nearest neighbors with the respective features used in training.

Feature importance calculated by SHAP is shown in Fig. 6.4 (A). Dependence on BMI-adjusted Cornell criteria is shown in Fig. 6.4 (B). The horizontal dashed line represents the cut-off between having a negative effect on being hypertensive (below the line) and a positive one (above the line). On the x-axis we see that participants with a value above 37 mV·kgr/m^2 (approximately) have a positive chance of being hypertensive. These values were calculated by SHAP on the RF model. Training the model to model to predict in which BP category each participant belonged did not give any meaningful results maybe because the data used were one-time measurements that were taken on different days than the ECG, thus having little association.

6.3 Discussion

To our knowledge, this is the first clinical study that exploits the promising potential of ML algorithms for the efficient and cost-effective opportunistic screening of arterial hypertension. We found specific basic clinical and ECG features that can be applied for point-of care detection of hypertensive population who will benefit from further evaluation and treatment. In our study, `age`, `BF`, `BMI_adj_cornell`, `R_aVL`, and `BMI_soko`, seemed to separate hypertensive patients from normotensive (Figures 6.3 and 6.4A). It is remarkable that using just these features our model can detect hypertension with a good accuracy. Our findings could be useful because, although hypertension is a leading preventable risk factor for premature death and disability worldwide, the proportion of awareness and treatment remains poor [103,104,79,93]. We showed that with familiar and easily obtainable clinical tools we can enhance the diagnostic efficacy and improve the detection of hypertension.

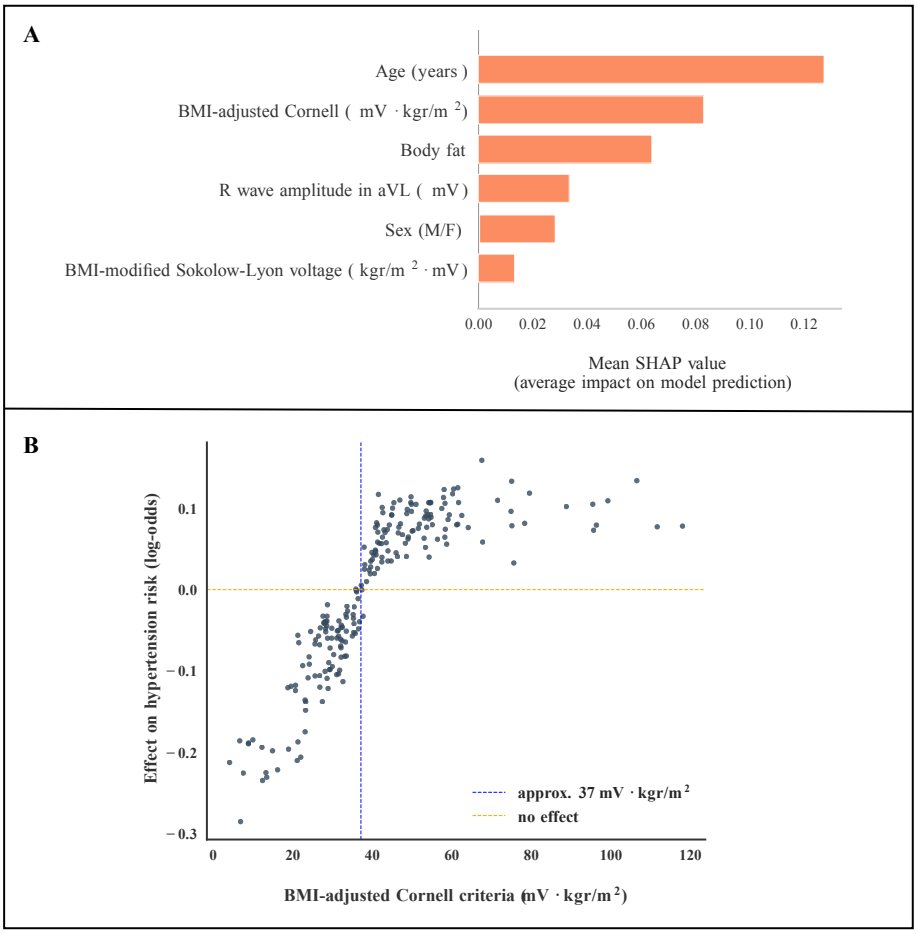

Fig. 6.4: Results in detecting hypertension by the Random Forest. (**A**) Feature importances calculated on the test set using Shapley Additive explanations (SHAP). Features are (BMI), BMIsoko, and BMIadjcornell. Body fat (BF) is defined by the formulae: [adult males] BF(%) = 1.20 · BMI + 0.23 · age - 10.8 - 5.4, and [adult females] BF(%) = 1.20 · BMI + 0.23 · age - 5.4. Sex is binary male/female (M/F). (**B**) Effect of BMI adjusted Cornell criteria on the risk of being hypertensive. Each dot in the plot is a participant whose BMI adjusted Cornell value is indicated on the x-axis. The values on the y-axis effectively indicate the effect of each participant's set of features in characterizing them as hypertensive.

Our study design has several strengths. First, the participants were carefully selected and did not have CVD since it largely eliminates other clinical parameters that could mislead our model. In this way, we improve the quality of input data and avoid various pitfalls that could arise due to the large diversity of pathological conditions that formed the basis for the training process. Second, our data were collected prospectively across real-world clinical settings and by many operators. There are limited data in the literature that attempt to associate BP level with ECG signals interpreted through ML algorithms [106,107,108]. However, most of the existing knowledge was derived from limited number of ECGs acquired from public databases such as *Physionet*. Our results are clear because our population was carefully chosen not to have CVD that could influence ECG features.

We present an ECG-based ML algorithm that can identify the existence of arterial hypertension by using easily obtained clinical data and ECG features in the clinical setting. This novel approach has the potential to serve as a cost-effective screening tool, empower clinicians to detect hypertensive participants and eliminate the effects of white coat and masked hypertension in the routine clinical practice. Our model contributes to the development of human-centered and autonomous technologies and can optimize patient-management.

Our study has some limitations. We could not control for every possible lifestyle factor, and there is possibility of residual confounding. Most of the people that were excluded did not meet the eligibility criteria or had one of the characteristics for exclusion. Only 15 refused to participate; this number is low and that reduces the potential bias and ensures the generalizability of our study findings.

7

Diagnostic performance of single-lead ECG in the detection of arterial hypertension

Hypertension was our focus in Chapter 6, and especially the fact that a large number of hypertensives are not aware of their condition. We showed that from basic clinical data and the use of the 12-lead ECG, we can identify participants with arterial hypertension.

In this Chapter we are trying to detect hypertension using anthropometric features and a single-ECG, namely of lead I.

ECG technology is now ubiquitous in smartwatches and other wearables and can record a single-lead ECG which is equivalent to lead I on a standard 12-lead ECG. So far wearable ECGs have consistently demonstrated their value in detecting arrhythmias such as atrial fibrillation.[109] Although still challenging, detecting cardiac diseases using a single or reduced number of leads has appeared in the literature.[110] In recent studies, single-lead ECGs were used to detect T-wave (due to ventricular repolarisation) morphology abnormalities,[111] to develop an automatic mental stress detection system based on ECG signals from smart T-shirts,[112] and to identify patients with atrial fibrillation-induced heart failure.[113]

The rapidly growing technology and popularity of wearables may offer the opportunity for detecting cardiovascular diseases using single-lead ECGs. This can increase dramatically HTN awareness in the general population, improve HTN control and transform the way health care is delivering. Our aim was to develop ML algorithms to detect HTN from single-lead ECGs in subjects without cardiovascular disease.

7.1 Feature selection

A 12-lead ECG of 10 seconds duration in resting position was performed on each subject, as described in Chapter 4. In order to simulate devices that record only a single lead, we kept the tracings of only lead I in our digital files for further processing.

The dataset was randomly partitioned into a train set of 1128 (90 %) data points, used directly to learn the parameters of the model, and a test set of 126 (10 %), consisting of a

part of the data the model had not seen before and was used exclusively for final performance evaluation of the models. Stratification for sex and history of hypertension, during the partition, ensured the two sets contained the same proportions of these two features. The ratio of hypertensive to normotensive patients in the total dataset was about 2, which did not, in our view, necessitate the use of techniques for imbalanced datasets.

7.2 Results

After careful screening 2156 healthy individuals, we enrolled 1254 consecutive subjects (Fig. 7.1). Of our participants, 715 (57.02 %) were female, 539 (42.98%) were male, and 831 (66.27%) were hypertensive. Overall, the mean age was 60.22±12.46 years old; 61.28±11.95 for females, and 58.8±12.98 for males. Based on BMI, 589 (46.97%) of them were obese, 475 (37.88%) were overweight, and 190 (15.15%) were within normal range. Compared with the normotensive group, the hypertensive group was older, had higher BMI, and tended to have slightly more female participation.

The comparative statistics for a range of anthropometric and ECG features between hypertensive and normotensive population are shown in Table 7.1. The prevalence of hypertensives in our study was 66.2%. Fig. 7.3 uses hierarchical clustering to show, along with descriptive names and clustering tendency, the 35 features in the model. In classifying hypertensive vs. normotensive, our RF model's accuracy was 81%, with specificity equal to 70%, sensitivity 88%, and area under the receiver operating characteristic curve (AUC/ROC) 0.82, as shown in Fig. 7.2.

Feature importance calculated by SHAP is shown in Fig. 7.4. As expected, age and BMI

Feature	Hypertensives			Normotensives			p value
	Mean	Std	Range	Mean	Std	Range	
Age, years	63.8	11.5	20.0 - 90.00	53.1	11.1	22.0 - 86.00	<0.001
BMI, kgr/m^2	31.4	5.4	18.8 - 56.64	28.4	5.5	17.6 - 50.47	<0.001
BMI-adjusted QRS complex area, ms·mV·kgr/m^2 (area under the QRS complex multiplied by BMI)	709.5	291.8	152.7 - 1846.01	554.3	280.1	75.9 - 1885.32	<0.001
Area under T wave divided by QRS complex area, (-)	1.0	0.5	0.0 - 3.25	1.3	0.7	0.1 - 5.23	<0.001
QRS axis frontal plane, degrees	13.3	31.9	-77.0 - 188.00	25.4	31.1	-82.0 - 128.00	<0.001
Corrected QT interval, ms	422.7	24.4	337.0 - 500.00	414.2	22.6	355.0 - 506.00	<0.001
PR interval duration, ms	168.5	25.6	99.0 - 277.00	160.4	23.3	112.0 - 341.00	<0.001
QT interval duration, ms	397.7	31.0	290.0 - 501.00	387.5	29.4	264.0 - 496.00	<0.001
R wave amplitude, mV	0.8	0.3	0.2 - 1.82	0.7	0.3	0.1 - 1.55	<0.001
ST segment, ms	112.6	43.2	0.0 - 289.00	101.6	35.3	0.0 - 229.00	<0.001
P wave duration	114.7	14.5	50.0 - 227.00	112.0	10.9	75.0 - 149.00	<0.001
T wave amplitude, mV	0.2	0.1	-0.4 - 0.56	0.2	0.1	-0.2 - 0.62	<0.001
Body surface area, m^2	2.0	0.2	1.3 - 3.08	1.9	0.2	1.4 - 2.79	<0.001
Height, cm	164.9	10.0	142.0 - 192.00	167.1	9.7	137.0 - 200.00	<0.001
Intrinsicoid deflection, ms	42.3	6.0	25.0 - 69.00	41.3	5.9	21.0 - 67.00	0.007
Area under S wave, ms·mV	1.2	2.0	0.0 - 14.17	1.3	2.1	0.0 - 23.32	0.008
QRS complex duration, ms	92.1	11.1	62.0 - 153.00	90.6	11.0	55.0 - 147.00	0.016
P axis in frontal plane, degrees	46.4	23.6	-80.0 - 116.00	50.1	26.9	-61.0 - 268.00	0.042
Heart rate, bpm	69.8	11.2	40.0 - 129.00	71.2	12.3	48.0 - 123.00	0.091
T wave, ms	192.9	39.4	56.0 - 357.00	195.3	30.4	94.0 - 372.00	0.278
S wave amplitude, mV	0.1	0.1	0.0 - 0.54	0.1	0.1	0.0 - 0.87	0.478
T axis frontal plane, degrees	39.2	35.9	-87.0 - 261.00	35.9	23.4	-58.0 - 154.00	0.49
Q wave, ms	15.1	8.1	0.0 - 55.00	15.6	7.7	0.0 - 40.00	0.648
P wave amplitude, mV	0.1	0.0	0.0 - 0.19	0.1	0.0	0.0 - 0.21	0.778
J point deflection, mV	0.0	0.0	-0.1 - 0.12	0.0	0.0	-0.1 - 0.12	0.949

BMI: body mass index.

Table 7.1: Characteristics and Comparative Statistics for Hypertensive and Normotensive Study Participants.

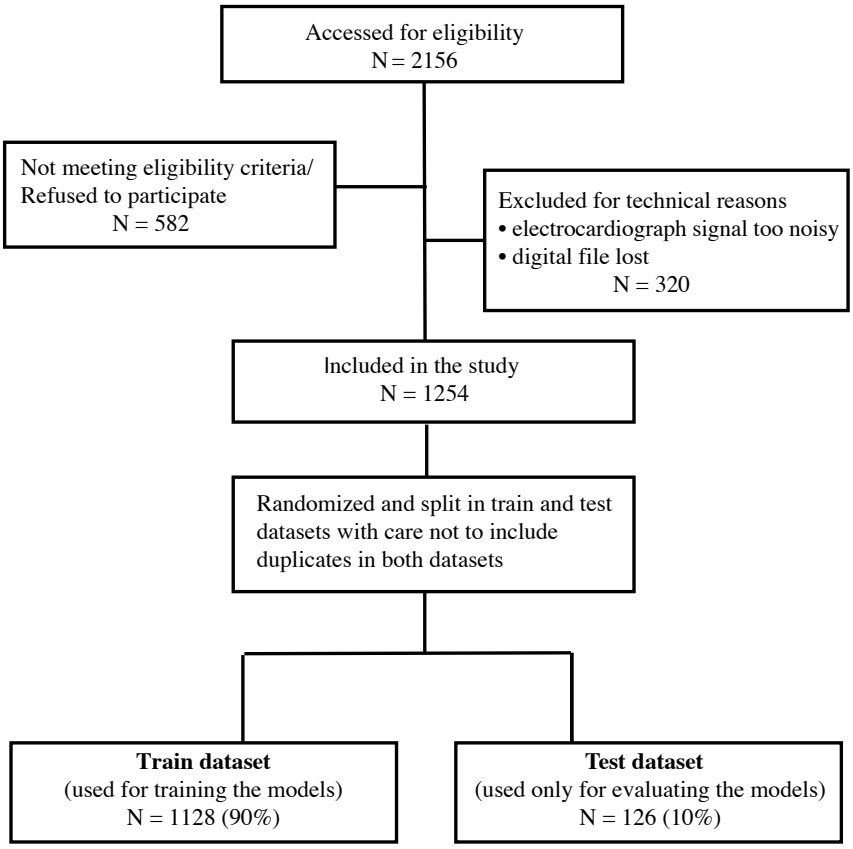

Fig. 7.1: Study flowchart.

were two of the most important features in our model. Other features that seemed to separate hypertensive patients from normotensive were the BMI-adjusted area under QRS complex area (defined as the area under the QRS complex multiplied by BMI), and the area under the T wave divided by QRS complex area. Other features that came as important, but less so, were the R-wave amplitude, the QRS axis in the frontal plane, the corrected QT interval duration, and the PR interval duration.

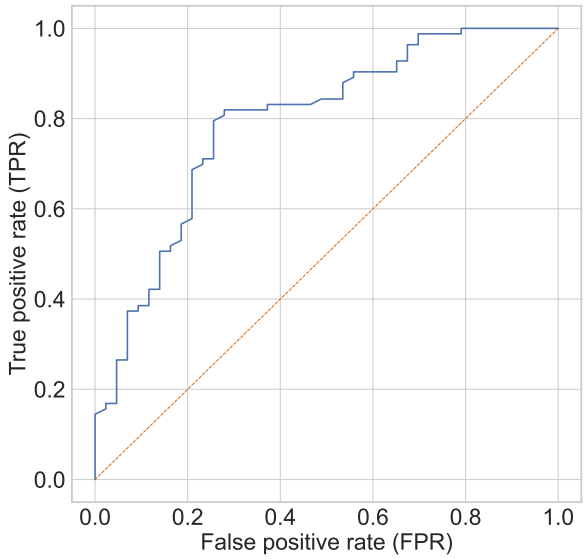

Fig. 7.2: Receiver operating characteristic (ROC) curve depicting the diagnostic performance of our Random Forest (RF) model, as the true positive rate (TPR) against the false positive rate (FPR). FPR is equal to 1 − specificity, and TPR is equal to the sensitivity. The AUC/ROC is 0.82

7.3 Discussion

In this prospective analysis of single-lead ECGs, we found that single-lead ECG features can be applied for the detection of arterial HTN. To our knowledge, this is the first clinical study that exploits the promising potential of ML algorithms for the efficient and cost-effective opportunistic screening of arterial HTN using a single-ECG reading and detect individuals who will benefit from further evaluation and treatment. According to our findings, age, BMI-adjusted area under total QRS area, body mass index, area under T wave divided by QRS complex area, seemed to separate hypertensive patients from normotensive.

Our results are significant given that HTN is a major public health issue and a lead-

ing preventable risk factor for premature death and disability worldwide.[114] Notably, a large proportion of hypertensive patients remained undiagnosed and unaware of their clinical condition.[99] Despite the increasing prevalence, the proportions of hypertension awareness and BP home monitoring are still low in the general population.[103,104,79,93] On the other hand, wearable devices and smartwatches are becoming increasingly popular and now have started to make a significant impact on the healthcare industry. Artificial intelligence-based ECGs analysis has been combined with wearable devices to investigate various cardiac pathologies and has already redefined assessment of several clinical conditions such as arrhythmias.[115] A lead I equivalent ECG strip can be obtained by several smartwatches and can provide very valuable clinical information. The applications of artificial intelligence on ECG are evolving rapidly with tremendous future implications on cardiovascular medicine.[62,63,64,61,65] ECG signals and patterns largely unrecognizable to human eye interpretation can be detected by ML algorithms making the ECG a more powerful, non-invasive clinical tool. Moreover, single-lead smartwatch ECGs have been shown to be accurate and provide a significant amount of information, such as heart rate, cardiac conduction, as well as rhythm. We showed that with familiar and easily obtainable single-lead ECG may enhance the diagnostic efficacy and improve the detection of hypertension. This might have a dramatic impact in the global HTN management.

The unawareness of HTN which is an important contributing factor for the inadequate control of the disease and often the diagnosis of HTN is challenging and demanding even in the office. BP measurements are not always optimally performed in the routine clinical practice or even skipped altogether. Most importantly, a modest elevation in BP demands a

confirmation of at least two occasions while the exclusion of clinical entities such as white coat effect or masked hypertension depends on out-of-office measurements.[79,93] The increase the diagnostic performance of HTN in the general population and the instantly derived screening at points-of care are significant for the management of the escalating burden of HTN.

Our study design has several strengths. First, the participants were carefully selected and did not have CVD since it largely eliminates other clinical parameters that could mislead our model. In this way, we improve the quality of input data and avoid various pitfalls that could arise due to the large diversity of pathological conditions that formed the basis for the training process. Second, our data were collected prospectively across real-world clinical settings and by many operators. There are limited data in the literature that attempt to associate HTN with ECG signals interpreted through ML algorithms.[106,107,108] Our group has developed ML models showing that detection of HTN through a 12-lead ECG is feasible.[116]

We present an ECG-based ML algorithm that can identify the existence of arterial hypertension by using easily obtained clinical data and single-lead ECG features in the clinical setting. Apart from the raised awareness, this novel approach has the potential to serve as a cost-effective screening tool, empower clinicians to detect hypertensive participants and eliminate the effects of white coat and masked hypertension in the routine clinical practice. Deep learning opens up new opportunities in health care quality and advancing personalized medicine at a low cost. Our model contributes to the development of human-centered and autonomous technologies and can optimize patient-management.

Our study has some limitations. We could not control for every possible lifestyle factor, and there is possibility of residual confounding. Nonetheless, our results are clear and mainly due to the fact that our population is carefully chosen and does not have other CVD that could influence ECG features.We did not perform coronary angiogram in all participants, and this may bias the outcomes. However, we believe that this bias is small since participants underwent a meticulous work out to exclude coronary artery disease while performing coronary angiogram in low probability participants would be unethical. Finally, more research is needed and the existing data should be validated different populations and ethnic groups.

The definition of hypertension considered office BP levels > 140 and/or 90 mmHg and was in accordance with the ESC/ESH guidelines for hypertension[79] instead of > 130 and/or 80 mmHg according to the ACC/AHA guidelines.[98] This might have an impact on our results.

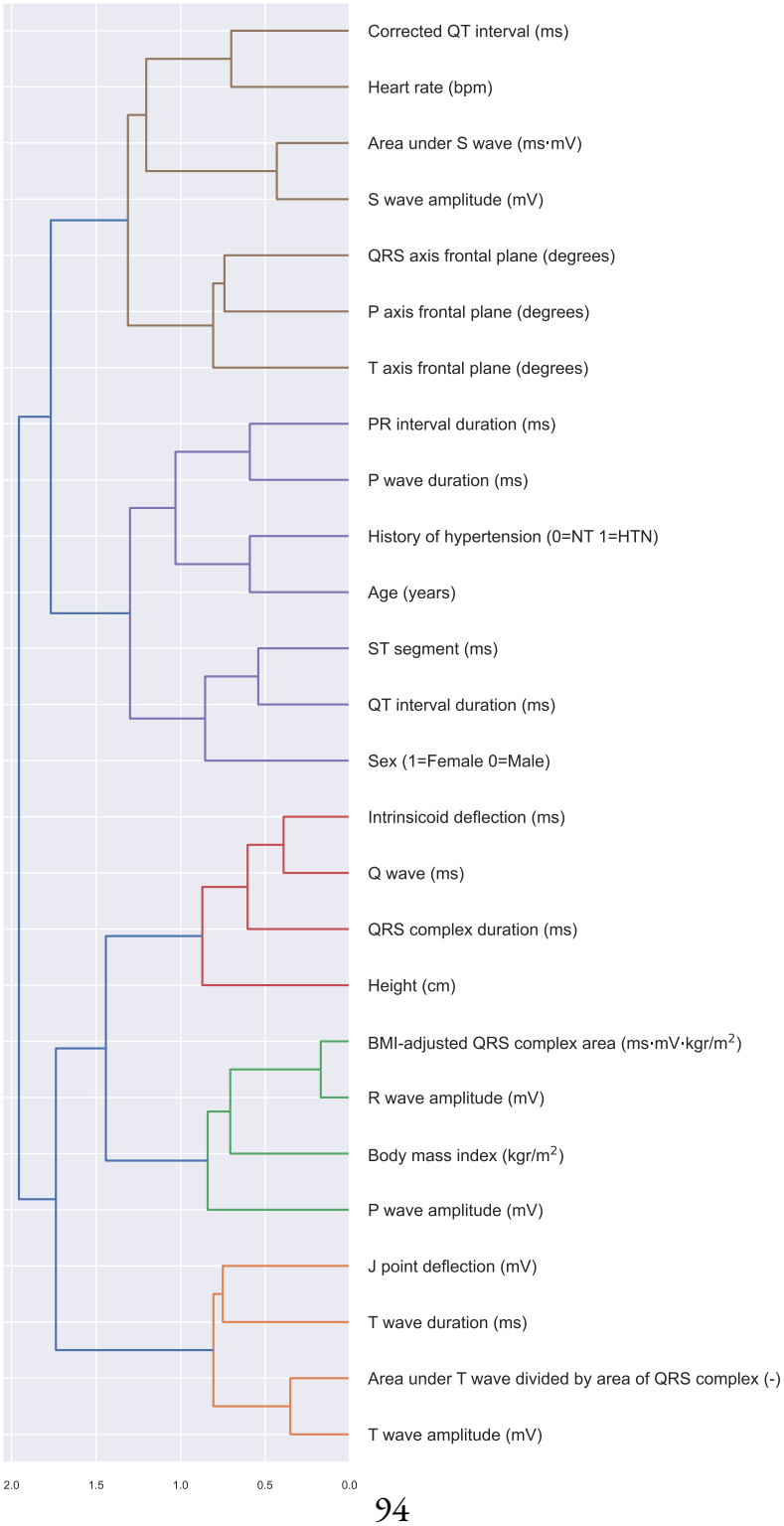

Fig. 7.3: Feature hierarchical clustering as a dendrogram

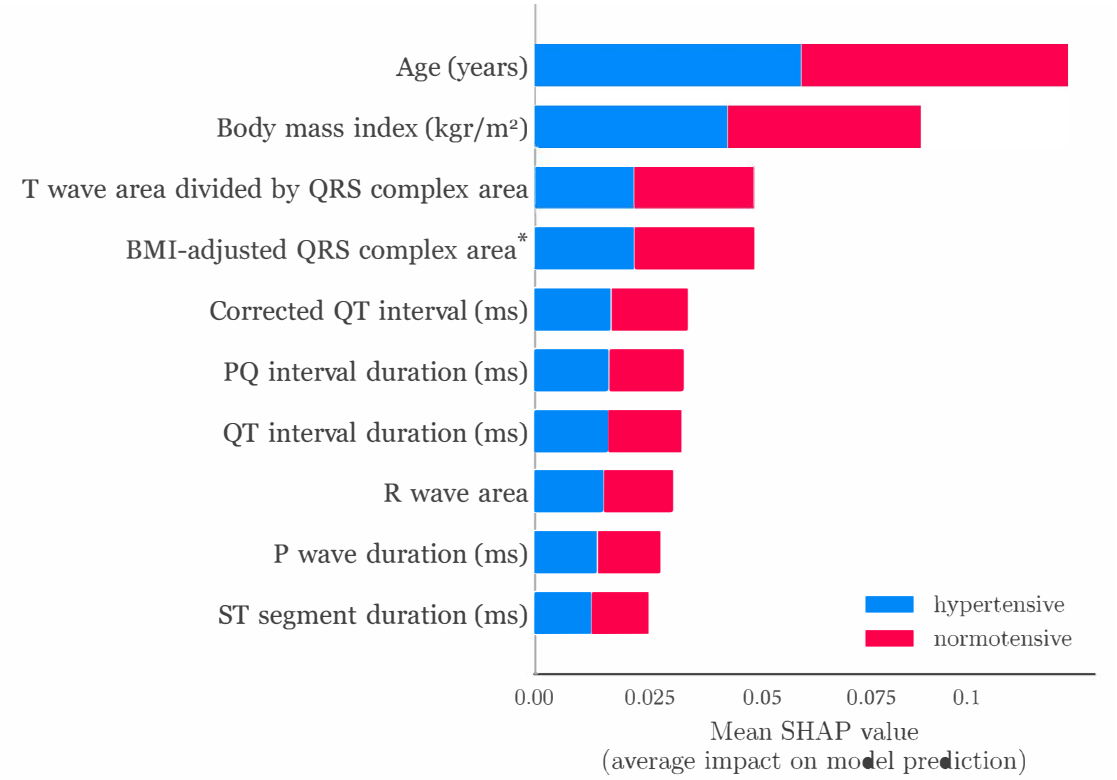

Fig. 7.4: Feature importance for the Random Forest model, calculated using *Shapley Additive explanations* (SHAP). Mean SHAP value, shown in x-axis, represents each feature's contribution in predicting hypertension. *BMI-adjusted QRS complex area is defined as the area under the QRS complex multiplied by BMI, in ms·mV·kgr/m^2.

Since all models are wrong, the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena.

George E. P. Box, British statistician, in his celebrated 1976 paper[117]

8

Equations of cardiac electrical propagation

Differential equations make the heart tick. Cardiac electrical propagation is often modeled as a reaction-diffusion process represented by sets of nonlinear ordinary (ODEs) or partial differential equations (PDEs). Computational models allow the study of action potential propagation in single cells, in a 1-dimensional cable of cells, in 2-dimensional slabs of tissue, or in 3-dimensional whole heart models. Alterations in the

electrical properties of ventricular tissue form the basis of ischemic arrhythmogenesis producing changes in the AP pulses and the ECG.[17,18] Such alterations, involving, for example, the remodeling of ionic currents due to changes in intracellular and extracellular ionic concentrations, has been studied in the literature.[19] Spatial heterogeneity such as cell-to-cell decoupling, occurring usually in later stages of ischemia, has also been shown experimentally to lead to propagation disruptions and a reduction in conduction velocity.[20]

Since Hodgkin and Huxley[118] proposed the first model for the AP in 1952, many ionic models have been proposed, each reproducing the AP in cardiac tissue, with few or more number of terms (introduced later in the Chapter).

8.1 The cable equations

Simulating transmembrane potentials from the level of a single myocyte up to measurable current on the body surface, is an intractable problem, if not in terms of representing them in mathematical terms, but because of computer memory and computational efficiency; there are thousands of ionic channels in a myocyte and billions of myocytes in the heart. We start with a simple depiction of the cardiac excitable tissue in one spatial dimension as a series of myocytes of length Δx electrically connected via gap junctions.[119] Fig. 8.1 shows a simplistic drawing of an "electrical circuit" comprised of a series of cells. Gap junctions are modeled as *resistors* having a resistance $r_g \Delta x$, where r_g is the resistance per unit length – larger than the resistance inside of the cell (cytoplasm) which we ignore. Each cell communicates with the extracellular environment via a) the membrane *capacitance* $C_m = c\Delta x$, where c is the membrane capacitance per unit length, and b) ion *channels* located on the

membrane, where $i_m \Delta x$ is the current per unit length flowing through the channel. The ion channels are nonlinear complicated structures, permeable only to specific ions.

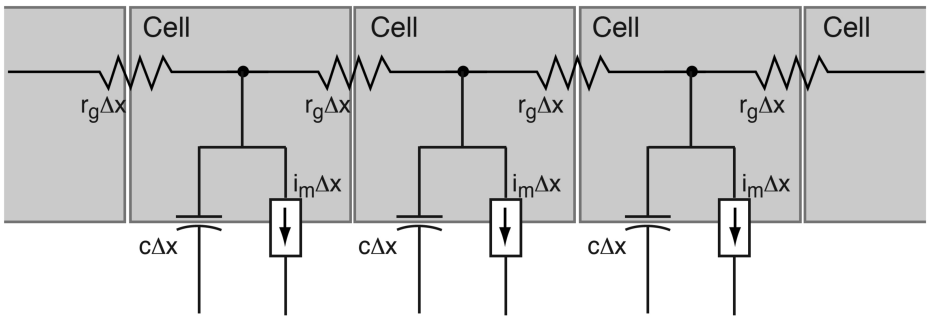

Fig. 8.1: Partial circuit representation of a series of excitable cells. Figure from[119]. Used with kind permission from Niels F. Otani.

We continue with the drawing in Fig. 8.2 which includes the extracellular environment. We set our coordinate system so that each cell center is at position x, its right neighbor at position $x + \Delta x$, and its left neighbor at position $x - \Delta x$. Since we are interested in voltages, we refer to the voltage inside the cell as $\Phi_i(x, t)$ (i for *inside*), and the voltage outside the cell as $\Phi_e(x, t)$ (e for *external*). This being a dynamical system, voltages are functions of both space x and time t; for brevity we will leave out the time dependence for now and add it later. We consider the gap junctions as functions of x (although they could also be time dependent), located at either $x + \Delta x/2$ or $x - \Delta x/2$, and having resistances of $r_g(x + \Delta x/2)\Delta x$ and $r_g(x - \Delta x/2)\Delta x$ respectively. We also introduce two stimulus currents, $i_{intracell}$ for the intracellular, and $i_{extracell}$ representing the extracellular, both currents per unit length.

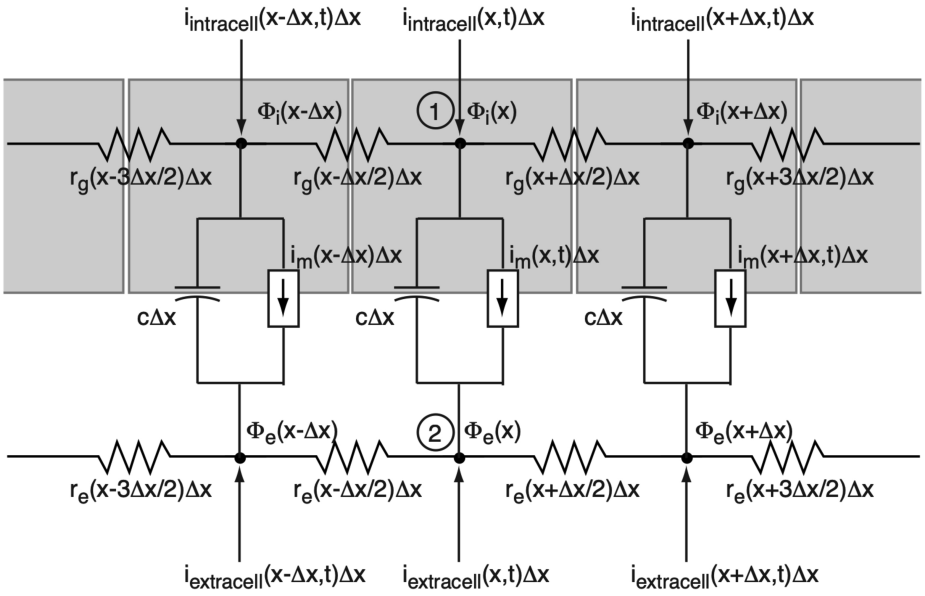

Fig. 8.2: Complete circuit representation of a series of excitable cells, including the intracellular and extracellular voltages, as well as internal and external stimulus currents. Figure from[119]. Used with kind permission from Niels F. Otani.

The currents entering and leaving node "1" in figure 8.2 are:

a) one leaving the node from the left

$$\frac{\Phi_i(x) - \Phi_i(x-\Delta x)}{r_g(x-\Delta x/2)\Delta x}\,, \tag{8.1}$$

b) one leaving the node from the right

$$\frac{\Phi_i(x) - \Phi_i(x+\Delta x)}{r_g(x+\Delta x/2)\Delta x}\,, \tag{8.2}$$

c) one flowing away through the membrane capacitor

$$C_{\mathrm{m}}\frac{dV_{\mathrm{m}}}{dt} = c\Delta x\frac{(\Phi_{\mathrm{i}}(x) - \Phi_{\mathrm{e}}(x))}{dt}\ , \tag{8.3}$$

where C_{m} is the capacitance, and V_{m} is the voltage drop across the membrane,

d) the intracellular going in (hence the negative sign)

$$-\mathrm{i}_{\mathrm{intracell}}(x)\Delta x\ , \tag{8.4}$$

e) one flowing away via the ion channels

$$\mathrm{i}_{\mathrm{m}}(x)\Delta x\ , \tag{8.5}$$

Adding Eq. (8.1) to Eq. (8.2) and rearranging the terms, we have

$$-\left[\frac{\Phi_{\mathrm{i}}(x+\Delta x) - \Phi_{\mathrm{i}}(x)}{\Delta x}\cdot\frac{1}{\mathrm{r}_{\mathrm{g}}(x+\Delta x/2)} \quad - \frac{\Phi_{\mathrm{i}}(x) - \Phi_{\mathrm{i}}(x-\Delta x)}{\Delta x}\cdot\frac{1}{\mathrm{r}_{\mathrm{g}}(x-\Delta x/2)}\right]\ . \tag{8.6}$$

Noticing that the first two terms of each addend are the derivatives at points $(x + \Delta x)$ and (x) respectively, and assigning

$$\mathrm{D}_{\mathrm{g}}(x) \equiv \frac{1}{\mathrm{r}_{\mathrm{g}}(x-\Delta x/2)}\ , \quad \text{and} \quad \mathrm{D}_{\mathrm{g}}(x+\Delta x) \equiv \frac{1}{\mathrm{r}_{\mathrm{g}}(x+\Delta x/2)}\ , \tag{8.7}$$

Eq. (8.6) becomes

$$-\left[\left.\frac{\Delta\Phi_{\mathrm{i}}}{\Delta x}\right|_{x+\Delta x}\cdot \mathrm{D_g}(x+\Delta x)-\left.\frac{\Delta\Phi_{\mathrm{i}}}{\Delta x}\right|_{x}\cdot \mathrm{D_g}(x)\right]\ . \tag{8.8}$$

Adding Eqs (8.3), (8.5), (8.8), and (8.4), and applying Kirchhoff's 1st law to node "1" (all currents should add to 0), we have

$$\begin{aligned}&\mathrm{c}\Delta x\frac{(\Phi_{\mathrm{i}}(x)-\Phi_{\mathrm{e}}(x))}{dt}+\mathrm{i_m}(x)\Delta x-\mathrm{i_{intracell}}(x)\Delta x\\&-\left[\left.\frac{\Delta\Phi_{\mathrm{i}}}{\Delta x}\right|_{x+\Delta x}\cdot \mathrm{D_g}(x+\Delta x)-\left.\frac{\Delta\Phi_{\mathrm{i}}}{\Delta x}\right|_{x}\cdot \mathrm{D_g}(x)\right]=0\end{aligned} \tag{8.9}$$

Dividing by $\mathrm{c}\Delta x$, and substituting $\mathrm{V_m}$ for $\Phi_{\mathrm{i}}(x)-\Phi_{\mathrm{e}}(x)$ we obtain

$$\begin{aligned}&\frac{\partial \mathrm{V_m}}{\partial t}(x)+\frac{\mathrm{i_m}(x)}{\mathrm{c}}-\frac{\mathrm{i_{intracell}}(x)}{\mathrm{c}}\\&-\frac{1}{\mathrm{c}\Delta x}\left[\frac{\partial\Phi_{\mathrm{i}}}{\partial x}(x+\Delta x)\cdot \mathrm{D_g}(x+\Delta x)-\frac{\partial\Phi_{\mathrm{i}}}{\partial x}(x)\cdot \mathrm{D_g}(x)\right]=0\end{aligned} \tag{8.10}$$

Substituting

$$\mathrm{D_g^*}=\frac{\mathrm{D_g}}{\mathrm{c}} \tag{8.11}$$

in Eq. (8.10) we have

$$\begin{aligned}(8.12)\qquad &\frac{\partial \mathrm{V_m}}{\partial t}(x) + \frac{\mathrm{i_m}(x)}{\mathrm{c}} - \frac{\mathrm{i_{intracell}}(x)}{\mathrm{c}} \\ &- \frac{1}{\Delta x} \cdot \left[\frac{\partial \Phi_\mathrm{i}}{\partial x}(x+\Delta x) \cdot \mathrm{D_g^*}(x+\Delta x) - \frac{\partial \Phi_\mathrm{i}}{\partial x}(x) \cdot \mathrm{D_g^*}(x) \right] = 0 \ .\end{aligned}$$

Noticing that

(8.13)

$$\lim_{\Delta x \to 0} \left\{ \frac{1}{\Delta x} \cdot \left[\frac{\partial \Phi_\mathrm{i}}{\partial x}(x+\Delta x) \cdot \mathrm{D_g^*}(x+\Delta x) - \frac{\partial \Phi_\mathrm{i}}{\partial x}(x) \cdot \mathrm{D_g^*}(x) \right] \right\} = \frac{\partial}{\partial x} \left[\frac{\partial \Phi_\mathrm{i}}{\partial x}(x) \cdot \mathrm{D_g^*} \right] ,$$

and substituting this result in Eq. (8.12), we obtain

$$(8.14)\qquad \frac{\partial \mathrm{V_m}}{\partial t}(x) + \frac{\mathrm{i_m}(x)}{\mathrm{c}} - \frac{\mathrm{i_{intracell}}(x)}{\mathrm{c}} - \frac{\partial}{\partial x} \left[\frac{\partial \Phi_\mathrm{i}}{\partial x}(x) \cdot \mathrm{D_g^*}(x) \right] = 0 \ .$$

The derivation for node "2", located outside the cell, is identical to the one we just performed for node "1", if we substitute $\mathrm{r_e} \to \mathrm{r_g}$, $\mathrm{i_{extracell}} \to \mathrm{i_{intracell}}$, $\Phi_\mathrm{e} \to \Phi_\mathrm{i}$, $\mathrm{i_m} \to -\mathrm{i_m}$, and $\mathrm{D_e^*} \to \mathrm{D_g^*}$, thus obtaining

$$(8.15)\qquad \frac{\partial(-\mathrm{V_m})}{\partial t}(x) - \frac{\mathrm{i_m}(x)}{\mathrm{c}} - \frac{\mathrm{i_{extracell}}(x)}{\mathrm{c}} - \frac{\partial}{\partial x} \left[\frac{\partial \Phi_\mathrm{e}}{\partial x}(x) \cdot \mathrm{D_e^*}(x) \right] = 0 \ .$$

If we concentrate on Eq. (8.14) and substitute $\Phi_\mathrm{i} = \mathrm{V_m} + \Phi_\mathrm{e}$ with the intent on producing

an equation that describes the behavior of the transmembrane potential V_m, we obtain

$$\begin{aligned}&\frac{\partial V_m}{\partial t}(x) + \frac{i_m(x)}{c} - \frac{i_{intracell}(x)}{c} \\ &- \frac{\partial}{\partial x}\left[\frac{\partial V_m}{\partial x}(x)\, D_g^*(x)\right] - \frac{\partial}{\partial x}\left[\frac{\partial \Phi_e}{\partial x}(x)\, D_g^*(x)\right] = 0\ .\end{aligned}$$

Ignoring the effects of the extracellular potential Φ_e, adding the explicit dependence on both x and t, and rearranging the terms, we have

$$\frac{\partial V_m}{\partial t}(x,t) = \frac{\partial}{\partial x}\left[\frac{\partial V_m}{\partial x}(x,t)\, D_g^*(x)\right] - \frac{i_m(x,t)}{c} + \frac{i_{intracell}(x,t)}{c}\ , \tag{8.16}$$

often called the *monodomain cable equation*, describing the AP propagation inside the cell in one dimension.

Each term in this equation is connected to parts of the system's behavior. The dynamic change in the membrane potential, $(\partial V_m/\partial t)(x,t)$, is effectively the charging of the capacitor – thinking of the membrane again as a capacitor – by the $i_m(x,t)$ current flowing via the membrane channels. One can make the connection with cell components in Fig. 2.5 in Chapter 2, where i_m could be the collective current flowing through the Na^+, Ca^{++}, and K^+ ion channels and maybe the transporters. This current has a complicated nonlinear behavior which can be defined in many ways depending on cardiac model. If, for example, we set

$$-\frac{i_m}{c} = \frac{1}{\varepsilon}\left(V_m - \frac{V_m^3}{3} - W\right)\ , \tag{8.17}$$

we have the *Fitzhugh-Nagumo*[120,121] equations, one of the simplest cardiac models:

$$\frac{\partial \mathrm{V_m}}{\partial t}(x,t) = \frac{1}{\varepsilon}\left(\mathrm{V_m} - \frac{\mathrm{V_m^3}}{3} - \mathrm{W}\right) ,$$

$$\frac{\partial \mathrm{W}}{\partial t} = \mathrm{V_m} - \beta \mathrm{W} + \gamma ,$$

where the small parameter ε is the ratio of the temporal scales between $\mathrm{V_m}$ and W, W can be considered as representing the slow ionic currents in the cell, and β and γ are parameters.

Continuing with the terms in Eq. (8.16), the current $\mathrm{i_{intracell}}(x,t)$ can be thought of as an external current applied to the cell at a given point in time and space. This concept will be useful later when we apply an external current to initiate the AP in our numerical calculations in Chapter 9.

Lastly, the term

$$\frac{\partial}{\partial x}\left[\mathrm{D_g^*}(x)\,\frac{\partial \mathrm{V_m}}{\partial x}(x,t)\right] ,$$

can be recognised as a diffusion term, with $\mathrm{D_g^*}(x)$ being a *spatially-dependent* diffusion coefficient. This term is at the center of our study. In Chapter 9, we will be changing this coefficient and observing the effects on AP propagation.

The Fitzhugh-Nagumo model is one of several electrophysiological models describing AP propagation across mammalian cardiac cells, with simplified ion currents.[122,123,124,125] More detailed ionic models include a large number of membrane currents with parameters whose value was measured in classic voltage-clamp or patch-clamp experiments, and

a larger number of gates. Complicated models include the Beeler-Reuter model[126], the Luo-Rudy model[127], and the TenTusscher-Noble-Noble-Panfilov model[128], which are based on direct experimental observations. These models are complex enough to capture some of the realistic cardiac behavior, yet they are less good in providing an essential phenomenological insight into the spatial dynamical behavior of the AP. Simpler models such as the four-variable Bueno-Cherry-Fenton model[129], and the three-variable cardiac Fenton-Karma model[130] (FK3V) provide insight with less computational burden.

We cannot solve these equations in any generic form, we may only see their behavior for a specific set of parameters by using numerical techniques; each set of parameters is connected to essential features of the model so the solutions shed light on those effects and physiological function. One-dimensional numerical simulations, being quick and efficient, enabled us to try out multiple different values for the relevant parameters and capture the changes in morphology. The property of conduction velocity being typically about two to three times faster along the length of the fiber compared to across its width, makes numerical calculations using 1D models a good first approach.

8.2 The Fenton-Karma model (1998)

Earlier in this Chapter, we derived the cable equations in a series of cells, and gave a brief review of existing cardiac models. Now, will take a closer look at the model which we will be implementing in our computational model, the three variable model by Fenton and Karma (FK) in one-dimension.

In his dissertation[131], Flavio Fenton introduced an extension to what was known as the

two-variable Karma model, namely the three-variable Fenton-Karma (FK3V) model of coupled reaction-diffusion equations on a 1-D cable of cells. We used this model in this work to produce pECG patterns relating to AP propagation. The model consists of the following three coupled PDEs,[130,122] shown below including the stimulus current, $J_{\text{stim}}(x,t)$, a rectangular pulse that initiates the AP:

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(\tilde{D}\nabla u\right) - J_{\text{fi}}(u;v) - J_{\text{so}}(u) - J_{\text{si}}(u;w) + J_{\text{stim}}(x,t)\ , \tag{8.18}$$

$$\frac{\partial v}{\partial t} = \Theta(u_c - u)\frac{1-v}{\tau_{\text{v}}^{-}(u)} - \Theta(u - u_{\text{c}})\frac{v}{\tau_{\text{v}}^{+}}\ , \tag{8.19}$$

$$\frac{\partial w}{\partial t} = \Theta(u_{\text{c}} - u)\frac{1-w}{\tau_{\text{w}}^{-}} - \Theta(u - u_c)\frac{w}{\tau_{\text{w}}^{+}}\ , \tag{8.20}$$

where, the transmembrane voltage $0 < u(x,t) < 1$, normalized to a dimensionless property, is defined to be $u \equiv (V - V_0)/(V_{\text{fi}} - V_0)$ where V is the un-normalized transmembrane potential, V_0 is the resting membrane potential, and V_{fi} is the Nernst potential of the fast inward current. The normalized threshold potential is given by u_{c}. The permeability of the channels in the cell membrane is regulated by the two gating variables $v(x,t)$ and $w(x,t)$. Gate state indicates whether ions can pass or not. Variable $v(x,t)$ denotes the fast inactivation gate which opens when the cell is not excited, and closes when it becomes excited. The closing time constant τ_v^+ corresponds to cell depolarization, and the opening time constant τ_v^- to cell repolarization. The $u-$dependent parameter $\tau_v^-(u)$ is given by

$$\tau_{\text{v}}^{-}(u) = \Theta(u - u_{\text{v}})\tau_{\text{v1}}^{-} + \Theta(u_{\text{v}} - u)\tau_{\text{v2}}^{-}. \tag{8.21}$$

This splitting allows the minimum diastolic interval, i.e., the excitable gap, controlled by τ_{v1}^{-}, to vary independently from the steepness of this curve, controlled by τ_{v2}^{-}. The voltage threshold $u_v < u_c$ controls the splitting. Variable w is the probability of a gate opening as described in the Hodgkin-Huxley model; τ_w^{+} and τ_w^{-} are the time constants for closing and opening of the gate, respectively. As a preview to the detailed analysis of the predictions of this numerical model, we show the typical behavior of the basic quantities, u, v, and w, with u being the voltage from which the pseudo-ECG can be constructed.

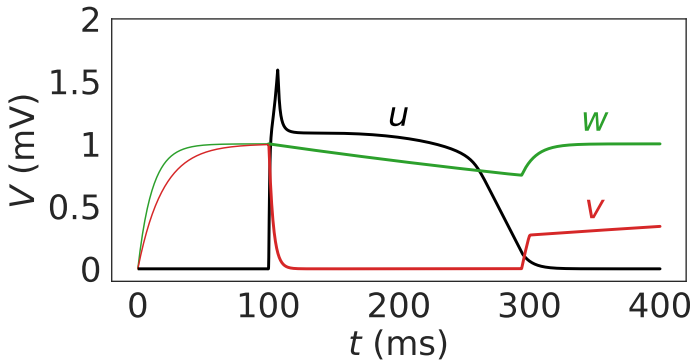

Fig. 8.3: AP as a function of time for a 1-D cable and for a 400 ms duration. Shown are the three variables of the FK3V model, the normalised membrane voltage u, and the gating variables v and w.

The scaled phenomenological ionic currents J_{fi}, J_{so}, and J_{si} (f meaning fast, and s slow), are related to the corresponding currents in units of mA through

$$J_{\text{c}} = \frac{I_{\text{c}}}{C_{\text{m}}(V_{\text{c}} - V_0)}, \tag{8.22}$$

where C_{m} is the membrane capacitance, and c can be either fi, or so, or si, where:

(a) J_{fi} corresponds to the fast inward sodium (Na^{+}) current, responsible for the depo-

larization of the membrane, and depending on the gating variable v. This gating variable is responsible for inactivation of the current after depolarization, and its reactivation after repolarization.

(b) J_{so} is a slow outward current analogous to the time-independent potassium (K^+) current; it is responsible for re-polarization of the cell membrane.

(c) J_{si} is a slow inward current, corresponding to the calcium (Ca^+) current, that balances I_{so} during the plateau phase of the action potential; this current depends on one gate variable w, responsible for its inactivation and reactivation. The above correspondence to the Na, K, and Ca currents, is an oversimplification, due to membrane dynamics being a lot more complex. The model, though, succeeds in capturing the minimal ionic complexity that underlies the membrane recovery processes. All currents are considered normalized. The expressions for the normalized currents read

$$J_{\mathrm{fi}}(u;v) = -\frac{v}{\tau_{\mathrm{d}}}\Theta(u-u_{\mathrm{c}})(1-u)(u-u_{\mathrm{c}}), \tag{8.23a}$$

$$J_{so}(u) = +\frac{u}{\tau_0}\Theta(u_{\mathrm{c}}-u) + \frac{1}{\tau_{textr}}\Theta(u-u_{\mathrm{c}}), \tag{8.23b}$$

$$J_{si}(u;w) = -\frac{w}{2\tau_{\mathrm{si}}}\left\{1+\tanh\left[k\left(u-u_{\mathrm{c}}^{\mathrm{si}}\right)\right]\right\}, \tag{8.23c}$$

where

$$\tau_{\mathrm{d}} = \frac{C_{\mathrm{m}}}{\bar{g}_{\mathrm{fi}}}. \tag{8.24}$$

The values of the parameters $\bar{g}_{\text{fi}}$, τ_0, τ_{r}, τ_{si}, k, and u_{c}^{si} are also given in Table I. The function $\Theta = \Theta(x)$, which appears repeatedly in Eqs (8.18) – (8.20) and Eqs (8.23a) – (8.23c), is the standard Heaviside step function defined by $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$. In this work, the modified Beeler-Reuter (MBR) model papameters are used. For illustrations of the AP, there is indication of whether the normalized or un-normalized quantities are depicted. Note: the parenthesis next to the symbol Θ, i.e., $\Theta(u - u_{\text{v}})$, is not a multiplicand but the argument of the function.

From Eq. (8.18) we can see that modeling the propagation of electrical impulses in cardiac tissue is affected by two distinct terms. The first term of the right hand side, includes the diffusion coefficient and encompasses the passive characteristics of the medium, such as its microscopic structure and cell-to-cell coupling via ion conducting gap junctions[133]. The second term, the sum of the ionic currents through the membrane channels (excluding the stimulus current), denotes the dynamic characteristic of the medium.

8.3 The Role of the Diffusion Coefficient

From the cable equation analysis, the effective voltage diffusion coefficient for homogeneous (healthy) tissue is given by

$$D_0 = \frac{1}{C_m \rho S_u}\,, \tag{8.25}$$

where C_m is the cell membrane capacitance, ρ is the longitudinal resistivity (attributed to the gap junctions), and S_u is the surface-to-volume ratio for the cell. The values of the pa-

Parameter	BR model	MBR model	MLR-I model
$\bar{g}_{\mathrm{fi}}$	4	4	5.8
τ_{r}	33.33	50	130
τ_{si}	29	44.84	127
τ_0	12.5	8.3	12.5
τ_{v}^{+}	3.33	3.33	10
τ_{v1}^{-}	1250	1000	18.2
τ_{v2}^{-}	19.6	19.2	18.2
τ_{w}^{+}	870	667	1020
τ_{w}^{-}	41	11	80
u_{c}	0.13	0.13	0.13
u_{v}	0.04	0.055	–
$u_{\mathrm{c}}^{\mathrm{si}}$	0.85	0.85	0.85
Other parameters			
C_{m}	$1\,\mu\mathrm{F/cm}^2$		
V_0	$-85\,mV$		
V_{fi}	$+15\,mV$		
k	10		

Table 8.1: Three different sets of model parameters that enter into the three-variable Fenton-Karma model (Eqs. (8.18) - (8.20)) from Ref. [132]. In this work, the modified Beeler-Reuter (MBR) model papameters are used.

rameters $C_m = 1\ \mu\mathrm{Fcm}^{-2}$, $\rho = 0.4\ \mathrm{k\Omega cm}$ (experimentally measured [129]), and $S_u = 5000\ \mathrm{cm}^{-1}$ provided in Table 8.1 for human ventricular cells, give the typical value $D_0 = 0.0012 \pm 0.0002\ \mathrm{cm}^2\mathrm{ms}^{-1}$ used often in literature. In this work we use various values for the diffusion coefficient, and by doing so, we can model various sub-cellular characteristics of cardiac electrical propagation, such as different gap junction resistance and cell membrane capacitance, and by extension, study conduction problems in the heart. My-

ocardial tissue is, of course, a very complex structure, and we hope to capture only a part of its behavior.

8.4 Summary

A surface EKG is recorded using two electrodes, an anode and a cathode, placed on the skin surface, far away from the heart itself. Therefore, the field of view of an EKG is the whole heart (Fig. 8.4 A). In contrast, the field of view of an activation recording using a catheter is much smaller, as the electrodes are closely spaced (2 – 3mm apart). The latter is called an electrogram (EGM) (Fig. 8.4 B). Image (A) is an outside view of the heart, whereas image (B) could be considered an inside view of the heart.

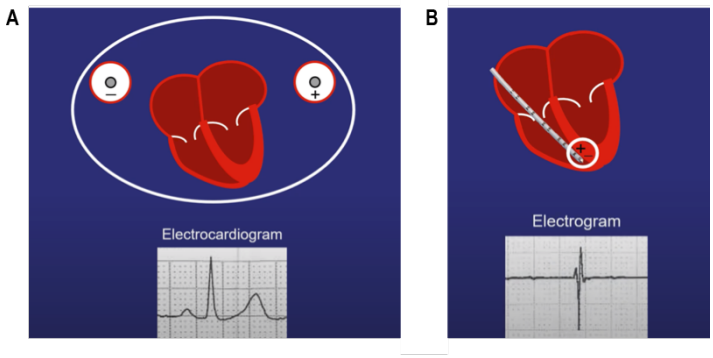

Fig. 8.4: The field of view of an EKG is the whole heart (**A**). In the plot of the lower portion of the image, one can clearly see the surface P wave, showing the atrial activity, the surface QRS complex, showing the ventricular depolarization, and lastly, the T wave, showing the ventricular repolarization. The field of view of an EGM is restrained to a local area of the heart (**B**). With a catheter placed in a small area of the left ventricular myocardium, we do not see any atrial activity or subsequent T wave.
Image adapted from a video by Dr. Joshua Cooper.[134] ECG: electrocardiogram, EGM: electrogram.

9

Numerical solution of the FK cardiac model

Differential equations are effective for describing various scientific and engineering problems depicting complex processes in natural and social sciences; examples include the FK3V, and dynamical systems describing climate modelling[135] or financial issues.[136] Solving them efficiently has been a long-standing challenge. Only a few simple ODEs or PDEs are solved analytically. For more complex PDEs, only numerical methods, such as

the Finite Difference Method (FDM) and the Finite Element Method (FEM), can be applied.

9.1 Numerical calculations

The diffusion coefficient plays an important role in the propagation of the AP. We study the role of the both when it is constant, and when it is allowed to vary spatially.

All numerical simulations for Eqs (8.18) – (8.20) along with Eqs (8.23a) – (8.23c), were performed on a theoretical 1D cable of cells, using the fourth order Runge-Kutta (RK4) algorithm with fixed timestep of $dt = 0.002$ ms. For spatial discretization of Eqs (8.18) – (8.20), the domain was divided into $N_x - 1$ elements with $N_x = 400$ at $x_i = (i-1)L/(N_x-1)$ ($i = 1, 2, ..., N_x$) which are separated by distance $dx = L/(N_x - 1) = 0.0075$ cm (about the length of a cardiac cell). Second order, centered finite difference formulas were used to discretize the first and second derivatives of the state variables wherever they appear in Eqs (8.18) – (8.20). The spatially discretized equations are given explicitly in the next subsection. The diffusion coefficient $\tilde{D}$ is, for some parts of the modeling, considered constant, and for some others, spatially dependent (inhomogeneous), $\tilde{D}(x)$, and is modeled as a double step-function controlled by two very steep `tanh` functions. Numerical calculations were implemented with `Python` code, and, unless otherwise specified, they were run for $150,000$ timesteps.

The boundary conditions at the ends of the cable are chosen to be those of zero-flux

(Neumann) type, i.e.,

$$\tilde{D}(x)\left.\frac{\partial u(x,t)}{\partial x}\right|_{x=0} = \tilde{D}(x)\left.\frac{\partial u(x,t)}{\partial x}\right|_{x=L} = 0, \tag{9.1}$$

where L is the length of the cable which, in what follows, is set everywhere equal to 3.0 cm. As explained above, the (inhomogeneous) diffusion coefficient $\tilde{D}(x)$ is practically a piecewise constant function which assumes the value D_0 and D_{scar} in the normal (healthy) and defected (scar) tissue region, respectively, as it is shown schematically in Fig. 9.5.

For a homogeneous (spatially constant) diffusion coefficient $\tilde{D}(x) = \tilde{D}$ along the cable, Eq. (8.18) becomes

$$\frac{\partial u}{\partial t} = \tilde{D} \cdot \frac{\partial^2 u}{\partial x^2} - J_{\text{fi}} - J_{\text{so}} - I_{\text{si}} + J_{\text{stim}} \ . \tag{9.2}$$

If the diffusion coefficient $\tilde{D}$ is allowed to depend on the spatial coordinate x, the first term on the right hand side of Eq. (8.18), i.e., $\nabla\left(\tilde{D}\nabla u\right)$ becomes

$$\frac{\partial}{\partial x}\left[\tilde{D}(x)\frac{\partial u(x)}{\partial x}\right] = \frac{\partial \tilde{D}(x)}{\partial x}\frac{\partial u(x,t)}{\partial x} + \tilde{D}(x)\frac{\partial^2 u(x,t)}{\partial x^2} \ . \tag{9.3}$$

From Eq. (9.3) we can see that we need the discrete form of the first and second spatial derivative of $u(x,t)$, as well as the first spatial derivative of $\tilde{D}(x)$. We use the following cen-

tered differences[137]

$$\frac{\partial u(x,t)}{\partial x} = \frac{u_{i+1}(t) - u_{i-1}(t)}{2\Delta x}, \tag{9.4}$$

$$\frac{\tilde{D}(x)}{\partial x} = \frac{\tilde{D}_{i+1} - \tilde{D}_{i-1}}{2\Delta x}, \tag{9.5}$$

$$\frac{\partial^2 u(x,t)}{\partial x^2} = \frac{u_{i+1}(t) - 2u_i(t) + u_{i-1}(t)}{\Delta x^2}\,. \tag{9.6}$$

Using Eqs (9.3) – (9.6), the spatially discretized system of Eqs (8.18) – (8.20) reads

$$\begin{aligned}\frac{\partial u_i}{\partial t} = \frac{\tilde{D}_{i+1} - \tilde{D}_{i-1}}{2\Delta x}\frac{u_{i+1} - u_{i-1}}{2\Delta x} + \tilde{D}(x_i)\frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} \\ -J_{\text{fi}}(u_i; v_i) - J_{\text{so}}(u_i) - J_{\text{si}}(u_i; w_i) + J_{\text{stim}}(x_i, t),\end{aligned} \tag{9.7}$$

$$\frac{\partial v_i}{\partial t} = \Theta(u_c - u_i)\frac{1 - v_i}{\tau_v^-(u_i)} - \Theta(u_i - u_c)\frac{v_i}{\tau_v^+}, \tag{9.8}$$

$$\frac{\partial w_i}{\partial t} = \Theta(u_c - u_i)\frac{1 - w_i}{\tau_w^-} - \Theta(u_i - u_c)\frac{w_i}{\tau_w^+}, \tag{9.9}$$

where it is implied that the discretized variables u_i, v_i, and w_i depend on time t.

We notice from the differential equations that the gating variables v and w can be solved analytically up to the point when the threshold is reached and the cell fires. We take advantage of that by including these equations our calculations.

$$v = (1 - e^{-t/\tau_{\text{v2}}^-}) \tag{9.10}$$

$$w = (1 - e^{-t/\tau_{\text{w}}^-}) \tag{9.11}$$

This is also verified from visually inspecting the plots for the solutions to these equations (see previous Chapter Fig. 8.3).

The stimulus current, which is necessary for the excitation of the AP pulse, is assumed to arise from physiological mechanisms of the heart. There is a large volume of works on the calculation of the ventricular AP in 1D[138,139,140,141] using various types of stimulus current functions $J_{\text{stim}}(x,t)$. The FK3V model has been also used for the calculation of the ventricular AP in two and three dimensions[142]. Also, mapping models have been used for the analysis of numerical results obtained through the FK3V model[143].

9.2 Propagation of the AP

As mentioned above, most of the research on propagation disruptions concentrates on the remodeling of ionic currents. We chose to concentrate on changing the diffusion coefficient and studying the effect on the pseudo-ECG. We chose to concentrate on varying the profiles of the voltage diffusion coefficient and consequently studying their effect on the calculated pseudo-ECG, regarding thus the suppression of electrical connection between cells as the primary cause of cardiac pathology.

Our theoretical cable of cells, of length $L = 3$ cm, is composed of 400 ventricular cells of a single cell type, connected via gap junctions. A stimulus current $J_{\text{stim}}(x,t)$ with an above-threshold amplitude is applied to the first cell at $x = 0$ and is spatially restricted to the first 15 cells. It is therefore assumed to excite a small region around the left end of the cable of length $L_{\text{exc}} = 15 \times$ (cell length). We take the cardiac cell length to be equal to the spatial discretization $dx = 0.0075$ cm. The stimulus current is taken to be a rectangular pulse of,

unless otherwise stated, amplitude $J_{\text{amp}} = 0.9$ mA and duration $\tau_{\text{p}} = 11$ ms.

Using Eq. (9.7) – (9.9), we have calculated numerically the ventricular AP propagating through ventricular tissue of length $L = 3$ cm as a function of time t. A small segment of the tissue/cable of length $L_{\text{exc}} = 0.11$ cm is initially excited through its left end, i.e., the segment from $x = 0$ to $x = L_{\text{exc}} = 0.11$ cm, using stimulus currents of amplitude $J_{\text{amp}} = 5$ mA and different durations τ_{p}. Typical AP pulse profiles (black curves) along with the associated stimulus currents (red curves) are shown in Fig. 9.1, monitored at two different locations on the cable, i.e., at $x \simeq 0.26$ cm (relatively close to the excited region, left panels) and $x \simeq 0.75$ cm (at one-fourth of the cable length as measured from $x = 0$, right panels). As it can be observed, the amplitude of the AP as well as its duration (action potential duration, APD) increases with increasing τ_p (from top to bottom). The latter, specifically, which is defined as the width of the pulse at 12% of its maximum amplitude (illustrated in (a) by the gray horizontal double-headed arrow), increases from 159 ms for $\tau_{\text{p}} = 1.0$ ms, to 194.4 ms for $\tau_{\text{p}} = 9.0$ ms, to 215.8 ms for $\tau_{\text{p}} = 9.0$ ms. Left and right panels, obtained by monitoring the AP pulses at different locations on the cable, also differ in that the former exhibit a sharp peak at a time instant corresponding to the end of the stimulus current pulse. This sharp peak decreases until it practically vanishes for locations on the cable relatively far from the excited region.

Similarly, in Fig. 9.2, the calculated action potential V as a function of time t is monitored at two different positions on the cable for stimulus currents of amplitude $J_{amp} = 4$ mA, duration $\tau_p = 4$ ms, and three different values of the initially excited segment at the left end of the cable of length L_{exc}, which extends from $x = 0$ to $x = L_{\text{exc}}$. As in Fig. 9.1,

the action potential is monitored at $x \simeq 0.25$ cm and $x \simeq 0.75$ cm (left and right panels, respectively). Again it is observed that, the duration of the action potential (APD) increases with increasing L_{exc}. Specifically, the APD increases from 173.9 ms for $L_{\text{exc}} = 0.075$ cm to 174.9 ms for $L_{\text{exc}} = 0.113$ cm, to 179.5 for $L_{\text{exc}} = 0.113$ cm. In both Figs. 9.1 and 9.2, the action potential exhibits the right characteristics in (e) and (f) panels, as long as the shape and the width (i.e., the APD) is concerned.

9.3 The pseudo-ECG

The analysis and interpretation of ECGs remains mostly empirical. The pseudo-ECG at a particular time-instant t is calculated numerically from the spatial profile of the AP at that time-instant on the cable using the expression [144,145,146,147,129] (for a thorough derivation see Ref. [148])

$$\Phi_e(\mathbf{x}^\star, t) = -K \int \nabla V(\mathbf{x}, t) \cdot \nabla \frac{1}{|\mathbf{x}^\star - \mathbf{x}|} d\mathbf{x}, \tag{9.12}$$

where $\nabla V(\mathbf{x}, t)$ is the spatial gradient of the ventricular AP, $K = 1.89$ mm^2 is a constant that depends on electrophysiological quantities, such as the radius of the fiber and the intracellular conductivity. The "electrode" measuring the voltage is at point $\mathbf{x}^\star$ of the fiber, and $|\mathbf{x}^\star - \mathbf{x}|$ is the distance from a source point $\mathbf{x}$ to a field point $\mathbf{x}^\star$, where $\mathbf{x}^\star > \mathbf{x}, \forall \mathbf{x}$. The temporal profile of the pseudo-ECG Φ_e constitutes an approximation for the ventricular component of the ECG, i.e., the pseudo-ECG generated at a hypothetical electrode which is located at a particular distance away from the last epicardial cell along the cable. As

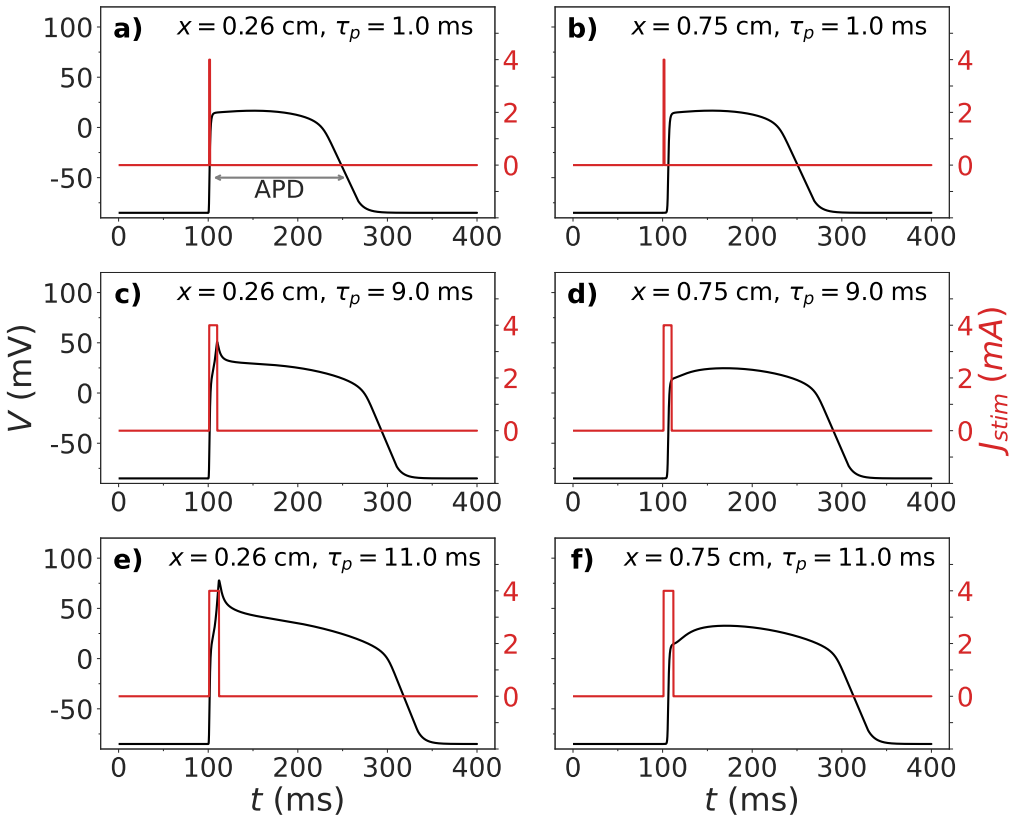

Fig. 9.1: The action potential V (black curves) in mV as a function of time t that is produced by a single rectangular current pulse depicted by the red curve (notice the small deflection in the beginning) for $D = 0.005$ cm^{2}ms^{-1}, $L_{\text{exc}} = 0.113$ cm, $J_{\text{amp}} = 4.0$ mA. The action potential duration (APD), indicated by the gray horizontal double-headed arrow, is measured for each plot. (a,b) $\tau_{\text{p}} = 1.0$ ms, and APD$= 159$ ms; (c,d) $\tau_{\text{p}} = 9.0$ ms, and APD$= 194.4$ ms; (e,f) $\tau_{\text{p}} = 11.0$ ms, and APD$= 215.8$ ms;. The action potential is monitored at positions $x \simeq 0.26$ cm (right panels), and at $x \simeq 0.75$ cm (left panels).

shown in Fig. (2.9), the ventricular potential contributes specifically to the formation of the QRS cluster and the T wave. The pseudo-ECG is, thus, expected to reproduce these features.

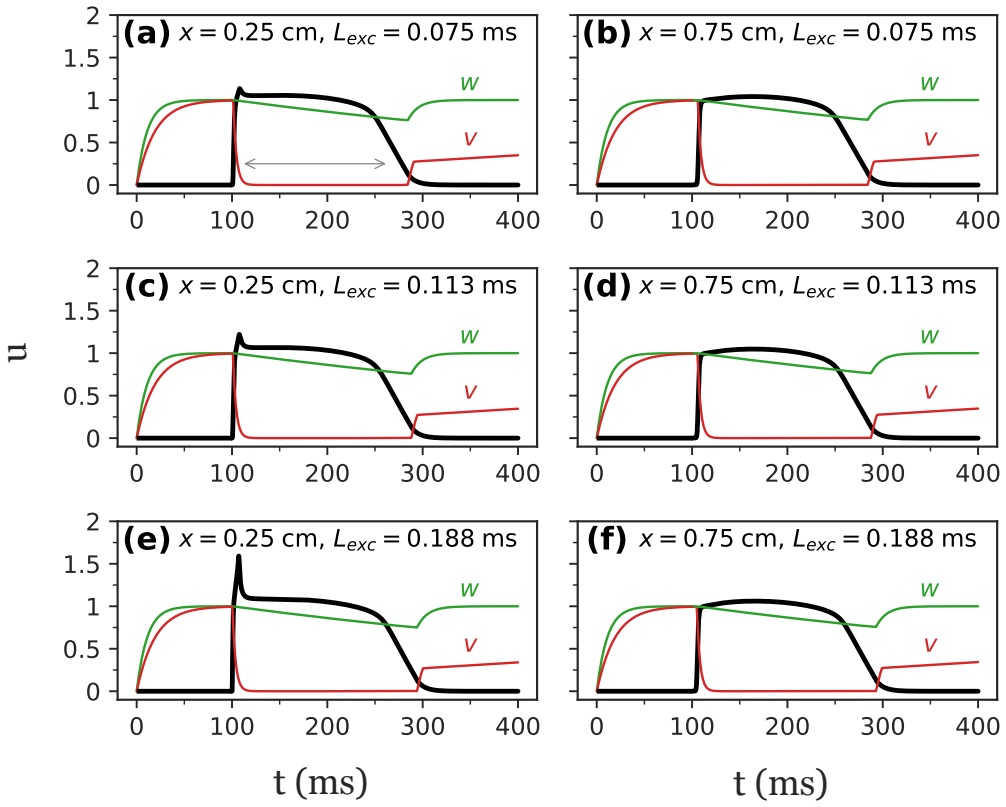

Fig. 9.2: The action potential u (black curves) in normalized units as a function of time t excited by a single rectangular pulse current with amplitude $J_{\text{amp}} = 4$ mA, and duration $\tau_{\text{p}} = 7$ ms. The action potential duration (APD), indicated by the gray horizontal double-headed arrow on the top left plot, is measured for each plot. (a,b) $L_{\text{exc}} = 0.075$ cm, and APD$= 173.9$ms; (c,d) $L_{\text{exc}} = 0.113$ cm, and APD$= 174.9$ms; (e,f) $L_{\text{exc}} = 0.188$ cm, and APD$= 179.5$ms. The action potential u is monitored at $x \simeq 0.25$ cm (right panels), and at $x \simeq 0.75$ cm (left panels).

In one dimension, Eq. (9.12) reads

$$\Phi_e(x^\star, t) = -K \int \frac{\partial V(x,t)}{\partial x} \left(\frac{\partial}{\partial x} \frac{1}{|x^\star - x|} \right) dx \,. \tag{9.13}$$

We calculate the pseudo-ECG for point outside the cell cable, so for $x^\star > L$, Eq. (9.13)

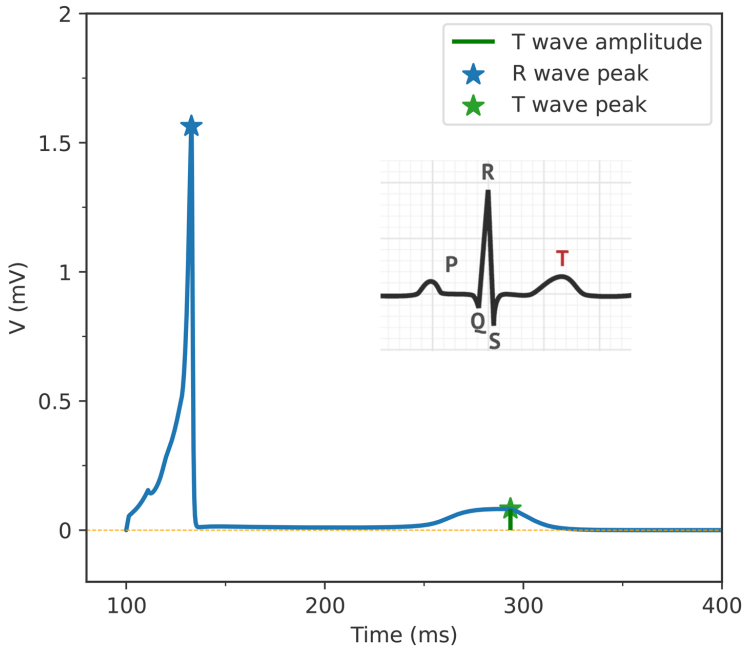

Fig. 9.3: Simulated pseudo-ECG as a function of time t calculated using the three-variable Fenton-Karma model . For comparison, a drawing of a real ECG is shown in the inset. We can clearly detect the R and T wave equivalents whose amplitudes we designate with the blue and green stars respectively.

becomes

$$\Phi_e(x^\star, t) = -K \int \frac{\partial V(x,t)}{\partial x} \frac{1}{(x^\star - x)^2}\, dx, \tag{9.14}$$

and is thus more easily calculated. Using the spatial profiles calculated from Eq. (9.7) - (9.9) at each time instant, we calculate $\Phi_e(L, t)$ which is the desired pseudo-ECG; in our calculations $x^\star = 3.37$ cm, while the cell cable length is $L = 3.0$ cm. As shown in Fig. 9.3, the T-wave has positive polarity and its amplitude is defined as the vertical distance

from $V = 0$. In general, T waves are considered positive when their deflection is upward, and negative when it is downward. For *biphasic T-waves* (waves with both an upward and a downward deflection), unless otherwise stated, the dominant deflection is chosen. In the small inset, a surface ECG which is recorded using two electrodes placed on the skin surface, away from the heart, is visually compared to the pseudo-ECG.

9.4 Constant Diffusion Coefficient

We first run our model with a spatially constant diffusion coefficient $\tilde{D} = D_0$. This can be regarded as an effective parameter, a mean value to account for the discontinuity defect part inserts. The height of the R-wave (see blue star in Fig. 9.3) in each pseudo-ECG denotes the value of the T-wave amplitude in units of mV; when plotted for 25 different values, as shown in Fig. 9.4, it appears to exhibit an exponential dependence on the effective diffusion coefficient $\tilde{D}$. Since repeated attempts to fit a single exponential curve using least squares failed, we tried using the sum of two exponentials via the *ansatz*

$$R_{\text{wa}} = Ae^{-\alpha\tilde{D}} + Be^{-\beta\tilde{D}} + C, \tag{9.15}$$

where R_{wa} is the R-wave amplitude, and A, α, B, β, and C, are parameters to be fitted. Using the ansatz (9.15), we obtained excellent fit using parameters $A = 0.79, \alpha = 115.36$, $B = 0.56, \beta = 1258.84$, and $C = 0.595$. In the exemplary fit shown by the blue line in Fig. 9.4, we notice a transition region around the value of 0.0012 ± 0.0002 cm^{2}ms^{-1} of the diffusion coefficient, which, as mentioned before, is an experimental value used frequently

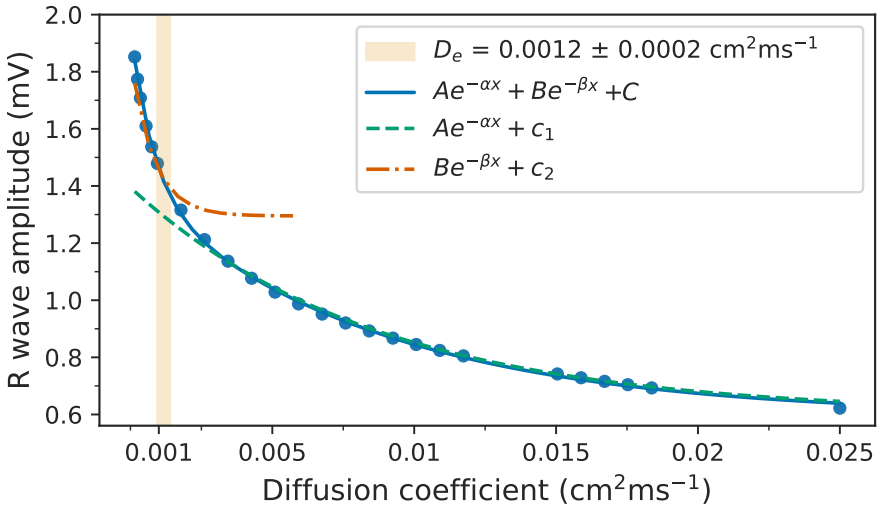

Fig. 9.4: R-wave amplitude as a function of the (homogeneous, spatially constant) voltage diffusion coefficient $\tilde{D} = D_0$, extracted from the calculated pseudo-ECGs using Eq. (9.13) for a total of 25 values for D_0 within the interval $0.0005 - 0.025$ cm^{2}ms^{-1} (blue curve). The root-mean-squared-error (RMSE) for this curve is $7.6 \cdot 10^{-03}$ mV, which is low enough to provide confidence in the solution. Other simulation parameters are: $J_{\text{amp}} = 10.0$ mA, $t_{\text{p}} = 10.0$ ms, $L_{\text{exc}} = 0.105$ cm, dx = 0.0075 cm, dt = 0.0013 ms, Nsteps = 230769. Curve fit parameters are: A = 0.79, α = 115.36, B = 0.56, β = 1258.84, and C = 0.595. The range of values D_{e} = (0.0012 $\pm$ 0.0002) cm^{2}ms^{-1} denoted by the shaded area, is a range of experimental values used frequently in the literature. The constants for the failed single exponentials are $c_1 = 0.6$ (green curve), and $c_2 = 1.3$ (red curve).

in literature. The transition region is identified by those values of $\tilde{D} = D_0$ for which the fitted single-exponential curves $Ae^{-\alpha\tilde{D}} + C_1$ (green-dashed curve) and $B\ e^{-\beta\tilde{D}} + C_2$ (red dashed-dotted curve) start diverging significantly from the numerical data (slightly above the experimental value of $\tilde{D} = D_e$. The algorithm was implemented using the `SciPy` python library.

9.5 Spatially-Dependent Diffusion Coefficient

In Fig. 9.5, we plot the profile of the diffusion coefficient $\tilde{D}(x)$ that contains a localized heterogeneity in the form of a defected (scar) region in which the conductance velocity has been significantly reduced due to reduced electrical connection between cells, i.e., a region in which the value of $\tilde{D}(x)$ has dropped to

$$D_{\text{scar}} = (1 + \lambda)\, D_0, \tag{9.16}$$

where D_0 is the value of $\tilde{D}(x)$ in the normal (healthy) region, and $-1 < \lambda < 0$. For example, for a cable length of $L = 3.0$ cm with $D_0 = 0.005$ cm^{2}ms^{-1}, a defected region of length $L_{\text{scar}} = 0.5$ cm and $\lambda = -0.8$ would have $D_{\text{scar}} = 0.001$ cm^{2}ms^{-1}. Hence, the spatially dependent voltage diffusion coefficient has the form

$$\tilde{D}(x) = \begin{cases} D_{scar}, & \text{if } x_{scar} < x < x_{scar} + L_{scar} \\ D_0, & \text{anywhere else.} \end{cases} \tag{9.17}$$

Note that in a recent work[149], a spatially and temporally diffusion coefficient was considered which encompasses conductance heterogeneities in the cardiac tissue induced by the dynamics of the gap junctions. Obviously, the adjustable parameters of the voltage diffusion profile is the starting point of the defected region x_{scar}, the spatial length of the scar tissue L_{scar}, and the percentage decrease λ which lowers $\tilde{D}(x)$ in the defected region. Using

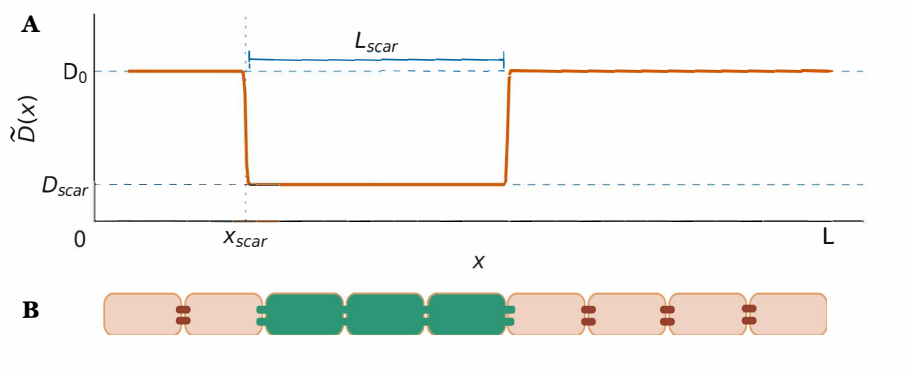

Fig. 9.5: A. Profile of the diffusion coefficient $\tilde{D}(x)$, where $\tilde{D}(x) > 0$, plotted against the position x on the cell strand ($0 < x < L$, where L is the length of the tissue). The plot depicts a localized defect of width L_{scar} and coefficient value D_{scar}, around the effective D_0 (normal tissue). Marking the start of the defected region (scar) is x_{scar}. **B**. Rendering of the corresponding cell strand for visualization purposes (not in scale). The green region corresponds to the scar tissue, while the rest of the strand is composed of normal tissue. Figure created with `matplotlib` using the following values: $D_0 = 0.005$ cm^{2}ms^{-1}, $L_{\text{scar}} = 1.125$ cm, $x_{\text{scar}} = 0.5$ cm, $\lambda = -0.8$, and $L = 3.0$ cm, $D_{\text{scar}} = 0.002$ mV (as derived from Eq. (9.16)).

Eqs (9.16) and (9.17) above, the spatially averaged diffusion coefficient is

$$< \tilde{D}(x) > \; = D_0 \left(1 + \lambda \frac{L_{\text{scar}}}{L}\right), \quad \text{where } \lambda < 0\,, \tag{9.18}$$

To study the polarization of the T-wave in a tissue containing a localized defect, a stimulus current in the form of a rectangular pulse of amplitude $J_{\text{amp}} = 0.9$ mA, and duration $\tau_{\text{p}} = 11$ ms, was applied at the first 15 cells of the cable, those which are at its left end ($x = 0$), whose length is $L_{\text{exc}} \simeq 0.011$ cm. Then, the pseudo-ECG is calculated from the spatio-temporal profile of the APs, and the maximum magnitude of the T-wave is identified. This procedure was repeated as a function of the width of the defect L_{scar} for three different values of the position of the onset of the defect x_{scar} and four values of the pa-

rameter λ. The results are presented in a compact way in Fig. 9.8. In all four subfigures, the diffusion coefficient in the healthy region is $D_0 = 0.005$ cm^2 ms^{-1}. The defect was modeled using a spatially dependent diffusion coefficient $\tilde{D}(x)$, whose characteristics were previously depicted in Fig. 9.5. For the results presented in Fig. 9.8, the defected region spans the interval from $x = x_{\text{scar}}$ to $x = x_{\text{scar}} + L_{\text{scar}}$. Within this interval, the diffusion coefficient is $\tilde{D} = L_{\text{scar}}$, with $\lambda = -0.8, -0.7, -0.6$, and -0.5 in Fig. 9.8(a), (b), (c), and (d), respectively. Obviously, the relation $x_{\text{scar}} + L_{\text{scar}} < L$ should hold in any case.

By inspection of Fig. 9.8 we observe that the curves for $x_{\text{scar}} = x_3 = 2$ cm (green curves) always remain on the positive side of the vertical axis, meaning that in this case there is no polarization inversion of the corresponding T-wave, and thus this is always positive. The same holds true for any other value of $x_{\text{scar}} > 2$, since the defected region constitutes only a relatively small part of the cable, which is of length $L = 3$ cm, so that it cannot affect significantly the spatio-temporal AP profile. It can be also be observed from Fig. 9.8(d) that all three curves remain on the positive sides of the vertical axis, and thus no T-wave inversion appears, due to the relatively small magnitude of λ. Indeed, the magnitude of λ in this case does not seem to be sufficiently high (or equivalently the defect is not sufficiently deep) to invert T waves. For slightly deeper defect, for $\lambda = -0.6$, as shown in Fig. 9.8(c), T-wave inversion is observed for $x_{\text{scar}} = x_0 = 0.5$ and $L_{\text{scar}} = 2.25$ (blue curve) but not for $x_{\text{scar}} = x_1 = 1.0$ or $x_{\text{scar}} = x_2 = 2.0$ (orange and green curves, respectively). The obvious reason is that in the latter cases the width of the defected region L_{scar} cannot reach such a high value as that in the former case ($L_{\text{scar}} = 2.25$). Moreover, as it can be observed from Figs. 9.8(a) and (b) the parts of the curves with $x_{\text{scar}} = x_0 = 0.5$ (blue curves) and

$x_{\text{scar}} = x_1 = 1.0$ (orange curves), respectively, with inverted (negative) T-wave become larger with decreasing λ. From these observations can thus be concluded that for fixed x_{scar}, deep and wide defected regions favor T-wave inversion. Furthermore, the value of L_{scar} at the transition from positive to negative T waves is lower in the orange curves ($x_{\text{scar}} = x_1 = 1.0$) than that in the blue curves ($x_{\text{scar}} = x_0 = 0.5$) as can be observed from Figs. 9.8(a) and (b). Thus, for fixed λ, defects with higher x_{scar} are capable to invert T-waves with lower L_{scar}. From the above remarks it becomes clear that the width, the depth, and the starting position of the defect contribute decisively to T-wave morphology.

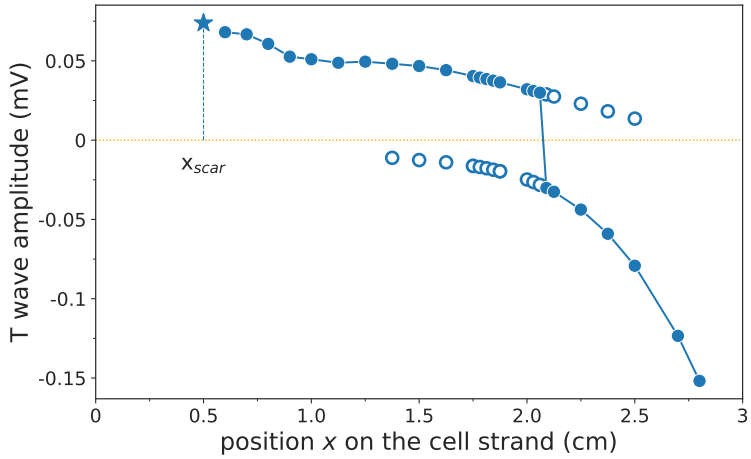

Fig. 9.6: T-wave maximae and minimae as a function of the width of the defected region L_{scar}, as obtained from the pseudo-ECG with inhomogeneous voltage diffusion coefficient $\tilde{D}(x)$ with $D_0 = 0.005$ cm^{2}ms^{-1} and $\lambda = -0.8$. The defected regions start at $x_{\text{scar}} = 0.5$ cm. The blue solid curve is a guide to the eye. The representation is the same as that in Fig. 9.8. The points depicted as blue solid circles are the ones chosen as the largest of the two in the biphasic wave. The hollow white circles show the amplitude of the other wave in the biphasic phase. The truly biphasic phase is limited to a few points around the blue vertical segment at $L_{\text{scar}} \simeq 1.6$. The purpose of this plot is to show that there is a transition phase during T inversion, where the wave has both positive (upward) and negative (downward) parts.

We should note that the transition from positive to negative T waves is realized through

a biphasic stage, with a minimum and a maximum of similar magnitude. This is consistent with the bibliography where it is reported that biphasic T waves usually evolve and are often followed by T-wave inversion with strongly suspected myocardial ischaemia [150]. There was no attempt made to trace the biphasic stage in Fig. 9.8, which is actually limited within a small interval around the transition point. Wherever two extremae appear in the calculated T-wave, only the higher of them is plotted. However, a typical biphasic stage of the calculated T-wave is illustrated below.

In Fig. 9.6, the maximum and the minimum of the T-wave (equivalently the maximum and the second maximum of the magnitude of the T-wave) are plotted as a function of the width of the defected region L_{scar}, for the parameters of the blue curve in Fig. 9.8(a). Recall that all points on that curve were obtained from the pseudo-ECG using voltage diffusion coefficient $\tilde{D}$ with $D_0 = 0.005$ cm^{2}ms^{-1}, $\lambda = -0.8$, and a defected region starting at $x_{\text{scar}} = x_0 = 0.5$ cm. The blue circles (filled and empty) have been obtained through numerical calculations while the (blue) solid curve is a guide to the eye, actually indicating the transition from positive to negative (inverted) T-wave. The filled and empty circles indicate maximae and second maximae (whenever they exist) of the T-wave magnitude. Note that the truly biphasic stage, for which the minimum and the maximum of the T-wave have approximately equal magnitude, is limited to a few (~ 5) points around the blue vertical segment indicating the T-wave inversion transition. Further away from that segment, e.g., at $x = 1.5$ cm, i.e., at $L_{\text{scar}} = 1.0$, the maximum of the T-wave has much larger magnitude of the minimum, and thus the positive character of the T-wave is dominant. Such cases are regarded as positive T-waves in Fig. 9.8. Correspondingly, cases in which the negative

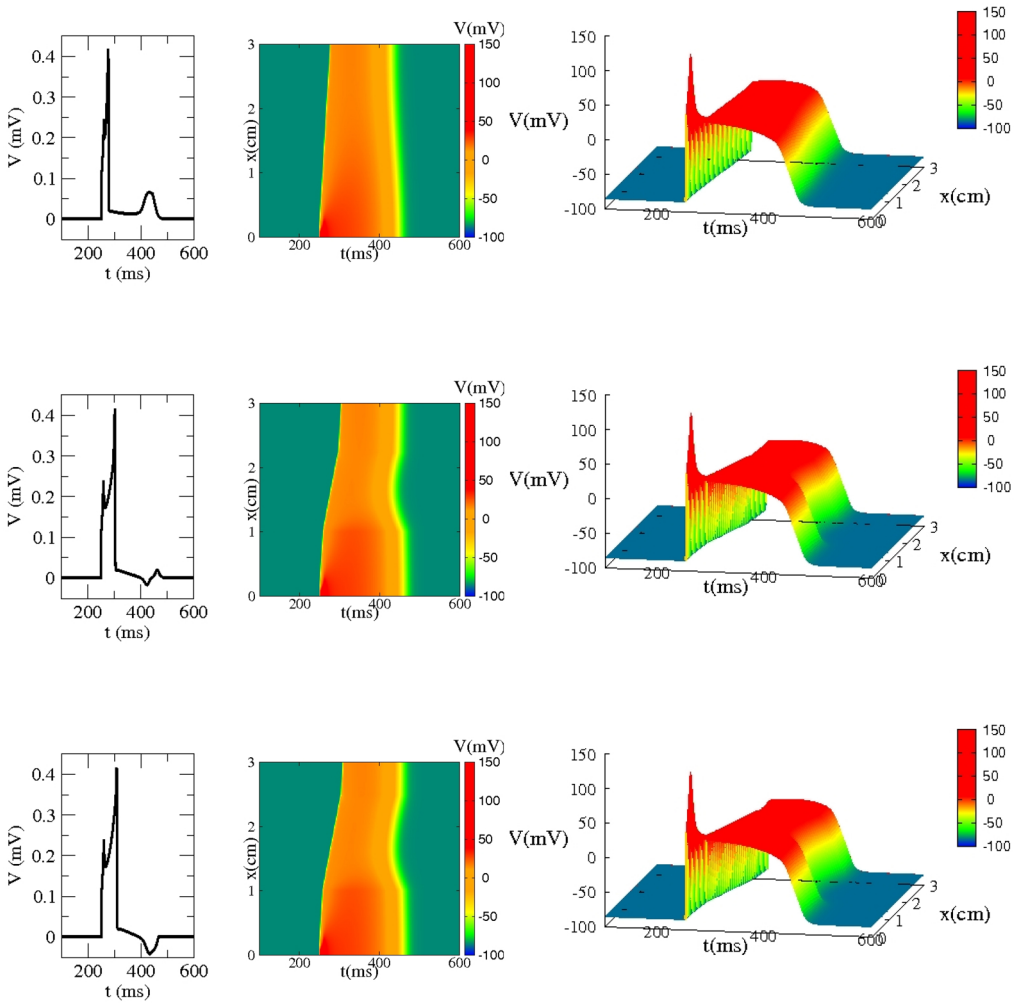

Fig. 9.7: Pseudo-ECGs as a function of time t, two-dimensional maps of the action potential on the $x - t$ plane, and three dimensional plots of the action potential on the $x - t$ plane, are shown on the left, middle, and right columns, respectively. The parameters, from top to bottom row are $L_{\text{scar}} = 0$ (first row), $L_{\text{scar}} = 1.25$ cm (second row), and $L_{\text{scar}} = 1.5$ cm (third row). The first, second, and third row show the case of pseudo-ECG with positive, biphasic, and negative (inverted) T-wave, respectively. These figures illustrate the effect of the defected region on the spatio-temporal profile of the action potential. The defected region in the second and third row starts at $L_{\text{scar}} = 0.5$, and the length of the cable is $L = 3$ cm in all three rows. Image used with kind permission from N. Lazarides.

(inverted) character of the T-wave is dominant are regarded as inverted T-waves in Fig. 9.8.

In Fig. 9.7, the pseudo-ECG as a function of time t, the map of the action potential on the $x - t$ plane, and the three-dimensional plot of the action potential on the $x - t$ plane are shown in three different widths L_{scar} of the defected region of the cardiac tissue, to illustrate its effect on the spatio-temporal profile of the action potential and eventually on the T-wave morphology. In the figure, from the first to third row (from top to bottom), the T-wave of the pseudo-ECG is positive, biphasic, and negative (inverted), respectively. The results shown on the first row have been obtained for a healthy tissue, that is, for an averaged diffusion coefficient $\tilde{D}$ which is homogeneous (without defected region, $\lambda = 0$). In this case, as shown in the map on the second column, the width of the action potential decreases monotonically as it propagates from the excitation region of the cable outwards. This is also apparent from the three-dimensional plot, where it is also clear that the amplitude of the pulse is not significantly affected during propagation. The sharp peak of the action potential profile, appearing in all three sub-figures in Fig. 9.7, is due to the action potential pulse being very close to or inside the excitation region of length $L_{exc} = 0.11$ cm of the cable. That peak however disappears after short time of propagation in all three cases.

The results shown in the second and third row have been obtained with a diffusion coefficient $\tilde{D} = \tilde{D}(x)$, as in Eq. (9.17), with $L_{\text{scar}} = 1.25$ cm and $L_{\text{scar}} = 1.5$ cm, respectively. AS it can be observed, the results in the second and third row are significantly affected by the existence of the defected region. In the second row the T-wave of the pseudo-ECG becomes biphasic, while the width of the action potential pulse is not any more monotonically decreasing during outward propagation. Instead, the pulse narrows substantially and

abruptly while it propagates into the defected region, and becomes wider after departing from it. That effect is also visible in the three-dimensional plot, where we may also observe that the amplitude of the AP is not significantly affected during propagation, even in the defected region. In the third row, the T-wave of the pseudo-ECG is inverted, becoming negative. The profile of the propagating AP pulse is in this case very similar to that shown in the second row, i.e., it narrows substantially and abruptly when entering the defected region and widens again when departing from it. In this case however the pulse narrows within a larger interval because of the larger $L_{\text{scar}} = 1.5$ cm. The three-dimensional plot is also very similar to that in the second row.

These figures clearly illustrate the effect of the defected region on the propagation of the action potential which in turn affect the pseudo-ECG and is capable of inverting the T-wave. In Fig. 9.7, no attempt was made to match observed ECG data. This would require to choose the constant K in Eq. (9.14) and other parameters appropriately. But this is outside the scope of this work which aims at showing qualitatively that spatially inhomogeneous voltage diffusion coefficients can account for the inversion of the T-wave, and also to account for the variation of the R-wave amplitude against a homogeneous (constant) diffusion coefficient $\tilde{D} = D_0$.

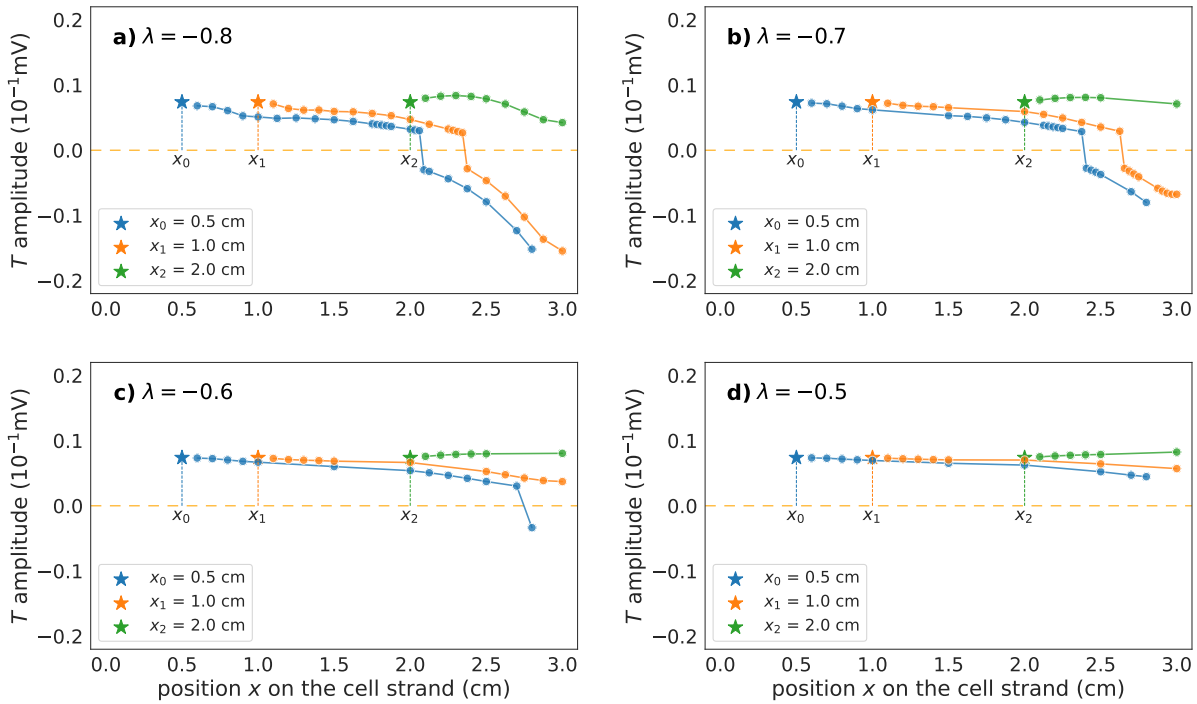

Fig. 9.8: T-wave morphology depicted as magnitude of the peak or the dip of the T-wave (in units of mV) for various starting positions and widths of the defected tissue. The variable $x\,(0 < x < L)$ is the position on the cell strand. The star at each $x_j, (j = 0, 1, 2)$ marks the position of the beginning of the defect for three different values of $x_{\text{scar}} = 0.5, 1.0$, and 2.0. All results are color-coded by this value. The location of each dot on the x-axis signifies the end point of the defected region at $x_{\text{scar}} + L_{\text{scar}}$, i.e., the distance along the x-axis of the dot from the star on the same curve represents the value of L_{scar}. Each plot has a different value for the parameter λ as indicated in the label on top. For all four subplots, the effective diffusion coefficient is $D_0 = 0.005$ cm^{2}ms^{-1}.

10

Conclusion and future work

10.1 Conclusion

Medical professionals follow their own mental model when making decisions. This model is "trained" by meticulous clinical history-taking, involving asking a series of questions and evaluating the answers, eventually refining the probabilities for a

specific disease. Past experiences with the same or other patients help them develop causal models to guide their decisions which are iteratively refined after the result of every intervention or observation.

AI models are supposed to replicate this process and assist with decision making. Can they do it? The answer would have been different 20 years ago. Today, AI algorithms are already part of the clinical evaluation of several medical modalities. There are papers that compare machine model performance with that of human readers. On one hand this shows the prospects of AI, on the other it contributes to a human-vs-machine perception of AI that makes some medical professionals worry about job security, and leaves them skeptical about AI's broader deployment. A more useful attitude would be machine and human working in a complementary way, leveraging each other's aptitudes. In this capacity, AI provides a tremendous opportunity for advancing medicine.

In this dissertation we showed the potential of ML modeling for the efficient and cost-effective diagnostic screening of abnormal LVG and cardiac remodeling through ECG. Specific clinical and ECG features including novel ECG markers assisted the model in predicting early pathological changes of LVG in patients without established CVD and detected the population who will benefit from a detailed echocardiographic evaluation.

We also showed that from basic clinical data and the use of ECG, we can identify participants with arterial hypertension which otherwise would be unaware of its existence. This ML model's purpose is to assess the risk and to prompt for further evaluation. The capability to detect a hypertensive individual immediately and efficiently only by using a simple tool as the ECG creates great potential in the management of hypertension. We showed

this using both the 12-lead ECG, and the single-lead ECG, the latter with an eye towards future use in ECG-enabled wearables.

In the computational part of this work, we simulated the dynamics of the action potential propagation in a cable and calculated a pseudo-ECG that reproduces the R wave and the T wave of an observed ECG. Our results connect the propagation of electrical (action) potentials within the cardiac tissue with the morphology of the pseudo-ECG, and by extension with what physicians actually observe i.e., the ECG. Specifically, our results reveal the dependence of the R-wave amplitude as a function of the (homogeneous) voltage diffusion coefficients and, most importantly they point towards an intimate relation between inhomogeneous diffusion coefficients (diffusion coefficients with defected regions) and the inversion (of the polarity) of the T-wave. The latter is often observed in cases of ischemia in ECG recordings by physicians.

AI is creating new opportunities to improve physicians' decision making and help with patient care. However, AI also creates ethical and privacy concerns. Populations not adequately represented in the training cohort will likely receive biased results. For example, state-of-the-art dermatology AI models substantially underperform on darker skin tones until they are fine-tuned on diverse data.[151] There is a whole field addressing ethics and fairness issues around AI, inductive or algorithmic biases, which is very important, too important to be treated lightly, so it's out of the scope of this work.

10.2 Future work: Modeling the FK cardiac model using neural networks

Our numerical solution to the FK model, shown in Eqs (8.18-8.20), by construction, produced a solution in the form of an array that contains the values of the desired functions, namely the voltage u and the gating variables v and w, at a selected group of points. Alternatively, there is a class of differential equation solvers that promise a mesh-free and time-continuous approach to solving forward and inverse problems governed by differential equations. In particular, they solve the governing coupled system of differential equations by optimizing the parameters of a deep neural network[152] (DNN) using a physics-based loss function. Neural networks are known to have function approximation capabilities and can produce solutions written in a differentiable, closed analytic form.[153] Modern DNNs, which are composed of a sequence of linear transformations and component-wise nonlinearities, provide such functions. DNN models are highly flexible—often having upwards of tens of thousands of parameters—but can still be trained and evaluated efficiently due to recent advances in parallelized hardware, automatic differentiation and stochastic optimization. More importantly, they enable integration of data and mathematical models within the same framework. This is known as a physics-informed neural network[154] (PINN). PINNs are effective especially for high-dimensional PDEs, for which traditional solvers are computationally expensive. Still, many times, conventional solvers have proved to be better. To train the PINN we need careful adjustment of the weights on the loss terms corresponding to PDE, boundary, interface, and initial conditions. E.g. for a PDE of the form

$$u_t = f(x, t, u, u_x, u_{xx}, \texttt{params}),$$

where, u_t denotes the first derivative w.r.t. time, etc., and `params` are other functions or coefficients that the solution depends on, e.g. a diffusion coefficient. The solution would then be a representation of a neural network $u = u_\theta$, trained by backpropagation to obtain gradients for its parameters θ.

While using the Runge-Kutta method resulted in an accurate solution, we did try out only a small number of values for the diffusion coefficient. Had we wanted to try out a large number of different values, or to change a host of other parameters in the forward problem, it would have been computationally tedious. A future neural network solution could be better, since it can provide a solution to the system of PDEs, for any coefficient value within that range, and the same applies for other parameters.

Another reason to utilize neural networks with PDEs is incorporating real data from observations, what is often called a *data-driven solution*.[154] That approach could solve the inverse problem; going from obtained observations/data to calibrating the parameters of the problem. In our case, a DNN would solve the inverse problem by incorporating data from real ECGs and discovering personalized parameters.

10.3 Future work: Uncovering latent features in raw ECG signals using neural networks

Although ensemble methods such as RF perform better on tabular data, deep learning has enabled tremendous progress for learning on images, language, and audio datasets. Treating the ECG as a picture, in the case of the single beat, or even as a time series if considered as a series of beats, and training a neural network on this data, could uncover features of the ECG connected and therefore predictive of certain cardiac conditions.

A

Annotation XML files

```
codeSystemName="MDC" />
<value xsi:type="CE" code="MDC_ECG_WAVC_PWAVE"
codeSystem="2.16.840.1.113883.6.24" codeSystemName="MDC" />
<support>
    <supportingROI>
        <code code="ROIPS" codeSystem="2.16.840.1.113883.5.4" />
        <component>
            <boundary>
                <code code="TIME_RELATIVE"
                codeSystem="2.16.840.1.113883.5.4"
                codeSystemName="ActCode" displayName="Relative Time"
                />
                <value xsi:type="IVL_PQ">
                    <low value="311" unit="ms" />
                    <high value="425" unit="ms" />
                </value>
```

Fig. A.1: Excerpt of the XML file showing the part where the start and end of the P wave are stored.

```
<annotationSet>
    <component>
        <annotation>
            <code code="MDC_ECG_HEART_RATE" codeSystem="2.16.840.1.113883.6.24" codeSystemName="MDC" />
            <value xsi:type="PQ" value="64" unit="bpm" />
        </annotation>
    </component>
    <component>
        <annotation>
            <code code="MDC_ECG_TIME_PD_P" codeSystem="2.16.840.1.113883.6.24" codeSystemName="MDC" />
            <value xsi:type="PQ" value="114" unit="ms" />
        </annotation>
    </component>
    <component>
        <annotation>
            <code code="MDC_ECG_TIME_PD_QRS" codeSystem="2.16.840.1.113883.6.24" codeSystemName="MDC" />
            <value xsi:type="PQ" value="79" unit="ms" />
        </annotation>
    </component>
    <component>
        <annotation>
            <code code="MDC_ECG_TIME_PD_RR" codeSystem="2.16.840.1.113883.6.24" codeSystemName="MDC" />
            <value xsi:type="PQ" value="956" unit="ms" />
        </annotation>
    </component>
    <component>
        <annotation>
            <code code="MDC_ECG_TIME_PD_PR" codeSystem="2.16.840.1.113883.6.24" codeSystemName="MDC" />
            <value xsi:type="PQ" value="175" unit="ms" />
        </annotation>
    </component>
    <component>
        <annotation>
            <code code="MDC_ECG_TIME_PD_PQ" codeSystem="2.16.840.1.113883.6.24" codeSystemName="MDC" />
            <value xsi:type="PQ" value="175" unit="ms" />
        </annotation>
    </component>
    <component>
        <annotation>
```

Fig. A.2: Excerpt from an XML file showing the part where the variable “MDC_ECG_HEART_RATE” stores the calculated heart rate, which, in this person, is 64 bpm. Also shown, among others, is the variable “MDC_ECG_TIME_PD_QRS” which stores the duration of the QRS complex as an average for all leads (in this case 79 ms). For definitions see Fig. 4.2 in Chapter 2.

B

Supplemental material

Feature	HTN			NT			p value
	Mean	Std	Range	Mean	Std	Range	
Systolic blood presure, mmHg	139.8	13.8	100.0 - 185.00	127.5	9.1	91.0 - 175.00	<0.001
Age, years	62.5	10.5	30.0 - 80.00	53.3	10.2	30.0 - 80.00	<0.001
Pulse blood pressure, mmHg	55.2	11.9	15.0 - 104.00	46.7	8.0	20.0 - 80.00	<0.001
Mean blood pressure	103.0	9.2	68.0 - 146.30	96.3	7.2	70.7 - 147.00	<0.001
Body mass index, kgr/m^2	31.4	5.4	18.8 - 56.64	28.1	5.2	17.6 - 48.87	<0.001
BMI-adjusted Cornell, mV·kgr/m^2	45.1	18.5	4.3 - 128.13	34.3	16.2	3.1 - 118.90	<0.001
Body fat, kgr/m^2	41.6	9.0	18.9 - 75.45	36.6	8.1	17.3 - 67.04	<0.001
Diastolic blood pressure, mmHg	84.6	8.9	42.0 - 127.00	80.7	7.6	52.0 - 133.00	<0.001
R wave amplitude in aVL, mV	0.6	0.3	0.0 - 1.91	0.5	0.3	0.0 - 1.61	<0.001
Weight, kgr	86.1	18.0	43.0 - 180.00	78.6	17.4	45.0 - 153.00	<0.001
Cornell criteria, mV	1.4	0.5	0.1 - 3.77	1.2	0.5	0.1 - 3.37	<0.001
Area under R wave in I, ms·mV	9.3	3.9	0.8 - 33.36	7.8	3.3	0.8 - 21.10	<0.001
QRS axis front, degrees°	13.6	32.4	-77.0 - 188.00	26.0	30.6	-82.0 - 88.00	<0.001
Corrected QT interval, ms	423.7	24.6	337.0 - 500.00	414.8	22.1	364.0 - 506.00	<0.001
P wave duration, ms	114.6	15.5	0.0 - 196.00	111.7	10.6	75.0 - 149.00	<0.001
PQ interval duration, ms	167.0	27.4	0.0 - 277.00	159.0	20.8	112.0 - 226.00	<0.001
QT interval duration, ms	397.9	31.7	290.0 - 501.00	387.5	28.7	312.0 - 496.00	<0.001
R wave amplitude in III, mV	0.2	0.2	0.0 - 1.37	0.3	0.3	0.0 - 1.74	<0.001
Planar frontal QRS-T angle, degrees°	37.0	35.3	0.0 - 178.00	26.2	25.3	0.0 - 168.00	<0.001
Area under R wave in aVF, ms·mV	3.9	3.6	0.0 - 25.39	4.7	3.7	0.0 - 23.51	<0.001
Area under T wave divided by QRS complex area	1.0	0.5	0.0 - 3.46	1.1	0.5	0.1 - 3.12	<0.001
Area under R wave in III, ms·mV	1.9	2.4	0.0 - 21.05	2.6	3.0	0.0 - 19.12	<0.001
BMI-modified Sokolow-Lyon voltage, kgr/m^2· mV	17.5	7.7	4.9 - 96.88	15.4	5.9	5.8 - 41.95	<0.001
BMI-adjusted Sokolow-Lyon voltage, mV	2.6	0.6	0.2 - 5.66	2.4	0.6	1.0 - 4.73	<0.001
Total QRS area in all leads, ms·mV	291.2	78.0	118.4 - 734.18	272.2	78.4	133.9 - 942.09	<0.001
S wave amplitude in V_5, mV	0.4	0.3	0.0 - 1.46	0.3	0.2	0.0 - 1.60	<0.001
T wave amplitude in V_5, mV	0.3	0.2	-0.7 - 1.02	0.3	0.2	-0.5 - 0.82	<0.001
S wave amplitude in V_3, mV	0.9	0.4	0.0 - 2.84	0.8	0.4	0.0 - 3.15	<0.001
QRS complex duration, ms	92.6	11.0	62.0 - 153.00	90.6	11.2	55.0 - 147.00	0.002
P wave amplitude in II, mV	0.1	0.0	0.0 - 0.28	0.1	0.0	0.0 - 0.41	0.003
Area under QRS interval in V_5, ms·mV	34.7	12.3	10.5 - 93.14	32.5	11.1	11.7 - 97.14	0.004
Q vs. S vector	0.9	0.6	-0.8 - 2.91	0.9	0.6	-0.8 - 2.65	0.008
J point deflection, mV	-0.0	0.0	-0.1 - 0.14	-0.0	0.0	-0.1 - 0.13	0.01
Q wave duration, ms	10.5	8.1	0.0 - 36.00	11.9	8.6	0.0 - 48.00	0.01
P axis in frontal plane, degrees°	46.6	22.5	-59.0 - 116.00	50.4	26.4	-61.0 - 268.00	0.011
Height, cm	165.4	9.9	142.0 - 192.00	166.9	9.8	137.0 - 200.00	0.039
Intrincicoid deflection in II, ms	41.9	6.7	5.0 - 95.00	41.1	6.7	4.0 - 64.00	0.042
Area under S wave in V_1, ms·mV	18.6	11.5	0.0 - 77.63	17.2	9.9	0.0 - 51.79	0.054
T wave duration, ms, ms	204.0	42.3	43.0 - 345.00	200.6	32.4	77.0 - 312.00	0.071
T wave amplitude in III, mV	0.0	0.1	-0.4 - 0.37	0.0	0.1	-0.3 - 0.57	0.077

BMI: body mass index, HTN: participants with hypertension, NT: participants with normal blood pressure (continues in next page).

Table B.1: (continued in next page).

Feature	HTN			NT			p value
	Mean	Std	Range	Mean	Std	Range	
Heart rate, bpm	69.9	11.6	40.0 - 129.00	71.1	12.0	48.0 - 109.00	0.108
R wave amplitude in V_2, mV	0.4	0.3	0.0 - 1.83	0.4	0.3	0.0 - 1.54	0.123
R-R interval duration, ms	884.7	140.5	464.0 - 1508.00	872.5	145.5	544.0 - 1360.00	0.128
T wave amplitude in aVL, mV	0.1	0.1	-0.3 - 0.43	0.1	0.1	-0.3 - 0.49	0.133
Q vs. S vector	-0.4	0.5	-2.8 - 1.31	-0.3	0.4	-1.9 - 1.26	0.227
Area under S wave in V_2, ms·mV	19.5	12.4	0.0 - 89.73	18.3	11.3	0.0 - 72.03	0.243
QRS-modified Sokolow-Lyon voltage, ms· mV	187.1	65.3	26.0 - 520.02	182.8	63.4	79.4 - 433.00	0.256
ST segment duration, ms	101.3	48.1	0.0 - 268.00	96.3	37.8	0.0 - 229.00	0.265
P wave amplitude in V_2, mV	0.0	0.0	0.0 - 0.18	0.0	0.0	0.0 - 0.19	0.281
T wave amplitude in V_2, mV	0.3	0.2	-0.3 - 1.12	0.3	0.2	-0.2 - 1.26	0.352
Quotient of RV2/RV5 amplitudes	0.4	0.3	0.0 - 1.99	0.4	0.3	0.0 - 1.71	0.438
R wave amplitude in V_6, mV	1.1	0.4	0.0 - 2.92	1.1	0.4	0.1 - 3.32	0.442
S wave amplitude in V_1, mV	0.7	0.3	0.0 - 2.56	0.7	0.3	0.0 - 1.58	0.454
Intrinsicoid deflection in V_5, ms	39.6	6.2	17.0 - 103.00	40.0	6.3	20.0 - 66.00	0.457
Vector V2/vector V5 adjusted for BMI	1.8	4.9	0.0 - 53.71	1.7	4.1	0.0 - 44.98	0.49
R wave amplitude in V_5, mV	1.2	0.5	0.0 - 3.21	1.2	0.5	0.2 - 2.83	0.514
T ratio	1.8	6.1	0.3 - 156.00	1.5	0.6	0.7 - 8.00	0.613
S wave amplitude in V_2, mV	0.8	0.4	0.0 - 3.03	0.8	0.4	0.1 - 2.37	0.783
Tallest R wave in limb leads, mV	0.9	0.3	0.3 - 1.96	0.9	0.3	0.3 - 2.07	0.824
Area under R wave in V_5, ms·mV	14.2	6.4	2.3 - 52.34	13.9	5.7	0.6 - 33.44	0.832
T axis frontal plane, degrees°	38.5	35.3	-87.0 - 258.00	36.2	23.4	-58.0 - 154.00	0.981
Sokolow-Lyon voltage, mV	2.0	0.6	0.2 - 4.86	2.0	0.6	0.8 - 4.33	0.991

BMI: body mass index, HTN: participants with hypertension, NT: participants with normal blood pressure.

Table B.1: Characteristics and Comparative Statistics for Hypertensive and Normotensive Study Participants.

References

[1] World Health Organization. Cardiovascular diseases, 2023 (accessed Oct 28, 2023).

[2] A Ganau, RB Devereux, MJ Roman, G de Simone, TG Pickering, PS Saba, P Vargiu, I Simongini, and JH Laragh. Patterns of left ventricular hypertrophy and geometric remodeling in essential hypertension. *J Am Coll Cardiol*, 19(7):1550–1558, 1992.

[3] MR Dweck, S Joshi, T Murigu, A Gulati, F Alpendurada, A Jabbour, A Maceira, I Roussin, DB Northridge, PJ Kilner, SA Cook, NA Boon, J Pepper, RH Mohiaddin, DE Newby, DJ Pennell, and SK Prasad. Left ventricular remodeling and hypertrophy in patients with aortic stenosis: insights from cardiovascular magnetic resonance. *J Cardiovasc Magn Reson*, 14(1):50–50, 2012.

[4] HM Krumholz, M Larson, and D Levy. Prognosis of left ventricular geometric patterns in the framingham heart study. *J Am Coll Cardiol*, 25(4):879–884, 1995.

[5] RB Devereux, K Wachtell, E Gerdts, K Boman, MS Nieminen, V Papademetriou, J Rokkedal, K Harris, P Aurup, and B DahlÃ¶f. Prognostic significance of left ventricular mass change during treatment of hypertension. *JAMA*, 292(19):2350–2356, 2004.

[6] Blausen.com staff. Blausen 0462 - anatomy of the heart, 2014.

[7] Augustus D. Waller. A demonstration of man of electromotive changes accompanying the hearts beat. *Annals of noninvasive electrocardiology*, 9(2):189–191, 2004.

[8] Einthoven W. Über die form des menschlichen electrocardiogramms. *Pfügers Archiv march 1895*, 1895.

[9] A Goldberger, ZD Goldberger, and A Shvilkin. *Section 3 - Quiz Master: Self-Assessment of Clinical ECG Skills*. Elsevier, ninth edition edition, 2018.

[10] *Braunwald's heart disease : a textbook of cardiovascular medicine*. Elsevier Saunders, Philadelphia, 9th ed. edition, 2012.

[11] Flavio H. Fenton. *Theoretical Investigation of Spiral and Scroll Wave Instabilities Underlying Cardiac Fibrillation*. Phd thesis, Northeastern University Graduate School of Arts and Sciences, Boston, MA, August 1999.

[12] Leonard S Lilly. *Pathophysiology of heart disease: an introduction to cardiovascular medicine*. Wolters Kluwer, Philadelphia, seventh edition edition, 2021.

[13] Macfarlane Peter W.; Janse Michiel; van Oosterom A.; Kligfield Paul; Pahlm Olle; Camm John. *Comprehensive Electrocardiology*. Springer London, London, 2 edition, 2011.

[14] S. Lilly. *Pathophysiology of heart disease : a collaborative project of medical students and faculty*. Wolters Kluwer, Philadelphia, edition 6 edition, 2016.

[15] Yumiko Kanzaki, Fumio Terasaki, Makoto Okabe, Shuichi Fujita, Takashi Katashima, Kaoru Otsuka, and Nobukazu Ishizaka. Three-dimensional architecture of cardiomyocytes and connective tissue in human heart revealed by scanning electron microscopy. *Circulation (New York, N.Y.)*, 122(19):1973–1974, 2010.

[16] Ary Louis Goldberger. *Clinical electrocardiography : a simplified approach*. Elsevier Mosby, Edinburgh, 7th ed. edition, 2006.

[17] Andrew L Wit. *The ventricular arrhythmias of ischemia and infarction : electrophysiological mechanisms*. Futura Pub. Co., Mount Kisco, NY, 1993.

[18] D. P. Zipes, J. Jalife, and W. G. Stevenson. *Cardiac electrophysiology : from cell to bedside*. Elsevier, Philadelphia, PA, 7th ed. edition, 2018.

[19] RM Shaw and Y Rudy. Electrophysiologic effects of acute myocardial ischemia: a theoretical study of altered cell excitability and action potential duration. *Cardiovascular research*, 35(2):256–272, 1997.

[20] H. J Jongsma and R Wilders. Gap junctions in cardiovascular disease. *Circulation research*, 86(12):1193–1197, 2000.

[21] Shlomo Mor-Yosef, Arnon Samueloff, Baruch Modan, Daniel Navot, and Joseph G Schenker. Ranking the risk factors for cesarean: Logistic regression analysis of a nationwide study. *Obstetrics and gynecology (New York. 1953)*, 75(6):944–947, 1990.

[22] Leijnen F van Veen F. Neural network zoo, 2019 (accessed Oct 28, 2023).

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

[24] Gareth (Gareth Michael) James. *An introduction to statistical learning : with applications in Python*. Springer texts in statistics. Springer, Cham, 1 edition, 2023.

[25] Rayan Krishnan, Pranav Rajpurkar, and Eric J Topol. Self-supervised learning in medicine and healthcare. *Nature biomedical engineering*, 6(12):1346–1352, 2022.

[26] Dani Kiyasseh, Tingting Zhu, and David A Clifton. Clocs: Contrastive learning of cardiac signals across space, time, and patients. *arXiv.org*, 2021.

[27] OpenAI. Chatgpt based on gpt-4. https://www.openai.com/, 2022. Accessed: Oct.28, 2023.

[28] Peter Lee, Sebastien Bubeck, and Joseph Petro. Benefits, limits, and risks of gpt-4 as an ai chatbot for medicine. *The New England journal of medicine*, 388(13):1233–1239, 2023.

[29] T Hastie, R Tibshirani, and J Friedman. *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York, NY, 2009.

[30] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.

[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[32] JD Hunter. Matplotlib: A 2d graphics environment. *Comput Sci Eng*, 9(3):90–95, 2007.

[33] The pandas development team. pandas-dev/pandas: Pandas, February 2020.

[34] L Breiman. Random forests. *Mach Learn*, 45(1):5–32, 2001.

[35] L Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[36] LÃ©o Grinsztajn, Edouard Oyallon, and GaÃ«l Varoquaux. Why do tree-based models still outperform deep learning on tabular data? *arXiv.org*, 2022.

[37] SM Lundberg, G Erion, H Chen, A DeGrave, JM Prutkin, B Nair, R Katz, J Himmelfarb, N Bansal, and SI Lee. Explainable ai for trees: From local explanations to global understanding. *CoRR*, abs/1905.04610, 2019.

[38] SM Lundberg, GG Erion, and SI Lee. Consistent individualized feature attribution for tree ensembles. *CoRR*, 2018.

[39] SM Lundberg and SI Lee. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*, pages 4765–4774, 2017.

[40] SM Lundberg, B Nair, MS Vavilala, M Horibe, MJ Eisses, T Adams, DE Liston, DK Low, SF Newman, J Kim, and SI Lee. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng*, 2(10):749–760, 2018.

[41] LJP van der Maaten and GE Hinton. Visualizing high-dimensional data using t-sne. *J Mach Learn Res*, 9(nov):2579–2605, 2008.

[42] W Xiong, J Droppo, X Huang, F Seide, M Seltzer, A Stolcke, D Yu, and G Zweig. Achieving human parity in conversational speech recognition. *arXiv.org*, 2017.

[43] Antoine Buetti-Dinh, Vanni Galli, SÃ¶ren Bellenberg, Olga Ilie, Malte Herold, Stephan Christel, Mariia Boretska, Igor V. Pivkin, Paul Wilmes, Wolfgang Sand, Mario Vera, and Mark Dopson. Deep neural networks outperform human expert's capacity in characterizing bioleaching bacterial biofilm composition. *Biotechnology Reports*, 22:e00321–e00321, 2019.

[44] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature (London)*, 529(7587):484–489, 2016.

[45] Pranav Rajpurkar, Jeremy Irvin, Robyn L. Ball, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis P. Langlotz, Bhavik N. Patel, Kristen W. Yeom, Katie Shpanskaya, Francis G. Blankenberg, Jayne Seekins, Timothy J. Amrhein, David A. Mong, Safwan S. Halabi, Evan J. Zucker, Andrew Y. Ng, and Matthew P. Lungren. Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists. *PLoS medicine*, 15(11):e1002686–e1002686, 2018.

[46] Rohan Gupta, Devesh Srivastava, Mehar Sahu, Swati Tiwari, Rashmi K. Ambasta, and Pravir Kumar. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular diversity*, 25(3):1315–1360, 2021.

[47] Mathieu Boniol, Teena Kunjumen, Tapas Sadasivan Nair, Amani Siyam, James Campbell, and Khassoum Diallo. The global health workforce stock and distribution in 2020 and 2030: a threat to equity and universal health coverage? *BMJ global health*, 7(6):e009316, 2022.

[48] Alejandra Barrero-Castillero, Brian K. Corwin, Deborah K. VanderVeen, and Jason C. Wang. Workforce shortage for retinopathy of prematurity care and emerging role of telehealth and artificial intelligence. *The Pediatric clinics of North America*, 67(4):725–733, 2020.

[49] Chayakrit Krittanawong, Kipp W. Johnson, Robert S. Rosenson, Zhen Wang, Mehmet Aydar, Usman Baber, James K. Min, W. H. Wilson Tang, Jonathan L. Halperin, and Sanjiv M. Narayan. Deep learning for cardiovascular medicine: a practical primer. *European heart journal*, 40(25):2058–2073, 2019.

[50] KC Siontis, X Yao, JP Pirruccello, AA Philippakis, and PA Noseworthy. How will machine learning inform the clinical care of atrial fibrillation. *Circ Res*, 127(1):155–169, 2020.

[51] K Seetharam, S Raina, and PP Sengupta. The role of artificial intelligence in echocardiography. *Curr Cardiol Rep*, 22(9), 2020.

[52] JD Lanzer, F Leuschner, R Kramann, RT Levinson, and J Saez-Rodriguez. Big data approaches in heart failure research. *Curr Heart Fail Rep*, 17(5):213–224, 2020.

[53] K Gilbert, C Mauger, AA Young, and A Suinesiaputra. Artificial intelligence in cardiac imaging with statistical atlases of cardiac anatomy. *Front Cardiovasc Med*, 7, 2020.

[54] W Nabi, A Bansal, and B Xu. Applications of artificial intelligence and machine learning approaches in echocardiography. *Echocardiography (Mount Kisco, N.Y.)*, 38(6):982–992, 2021.

[55] A Lin, M KolossvÃ¡ry, M Motwani, I IÅ¡gum, Pl Maurovich-Horvat, PJ Slomka, and D Dey. Artificial intelligence in cardiovascular imaging for risk stratification in coronary artery disease. *Radiol Cardiothorac Imaging*, 3(1):e200512–e200512, 2021.

[56] Matthew W. Segar, Kershaw V. Patel, Colby Ayers, Mujeeb Basit, W.H. Wilson Tang, Duwayne Willett, Jarett Berry, Justin L. Grodin, and Ambarish Pandey. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning based unsupervised cluster analysis. *European journal of heart failure*, 22(1):148–158, 2020.

[57] Damini Dey, Piotr J. Slomka, Paul Leeson, Dorin Comaniciu, Sirish Shrestha, Partho P. Sengupta, and Thomas H. Marwick. Artificial intelligence in cardiovascular imaging jacc state-of-the-art review. *Journal of the American College of Cardiology*, 73(11):1317–1335, 2019.

[58] Partho P Sengupta and Y Chandrashekhar. From conventional deep learning to gpt: Ai's emergent power for cardiac imaging. *JACC. Cardiovascular imaging*, 16(8):1129–1131, 2023.

[59] MÃ¡rton Tokodi, BÃ¡lint Magyar, AndrÃ¡s SoÃ³s, Masaaki Takeuchi, MÃ¡tÃ© Tolvaj, BÃ¡lint KÃ¡roly Lakatos, Tetsuji Kitano, Yosuke Nabeshima, Alexandra FÃ¡biÃ¡n, Mark Bence Szigeti, AndrÃ¡s HorvÃ¡th, BÃ©la Merkely, and Attila KovÃcs. Deep learning-based prediction of right ventricular ejection fraction using 2d echocardiograms. *JACC. Cardiovascular imaging*, 16(8):1005–1018, 2023.

[60] Florin-Cristian Ghesu, Bogdan Georgescu, Yefeng Zheng, Sasa Grbic, Andreas Maier, Joachim Hornegger, and Dorin Comaniciu. Multi-scale deep reinforcement learning for real-time 3d-landmark detection in ct scans. *IEEE transactions on pattern analysis and machine intelligence*, 41(1):176–189, 2019.

[61] KC Siontis, PA Noseworthy, ZI Attia, and PA Friedman. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*, 18(7):465–478, 2021.

[62] PE Vardas, FW Asselbergs, M van Smeden, and P Friedman. The year in cardiovascular medicine 2021: digital health and innovation. *Eur Heart J*, 2022.

[63] ZI Attia, DM Harmon, ER Behr, and PA Friedman. Application of artificial intelligence to the electrocardiogram. *Eur Heart J*, 42(46):4717–4730, 2021.

[64] KC Siontis, K Liu, JM Bos, ZI Attia, M Cohen-Shelly, AM Arruda-Olson, N Zanjirani Farahani, PA Friedman, PA Noseworthy, and MJ Ackerman. Detection of hypertrophic cardiomyopathy by an artificial intelligence electrocardiogram in children and adolescents. *Int J Cardiol*, 340:42–47, 2021.

[65] P Bachtiger, CF Petri, FE Scott, S Ri Park, MA Kelshiker, HK Sahemey, B Dumea, R Alquero, PS Padam, IR Hatrick, A Ali, M Ribeiro, WS Cheung, N Bual, B Rana, M Shun-Shin, DB Kramer, A Fragoyannis, D Keene, CM Plymen, and NS Peters. Point-of-care screening for heart failure with reduced ejection fraction using artificial intelligence during ecg-enabled stethoscope examination in london, uk: a prospective, observational, multicentre study. *Lancet Digit Health*, 2022.

[66] Rachael Hagan, Charles J. Gillan, and Fiona Mallett. Comparison of machine learning methods for the classification of cardiovascular disease. *Informatics in medicine unlocked*, 24:100606, 2021.

[67] Awni Y. Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H. Tison, Codie Bourn, Mintu P. Turakhia, and Andrew Y. Ng. Cardiologist-level arrhythmia

detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1):65–69, 2019.

[68] Zachi I Attia, Peter A Noseworthy, Francisco Lopez-Jimenez, Samuel J Asirvatham, Abhishek J Deshmukh, Bernard J Gersh, Rickey E Carter, Xiaoxi Yao, Alejandro A Rabinstein, Brad J Erickson, Suraj Kapa, and Paul A Friedman. An artificial intelligence-enabled ecg algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet (British edition)*, 394(10201):861–867, 2019.

[69] Fernando De la Garza Salazar, Maria Elena Romero Ibarguengoitia, JosÃ© RamÃ³n Azpiri LÃ³pez, and Arnulfo GonzÃ¡lez CantÃº. Optimizing ecg to detect echocardiographic left ventricular hypertrophy with computer-based ecg data and machine learning. *PloS one*, 16(11):e0260661–e0260661, 2021.

[70] N Kagiyama, M Piccirilli, N Yanamala, S Shrestha, PD Farjo, G Casaclang-Verzosa, WM Tarhuni, N Nezarat, MJ Budoff, J Narula, and PP Sengupta. Machine learning assessment of left ventricular diastolic function based on electrocardiographic features. *J Am Coll Cardiol*, 76(8):930–941, 2020.

[71] Marc Strik, ThÃ©o Caillol, F Daniel Ramirez, Saer Abu-Alrub, Hugo Marchand, Nicolas Welte, Philippe Ritter, Michel HaÃ¯ssaguerre, Sylvain Ploux, and Pierre Bordachar. Validating qt-interval measurement using the apple watch ecg to enable remote monitoring during the covid-19 pandemic. *Circulation (New York, N.Y.)*, 142(4):416–418, 2020.

[72] Joel M. Raja, Carol Elsakr, Sherif Roman, Brandon Cave, Issa Pour-Ghaz, Amit Nanda, Miguel Maturana, and Rami N. Khouzam. Apple watch, wearables, and heart rhythm: where do we stand? *Annals of translational medicine*, 7(17):417–417, 2019.

[73] Jianzheng Li, Jialong Zhang, Yizhou Jiang, Chongyuan Ren, Ran Guo, Yu Ma, and Yajie Qin. A flexible and miniaturized chest patch for real-time ppg/ecg/bio-z monitoring. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, volume 2022, pages 4312–4315, Piscataway, 2022. The Institute of Electrical and Electronics Engineers, Inc. (IEEE).

[74] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. 2013.

[75] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *International journal of computer vision*, 128(2):336–359, 2020.

[76] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on knowledge discovery and data mining*, pages 1135–1144. ACM, 2016.

[77] William Press. *Numerical recipes : the art of scientific computing*. Cambridge University Press, Cambridge [Cambridgeshire] ; New York, 1986.

[78] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[79] Bryan Williams, Giuseppe Mancia, Wilko Spiering, Enrico Agabiti Rosei, Michel Azizi, Michel Burnier, Denis Clement, Antonio Coca, Giovanni De Simone, Anna Dominiczak, Thomas Kahan, Felix Mahfoud, Josep Redon, Luis Ruilope, Alberto Zanchetti, Mary Kerins, Sverre Kjeldsen, Reinhold Kreutz, Stephane Laurent, Gregory Y.H Lip, Richard McManus, Krzysztof Narkiewicz, Frank Ruschitzka, Roland Schmieder, Evgeny Shlyakhto, Konstantinos Tsioufis, Victor Aboyans, and Ileana Desormais. 2018 practice guidelines for the management of arterial hypertension of the european society of hypertension and the european society of cardiology : Esh/esc task force for the management of arterial hypertension. *Journal of hypertension*, 36(12):2284–2309, 2018.

[80] TH Marwick, TC Gillebert, G Aurigemma, J Chirinos, G Derumeaux, M Galderisi, J Gottdiener, B Haluska, E Ofili, P Segers, R Senior, RJ Tapp, and JL Zamorano. Recommendations on the use of echocardiography in adult hypertension: A report from the european association of cardiovascular imaging (eacvi) and the american society of echocardiography (ase). *J Am Soc Echocardiogr*, 28(7):727–754, 2015.

[81] Jan A. Weir CB. Bmi classification percentile and cut off points, 2023 (accessed Oct 28, 2023).

[82] WHO. Obesity: Preventing and managing the global epidemic - introduction: Report of a who consultation. *OBESITY: PREVENTING AND MANAGING THE GLOBAL EPIDEMIC*, 894:1–253, 2000.

[83] R D Mosteller. Simplified calculation of body-surface area. *The New England journal of medicine*, 317(17):1098–1098, 1987.

[84] P Deurenberg, JA Weststrate, and JC Seidell. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. *Br J Nutr*, 65(2):105–114, 1991.

[85] Unknown author(s). Detection of left ventricular hypertrophy on the ecg through machine learning with a focus on obesity. *European heart journal. Digital health*, 3(4), December 2022.

[86] Satoshi Kurisu, Kazuhiro Nitta, Yoji Sumimoto, Hiroki Ikenaga, Ken Ishibashi, Yukihiro Fukuda, and Yasuki Kihara. Implications of electrocardiographic frontal qrs axis on left ventricular diastolic parameters derived from electrocardiogram-gated myocardial perfusion single photon emission computed tomography. *Annals of nuclear medicine*, 32(6):404–409, 2018.

[87] Aapo L. Aro, Heikki V. Huikuri, Jani T. Tikkanen, M. Juhani Junttila, Harri A. Rissanen, Antti Reunanen, and Olli Anttonen. Qrs-t angle as a predictor of sudden cardiac death in a middle-aged general population. *Europace (London, England)*, 14(6):872–876, 2012.

[88] PW Macfarlane. The frontal plane qrs-t angle. *Europace*, 14(6):773–775, 2012.

[89] Marc R. Dweck, Sanjiv Joshi, Timothy Murigu, Ankur Gulati, Francisco Alpendurada, Andrew Jabbour, Alicia Maceira, Isabelle Roussin, David B. Northridge, Philip J. Kilner, Stuart A. Cook, Nicholas A. Boon, John Pepper, Raad H. Mohiaddin, David E. Newby, Dudley J. Pennell, and Sanjay K. Prasad. Left ventricular remodeling and hypertrophy in patients with aortic stenosis: insights from cardiovascular magnetic resonance. *Journal of cardiovascular magnetic resonance*, 14(1):50–50, 2012.

[90] Anil Verma, Alessandra Meris, Hicham Skali, Jalal K. Ghali, J. Malcolm O. Arnold, Mikhail Bourgoun, Eric J. Velazquez, John J. V. McMurray, Lars Kober, Marc A. Pfeffer, Robert M. Califf, and Scott D. Solomon. Prognostic implications of left ventricular mass and geometry following myocardial infarction the valiant (valsartan in acute myocardial infarction) echocardiographic study. *JACC. Cardiovascular imaging*, 1(5):582–591, 2008.

[91] M Bressman, AY Mazori, E Shulman, JJ Chudow, Y Goldberg, JD Fisher, KJ Ferrick, M Garcia, L Di Biase, and A Krumerman. Determination of sensitivity and specificity of electrocardiography for left ventricular hypertrophy in a large, diverse patient population. *Am J Med*, 133(9):e495–e500, 2020.

[92] D Levy, SB Labib, KM Anderson, JC Christiansen, WB Kannel, and WP Castelli. Determinants of sensitivity and specificity of electrocardiographic criteria for left ventricular hypertrophy. *Circulation*, 81(3):815–820, 1990.

[93] PK Whelton, RM Carey, WS Aronow, DE Casey, KJ Collins, C Dennison Himmelfarb, SM DePalma, S Gidding, KA Jamerson, DW Jones, EJ MacLaughlin, P Muntner, B Ovbiagele, SC Smith, CC Spencer, RS Stafford, SJ Taler, RJ Thomas, KA Williams, JD Williamson, and JT Wright. 2017 acc/aha/aapa/abc/acpm/ags/apha/ash/aspc/nma/pcna guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: A report of the american college of cardiology/american heart association task force on clinical practice guidelines. *J Am Coll Cardiol*, 71(19):e127–e248, 2018.

[94] H Afify, HL Lee, EZ Soliman, and MJ Singleton. Prognostic significance of body mass index adjusted criteria for left ventricular hypertrophy. *J Clin Hypertens (Greenwich)*, 22(8):1476–1483, 2020.

[95] G Lematre, F Nogueira, and CK Aridas. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res*, 18(17):1–5, 2017.

[96] R Sparapani, NM Dabbouseh, D Gutterman, J Zhang, H Chen, DA Bluemke, JA. C Lima, GL Burke, and EZ Soliman. Detection of left ventricular hypertrophy using bayesian additive regression trees: The mesa (multi-ethnic study of atherosclerosis). *J Am Heart Assoc*, 8(5):e009959, 2019.

[97] CS Park, JB Park, Y Kim, YE Yoon, SP Lee, HK Kim, YJ Kim, GY Cho, DW Sohn, and SH Lee. Left ventricular geometry determines prognosis and reverse j-shaped relation between blood pressure and mortality in ischemic stroke patients. *JACC Cardiovasc Imaging*, 11(3):373–382, 2017.

[98] PK Whelton, RM Carey, WS Aronow, DE Casey, KJ Collins, C Dennison Himmelfarb, SM DePalma, S Gidding, KA Jamerson, DW Jones, EJ MacLaughlin, P Muntner, B Ovbiagele, SC Smith, CC Spencer, RS Stafford, SJ Taler, RJ Thomas, KA Williams, JD Williamson, and JT Wright. 2017 acc/aha/aapa/abc/acpm/ags/apha/ash/aspc/nma/pcna guideline for the prevention, detection, evaluation,

and management of high blood pressure in adults: Executive summary: A report of the american college of cardiology/american heart association task force on clinical practice guidelines. *Hypertension*, 71(6):1269–1324, 2018.

[99] AS Go, D Mozaffarian, VL Roger, EJ Benjamin, JD Berry, WB Borden, DM Bravata, S Dai, ES Ford, CS Fox, S Franco, HJ Fullerton, C Gillespie, SM Hailpern, JA Heit, VJ Howard, MD Huffman, BM Kissela, SJ Kittner, DT Lackland, JH Lichtman, LD Lisabeth, D Magid, GM Marcus, A Marelli, DB Matchar, DK McGuire, ER Mohler, CS Moy, ME Mussolino, G Nichol, NP Paynter, PJ Schreiner, PD Sorlie, J Stein, TN Turan, SS Virani, ND Wong, D Woo, and MB Turner. Executive summary: Heart disease and stroke statistics: 2013 update: A report from the american heart association. *Circulation*, 127(1):143–146, 2013.

[100] CMM Lawes, SV Hoorn, and A Rodgers. Global burden of blood-pressure-related disease, 2001. *Lancet (British edition)*, 371(9623):1513–1518, 2008.

[101] GA Roth, GA Mensah, CO Johnson, G Addolorato, E Ammirati, LM Baddour, NC Barengo, AZ Beaton, EJ Benjamin, CP Benziger, A Bonny, M Brauer, M Brodmann, TJ Cahill, and Carapetis. Global burden of cardiovascular diseases and risk factors, 1990-2019 update from the gbd 2019 study. *J Am Coll Cardiol*, 76(25):2982–3021, 2020.

[102] NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *Lancet (British edition)*, 398(10304):957–980, 2021.

[103] CK Chow, KK Teo, S Rangarajan, S Islam, R Gupta, A Avezum, A Bahonar, J Chifamba, G Dagenais, R Diaz, K Kazmi, F Lanas, L Wei, P Lopez-Jaramillo, L Fanghong, NH Ismail, T Puoane, A Rosengren, A Szuba, A Temizhan, A Wielgosz, R Yusuf, A Yusufali, M McKee, L Liu, P Mony, and S Yusuf. Prevalence, awareness, treatment, and control of hypertension in rural and urban communities in high-, middle-, and low-income countries. *JAMA*, 310(9):959–968, 2013.

[104] JS Lindholt and R SÃžgaard. Population screening and intervention for vascular disease in danish men (viva): a randomised controlled trial. *Lancet (British edition)*, 390(10109):2256–2265, 2017.

[105] G Parati, C Lombardi, M Pengo, G Bilo, and JE Ochoa. Current challenges for hypertension management: From better hypertension diagnosis to improved patients' adherence and blood pressure control. *Int J Cardiol*, 331:262–269, 2021.

[106] LP Yao and Z Pan. Cuff-less blood pressure estimation from photoplethysmography signal and electrocardiogram. *Phys Eng Sci Med*, 44(2):397–408, 2021.

[107] S Baker, W Xiang, and I Atkinson. A hybrid neural network for continuous and non-invasive estimation of blood pressure from raw electrocardiogram and photoplethysmogram waveforms. *Comput Methods Programs Biomed*, 207:106191–106191, 2021.

[108] M Sharma, JS Rajput, RS Tan, and UR Acharya. Automated detection of hypertension using physiological signals: A review. *Int J Environ Res Public Health*, 18(11):5838, 2021.

[109] Marco V Perez, Kenneth W Mahaffey, Haley Hedlin, John S Rumsfeld, Ariadna Garcia, Todd Ferris, Vidhya Balasubramanian, Andrea M Russo, Amol Rajmane, Lauren Cheung, Grace Hung, Justin Lee, Peter Kowey, Nisha Talati, Divya Nag, Santosh E Gummidipundi, Alexis Beatty, Mellanie True Hills, Sumbul Desai, Christopher B Granger, Manisha Desai, and Mintu P Turakhia. Large-scale assessment of a smartwatch to identify atrial fibrillation. *The New England journal of medicine*, 381(20):1909–1917, 2019.

[110] Santiago JimÃ©nez-Serrano, Miguel Rodrigo, Conrado J Calvo, JosÃ© Millet, and Francisco Castells. From 12 to 1 ecg lead: multiple cardiac condition detection mixing a hybrid machine learning approach with a one-versus-rest classification strategy. *Physiological measurement*, 43(6):64003, 2022.

[111] Bambang Tutuko, Annisa Darmawahyuni, Siti Nurmaini, Alexander Edo Tondas, Muhammad Naufal Rachmatullah, Samuel Benedict Putra Teguh, Firdaus Firdaus, Ade Iriani Sapitri, and Rossi Passarella. Dae-convbilstm: End-to-end learning single-lead electrocardiogram signal for heart abnormalities detection. *PloS one*, 17(12):e0277932–e0277932, 2022.

[112] Md Belal Bin Heyat, Faijan Akhtar, Syed Jafar Abbas, Mohammed Al-Sarem, Abdulrahman Alqarafi, Antony Stalin, Rashid Abbasi, Abdullah Y. Muaad, Dakun Lai, and Kaishun Wu. Wearable flexible electronics based cardiac electrode for researcher mental stress detection system using machine learning models on single lead electrocardiogram signal. *Biosensors (Basel)*, 12(6):427, 2022.

[113] Giorgio Luongo, Felix Rees, Deborah Nairn, Massimo W Rivolta, Olaf DÃ¶ssel, Roberto Sassi, Christoph Ahlgrim, Louisa Mayer, Franz-Josef Neumann, Thomas Arentz, Amir Jadidi, Axel Loewe, and BjÃ¶rn MÃŒller-Edenborn. Machine learning using a single-lead ecg to identify patients with atrial fibrillation-induced heart failure. *Frontiers in cardiovascular medicine*, 9:812719–812719, 2022.

[114] World Health Organization. A global brief on hypertension : silent killer, global public health crisis: World health day 2013, 2023 (accessed Oct 28, 2023).

[115] G Hindricks, T Potpara, N Dagres, E Arbelo, JJ Bax, C BlomstrÃ¶m-Lundqvist, G Boriani, M Castella, GA Dan, PE Dilaveris, Laurent Fauchier, Gerasimos Filippatos, Jonathan M Kalman, Mark La Meir, Deirdre A Lane, Jean-Pierre Lebeau, Maddalena Lettino, Gregory Y H Lip, Fausto J Pinto, G Neil Thomas, Marco Valgimigli, Isabelle C Van Gelder, Bart P Van Putte, and Caroline L Watkins. 2020 esc guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the european association for cardio-thoracic surgery (eacts): The task force for the diagnosis and management of atrial fibrillation of the european society of cardiology (esc) developed with the special contribution of the european heart rhythm association (ehra) of the esc. *Eur Heart J*, 42(5):373–498, 2021.

[116] Eleni Angelaki, Georgios D. Barmparis, George Kochiadakis, Spyros Maragkoudakis, Eirini Savva, Emmanuel Kampanieris, Spyros Kassotakis, Petros Kalomoirakis, Panos Vardas, Giorgos P. Tsironis, and Maria E. Marketou. Artificial intelligence-based opportunistic screening for the detection of arterial hypertension through ecg signals. *Journal of hypertension*, 40(12):2494–2501, 2022.

[117] George E. P. Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976.

[118] A.L. Hodgkin and A.F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Bulletin of mathematical biology*, 52(1):25–71, 1990.

[119] Niels F. Otani. Bidomain model of action potential propagation, extracellular fields, and cardiac defibrillation, 2003 (accessed Oct 28, 2023).

[120] Richard FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical journal*, 1(6):445–466, 1961.

[121] J. Nagumo, S. Arimoto, and S. Yoshizawa. An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, 50(10):2061–2070, 1962.

[122] S. Alonso, M. Bär, and B. Echebarria. Nonlinear physics of electrical wave propagation in the heart: a review. *Rep. Prog. Phys.*, 79:096601 (56pp), 2016.

[123] P. C. Franzone, L. F. Pavarino, and S. Scacchi. Mathematical cardiac electrophysiology. *Springer International Publishing Switzerland*, 2014.

[124] Yanyan Claire Ji and F. H. Fenton. Numerical solutions of reaction-diffusion equations: Application to neural and cardiac models. *Am. J. Phys.*, 84:626–638, 2016.

[125] S. Golemati and K. S. Nikita (Eds). Cardiovascular computing-methodologies and clinical applications. *Springer Nature Singapore Pte Ltd. 2019*, 2019.

[126] G. W. Beeler and H. Reuter. Reconstruction of the action potential of ventricular myocardial fibres. *J. Physiol.*, 268:177–210, 1977.

[127] Ching-Hsing Luo and Yoram Rudy. A model of the ventricular cardiac action potential: Depolarization, repolarization, and their interaction. *Circulation Research*, 68:1501–1526, 1991.

[128] K. H. W. J. Ten Tusscher, D. Noble, P. J. Noble, and A. V. Panfilov. A model for human ventricular tissue. *Am. J. Physiol. Heart Circ. Physiol.*, 286:H1573–H1589, 2004.

[129] A. Bueno-Orovio, E. M. Cherry, and F. H. Fenton. Minimal model for human ventricular action potentials in tissue. *Journal of Theoretical Biology*, 253:544–560, 2008.

[130] F. Fenton and A. Karma. Vortex dynamics in three-dimensional continuous myocardium with fiber rotation: Filament instability and fibrillation. *Chaos*, 8:20–47, 1998.

[131] Flavio H. Fenton. *Theoretical Investigation of Spiral and Scroll Wave Instabilities Underlying Cardiac Fibrillation*. Phd thesis, Northeastern University, Boston, MA, 1999.

[132] Flavio Fenton and Alain Karma. Erratum: Vortex dynamics in three-dimensional continuous myocardium with fiber rotation: Filament instability and fibrillation. *Chaos*, 8:879, 1998.

[133] Stephan Rohr. Role of gap junctions in the propagation of the cardiac action potential. *Cardiovascular Research*, 62(2):309–322, 05 2004.

[134] Joshua Cooper. Intro to intra-cardiac electrograms and the ep lab, 2022 (accessed November 7, 2022).

[135] Sergei Soldatenko. Predictability in deterministic dynamical systems with application to weather forecasting and climate modelling. IntechOpen, 2017.

[136] Wei-Ching Chen. Dynamics and control of a financial system with time-delayed feedbacks. *Chaos, solitons and fractals*, 37(4):1198–1207, 2008.

[137] Hans Petter Langtangen and Svein Linge. *Finite Difference Computing With Pdes : a Modern Software Approach*, volume 16 of *Texts in Computational Science and Engineering*. Springer Open, Cham, 2017.

[138] M. D. Lesh, M. Pring, and J. F. Spear. Cellular uncoupling can unmask dispersion of action potential duration in ventricular myocardium: A computer modeling study. *Circulation Research*, 65(5):1426–1440, 1989.

[139] J. W. Cain, E. G. Tolkacheva, D. G. Schaeffer, and D. J. Gauthier. Rate-dependent propagation of cardiac action potentials in a one-dimensional fiber. *Phys. Rev. E*, 70(6):061906, 2004.

[140] R. A. Oliver and W. Krassowska. Reproducing cardiac restitution properties using the fenton-karma membrane model. *Annals of Biomedical Engineering*, 33(7):907–911, 2005.

[141] A. Peňaranda, I. R. Cantalapiedra, J. Bragard, and B. Echebarria. Cardiac dynamics: a simplified model for action potential propagation. *Theoretical Biology and Medical Modelling*, 9:50, 2012.

[142] F. H. Fenton, E. M. Cherry, H. M. Hastings, and S. J. Evans. Multiple mechanisms of spiral wave breakup in a model of cardiac electrical activity. *Chaos*, 12(3):852–892, 2002.

[143] E. G. Tolkacheva, D. G. Schaeffer, D. J. Gauthier, and C. C. Mitchell. Analysis of the fenton-karma model through an approximation by a one-dimensional map. *Chaos (Woodbury, N.Y.)*, 12(4):1034–1042, 2002.

[144] K. Gima and Y. Rudy. Ionic current basis of electrocardiographic waveforms - a model study. *Circ. Res.*, 90(8):889–896, 2002.

[145] R. H. Clayton and A. V. Holden. Propagation of normal beats and re-entry in a computational model of ventricular cardiac tissue with regional differences in action potential shape and duration. *Progress in Biophysics and Molecular Biology*, 85:473–499, 2004.

[146] O. V. Aslanidi, R. H. Clayton, J. L. Lambert, and A. V. Holden. Dynamical and cellular electrophysiological mechanisms of ecg changes during ischaemia. *Journal of Theoretical Biology*, 237:369–381, 2005.

[147] K. Q. Wang, Y. F. Yuan, Y. Y. Tang, and H. Zhang. Simulated ecg waveforms in long qt syndrome based on a model of human ventricular tissue. In *2006 Computers in Cardiology*, pages 673–676, 2006.

[148] Robert Plonsey and Roger C Barr. *Bioelectricity: A Quantitative Approach*. Springer Nature, Netherlands, 3 edition, 2007.

[149] J. Bragard, A. Witt, D. Laroze, C. Hawks, J. Elorza, I. R. Cantalapiedra, A. Penaranda, and B. Echebarria. Conductance heterogeneities induced by multistability in the dynamics of coupled cardiac gap junctions. *Chaos*, 31:073144, 2021.

[150] K. Channer and F. Morris. Myocardial ischaemia. *BMJ*, 324(7344):1023–1026, 2002.

[151] Roxana Daneshjou, Kailas Vodrahalli, Roberto A. Novoa, Melissa Jenkins, Weixin Liang, Veronica Rotemberg, Justin Ko, Susan M. Swetter, Elizabeth E. Bailey, Olivier Gevaert, Pritam Mukherjee, Michelle Phung, Kiana Yekrang, Bradley Fong, Rachna Sahasrabudhe, Johan A. C. Allerup, Utako Okata-Karigane, James Zou, and Albert S. Chiou. Disparities in dermatology ai performance on a diverse, curated clinical image set. *Science advances*, 8(32):eabq6147–eabq6147, 2022.

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

[153] I.E. Lagaris, A. Likas, and D.I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE transactions on neural networks*, 9(5):987–1000, 1998.

[154] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *arXiv.org*, 2017.

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, was created by the author of this thesis and DALL·E (OpenAI) using the prompt "a heart knitted with red yarn". A template that can be used to format a PhD thesis with this look and feel has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu.