

Unraveling the Electro- Pathophysiology of Atrial Fibrillation

Automated Atrial Fibrillation Analysis in Post- Operative Electrocardiograms

F.J. Wesselius

Technical Medicine – Sensing and Stimulation

September 2019 – July 2020



Unraveling the Electro- Pathophysiology of Atrial Fibrillation

Automated Atrial Fibrillation Analysis in Post-Operative Electrocardiograms

by

F.J. Wesselius

General information

Master program (track)

Technical Medicine (Sensing and Stimulation)

Project duration

September 2019 – July 2020

Student number

4374290

Supervisors

Clinical supervisor

Prof. Dr. N.M.S. de Groot

Technical supervisor

Dr. M.C. Roos

Daily supervisor

Drs. M.S. van Schie

Masters Graduation Committee

Core members

Prof. Dr. Ir. J. Harlaar

TU Delft, chairman

Prof. Dr. N.M.S. de Groot

Erasmus MC, medical supervisor

Dr. M.C. Roos

Erasmus MC, technical supervisor

Extra members and advisors

Dr. Y.J.H.J. Taverne

Erasmus MC, cardiothoracic surgeon

Drs. M.S. van Schie

Erasmus MC, daily supervisor

Rotterdam, Friday 19 June 2020

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.





Preface

Preface

Almost six years ago my journey to become a Technical Physician started. Together with a group of around one hundred enthusiastic fellow students I decided to enroll for the new bachelor's program Clinical Technology at TU Delft, Erasmus MC (Rotterdam) and LUMC (Leiden). Over the years, I lost count of how many times we were called "the guinea pigs of the program". After three years of education, I proudly graduated for the bachelor's program. After a short summer break, my education continued with the, also new, master's program Technical Medicine at the TU Delft, Erasmus MC and LUMC. The track *Sensing and Stimulation* started in September 2017 with only four fellow students and me. Not much later, in February, the fifth fellow student joined. Instead of "guinea pigs", more and more teachers called us "the pioneers of the program".

It was the same February that we got education on the subject of this master thesis. Alfred Schouten from TU Delft and Natasja de Groot from Erasmus MC introduced us to the methods of signal processing and cardiac electrophysiology. Not only did I really enjoy these subjects, but I also noted that Technical Medicine could play, and did already play, a major role in connecting these fields. In September 2018, the clinical internships started. Instead of attending classes, we learned both research skills and medical skills in the huge medical centers in Rotterdam and Leiden. After two internships at the rehabilitation center in Rotterdam and the Heart Lung Center in Leiden, I decided to return to the fields of cardiac electrophysiology and signal processing. My third internship took place at the department of translational electrophysiology under the supervision of Natasja. At the end of the internship, it came as no surprise when I asked whether it was possible to arrange my graduate internship at the department, after my fourth clinical internship at the transplantation center at the Erasmus MC.

In September 2019, I returned to the familiar 19th floor of the EE-building at the Erasmus MC. The year started with further developing my academic writing skills, followed by writing a systematic review, and ending with working on my thesis project. During all these phases, as well as during my third clinical internship, I was helped by the wonderful people from the research group, who I would like to thank for their time, help and motivation they gave me. First, I want to thank the group of PhD students, in particular Mathijs and Rohit, for their hours and hours of help and their infinite patience. Furthermore, I want to thank Natasja and Maarten for their time to discuss the progress and results of my project, but most of all for their dedication to help me improve my clinical and research skills. Lastly, I would like to thank some other people as well.

Thank you, Agnes Muskens, for introducing me to the clinical departments.

Thank you, Paul Knops, for helping with the technical issues.

Thank you, Yannick Taverne and Jaap Harlaar, for being part of my graduation committee.

Thank you, Mitchel, Fleur, Amne, Frederique, and Siri, for having started this journey together three years ago.

Thank you, my family, my friends, and my dog, for supporting me and giving me motivation to complete this thesis.

And last but not least, I want to thank my girlfriend, Marloes, for always being there and for her constant support.

Although my journey at TU Delft will be ending, my journey at the research department will continue. I am looking forward to continuing my journey as a PhD student and working with everyone at the research department.

*F.J. Wesselius
Rotterdam, Friday 19 June 2020*



Summary

Summary

Introduction Atrial fibrillation (AF) is the most common age-related, progressive tachyarrhythmia in the USA and in European countries. AF is associated with an increased risk of stroke, heart failure, impaired cognitive function, and increased mortality. An obstacle for optimal diagnosis and treatment is the relatively unknown (electro-)pathophysiology of AF. In combination with intra-operative cardiac mapping, accurate analysis of the AF burden using post-operative continuous rhythm registrations might provide great insight into the underlying mechanisms of AF development. However, manual analysis of these continuous rhythm registrations is both time-consuming and subject to interpretation. Therefore, the aim of this study is to develop an automated AF detection algorithm for use in the research setting.

Methods Using 6,400 manually annotated 30-seconds electrograms (ECGs) derived from the post-operative continuous rhythm registrations in the Erasmus Medical Center (Rotterdam), and 192 annotated records from standard MIT-BIH ECG databases, a classifier was developed with three output classes: *AF*, *No AF*, and *Unusable* (due to noise/artefacts). QRS-complexes were detected using a method based on the Pan-Tompkins algorithm. Subsequently, P- and T-waves were detected and features were extracted, which can be grouped into eight groups: RR-interval characteristics, peak-interval characteristics, amplitude characteristics, P-wave characteristics, T-wave characteristics, QRS-morphology characteristics, autocorrelation characteristics, and noise. Multiple classifiers were trained using a training set containing 4,800 post-operative ECGs and a hidden test set containing the remaining 1,600 post-operative ECGs. The optimal classifier in terms of accuracy was further optimized.

Results Optimal classification was achieved using boosted decision trees. For the hidden test set, this resulted in an accuracy of 96.44% (95% CI: 95.41% - 97.24%) for detection of AF with a false negative rate of 2.8% (95% CI: 1.5% - 4.9%) and a false positive rate of 3.8% (95% CI: 2.9% - 5.1%). Of all 74 misclassifications, 36 (49%) were made in the group with irregular rhythms without AF. Classification was mainly based on the RR-interval characteristics.

Conclusion An automated AF classifier based on post-operative continuous rhythm registrations for use in the research setting was proposed. Careful use of the classifier in combination with manual validation of detected AF segments makes the classifier suitable for supervised research purposes.

Index terms

post-operative atrial fibrillation, telemetry, classification, algorithms, machine learning



Table of Contents

Table of Contents

Preface	2
Summary	4
1 Introduction to Cardiac Anatomy, Atrial Fibrillation, and Electrocardiograms	11
Anatomy and physiology of the heart	12
Cardiac conduction during sinus rhythm	13
Cardiac action potentials	13
Impulse propagation	14
Pacemaker activity	15
Electrocardiogram	15
Atrial fibrillation	16
Pathophysiology	16
Diagnosis and treatment	17
Mechanisms underlying atrial fibrillation	17
Post-operative atrial fibrillation	17
Cardiac mapping	18
Focal activation patterns and endo-epicardial asynchrony	18
Goals and objectives of this study	19
2 Automated detection of Atrial Fibrillation	21
Introduction	22
Methods	23
Data	23
Phase 1: Artefact detection	24
Phase 2: Peak detection	24
Phase 3: Feature extraction	33
Phase 4: Training classifiers	41
Phase 5: Testing classifiers	41
Statistical analysis	41
Results	46
Peak detection	46
Classifier performance	47
Feature importance	47
Discussion	50
Peak detection	50

Predictive validity	51
AF segments to AF episodes	51
Other irregular rhythms	52
Limitations and future research	52
Conclusion	53
References	55
Appendices	63
Appendix I. Introduction to signal analysis and filter design using the z-transform	64
Appendix II. Peak prominence – An example	67
Appendix III. The sigmoid function	68
Appendix IV. Calculating the standard deviation of distances from points to lines through the origin using the rotation matrix – An example	70
Appendix V. Shannon Entropy – An example	73
Appendix VI. Detection of QRS-complexes	74
Appendix VII. Detection of P- and T-waves	77
Appendix VIII. Boosted decision trees & AdaBoost	82
Appendix IX. Bayesian optimization	86
Appendix X. Relative estimated feature importance	90

1



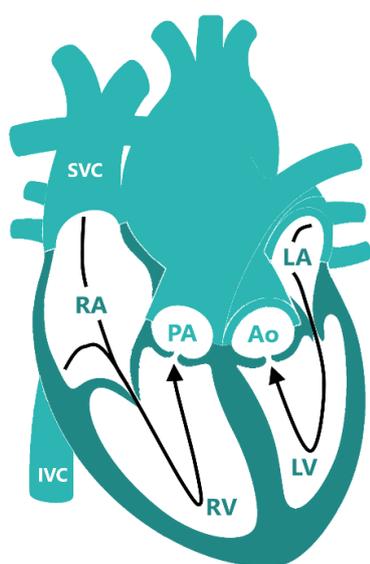
**Introduction to
Cardiac Anatomy, Atrial
Fibrillation, and
Electrocardiograms**

Introduction to Cardiac Anatomy, Atrial Fibrillation, and Electrocardiograms

Atrial fibrillation (AF) is the most common age-related, progressive tachyarrhythmia in the USA and in European countries. In Australia, Europe and the USA, the current estimated prevalence of AF is 1-4%. In 2060, the prevalence is expected to have increased 2.3-fold.¹ Furthermore, the incidence of AF after cardiac surgery lies between 10 and 60%.² AF is associated with serious complications such as stroke, heart failure, impaired cognitive function, prolonged hospital stay, and increased mortality.^{3,4} Therefore, optimal diagnosis and treatment is of the uttermost importance. However, electro-physiological mechanisms underlying AF have still not been fully unraveled.

Anatomy and physiology of the heart

To understand the mechanisms potentially underlying AF, knowledge of the anatomy and physiology of the heart is essential. The main function of the heart is to pump blood through the body. The blood flow through the heart is visualized in Figure 1. The right side of the heart receives oxygen-poor blood from the body through the superior and inferior venae cavae (SVC and IVC). The oxygen-poor blood enters the right atrium (RA). The tricuspid valve separates the RA from the right ventricle (RV). Blood flows through this valve to the RV, from where it is pumped through the pulmonary valve to the pulmonary artery (PA). The PA leads to the lungs, where the blood is oxygenated. The oxygen-rich blood arrives in the left atrium (LA) through the pulmonary veins (PV). The LA and the left ventricle (LV) are separated by the mitral valve. Oxygen-rich blood enters the LV, from where it is pumped through the aortic valve to the aorta. The aorta leads the oxygen-rich blood to the rest of the body, where oxygen is used. The cycle continues with oxygen-poor blood entering the SVC and IVC again and the circle is completed.⁵



Blood flow through the heart*

Right side (left side on figure)

1. SVC and IVC → RA
2. RA → RV
3. RV → PA → Lungs

Left side (right side on figure)

1. Pulmonary veins → LA
2. LA → LV
3. LV → Ao → Body

* SVC = superior vena cava; IVC = inferior vena cava;
RA = right atrium; RV = right ventricle;
PA = pulmonary artery; LA = left atrium;
LV = left ventricle; Ao = aorta;

Figure 1 – Blood flow through the heart

Cardiac conduction during sinus rhythm

Electrical activation of cardiomyocytes facilitates contraction of the heart muscle. Cardiac electrophysiology describes the generation and conduction of action potentials through the heart muscle. In 2012, Lederer extensively described the cardiac electrophysiology during normal sinus rhythm (SR).⁶ Cardiac action potentials originate from the sinoatrial (SA) node, which is located in the RA. Bachman's bundle facilitates conduction from the RA to the LA, hence also the LA is activated from the SA node. Due to the presence of the fibrous atrioventricular ring between atria and ventricles, no direct conduction from atria to ventricles is possible. The sole pathway for the impulse to travel from atria to ventricles is the atrioventricular (AV) node. Compared to the conduction velocity in the atria (~1m/s), the conduction velocity through the AV node is 20-fold lower (~0.05m/s), hence a delay is caused between contraction of the atria and contraction of the ventricles. From the AV node, the impulse travels over the interventricular septum, which is depolarized from left to right, using the bundle of His and the left and right bundle branch. Subsequently the left and right ventricle are activated from apex to base using the Purkinje fibers.

Cardiac action potentials

Cardiac action potentials at cellular level form the basis of electrical activation. These action potentials are caused by four membrane currents, resulting in depolarization and repolarization. Ca^{2+} current (I_{Ca}) causes depolarization in the SA node and AV node and facilitates contraction in cardiomyocytes. In contrast, Na^{+} current (I_{Na}) solely causes depolarization in the atrial and ventricular muscle and Purkinje fibers. K^{+} current (I_{K}) repolarizes all cardiomyocytes. Pacemaker current (I_{f}) facilitates pacemaker activity in the SA node, AV node and Purkinje fibers. Pacemaker activity results in generation of new cardiac action potentials.

As illustrated in Figure 2a-b for sinus node cells and ventricular muscle cells, five phases of the membrane potential during the cardiac action potential are distinguished:

Phase 0. Depolarization

The membrane potential increases due to an inward I_{Ca} (slow increase) or a combined inward I_{Ca} and I_{Na} (fast increase).

Phase 1. Early repolarization

The membrane potential decreases rapidly due to inactivation of I_{Na} and/or I_{Ca} .

Phase 2. Plateau phase

The membrane potential is constant, due to inflow of Ca^{2+} and outflow of K^{+} .

Phase 3. Repolarization

The membrane potential decreases again to the resting potential depending on an outward I_{K} .

Phase 4. Resting phase

In SA and AV nodal cells, pacemaker activity is produced by changing I_{K} , I_{Ca} and I_{f} . In Purkinje fibers, pacemaker activity is produced solely by I_{f} .

During depolarization and repolarization, cardiac cells become refractory to other activation. When Na^+ - and Ca^{2+} -currents are inactivated, new stimuli have no effect on the action potential. This phase is called the effective refractory period. After this, the relative refractory period starts, when the Na^+ - and Ca^{2+} -currents recover. During this phase, action potentials can be produced, but will be smaller due to the outflow of K^+ .

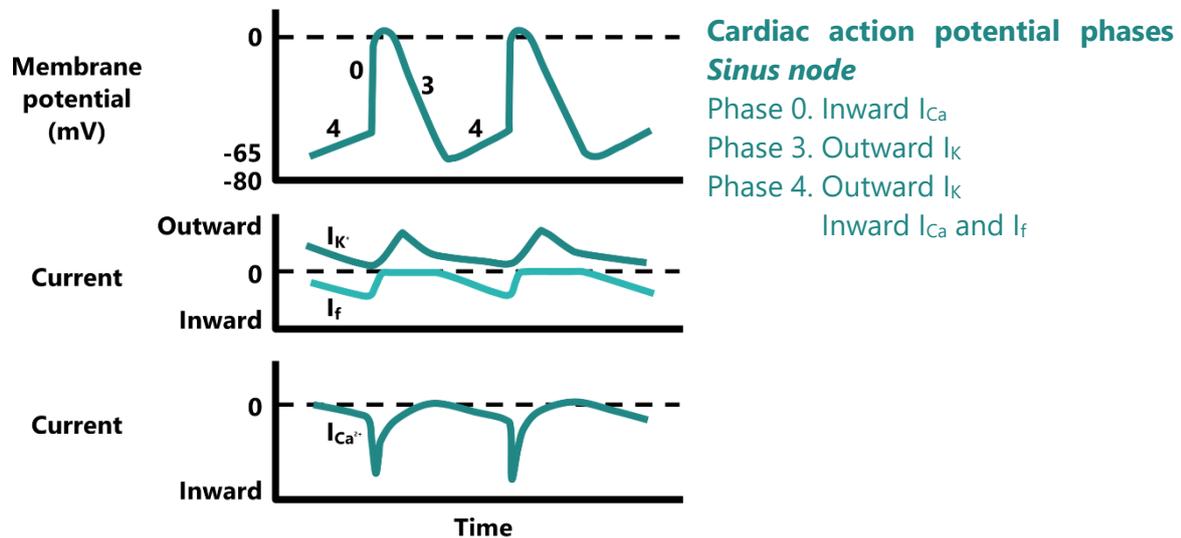


Figure 2a – Cardiac action potential phases in sinus node cells, as further explained in text (adapted version of figure by Boron and Boulpaep⁶)

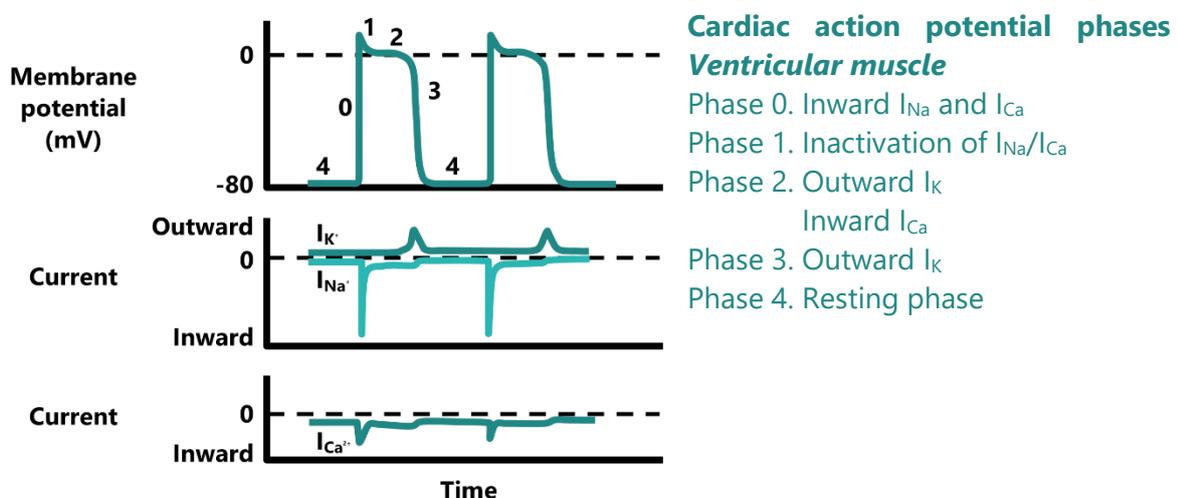


Figure 2b – Cardiac action potential phases in ventricular muscle cells, as further explained in text (adapted version of figure by Boron and Boulpaep⁶)

Impulse propagation

Cardiac cells are linked through gap junctions, which enable electrical current to flow between cells. When there is a voltage difference between two cells, an electrical current will flow, which is inversely proportional to the resistance of the gap junction and proportional to the voltage difference. The small depolarizing currents between cardiac cells may result in neighboring cells reaching their excitation threshold and hence firing an action potential. This way, impulses propagate through the heart.

Pacemaker activity

Pacemaker activity originates from three cardiac structures: SA node, AV node and Purkinje fibers. In rest, the SA node has an intrinsic rate of ~60 beats per minute (bpm) or faster. A decreasing outward current (I_k) and two increasing inward currents (I_{Ca} and I_f) result in slow depolarization. Although the intrinsic rate of the SA node is ~60bpm, these membrane currents are affected by, among others, acetylcholine, epinephrine, and norepinephrine. Also, therapy focused on controlling the heart rate is aimed at these currents, hence the firing rate of the SA node can be subject to external conditions. The intrinsic rate of the AV node is ~40bpm. Since the intrinsic rate is lower than the rate of the SA node, the heart rate is not set by the AV node. However, in case the SA node fails, the AV nodal rhythm characterizes the heart rhythm. Finally, the Purkinje fibers have the slowest intrinsic pacemaker rate of ~20bpm or less.

Electrocardiogram

In the 1900s, Willem Einthoven laid the foundation for the currently standard clinical tool to measure the electrical activity of the heart, the electrocardiogram (ECG).^{6, 7}

Electrode measurements

Using electrodes, the extracellular voltage differences are measured. During the depolarization phase, positive ions enter the cardiac myocytes, hence result in a positive deflection when measured at the intracellular level. Therefore, the difference in *intracellular* voltages for an approaching depolarization wave (to the positive electrode) will be negative. The *extracellular* current is equal to the intracellular current, but opposite in direction. Therefore, when measuring a positive voltage difference using ECG, an approaching depolarization wave is measured.

Constructing an electrocardiogram

Using two electrodes on the upper extremities, two electrodes on the lower extremities (one ground electrode), and six electrodes on standard locations across the chest, a 12-lead ECG can be obtained, as visualized in Figure 3. A lead is composed of one positive electrode (+) and one or multiple negative electrodes (-). The six limb leads define axes with certain angles in the frontal plane, where 0° is defined as the horizontal line pointing left and positive angles are clockwise rotations:

- Lead I.* Right upper extremity to left upper extremity; 0°.
- Lead II.* Right upper extremity to left lower extremity; 60°.
- Lead III.* Left upper extremity to left lower extremity; 120°.
- Lead aVR.* Middle of the heart to right upper extremity; -150°.
- Lead aVL.* Middle of the heart to left upper extremity; -30°.
- Lead aVF.* Middle of the heart to left lower extremity; 90°.

aVR, aVL and aVF are augmented leads, comparing one limb electrode to the average of the other two. This way, angles of -150°, -30° and 90° are constructed.

Additionally, six precordial leads V1 to V6 are constructed using the average of the three limb electrodes in combination with the six chest electrodes. The average of three limb electrodes represents the middle of the heart. Comparing each of the six chest electrodes

with the average of the limb electrodes provides insight in the transversal plane, instead of the frontal plane, as visualized in the right panel of Figure 3.

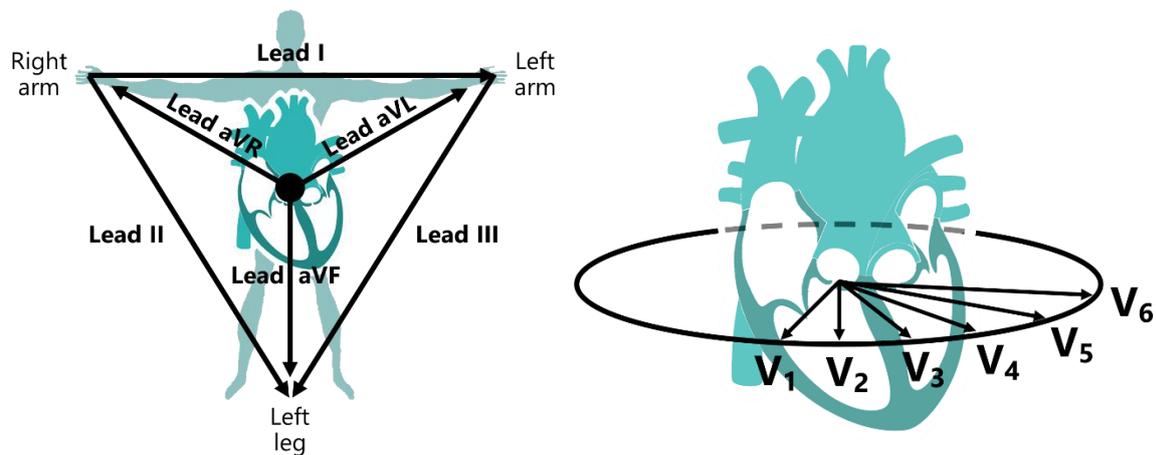


Figure 3 – Standard 12 ECG leads; left: limb leads and augmented leads, right: precordial leads

Interpretation of the electrocardiogram

The ECG is characterized by five waves: P-, Q-, R-, S- and T-waves, as visualized for lead II in Figure 4.⁶ P-waves, the QRS-complex and the T-wave are caused by atrial depolarization, ventricular depolarization, and ventricular repolarization, respectively. Atrial repolarization does not cause a specific wave in the ECG, because it coincides with ventricular depolarization.

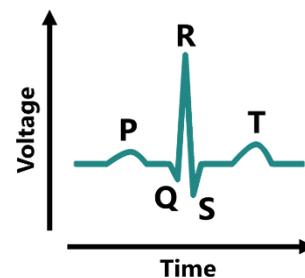


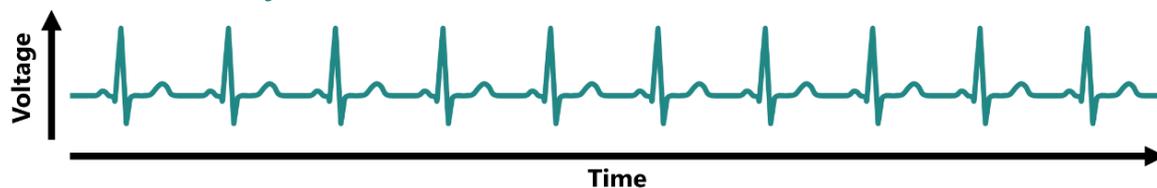
Figure 4 – Peaks in ECG (lead II)

Atrial fibrillation

Pathophysiology

AF is characterized by continuous rapid chaotic atrial activity and an irregular ventricular rhythm.⁸ Since the AV node slows down the potential between atria and ventricles, the high atrial frequency is not propagated to the ventricles. As visualized in Figure 5, this results in irregular intervals between R-waves and absence of P-waves.

A) Normal sinus rhythm



B) Atrial fibrillation

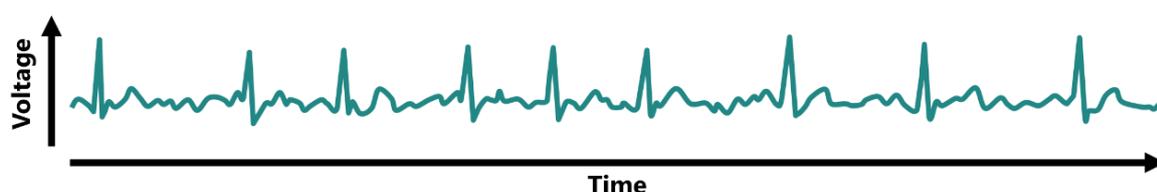


Figure 5 – A) ECG (lead II) showing 10 beats of normal sinus rhythm; B) ECG (lead II) showing atrial fibrillation characterized by absence of P-waves and irregular intervals between R-waves.

Diagnosis and treatment

Typical symptoms of AF include palpitations, tachycardia, fatigue, weakness, dizziness, light-headedness, reduced exercise capacity, increased urination, and mild dyspnea. In combination with the medical complaints of the patient and echocardiography, ECG is used to determine whether a patient has AF or does not have AF.⁹ AF is classified into four categories based on duration and treatment goals: paroxysmal (<7 days), persistent (>7 days), long-standing persistent (>12 months), and permanent (no longer pursuing rhythm control).¹⁰

Treatment options for AF can be subdivided into two categories: rate control and rhythm control. Treatment focused on rate control includes drugs that block the AV node (beta blockers) in order to control the heart rate.^{9, 11} Treatment focused on rhythm control includes anti-arrhythmic drugs, catheter ablation and/or surgical treatments in order to recover SR.^{9, 11} Another way to recover SR is using electrical or chemical cardioversion.⁹ Rate control is the initial therapy of choice in asymptomatic patients, because of possible side effects of the antiarrhythmic drugs and non-inferior outcome rates compared to rhythm control.^{9, 11, 12} Rhythm control is preferred in patients where rate control fails, in symptomatic patients and in patients less than 65 year old.^{9, 11}

Mechanisms underlying atrial fibrillation

Exact mechanisms underlying AF are not yet entirely understood. Multiple theories have been proposed, discussing both initiation and persistence of AF. Single foci can be a trigger for the initiation of AF. These foci are mostly found near the PV, but appear throughout the entire atria.¹³ However, single foci do not explain the persistence of AF. Some widely supported theories for the persistence of AF are the multiple wavelet hypothesis and the re-entrant circuits and rotors theories.¹⁴ During AF, multiple activation waves are propagating through the atrial wall. These waves encounter areas of different refractoriness, which might result in re-entrance.¹⁴ These re-entrant circuits are characterized by short cycle lengths, which may result in the development of functional blocks and slow conduction.¹³ Also, when multiple wave fronts collide, this might result in the creation of new wave fronts, hence result in persistence of AF. These mechanisms are supported by atrial and electrical remodeling. Atrial remodeling results in structural changes of the atria that can lead to the initiation and persistence of AF. Electrical remodeling entails changes in refractoriness of the heart tissue, hypothesized to occur due to changes in calcium channels.¹⁵ Also, the autonomic nervous system plays a role in the initiation and persistence of AF.¹³ All these mechanisms are supported by several studies. Most likely, AF is not caused by a single mechanism, but by a combination of multiple mechanisms.¹⁶

Post-operative atrial fibrillation

After surgery, 0.3 to 29% of the patients develop post-operative AF (PoAF).² Most patients developing PoAF do so immediately post-operatively or on the second day after surgery.¹⁷ Most of the PoAF episodes are self-limited within one day after the onset and few patients experience more than two episodes.^{18, 19} After cardiac surgery – more specifically coronary artery bypass grafting (CABG) and valve surgery – the reported incidence of AF lies

between 10 and 60%.² Abnormal atrial conduction and refractoriness caused by incisions and ischemia most likely are key factors in the higher incidence after cardiac surgery compared to other surgical procedures.²⁰ However, also differences in patient population (e.g. age and pre-existing comorbidities), surveillance and detection in the peri- and post-operative setting contribute to the higher incidence of PoAF.¹⁸

Cardiac mapping

Recently, mapping of atrial conduction patterns using advanced techniques has provided more insights in the electro-pathophysiology of AF. One of the methods to perform cardiac mapping was developed in the Erasmus Medical Center (Erasmus MC) in Rotterdam (The Netherlands). Using arrays containing 128 or 192 unipolar electrodes, the atria are mapped at high spatial resolution and 1kHz temporal resolution during open-heart surgery.²¹ Using the unipolar electrograms recorded at every electrode, local activation times can be determined and visualized, which makes it possible to follow the propagation of waves through the atria. More recently, simultaneous endo- and epicardial mapping was introduced. Two similar electrode arrays are used, as shown in Figure 6. By positioning one electrode array at the endocardium (inner layer of the heart), and another electrode on the exact opposite epicardium (outer layer of the heart), conduction patterns in both the endo- and epicardium can be analyzed simultaneously.²²



Figure 6 – Electrode arrays used for simultaneous endo-epicardial mapping

Focal activation patterns and endo-epicardial asynchrony

One of the observations done during AF are focal activation patterns (FAPs), also known as breakthrough waves. These FAPs are defined as activation appearing at a local site within the recording area with a radial activation pattern.²³ Various mechanisms underlying these FAPs have been proposed, including transmural conduction (from epi- to endocardium or vice versa) and ectopic focal discharge.²⁴ In 2010, an association between the presence of FAPs and the persistence of AF was revealed. During high-resolution epicardial mapping, a four-fold higher incidence of FAPs was observed in patients with longstanding persistent AF compared to patients with pacing-induced AF (respectively 0.47 vs. 0.11/AF-cycle/cm²).²⁵ Therefore, FAPs are assumed to play a key role in the underlying mechanisms of AF.

During recent years, large amounts of mapping data have been analyzed, resulting in several new insights about the electro-pathophysiology of AF. One major breakthrough was the proof of endo-epicardial asynchrony during AF, which implies the asynchronous activation of the endocardium and epicardium.²⁶ This finding resulted in the hypothesis of possible transmural wave propagation from the endocardium to the epicardium and vice versa. Also, epicardial FAPs were described during SR, showing rS- and RS-morphology. It is important to note that these R- and S-waves are not the waves as described for QRS-complexes in the ECG. For unipolar electrograms, R-waves indicate wave propagation towards the electrode and S-waves indicate wave propagation in a direction away from the electrode. Therefore, rS- and RS-morphologies of these FAPs again indicate transmural conduction towards the epicardium.²¹ Based on the wide spatial distribution of FAPs, non-repetitiveness, non-prematurity and breakthrough potentials with R-waves, transmural conduction is increasingly assumed to be part of the underlying mechanism of FAPs.²⁷

Goals and objectives of this study

Analysis of AF episodes in post-operative continuous rhythm registrations is essential to draw conclusions on the relation between intra-operative mapping and PoAF. Currently available AF detection algorithms are too sensitive to other irregular rhythms and artefacts. The alternative is to analyze the continuous rhythm registrations manually. However, due to the large amount of data, this is both time-consuming and subject to interpretation. This study focusses on the development of an AF classifier to optimize automated AF detection and get information on the AF burden in terms of episode duration, number of episodes, and proportion of time an individual is in AF during the monitoring period.

2



Automated Detection of Atrial Fibrillation

Automated Detection of Atrial Fibrillation

Introduction

Patients undergoing cardiac surgery commonly develop atrial fibrillation (AF) post-operatively.² Continuous rhythm registrations of hospitalized patients are acquired using telemetry devices. Accurate detection of AF episodes in these continuous rhythm registrations is essential to improve treatment of patients. Additionally, accurate detection of AF segments provides further information on the AF burden in terms of AF episode duration, number of episodes per day, and proportion of time an individual is in AF during the monitoring period.²⁸ This data not only supports treatment, but also provides new outcome measures which can be used in scientific research.

Manual analysis of long-term continuous rhythm registrations is both time-consuming and subject to interpretation. To detect AF in telemetry data, an accurate automated detection algorithm is essential. Several available detection algorithms which are currently being used for automated classification of continuous rhythm registrations fail to accurately detect AF episodes in post-operative rhythm registrations, due to the presence of other irregular rhythms (e.g. atrial and ventricular premature beats), noise, and artefacts due to movement.

A systematic literature study has shown that research on automated detection methods for AF in surface electrocardiogram (ECG) measurements has been increasing over the past years.²⁹ However, no single classification algorithm or feature selection method was found superior on any given data set. Many studies report testing accuracies for AF detection ranging up to 100%, using widely available ECG databases for training and validation of the classification methods. In contrast to telemetry data, these standard ECG databases mostly contain clean ECG records without artefacts. Due to movement of the patient, poor contact between electrode and skin, and interferences from other devices, noise and artefacts are major obstacles in the analysis of post-operative continuous rhythm registrations.

Therefore, the goal of this study is to develop an accurate classification method for detection of AF episodes and calculation of AF burden in post-operative continuous rhythm registrations. The research setting differs from the clinical setting, hence the intended use should be considered when developing a classifier. Requirements were set based on the false positive rate (FPR) and false negative rate (FNR). FPR is calculated by dividing the amount of false positive classifications by the total amount of samples which are negative for the condition: $(100\% \cdot) \Sigma(\text{false positive})/\Sigma(\text{condition negative})$. FNR is calculated by dividing the amount of false negative classifications by the total amount of samples which are positive for the condition: $(100\% \cdot) \Sigma(\text{false negative})/\Sigma(\text{condition positive})$. Multiple studies show a relation between the AF burden and the risk of stroke and systemic embolism.^{30, 31} Therefore, missing a single short AF episode would not impact the treatment considerations in the clinical setting. Also, manual validation of detected AF segments to screen for false positive classifications is time-consuming, hence

unfeasible in clinical practice. Therefore, in the clinical setting, a high FPR would be problematic, since the predicted AF burden would be higher, potentially resulting in overtreatment. On the other hand, a minor amount of false negative classifications would be accepted, since this would not impact the treatment considerations. In the research setting, the aim is to unravel the exact electro-pathophysiology underlying AF. Single short AF episodes could also contribute to new insights. In addition, manual validation of AF episodes is more feasible in a research setting. For example, for every AF episode, a certain number of random ECG segments could be manually validated per hour of AF duration. Therefore, a higher FPR is accepted. On the other hand, the aim would be to detect all AF episodes, hence the FNR should be lower. Since the aim of the current study was the development of an AF detection algorithm for a research setting, requirements to the classifier's performance were arbitrarily set to an FPR of at most 5% and an FNR of at most 3%.

Methods

Data

The AF detection algorithm was developed using post-operative continuous rhythm registrations obtained from patients after cardiac surgery for valvular heart disease and/or ischemic heart disease in the Erasmus Medical Center, Rotterdam. This data consists of 6,400 standard 30-second 12-lead recordings (sample rate 200Hz), but only lead II was used in the present study. Manual annotations for *AF*, *No AF* and *Unusable* (due to noise/artefacts) segments were available. Additionally, the ECG segments without AF were annotated as regular or irregular, based on the assessor's judgement.

Validation of peak detection was performed using five standard MIT-BIH databases³²:

- ❖ **Normal Sinus Rhythm Database (NSRDB)**³²
containing 18 24-h ECGs with normal sinus rhythm with annotated ventricular activity,
- ❖ **Long-term ECG Database (LTDB)**³²
containing 7 24-h ECGs with annotated ventricular activity,
- ❖ **Atrial Fibrillation Database (AFDB)**^{32, 33}
containing 23 10-h ECGs with annotated ventricular activity,
- ❖ **Arrhythmia Database (MITDB)**^{32, 34}
containing 48 ½-h ECGs with annotated ventricular activity, and
- ❖ **QT Database (QTDB)**^{32, 35}
containing 105 15-minute ECGs, of which 96 contain atrial and ventricular annotations.

Development of the proposed AF classifier consisted of five phases:

Phase 1. Artefact detection

Phase 2. Peak detection

Phase 3. Feature extraction

Phase 4. Training classifiers

Phase 5. Testing classifier

All data processing was performed using the raw output of the ECG recording system in arbitrary units. The same steps apply for data converted to (milli)volt. The algorithm uses ECG segments with a duration of 30 seconds. Therefore, if the ECG has a longer duration, it should be split into multiple segments of exactly 30 seconds.

Phase 1: Artefact detection

First, artefacts in the ECG records resulting in unusable signal were detected and excluded from analysis. Two types of these artefacts were present in the records:

- Type 1. Clipping or no signal; characterized by a rapid increase/decrease in signal amplitude, a series of constant values, and a rapid change of signal amplitude in opposite direction. Clipping occurs when the measured voltage is outside the measurement range of the device. The source of these artefacts is, e.g. no contact between electrode and skin (extremely low voltage output resulting in no signal) or cardioversion (extremely high voltage output resulting in clipping).
- Type 2. Single spikes; characterized by a rapid increase/decrease in signal amplitude followed by a rapid change of signal amplitude in opposite direction. An example of the source of these artefacts is cardiac pacing.

Artefacts were commonly surrounded by either a distorted QRS-complex due to baseline drift or incomplete QRS-complexes. Therefore, artefacts were excluded with a margin of 300ms (= 60 samples with a sampling rate of 200Hz) on both sides. Other artefacts, caused by for example slight patient movement, are excluded in later phases, since these segments could contain useful ECG signal.

Phase 2: Peak detection

Peak detection was split into four phases. First, the QRS-complexes were detected. Using the location of these QRS-complexes, P- and T-waves were located between subsequent QRS-complexes. Next, peak detection was validated using the annotations in the MIT-BIH databases. Lastly, segments containing excessive noise were excluded from analysis.

Phase 2a: QRS-complex detection

Pan-Tompkins Algorithm

Windows containing QRS-complexes were detected based on the algorithm as described by Pan and Tompkins, also known as the Pan-Tompkins algorithm.³⁶ The aim of the Pan-Tompkins algorithm is to amplify the QRS-complexes and reduce other elements (P-waves, T-waves and noise). The algorithm consists of four steps, as visualized in Figure 7. A bandpass filter is applied to eliminate high-frequency noise and baseline drift from the signal. Using a derivative filter and squaring the resulting signal, segments with steep slopes are amplified. Lastly, moving-window integration is used to detect the edges of the QRS-windows. Original parameters used by Pan and Tompkins for the bandpass filter are lower and higher cut-off frequencies of 5 and 15Hz, respectively. Derivative filtering was performed using a standard five-point derivative filter. The original Pan-Tompkins algorithm uses a derivative filter with difference equation $y[nT] = \frac{1}{8T}(-x[nT - 2T] - 2x[nT - T] + 2x[nT + T] + x[nT + 2T])$, where T is the sampling period, $x[n]$ is the original

signal and y is the filtered signal. The corresponding transfer function in the z -domain is $H(z) = \frac{1}{8T}(-z^{-2} - 2z^{-1} + 2z + z^2)$, where T is the sampling period and z is a complex variable defined as $re^{j\Omega}$. r and Ω describe the magnitude and the angle of z , respectively.* The size of the original moving window was set to 150ms.

Thakor et al. have analyzed the power spectrum of ECGs and have shown that most of the energy of the QRS complex is between 3Hz and 40Hz.^{37, 38} The cut-off frequencies for the bandpass filter resulting in an optimal signal-to-noise ratio depend on both the power spectrum of the QRS-complexes and the power spectrum of the noise. Empirical analysis showed that a higher cut-off frequency of 25Hz for the bandpass filter and a moving window of 200ms improved the detection rate of QRS-complexes in the records used in this study.

The outcome of the Pan-Tompkins algorithm (ECG_{P-T}) provides insight into the temporal locations of QRS-complexes. QRS-windows were extracted by applying a threshold set to the overall mean of ECG_{P-T} . QRS-windows were extracted as windows with ECG_{P-T} larger than this threshold.

R-peak detection

For R-peak detection within the QRS-windows, peak prominence was used. Peak prominence is a measure for the peak amplitude relative to other surrounding peaks.[†] Prominence was used instead of amplitude, to avoid mistakenly detecting high-frequency noise as R-peaks. Peak amplitude of peaks caused by noise can be high. However, peak prominence of such peaks is in general lower, because the peak amplitude is similar to the amplitudes of surrounding peaks. Within each QRS-window, one R-peak was detected, defined as the positive peak with highest prominence, as illustrated in the middle panel of Figure 8.

Missing QRS-complexes

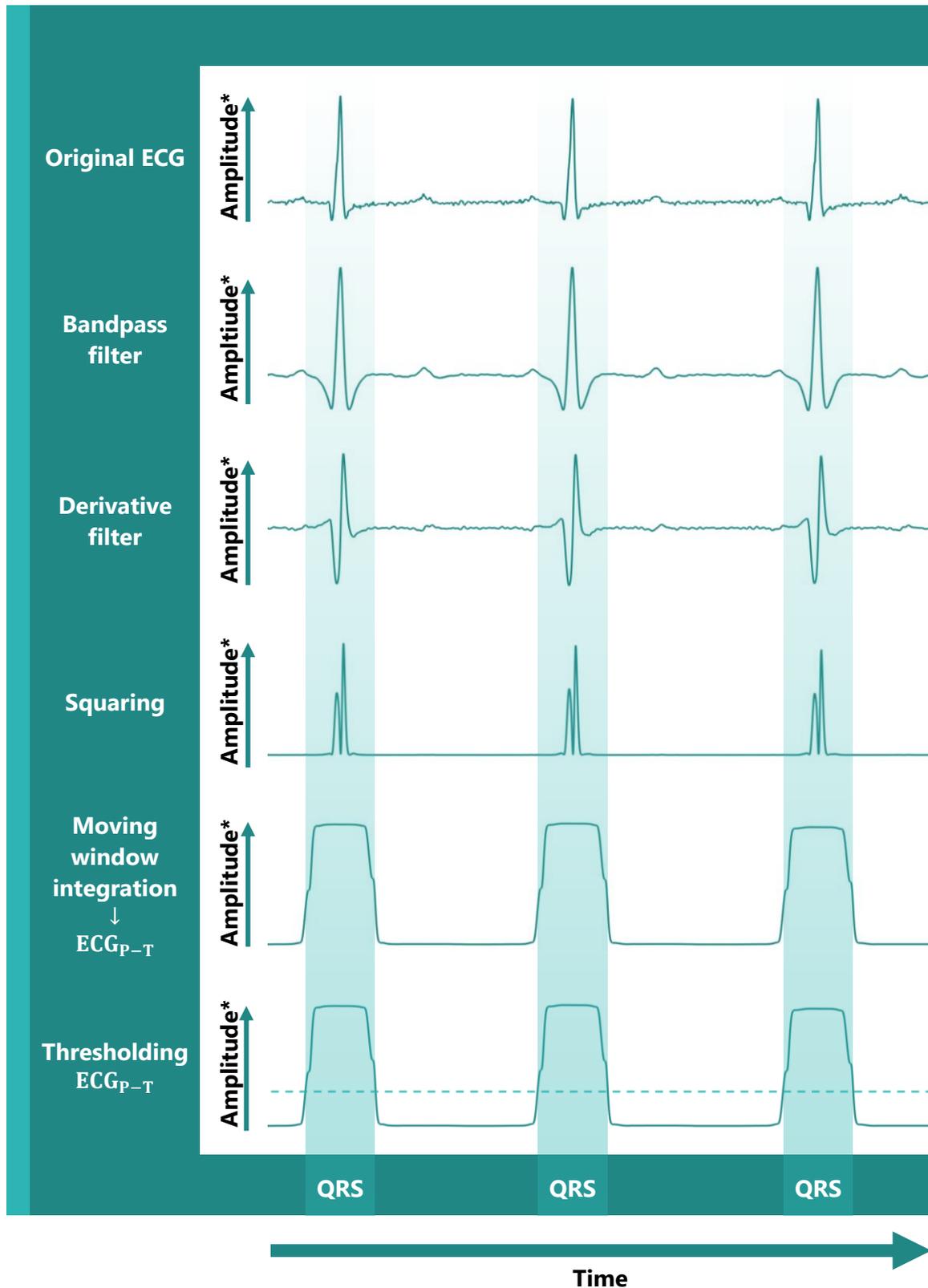
Time intervals between R-peaks were calculated. In case the interval between subsequent R-peaks was larger than 120% of the median time interval between R-peaks in the entire 30-seconds segment, ECG_{P-T} between these two subsequent R-peaks was again searched for R-peaks with a threshold which was 30% lower than the original threshold.

Q- and S-peak detection

As visualized in the lower panel of Figure 8, Q- and S-peak detection was done in a similar way as R-peak detection. Within each QRS-window, the Q-peak was detected prior to the R-peak, as first minimum with a prominence larger than 95% of the maximum prominence of all minima prior to the R-peak. The S-peak was detected after the R-peak, as first minimum with a prominence larger than 95% of the maximum prominence of all minima following the R-peak. No Q- or S-peak was detected if no local minimum was present prior to the R-peak or after the R-peak, respectively.

* Further information on filter design and the z -transform is provided in Appendix I.

† Further explained with an example in Appendix II.



**For illustrative purpose only; scaling is not equal between the different panels*

Figure 7 – Pan-Tompkins algorithm used to amplify and detect windows with QRS-complexes, as further explained in text.

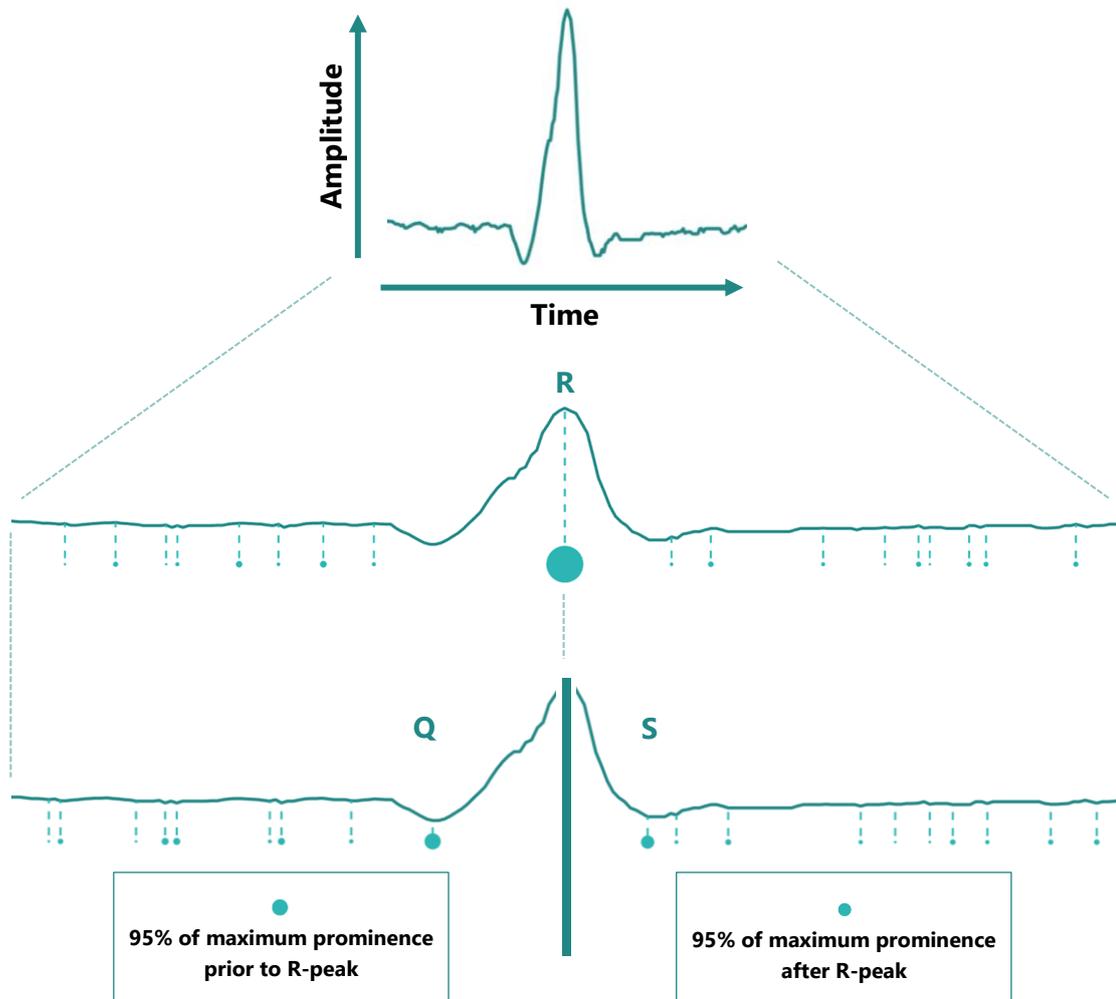


Figure 8 – Detection of Q-, R- and S-peaks using peak prominence. Upper panel: original ECG; middle panel: detection of R-peak as the local maximum with the highest prominence; lower panel: detection of Q- and S-peaks as the first local minima (starting from the R-peak) with a prominence of at least 95% of the maximum prominence prior to and after the R-peak, respectively.

Adaptive thresholding

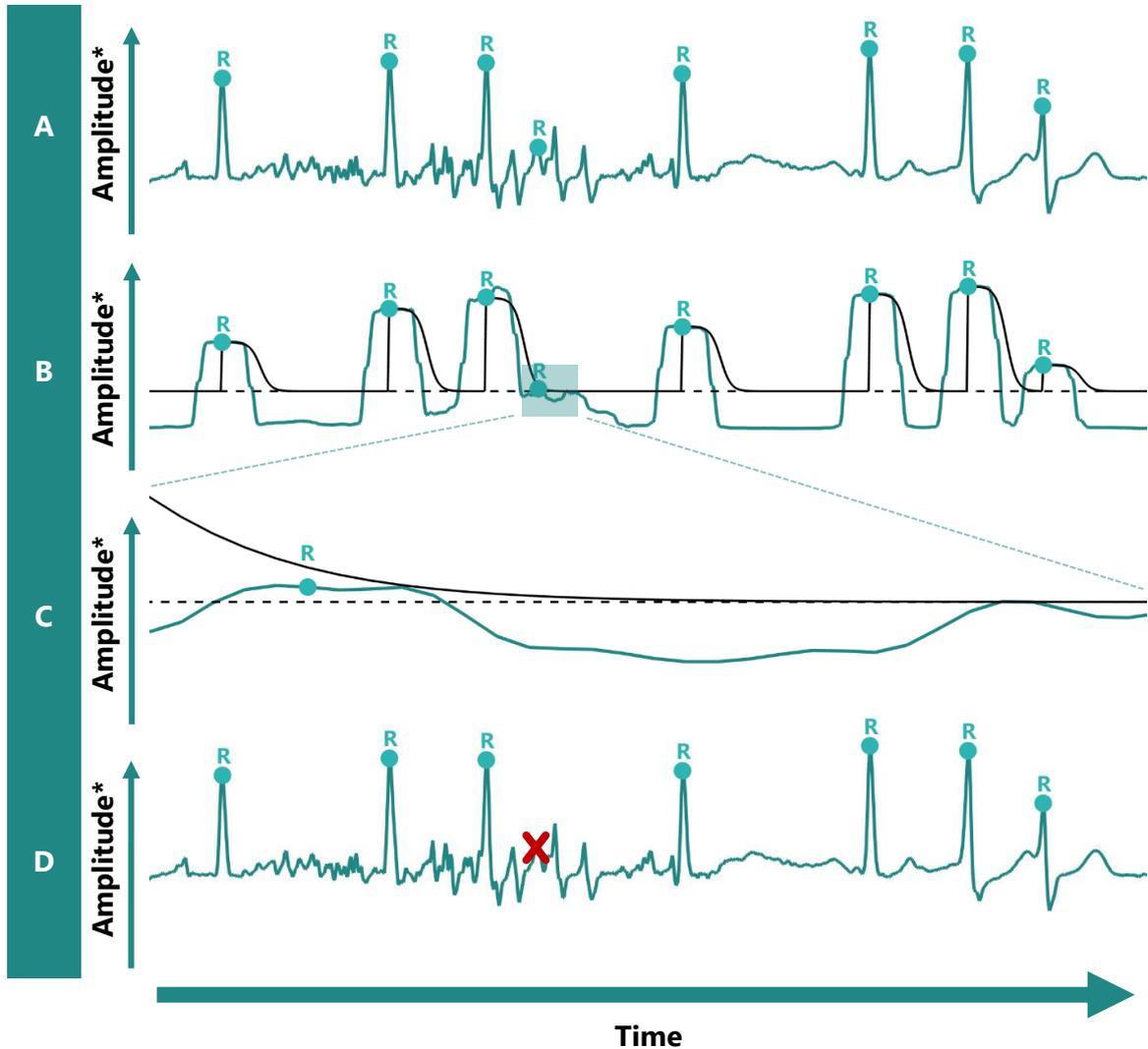
In a short time interval after ventricular activation, cardiomyocytes are in a refractory period. During this period, the inward sodium and calcium currents of the cardiomyocytes are inactivated, which prevents reactivation by action potentials.⁶ Therefore, detecting another QRS-complex in this time interval is highly unlikely. In panel A of Figure 9, an ECG segment with an incorrectly classified R-peak is shown. Using the locations of all detected R-peaks, correction was applied for such false positive detections. Again, a threshold was applied to ECG_{p-T} . As visualized in panel B of Figure 9, instead of simply using the mean of ECG_{p-T} , the new threshold was set based on the location of the previously detected R-peaks, following the sigmoid function[‡]:

$$T_{adaptive}[m] \Big|_{R_n}^{R_{n+1}} = \frac{-(ECG_{p-T}[R_n] - T_{old}[R_{n+1}])}{1 + e^{-B \cdot (m - R_n - C)}} + T_{old}[m] + (ECG_{p-T}[R_n] - T_{old}[R_{n+1}]) \quad (\text{Eq. 1})$$

- $T_{adaptive}$ = adaptive threshold
- R_n = location of previous R-peak
- R_{n+1} = location of next R-peak
- m = distance from R_n in unit of time samples
- ECG_{p-T} = result of the Pan-Tompkins algorithm
- T_{old} = threshold as previously used between R_n and R_{n+1}
- B = scaling factor; constant value, arbitrarily set at 50
- C = offset in time; constant value, arbitrarily set at 0.15)

The sigmoid curve was drawn from an R-peak to the next R-peak, as visualized in panel B of Figure 9. The denominator in the equation represents the difference between the lowest point of the threshold and the highest point of the adaptive threshold (= ECG_{p-T} at the previous R-peak). The numerator represents the slope and delay of the sigmoid curve. The added factor lifts the sigmoid curve to the level of the old threshold. As m tends to infinity, the numerator will be almost 1. Therefore, $T_{adaptive}$ will be almost equal to the old threshold. In reality this will not happen, since the equation is evaluated for m from an R-peak location to the next R-peak location. As m tends to R_n , the numerator will be large compared to the denominator. This will result in a $T_{adaptive}$ which is almost equal to $ECG_{p-T}[R_n]$. Detected QRS-complexes of which ECG_{p-T} was below the adaptive threshold were excluded, as visualized in panel C and D of Figure 9. After exclusion of a QRS-complex, the adaptive threshold was updated to correct for the removed QRS-complex.

[‡] Further information on the sigmoid function is provided in Appendix III.



** For illustrative purpose only; scaling is not equal between the different panels*

Figure 9 – Applying the adaptive threshold. Panel A shows an ECG segment with eight detected R-peaks, of which the fourth is incorrectly detected as R-peak. Panel B shows the corresponding result of the Pan-Tompkins algorithm. The dashed black line (---) shows the original threshold for detecting windows for QRS-detection. The solid black line (—) shows the adapted threshold. Panel C zooms in on the misclassified R-peak. The point is above the original threshold, but underneath the adapted threshold, hence as visualized in panel D, this R-peak is excluded.

Phase 2b: P- and T-wave detection

The algorithm for P- and T-wave detection was based on the method as described by Elgendi et al.³⁹

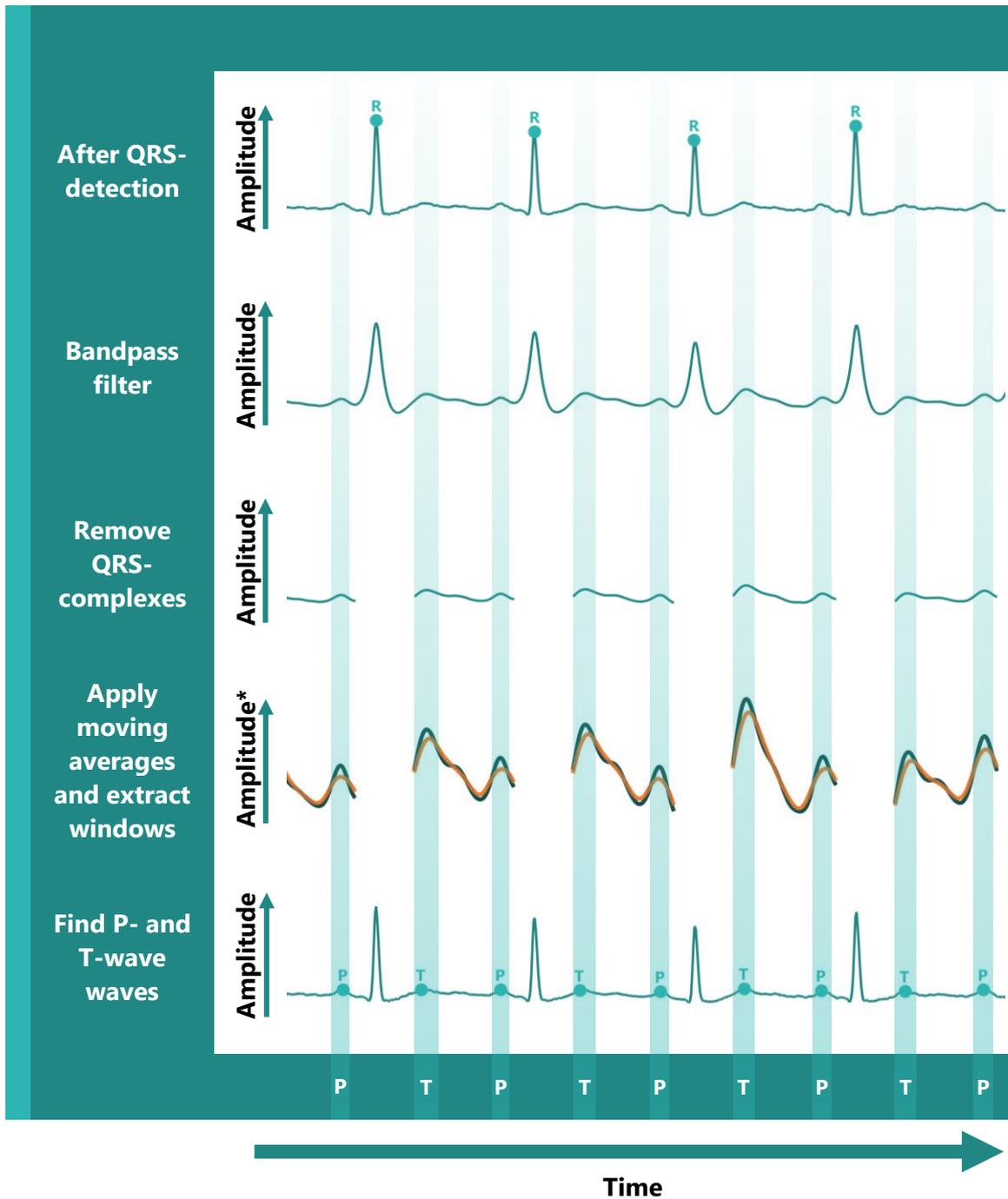
In Figure 10, the methods for detection of P- and T-waves are shown. First, a second-order Butterworth filter was applied with lower and higher cut-off frequencies of 0.5Hz and 6Hz, respectively. These cut-off frequencies were chosen, because the main frequencies of P- and T-waves are between these boundaries.³⁹ QRS-complexes were then removed from the signal. Using two moving-average windows with different lengths on the original ECG signal, windows potentially containing P- and T-waves were detected. The normal width of the P-wave is 110ms.³⁹ The size of the first moving window was set to half of the normal P-wave width (55ms). The second moving window was used as a threshold. The size of the second moving window was set to 110ms, resulting in lower values for P- and T-waves compared to the first moving window.

Next, potential P- and T-windows were extracted where the first moving average was larger than the second moving average. Noise can also be characterized by short peaks, thus can be mistakenly identified as P- or T-waves. To avoid such false positive detections, windows shorter than 25% of the normal width of the P-wave (110ms) were excluded from P- and T-wave detection.

Subsequently, the bandpass-filtered ECG was used for peak detection. In the P- and T-windows between subsequent QRS-complexes, all minima and maxima were determined. In general, T-waves have higher amplitudes compared to P-waves. Therefore, T-waves were detected in the window with the largest minimum or maximum, whichever had a larger amplitude. If no peaks were detected within the windows, no P- and T-waves were detected. Windows between a detected T-wave and the next QRS-complex were used for P-wave detection. First, all maxima and minima within these windows were detected. Next, maxima or minima with an amplitude larger than 80% of the maximum peak amplitude were detected as P-wave.

Phase 2c: Validation of peak detection

The locations of the detected peaks were compared to the annotated locations in the records of the standard MIT-BIH databases (NSRDB, AFDB, LTDB, and MITDB). These databases contain annotations for ventricular activity. The normal QRS-width is 100ms (± 20 ms).³⁹ Therefore, an annotation and a detected R-peak were considered similar when they occurred within 100ms. Only one detected R-peak was matched with each annotation, and vice versa. Therefore, if many R-peaks were detected within a short time interval and only one annotation was present in the same time interval, only one R-peak was labelled as correct. Validation of P- and T-wave detection was performed using the QTDB. The same approach was used as previously described for validation of R-peak detection. However, for T-wave detection, the accepted time difference was set to 150ms instead of 100ms, because the normal duration of T-waves can be longer than 100ms.



**The results of the moving means are amplified to better illustrate the differences between the two moving windows.*

Figure 10 – P- and T-wave detection, as further explained in text. The results of the moving mean with a window of 55ms and 110ms are shown in dark green and orange, respectively.

Phase 2d: Noise detection

Noise detection was performed using all local minima and maxima in the ECG signal. For each R-peak, prominences of local minima and maxima in the surrounding 4 seconds (2 seconds prior to the R-peak and 2 seconds after the R-peak) were determined. The 85th percentile of all prominences was calculated. R-peaks with a prominence smaller than two times the 85th percentile were excluded for further analysis and were marked as noise.

Additional R-peaks were excluded based on noise peak prominence in the close surrounding of the R-peaks. The local minima and maxima in the surrounding 2 seconds of an R-peak (1 second prior to the R-peak and 1 second after the R-peak) were detected, excluding detected P-, Q-, R-, S- and T-peaks and thereby assuming all detected minima and maxima were caused by noise. The 85th percentile of the prominences of these local minima and maxima was calculated. Two types of noise segments are determined, as visualized in Figure 11 and 12: 1) noise segments in which R-peaks cannot be accurately detected, and 2) noise segments in which P- and T-waves cannot be accurately detected.

Segments with R-peaks which have prominences smaller than *three* times the 85th percentile of prominences of all local minima and maxima, which are not P-, Q-, R-, S- and T-peaks, were marked as the first type of noise and excluded for QRS-complex detection and P- and T-wave detection. Segments with R-peaks which have prominences smaller than *ten* times the 85th percentile of the prominences of all minima and maxima, which are not P-, Q-, R-, S- and T-peaks, were marked as the second type of noise and excluded for P- and T-wave detection.

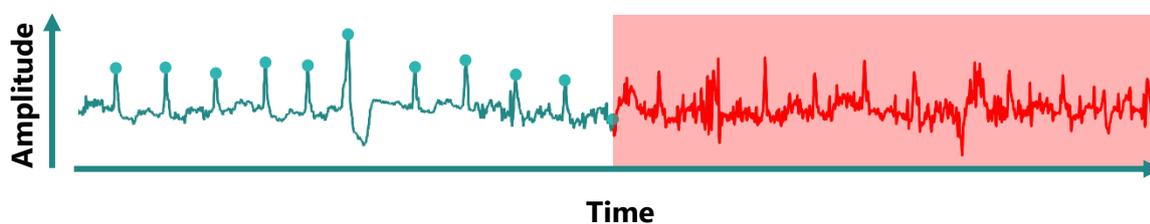


Figure 11 – Example of ECG segment excluded for QRS-detection. The green line represents an included segment, the red line in the shaded area represents the signal in the subsequent segment which was excluded for QRS-detection. R-peaks after excluding noise segments are represented by the dots (•).

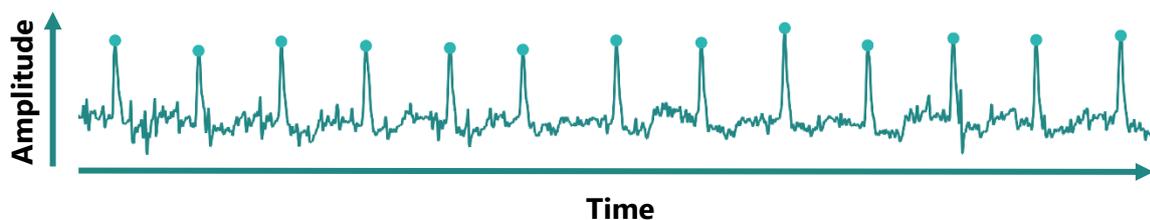


Figure 12 – Example of ECG segment excluded for P- and T-wave detection. R-peaks after excluding noise segments are represented by the dots (•).

Phase 3: Feature extraction

Using the detected P-, Q-, R-, S- and T-peaks, 57 features were extracted from the ECG records, as shown in Table 3 (p. 44). Features were extracted from each ECG segment of 30 seconds and could be divided into eight groups:

1. RR-interval characteristics
2. Peak-interval characteristics
3. Amplitude characteristics
4. P-wave characteristics
5. T-wave characteristics
6. QRS-morphology characteristics
7. Autocorrelation characteristics
8. Noise

RR-interval characteristics

Excluding irregular rhythms other than atrial fibrillation

A total of ten features were based on irregularity of RR-intervals. The RR-interval was defined as the time between two consecutive R-peaks. Continuous rhythm registrations not only showed RR-interval irregularity caused by AF, but also caused by premature atrial complexes (PACs), premature ventricular complexes (PVCs) and beats missed by peak detection. Therefore, before extracting the features based on RR-intervals, these beats were excluded from analysis using characteristics of the Poincaré plot.

A Poincaré plot is constructed by plotting each RR-interval against the next RR-interval. When all RR-intervals are similar, points will be very close to each other. If RR-intervals are changing slowly, the points in the plot will be close to the -45° -line through the origin, on which consecutive RR-intervals are equal. Otherwise, if RR-intervals are changing rapidly, the points in the Poincaré plot will move away from this -45° -line through the origin. Detection of PVCs, PACs, and missing beats was based on the specific patterns in the Poincaré plot, as described by Park et al.⁴⁰ First, the central point of the Poincaré plot was determined as the median of all RR-intervals of the entire 30-seconds ECG. Poincaré plots of missing beats, PACs, and PVCs are visualized in Figure 13. Missing beats and PACs are characterized by a triangle shape, whereas PVCs are characterized by a wedge shape in the Poincaré plot. For detection of PVCs, PACs and missing beats, an RR-interval preceding the analyzed beat (RR_i), and the preceding (RR_{i-1}) and succeeding RR-interval (RR_{i+1}) were used.

Detection criteria for missing beats were:

- ❖ Distance between (RR_{i-1}, RR_i) or (RR_i, RR_{i+1}) and the central point of all RR-intervals is more than 25% of the distance from the origin to this central point,
- ❖ The central point of RR_{i-1}, RR_i and RR_{i+1} is in the upper right plane with respect to the central point of all RR-intervals, and
- ❖ The distance between point (RR_i, RR_{i+1}) and the central point of all RR-intervals and the distance between point (RR_{i-1}, RR_i) differ no more than 10%.

Detection criteria for PACs were:

- ❖ Distance between (RR_{i-1}, RR_i) or (RR_i, RR_{i+1}) and the central point of all RR-intervals is more than 15% of the distance from the origin to this central point,
- ❖ RR_i is smaller than RR_{i-1} ,
- ❖ RR_{i+1} and RR_{i-1} differ no more than 10%.
- ❖ The central point of RR_{i-1} , RR_i and RR_{i+1} is in the lower left plane with respect to the central point of all RR-intervals, and
- ❖ The distance between point (RR_i, RR_{i+1}) and the central point of all RR-intervals and the distance between point (RR_{i-1}, RR_i) differ no more than 10%.

Detection criteria for PVCs were:

- ❖ Distance between (RR_{i-1}, RR_i) or (RR_i, RR_{i+1}) and the central point of all RR-intervals is more than 10% of the distance from the origin to this central point,
- ❖ RR_{i-1} differs no more than 10% from the central point of all RR-intervals,
- ❖ The sum of RR_i and RR_{i+1} equals two times RR_{i-1} with a margin of 20%,
- ❖ RR_i is smaller than RR_{i-1} , and
- ❖ RR_i is smaller than RR_{i+1} .

In case of missing beats, PACs, and PVCs, the preceding RR-interval was excluded from analysis. Additionally, in case of PVCs, the RR-interval following the PVC was excluded, because the interval from PVC to the next beat is longer due to a compensatory pause. This pause is caused by failed conduction of a sinus beat to the ventricles due to early depolarization of the ventricles as a result of the PVC. PACs may also cause a compensatory pause due to early depolarization of the sinus node. However, the exact duration of this pause is unpredictable, while in case of PVCs, the sinus node is not disrupted and the next beat will follow the next sinus node activation.

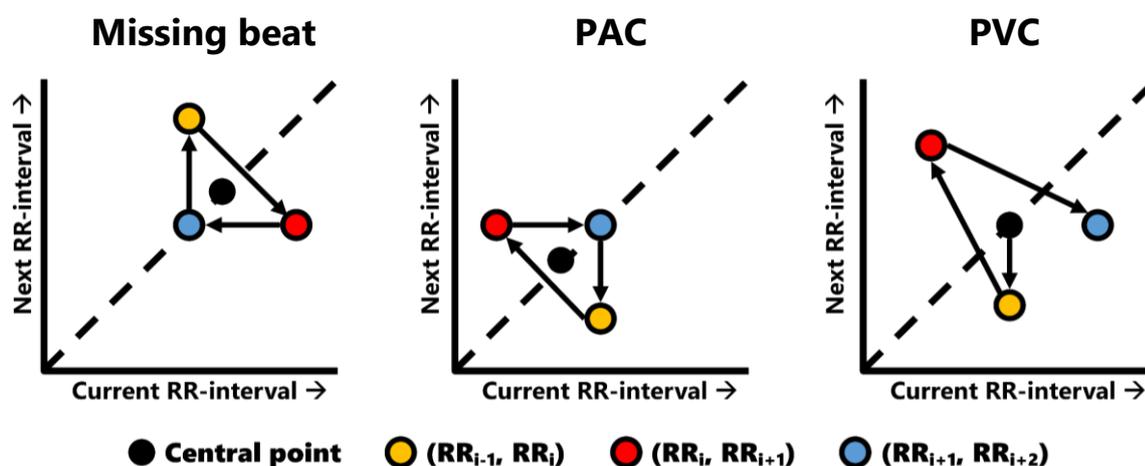


Figure 13 – Poincaré plots of missing beat (left panel), premature atrial complex (PAC) (middle panel) and premature ventricular complex (PVC) (right panel). The dashed line (- -) indicates the -45°-line where successive RR-intervals are equal. RR_i = RR-interval preceding the i -th beat; RR_{i-1} = RR-interval succeeding the i -th beat; RR_{i+2} = RR-interval succeeding RR_{i+1} .

Feature extraction

Next, a new Poincaré plot was made without the eliminated PACs, PVCs, and missing beats. Figure 14 shows a Poincaré plot of sinus rhythm (left panel) and a Poincaré plot of atrial fibrillation (right panel). Features extracted from this plot were the standard deviations of distances of points to the 45°-line and to the -45°-line, as visualized in Figure 14.[§] A large standard deviation of distances to the -45°-line indicates successive RR-intervals are different, hence this is an indicator for fast changes in RR-interval. A large standard deviation of distances to the 45°-line without a large standard deviation of distances to the -45°-line indicates successive RR-intervals are similar, but all RR-intervals together show a large variation, hence this is an indicator for slow changes in RR-intervals.

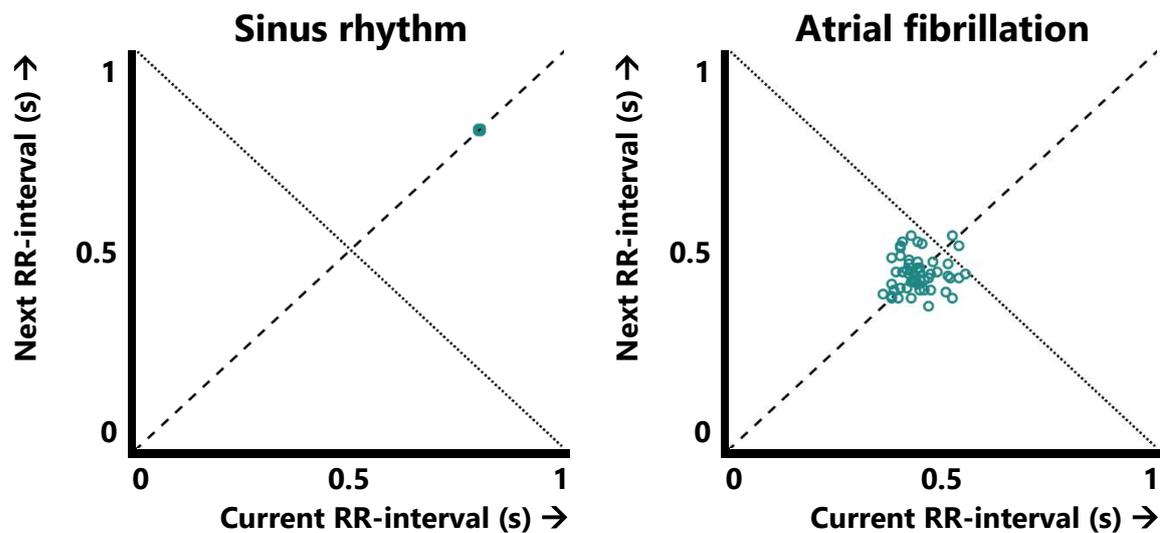


Figure 14 – Poincaré plots of an ECG of 30 seconds showing sinus rhythm (left) and atrial fibrillation (right). Points (o) indicate two successive RR-intervals, the dashed lines (- - -) and the dotted lines (···) indicate the -45°-line and the 45°-line, respectively, used for calculation of standard deviation, as further explained in text.

Furthermore, the median and interquartile range (IQR) of the intervals between consecutive R-peaks were calculated. The number of beats per minute was determined using the reciprocal of the median RR-interval. Also, the median and IQR of the intervals between an R-peak and the R-peak which was two beats later were calculated. As another measure of RR-interval irregularity, the median and IQR of the ratio between the RR-interval preceding an R-peak and the RR-interval succeeding an R-peak were calculated. Lastly, Shannon entropy of the RR-intervals was calculated as a measure for the variation of RR-intervals. Shannon entropy describes the information content of a variable and is defined as⁴¹:

$$\text{Shannon entropy} = - \sum_{i=1}^M p(i) \log_2 p(i) \quad (\text{Eq. 2})$$

where M equals the number of discrete values the variable can take and $p(i)$ is the

[§] Further explained with an example in Appendix IV.

probability density function of variable x for the i^{th} value.** For implementation in the AF classifier, a histogram of the RR-intervals was made with a bin width of 100ms. Next, the histogram was normalized to sum to one. The result was used as a probability distribution for calculation of Shannon entropy. The probability distributions of an ECG with AF and without AF are visualized in Figure 15. If all RR-intervals are similar, Shannon entropy will be low. Otherwise, if RR-intervals vary, Shannon entropy will be higher. In these examples, the Shannon entropy of the RR-intervals during sinus rhythm and atrial fibrillation are 0.76 bits and 1.3 bits, respectively.

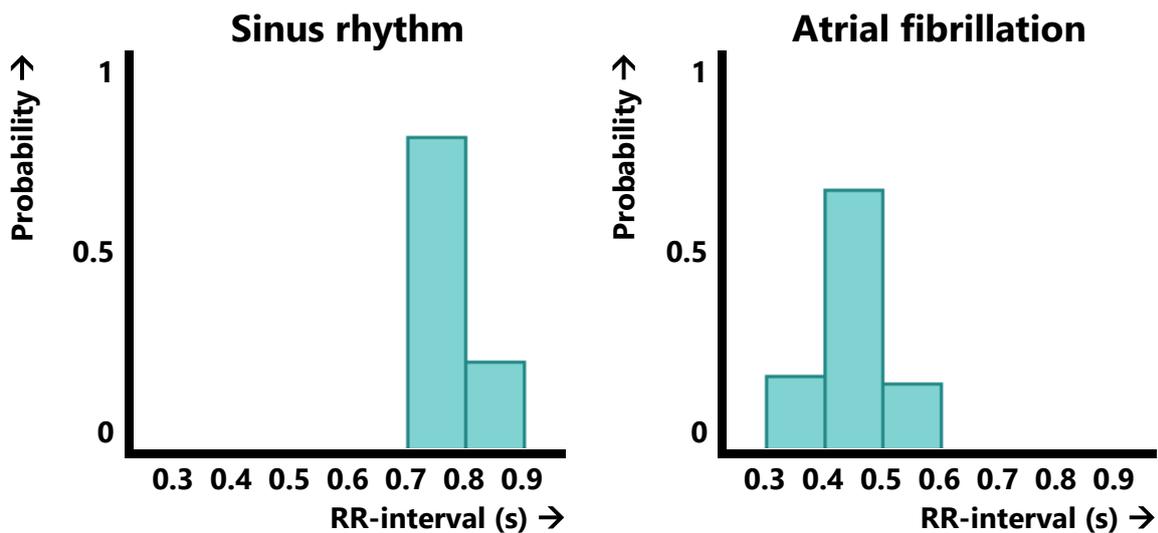


Figure 15 – Probability distributions of an ECG of 30 seconds showing sinus rhythm (left) and atrial fibrillation (right).

Peak-interval characteristics

Ten features based on peak intervals were extracted from the ECG segments. Of each interval, median and IQR were used. Calculated intervals were P-peak to Q-peak (PQ-interval), Q-peak to S-peak (QRS-width), Q-peak to T-peak (QT-time), and S-peak to T-peak (ST-time). Additionally, the QT-interval was corrected for heart rate using Bazett's formula, which describes the observed relation between the QT-interval and the RR-interval: $QT_c = QT\text{-interval} / \sqrt{RR\text{-interval}}$.⁴²

Amplitude characteristics

Based on peak-to-peak amplitudes, twelve features were extracted. Median and IQR of amplitudes from Q-peak to R-peak, R-peak to S-peak, absolute differences between amplitudes of consecutive R-peaks (RR-amplitude), an R-peak to the R-peak two peaks later, and the ratio between RR-amplitude preceding an R-peak and RR-amplitude succeeding an R-peak were calculated. Furthermore, similarly to RR-intervals, Poincaré features were calculated. Again, the standard deviation of distances of points to the 45°-line and to the -45°-line were extracted.

** Further explained with an example in Appendix V.

P- and T-wave characteristics

Thirteen characteristics of P- and T-waves were extracted from each ECG segment. The percentage of R-peaks without P-waves, with one P-wave, and with multiple P-waves were calculated. Also, the percentage of R-peaks without T-waves and with one T-wave were calculated. Furthermore, median and IQR of the P-wave amplitude and T-wave amplitude were extracted. Additionally, the median and IQR of the ratio between the amplitude of the P-wave and amplitude of the QRS-complex, and the ratio between the amplitude of the T-wave and amplitude of the QRS-complex were computed. Lastly, the 85th percentile of the ratio between the amplitude of the P-wave and amplitude of the QRS-complex was calculated as a measure for the amplitude of P-waves with high amplitudes.

QRS-morphology characteristics

As visualized in Figure 17, using the cross-correlation between all QRS-complexes in the ECG segment, QRS-complexes with similar QRS-morphologies were detected. First, the ECG signal was filtered to eliminate noise using a second-order Butterworth filter with lower and higher cut-off frequencies of 2Hz and 30Hz, respectively. These cut-off frequencies were arbitrarily set to eliminate noise but keep the morphology of the QRS-complexes intact. Next, for each detected QRS-complex, a segment starting 100ms before the R-peak and ending 100ms after the R-peak was created. Before calculating the correlation coefficients between segments, segments were optimally aligned using the cross-correlation between segments. Cross-correlation was calculated with delays from -50ms up to 50ms. Correlation coefficients were calculated for the delay for which the segments were most similar. In Figure 16, the relation between correlation coefficients and QRS-morphology is illustrated using examples with correlation coefficients ranging from 0.3 to 1.0. If the correlation coefficient was higher than an arbitrary threshold of 0.8, QRS-morphologies were assumed to be similar. The number of different morphologies, percentage of QRS-complexes with the most common morphology, and percentage of QRS-complexes with the second most common morphology were extracted as features.

Autocorrelation characteristics

Another measure for signal irregularity is the autocorrelation, as visualized in Figure 18. First, the signal was filtered using a second-order Butterworth filter with lower and higher cut-off frequencies of 0.5Hz and 10Hz, respectively. These cut-off frequencies differ from the frequencies used to detect QRS-morphology. For this calculation, the focus was only on the location of the QRS-complexes and not on the exact morphological differences. Therefore, the high frequencies in the ECG were filtered out. The normalized autocorrelation of the ECG was calculated with a maximum delay of 15 seconds, which is half of the duration of the input ECG. Next, the moving maximum of the normalized autocorrelation was calculated with a window size of 500ms. The result was used for peak detection. The central peak was excluded since the autocorrelation is always optimal with a delay of 0ms. Seven features were extracted: median and IQR of the autocorrelation, number of peaks in the autocorrelation, median and IQR of the amplitude of peaks in the autocorrelation, and median and IQR of the time interval between peaks in the autocorrelation.

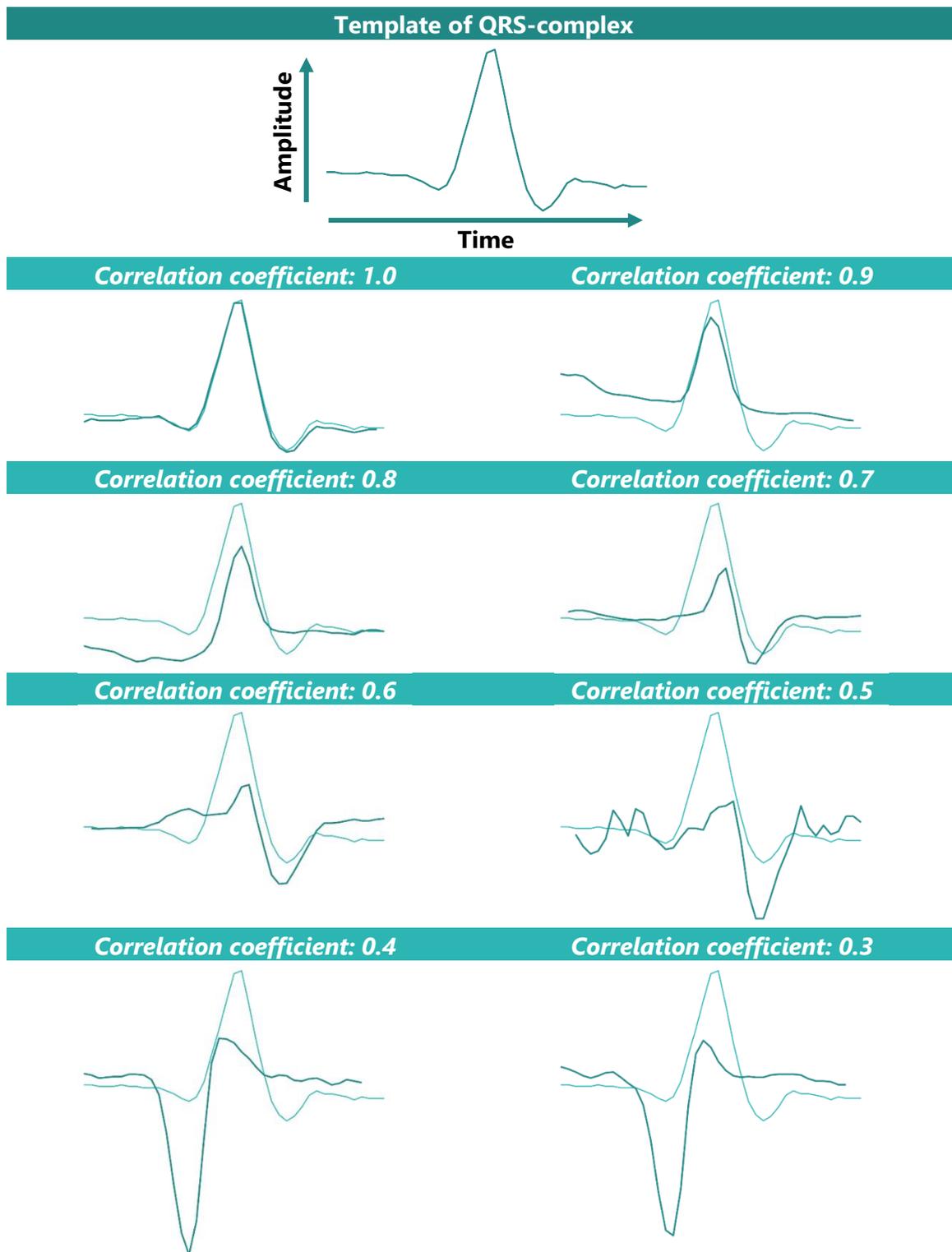


Figure 16 – Examples of QRS-complexes from post-operative telemetry with correlations ranging from 0.3 to 1.0. The upper figure contains the ECG which was (for this example) used as a template to calculate correlation coefficients. The other figures show the template in light green and the other segment in dark green. QRS-complexes with correlation coefficients of at least 0.8 were assumed to have similar QRS-morphologies. Scales of all figures are equal.

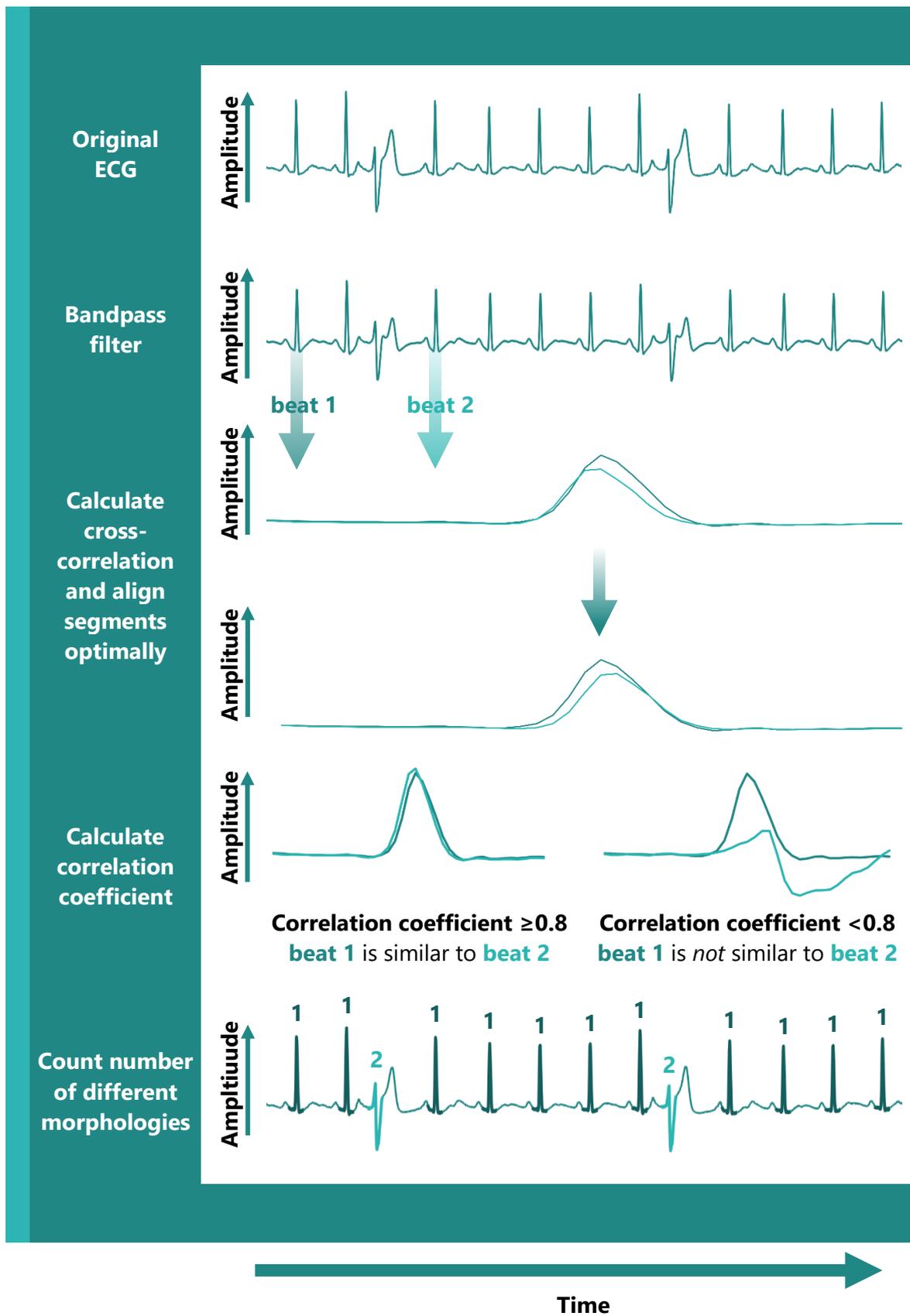


Figure 17 – Differentiating QRS-morphologies, as further explained in text.

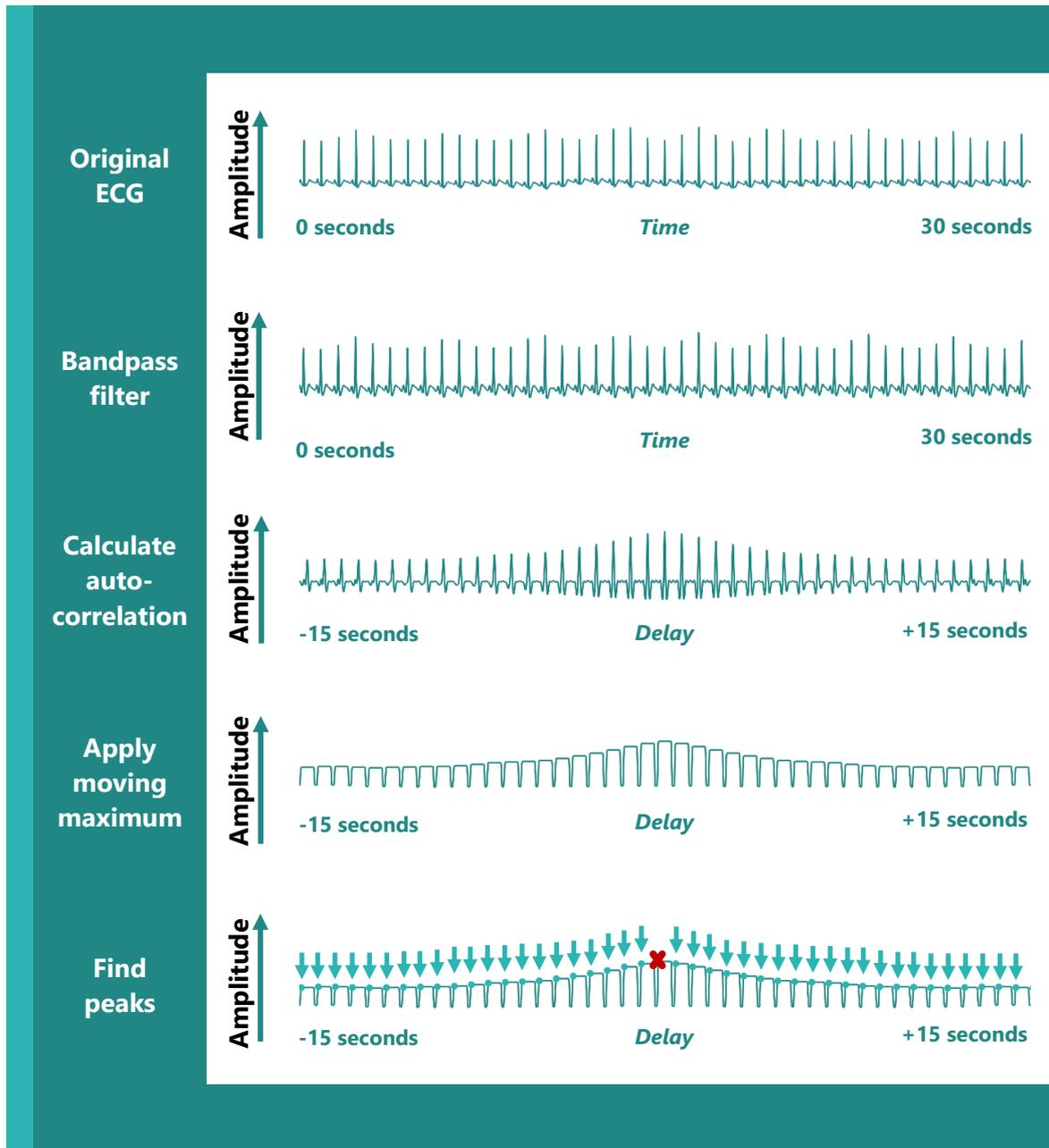


Figure 18 – Calculating autocorrelation of ECG segments, as further explained in text. The red cross (X) indicates the center peak which was excluded from analysis.

Noise

To classify noise segments, for each ECG segment, the percentage of excluded samples for R-peak detection was calculated as a feature.

Phase 4: Training classifiers

Classifiers were trained to classify 30-seconds ECG segments into one of three classes: *AF*, *No AF*, and *Unusable*. Training was performed using the Classification Learner Application in Matlab (R2019b with Statistics and Machine Learning Toolbox 11.6) using 10-fold cross-validation. The input for training consisted of feature vectors containing all 57 features for 4,800 randomly selected 30-seconds ECG segments. All classes were represented equally (1,600 segments per class). The *No AF*-class was filled with 800 ECG segments with a regular rhythm and 800 ECG segments with an irregular rhythm, but without AF (e.g. ECGs showing PACs, PVCs, or sinus arrhythmia). The remaining 1,600 ECG segments were kept hidden for testing. Additionally, a vector containing the manual annotations of the 4,800 training segments was added as reference. Initially, all available classification methods were applied to gain insight into the best method. The optimal classifier in terms of accuracy was selected for final analysis.

Phase 5: Testing classifier

The remaining 1,600 30-seconds ECG segments were used to test the classifier's performance in terms of overall accuracy. Furthermore, accuracy, recall and FNR, and specificity and FPR were calculated for detection of AF specifically. The testing set consisted of 400 records with AF, 400 records with a regular rhythm, 400 records with an irregular rhythm (without AF), and 400 unusable records due to noise and/or artefacts.

Statistical analysis

Performance measures for both peak detection and AF classification were calculated. As listed in Table 1, the outcome measures for peak detection were **recall**, **precision**, and the **harmonic mean of recall and precision (F1-score)**:

$$\text{F1-score} = 2 \cdot \frac{\text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}} \quad (\text{Eq. 3})$$

Recall is a measure for the number of peaks that were correctly detected by the algorithm, compared to the number of peaks that were manually annotated. Likewise, **precision** is a measure for the number of peaks that were correctly detected by the algorithm, compared to the total number of both correctly and incorrectly detected peaks. The arithmetic mean (A) of a dataset a with n samples is defined as:

$$A = \frac{1}{n} \sum_{i=1}^n a_i \quad (\text{Eq. 4})$$

Calculating the arithmetic mean of **recall** and **precision** would result in an overestimation of the classifier performance. For example, if the algorithm detected only one R-peak and it was correctly detected, the **precision** would be 1. However, if 100,000 R-peaks were

manually annotated, the **recall** would be 0.00001. The arithmetic mean would then be 0.5, which overestimates the performance of the algorithm.

Instead, the **harmonic mean** (H) of a dataset (a) with n samples is mainly affected by the minimum value. The definition is as follows:

$$H = \frac{n}{\sum_{i=1}^n a_i^{-1}} \quad (\text{Eq. 5})$$

The example with an arithmetic mean of 0.5 results in a **harmonic mean** of 0.00002, which gives a better representation of the actual performance of the algorithm in this example.

Outcome measures for AF detection were **accuracy**, **sensitivity** and **FNR**, and **specificity** and **FPR**. **Accuracy** is a measure for the correctly classified records compared to the total number of analyzed records. **Sensitivity** is a measure for the number of records correctly classified as AF, compared to the total number of records with manually annotated AF. **Specificity** is a measure for the number of records correctly classified as not containing AF, compared to the total number of records without manually annotated AF. **FNR** is the proportion of the number of records with manually annotated AF, which were incorrectly classified as not containing AF. **FPR** is the proportion of the number of records without manually annotated AF, which were incorrectly classified as containing AF.

The binomial proportion 95% confidence interval (95% CI) was calculated with the Wilson score using the standard definition⁴³:

$$\frac{2np + z^2 \pm z\sqrt{z^2 + 4npq}}{2 \cdot (n + z^2)} \quad (\text{Eq. 6})$$

n = sample size

p = proportion

z = standard normal deviate (≈ 1.96 for 95% CI)

$q = 1 - p$

For **accuracy**, **recall**, **precision**, **sensitivity**, **specificity**, **FNR** and **FPR** parameters are summed up in Table 2. For the **F1-score**, the lower and upper confidence bound were calculated as the harmonic mean of the lower confidence bounds and upper confidence bounds of recall and precision, respectively.

Table 1 – Performance measures

		Classifier output		Measure	Calculation
		+	-		
Manual annotations	+	True positive (TP)	False negative (FN)	Accuracy	$\frac{TP + FN}{TP + FN + FP + TN}$
	-	False positive (FP)	True negative (TN)	Recall	$\frac{TP}{TP + FN}$
				Precision	$\frac{TP}{TP + FP}$
				Sensitivity	$\frac{TP}{TP + FN}$
				FNR	$\frac{FN}{TP + FN}$
				Specificity	$\frac{TN}{FP + TN}$
				FPR	$\frac{FP}{FP + TN}$
				F1-score	$2 \cdot \frac{\text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}}$

Table 2 – Parameters for Wilson score

	n	p	z	q
Accuracy	$TP + FN + FP + TN$	$\frac{TP + FN}{TP + FN + FP + TN}$	1.9600	$1 - \frac{TP + FN}{TP + FN + FP + TN}$
Recall	$TP + FN$	$\frac{TP}{TP + FN}$	1.9600	$1 - \frac{TP}{TP + FN}$
Precision	$TP + FP$	$\frac{TP}{TP + FP}$	1.9600	$1 - \frac{TP}{TP + FP}$
Sensitivity	$TP + FN$	$\frac{TP}{TP + FN}$	1.9600	$1 - \frac{TP}{TP + FN}$
FNR	$TP + FN$	$\frac{FN}{TP + FN}$	1.9600	$1 - \frac{FN}{TP + FN}$
Specificity	$FP + TN$	$\frac{TN}{FP + TN}$	1.9600	$1 - \frac{TN}{FP + TN}$
FPR	$FP + TN$	$\frac{FP}{FP + TN}$	1.9600	$1 - \frac{FP}{FP + TN}$
F1-score	Not applicable; lower and upper confidence bound were calculated as the harmonic mean of the lower confidence bounds and upper confidence bounds of recall and precision, respectively.			

TP = number of true positive classifications; FN = number of false negative classifications; FP = number of false positive classifications; TN = number of true negative classifications;

Table 3 – Features used in atrial fibrillation classifier

<p>1. RR-interval characteristics (10 features)</p> <ul style="list-style-type: none"> • Number of beats per minute (bpm) • Interval between R_i and R_{i+1} (seconds)* • Interval between R_{i-1} and R_{i+1} (seconds)* • Ratio between the RR-interval preceding an R-peak and the RR-interval succeeding an R-peak* • Poincaré plot: standard deviation of distances of points to the 45°-line and to the -45°-line • Shannon entropy of the RR-intervals
<p>2. Peak-interval characteristics (10 features)</p> <ul style="list-style-type: none"> • Interval between P-peak and Q-peak (seconds)* • Interval between Q-peak and S-peak (seconds)* • Interval between Q-peak and T-peak (seconds)* • Corrected interval between Q-peak and T-peak using Bazett's formula (seconds)* • Interval between S-peak and T-peak (seconds)*
<p>3. Amplitude characteristics (12 features)</p> <ul style="list-style-type: none"> • Amplitude from Q-peak to R-peak* • Amplitude from R-peak to S-peak* • Difference in amplitude of R_i and R_{i+1}* • Difference in amplitude of R_{i-1} and R_{i+1}* • Ratio between the amplitude differences of $R_{i-1}R_i$ and R_iR_{i+1}* • Poincaré plot of differences in RR-amplitude: standard deviation of distances of points to the 45°-line and to the -45°-line
<p>4. P-wave characteristics (8 features)</p> <ul style="list-style-type: none"> • Percentage of R-peaks without P-waves, with one P-wave, and with multiple P-waves • Amplitude of P-wave* • Ratio between amplitudes of P-wave and QRS-complex* + 85th percentile of amplitudes of this ratio
<p>5. T-wave characteristics (6 features)</p> <ul style="list-style-type: none"> • Percentage of R-peaks without T-waves, and with one T-wave • Amplitude of T-wave* • Ratio between amplitudes of T-wave and QRS-complex*
<p>* median and interquartile range included as features</p>

bpm = beats per minute; R_i = the i -th R-peak; R_{i-1} = the R-peak preceding the i -th R-peak; R_{i+1} = the R-peak succeeding the i -th R-peak;

Table 3 – Features used in atrial fibrillation classifier (continued)

6. QRS-morphology characteristics (3 features)
<ul style="list-style-type: none">• Number of different QRS-morphologies• Percentage of QRS-complexes with the most common morphology• Percentage of QRS-complexes with the second most common morphology
7. Autocorrelation characteristics (7 features)
<ul style="list-style-type: none">• Autocorrelation of ECG segment*• Ratio between number of peaks in the autocorrelation and number of QRS-complexes• Amplitude of peaks in the autocorrelation*• Time interval between peaks in the autocorrelation (seconds)*
8. Noise (1 feature)
<ul style="list-style-type: none">• Percentage of excluded samples for R-peak detection
<i>* median and interquartile range included as features</i>

Results

Peak detection

Detection of QRS-complexes

Detection of QRS-complexes was validated using the four standard MIT-BIH databases (NSRDB, AFDB, LTDB and MITDB), containing a total of 96 records and 3,636,367 annotated QRS-complexes. The total amount of detected QRS-complexes was 3,650,232, of which 3,572,255 detections were true positive. 77,977 detections were false negative, and 64,112 QRS-complexes were missed. This results in an overall recall of 98.237% and precision of 97.864%. The F1-score is 98.050%. Results per database are summarized in Table 4. An entire list of detections per record is provided in Appendix VI.

Table 4 – Test characteristics for detection of QRS-complexes

	Annotated	Detected	TP	FP	FN
NSRDB	1,729,630	1,737,076	1,727,469	9,607	2,161
AFDB	1,128,561	1,143,166	1,075,917	67,249	52,644
LTDB	668,682	660,548	660,037	511	8,645
MITDB	109,494	109,442	108,832	610	662
Overall	3,636,367	3,650,232	3,572,255	77,977	64,112

	Recall (95% CI)	Precision (95% CI)	F1-score (95% CI)
NSRDB	99.875% (99.870% - 99.880%)	99.447% (99.436% - 99.458%)	99.661% (99.652% - 99.669%)
AFDB	95.335% (95.296% - 95.374%)	94.117% (94.074% - 94.160%)	94.722% (94.681% - 94.763%)
LTDB	98.707% (98.680% - 98.734%)	99.923% (99.916% - 99.929%)	99.311% (99.294% - 99.328%)
MITDB	99.395% (99.348% - 99.440%)	99.443% (99.397% - 99.485%)	99.419% (99.372% - 99.462%)
Overall	98.237% (98.223% - 98.250%)	97.864% (97.849% - 97.879%)	98.050% (98.036% - 98.064%)

NSRDB = Normal Sinus Rhythm Database; AFDB = Atrial Fibrillation Database; LTDB = Long-term ECG Database; MITDB = Arrhythmia Database; TP = True Positive, FP = False Positive; FN = False Negative; F1 = F1-score;

P- and T-wave detection

Detection of P- and T-waves was validated using the MIT-BIH QT Database, containing 96 records with annotated QRS-complexes and P- and T-waves. A total of 91,607 P-waves and 117,632 T-waves were annotated. The detection algorithm resulted 102,289 P-waves and 97,458 T-waves, of which 19,543 and 7,263 detections were false positive, respectively. The algorithm missed 8,861 P-waves and 27,437 T-waves. This results in a recall of 90.327% (95% CI: 90.134% - 90.517%) and 76.676% (95% CI: 76.433% - 76.916%), precision of 80.894% (95% CI: 80.652% - 81.134%) and 92.548% (95% CI: 92.381% - 92.710%), and F1-score of 85.351% (95% CI: 85.130% - 85.569%) and 83.867% (95% CI: 83.654% - 84.078%) for P-wave detection and T-wave detection, respectively. An entire list of detections per record is provided in Appendix VII.

Classifier performance

The optimal classification method was using boosted decision trees; hence this method was used for further training. In Matlab, the AdaBoost algorithm is used to train this classifier.^{††} The AdaBoost algorithm minimizes the final classification error by training multiple decision trees iteratively. Given the classification error of the trained decision tree, a weight is assigned to the decision tree. Better performing decision trees will be assigned a higher weight. The next decision tree is trained using another custom set of training data, in which previously misclassified samples are more frequently represented.⁴⁴ Hyperparameters of the model are the number of decision trees, the number of splits per decision tree, and the learning rate. Using Bayesian optimization with an expected improvement acquisition function, the error as a function of the hyperparameters was estimated and the hyperparameters were optimized.^{‡‡} The learning rate was fixed at 0.1.

The overall classification accuracy as calculated using 10-fold cross-validation was 96.38% (95% CI: 95.81% – 96.87%). For AF detection specifically, the accuracy was 97.42% (95% CI: 96.93% – 97.83%). Other performance measures are summarized in Table 5. Furthermore, the receiver operating characteristic (ROC) curve for AF detection is visualized in Figure 19, showing an area under the curve (AUC) of 0.99.

The overall classification accuracy when applying the classifier to the hidden test set was 95.38% (95% CI: 94.23% – 96.30%). For AF detection specifically, the accuracy was 96.44% (95% CI: 95.41% – 97.24%). Other performance measures are summarized in Table 6. Of all 74 misclassifications, 36 (49%) were made in the group with irregular rhythms without AF.

Feature importance

Figure 20 visualizes the estimated feature importance based on the reduced risk of errors due to adding the features to the decision trees.⁴⁵ The feature with the highest importance was the percentage of excluded samples for R-peak detection. Because this feature is only expected to impact the classification of a segment as noise/artefact, this feature was excluded from Figure 20. As expected, the feature with the second highest importance was based on RR-interval variability. Moreover, the variation in interval between P-wave and Q-peak and the percentage of R-peaks with one P-wave were estimated to be of high importance to classify ECG segments correctly. An entire list containing the estimated importance for all features is provided in Appendix X.

^{††} Further information on boosted decision trees is provided in Appendix VIII.

^{‡‡} Further information on Bayesian optimization is provided in Appendix IX.

Table 5 – Classifier performance (using 10-fold cross-validation)

		Classifier output		
		AF	No AF	Un-usable
Manual annotations	AF	1546	45	9
	No AF	63	1518	19
	Un-usable	7	31	1562

AF detection	%	95% CI
Accuracy	97.42%	96.93% - 97.83%
Value of negative classification		
Sensitivity	96.63%	95.62% - 97.40%
FNR	3.4%	2.6% - 4.4%
Value of positive classification		
Specificity	97.81%	97.25% - 98.26%
FPR	2.2%	1.7% - 2.8%

AF = atrial fibrillation; FNR = false negative rate; FPR = false positive rate

Table 6 – Classifier performance (using hidden test set)

		Classifier output		
		AF	No AF	Un-usable
Manual annotations	AF	389	8	3
	No AF (reg.)	7	388	5
	No AF (irreg.)	33	364	3
	Un-usable	6	9	385

AF detection	%	95% CI
Accuracy	96.44%	95.41% - 97.24%
Value of negative classification		
Sensitivity	97.3%	95.1% - 98.5%
FNR	2.8%	1.5% - 4.9%
Value of positive classification		
Specificity	96.17%	94.92% - 97.11%
FPR	3.8%	2.9% - 5.1%

AF = atrial fibrillation; reg. = regular; irreg. = irregular; FNR = false negative rate; FPR = false positive rate

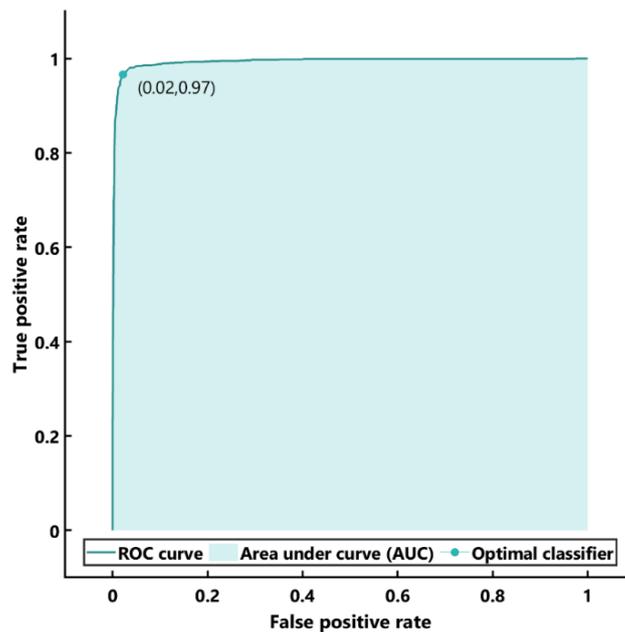


Figure 19 – Receiver operating characteristic (ROC) curve for AF detection. The calculated area under the curve (AUC) was 0.99.

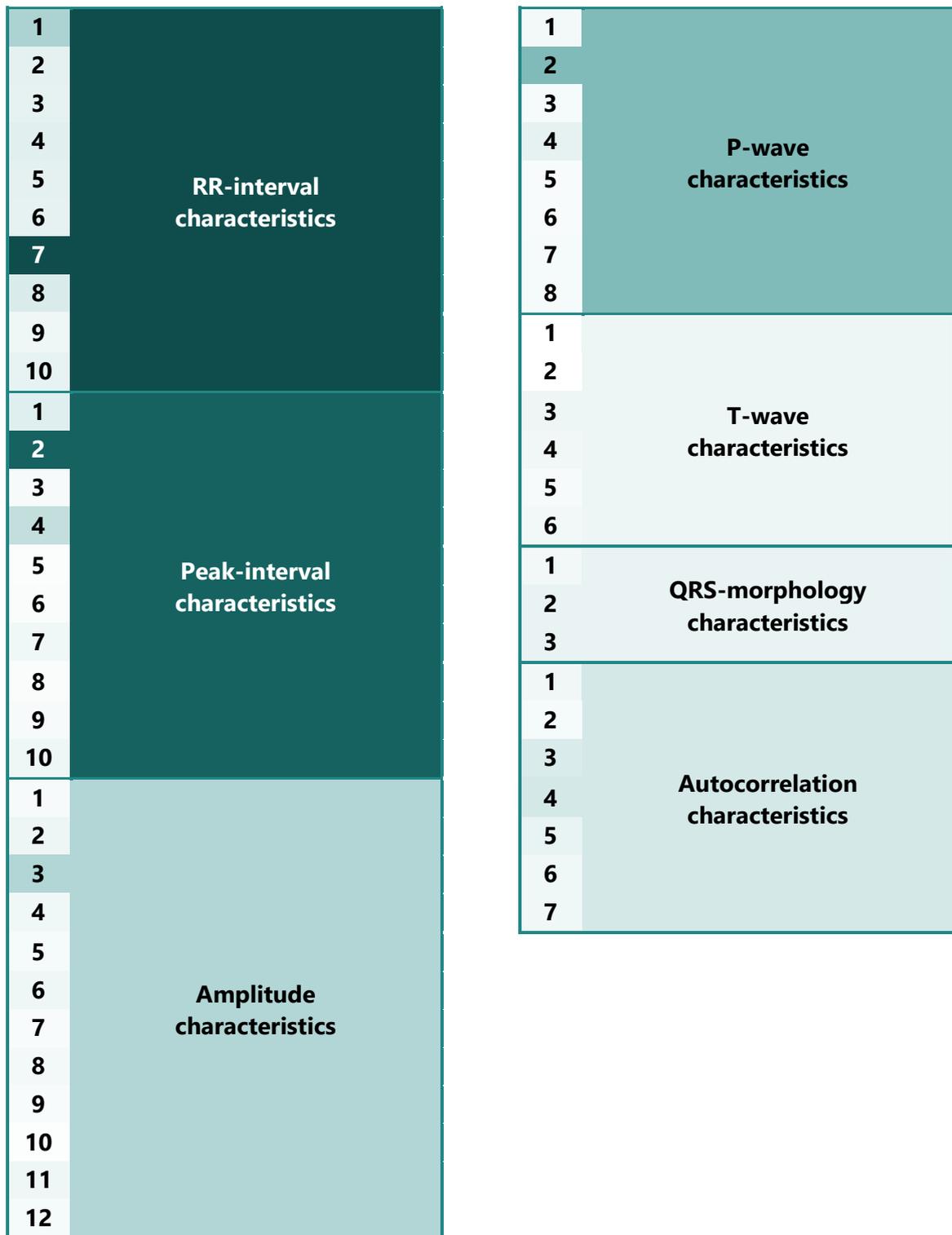


Figure 20 – Feature importance. Darker color indicates higher importance. Single features are visualized in the left (narrow) columns. The exact order of the features and results are provided in Appendix X. Maximum feature importance of feature groups is visualized in the right (wide) columns. The noise feature is not included in this figure, because of the extremely high estimated importance caused by classifying noise signal.

Discussion

The goal of this study was to develop a classifier for automated detection of AF with an FPR of at most 5% and an FNR of at most 3%. A novel algorithm was introduced for AF detection, using a combination of common ECG features (e.g. RR-intervals and presence of P-waves) and more complex features (e.g. QRS-morphology and autocorrelation). Whereas most studies use standard ECG databases containing relatively clean ECG data for development of an algorithm, the current classifier was trained and tested using manually annotated continuous rhythm registrations obtained in the post-operative setting. This method resulted in a classifier with an overall accuracy of 95.37%, and for detection of AF a low false negative rate of 2.75% and a medium false positive rate of 3.83%. As expected beforehand, classification was mainly based on RR-interval irregularity.

Peak detection

QRS-complex detection

Detection of QRS-complexes using a method based on the Pan-Tompkins algorithm resulted in a high recall and precision in the MIT-BIH databases containing long-term ECGs with both normal sinus rhythm (NSRDB) and more irregular rhythms (LTDB). Furthermore, detection of QRS-complexes in short ECGs containing various rhythm disorders was accurate with an F1-score of 99.4%. Detection of QRS-complexes in the MIT-BIH database containing long-term ECG recordings of patients with AF (AFDB) was less accurate with an F1-score of 94.7%. This is most likely due to the different methodology for annotating the beats in this database. Instead of manually annotating the beats, beat annotations in the AFDB are generated using an automated detector. Not all records have been manually corrected, which in some records results in inaccurate beat annotations. In particular, as visualized in Figure 21, manual inspection of record 07162 (recall: 22.1%; precision: 21.6%; F1-score: 21.9%) of the AFDB shows inaccurate placement of the annotations.

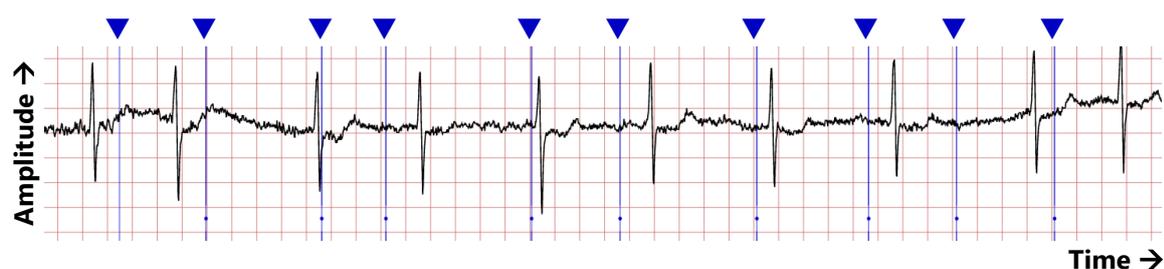


Figure 21 – Screenshot of the first ten seconds of record 07162 from the MIT-BIH Atrial Fibrillation Database.⁴⁶ Blue arrows and vertical lines indicate annotated QRS-complexes, which were not manually corrected.

P- and T-wave detection

AF is characterized by the absence of P-waves, hence accurate detection of P-waves is essential. The proposed algorithm detected 90.3% of all annotated P-waves and 92.5% of detected T-waves were annotated T-waves. However, the algorithm also mistakenly detected additional P-waves and missed a substantial number of T-waves. Although the detection of P- and T-waves does not perform optimally, these error rates are most likely acceptable for AF detection. Detection of T-waves is used to prevent the algorithm

mistakenly detecting T-waves as P-waves. Therefore, it is expected that missing T-waves only has minor impact on the final AF classification. On the other hand, if there is a P-wave, it should be detected by the algorithm to ensure accurate AF classification. Therefore, for P-wave detection a high recall is preferred over a high precision. Furthermore, to mitigate the effect of the low precision of P-wave detection on the AF classifier, amplitude of the P-wave relative to the amplitude of the R-peak was added as a feature. Also, false positive detection of P-waves in ECG showing AF likely results in larger variance in the interval between P-wave and QRS-complex. This might explain the relatively high importance of this feature in the final AF classifier.

Predictive validity

The proposed classifier has a low FNR, hence few ECG segments with AF will be missed. However, 33 out of 400 ECG records in the test set showing other irregular heart rhythms were incorrectly classified as containing AF.

The aim of the current study was to train a classifier for use in the research setting. Therefore, as stated before, focus was on reducing the FNR. The FPR of 3.83% is still high, hence screening of the AF segments is still required to get an accurate estimation of the AF burden. However, since the FNR is only 2.75%, few ECG records containing AF are missed. Therefore, only screening the segments with detected AF already results in a decent estimation of the AF burden.

Further development is essential to use the proposed classification model unsupervised and/or in clinical practice, because requirements differ from the supervised research setting. Also, the calculated performance measures cannot be directly translated to a single patient. The test set consisted of segments from many patients. Therefore, it can be assumed that the segments are not related. However, in a single patient, if one segment is misclassified as AF, a subsequent similar segment will most likely be misclassified. In a single patient, this could lead to overestimation or underestimation of the AF burden, depending on the type of error (false positive or false negative). Furthermore, provided accuracy measures solely apply to the detection of 30-seconds segments containing AF. Therefore, these measures do not directly translate to a high accuracy for detecting the start and end times, and duration of AF episodes.

AF segments to AF episodes

The output of this algorithm is the classification of a single 30-seconds ECG segment into one of three classes (AF, no AF, unusable). Therefore, the result of analyzing the telemetry data of an entire patient results in a series of classifications for each segment of 30 seconds. However, this does not yet result in the demarcation of AF episode boundaries. Additional analysis should provide information on the exact start and end times of AF episodes.

When analyzing a long-term ECG, a parameter which can be changed is the time between 30-seconds ECG segments which are analyzed. One option is to not allow overlap between analyzed segments. In this case, a next segment starts where the current segment ends. Another option would be to allow overlap between analyzed ECG segments. A combination of these methods most likely results in the fastest accurate estimation of the AF burden. A

first and rough estimate of the AF burden can be made using no or minor overlap between segments. Subsequently, the algorithm could be applied with more overlap to segments in the area with detected AF, resulting in more detailed information on the start and end times of the AF episode.

Furthermore, the definition for an AF episode should be clarified. Currently, the definition of a diagnostic AF episode is provided in the ESC Clinical Practice Guidelines.⁴⁷ The duration of a diagnostic AF episode should be at least 30 seconds. However, the guidelines do not include time between subsequent AF episodes as a factor for duration of the AF episode. For example, if telemetry shows two AF episodes with durations of hours, with only a few minutes without AF in between, it is still unclear whether this should be counted as single episode, or as two separate episodes. The same applies to AF segments with bad signal in between. A novel method for both cases is implemented in the current algorithm. First, AF episodes with noise or artefacts in between are merged if the total time of the episode does not increase more than 5%. Subsequently, AF episodes are merged if the time between episodes is less than an arbitrary threshold of 5 minutes.

Currently, validation of AF episode detection is not feasible, because this would require large amounts of manually annotated telemetry data with start and end times of AF episodes. Therefore, this definition of the AF episode is still experimental.

Other irregular rhythms

Almost half of the misclassifications were due to irregular rhythms without AF (49%). This further illustrates the fact that irregularity of RR-intervals is the main factor which is used in the classifier. Furthermore, the percentage of R-peaks with one P-wave is one of the major features used to classify features correctly. The current algorithm showed an F1-score for P-wave detection of 85.351%. This test was performed on data from the MIT-BIH QT database, which contains ECG signals which are, in general, without much noise and artefacts. In contrast, continuous rhythm registration can contain large noisy segments or artefacts. Therefore, P-wave detection will most likely be less accurate in the telemetry data compared to the MIT-BIH QT database. Inaccurate P-wave detection might be a factor resulting in other irregular rhythms being incorrectly detected as AF segments. Furthermore, only 4,800 30-seconds ECG segment were used to train the classifier. More segments might be needed to increase the generalizability of the classifier to other datasets.

Limitations and future research

Although a next step has been taken in the automated AF detection in post-operative telemetry data, the algorithm is not yet suitable for unsupervised and/or clinical use. In clinical use, the current FPR would lead to an overestimation of AF burden, potentially resulting in overtreatment. For the same reason, further improvements should be made before using the algorithm unsupervised in a research setting. Improvements should be focused on differentiation between AF and other arrhythmias, more accurate detection of atrial activity, and identification of AF episode boundaries.

Differentiation between atrial fibrillation and other arrhythmias

The most important features of the classifier are based on irregularity of RR-intervals. However, this irregularity can also be caused by other arrhythmias or inaccuracies in R-peak detection. Single PACs, PVCs, and missing beats were already excluded for RR-interval analyses. However, the accuracy of detecting these beats has not been validated yet. Also, the exclusion criteria all assumed a normal beat following and preceding the PAC, PVC, or missing beat. Therefore, when a PAC, PVC, or missing beat is followed by another irregular beat, the PAC, PVC, or missing beat will not be detected as such. Detection of these irregularities other than AF should be optimized. The definition of AF states that an episode should last at least 30 seconds. For other arrhythmias, the definition is based on the percentage of abnormal beats or number of abnormal beats per hour (e.g. PVCs⁴⁸). Therefore, instead of classifying 30-seconds segments with these other arrhythmias, beat-level classification might provide better insights into these arrhythmias.⁴⁹ The percentage of abnormal beats could be used as a feature for the AF classifier.

Detection of atrial activity

Accuracy of P- and T-wave detection might be too low for ECG segments containing noise and artefacts. Therefore, detection of atrial activity should be optimized. However, it remains a question whether the current equipment is suitable for measuring atrial activity accurately with a lot of noise and artefacts. The citation "garbage in, garbage out" might apply to detecting atrial activity in these ECG segments; the quality of the input signal might be too low to accurately detect atrial activity. If this is the case, the first step should be to optimize data acquisition as much as possible in the post-operative clinical setting (e.g. ensure good electrode-to-skin contact and use quality electrodes).

Identification of atrial fibrillation episode boundaries

Another limitation of the proposed algorithm is the use of a classifier to classify short 30-seconds ECG segments. This implies surrounding segments are ignored and segments are independent of each other. This is not the case for consecutive ECG segments. The current classification algorithm does not take into account the state of previous ECG segments (*AF*, *No AF*, or *Unusable*). For detection of start and end times of AF episodes, use of state transition diagrams, as previously described for rhythm identification of ventricular arrhythmias, might increase the accuracy.⁴⁹ This way, instead of classifying single ECG segments, the states of previous ECG segments could be used as input for the algorithm. Depending on the state of previous ECG segments, the likelihood of changing to another state could be determined.

Conclusion

An automated AF classifier based on post-operative continuous rhythm registrations for use in the research setting was proposed. An FPR and FNR of 3.8% and 2.8%, respectively, were reached. Main features for detection of AF are based on RR- and PQ-interval irregularities, and P-wave characteristics. Careful use of the classifier in combination with manual validation of the detected AF segments makes the classifier suitable for supervised research purposes.



References

References

1. Zulkifly H, Lip GYH and Lane DA. Epidemiology of atrial fibrillation. *Int J Clin Pract.* 2018;72:e13070.
2. Maesen B, Nijs J, Maessen J, Allesie M and Schotten U. Post-operative atrial fibrillation: a maze of mechanisms. *Europace.* 2012;14:159-74.
3. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, Gillum RF, Kim Y-H, McAnulty JH, Jr., Zheng Z-J, Forouzanfar MH, Naghavi M, Mensah GA, Ezzati M and Murray CJL. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation.* 2014;129:837-847.
4. Gudbjartsson T, Helgadottir S, Sigurdsson MI, Taha A, Jeppsson A, Christensen TD and Riber L. New-onset postoperative atrial fibrillation after heart surgery. *Acta Anaesthesiol Scand.* 2019.
5. Moore KL, Agur AMR and Dalley AF. Heart and Great Vessels *Essential Clinical Anatomy.* 5th ed. Baltimore: Wolters Kluwer Health; 2015: 81-110.
6. Lederer WJ. Cardiac Electrophysiology and the Electrogram. In: W. F. Boron and E. L. Boulpaep, eds. *Medical Physiology: A Cellular and Molecular Approach.* 2nd ed. Philadelphia: Elsevier Saunders; 2012: 504-528.
7. Einthoven W. Die galvanometrische Registrierung des menschlichen Elektrokardiogramms, zugleich eine Beurtheilung der Anwendung des Capillar-Elektrometers in der Physiologie. *Archiv für die gesamte Physiologie des Menschen und der Tiere.* 1903;99:472-480.
8. Dubin D. *Rapid Interpretation of EKG's.* 6th ed. Fort Myers, Florida: COVER Publishing Company; 2000.
9. Kumar K. Overview of atrial fibrillation. 2019. [ONLINE] Available at: <https://www.uptodate.com/contents/overview-of-atrial-fibrillation> [Accessed 8 January, 2020].
10. January CT, Wann LS, Alpert JS, Calkins H, Cigarroa JE, Cleveland JC, Jr., Conti JB, Ellinor PT, Ezekowitz MD, Field ME, Murray KT, Sacco RL, Stevenson WG, Tchou PJ, Tracy CM, Yancy CW and Members AATF. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the Heart Rhythm Society. *Circulation.* 2014;130:2071-104.
11. Kumar K and Manning WJ. Rhythm control versus rate control in atrial fibrillation. 2019. [ONLINE] Available at: <https://www.uptodate.com/contents/rhythm-control-versus-rate-control-in-atrial-fibrillation> [Accessed 8 January, 2020].
12. Van Gelder IC, Hagens VE, Bosker HA, Kingma JH, Kamp O, Kingma T, Said SA, Darmanata JI, Timmermans AJ, Tijssen JG, Crijns HJ and Rate Control versus Electrical Cardioversion for Persistent Atrial Fibrillation Study G. A comparison of rate control and rhythm control in patients with recurrent persistent atrial fibrillation. *N Engl J Med.* 2002;347:1834-40.
13. Olshansky B and Arora R. Mechanisms of Atrial Fibrillation. 2019. [ONLINE] Available at: <https://www.uptodate.com/contents/mechanisms-of-atrial-fibrillation> [Accessed 8 January, 2020].
14. Yamazaki M and Jalife J. Pathophysiology of atrial fibrillation: From initiation to maintenance. *Journal of Arrhythmia.* 2012;28:129-139.

15. Brundel BJ, Van Gelder IC, Henning RH, Tieleman RG, Tuinenburg AE, Wietes M, Grandjean JG, Van Gilst WH and Crijns HJ. Ion channel remodeling is related to intraoperative atrial effective refractory periods in patients with paroxysmal and persistent atrial fibrillation. *Circulation*. 2001;103:684-90.
16. Waks JW and Josephson ME. Mechanisms of Atrial Fibrillation - Reentry, Rotors and Reality. *Arrhythm Electrophysiol Rev*. 2014;3:90-100.
17. Melby SJ, George JF, Picone DJ, Wallace JP, Davies JE, George DJ and Kirklin JK. A time-related parametric risk factor analysis for postoperative atrial fibrillation after heart surgery. *J Thorac Cardiovasc Surg*. 2015;149:886-92.
18. Bessissow A, Khan J, Devereaux PJ, Alvarez-Garcia J and Alonso-Coello P. Postoperative atrial fibrillation in non-cardiac and cardiac surgery: an overview. *Journal of thrombosis and haemostasis : JTH*. 2015;13 Suppl 1:S304-12.
19. Mathew JP, Fontes ML, Tudor IC, Ramsay J, Duke P, Mazer CD, Barash PG, Hsu PH, Mangano DT, Research ftlotl, Foundation E and Group tMSoPIR. A Multicenter Risk Index for Atrial Fibrillation After Cardiac Surgery. *JAMA*. 2004;291:1720-1729.
20. Maisel WH, Rawn JD and Stevenson WG. Atrial Fibrillation after Cardiac Surgery. *Annals of Internal Medicine*. 2001;135:1061-1073.
21. Mouws E, Lanters EAH, Teuwen CP, van der Does L, Kik C, Knops P, Bekkers JA, Bogers A and de Groot NMS. Epicardial Breakthrough Waves During Sinus Rhythm: Depiction of the Arrhythmogenic Substrate? *Circ Arrhythm Electrophysiol*. 2017;10.
22. Knops P, Kik C, Bogers AJ and de Groot NM. Simultaneous endocardial and epicardial high-resolution mapping of the human right atrial wall. *J Thorac Cardiovasc Surg*. 2016;152:929-31.
23. Holm M, Johansson R, Brandt J, Luhrs C and Olsson SB. Epicardial right atrial free wall mapping in chronic atrial fibrillation. Documentation of repetitive activation with a focal spread--a hitherto unrecognised phenomenon in man. *Eur Heart J*. 1997;18:290-310.
24. Eckstein J, Zeemering S, Linz D, Maesen B, Verheule S, van Hunnik A, Crijns H, Allessie MA and Schotten U. Transmural conduction is the predominant mechanism of breakthrough during atrial fibrillation: evidence from simultaneous endo-epicardial high-density activation mapping. *Circ Arrhythm Electrophysiol*. 2013;6:334-41.
25. de Groot NM, Houben RP, Smeets JL, Boersma E, Schotten U, Schalij MJ, Crijns H and Allessie MA. Electropathological substrate of longstanding persistent atrial fibrillation in patients with structural heart disease: epicardial breakthrough. *Circulation*. 2010;122:1674-82.
26. de Groot N, van der Does L, Yaksh A, Lanters E, Teuwen C, Knops P, van de Woestijne P, Bekkers J, Kik C, Bogers A and Allessie M. Direct Proof of Endo-Epicardial Asynchrony of the Atrial Wall During Atrial Fibrillation in Humans. *Circ Arrhythm Electrophysiol*. 2016;9.
27. de Groot NMS and Allessie MA. Pathophysiology of atrial fibrillation: Focal patterns of activation. *Pacing Clin Electrophysiol*. 2019;42:1312-1319.
28. Chen LY, Chung MK, Allen LA, Ezekowitz M, Furie KL, McCabe P, Noseworthy PA, Perez MV and Turakhia MP. Atrial Fibrillation Burden: Moving Beyond Atrial Fibrillation as a Binary Entity: A Scientific Statement From the American Heart Association. *Circulation*. 2018;137:e623-e644.

29. Wesselius FJ, van Schie MS, Roos MC and de Groot NMS. Classification Algorithms for Atrial Fibrillation using Surface Electrocardiograms: A Systematic Review. 2020. delivered 15 January 2020.
30. Go AS, Reynolds K, Yang J, Gupta N, Lenane J, Sung SH, Harrison TN, Liu TI and Solomon MD. Association of Burden of Atrial Fibrillation With Risk of Ischemic Stroke in Adults With Paroxysmal Atrial Fibrillation: The KP-RHYTHM Study. *JAMA Cardiology*. 2018;3:601-608.
31. Kaplan RM, Koehler J, Ziegler PD, Sarkar S, Zweibel S and Passman RS. Stroke Risk as a Function of Atrial Fibrillation Duration and CHA₂DS₂-VASc Score. *Circulation*. 2019;140:1639-1646.
32. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK and Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. 2000;101:E215-20.
33. Moody GB and Mark RG. A new method for detecting atrial fibrillation using R-R intervals. *Computers in Cardiology*. 1983;10:227-230.
34. Moody GB and Mark RG. The impact of the MIT-BIH arrhythmia database. *IEEE Eng Med Biol Mag*. 2001;20:45-50.
35. Laguna P, Mark RG, Goldberger AL and Moody GB. A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG. *Computers in Cardiology*. 1997;24:673-676.
36. Pan J and Tompkins WJ. A real-time QRS detection algorithm. *IEEE Trans Biomed Eng*. 1985;32:230-6.
37. Sahambi JS, Tandon SN and Bhatt RKP. Using wavelet transforms for ECG characterization. An on-line digital signal processing system. *IEEE Engineering in Medicine and Biology Magazine*. 1997;16:77-83.
38. Thakor NV, Webster JG and Tompkins WJ. Estimation of QRS Complex Power Spectra for Design of a QRS Filter. *IEEE Transactions on Biomedical Engineering*. 1984;BME-31:702-706.
39. Elgendi M, Meo M and Abbott D. A Proof-of-Concept Study: Simple and Effective Detection of P and T Waves in Arrhythmic ECG Signals. *Bioengineering (Basel)*. 2016;3:26.
40. Park J, Lee S and Jeon M. Atrial fibrillation detection by heart rate variability in Poincare plot. *BioMedical Engineering OnLine*. 2009;8:38.
41. Dharmapran D, Dykes L, McGavigan AD, Kuklik P, Pope K and Ganesan AN. Information Theory and Atrial Fibrillation (AF): A Review. *Frontiers in Physiology*. 2018;9.
42. Bazett HC. An Analysis of the Time-Relations of Electrocardiograms. *Annals of Noninvasive Electrocardiology*. 1997;2:177-194.
43. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med*. 1998;17:857-72.
44. MathWorks. Ensemble Algorithms. 2020. [ONLINE] Available at: <https://nl.mathworks.com/help/stats/ensemble-algorithms.html> [Accessed 14 April, 2020].
45. MathWorks. Estimates of predictor importance for classification tree. 2020. [ONLINE] Available at:

- <https://nl.mathworks.com/help/stats/compactclassificationtree.predictorimportance.html> [Accessed 25 May, 2020].
46. PhysioNet. MIT-BIH Atrial Fibrillation Database. 2000. [ONLINE] Available at: <https://physionet.org/content/afdb/1.0.0/> [Accessed June 12, 2020].
 47. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, Castella M, Diener H-C, Heidbuchel H, Hendriks J, Hindricks G, Manolis AS, Oldgren J, Popescu BA, Schotten U, Van Putte B, Vardas P and Group ESD. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *European Heart Journal*. 2016;37:2893-2962.
 48. Al-Khatib SM, Stevenson WG, Ackerman MJ, Bryant WJ, Callans DJ, Curtis AB, Deal BJ, Dickfeld T, Field ME, Fonarow GC, Gillis AM, Granger CB, Hammill SC, Hlatky MA, Joglar JA, Kay GN, Matlock DD, Myerburg RJ and Page RL. 2017 AHA/ACC/HRS Guideline for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death. *Circulation*. 2018;138:e272-e391.
 49. Pathangay V and Rath SP. Arrhythmia detection in single-lead ECG by combining beat and rhythm-level information. *Conf Proc IEEE Eng Med Biol Soc*. 2014;2014:3236-9.
 50. Franklin GF, Powell JD and Emami-Naeini A. *Digital Control Feedback Control of Dynamic Systems*. 7th ed. Harlow: Pearson Education Limited; 2015: 610-656.
 51. Hsu HP. The z-Transform and Discrete-Time LTI Systems *Signals and Systems*. 3rd ed. New York: McGraw-Hill Education; 2014: 148-192.
 52. MathWorks. Prominence. 2020. [ONLINE] Available at: <https://www.mathworks.com/help/signal/ug/prominence.html> [Accessed 25 May, 2020].
 53. MathWorks. Rotation matrix for rotations around z-axis. 2020. [ONLINE] Available at: <https://www.mathworks.com/help/phased/ref/rotz.html> [Accessed 29 May, 2020].
 54. Hsu HP. Review of Probability *Signals and Systems*. 3rd ed. New York: McGraw-Hill Education; 2014: 459-477.
 55. MathWorks. Growing Decision Trees. 2020. [ONLINE] Available at: <https://www.mathworks.com/help/stats/growing-decision-trees.html> [Accessed 10 June, 2020].
 56. Mather P and Tso B. Gini Impurity Index *Classification Methods for Remotely Sensed Data*. 2nd ed. Boca Raton: CRC Press; 2016: 188-189.
 57. Li Y, Gao J, Li Q and Fan W. Ensemble Learning. In: C. C. Aggarwal, ed. *Data Classifications: Algorithms and Applications*. 1st ed. Boca Raton: CRC Press; 2015: 483-510.
 58. Albon C. Adaboost Classifier. 2017. [ONLINE] Available at: https://chrisalbon.com/machine_learning/trees_and_forests/adaboost_classifier/ [Accessed June 10, 2020].
 59. MathWorks. Hyperparameter Optimization in Classification Learner App. 2020. [ONLINE] Available at: <https://www.mathworks.com/help/stats/hyperparameter-optimization-in-classification-learner-app.html> [Accessed 10 June, 2020].
 60. Frazier PI. A Tutorial on Bayesian Optimization. 2018. [ONLINE] Available at: <https://arxiv.org/pdf/1807.02811.pdf> [Accessed June 11, 2020].
 61. MathWorks. Bayesian Optimization Algorithm. 2020. [ONLINE] Available at: <https://www.mathworks.com/help/stats/bayesian-optimization-algorithm.html> [Accessed 11 June, 2020].

62. Turner R. Gaussian Processes. 2016. Lecture, Machine Learning Tutorial Series, Imperial College London, delivered November 11, 2016.
63. Rasmussen CE and Williams CKI. *Gaussian Processes for Machine Learning*. Cambridge: MIT Press; 2006.
64. Hoffman MW. Bayesian Optimization. 2018. Lecture, Uncertainty in Artificial Intelligence 2018, delivered August 6, 2018.



Appendices

Appendix I. Introduction to signal analysis and filter design using the z-transform

From analog to digital

Electrocardiograms (ECGs) are measured using dedicated measuring devices. Electrocardiograms cannot be measured continuously and therefore measurements are taken with a certain sampling rate using an analog-to-digital converter. The sampling rate is defined as the number of samples per second (Hz). The output of the device is therefore not described by a continuous function $x(t)$, where t is the time, but by a discrete function $x[n]$, where n is the number of samples. Figure 22 shows an example of a continuous-time signal and the associated discrete-time signal with a sampling rate of 30Hz. As can be seen from the figure, slow changes are recognizable in the discrete-time signal, but fast changes are missed as a result of the sampling rate. To reconstruct the original signal perfectly, the Nyquist theorem states that the sampling rate must be twice the highest frequency component of the signal.⁵⁰

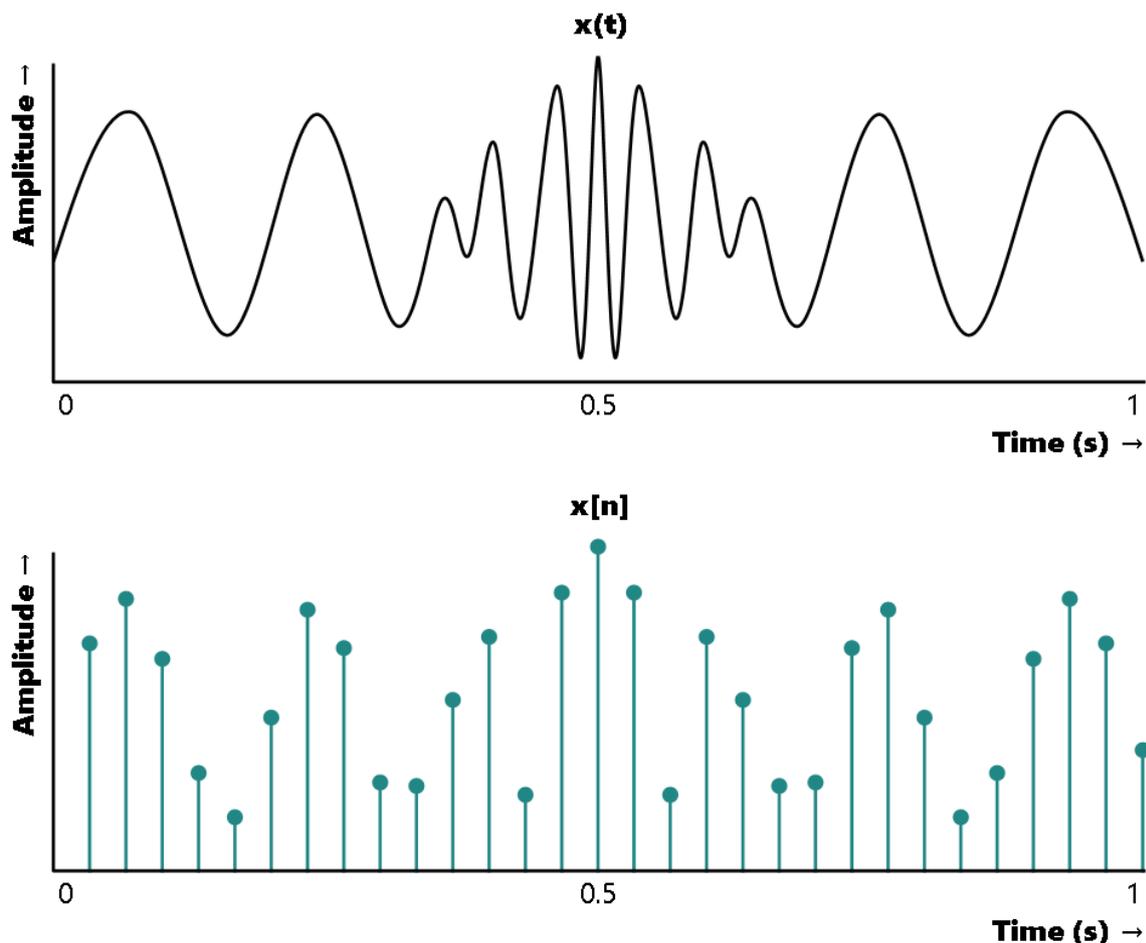


Figure 22 – Example of continuous-time signal ($x(t)$, upper panel) and the associated discrete-time signal ($x[n]$, lower panel) with a sampling rate of 30Hz.

Discrete-time signals in the z-domain

Instead of describing a discrete-time signal $x[n]$, the signal can also be described in the z -domain, where z is a complex variable.⁵¹ This way, difference equations are converted into more straightforward algebraic equations. The z -transform $X(z)$ of a discrete-time signal $x[n]$ is defined as:

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n} \quad (\text{Eq. A1.1})$$

where z is a complex variable defined as $re^{j\Omega}$. r and Ω describe the magnitude and the angle of z , respectively.

For example, $x[n] = \{3, 7, -2, 5, 8\}$ for $n = \{-1, 0, 1, 2, 3\}$ transforms into

$$X(z) = 3z + 7 - 2z^{-1} + 5z^{-2} + 8z^{-3} \quad (\text{Eq. A1.2})$$

The z -transform has several properties, of which the time-shift property, the linearity property and the convolution property are most relevant for filter design.⁵¹ The time-shift property states that if $X(z)$ is the z -transform of $x[n]$, the z -transform of $x[n - n_0]$ equals $z^{-n_0}X(z)$. The linearity property states that if the z -transforms of $x_1[n]$ and $x_2[n]$ are $X_1(z)$ and $X_2(z)$, respectively, the sum of x_1 and x_2 has a z -transform of $X_1(z) + X_2(z)$. The convolution property states that if the z -transforms of $x_1[n]$ and $x_2[n]$ are $X_1(z)$ and $X_2(z)$, respectively, the z -transform of the convolution $x_1[n] * x_2[n]$ equals $X_1(z)X_2(z)$.

Filter design using the z-transform

Filters are applied using a convolution between the original signal $x[n]$ and the filter $h[n]$, resulting in $y[n]$, which can be written as a sum:

$$y[n] = \sum_{m=-\infty}^{\infty} x[m] \cdot h[n - m] \quad (\text{Eq. A1.3})$$

where m introduces a delay of the signal.

In the z -domain, the convolution can be rewritten as a multiplication between the z -transform of the original signal, $X(z)$, and the transfer function of the filter, $H(z)$:

$$Y(z) = X(z)H(z) \quad (\text{Eq. A1.4})$$

Filters are commonly described using linear constant-coefficient difference equations of the form:

$$\sum_{k=0}^N a_k y[n - k] = \sum_{k=0}^N b_k x[n - k] \quad (\text{Eq. A1.5})$$

where $x[n]$ is the original signal, $y[n]$ is the output signal, k introduces a delay of the signal, and a_k and b_k are the filter coefficients.

Using the z-transform, this can be written as:

$$\sum_{k=0}^N a_k z^{-k} Y(z) = \sum_{k=0}^N b_k z^{-k} X(z) \quad (\text{Eq. A1.6})$$

$$Y(z) \sum_{k=0}^N a_k z^{-k} = X(z) \sum_{k=0}^N b_k z^{-k} \quad (\text{Eq. A1.7})$$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^N b_k z^{-k}}{\sum_{k=0}^N a_k z^{-k}} \quad (\text{Eq. A1.8})$$

For example, the derivative filter used in the Pan-Tompkins algorithm has a difference equation of $y[nT] = \frac{1}{8T}(-x[nT - 2T] - 2x[nT - T] + 2x[nT + T] + x[nT + 2T])$, which can be written as $8T \cdot y[nT] = -x[nT - 2T] - 2x[nT - T] + 2x[nT + T] + x[nT + 2T]$, where x is the input signal and y is the output signal, and T is the sampling period.³⁶ Using the z-transform, this can be written as:

$$Y(z) \cdot (8T) = X(z) \cdot (-z^{-2} - 2z^{-1} + 2z^1 + z^2) \quad (\text{Eq. A1.9})$$

Therefore,

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{8T}(-z^{-2} - 2z^{-1} + 2z^1 + z^2) \quad (\text{Eq. A1.10})$$

Appendix II. Peak prominence – An example

In Figure 23 the measurement of peak prominence is visualized and explained. For calculation of peak prominence, first a window is determined around each peak. A horizontal line from the peak extending to the left and right is drawn. The intersections of this line with the signal on the left and right of the peak are the boundaries of the window for that peak. Within this window, prominence is defined as⁵²:

$$\text{Prominence} = \min(A_{max} - A_{min,L}, A_{max} - A_{min,R}) \quad (\text{Eq. A2.1})$$

A_{max} = Peak amplitude

$A_{min,L}$ = Minimum amplitude preceding the peak

$A_{min,R}$ = Minimum amplitude succeeding the peak

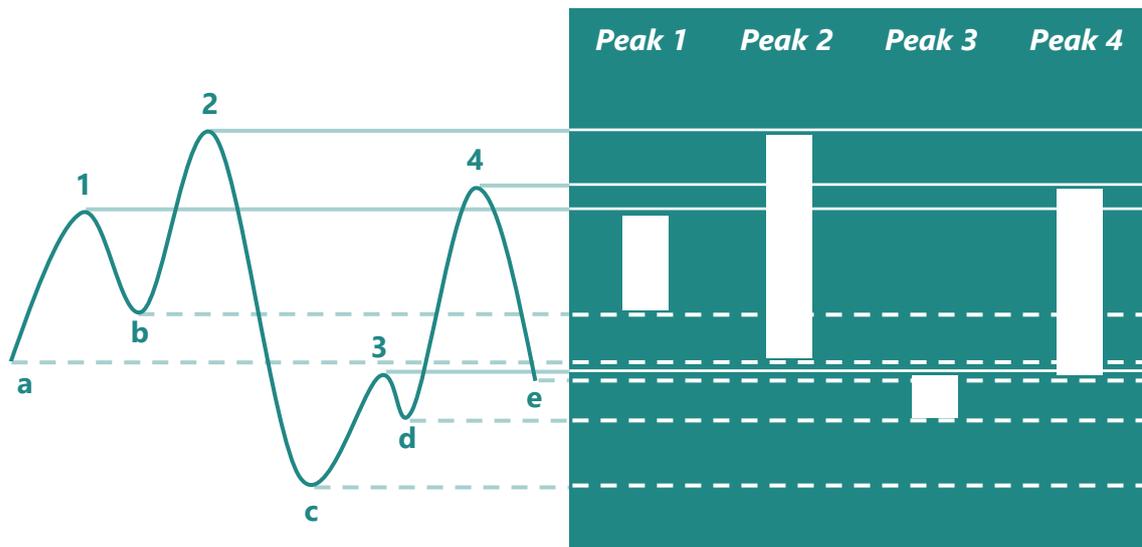


Figure 23 – Peak prominence measurement. Peak prominence is defined as the minimum difference between a local maximum and the minima on the left side of the peak and the right side of the peak, within a window demarcated by the first preceding and succeeding samples of which the signal is equal or higher than the local maximum. Four examples are visualized:

Peak 1: the window of peak 1 starts at the beginning of the signal and ends at the line segment between local minimum *b* and peak 2. The minimum left from the peak is at point *a* and the minimum right from the peak is at point *b*. The difference between peak 1 and local minimum *b* is smallest, hence the prominence of peak 1 equals the difference between peak 1 and local minimum *b*.

Peak 2: the window of peak 2 starts at the beginning of the signal and ends at the end of the signal. The minimum left from the peak is at point *a* and the minimum right from the peak is at point *c*. The difference between peak 2 and local minimum *a* is smallest, hence the prominence of peak 2 equals the difference between peak 2 and local minimum *a*.

Peak 3: the window of peak 3 starts at the line segment between peak 2 and local minimum *c* and ends at the line segment between local minimum *d* and peak 4. The minimum left from the peak is at point *c* and the minimum right from the peak is at point *d*. The difference between peak 3 and local minimum *d* is smallest, hence the prominence of peak 3 equals the difference between peak 3 and local minimum *d*.

Peak 4: the window of peak 4 starts at the line segment between peak 2 and local minimum *c* and ends at the end of the signal. The minimum left from the peak is at point *c* and the minimum right from the peak is at point *e*. The difference between peak 4 and local minimum *e* is smallest, hence the prominence of peak 4 equals the difference between peak 4 and local minimum *e*.

Appendix III. The sigmoid function

Sigmoid functions, or logistic functions, are mainly used for the prediction of probabilities. The standard sigmoid function is the function:

$$y(x) = \frac{1}{1 + e^{-x}} \quad (\text{Eq. A3.1})$$

Figure 24 shows a plot of this sigmoid function. When x tends to $-\infty$, y will be almost 0; when x equals 0, y will be 0.5; and when x tends to $+\infty$, y will be almost 1.

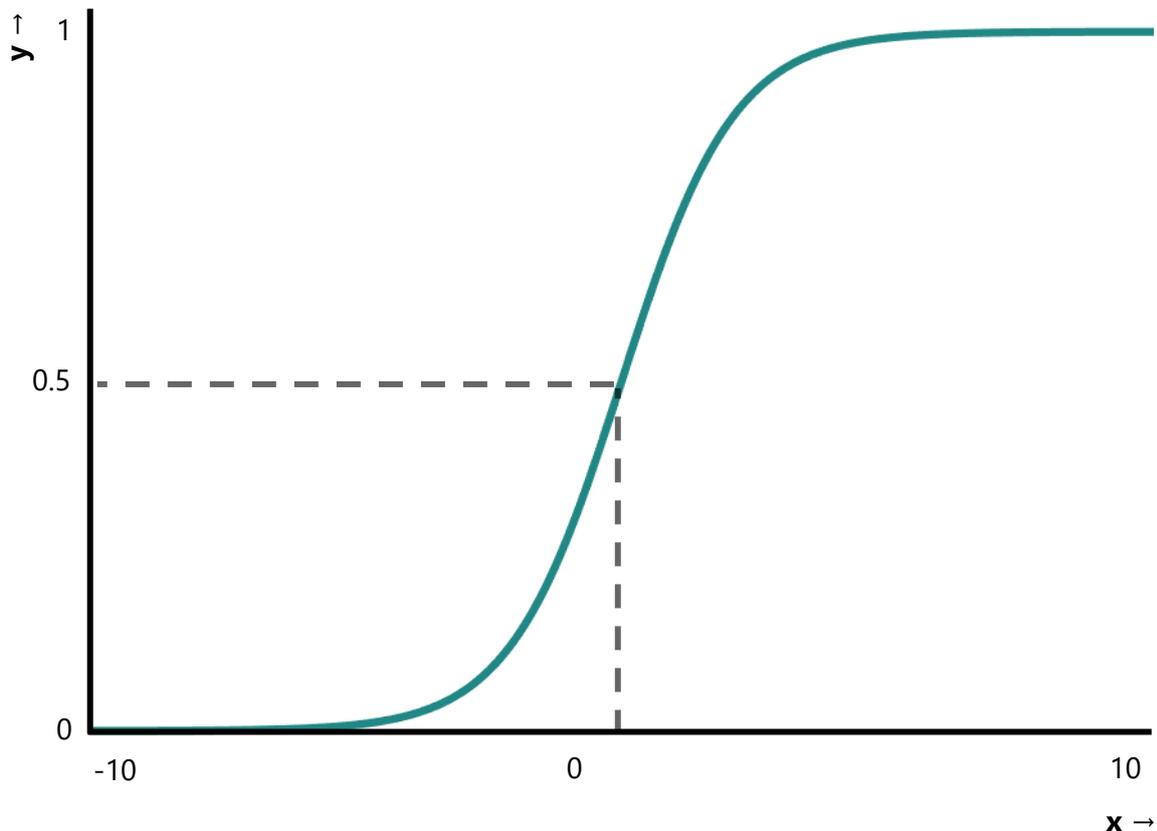


Figure 24 – Standard sigmoid function $y(x) = 1/(1 + e^{-x})$ for x from -10 to 10.

The sigmoid function can be manipulated by adding parameters to the function:

$$y(x) = A * \frac{1}{1 + e^{-B*(x-C)}} + D \quad (\text{Eq. A3.2})$$

A is a scaling factor for the y -axis, B is a scaling factor for the x -axis, C adds an offset to the value of x (i.e. shifts along the x -axis), and D adds an offset to the value of y (i.e. shifts along the y -axis). Examples for all four parameters are provided in Figure 25. It is important to note that if either A or B is negative, the function has a negative slope instead of a positive slope. If A is negative, the function is mirrored about the x -axis, whereas if B is negative, the function is mirrored about the y -axis. The parameters can be combined to create sigmoid functions with the desired slope and offset.

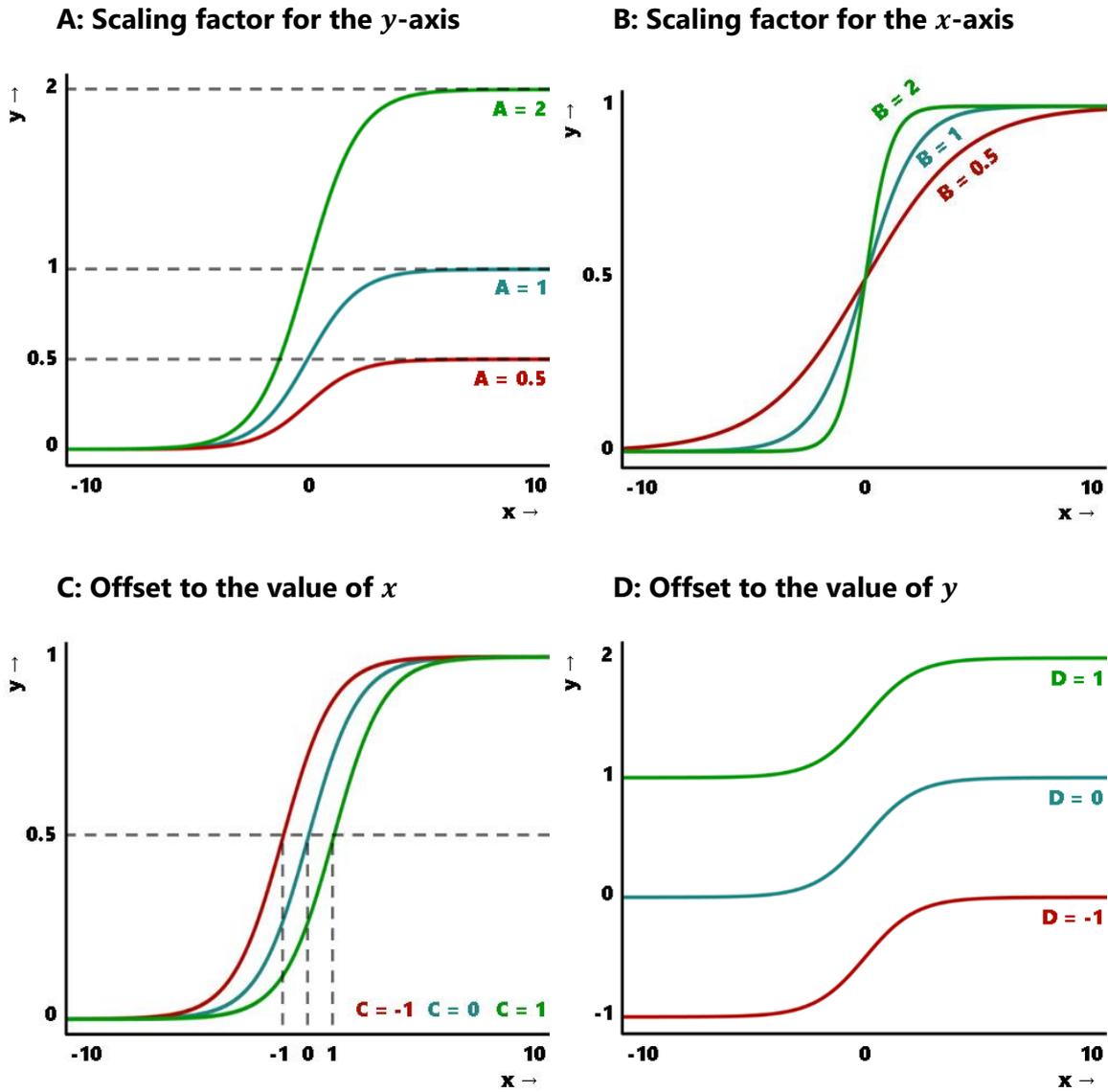


Figure 25 – Sigmoid function $y(x) = A * 1 / (1 + e^{(-B * (x - C))}) + D$ for x from -10 to 10. Default values of parameters A , B , C and D were 1, 1, 0, and 0, respectively. A is a scaling factor for the y-axis, B is a scaling factor for the x-axis. C adds an offset to the value of x (i.e. shifts along the x-axis). D adds an offset to the value of y (i.e. shifts along the y-axis). Each panel shows the effect of changing one parameter, while leaving the other parameters constant.

Appendix IV. Calculating the standard deviation of distances from points to lines through the origin using the rotation matrix – An example

To calculate standard deviations of distances from points to perpendicular lines through the origin, the standard coordinate system is rotated in the Euclidean space using the rotation matrix⁵³:

$$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad (\text{Eq. A4.1})$$

By definition, counterclockwise rotations have negative angles and clockwise rotations have positive angles. In Figure 26, an example is shown with a counterclockwise rotation of 45° ($=-\pi/4$ radians). Multiplying the rotation matrix with the original x - and y -coordinates results in the transformed x - and y -coordinates:

$$\begin{bmatrix} x_{\text{transformed}} \\ y_{\text{transformed}} \end{bmatrix} = R\left(-\frac{\pi}{4}\right) \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \cdot \cos\left(-\frac{\pi}{4}\right) - y \cdot \sin\left(-\frac{\pi}{4}\right) \\ x \cdot \sin\left(-\frac{\pi}{4}\right) + y \cdot \cos\left(-\frac{\pi}{4}\right) \end{bmatrix} \quad (\text{Eq. A4.2})$$

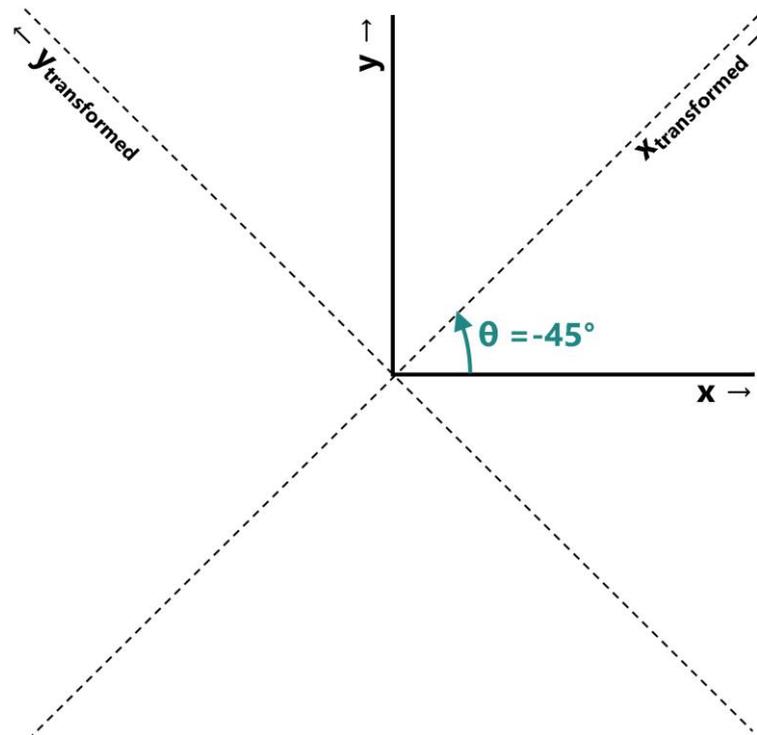


Figure 26 – Example of rotating the standard coordinate system 45° counterclockwise

An illustrative example is provided in Figures 27a-c. Figure 27a shows the example as used in the methods section of this thesis (RR-interval characteristics – Feature extraction, p. 35). Using Eq. A4.2, the coordinate system is rotated 45° counterclockwise, as visualized in Figure 27b. This way, the horizontal and vertical axes are parallel to the -45°- and 45°-line, respectively.

The standard deviation σ is defined as⁵⁴:

$$\sigma(\mathbf{x}) = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu_x)^2}{n - 1}} \quad (\text{Eq. A4.3})$$

where \mathbf{x} is the input vector, n is the sample size, and μ_x is the mean of the input vector \mathbf{x} .

The standard deviation is independent of the mean of the input vector, because μ_x is subtracted from each value in \mathbf{x} . Therefore, as shown in Figure 27c, the standard deviation of distances of points to any line parallel to the horizontal axis can be calculated as the standard deviation of $y_{transformed}$. Similarly, the standard deviation of distances of points to any line parallel the vertical axis can be calculated as the standard deviation of $x_{transformed}$. In this case, this means that the distances of points to the -45°- and 45°-line can be calculated as the standard deviations of $y_{transformed}$ and $x_{transformed}$, respectively.

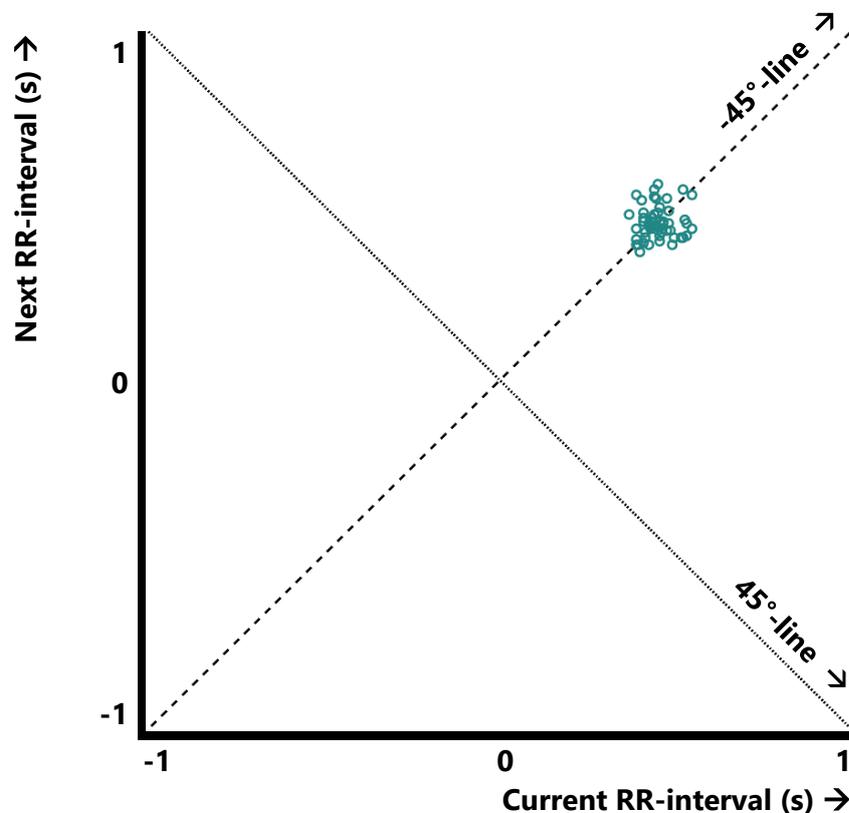


Figure 27a – Poincaré plot of RR-intervals during atrial fibrillation, as also described in the thesis (Figure 14, p. 35). Points (o) indicate two successive RR-intervals, the dashed line (- - -) and the dotted line (···) indicate the -45°-line and the 45°-line, respectively, as described in text (Feature extraction, p. 35).

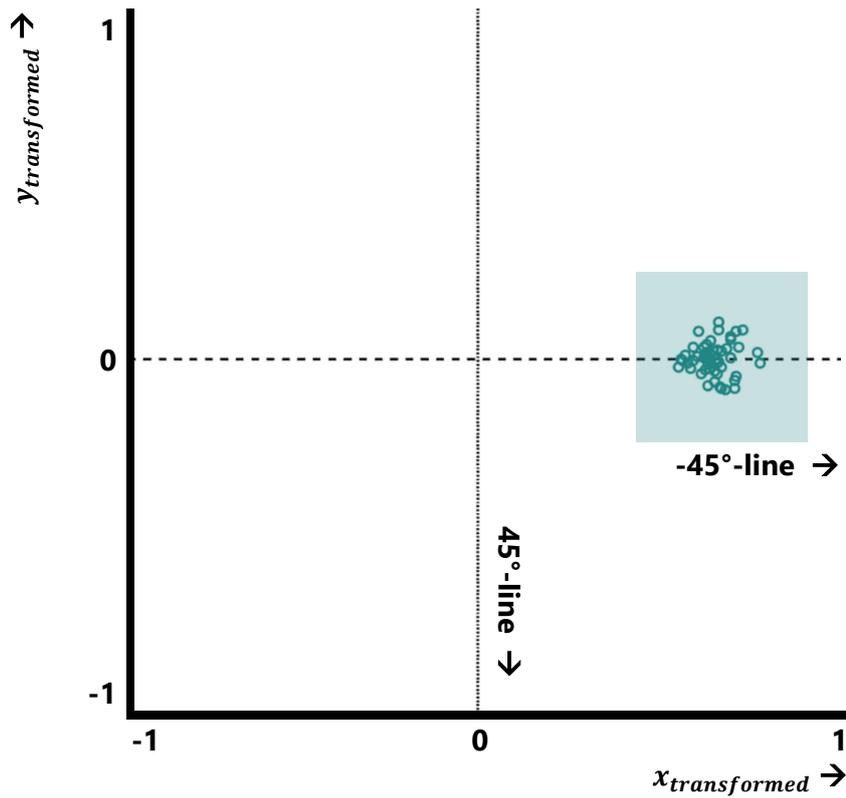


Figure 27b – Poincaré plot of RR-intervals during atrial fibrillation after rotating the coordinate system 45° counterclockwise. Points (o) indicate two successive RR-intervals, the dashed line (- - -) and the dotted line (· · ·) indicate the -45° -line and the 45° -line, respectively, as described in text (Feature extraction, p. 35). The shaded area is further visualized in Figure 27c.

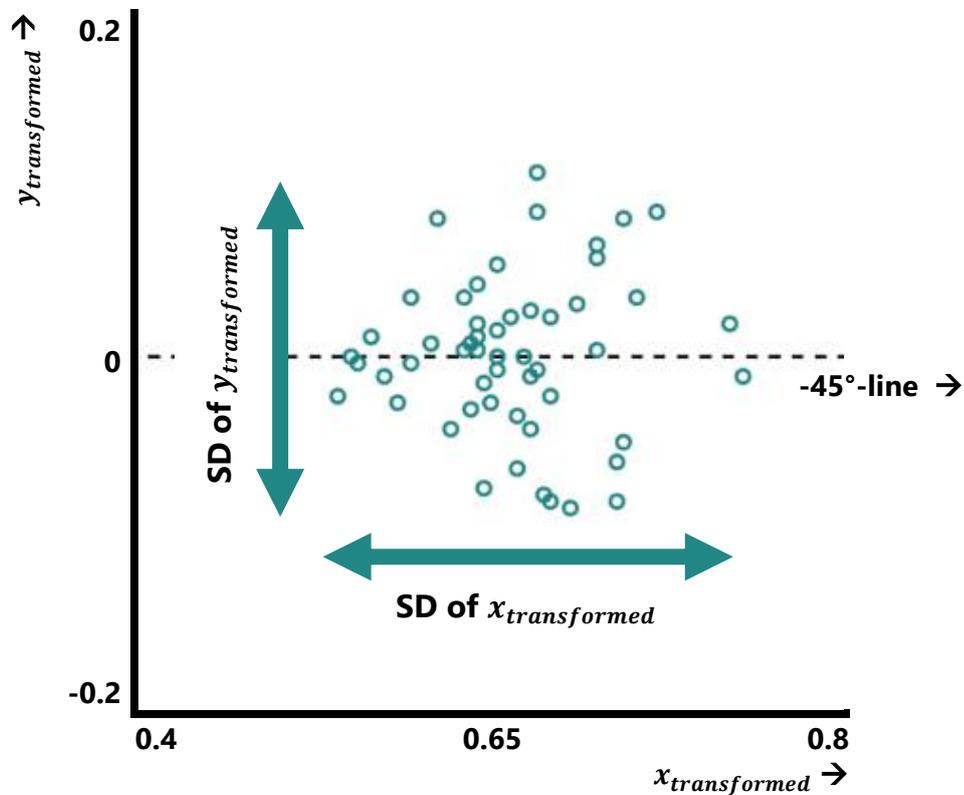


Figure 27c – Calculating the standard deviations (SD) of $x_{transformed}$ and $y_{transformed}$. Points (o) indicate two successive RR-intervals, and the dashed line (- - -) indicates the -45° -line, as described in text (Feature extraction, p. 35). The 45° -line is not visualized as a result of horizontal scaling.

Appendix V. Shannon Entropy – An example

Shannon entropy describes the information content of a variable x and is defined as⁴¹:

$$\text{Shannon entropy} = - \sum_{i=1}^M p(i) \log_2 p(i) \quad (\text{Eq. A5.1})$$

where M equals the number of discrete values the variable can take and $p(i)$ is the probability density function of variable x for the i^{th} value.

Shannon entropy is given in the amount of information in bits. This is most easily explained using a coin toss with the outcome of the coin toss as the variable x . In this case, two possible outcomes are heads and tails. Both outcomes have equal chances (1/2). Therefore, the Shannon entropy will be:

$$-\left(\frac{1}{2} \cdot \log_2 \left(\frac{1}{2}\right) + \frac{1}{2} \cdot \log_2 \left(\frac{1}{2}\right)\right) = 1 \text{ bit} \quad (\text{Eq. A5.2})$$

Likewise, using a fair dice with six different sides, six outcomes of the dice are possible. All outcomes have equal chances (1/6). Therefore, the Shannon entropy will be:

$$\begin{aligned} -\left(\frac{1}{6} \cdot \log_2 \left(\frac{1}{6}\right) + \frac{1}{6} \cdot \log_2 \left(\frac{1}{6}\right) + \frac{1}{6} \cdot \log_2 \left(\frac{1}{6}\right) + \frac{1}{6} \cdot \log_2 \left(\frac{1}{6}\right) \right. \\ \left. + \frac{1}{6} \cdot \log_2 \left(\frac{1}{6}\right) + \frac{1}{6} \cdot \log_2 \left(\frac{1}{6}\right)\right) \approx 2.59 \text{ bits} \end{aligned} \quad (\text{Eq. A5.3})$$

Lastly, if all sides of the dice have the similar outcome, the Shannon entropy will be:

$$-(1 \cdot \log_2(1)) = 0 \text{ bits} \quad (\text{Eq. A5.4})$$

The higher value for the fair dice with six different sides indicates that a throw with this dice will give 2.59 bits of information. Using the last dice, the outcome is already known beforehand, hence throwing the dice would give 0 bits of information.

Appendix VI. Detection of QRS-complexes

Results per record

Table 7 – Detection of QRS-complexes, results per record

Record	Annotated	Detected	True positive	False positive	False negative
MIT-BIH Normal Sinus Rhythm Database (NSRDB)					
16265	100243	100480	100243	237	0
16272	87758	91204	85725	5479	2033
16273	89845	89845	89838	7	7
16420	102067	102102	102058	44	9
16483	104334	104343	104326	17	8
16539	108282	108291	108279	12	3
16773	81989	81987	81987	0	2
16786	101615	101633	101614	19	1
16795	86872	86904	86870	34	2
17052	87356	87570	87355	215	1
17453	100658	100684	100656	28	2
18177	115911	116194	115894	300	17
18184	102313	102363	102303	60	10
19088	97961	99622	97960	1662	1
19090	81391	81571	81387	184	4
19093	75106	75187	75106	81	0
19140	96596	96757	96596	161	0
19830	109333	110339	109272	1067	61
Overall	1729630	1737076	1727469	9607	2161
MIT-BIH Atrial Fibrillation Database (AFDB)					
04015	44005	44430	42944	1486	1061
04043	61915	63218	61788	1430	127
04048	39934	40109	39795	314	139
04126	42860	43662	42190	1472	670
04746	47873	49059	47787	1272	86
04908	61760	63136	61030	2106	730
04936	53646	54981	50786	4195	2860
05091	36793	35126	34036	1090	2757
05121	49881	47712	45824	1888	4057
05261	45534	46524	45348	1176	186
06426	55155	54385	52309	2076	2846
06453	34837	35122	34296	826	541
06995	55189	56357	54998	1359	191
07162	39298	40302	8698	31604	30600
07859	60266	61669	57094	4575	3172
07879	56594	57887	56302	1585	292

Record	Annotated	Detected	True positive	False positive	False negative
07910	36599	37809	36590	1219	9
08215	43356	44195	43219	976	137
08219	59293	60714	58772	1942	521
08378	45515	46308	44779	1529	736
08405	58856	60313	58525	1788	331
08434	39850	40753	39752	1001	98
08455	59552	59395	59055	340	497
Overall	1128561	1143166	1075917	67249	52644
MIT-BIH Long-Term ECG Database (LTDB)					
14046	115278	108264	108143	121	7135
14134	49632	49637	49625	12	7
14149	144818	144463	144355	108	463
14157	88104	88090	88057	33	47
14172	66006	65732	65626	106	380
14184	101490	101405	101405	0	85
15814	103354	102957	102826	131	528
Overall	668682	660548	660037	511	8645
MIT-BIH Arrhythmia Database (MITDB)					
100	2273	2272	2272	0	1
101	1865	1868	1864	4	1
102	2187	2187	2184	3	3
103	2084	2083	2083	0	1
104	2229	2220	2200	20	29
105	2572	2597	2560	37	12
106	2027	1959	1959	0	68
107	2137	2134	2134	0	3
108	1763	1759	1730	29	33
109	2532	2521	2521	0	11
111	2124	2124	2123	1	1
112	2539	2540	2539	1	0
113	1795	1794	1794	0	1
114	1879	1878	1875	3	4
115	1953	1952	1952	0	1
116	2412	2386	2384	2	28
117	1535	1536	1535	1	0
118	2278	2279	2278	1	0
119	1987	1987	1987	0	0
121	1863	1861	1861	0	2
122	2476	2476	2476	0	0
123	1518	1518	1518	0	0
124	1619	1618	1618	0	1
200	2601	2607	2593	14	8

Record	Annotated	Detected	True positive	False positive	False negative
201	1963	1917	1917	0	46
202	2136	2125	2125	0	11
203	2980	2918	2885	33	95
205	2656	2652	2647	5	9
207	1860	2048	1795	253	65
208	2955	3051	2898	153	57
209	3005	3005	3005	0	0
210	2650	2605	2600	5	50
212	2748	2748	2748	0	0
213	3251	3241	3241	0	10
214	2262	2257	2255	2	7
215	3363	3356	3355	1	8
217	2208	2204	2202	2	6
219	2154	2152	2152	0	2
220	2048	2047	2047	0	1
221	2427	2415	2415	0	12
222	2483	2467	2465	2	18
223	2605	2579	2579	0	26
228	2053	2073	2039	34	14
230	2256	2256	2256	0	0
231	1571	1571	1571	0	0
232	1780	1783	1779	4	1
233	3079	3068	3068	0	11
234	2753	2748	2748	0	5
Overall	109494	109442	108832	610	662

Appendix VII. Detection of P- and T-waves

Results per record

Table 8 – Detection of P- and T-waves, results per record

Record	Annotated	Detected	True positive	False positive	False negative
P-wave detection using the MIT-BIH QT Database (QTDB)					
100	1129	1133	1128	5	1
103	1047	1748	862	886	185
114	817	964	749	215	68
116	1019	1247	978	269	41
117	765	765	764	1	1
123	751	756	751	5	0
213	1573	1631	1531	100	42
223	1042	1373	966	407	76
230	1064	1122	746	376	318
231	729	742	716	26	13
233	1144	1274	1103	171	41
301	1282	1269	1136	133	146
302	1497	1497	1493	4	4
306	1036	1099	1029	70	7
307	849	855	848	7	1
308	1202	1278	1175	103	27
803	953	1020	950	70	3
808	899	903	896	7	3
811	694	731	686	45	8
820	1073	1156	1070	86	3
821	1447	1236	1196	40	251
840	931	1296	832	464	99
847	784	840	776	64	8
853	988	1099	706	393	282
871	911	961	895	66	16
872	952	984	950	34	2
873	342	965	265	700	77
883	891	1037	234	803	657
891	1045	1296	806	490	239
16265	712	1118	709	409	3
16272	846	850	838	12	8
16273	818	1218	812	406	6
16420	1062	1062	1061	1	1
16483	1085	1085	1085	0	0
16539	916	926	915	11	1
16773	990	1016	984	32	6

Record	Annotated	Detected	True positive	False positive	False negative
16786	921	1023	919	104	2
16795	759	769	758	11	1
17453	1045	1046	1040	6	5
e0104	799	801	798	3	1
e0106	894	895	892	3	2
e0107	248	812	183	629	65
e0110	854	886	589	297	265
e0111	906	905	904	1	2
e0112	656	777	547	230	109
e0114	413	704	387	317	26
e0116	511	746	196	550	315
e0121	1430	1427	1421	6	9
e0122	1412	1412	1412	0	0
e0124	781	1121	779	342	2
e0126	925	1367	869	498	56
e0129	664	690	469	221	195
e0133	839	838	838	0	1
e0136	661	987	516	471	145
e0166	809	871	799	72	10
e0170	893	911	886	25	7
e0203	1229	1345	1189	156	40
e0210	1061	1091	1057	34	4
e0211	1551	1574	1550	24	1
e0303	1042	1075	1040	35	2
e0405	1203	1307	910	397	293
e0406	957	994	947	47	10
e0409	1648	792	767	25	881
e0411	1144	1240	1134	106	10
e0509	1001	1134	999	135	2
e0603	867	886	504	382	363
e0604	1030	1178	1010	168	20
e0606	1435	1445	1431	14	4
e0607	1181	1181	1180	1	1
e0609	1119	1126	1110	16	9
e0612	739	750	698	52	41
e0704	1092	1176	888	288	204
30	1008	1046	957	89	51
31	1038	1140	974	166	64
32	1192	1191	1097	94	95
33	523	545	522	23	1
34	582	914	543	371	39
38	189	18	0	18	189

Record	Annotated	Detected	True positive	False positive	False negative
39	1147	1169	1112	57	35
40	823	1024	350	674	473
41	920	1361	568	793	352
42	412	1243	120	1123	292
43	1175	1428	1046	382	129
44	1040	1308	981	327	59
45	739	981	590	391	149
46	612	890	182	708	430
47	883	926	869	57	14
48	1335	1400	1330	70	5
49	751	785	132	653	619
51	747	1009	728	281	19
52	1402	1437	1368	69	34
17152	1625	1626	1625	1	0
14046	885	1258	882	376	3
14157	910	1076	878	198	32
14172	659	675	657	18	2
15814	1006	1005	978	27	28
Overall	91607	102289	82746	19543	8861
T-wave detection using the MIT-BIH QT Database (QTDB)					
100	1981	1127	1113	14	868
103	1048	1048	1047	1	1
114	963	857	827	30	136
116	1188	1185	1176	9	12
117	765	765	765	0	0
123	771	756	749	7	22
213	1768	1633	1608	25	160
223	2098	1304	1284	20	814
230	1248	1077	867	210	381
231	883	732	715	17	168
233	1538	1325	1229	96	309
301	1664	1241	1238	3	426
302	1513	1498	1497	1	16
306	1045	1039	741	298	304
307	859	852	850	2	9
308	1306	1266	1260	6	46
803	1037	1025	960	65	77
808	941	902	183	719	758
811	700	703	700	3	0
820	1730	1142	972	170	758
821	1558	1553	1549	4	9
840	1306	1178	1175	3	131

Record	Annotated	Detected	True positive	False positive	False negative
847	814	802	776	26	38
853	1302	1111	854	257	448
871	1057	914	912	2	145
872	1026	988	958	30	68
873	865	859	857	2	8
883	1163	892	876	16	287
891	1629	1266	1231	35	398
16265	1032	1031	1030	1	2
16272	849	850	849	1	0
16273	1112	1110	1110	0	2
16420	1175	1063	1062	1	113
16483	1084	1086	1083	3	1
16539	931	922	919	3	12
16773	1064	1008	1005	3	59
16786	925	924	923	1	2
16795	759	760	759	1	0
17453	1143	1047	1045	2	98
e0104	1200	801	738	63	462
e0106	879	896	856	40	23
e0107	845	812	659	153	186
e0110	906	872	869	3	37
e0111	1137	906	888	18	249
e0112	685	684	514	170	171
e0114	926	698	517	181	409
e0116	840	558	542	16	298
e0121	1431	1427	1423	4	8
e0122	1414	1414	1412	2	2
e0124	1181	1120	1118	2	63
e0126	1772	944	944	0	828
e0129	756	675	618	57	138
e0133	873	839	822	17	51
e0136	855	809	805	4	50
e0166	923	811	806	5	117
e0170	900	897	895	2	5
e0203	2271	1246	1227	19	1044
e0210	1080	1063	1061	2	19
e0211	1593	1574	1573	1	20
e0303	1363	1044	1042	2	321
e0405	1420	1215	1213	2	207
e0406	975	959	957	2	18
e0409	1780	775	774	1	1006
e0411	1292	1201	1201	0	91

Record	Annotated	Detected	True positive	False positive	False negative
e0509	1025	1023	1019	4	6
e0603	1620	869	867	2	753
e0604	1035	1031	1029	2	6
e0606	1453	1441	1438	3	15
e0607	2122	1177	1169	8	953
e0609	1269	1124	1123	1	146
e0612	769	751	749	2	20
e0704	1786	1089	1089	0	697
30	1152	1011	1007	4	145
31	1113	1085	582	503	531
32	1228	1191	971	220	257
33	525	526	472	54	53
34	907	897	894	3	13
38	1559	527	525	2	1034
39	1981	1170	1153	17	828
40	1401	1069	1066	3	335
41	1285	1365	783	582	502
42	2025	1247	1159	88	866
43	587	1349	445	904	142
44	1476	1305	872	433	604
45	857	971	723	248	134
46	1259	838	546	292	713
47	1593	885	761	124	832
48	2373	1395	1329	66	1044
49	1411	826	690	136	721
51	415	749	227	522	188
52	2188	1407	1380	27	808
17152	1723	30	27	3	1696
14046	1262	1258	1238	20	24
14157	1340	1079	1057	22	283
14172	732	663	578	85	154
15814	1024	1029	999	30	25
Overall	117632	97458	90195	7263	27437

Appendix VIII. Boosted decision trees & AdaBoost

Decision trees

Decision trees consist of nodes, which are connected through branches. An example of a decision tree is shown in Figure 28. The tree starts with a root node at the top and ends with leaf nodes at the bottom. Between the leaf nodes and the root node can be internal nodes. All nodes are connected through branches, which represent outcomes of the decision rule as defined for the prior node.

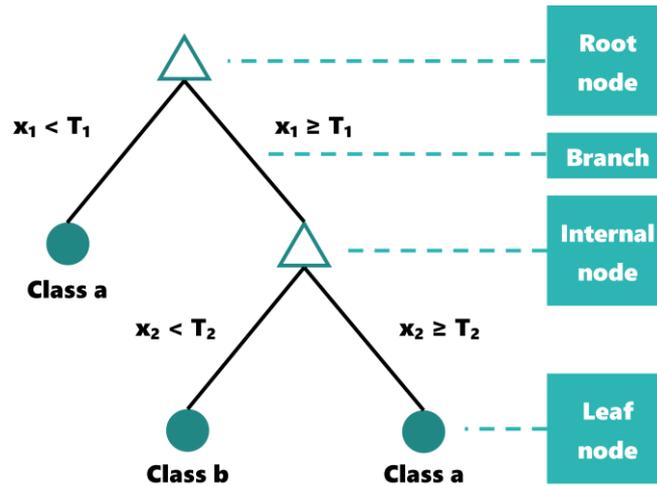


Figure 28 – Example of a decision tree with one root node, one internal node, three leaf nodes and four branches. x_1 and x_2 are the features, T_1 and T_2 are the thresholds for the decision rules and a and b are the output classes of the tree.

Decision trees can be automatically generated by finding the best splits for each node. A split describes the initial node, the two branches with decision rules and the two connected nodes at a lower level in the decision tree. Each split is based on one feature from the feature set. Splits are made based on the splitting criterion. By default, Matlab uses the Gini's diversity index as splitting criterion.⁵⁵ The Gini's diversity index (G) is calculated using the following equation⁵⁶:

$$G = 1 - \sum_{i=1}^n (p_i)^2 \quad (\text{Eq. A8.1})$$

where p_i is the probability of an object being classified to a particular class i and n is the total number of classes. This Gini's diversity index is calculated for both branches originating at a node and the weighted average of both Gini's diversity indices is used to determine the optimal split.

An example is visualized in Figure 29, where the split based on a feature x with a certain decision rule is evaluated: $x \leq v \rightarrow$ go left, $x > v \rightarrow$ go right. Fifteen objects are used as input: eight objects from the positive class and seven objects from the negative class. Seven objects from the positive class satisfy the decision rule for the left branch and three objects from the negative class satisfy the decision rule for the left branch. The other objects satisfy

the decision rule for the right branch. First the Gini's diversity index is calculated for the initial root node, next the left branch ($x \leq v$) is evaluated and lastly the right branch ($x > v$) is evaluated.

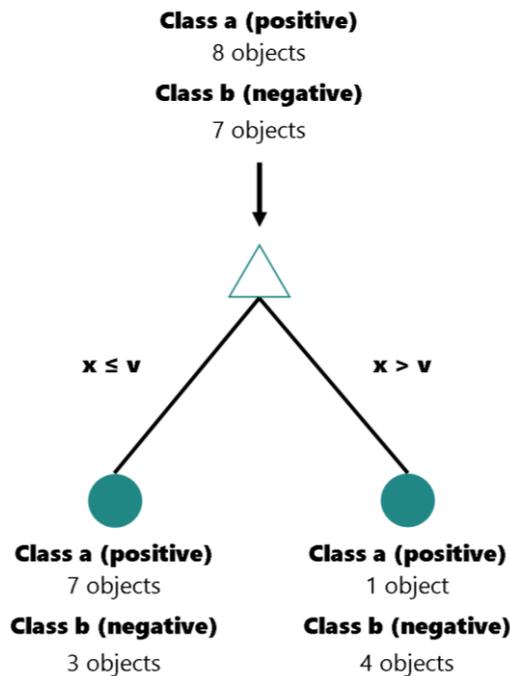


Figure 29 – Example of a decision tree with one root node and two leaf nodes. If the condition $x \leq v$ is satisfied, the left branch is taken, otherwise the right branch is taken. Fifteen objects are used as input; eight objects from the positive class and seven objects from the negative class. Seven objects from the positive class take the left branch and three objects from the negative class take the left branch. The other objects take the right branch.

The initial root node shows eight times the positive class and seven times the negative class. Therefore, the Gini's diversity index of the node ($G_{initial}$) is:

$$G_{initial} = 1 - \left(\left(\frac{7}{15} \right)^2 + \left(\frac{8}{15} \right)^2 \right) \approx 0.50 \quad (\text{Eq. A8.2})$$

The left branch shows seven times the positive class and three times the negative class. Therefore, the Gini's diversity index of this branch (G_{left}) is:

$$G_{left} = 1 - \left(\left(\frac{7}{10} \right)^2 + \left(\frac{3}{10} \right)^2 \right) = 0.42 \quad (\text{Eq. A8.3})$$

The right branch shows one time the positive class and four times the negative class. Therefore, the Gini's diversity index of this branch (G_{right}) is:

$$G_{right} = 1 - \left(\left(\frac{1}{5} \right)^2 + \left(\frac{4}{5} \right)^2 \right) = 0.32 \quad (\text{Eq. A8.4})$$

Combining these values, the Gini's diversity index of the decision made at the node based on feature x with this decision rule equals:

$$G_{new} = \frac{10}{15} \cdot G_{left} + \frac{5}{15} \cdot G_{right} \approx 0.39 \quad (\text{Eq. A8.5})$$

The decrease in the Gini's diversity index by using feature x with the decision rule is used as a measure for decrease in impurity. When generating a classification tree, this value is maximized. In this case, the decrease in Gini's diversity index is:

$$\Delta G = G_{initial} - G_{new} = 0.50 - 0.39 = 0.11 \quad (\text{Eq. A8.6})$$

Training a decision tree is done using the following steps⁵⁵:

1. Find the best split for each feature.

Values of the feature are sorted from smallest to largest. Each value is used as a split point. The split point is the "threshold" used for the node (e.g. when the split point is called v ; if $x \leq v$ take the left branch, else take the right branch). The splitting criterion is maximized for each feature.

2. Find the best split for the node.

The best split is defined as the feature and corresponding split point which maximize the splitting criterion.

3. Split the node using the split found in step 2.

This process is repeated until at least one of the following stopping rules has been reached, which are standardly used by Matlab⁵⁵:

- **The node is pure (i.e. only observations of one class are available for a node).**
- **The number of observations at the current node is below a set threshold.**
- **The number of observations in any of the generated nodes is below a set threshold.**
- **The number of nodes will exceed a threshold, which is set manually.**

Boosted decision trees

Using boosted decision trees, multiple decision trees are generated iteratively using different subsets of the dataset. By default, Matlab uses the AdaBoost algorithm to train these decision trees. In short, the AdaBoost algorithm generates a decision tree, assigns a weight to the decision tree based on the classification error, and generates a new subset of the data with misclassified data being more represented. The new subset of data is used to generate the next decision tree. This is repeated until a stopping criterion is satisfied (e.g. a maximum number of iterations).⁵⁷

Weight of the decision tree

The weight α_t for a decision tree h_t with a classification error rate $\varepsilon(h_t)$ is calculated as⁵⁷:

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon(h_t)}{\varepsilon(h_t)} \right) \quad (\text{Eq. A8.7})$$

The error rate $\varepsilon(h_t)$ of the decision tree h_t is calculated as the number of misclassifications divided by the number of total classifications, taking into account the weights of the different objects which are classified. For the first iteration, weights of all objects are equal.

This equation results in a high weight for decision trees for which $\varepsilon(h_t)$ is small. Oppositely, if $\varepsilon(h_t)$ tends to 1 (i.e. all classifications are wrong), the decision tree will be assigned a large negative weight.

New subset of data

When initiating the training of boosted decision trees, the initial weight for all objects ($W_1(i)$) is set to $1/n$, where n is the total number of objects. This indicates that for the first iteration, all objects contribute equally to the training of the decision tree. After training, the weights of the objects will be updated based on the object weights for the trained decision tree and the weight of the trained decision tree (α_t)⁵⁷:

$$W_{t+1}(i) = \begin{cases} W_t(i)e^{-\alpha_t} & \text{if object } i \text{ was correctly classified} \\ W_t(i)e^{\alpha_t} & \text{if object } i \text{ was incorrectly classified} \end{cases} \quad (\text{Eq. A8.8})$$

This results in an increased weight for objects which were incorrectly classified and a decreased weight for objects which were correctly classified. The weights are normalized to sum to one. The new subset of data for the next iteration is obtained. The weights are then used for training the next decision tree and to calculate the error rate.

Hyperparameters

Hyperparameters of a model are the number of decision trees to be generated, the number of splits per decision tree and the learning rate.⁴⁴ The learning rate is used as a scaling factor for changing the object weights.⁵⁸ With small learning rates, the difference $W_t(i)$ and $W_{t+1}(i)$ will be kept smaller, to avoid overfitting of the classifier to the dataset. However, the training process will take longer to converge to the optimal solution, since training is slowed.

Appendix IX. Bayesian optimization

For this thesis, a boosted decision trees model was used with the number of decision trees and the number of splits per decision tree as optimizable hyperparameters. Hyperparameters of classifiers can be optimized using multiple methods. The most straightforward method is trying all possible inputs within a certain range for the hyperparameters, training the classifier, and determining the error for each input. However, this is extremely time-intensive when training complex classifiers. Therefore, the default method in Matlab, Bayesian optimization, is a more efficient one.⁵⁹

The aim of Bayesian optimization is estimating the vector of inputs \mathbf{x} for which a function $f(\mathbf{x})$ reaches its optimum, without calculating $f(\mathbf{x})$ for all values of \mathbf{x} .^{60, 61} For the boosted decision tree model, $f(\mathbf{x})$ describes the quality of the classifier using certain inputs for hyperparameters (\mathbf{x}). The quality of the boosted decision trees model is described by the number of incorrect classifications, which is the error function. It is important to note that in case of the error function, *increased* quality indicates a *decrease* in the error function. Therefore, although the aim is to reach a *minimum* error, the function $f(\mathbf{x})$ should be *maximized*. Bayesian optimization is an algorithm that works with a surrogate model $\hat{f}(\mathbf{x})$ of the true function $f(\mathbf{x})$ and an acquisition function $\mu(\mathbf{x})$, where \mathbf{x} is a vector containing the hyperparameters. $\mu(\mathbf{x})$ describes the method for selecting points \mathbf{x} for which $f(\mathbf{x})$ should be evaluated.

Gaussian processes

The surrogate model $\hat{f}(\mathbf{x})$ has an uncertainty for each value of \mathbf{x} , which is represented by a Gaussian process. A Gaussian process describes the probability distribution function over all possible functions that fit a set of observations, as visualized in Figure 30.⁶⁰ In other words, the Gaussian process is an indicator for the certainty that the surrogate model $\hat{f}(\mathbf{x})$ equals the true function $f(\mathbf{x})$ at a point \mathbf{x} . The Gaussian process is described by a mean function containing the expected values of $\hat{f}(\mathbf{x})$, and the covariance matrix for observations of $f(\mathbf{x})$.

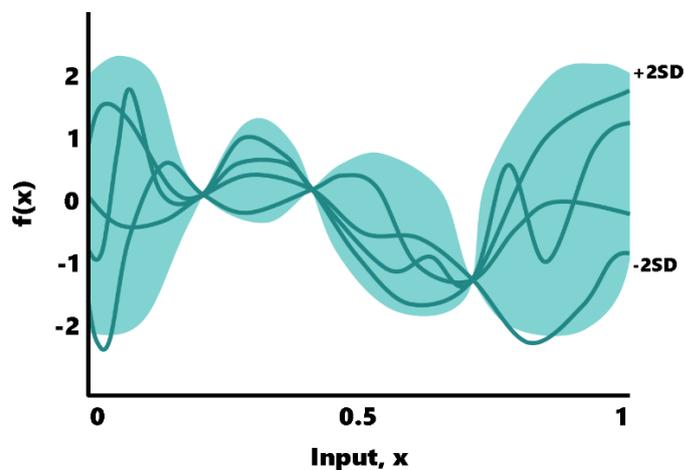


Figure 30 - Example of a Gaussian process. The lines show functions drawn from the set of functions described by the Gaussian process. The shaded regions show for each value of x twice the standard deviation (SD).

Covariance matrix

A covariance matrix for two variables \mathbf{z}_1 and \mathbf{z}_2 can be described as⁵⁴:

$$\Sigma(x_1, x_2) = \begin{bmatrix} \text{cov}(\mathbf{z}_1, \mathbf{z}_1) & \text{cov}(\mathbf{z}_2, \mathbf{z}_1) \\ \text{cov}(\mathbf{z}_1, \mathbf{z}_2) & \text{cov}(\mathbf{z}_2, \mathbf{z}_2) \end{bmatrix} = \begin{bmatrix} \text{var}(\mathbf{z}_1) & \text{cov}(\mathbf{z}_2, \mathbf{z}_1) \\ \text{cov}(\mathbf{z}_1, \mathbf{z}_2) & \text{var}(\mathbf{z}_2) \end{bmatrix} \quad (\text{Eq. A9.1})$$

where $\text{cov}(\mathbf{z}_1, \mathbf{z}_2)$ is the covariance between \mathbf{z}_1 and \mathbf{z}_2 , and $\text{var}(\mathbf{z}_1)$ is the variance of \mathbf{z}_1 .

The covariance between two variables \mathbf{z}_1 and \mathbf{z}_2 is defined as⁵⁴:

$$\text{cov}(\mathbf{z}_1, \mathbf{z}_2) = E[(\mathbf{z}_1 - E(\mathbf{z}_1))(\mathbf{z}_2 - E(\mathbf{z}_2))] \quad (\text{Eq. A9.2})$$

where E is the expected value (or mean). By definition, $\text{cov}(\mathbf{z}_1, \mathbf{z}_2) = \text{cov}(\mathbf{z}_2, \mathbf{z}_1)$.

The variance of a variable \mathbf{z}_1 is defined as⁵⁴:

$$\text{var}(\mathbf{z}_1) = E[(\mathbf{z}_1 - E(\mathbf{z}_1))^2] \quad (\text{Eq. A9.3})$$

Therefore, by definition $\text{cov}(\mathbf{z}_1, \mathbf{z}_1) = \text{var}(\mathbf{z}_1)$.

Radial Basis Function (RBF) kernel

The covariance matrix for the Gaussian process is generated using a Radial Basis Function (RBF) kernel $K(x_1, x_2)$ showing a Gaussian distribution^{62, 63}:

$$K(x_1, x_2) = \sigma^2 \cdot e^{-\frac{1}{2l^2}(x_1 - x_2)^2} \quad (\text{Eq. A9.4})$$

where $x_1 - x_2$ is the difference in hyperparameters between two observations x_1 and x_2 , and σ and l are new hyperparameters of the RBF kernel describing the vertical scaling and the horizontal scaling of the Gaussian distribution, respectively. Therefore, if $x_1 - x_2$ tends to 0, this will result in $K(x_1, x_2)$ being close to σ^2 . However, if $x_1 - x_2$ tends to ∞ , $K(x_1, x_2)$ will be close to 0. In practice, this implies that two points x_1 and x_2 with only minor differences in hyperparameters have a high covariance, whereas two points x_1 and x_2 with large differences in hyperparameters have a low covariance. Furthermore, a smaller l will result in steeper slopes in $K(x_1, x_2)$ and σ impacts the maximum value of $K(x_1, x_2)$.

Using the RBF, the covariance matrix can be calculated as^{62, 63}:

$$\Sigma(x_1, x_2) = K(x_1, x_2) + I\sigma_y^2 \quad (\text{Eq. A9.5})$$

where $I\sigma_y^2$ describes how well the estimated function $\hat{f}(x)$ fits through all observations of $f(x)$.

Predicting

Using this covariance matrix, predictions can be made. Predictions of $f(x)$ are made for a finite amount of queried points x . The predicted values are stored in a vector \mathbf{y}_1 . First, observations of $f(x)$ are done, which are stored in a vector \mathbf{y}_2 . Bayes' rule then states that if we know \mathbf{y}_2 , the probability of \mathbf{y}_1 given \mathbf{y}_2 ($p(\mathbf{y}_1|\mathbf{y}_2)$) is defined as^{62, 63}:

$$p(\mathbf{y}_1|\mathbf{y}_2) = \frac{p(\mathbf{y}_2|\mathbf{y}_1)p(\mathbf{y}_1)}{p(\mathbf{y}_2)} = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)} \quad (\text{Eq. A9.6})$$

where $p(\mathbf{y}_1, \mathbf{y}_2)$ is the prior probability of \mathbf{y}_1 and \mathbf{y}_2 , and $p(\mathbf{y}_n)$ the prior probability of \mathbf{y}_n .

Using this property together with the covariance matrix of \mathbf{y}_1 and \mathbf{y}_2 , the predictive mean can be calculated. The covariance matrix of \mathbf{y}_1 and \mathbf{y}_2 is defined as:

$$\begin{bmatrix} A & C \\ C^T & B \end{bmatrix} = \begin{bmatrix} \text{var}(\mathbf{y}_1) & \text{cov}(\mathbf{y}_2, \mathbf{y}_1) \\ \text{cov}(\mathbf{y}_1, \mathbf{y}_2) & \text{var}(\mathbf{y}_2) \end{bmatrix} \quad (\text{Eq. A9.7})$$

Then, as shown by Rasmussen et al., the predictive mean of \mathbf{y}_1 given \mathbf{y}_2 is calculated as⁶³:

$$\mu_{\mathbf{y}_1|\mathbf{y}_2} = \mu_{\mathbf{y}_1} + CB^{-1}(\mathbf{y}_2 - \mu_{\mathbf{y}_2}) \quad (\text{Eq. A9.8})$$

where $\mu_{\mathbf{y}_1}$ is the mean of \mathbf{y}_1 and $\mu_{\mathbf{y}_2}$ is the mean of \mathbf{y}_2 .

Next, the predictive covariance of \mathbf{y}_1 given \mathbf{y}_2 is calculated as⁶³:

$$\Sigma_{\mathbf{y}_1|\mathbf{y}_2} = A - CB^{-1}C^T \quad (\text{Eq. A9.9})$$

where A indicates the uncertainty without having done any observations (since \mathbf{y}_2 is not involved in A), and $CB^{-1}C^T$ describes a reduction in uncertainty as a result of the observations in \mathbf{y}_2 . Using this method, predictions of $f(\mathbf{x})$ can be done and the uncertainty of the predictions can be calculated.

Acquisition function

The surrogate model is optimized by calculating the true values of $f(\mathbf{x})$ at certain values for \mathbf{x} , as visualized in Figure 31.^{62, 63} The values of \mathbf{x} for which $f(\mathbf{x})$ should be evaluated are determined by the acquisition function. An acquisition function can be set to explore or to exploit. An acquisition function focused on exploration tries to decrease uncertainty in regions with a high uncertainty, whereas an acquisition function focused on exploitation tries to decrease uncertainty in regions with optimal values of $\hat{f}(\mathbf{x})$.⁶⁴ In other words, exploration is used to find an optimum in a new region, whereas exploitation is used to find the precise location of the optimum. In practice, acquisition functions are designed to both explore and exploit. Again, it is important to note that in case of the error function, increased quality $f(\mathbf{x})$ indicates a *decrease* in the error function.

By default, Matlab uses an acquisition function based on the expected improvement (EI)⁵⁹:

$$EI_n(\mathbf{x}) = E_n \left[[\hat{f}(\mathbf{x}) - f_n^*]^+ \right] \quad (\text{Eq. A9.10})$$

where n is the iteration, E_n is the expected value (or mean) for the n^{th} iteration, $\hat{f}(\mathbf{x})$ is the estimated quality of the surrogate model with input \mathbf{x} , and f_n^* is the maximum quality of $\hat{f}(\mathbf{x})$ for all *evaluated* values of \mathbf{x} . Therefore, $\hat{f}(\mathbf{x}) - f_n^*$ is the expected difference in quality between a point \mathbf{x} and the previous best. The $+$ indicates that only the positive part is considered. In other words, $a^+ = \max(a, 0)$.

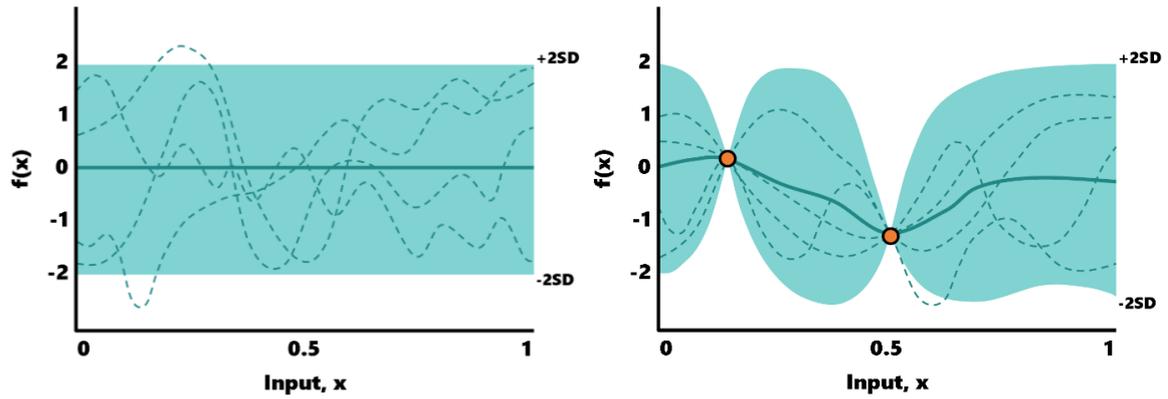


Figure 31 – Example of a Gaussian process with observed values of $f(x)$ at point x . The left panel shows functions from the prior distribution. The right panel shows the functions of the posterior distribution fitting the two observed values of $f(x)$ for two points x . The solid line (–) is the mean prediction. The dashed lines show functions drawn from the set of functions described by the Gaussian process. The orange points indicate the evaluated points of $f(x)$. The shaded regions show for each value of x twice the standard deviation (SD). Adapted version of figure by Rasmussen et al.⁶³

The acquisition function further considers the probability of $\hat{f}(x)$ resulting in a better quality than f_n^* . As described by Frazier, this results in the following acquisition function $\alpha(x)$ ⁶⁰:

$$\alpha_n(x) = E_n[\Delta_n(x)^+] + \sigma_n(x)\varphi\left(\frac{\Delta_n(x)}{\sigma_n(x)}\right) - |\Delta_n(x)|\Phi\left(\frac{\Delta_n(x)}{\sigma_n(x)}\right) \quad (\text{Eq. A9.11})$$

where n is the iteration, E_n is the expected value (or mean) for the n^{th} iteration, $\Delta_n(x) = \hat{f}(x) - f_n^*$, $\sigma_n(x)$ is the standard deviation of the probability distribution function for the n^{th} iteration, φ is the standard normal density function, and Φ is the standard normal distribution function.

Using the acquisition function, a point x is determined, for which the quality of the model is most likely to improve maximally. Next, the real function $f(x)$ is evaluated for that point x and the predicted means and covariances of the Gaussian process are updated. Using this updated Gaussian process, the acquisition function is determined and so we come full circle again.⁶⁰ This process is repeated until either a maximum number of iterations, a maximum amount of time, or a manual stopping criterion has been reached.

Summary of Bayesian optimization

In short, Bayesian optimization can be used to estimate the optimal hyperparameters for a model without evaluating the model for all possible hyperparameters. Bayesian optimization fits a Gaussian process over the data, describing the expected values and uncertainties for each queried x . Using these expected values and uncertainties, a point x is selected for which the acquisition function is optimal. For this point x , the real value of $f(x)$ is determined. Next, the Gaussian process is updated, as this new value results in less uncertainty. Finally, the new expected values and uncertainties are used to update the acquisition function and the next iteration starts.

Appendix X. Relative estimated feature importance

Table 9 – Estimated feature importance for all features relative to the maximum estimated feature importance (excluding noise), which is the IQR of the ratio between the RR-interval preceding an R-peak and the RR-interval succeeding an R-peak.

Features	Relative estimated feature importance
RR-interval characteristics	
1. Number of beats per minute (bpm)	0.3020
2. Median interval between R_i and R_{i+1} (seconds)	0.1139
3. IQR of interval between R_i and R_{i+1} (seconds)	0.1040
4. Median interval between R_{i-1} and R_{i+1} (seconds)	0.1289
5. IQR interval between R_{i-1} and R_{i+1} (seconds)	0.1126
6. Median ratio between the RR-interval preceding an R-peak and the RR-interval succeeding an R-peak	0.1107
7. IQR of ratio between the RR-interval preceding an R-peak and the RR-interval succeeding an R-peak	1 (maximum)
8. Poincaré plot: standard deviation of distances of points to the -45°-line	0.1586
9. Poincaré plot: standard deviation of distances of points to the 45°-line	0.0741
10. Shannon entropy of the RR-intervals	0.1036
Peak-interval characteristics	
1. Median interval between P-peak and Q-peak (seconds)	0.1351
2. IQR of interval between P-peak and Q-peak (seconds)	0.9044
3. Median interval between Q-peak and S-peak (seconds)	0.0374
4. IQR of interval between Q-peak and S-peak (seconds)	0.2348
5. Median interval between Q-peak and T-peak (seconds)	0.0201
6. IQR of interval between Q-peak and T-peak (seconds)	0.0248
7. Median corrected interval between Q-peak and T-peak (seconds)	0.0656
8. IQR of corrected interval between Q-peak and T-peak (seconds)	0.0232
9. Median interval between S-peak and T-peak (seconds)	0.0376
10. IQR of interval between S-peak and T-peak (seconds)	0.0754
Amplitude characteristics	
1. Median amplitude from Q-peak to R-peak	0.0396
2. IQR of amplitude from Q-peak to R-peak	0.0800
3. Median amplitude from R-peak to S-peak	0.2809
4. IQR of amplitude from R-peak to S-peak	0.0684
5. Median difference in amplitude of R_i and R_{i+1}	0.0441
6. IQR of difference in amplitude of R_i and R_{i+1}	0.0731
7. Median difference in amplitude of R_{i-1} and R_{i+1}	0.0354
8. IQR of difference in amplitude of R_{i-1} and R_{i+1}	0.0525
9. Median ratio between the amplitude differences of $R_{i-1}R_i$ and R_iR_{i+1}	0.0349
10. IQR of ratio between the amplitude differences of $R_{i-1}R_i$ and R_iR_{i+1}	0.0152

Features	Relative estimated feature importance
11. Poincaré plot of differences in RR-amplitude: standard deviation of distances of points to the -45°-line	0.0715
12. Poincaré plot of differences in RR-amplitude: standard deviation of distances of points to the 45°-line	0.0441
P-wave characteristics	
1. Percentage of R-peaks without P-waves	0.0524
2. Percentage of R-peaks with one P-wave	0.4394
3. Percentage of R-peaks with multiple P-waves	0.0444
4. Median amplitude of P-wave	0.1082
5. IQR of amplitude of P-wave	0.0339
6. 85 th percentile of ratio between amplitudes of P-wave and QRS-complex	0.0366
7. Median ratio between amplitudes of P-wave and QRS-complex	0.0426
8. IQR of ratio between amplitude of P-wave and QRS-complex	0.0449
T-wave characteristics	
1. Percentage of R-peaks without T-wave	0.0020
2. Percentage of R-peaks with one T-wave	0.0022
3. Median amplitude of T-wave	0.0800
4. IQR of amplitude of T-wave	0.0617
5. Median ratio between amplitudes of T-wave and QRS-complex	0.0400
6. IQR of ratio between amplitudes of T-wave and QRS-complex	0.0563
QRS-morphology characteristics	
1. Number of different QRS-morphologies	0.0501
2. Percentage of QRS-complexes with the most common morphology	0.0956
3. Percentage of QRS-complexes with the second most common morphology	0.1084
Autocorrelation characteristics	
1. Median autocorrelation of ECG segment	0.0628
2. IQR of autocorrelation of ECG segment	0.0435
3. Ratio between number of peaks in the autocorrelation and number of QRS-complexes	0.1549
4. Median amplitude of peaks in the autocorrelation	0.1774
5. IQR of amplitude of peaks in the autocorrelation	0.0979
6. Median time interval between peaks in the autocorrelation (seconds)	0.0648
7. IQR of time interval between peaks in the autocorrelation (seconds)	0.0601
Noise (not included in Figure 20)	
Percentage of excluded samples for R-peak detection	3.6815

