Evaluating Smartwatch Detection of Atrial Fibrillation

Master's thesis in Engineering Mathematics and Computational Science

ALBIN EDEGRAN

DEPARTMENT OF MATHEMATICAL SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Master's thesis 2025

Evaluating Smartwatch Detection of Atrial Fibrillation

ALBIN EDEGRAN

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
Chalmers University of Technology
Gothenburg, Sweden 2025

Evaluating Smartwatch Detection of Atrial Fibrillation
ALBIN EDEGRAN

Supervisor: Linda Johnson, Lund University
Examiner: Larisa Beilina, Department of Mathematical Sciences, Chalmers and University of Gothenburg

Master's Thesis 2025
Department of Mathematical Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2025

Evaluating Smartwatch Detection of Atrial Fibrillation
Albin Edegran
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

Smartwatches, such as the Apple Watch, are increasingly used for monitoring and detecting serious health conditions like atrial fibrillation (AF), a common arrhythmia with significant health risks. However, to avoid burdening the healthcare sector with false alarms, their detection algorithms are designed to minimize false positives, a choice that may compromise their sensitivity to sparse or intermittent AF. While previous studies have analyzed the specificity of these alerts, the rate of false negatives remains underreported. This thesis aims to quantify this performance gap by evaluating the Apple Watch's photoplethysmography-based irregular-pulse notification and its rate of false negatives. Using a large clinical ambulatory ECG dataset, we modeled and simulated long-term AF episode patterns via two distinct stochastic methods: a two-state discrete-time Markov chain and a continuous-time alternating bivariate Hawkes process. Analysis of the simulated data revealed that a significant proportion of individuals, particularly those with low-to-moderate AF burden (0.5%–5.5%), risk going undetected for extended periods. For example, for a burden of 0.5%–1.5%, over 20% of cases remained undetected after five years, and for a burden of 3.5%–4.5%, over 5% remained undetected after five years of monitoring. The findings, consistent across both modeling approaches, demonstrate that detection is highly dependent on the temporal pattern of AF episodes and not just the AF burden. They also suggest that the Apple Watch algorithm has reduced sensitivity to certain AF episode patterns, and users should be cautious not to interpret the absence of an alert as confirmation of a normal rhythm.

Keywords: Atrial fibrillation, Smartwatch, Markov chain, Hawkes process

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Linda Johnson, whose expert guidance and support were instrumental in making this project possible. I am also deeply grateful to the team at MEDICALgorithmics for the opportunity to conduct my master's thesis with them and for providing invaluable data. Special thanks go to my friends Bekir Fazlija and Benjamin Elm Jonsson for their insightful feedback and helpful suggestions. Finally, my heartfelt thanks to my friends and family for their unwavering support throughout this project.

Albin Edegran, Gothenburg, June 2025

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AF	Atrial Fibrillation
AI	Artificial Intelligence
AV	Atrioventricular
CDF	Cumulative Distribution function
DRAI	DeepRhythmAI
ECG	Electrocardiogram
EDF	Empirical Distribution Function
KS	Kolmogorov-Smirnov
MLE	Maximum Likelihood Estimation
PPG	Photoplethysmography
SA	Sinoatrial
SEM	Standard Error of the Mean
SR	Sinus Rhythm

Contents

List of Figures

List of Tables

1

Introduction

Smartwatches equipped with photoplethysmography (PPG) have become increasingly common over-the-counter devices for tracking heart rate, blood oxygen levels, and other health-related metrics. Many also include the ability to screen for irregular rhythms associated with atrial fibrillation (AF), the most common heart rhythm disorder [1]. The stakes for detecting AF are high. Affecting over nine million Europeans as of 2010 [2], AF is associated with a five-fold increase in the risk of stroke, and substantially raises the risk of heart failure, dementia, and death [3]. Early detection is therefore critical. However, designing an AF alert system for a consumer device presents a paradoxical challenge. False alerts cause unnecessary stress for users and create a need for downstream monitoring with electrocardiogram (ECG) devices, which increases the costs and risks overburdening the healthcare system. Given that these devices are worn by millions of individuals, most of whom are healthy, this risk makes minimizing false positives a primary design goal, rather than maximizing detection. Consequently, many smartwatches employ algorithms that are designed for a high specificity, often by requiring multiple, separate detections of an irregular pulse over time before triggering an alert. While this approach may be effective at reducing the number of false positives, its impact on false negatives, the number of AF cases that go undetected, is significantly underreported.

The Apple Watch, a market leader as of 2024 [4], is one example of a smartwatch using an algorithm of this design. Its algorithm analyzes a series of one-minute PPG measurements, each spaced 15 minutes apart, to detect irregular rhythm that could potentially be caused by AF. Only if five out of six consecutive readings are classified as irregular rhythm does the user receive a notification. Many other arrhythmias can cause short (seconds to minutes) episodes of irregular rhythm, but AF is the most common sustained arrhythmia, and thus this system, while minimizing the false positive rate, is also efficient in terms of how much processor power the rhythm monitoring requires. The design has been validated in a large prospective study [5], which showed that the irregular rhythm notifications were highly specific. However, because this study only investigated individuals who received an alert, the number of false negatives remains unknown. This knowledge gap, coupled with an algorithm requiring multiple spaced-out segments to be irregular, raises concerns about its effectiveness for individuals with sparse or shorter episodes of AF. This potential limitation is significant, since it has been shown that patients with short AF episodes can run the same risk of stroke as those with longer episodes [6]. Furthermore,

individuals who wear the Apple Watch may develop a false sense of security, thinking it would alert them if they had an arrhythmia, and thus be less likely to seek appropriate care in time.

To the best of our knowledge, the rate of these false negatives for the Apple Watch's irregular rhythm notification has not been systematically studied. This thesis aims to bridge this knowledge gap by estimating the sensitivity of the Apple Watch's PPG-based AF detection. To achieve this, we simulate the performance of the Apple Watch algorithm on a broad range of AF episode patterns, derived from a large clinical dataset of ambulatory ECGs. Since multi-year-long AF episode patterns are unavailable, we employ two distinct stochastic models, a two-state Markov chain and an alternating bivariate Hawkes process, to simulate realistic, multi-year AF patterns.

1.1 Related works

There are several studies that are relevant to this work. In this section, two of the most important ones will be highlighted. The first is Apple's study on the effectiveness of the Apple Watch's alerts, and the second is a study focusing on how AF episode patterns can best be modeled.

The Apple Heart Study was a large-scale prospective study conducted by Apple in collaboration with Stanford Medicine aiming to evaluate the effectiveness of the Apple Watch in detecting AF using the PPG sensor [5]. Over 400,000 participants were recruited over a period of eight months, with a median monitoring duration of 117 days per participant. The participants self-reported that they had no prior knowledge of AF. In total, 2,161 participants received an irregular pulse notification. Out of these 2,161 users, 450 received follow-up ECG monitoring using a wearable ECG patch, a small device that records heart rhythm. The ECG patch also detected AF in 34% of the cases, and the positive predictive value for observing AF at the same time as an irregular pulse notification was 0.84 [5]. A later sub-study on 226 participants who all wore the ECG patch for approximately one week showed that out of 57 participants who got an alert from the smartwatches, 45 showed concordant AF on the ECG patch [7].

Another relevant line of research concerns the modeling of temporal AF episode patterns. In [8], a point process, the alternating bivariate Hawkes process was proposed for modeling AF episode patterns. The process is an extension of the classic Hawkes process, which is a type of self-exciting point process that has found success in multiple fields, such as finance and seismology [9]. In [8] it is shown that the alternating bivariate Hawkes process can be used to model many types of AF episode patterns. On three datasets comprising 59 individuals in total, this model provided a good fit for 40 patients, a better performance than earlier models. For example, in [10] the Weibull model was shown to achieve a good fit for only five out of nine patients.

2

Background

This chapter contains the background necessary to understand the basics of what AF is and how the Apple Watch tries to detect it. It also explains how the Apple Watch algorithm works. Understanding these aspects is essential to interpret the data and methods used in the thesis. The chapter first describes the physiological mechanisms of the heart and atrial fibrillation, followed by an overview of the Apple Watch's AF detection features and algorithms.

2.1 The heart and atrial fibrillation

During sinus rhythm (SR), the normal rhythm of the heart, the atria contract first, pushing blood into the ventricles. The ventricles then contract, propelling blood to the lungs and the rest of the body. These contractions are triggered by an electrical impulse that originates in the sinoatrial (SA) node, which is a specialized cluster of cells located in the right atrium that acts as the heart's natural pacemaker [11]. From the SA node, the impulse spreads through both atria, before reaching the atrioventricular (AV) node. The atria and the ventricles are electrically isolated from each other, and the AV node is the only point where a signal can pass from the atria to the ventricles. In the AV node there is a slight delay in conduction, which allows the ventricles to be filled by the contraction of the atria. Once the impulse is propagated through the AV node, it causes the ventricles to contract, which pushes blood to the aorta, and from there to the rest of the body. Following each impulse, the cells in the AV node and the ventricles enter a refractory period, where they are unresponsive to new electrical signals [11]. A diagram of the conduction system of the heart is shown in Figure 2.1.

During AF, the orderly electrical impulses inside the atria are replaced with rapid, disorganized, and chaotic electrical activity. As a result, the atria start to twitch or 'fibrillate' rather than maintain a regular rhythm. While the refractory period of the AV node prevents the ventricles from firing at the same chaotic rate, the impulses that do pass through are irregularly timed and lead to an irregular heart rhythm [13].

The electrical activity of the heart is often recorded using an ECG to help doctors understand what is happening inside the heart. The ECG works by placing elec-

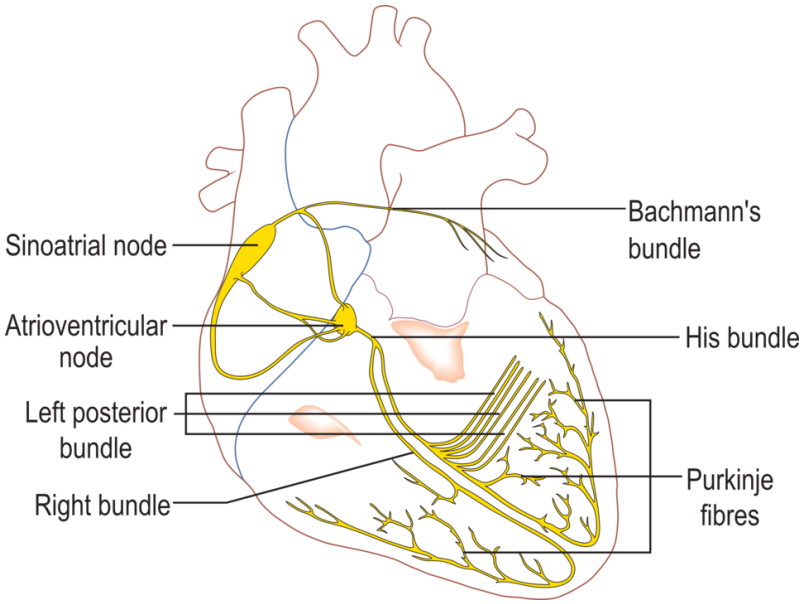


Figure 2.1: Diagram of the heart's conduction system [12].

trodes on specific parts of the chest, arms, and legs to measure the difference in electrical potential between these points. The voltage differences are recorded over time and graphed. The appearance of the signal varies depending on the leads used to record the electrical activity. An example of ECG readings during SR and AF is shown in Figure 2.2.

Figure 2.2: A short ECG strip showing AF in the top panel and SR in the bottom panel [14]. The arrow in the bottom panel points to the p-wave, which represents atrial depolarization. During AF, the p-wave is absent and replaced by erratic fibrillatory (f) waves.

2.1.1 Prevalence and dangers

Atrial fibrillation is one of the most common arrhythmias overall and is the one for which patients are most frequently treated and hospitalized [3]. In 2010, its estimated prevalence ranged from 3–6 million cases in the United States, to over nine million in Europe [2]. The number of cases is also predicted to rise in part due to the increasing life expectancy. The incidence risk increases with age, and the arrhythmia is most common among the elderly [3]. In [15], the median age at

diagnosis was reported to be 72 years based on data from the Danish nationwide registry between 1995 and 2021. Other risk factors include gender, with men being at larger risk, as well as weight, family history of AF, and sleep apnea [3].

AF is coupled with many dangers, the most prominent being a five-fold increase in the risk of stroke, and twice the risk of all-cause mortality and cognitive dysfunction [3]. Other dangers are heart failure and sudden death. The severity and amount of symptoms caused by AF can range from very little to very severe. AF episodes lasting seven days or longer are typically classified as persistent, while those lasting less than seven days are considered paroxysmal. Furthermore, the AF burden is defined as the percentage of time spent in AF. The clinical definition of AF also requires an episode to be longer than 30 seconds to be classified as AF. This limit is, however, somewhat arbitrary, and it has been shown that having AF-like supraventricular beats that lack a P-wave is a strong indicator for later AF onset [6].

AF can be everything from asymptomatic to functionally disabling. Common symptoms are palpitations, meaning the sensation of a fast or fluttering heart, fatigue, lightheadedness, and effort intolerance [3]. There is also a significant risk that since many of those with AF experience the onset at a late age, they may assume that the symptoms are simply due to their advanced age and avoid seeking medical attention. Even patients who experience only mild to no symptoms can get severe complications such as stroke, and stroke can be the first manifestation of AF [3].

The risks associated with AF can be reduced with the proper treatment. Anticoagulants can be used to reduce the risk of stroke by up to 70%, and rate or rhythm control drugs can help slow the heart rate and restore sinus rhythm, reducing symptoms. This makes early detection important [3].

2.2 The Apple Watch algorithm

The Apple Watch is equipped with two methods for detecting AF: PPG-based detection and ECG-based detection. PPG is a technology that uses light to monitor different properties of the heart and blood, such as the heart rate, blood oxygen saturation, and more [16]. It works by emitting light into the tissue and measuring the amount that is reflected using a photodetector. Blood absorbs more light than surrounding tissue, which leads to a higher absorption when the heart is in its systolic phase, where blood is pumped out from the heart. The amount of light which is absorbed over time is recorded, and is called a tachogram. The Apple Watch can use PPG tachograms to look for heart arrhythmias, such as AF. The feature is opt-in and uses an algorithm that analyzes a series of one-minute tachograms to determine if it should issue an alert, warning the user of potential AF [7]. The algorithm used for this detection is outlined below.

Algorithm 1 Apple Watch Algorithm

```
1: while true do
2:     Wait until user is not in motion
3:     Analyze a new one-minute tachogram
4:     if tachogram indicates AF then
5:         Initialize AF counter ← 1
6:         Initialize SR counter ← 0
7:         Record start time ← current time
8:         while true do
9:             Wait 15 minutes, then wait until user is not in motion
10:            if current time − start time > 48 hours then
11:                Break from inner loop
12:            end if
13:            Acquire and analyze a new one-minute tachogram
14:            if tachogram indicates AF then
15:                Increment AF counter
16:            else
17:                Increment SR counter
18:            end if
19:            if AF counter = 5 then
20:                Trigger irregular pulse alert
21:                Stop algorithm
22:            else if SR counter = 2 then
23:                Break from inner loop
24:            end if
25:        end while
26:    end if
27:    Wait two hours
28: end while
```

The algorithm that Apple uses to check each one-minute tachogram for irregular rhythm is proprietary [7]. Furthermore, since PPG readings are sensitive to motion they are not always able to capture an adequate signal. The Apple Watch deals with this by delaying the tachograms if the wearer is in motion using the watch's built-in accelerometer.

The reliability of the Apple Watch's AF detection has been supported by findings from the Apple Heart Study and its follow-up sub-study. These studies showed that the positive predictive value of receiving an irregular rhythm notification and actually having AF is high [5], [7]. However, since only patients who received an alert from their smartwatch were selected for further monitoring with the patches, these studies lack negative cases and therefore cannot be used to determine the sensitivity or specificity of the watch's AF detection [3].

The ECG-based detection, available on Series 4 and later, uses an electrode in the crown of the Apple Watch along with one on the underside that touches the wrist. By using Apple's ECG app and placing a finger from the opposite hand on the crown, a 30 seconds long, two-lead ECG recording can be taken. A proprietary algorithm then classifies the ECG trace as SR, AF, or inconclusive [7]. Since the ECG-based detection requires the user to manually initiate a test and cannot monitor continuously in the background, it is unlikely to be useful for detecting AF unless the user is experiencing symptoms. This is why it is not the focus of this study.

3

Theory

This chapter presents the theoretical background necessary to understand the mathematics behind the two models used to simulate AF episode patterns in this thesis. Specifically, it outlines the theory of the two stochastic processes employed: discrete-time Markov chains and the Hawkes process. The section on Markov chains draws primarily from [17], while the theory on Hawkes processes is based on [9].

3.1 Markov chains

Markov chains are stochastic processes with the property that the future evolution of the chain is conditionally independent of the past, given the present state. This property is known as the Markov property. Many real-world systems can be effectively modeled using Markov chains [18], and assuming the Markov property often leads to more tractable calculations [17]. The property is formalized in the following definition.

Definition 3.1.1 (Markov Chain)**.** *A Markov chain is a sequence of random variables $X_0, X_1, X_2, \ldots$ taking values in a discrete set $\mathcal{S}$, called the state space with the property that*

$$\begin{aligned} &P(X_{n+1} = j \mid X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = i) \\ &\quad = P(X_{n+1} = j \mid X_n = i). \end{aligned} \tag{3.1}$$

for all $x_0, \ldots, x_{n-1}, i, j \in \mathcal{S}$, and $n \geq 0$.

This thesis considers only discrete-time Markov chains with finite state spaces. Owing to the Markov property, all transition probabilities depend solely on the current state i and the next state j. These probabilities can be collected into a matrix $\mathbf{P}$, known as the transition matrix of the chain.

Definition 3.1.2 (Transition matrix)**.** *Let $(X_n)_{n\geq 0}$ be a Markov chain with finite state space $\mathcal{S} = \{1, 2, \ldots, N\}$. The transition matrix $\mathbf{P} = (P_{ij}) \in \mathbb{R}^{N\times N}$ is defined by*

$$P_{ij} = \Pr(X_{n+1} = j \mid X_n = i), \quad \textit{for all} \quad i, j \in \mathcal{S}.$$

Thus the entry P_{ij} gives the probability of transitioning from state i to state j in one time step. To compute the probability of transitioning between states over multiple steps, one considers powers of the transition matrix: the element $(\mathbf{P}^n)_{ij}$ corresponds to the probability of moving from i to j in n steps [17].

The structure of the transition matrix $\mathbf{P}$ determines important properties of the Markov chain. For example, if a set of states are all reachable from one another, they form a communicating class. Additionally, a Markov chain is said to be irreducible if all states belong to the same communicating class.

3.1.1 Long-term behavior

The long-term behavior of a Markov chain is determined by the properties of its states, in particular whether they are recurrent or transient. If a state, when left, is bound to eventually be returned to, that state is said to be recurrent. If the expected time to return is also finite then the state is positive recurrent. For Markov chains with finite state spaces, as are the ones in this thesis, any recurrent state is also positively recurrent. A key characteristic of Markov chains where all states are positive recurrent is the existence of a unique stationary distribution.

Theorem 3.1.1 (Existence and Uniqueness of the Stationary Distribution)**.** *Let* $\mathbf{P}$ *be the transition matrix of a finite, irreducible, and positive recurrent Markov chain. Then there exists a unique row vector* $\boldsymbol{\pi} \in \mathbb{R}^N$ *such that*

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}, \quad \textit{and} \quad \sum_{i=1}^{N} \pi_i = 1. \tag{3.2}$$

The vector $\boldsymbol{\pi}$ *is called the stationary distribution and describes the long-term proportion of time the chain spends in each state.*

Not all Markov chains have all positive recurrent states. If a state instead, as time goes to infinity, is visited only a finite number of times then that state is said to be transient. Conversely, a state is absorbing if, once entered, it cannot be left. Markov chains containing at least one absorbing state are called absorbing chains. A simple example of an absorbing chain with two transient states, a and b, and an absorbing state c is shown in Figure 3.1, and the corresponding transition matrix $\mathbf{P}^*$ is given by (3.3).

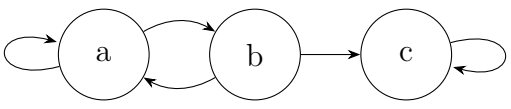


Figure 3.1: A three-state Markov chain with two transient states a and b, and one absorbing state c.

$$\mathbf{P}^* = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.3}$$

For absorbing chains it is interesting to look at the expected number of steps until the chain enters an absorbing state. This can be done using the fundamental matrix, as shown in the following theorem.

Theorem 3.1.2 (Fundamental Matrix)**.** *Consider an absorbing Markov chain with n transient states and m absorbing states. Let its transition matrix $\mathbf{P}$ be written in the canonical form*

$$\mathbf{P} = \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix},$$

where

- $\mathbf{Q}$ *is the $n \times n$ submatrix of transition probabilities between the transient states,*
- $\mathbf{R}$ *is the $n \times m$ submatrix of transition probabilities from transient states to absorbing states,*
- $\mathbf{0}$ *is a $m \times n$ zero matrix, and*
- $\mathbf{I}_m$ *is the $m \times m$ identity matrix.*

Let $\mathbf{F}$ be a $n \times n$ matrix indexed by the transient states, where F_{ij} is the expected number of times the chain is in state j before being absorbed, given that it started in state i. Then $\mathbf{F}$, called the fundamental matrix, is given by

$$\mathbf{F} = (\mathbf{I}_n - \mathbf{Q})^{-1}.$$

The result is proved as follows. Let $\mathcal{T}$ denote the set of transient states, F_{ij} the expected number of visits to transient state j given that the chain starts in transient state i, and define δ_{ij} to be the Kronecker delta, i.e., $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. Assume $i, j \in \mathcal{T}$. By conditioning on the first step, we obtain the recurrence relation

$$F_{ij} = \delta_{ij} + \sum_{k \in \mathcal{T}} Q_{ik} F_{kj}.$$

In other words, the expected number of visits to a state j from i is the same as the expected number of visits from the next state, k, times the probability of moving from i to k, plus one if the chain started in j. In matrix form, this becomes

$$\mathbf{F} = \mathbf{I}_n + \mathbf{Q}\mathbf{F},$$

rearranging the terms gives

$$(\mathbf{I}_n - \mathbf{Q})\mathbf{F} = \mathbf{I}_n,$$

and since absorbing chains eventually leave the transient states with probability 1, then $\mathbf{Q}^k \to 0$ as $k \to \infty$. Per the definition of a Neumann series the matrix $(\mathbf{I}_n - \mathbf{Q})$ is guaranteed to be invertible [17]. Solving this linear system yields a final expression for the fundamental matrix

$$\mathbf{F} = (\mathbf{I}_n - \mathbf{Q})^{-1}.$$

The expected number of steps until the chain reaches an absorbing state, for a chain started in state i, is given by the row sum of the fundamental matrix

$$a_i = \sum_{j \in \mathcal{T}} F_{ij},$$

where a_i is the expected number of steps to absorption for a chain started in i.

3.1.2 Maximum likelihood estimation

One way to estimate the elements of the transition matrix from a sequence of observed transitions is by maximum likelihood estimation (MLE). The following theorem defines the maximum likelihood estimator for a Markov chain.

Theorem 3.1.3 (Maximum Likelihood Estimator for Markov Chains)**.** *Let $\mathcal{S} = \{1, 2, 3, \ldots, N\}$ be a finite state space, and let $X = \{x_0, x_1, x_2, \ldots, x_n\}$ be an observed sequence of states generated by a Markov chain with transition matrix $\mathbf{P}$. Let N_{ij} denote the number of observed transitions out of state i to state j, given by*

$$N_{ij} = \sum_{t=0}^{n-1} \mathbf{1}\{x_t = i,\, x_{t+1} = j\}.$$

Then the maximum likelihood estimator of the transition probability P_{ij} is

$$\hat{P}_{ij} = \frac{N_{ij}}{\sum_{j \in \mathcal{S}} N_{ij}}, \tag{3.4}$$

i.e., the proportion of observed transitions from state i that went to state j.

The proof proceeds as follows. The likelihood of observing the sequence of states X, given a transition matrix $\mathbf{P}$, is

$$L(X \mid \mathbf{P}) = \prod_{i,j \in \mathcal{S}} P_{ij}^{N_{ij}}.$$

Taking the logarithm yields the log-likelihood

$$\ell(X \mid \mathbf{P}) = \sum_{i,j \in \mathcal{S}} N_{ij} \ln P_{ij}.$$

We also impose the constraint that each row of the transition matrix must sum to one

$$\sum_{j \in \mathcal{S}} P_{ij} = 1, \quad \text{for all} \quad i \in \mathcal{S}. \tag{3.5}$$

To maximize the log-likelihood subject to this constraint, we use Lagrange multipliers. The Lagrangian becomes

$$\mathcal{L}(X, \lambda \mid \mathbf{P}) = \sum_{i,j \in \mathcal{S}} N_{ij} \ln P_{ij} - \sum_{i \in \mathcal{S}} \lambda_i \left(\sum_{j \in \mathcal{S}} P_{ij} - 1 \right), \tag{3.6}$$

where λ_i is the Lagrange multiplier associated with the constraint for row i. Taking the derivative of $\mathcal{L}$ with respect to P_{ij} and setting it to zero gives

$$\frac{\partial \mathcal{L}}{\partial P_{ij}} = \frac{N_{ij}}{P_{ij}} - \lambda_i = 0 \quad \Rightarrow \quad P_{ij} = \frac{N_{ij}}{\lambda_i}. \tag{3.7}$$

Combining this with the constraint in (3.5), we have

$$\sum_{j \in \mathcal{S}} P_{ij} = 1 \quad \Rightarrow \quad \sum_{j \in \mathcal{S}} \frac{N_{ij}}{\lambda_i} = 1 \quad \Rightarrow \quad \sum_{j \in \mathcal{S}} N_{ij} = \lambda_i, \tag{3.8}$$

which, when substituted back into (3.7), gives the final result

$$\hat{P}_{ij} = \frac{N_{ij}}{\sum_{j \in \mathcal{S}} N_{ij}}.$$

3.1.3 Simulation

Simulating Markov processes is relatively straightforward. First, select a starting state and find the corresponding row in the transition matrix, which gives the transition probabilities for moving to each possible next state. Then, sample the next state according to these probabilities. Repeat this process until the chain is of the desired length. A simple outline of the algorithm is given below.

Algorithm 2 Simulating a Markov chain

Require: Simulation horizon $T > 0$
Require: Transition matrix $\mathbf{P}$ on state space $\mathcal{S}$
1: Choose an initial state $x_0 \in \mathcal{S}$
2: **for** $t = 1$ to T **do**
3: Sample $x_t \sim \mathbf{P}(x_{t-1}, \cdot)$
4: **end for**
5: **return** $\{x_t\}_{t=0}^{T}$

3.2 Hawkes process

This section will describe another stochastic process, the Hawkes process and the theory necessary to understand how a variation of this process, the alternating bivariate Hawkes process, can be used to model paroxysmal AF.

3.2.1 Point Processes and counting processes

A counting process is a stochastic process $N(t)_{t\geq 0}$ that satisfy the following conditions [19],

(i) $N(0) \geq 0$,

(ii) $N(t) \in \mathbb{N}_0$,

(iii) $N(s) \leq N(t), \quad s < t$,

(iv) For $s < t$, $N(t) - N(s)$ equals the number of events in the interval $(s, t]$.

It can be seen as the cumulative count of the number of events up to the current time. If one instead looks at the sequence of random event-times $T = \{t_1, t_2, \dots\}$ themselves, it can be considered a point process. Furthermore, let $\mathcal{H}(t) : t \geq 0$ denote the history of the arrival-times up to a time t. Formally, $\mathcal{H}(\cdot)$ is a filtration, that is, a non-decreasing family of σ-algebras representing the information available up to each time t. However, in this context, it is sufficient to understand it as the history of arrival times up to time t. Then, if the associated point process is not explosive, in that it can't have infinite points in a finite time, and if there exists a function $\lambda^*(t)$ such that:

$$\lambda^*(t) = \lim_{h \searrow 0} \frac{\mathbb{E}\left[N(t+h) - N(t) \mid \mathcal{H}(t)\right]}{h},$$

then $\lambda^*(t)$ is called the conditional intensity function of the counting process $N(t)$. In such cases, the conditional intensity fully characterizes the statistical behavior of the process [9]. Intuitively, the conditional intensity function can be interpreted as the instantaneous rate at which events occur at time t. The integral of the conditional intensity function over an interval, gives the expected number of events

in that interval. The simplest case is the process where the conditional intensity is equal to a constant such as

$$\lambda^*(t) = \mu,$$

then $N(t)_{t\geq 0}$ is said to be a Poisson process.

3.2.2 Univariate Hawkes process

Suppose we have a process with a conditional intensity function

$$\lambda^*(t) = \mu + \int_0^t f(t-u)\,\mathrm{d}N(u).$$

where $\mu > 0$ is known as the background intensity and $f : [0 : \infty) \to [0, \infty)$ is the excitation kernel. Such a process is known as the Hawkes Process [9]. An equivalent way to define it is by instead using the event-timings as a point process rather than a counting process. The conditional density for the Hawkes process then becomes

$$\lambda^*(t) = \mu + \sum_{t_i<t} f(t-t_i), \tag{3.9}$$

where t_i denote the event times occuring before time t. A common choice for the excitation kernel $f(t)$ is to use a function with exponential decay, such as $f(t) = \alpha\beta \mathrm{e}^{-\beta t}$, which gives the conditional intensity

$$\lambda^*(t) = \mu + \sum_{t_i<t} \alpha\beta \mathrm{e}^{-\beta(t-t_i)}, \tag{3.10}$$

where $\alpha > 0$ and $\beta > 0$ are constants that can be interpreted as the expected number of offspring from any single event and the decay rate of the excitations, respectively.

If we are working with two point processes that influence each other, the Hawkes process can be extended to a bivariate form. This bivariate Hawkes process will have two conditional intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$, one for each point process, and the conditional intensities will contain additional terms that account for the cross-excitation between two processes. Let $N_1(t)$ and $N_2(t)$ be two counting processes with conditional intensities $\lambda_1^*(t)$ and $\lambda_2^*(t)$, respectively. Furthermore, let $t_i^{(1)}$ and $t_j^{(2)}$ denote the event times occuring before time t in the first and second counting process, respectively. Using exponential kernels, the model becomes

$$\lambda_1^*(t) = \mu_1 + \sum_{t_i^{(1)}<t} \alpha_{11}\beta_{11}\mathrm{e}^{-\beta_{11}(t-t_i^{(1)})} + \sum_{t_j^{(2)}<t} \alpha_{12}\beta_{12}, \mathrm{e}^{-\beta_{12}(t-t_j^{(2)})},$$
$$\lambda_2^*(t) = \mu_2 + \sum_{t_j^{(2)}<t} \alpha_{22}\beta_{22}\mathrm{e}^{-\beta_{22}(t-t_j^{(2)})} + \sum_{t_i^{(1)}<t} \alpha_{21}\beta_{21}, \mathrm{e}^{-\beta_{21}(t-t_i^{(1)})}, \tag{3.11}$$

where $\mu_1, \mu_2 > 0$ are the baseline intensities. The parameters $\alpha_{ij} \geq 0$, for $i, j \in \{1, 2\}$ represent the expected number of new events for process i due to an event in process j. The parameters $\beta_{ij} > 0$, for $i, j \in \{1, 2\}$ are the corresponding decay rates of the excitations of process i from process j. When $i = j$ the terms model self-excitation, and when $i \neq j$, they represent cross-excitation.

3.2.3 Alternating bivariate Hawkes process

The alternating bivariate Hawkes process, introduced in [8], is a variation of the bivariate Hawkes process with an adaptation that makes it more suitable for paroxysmal AF modeling. The alternating bivariate Hawkes process imposes a restriction on the counting processes such that the events in each point process have to come alternate of each other, for example

$$t_1^{(1)} < t_1^{(2)} < t_2^{(1)} < t_2^{(2)} < t_3^{(1)} < t_3^{(2)} < \ldots,$$

where $t_n^{(m)}$ denotes the time of the nth event in the process $N_m(t)$, for $m = 1, 2$. This alternating structure can be enforced by multiplying the conditional intensities with an occurrence function

$$o_1(t) = \begin{cases} 1, & N_1(t) = N_2(t), \\ 0, & \text{otherwise}, \end{cases} \tag{3.12}$$
$$o_2(t) = \begin{cases} 1, & N_2(t) \neq N_1(t), \\ 0, & \text{otherwise}. \end{cases} \tag{3.13}$$

In [8], the alternating bivariate Hawkes process included two constants d_1, d_2, that defined the minimum duration of respective episode before a transition was possible. However, since the dataset used in this study had no such constraint, these were omitted by setting them to 0. The modified conditional intensities are defined as

$$\tilde{\lambda}_1^*(t) = \lambda_1^*(t)o_1(t),$$
$$\tilde{\lambda}_2^*(t) = \lambda_2^*(t)o_2(t). \tag{3.14}$$

This modification ensures that the process alternates between the two states.

3.2.4 Likelihood

Fitting the alternating bivariate Hawkes process to data requires a likelihood function. For a simple point process, meaning a point process where no two points coincide, the likelihood function is outlined in the following theorem [9].

Theorem 3.2.1 (Point Process Likelihood)**.** *Let $N(t)$ be a simple point process on the interval $[0, \tau]$, with the associated conditional intensity function $\lambda(t)$. Let $T = \{t_1, \dots, t_n\}$ denote the observed event times within this interval. Then, the likelihood of observing the event times T given the intensity function λ, is*

$$L(T \mid \lambda) = \left[\prod_{i=1}^{n} \lambda(t_i)\right] \cdot \exp\left(-\int_0^\tau \lambda(s)\,\mathrm{d}s\right).$$

The theorem can be extended to multivariate point processes.

Theorem 3.2.2 (Multivariate Point Process Likelihood)**.** *Let $\boldsymbol{N}(t)$ be a collection of m simple point processes $(N_1(t), \dots, N_m(t))$, on the interval $[0, \tau]$, with the joint associated conditional intensity $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)$. Let $\boldsymbol{T} = \left(T^{(1)}, \dots, T^{(m)}\right)$ be the observed points from the corresponding point processes in the interval. Then, the likelihood of observing the event times $\boldsymbol{T}$, is*

$$L(\boldsymbol{T} \mid \boldsymbol{\lambda}) = \prod_{i=1}^{m} \left[\left(\prod_{t_j^{(i)} \in T^{(i)}} \lambda_i(t_j^{(i)})\right) \cdot \exp\left(-\int_0^\tau \lambda_i(s)\mathrm{d}s\right)\right].$$

The theorems are proved in [9]. The multivariate point process likelihood theorem can be used to describe the likelihood of the alternating bivariate point process. Let the observed event times for the two point processes be $T^{(1)} = \{t_1^{(1)}, t_2^{(1)}, \dots\}$ and $T^{(2)} = \{t_1^{(2)}, t_2^{(2)}, \dots\}$. Furthermore, let θ be a vector containing all of the model constants, and let $\tilde{\lambda}_1^*(t)$ and $\tilde{\lambda}_2^*(t)$ be the conditional intensity functions of the alternating bivariate Hawkes process as defined in (3.14). Let $M = \{(1,2),(2,1)\}$ denote the set of index pairs indicating the currently active process and its counterpart, and finally also introduce I^m as the set of intervals in which the occurrence function $o_m(t)$ is equal to 1. For example

$$\begin{aligned} I^{(1)} &= \{[0, t_1^{(1)}), [t_1^{(2)}, t_2^{(1)}), \dots\}, \\ I^{(2)} &= \{[t_1^{(1)}, t_1^{(2)}), [t_2^{(1)}, t_2^{(2)}), \dots\}. \end{aligned} \tag{3.15}$$

Then, the likelihood function of the alternating bivariate Hawkes process can be written as

$$L(T^{(1)}, T^{(2)} \mid \theta) = \prod_{(m,\tilde{m})\in M} \left[\left(\prod_{t_i^{(m)} \in T^{(m)}} \tilde{\lambda}_m^*(t_i^{(m)}) \right) \cdot \exp\left(- \sum_{(u,v)\in I^{(m)}} \int_u^v \tilde{\lambda}_m^*(t)\,\mathrm{d}t \right) \right], \tag{3.16}$$

From this, we now derive the likelihood of the alternating bivariate Hawkes process with exponential kernels. Taking the logarithm of the likelihood function yields

$$\ell(T^{(1)}, T^{(2)} | \theta) = \sum_{(m,\tilde{m})\in M} \left[\sum_{t_i^{(m)} \in T^{(m)}} \ln\left(\tilde{\lambda}_m^*(t_i^{(m)})\right) - \sum_{(u,v)\in I^{(m)}} \int_u^v \tilde{\lambda}_m^*(t)\mathrm{d}t \right]. \tag{3.17}$$

Inserting (3.11) into the first term of the inner sum in (3.17) yields

$$\begin{aligned} \sum_{t_i^{(m)} \in T^{(m)}} \ln\left(\tilde{\lambda}_m^*(t_i^{(m)})\right) = \sum_{t_i^{(m)} \in T^{(m)}} \ln\Big(\mu_m &+ \sum_{t_j^{(m)} < t_i^{(m)}} \alpha_{mm}\beta_{mm} \exp\big(-\beta_{mm}(t_i^{(m)} - t_j^{(m)})\big) \\ &+ \sum_{t_j^{(\tilde{m})} < t_i^{(m)}} \alpha_{m\tilde{m}}\beta_{m\tilde{m}} \exp\big(-\beta_{m\tilde{m}}(t_i^{(m)} - t_j^{(\tilde{m})})\big)\Big), \end{aligned} \tag{3.18}$$

and inserting (3.11) into the second term of the inner sum in (3.17) gives

$$\begin{aligned} &\sum_{(u,v)\in I^{(m)}} \int_u^v \tilde{\lambda}_m^*(t)\mathrm{d}t \\ &\quad= \sum_{(u,v)\in I^{(m)}} \int_u^v \left(\mu_m + \sum_{t_j^{(m)} < t} \alpha_{mm}\beta_{mm} \exp\Big(-\beta_{mm}(t - t_j^{(m)})\Big) \right. \\ &\qquad\qquad \left. + \sum_{t_j^{(\tilde{m})} < t} \alpha_{m\tilde{m}}\beta_{m\tilde{m}} \exp\Big(-\beta_{m\tilde{m}}(t - t_j^{(\tilde{m})})\Big) \right) \mathrm{d}t \\ &\quad= \sum_{(u,v)\in I^{(m)}} \left((v-u)\mu_m + \int_u^v \sum_{t_j^{(m)} < t} \alpha_{mm}\beta_{mm} \exp\Big(-\beta_{mm}(t - t_j^{(m)})\Big)\mathrm{d}t \right. \\ &\qquad\qquad \left. + \int_u^v \sum_{t_j^{(\tilde{m})} < t} \alpha_{m\tilde{m}}\beta_{m\tilde{m}} \exp\Big(-\beta_{m\tilde{m}}(t - t_j^{(\tilde{m})})\Big)\mathrm{d}t \right). \end{aligned} \tag{3.19}$$

To proceed with evaluating the likelihood, we now compute the integrals. Starting with the self-excitation term

$$\begin{aligned}
&\int_u^v \sum_{t_j^{(m)}<t} \alpha_{mm}\beta_{mm}\exp(-\beta_{mm}(t-t_j^{(m)}))\mathrm{d}t \\
&\quad= \sum_{t_j^{(m)}\le u} \int_u^v \alpha_{mm}\beta_{mm}\exp(-\beta_{mm}(t-t_j^{(m)}))\mathrm{d}t, \\
&\quad= \sum_{t_j^{(m)}\le u} \alpha_{mm}\beta_{mm}\int_u^v \exp(-\beta_{mm}(t-t_j^{(m)}))\mathrm{d}t, \\
&\quad= -\sum_{t_j^{(m)}\le u} \alpha_{mm}\Big(\exp(-\beta_{mm}(v-t_j^{(m)})) - \exp(-\beta_{mm}(u-t_j^{(m)}))\Big). \qquad (3.20)
\end{aligned}$$

The order of summation and integration can be interchanged. This step is justified by Tonelli's Theorem [20], which states that for a sequence of non-negative, measurable functions $f_n(t)$, the integral of their sum is equal to the sum of their integrals:

$$\int \sum_n f_n(t)\,dt = \sum_n \int f_n(t)\,dt.$$

Since each term $\alpha_{mm}\beta_{mm}\exp(-\beta_{mm}(t-t_j^{(m)}))$ in our sum is non-negative, the theorem applies. The cross-excitation term is computed similarly to the self-excitation

$$\begin{aligned}
&\int_u^v \sum_{t_j^{(\tilde{m})}<t} \alpha_{m\tilde{m}}\beta_{m\tilde{m}}\exp(-\beta_{m\tilde{m}}(t-t_j^{(\tilde{m})}))\mathrm{d}t \\
&\quad= -\sum_{t_j^{(\tilde{m})}\le u} \alpha_{m\tilde{m}}\Big(\exp(-\beta_{m\tilde{m}}(v-t_j^{(\tilde{m})})) - \exp(-\beta_{m\tilde{m}}(u-t_j^{(\tilde{m})}))\Big). \qquad (3.21)
\end{aligned}$$

Next, we introduce the functions $C_{uv}(t)$ and $C_{uv}^*(t)$ to keep the final form more compact

$$\begin{aligned}
C_{uv}(t) &= \sum_{t_j^{(v)}<t} \alpha_{uv}\exp\left(-\beta_{uv}(t-t_j^{(v)})\right), \\
C_{uv}^*(t) &= C_{uu}(t) + C_{uv}(t). \qquad (3.22)
\end{aligned}$$

Combining the results from (3.20) and (3.21), and substituting them into (3.19), we obtain a simplified expression for the integral term. Together with the first term in (3.18), this is then inserted into (3.17), yielding the final form of the log-likelihood for the alternating bivariate Hawkes process

$$\ell(T^{(1)}, T^{(2)} \mid \theta) = \sum_{(m,\tilde{m})\in M} \left\{ \sum_{t_i^{(m)} \in T^{(m)}} \ln\left(\mu_m + \beta_{mm} C_{mm}(t_i^{(m)}) + \beta_{m\tilde{m}} C_{m\tilde{m}}(t_i^{(m)})\right) + \sum_{(u,v)\in I^{(m)}} \left((v-u)\mu_m - C^*_{m\tilde{m}}(v) + C^*_{m\tilde{m}}(u)\right)\right\}. \tag{3.23}$$

This likelihood function can be optimized using numerical optimization methods.

3.2.5 Ogata's thinning algorithm

Ogata's thinning algorithm is a widely used method for simulating point processes such as the Hawkes process [21]. Starting from the current time t_0, an upper bound λ_0 of the intensity function until the next event is computed. A proposed event time t_p is then sampled from an exponential distribution with rate λ_0. To decide whether to accept this proposed event, the intensity at time t_p, denoted $\lambda_p = \lambda^*(t_p)$ is evaluated, and a uniform random variable $r \sim \mathcal{U}(0,1)$ is drawn. The event at t_p is accepted if $r \leq \lambda_p/\lambda_0$. Regardless of acceptance, the current time is updated to t_p. A description of the algorithm adapted to the alternating bivariate Hawkes process is presented below.

Algorithm 3 Ogata's Thinning Algorithm for alternating bivariate Hawkes processes with exponential kernels.

Require: Simulation horizon $T > 0$
Require: Conditional intensity functions $\lambda_1^*(t)$, $\lambda_2^*(t)$
Initialize $t_0 \leftarrow 0$
Initialize event lists: $\mathcal{T}^{(1)} \leftarrow [\,]$, $\mathcal{T}^{(2)} \leftarrow [\,]$
while $t_0 < T$ **do**
    for $m \in \{1, 2\}$ **do**
        Compute intensity upper bound $\lambda_0 = \lambda_m^*(t_0)$
        Sample $u \sim \mathcal{U}(0,1)$
        Propose event time $t_p = t_0 + \frac{1}{\lambda_0} \log\left(\frac{1}{u}\right)$
        Compute $\lambda_p = \lambda_m^*(t_p)$
        Sample $r \sim \mathcal{U}(0,1)$
        if $r \leq \frac{\lambda_p}{\lambda_0}$ **then**
            Append t_p to $\mathcal{T}^{(m)}$
        end if
        Update $t_0 \leftarrow t_p$
    end for
end while
return $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}$

3.2.6 Compensators and goodness of fit

When working with point processes, the integrated conditional intensity function is often a useful tool. It is known as the compensator and is defined as

$$\Lambda(t) = \int_0^t \lambda^*(s)\mathrm{d}s.$$

A key application of the compensator is the random time change theorem, which allows us to transform a realization of a point process into an equivalent series of points drawn from a unit-rate Poisson process. The following theorem is stated as in [9].

Theorem 3.2.3 (Random Time Change Theorem)**.** *Say* $\{t_1, t_2, \dots, t_k\}$ *is a realization over time* $[0, T]$ *from a point process with conditional intensity function* $\lambda^*(\cdot)$*. If* $\lambda^*(\cdot)$ *is positive over* $[0, T]$ *and* $\Lambda(T) < \infty$ *almost surely, then the transformed points* $\{\Lambda(t_1), \Lambda(t_2), \dots, \Lambda(t_k)\}$ *form a Poisson process with unit rate.*

This transformed series of points can be used in a goodness of fit analysis, using standard methods for Poisson processes. This is confirmed by the following theorem, which also is stated as in [9].

Theorem 3.2.4 (Residual Analysis)**.** *Consider an unbounded, increasing sequence of time points* $\{t_1, t_2, \dots\}$ *in the half-line* $(0, \infty)$*, and a monotonic, continuous compensator* $\Lambda(\cdot)$ *such that* $\lim_{t\to\infty} \Lambda(t) = \infty$ *almost surely. The transformed sequence* $\{t_1^*, t_2^*, \dots\} = \{\Lambda(t_1), \Lambda(t_2), \dots\}$ *, whose counting process is denoted* $N^*(t)$ *is a realization of a unit rate Poisson process if and only if the original sequence* $\{t_1, t_2, \dots\}$ *is a realization from the point process defined by* $\Lambda(\cdot)$*.*

For the alternating bivariate Hawkes process, we can apply the random time change theorem to our two series of events to get two new series which under a good model fit, should resemble realizations from a Poisson distribution. In our case the time rescaled series of events can be computed using the compensator

$$\Lambda_m(t) = \int_0^t \tilde{\lambda}_m^*(s)\mathrm{d}s,$$

for $m \in \{1, 2\}$. Since the interarrival times of a unit-rate Poisson process are independent and exponentially distributed then the transformed interarrival times

$$\{\tau_1^{(m)}, \tau_2^{(m)}, \tau_3^{(m)}, \dots\} = \{\Lambda(t_1^m) - 0, \Lambda(t_2^{(m)}) - \Lambda(t_1^{(m)}), \Lambda(t_3^{(m)}) - \Lambda(t_2^{(m)}), \dots\},$$

should follow $\tau_i^{(m)} \sim \text{Exp}(1)$. These can be transformed one step further into uniformly distributed random variables by taking the transformation

$$z_i^{(m)} = 1 - \mathrm{e}^{-\tau_i^{(m)}},$$

such that $z_i^{(m)} \sim \mathcal{U}(0,1)$, assuming our data is distributed according to the process. A nonparametric approach to evaluating the goodness of fit is the Kolmogorov-Smirnov (KS) test, which was used to assess the fit of the alternating bivariate Hawkes process to AF patterns in [8]. The KS test compares the empirical distribution function (EDF) of some data, to the cumulative distribution function (CDF) of the reference distribution. The CDF of the uniform distribution is a 45-degree line, thus, for a good fit, the EDF of our transformed data should lie close to this diagonal. The test statistic d_{KS}, or the KS-distance, is defined as the maximum difference between the EDF and CDF. Smaller values indicate a better fit.

4

Methods

This chapter presents the methodology for evaluating the Apple Watch's AF detection algorithm. It first introduces the dataset and the data preprocessing techniques. Then, some limitations of the approach used are explained. Finally the chapter provides a detailed explanation of how the two modeling approaches, the Markov chain and the Hawkes process, are used to assess the algorithm's performance.

4.1 Dataset

In this thesis, a dataset of patients examined using PocketECG in the United States in 2021 is used. PocketECG is a wearable long-term ambulatory ECG monitoring device designed by MEDICALgorithmics, Warsaw, Poland. The ECG recordings were labeled by DeepRhythmAI (DRAI), an artificial intelligence (AI) algorithm developed by MEDICALgorithmics, which has been proven to have a higher negative predictive value than human technicians [22]. The full dataset consisted of 32,162 patients with a mean ($\pm$ standard deviation) age of 59.3 ± 18.8 years, and 40.6% were men. From this dataset, only patients with an AF burden greater than 0% but less than 100% were selected, resulting in a study population of 2,486 patients, with a mean age of 69.8 ± 12.3 years and 51.9% men. The average monitoring duration in this subset was 18.2 ± 10.0 days, and the maximum monitoring duration is 30 days. Since the alternating bivariate Hawkes algorithm has more parameters, it also requires more data to achieve a reliable fit. In [8], the authors evaluated how the number of transitions impacted the quality of the fit, and chose 20 transitions in total as a minimum. A similar criterion was applied in this thesis. When fitting the alternating bivariate Hawkes process, the dataset was filtered to only include patients with at least 10 transitions in each direction. Furthermore, because high burden patients were unlikely to be missed by the watch, they were not of interest for the analysis, and only those with a burden of $\leq 20\%$ were included. This resulted in a sample of 1,087 patients with a mean age of 69.9 ± 12.4 years and 49.9% men.

The original dataset, which consisted of beats labeled with different arrhythmias by DRAI, was first preprocessed to identify episodes of consecutive AF or SR beats. Episodes lacking valid measurements or containing other unrelated arrhythmias were labeled as missing. For the Markov method, the data also needed to be discretized into one-minute segments. This was achieved by dividing each day into one-minute

intervals and labeling each interval according to the majority rhythm, AF or SR. Intervals that did not contain a full minute of data were discarded.

4.2 Limitations

Because the algorithm Apple uses to classify each PPG segment as normal or arrhythmic is proprietary, the exact performance could not be verified. Instead, it was assumed that if a one-minute interval contained at least 30 seconds of AF in total, then Apple's algorithm would be able to correctly identify it as arrhythmic. Furthermore, since simulating the wear and activity patterns of the smartwatches would add an additional layer of complexity, results were calculated under the assumption that the watch is worn all the time, and that it can always read an adequate signal.

4.3 Markov chain approach

The first method relied on discretizing the data into one-minute segments, and modeling the transitions between states as a Markov chain. First, a two-state Markov transition matrix was fitted to the discretized data of each patient separately, using MLE. The transition probabilities of a two-state Markov chain are given in (4.1). In our case, let p denote the probability of transitioning from SR to AF, and q the probability of transitioning from AF to SR. Then the transition matrix $\mathbf{P}$ is defined as

$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}. \tag{4.1}$$

Using the MLE formulas in (3.4), the transition probabilities p and q were estimated as

$$p = \frac{N_{\mathrm{SR}\to\mathrm{AF}}}{N_{\mathrm{SR}\to\mathrm{SR}} + N_{\mathrm{SR}\to\mathrm{AF}}},$$
$$q = \frac{N_{\mathrm{AF}\to\mathrm{SR}}}{N_{\mathrm{AF}\to\mathrm{AF}} + N_{\mathrm{AF}\to\mathrm{SR}}},$$

where $N_{\mathrm{SR}\to\mathrm{AF}}$ is the number of transitions from SR to AF observed in the data, and similarly for the other terms. Given the transition probabilities between each one-minute segment and assuming the Markov property, we can construct a Markov chain that simulates the Apple Watch's algorithm. However, because the Apple Watch waits a minimum of 15 minutes between each one-minute PPG, we first need the transition probabilities for a 15 minute interval. Under the Markov assumption, this can be computed by taking the 15th power of the transition matrix

$$\mathbf{R} = \mathbf{P}^{15},$$

where **R** is the transition matrix for a 15 minute wait. Using this, a complete Markov chain for the Apple Watch algorithm is constructed, as shown in Figure 4.1.

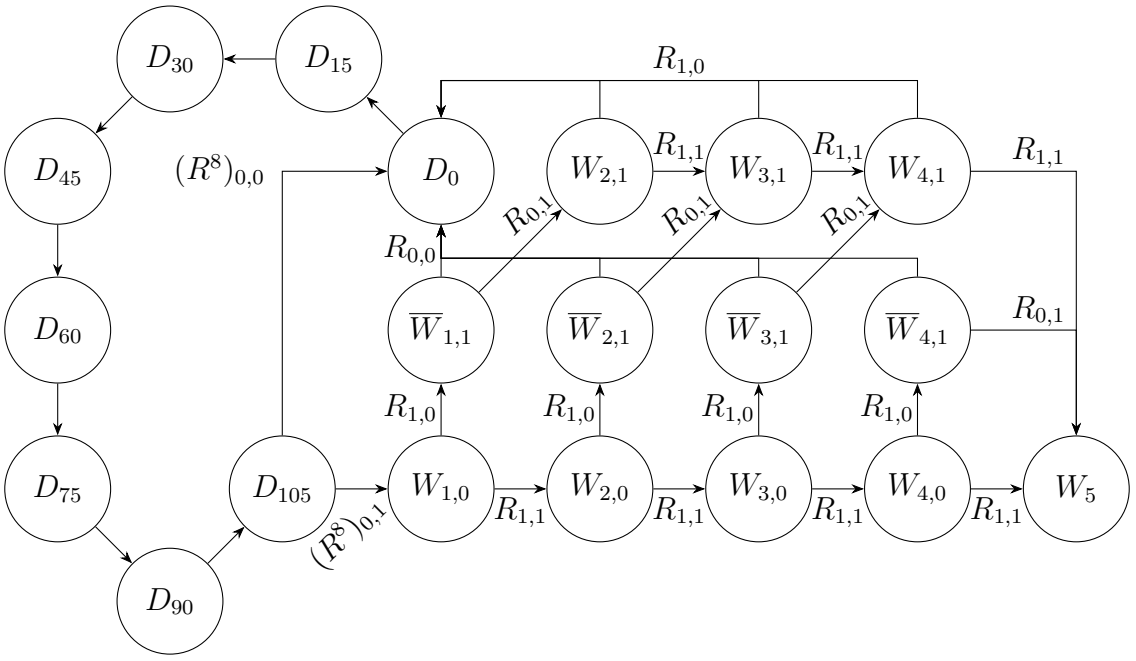


Figure 4.1: Markov chain representation of the Apple Watch algorithm. **R** is the 15-minute transition matrix. The algorithm keeps track of the number of found AF segments and the number of SR segments found in the first and second index of the $W_{x,y}$ states. In the states $\overline{W}_{x,y}$ the overline indicates that the previous recorded tachogram was classified as having a regular rhythm. If two SR segments are recorded before five AF, or if the first segment is SR after a two hour sleep, the algorithm enters another two hour sleep. This sleep is represented by the D_t states where t is the time slept so far. If the chain reaches the absorbing state W_5, the watch notifies the user of potential AF. Each transition represents 15 minutes of elapsed time.

The Markov chain representation of the Apple Watch algorithm consists of twenty states, with the final state, where the smartwatch issues an alert, being absorbing. Given any pair of transition probabilities p and q, two key metrics can be computed analytically. First, the average proportion of time spent in AF can be found from the stationary distribution, (3.2), of the two-state Markov chain described in (4.1). By solving for the stationary distribution $\boldsymbol{\pi}$, we obtain

$$\boldsymbol{\pi} = \left[\frac{q}{p+q}, \frac{p}{p+q}\right], \tag{4.2}$$

where the second entry gives the average time spent in the AF state, which is the same as the AF burden. Second, the expected time until an alert is issued can be determined using the fundamental matrix of the full Markov chain representation of the Apple Watch algorithm shown in Figure 4.1.

While the theoretical values of these metrics are known for any given p and q, estimating their distribution in the population is necessary to draw meaningful conclusions at the population level. To this end, MLE was applied to compute individual estimates of the transition rates, p and q, for each patient in the dataset. The resulting estimates were then used to fit lognormal distributions for both parameters, which serve as approximations of the underlying population-level distributions. However, due to the recording duration being limited to 30 days in the dataset, the smallest observable transition rate corresponds to one transition in 30 days. This constraint limits the ability to capture patients with extremely sparse AF episodes. To assess sensitivity to this limitation, two approaches were evaluated. In the first approach, the lognormal distributions were extrapolated beyond the observable range, and in the second approach a lower bound was imposed to restrict samples to the observable range. While both approaches were explored, the bounded method was used in subsequent analysis to avoid relying on extrapolations.

Using the estimated population level distributions of p and q, a Monte Carlo simulation with 30 million iterations was performed to estimate the Apple Watch's detection rate in a realistic population. Each iteration followed these steps:

(i) Sample transition probabilities p and q for the two-state Markov chain.

(ii) Simulate a five-year sequence of transitions between the two states.

(iii) Apply the Apple Watch detection algorithm to the simulated sequence, and record the time of the first alert.

(iv) Calculate the corresponding AF burden using the stationary distribution of the Markov chain.

The results were used to generate plots illustrating the number of undetected AF cases over time, stratified by the AF burden.

4.4 Hawkes process approach

For the Hawkes process, parameters were estimated using numerical optimization via the `minimize` function from the `SciPy` library in Python. Specifically, the negative log-likelihood of the alternating bivariate Hawkes process in (3.23) was minimized using the `L-BFGS-B` algorithm, subject to the following constraints

$$\begin{aligned} \mu_i &\geq 10^{-16} \quad \text{for} \quad i \in \{1,2\}, \\ \alpha_{ij} &\geq 0 \quad \text{for} \quad i,j \in \{1,2\}, \\ \beta_i &\geq 10^{-5} \quad \text{for} \quad i \in \{1,2\}. \end{aligned}$$

The constraint on μ_i ensures a strictly positive baseline intensity, while the lower bound on β_i prevents the excitation half-life from exceeding approximately 20 hours.

This was necessary because the likelihood surface turned out to often be relatively flat with respect to β_i, which could otherwise lead to convergence to implausibly low values. As in [8], the number of parameters was reduced by setting $\beta_{11} = \beta_{12} = \beta_1$ and $\beta_{21} = \beta_{22} = \beta_2$. The constraint on α_{ij} keeps the branching ratio non-negative for all excitations.

Given the relatively high dimensionality of the parameter space and the comparatively limited number of samples, the parameter distributions at the population level were not modeled directly. Instead, the fitted parameter sets were treated empirically, and each patient's parameters were assumed to represent a possible realization from the target distribution.

Using Ogata's thinning algorithm, a large number of five-year sequences of heart rhythm patterns were simulated for each parameter set. The Apple Watch algorithm was then applied to each sequence to determine the time until the first alert. Between 50 and 1000 sequences were simulated and analyzed per patient, with the exact number varying depending on the computational intensity of that AF pattern. To avoid bias from unequal simulation counts, the results were weighted such that each patient contributed equally to the final results. AF burden was computed separately for each parameter set by simulating a 10-year long sequence and calculating the proportion of time spent in AF. Finally, the results were stratified by AF burden, and the proportion of undetected AF cases was visualized over time.

5

Results

This chapter presents the results from the two models used to evaluate the Apple Watch's AF detection algorithm. The findings are organized by model, starting with the results from the Markov chain approach, and followed by the findings from the alternating bivariate Hawkes process. The chapter concludes with a comparison of the two approaches.

5.1 Markov chain results

The number of transitions between AF and SR was counted for each of the patients in the dataset, and the transition probabilities were estimated using MLE. The results are scattered in Figure 5.1.

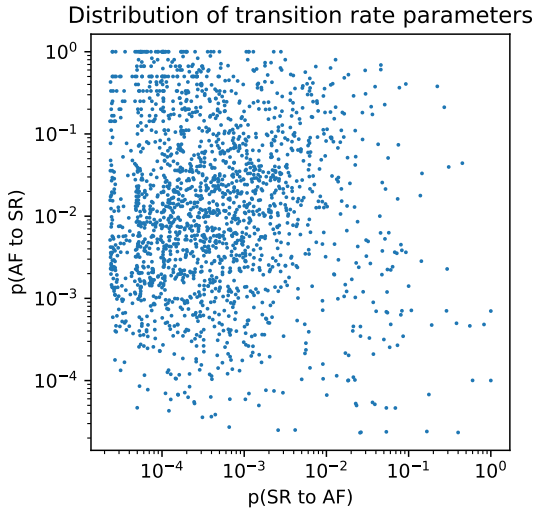


Figure 5.1: Scatterplot of the transition-rate parameter estimates obtained using MLE for the Markov chain. The axes are log-scaled.

The Pearson correlation coefficient between the two transition probabilities is 0.038, with a p-value of 0.096, which indicates that there is no statistically significant correlation. Therefore, to model the distribution of the transition rates, two independent lognormal distributions were used. The fitted lognormal distributions are plotted in Figure 5.2.

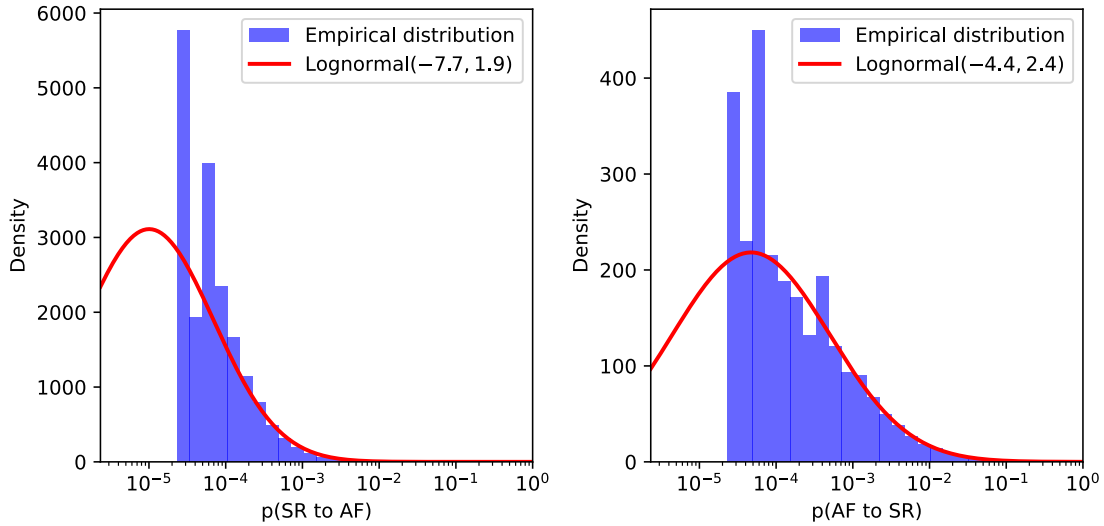
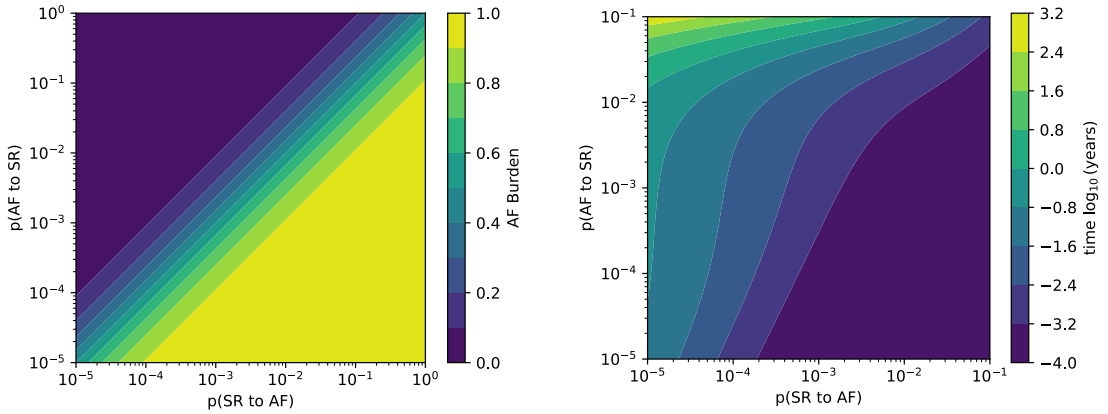


Figure 5.2: Histograms as well as fitted lognormal distributions to the data. The fitted distributions are of the form Lognormal(μ, σ), with the parameters shown in the figure legend.

It should be noted that since the longest observation times are one month, the minimal transition rate possible in this dataset is an expected rate of one transition in 30 days. This gives a transition probability of $1/(30 \cdot 24 \cdot 60) \approx 2.31 \cdot 10^{-5}$, which explains why the histogram cuts off at that location. Furthermore, this also explains the observed spikes near the lower end of the histogram, since cases with AF episodes that are more sparse than one in 30 days still contribute to the first histogram bin if a single AF event occurs within the first 30 days of monitoring. Similarly, the second peak is caused by the cases where we have exactly two transitions in 30 days.

To analyze how the different transition rates affect the burden and the expected time to an alert these were plotted in Figure 5.3. From these figures we can see that the expected time to alert varies greatly even for parameter combinations that share the same AF burden. In general the expected time to detection increases the shorter the AF episodes are and decreases with an increased frequency of transitioning to AF.



(a) AF burden for different pairs of the parameters p and q.

(b) Expected time until an alert by the Apple Watch for different pairs of the parameters p and q.

Figure 5.3: Visualizations of how the transition rate parameters affect the AF burden and the expected time to an alert from the Apple-watch. The burden is calculated from the stationary distribution of a two-state Markov chain. The expected time to detection is calculated from the fundamental matrix of the Markov chain in Figure 4.1, and assuming that the Markov chain starts in state $W_{1,0}$ with a probability p_b and D_0 with a probability $1 - p_b$ where p_b is the AF burden.

Another perspective on the expected time to detection comes from reparametrizing the axes. Instead of using the transition probabilities p and q directly, the axes are transformed to represent the AF burden, $p/(p+q)$, and a scale parameter, $p+q$, which controls the overall transition rate. Figure 5.4 shows the expected detection time under this transformation, along with estimated transition rates for the patients in the data, transformed accordingly.

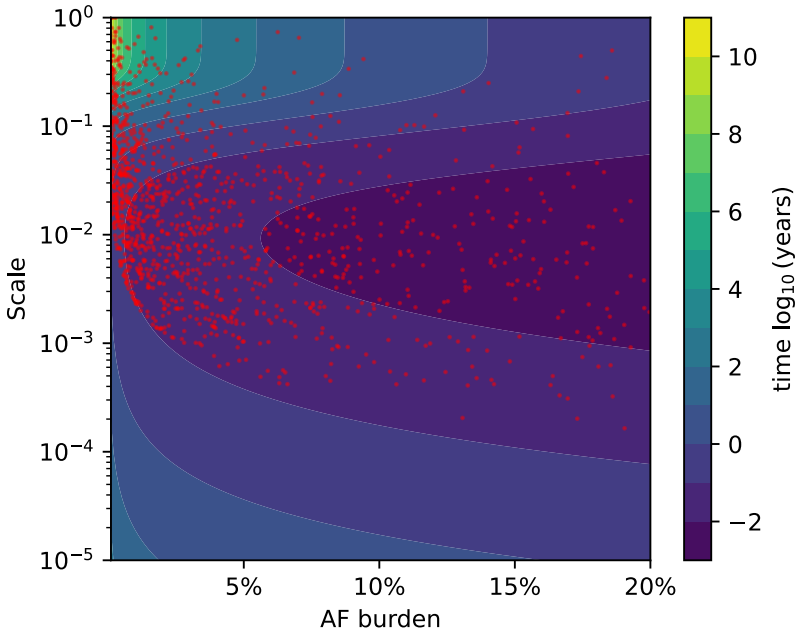


Figure 5.4: Expected time to detection across combinations of AF burden and transition scale. The scattered dots represent the estimated transition rates from the data, reparametrized into burden and scale.

The plot confirms that the expected detection time varies considerably, even when the burden is held constant. For most burdens, the shortest expected detection times occur at a transition scale of around 1%. However, the scattered data show that many individuals experience AF patterns outside this easily detectable range.

Following the theoretical analysis, the detection performance was quantified by simulating the algorithm's performance on a wide range of AF patterns. The results are presented in survival plots, stratified by the AF burden, which show how many AF cases remain undetected over a period of five years. Figure 5.5 shows the primary results, based on the model where transition rates were constrained to the observed data range. To evaluate the impact of this constraint, Figure 5.6 compares the outcome of the non-extrapolated and extrapolated models. The exact numerical data for each year are detailed in Tables 5.1 and 5.2.

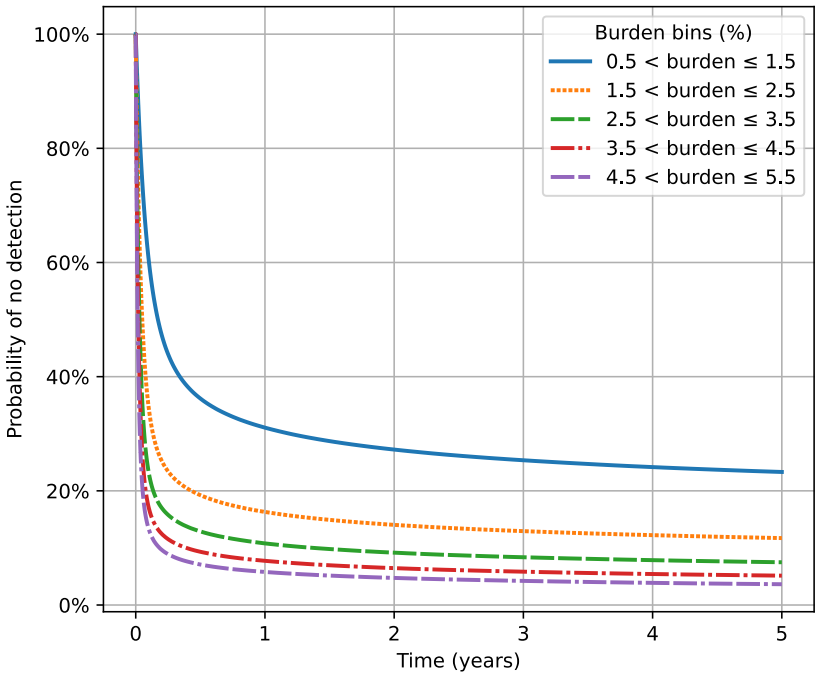


Figure 5.5: Kaplan-Meier plots for the detection time for different burdens. The transition rates were samples from the lognormal distributions with a minimum transition rate of 1 in 30 days to avoid extrapolation.

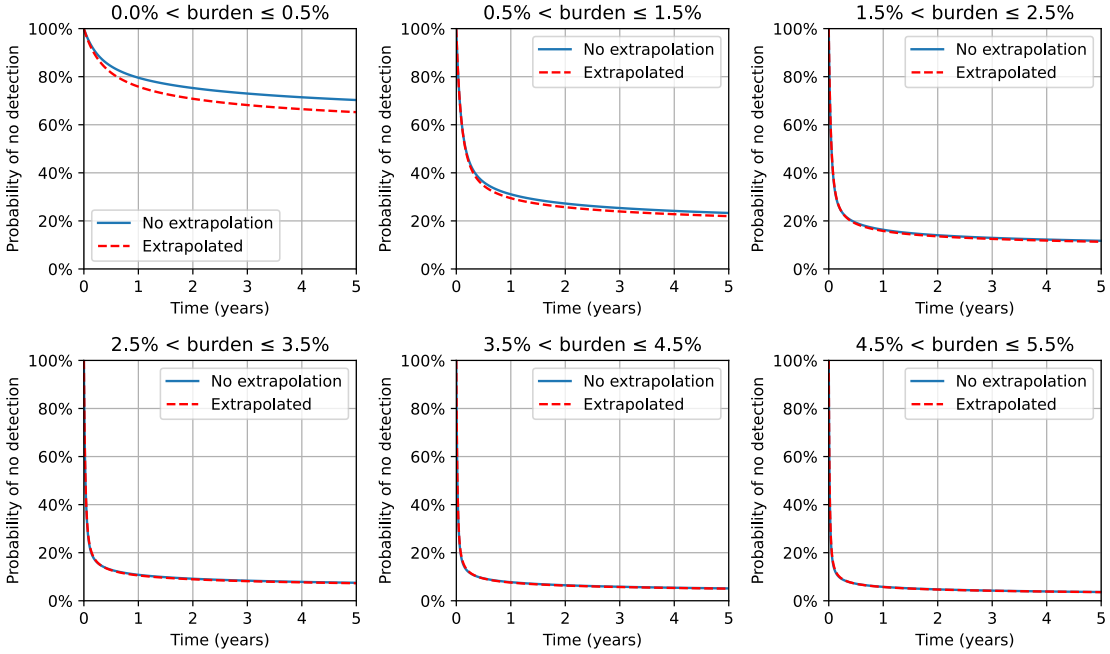


Figure 5.6: Comparison between drawing any p and q from the lognormal distributions and limiting the transition rates to the region where we have observed data.

Table 5.1: Percent not alerted over time for different AF burdens. Parameters drawn without extrapolating.

Years	0.5%–1.5%	1.5%–2.5%	2.5%–3.5%	3.5%–4.5%	4.5%–5.5%
1	31.1	16.3	10.8	7.7	5.8
2	27.2	14.0	9.2	6.5	4.7
3	25.4	12.9	8.4	5.8	4.2
4	24.2	12.2	7.9	5.4	3.9
5	23.3	11.7	7.5	5.1	3.6

Table 5.2: Percent not alerted over time for different AF burdens. Parameters drawn with extrapolation.

Years	0.5%–1.5%	1.5%–2.5%	2.5%–3.5%	3.5%–4.5%	4.5%–5.5%
1	29.4	15.8	10.6	7.6	5.7
2	25.7	13.6	8.9	6.3	4.7
3	23.9	12.5	8.2	5.7	4.2
4	22.8	11.8	7.7	5.3	3.8
5	22.0	11.3	7.3	5.0	3.6

5.2 Hawkes process results

The individual distributions of the parameters of the fitted alternating bivariate Hawkes processes are shown in histograms in Figure 5.7. Using these fitted parameters the performance of the Apple Watch algorithm was simulated, with the results presented as a survival plot in Figure 5.8. The plot is stratified by the AF burden, and Table 5.3 provides the number of patients within each burden category. The exact values for each year can also be found in Table 5.4.

Table 5.3: Number of patients with an AF burden within the ranges of the bins for the Hawkes process.

Bin Range	Count
0%–0.5%	392
0.5%–1.5%	180
1.5%–2.5%	100
2.5%–3.5%	81
3.5%–4.5%	42
4.5%–5.5%	40

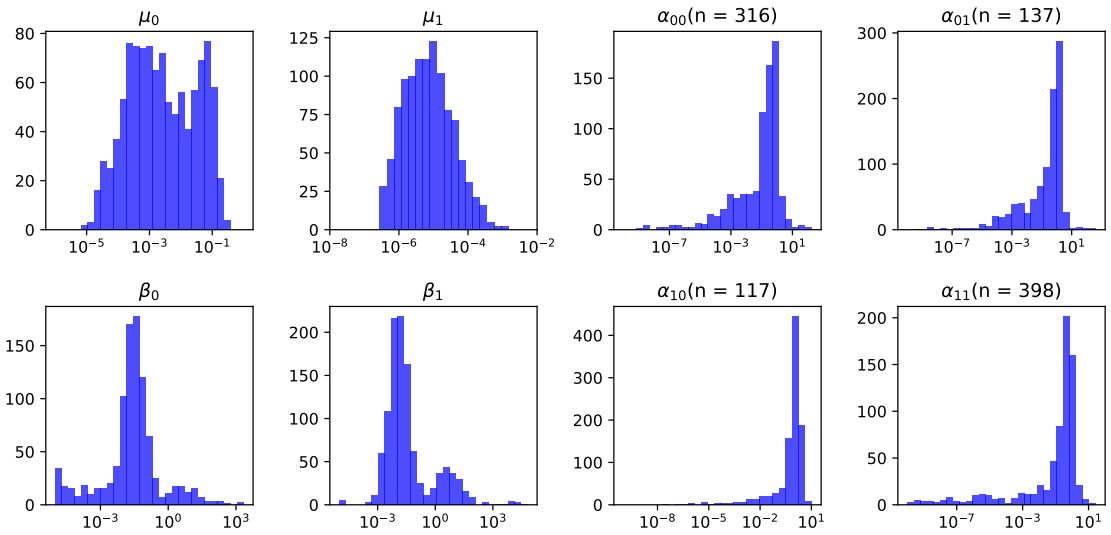


Figure 5.7: Parameter distributions for the fitted Hawkes processes. For the plots containing α, the number of samples where the parameter was zero, and therefore is not included in the histogram, is displayed above the plot.

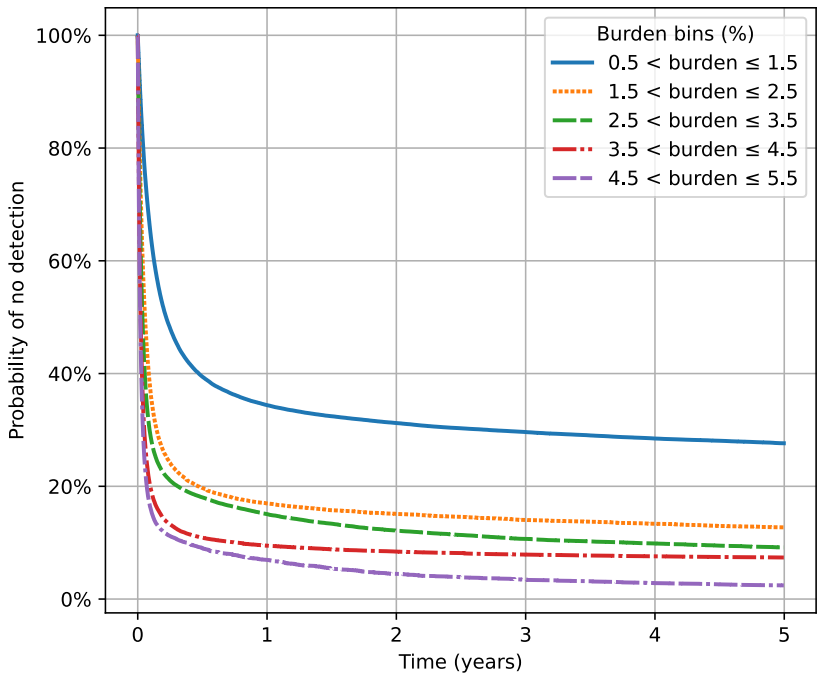


Figure 5.8: Kaplan-Meier survival plot for the Apple Watch algorithm on data simulated using the alternating bivariate Hawkes process.

Table 5.4: Percent not alerted over time for different AF burdens. Based on the simulations using the alternating bivariate Hawkes process.

Years	0.5%–1.5%	1.5%–2.5%	2.5%–3.5%	3.5%–4.5%	4.5%–5.5%
1	34.4	17.0	15.1	9.5	6.9
2	31.2	15.1	12.1	8.4	4.5
3	29.6	14.0	10.7	7.9	3.4
4	28.5	13.4	9.9	7.6	2.8
5	27.6	12.7	9.2	7.4	2.4

5.2.1 Goodness of fit

The goodness of fit was evaluated using the KS distances. The mean and the median KS distances, as well as confidence intervals produced using the standard error of the mean and a bootstrap respectively is shown in Figure 5.9.

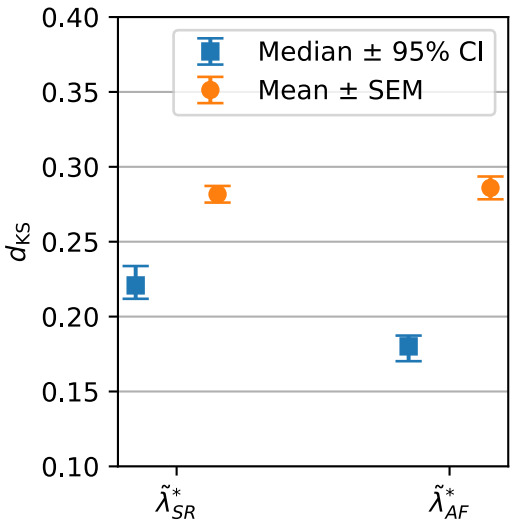


Figure 5.9: Median and mean of the KS distances as well as error bars with a bootstrapped 95% confidence interval for the median and the standard error of the mean.

In the ideal case, the KS distances would be close to zero for all samples. In this analysis, the average KS distance is slightly below 0.3, suggesting that while the alternating bivariate Hawkes process does not perfectly capture all AF patterns, it still provides a reasonably good approximation across the patient population.

Table 5.5: Metrics for the KS-distance for the two point processes with conditional intensities $\tilde{\lambda}^*_{SR}$ and $\tilde{\lambda}^*_{AF}$.

Metric	$\tilde{\lambda}^*_{SR}$	$\tilde{\lambda}^*_{AF}$
Mean	0.28	0.29
SEM	0.006	0.008
Median	0.22	0.18
Median 95% CI	(0.21, 0.23)	(0.17, 0.19)

5.3 Comparison

In order to more easily compare the results for the two methods, they were plotted together in Figure 5.10.

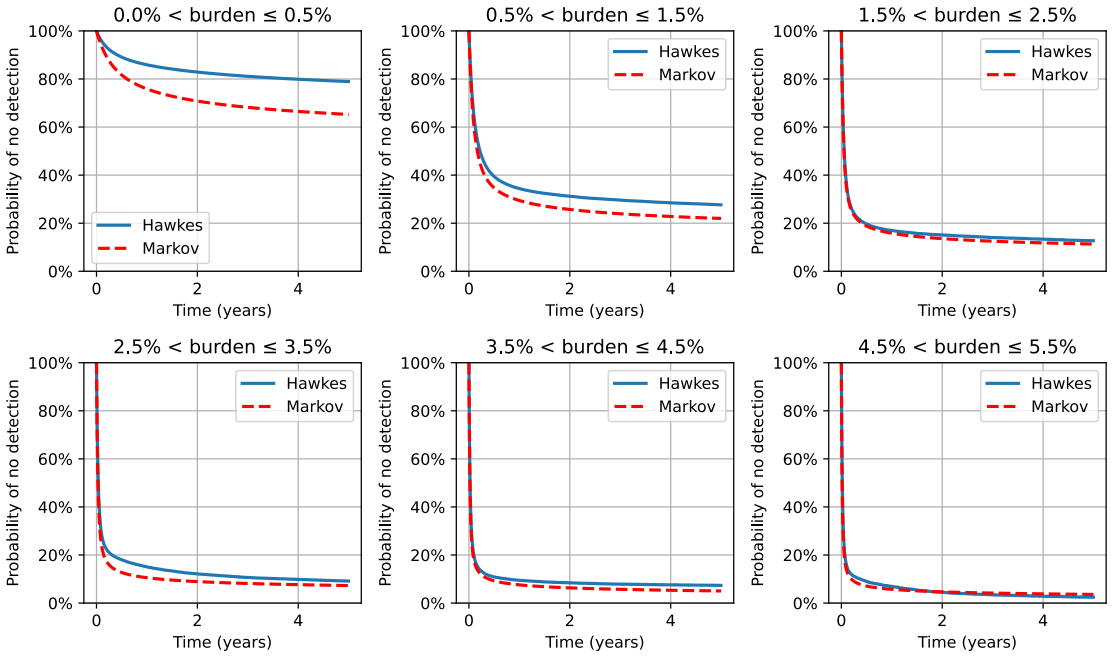
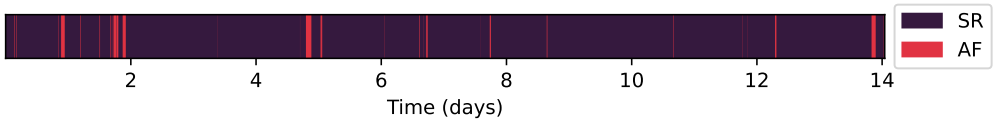
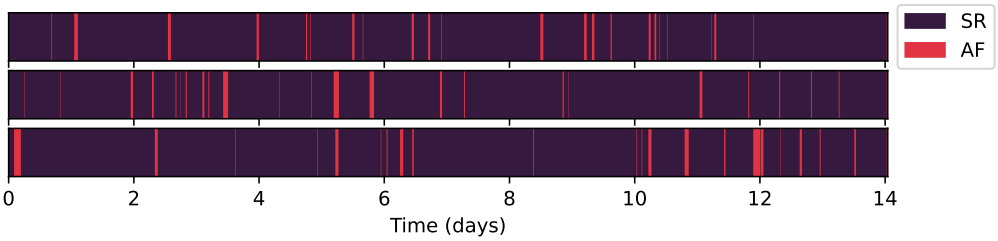


Figure 5.10: Combined results from the simulation using the alternating bivariate Hawkes process and the Markov chain. For the Markov chain, parameters without extrapolation were used.

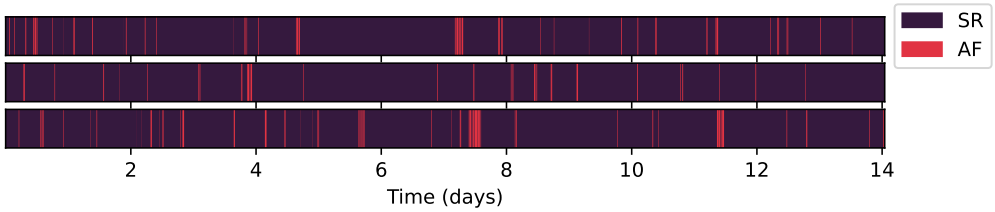
To better illustrate the model's performance, examples of AF patterns and simulations from the two models based on that data are shown below. The first example shows the real AF pattern for a patient with an estimated burden of 3.3%, along with simulations from the Hawkes process and the Markov chain, as shown in Figure 5.11. Similarly, the original signal as well and corresponding simulations for a patient with an estimated burden of 6.1% are presented in Figure 5.12.



(a) Real data

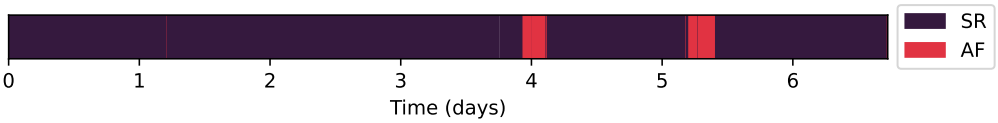


(b) Three different simulated sequences using the two state Markov chain

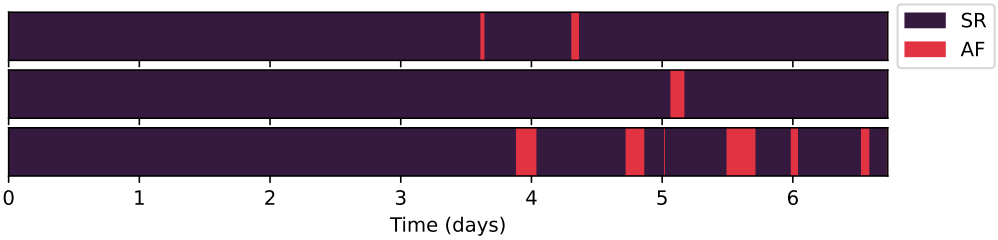


(c) Three simulated sequences using the alternating bivariate Hawkes process.

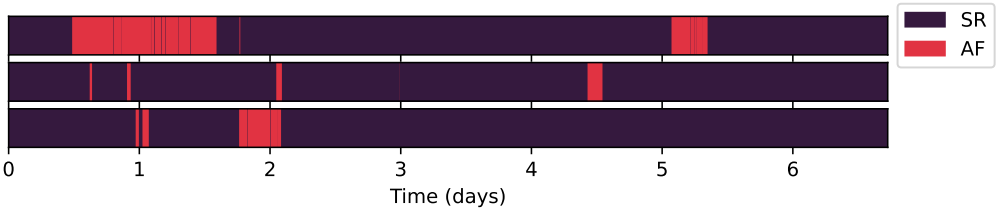
Figure 5.11: Plots of the AF state over time, and simulated sequences of the same length, for a patient with 3.3% AF burden.



(a) Real data



(b) Three different simulated sequences using the two state Markov chain



(c) Three simulated sequences using the alternating bivariate Hawkes process.

Figure 5.12: Plots of the AF state over time, and simulated sequences of the same length, for a patient with 6.1% AF burden.

6

Discussion

The results of this study show that for a relatively large percent of Apple Watch wearers, the PPG-based arrhythmia detection of the watch is likely not enough to reliably warn them if they were to develop AF. About a third of patients with a burden of between 0.5%–1.5% AF will remain undetected after a year of wearing the watch, and more than one in five will not be detected even after wearing the watch for five years. The detection rates did improve with an increasing burden, but difficult patterns seemed to still exist even with a comparatively high burden of 3.5%–4.5%. At this burden the Markov chain approach showed 5% without detections after five years, and the Hawkes process method showed an even higher number at 7.4% undetected after five years.

The results also show that the different patterns of paroxysmal AF have a large impact on the detection rate of the watch, with the majority of all detection occurring at the start of the simulation. This is easiest confirmed by the plot in Figure 5.4, where we can see that the expected times to the first detection can vary vastly, even for cases with the same AF burden. The fact that different AF-patterns are varyingly difficult for the algorithm to detect also explains why about 90% of all the cases were detected in the first year for the 2.5%–3.5% AF burden bin and Markov algorithm, but only 25% of the remaining cases were found in the subsequent four years, a much slower rate for the more difficult patterns.

The two approaches to modeling the AF patterns achieved similar results, which strengthens the drawn conclusions. In general, the Hawkes processes predicted a slightly worse detection rate, with the largest discrepancy being at the lowest burdens. Nonetheless, they both have advantages and disadvantages. The Markov model is the simpler of the two, partly due to using data that already had been discretized into one-minute segments, and partly due to only relying on two parameters and the assumption of memorylessness. Furthermore, while the temporal patterns of AF are generally not memoryless [8], the Markov model remains worth studying due to its numerical stability and the existence of exact solutions for burden and expected detection time. Conversely, the alternating bivariate Hawkes process offers its own strengths. By not discretizing the data beforehand, it retains more detailed information to fit to. Additionally, the increased complexity of the model allows for a more precise fit, since it is able to model excitations that can cause clustering behaviors of AF episodes, a phenomenon that has been shown to exist in previous

studies [8]. However, this additional complexity also comes with drawbacks. The requirement of more transitions in the dataset reduced the size of the available data and may, in particular, reduce the number of patients with very sparse AF episodes, which could affect the results. Furthermore, the much more complex likelihood function required numerical optimization techniques, and was relatively flat with respect to the excitation decay rates, for some patients, making the fitted parameters less reliable.

The overall fit of the alternating bivariate Hawkes process was not as strong as what was found in [8], where the average KS distance was reported to range from 0.14–0.23 and 0.11–0.17 on different datasets for $\tilde{\lambda}^*_{\text{SR}}$ and $\tilde{\lambda}^*_{\text{AF}}$ respectively. In this study, the mean KS distances were higher, at around 0.28 and 0.29. The difference is likely due to differences in the datasets. The previous study used shorter datasets with recording durations ranging from an average of 10 hours to 4 days across different datasets. Furthermore, those datasets were smaller with each containing fewer than 30 recordings. In contrast, our dataset consisted of over 1000 recordings with an average monitoring duration of about 18 days. It is also possible that there were differences in how the data was collected or annotated, as this study used AI annotations.

The results rely on several assumptions. The first assumption being that the Apple Watch would mark a one-minute tachogram as arrhythmic if at least 50% of the recording contained an arrhythmia. Due to the actual algorithm being proprietary, this assumption cannot be verified. However, since the clinical definition of AF requires the episode to be at least 30 seconds long, and since our assumption only requires 30 seconds in total per one-minute tachogram to classify it as arrhythmic, the assumption is likely conservative. Secondly, the Apple Watch relies on the user wearing the watch and being still in order to capture tachograms, while this study assumes that the clock is always worn, and that it is always able to capture an adequate signal. In reality all users will have to take off the watch when charging it, and many will also not wear it at night, or only wear it when leaving home. This could significantly reduce the time that the watch is able to capture tachograms, and any movement would reduce the time available for measurements even further. This assumption therefore likely makes us underestimate the minimum time to alert.

6.1 Conclusion and future prospects

In summary, while the Apple Watch has the ability to detect and alert users of potential AF, it is not always reliable. People experience AF in varying patterns, and the watch's ability to detect the arrhythmia varies significantly depending on the pattern. As a result, people with a relatively low AF burden are at risk of slipping through the cracks of the algorithm and may go undetected by the watch, even after several years of paroxysmal AF. At the same time, these users may develop a false sense of security by wearing a watch on which they have enabled arrhythmia detection, and therefore potentially delay or avoid seeking medical attention. This is concerning since they still have an increased risk of stroke, which could be reduced

with the right treatment.

Future research on the topic could aim to refine the models used to simulate the AF patterns, as well as to extend the models to also simulate realistic activity and wear patterns. Furthermore, it would also be valuable to explore the performance of the algorithms used by other smartwatches, and to research if the algorithms could be improved in regards to their sensitivity for low-burden AF without a substantial increase in false positives.

Bibliography

[1] P. Kirchhof, S. Benussi, D. Kotecha, et al., "2016 esc guidelines for the management of atrial fibrillation developed in collaboration with eacts," European Heart Journal, vol. 37, no. 38, pp. 2893–2962, Aug. 2016, ISSN: 1522-9645. DOI: 10.1093/eurheartj/ehw210.

[2] J. Kornej, C. S. Börschel, E. J. Benjamin, and R. B. Schnabel, "Epidemiology of atrial fibrillation in the 21st century: Novel methods and new insights," Circulation Research, vol. 127, no. 1, pp. 4–20, Jun. 2020, ISSN: 1524-4571. DOI: 10.1161/circresaha.120.316340.

[3] P. Libby, Braunwald's heart disease, 2 vol set, 12th ed., R. O. Bonow, D. L. Mann, G. F. Tomaselli, D. L. Bhatt, S. D. Solomon, and E. Braunwald, Eds. Philadelphia, PA: Elsevier - Health Sciences Division, Nov. 2021.

[4] C. Research, Global smartwatch shipments in 2024, Mar. 2025.

[5] M. V. Perez, K. W. Mahaffey, H. Hedlin, et al., "Large-scale assessment of a smartwatch to identify atrial fibrillation," New England Journal of Medicine, vol. 381, no. 20, pp. 1909–1917, Nov. 2019, ISSN: 1533-4406. DOI: 10.1056/nejmoa1901183.

[6] L. S. Johnson, A. P. Persson, P. Wollmer, S. Juul-Möller, T. Juhlin, and G. Engström, "Irregularity and lack of p waves in short tachycardia episodes predict atrial fibrillation and ischemic stroke," Heart Rhythm, vol. 15, no. 6, pp. 805–811, Jun. 2018, ISSN: 1547-5271. DOI: 10.1016/j.hrthm.2018.02.011.

[7] Apple Inc., "Using apple watch for arrhythmia detection," Apple Inc., Tech. Rep., 2020, White paper.

[8] M. Henriksson, A. Martin-Yebra, M. Butkuviene, et al., "Modeling and estimation of temporal episode patterns in paroxysmal atrial fibrillation," IEEE Transactions on Biomedical Engineering, vol. 68, no. 1, pp. 319–329, Jan. 2021, ISSN: 1558-2531. DOI: 10.1109/tbme.2020.2995563.

[9] P. J. Laub, Y. Lee, and T. Taimre, The elements of Hawkes processes. Cham, Switzerland: Springer Nature, Jan. 2022.

[10] W. F. Kaemmerer, M. S. Rose, and R. Mehra, "Distribution of patients' paroxysmal atrial tachyarrhythmia episodes: Implications for detection of treatment efficacy," Journal of Cardiovascular Electrophysiology, vol. 12, no. 2, pp. 121–130, Feb. 2001, ISSN: 1540-8167. DOI: 10.1046/j.1540-8167.2001.00121.x.

[11] H. Dobrzynski, R. H. Anderson, A. Atkinson, et al., "Structure, function and clinical relevance of the cardiac conduction system, including the atrioventricular ring and outflow tract tissues," Pharmacology & Therapeutics, vol. 139, no. 2, pp. 260–288, Aug. 2013, ISSN: 0163-7258. DOI: 10.1016/j.pharmthera.2013.04.010.

[12] Madhero88, Conduction system of the heart, Licensed under CC BY 3.0, 2010.

[13] R. Langendorf, A. Pick, and L. N. Katz, "Ventricular response in atrial fibrillation: Role of concealed conduction in the av junction," Circulation, vol. 32, no. 1, pp. 69–75, Jul. 1965, ISSN: 1524-4539. DOI: 10.1161/01.cir.32.1.69.

[14] J. Heuser, Scheme of atrial fibrillation (top) and sinus rhythm (bottom), Accessed: 2025-05-05, Dec. 17, 2005.

[15] C. Paludan-Müller, O. B. Vad, N. K. Stampe, et al., "Atrial fibrillation: Age at diagnosis, incident cardiovascular events, and mortality," European Heart Journal, vol. 45, no. 24, pp. 2119–2129, Apr. 2024, ISSN: 1522-9645. DOI: 10.1093/eurheartj/ehae216.

[16] K. B. Kim and H. J. Baek, "Photoplethysmography in wearable devices: A comprehensive review of technological advances, current challenges, and future directions," Electronics, vol. 12, no. 13, p. 2923, Jul. 2023, ISSN: 2079-9292. DOI: 10.3390/electronics12132923.

[17] R. P. Dobrow, Introduction to stochastic processes with R. Nashville, TN: John Wiley & Sons, Feb. 2016.

[18] W.-K. Ching and M. K. Ng, Markov chains (International Series in Operations Research & Management Science), 2006th ed. New York, NY: Springer, Dec. 2005.

[19] S. M. Ross, Stochastic Processes, 2nd ed. Nashville, TN: John Wiley & Sons, Jan. 1995.

[20] D. L. Cohn, Measure theory (Birkhäuser Advanced Texts Basler Lehrbücher), 2nd ed. Cambridge, MA: Springer, Jul. 2013.

[21] Y. Ogata, "On lewis' simulation method for point processes," IEEE Transactions on Information Theory, vol. 27, no. 1, pp. 23–31, 1981.

[22] L. S. Johnson, P. Zadrozniak, G. Jasina, et al., "Artificial intelligence for direct-to-physician reporting of ambulatory electrocardiography," Nature Medicine, vol. 31, no. 3, pp. 925–931, Feb. 2025, ISSN: 1546-170X. DOI: 10.1038/s41591-025-03516-x.

DEPARTMENT OF MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY