



MONIKA BUTKUVIENĖ

**CHARACTERIZATION
OF TEMPORAL
EPISODE PATTERNS
IN PAROXYSMAL
ATRIAL
FIBRILLATION**

DOCTORAL DISSERTATION

K a u n a s
2 0 2 2

KAUNAS UNIVERSITY OF TECHNOLOGY

MONIKA BUTKUVIENĖ

CHARACTERIZATION OF
TEMPORAL EPISODE PATTERNS IN
PAROXYSMAL ATRIAL FIBRILLATION

Doctoral dissertation

Technological Sciences, Electrical and Electronic Engineering (T 001)

Kaunas, 2022

This doctoral thesis was prepared at Kaunas University of Technology, Department of Biomedical Engineering Institute during the period of 2018–2022. The studies were supported by the Research Council of Lithuania.

Scientific Supervisor:

Dr. Andrius PETRĖNAS (Kaunas University of Technology, Technological Sciences, Electrical and Electronic Engineering, T 001)

Edited by: English language editor dr. Armandas Rumšas (Publishing House *Technologija*), Lithuanian language editor Aurelija Gražina Rukšaitė (Publishing House *Technologija*).

Dissertation Defence Board of Electrical and Electronic Engineering Science Field:

Prof. dr. Arminas RAGAUSKAS (Kaunas University of Technology, Technological Sciences, Electrical and Electronic Engineering, T 001) – **chairperson**;

Prof. dr. Tomas KAZAKEVIČIUS (Lithuanian University of Health Sciences, Biomedical Sciences, Medicine, M 001);

Prof. habil. dr. Arūnas LUKOŠEVIČIUS (Kaunas University of Technology, Technological Sciences, Electrical and Electronic Engineering, T 001);

Dr. Michela MASÈ (University of Trento, Italy, Biomedical Sciences, Biophysics, N 011);

Prof. dr. Algimantas VALINEVIČIUS (Kaunas University of Technology, Technological Sciences, Electrical and Electronic Engineering, T 001).

The official defence of the dissertation will be held at 10 a.m. on 2 September, 2022 at the public meeting of Dissertation Defence Board of Electrical and Electronic Engineering Science Field in M7 Hall at the Campus Library of Kaunas University of Technology.

Address: Studentų St. 48, 51367 Kaunas, Lithuania.

Tel. no. (+370) 37 30 00 42; fax. (+370) 37 32 41 44; e-mail doktorantura@ktu.lt

Doctoral dissertation was sent on 2 August, 2022.

The doctoral dissertation is available on the internet at <http://ktu.edu> and at the library of Kaunas University of Technology (K. Donelaičio St. 20, 44239 Kaunas, Lithuania).

KAUNO TECHNOLOGIJOS UNIVERSITETAS

MONIKA BUTKUVIENĖ

PAROKSIZMINIO PRIEŠIRDŽIŲ VIRPĖJIMO
EPIZODŲ PROFILIŲ CHARAKTERIZAVIMAS

Daktaro disertacija

Technologijos mokslai, elektros ir elektronikos inžinerija (T 001)

Kaunas, 2022

Disertacija rengta 2018–2022 metais Kauno technologijos universiteto Biomedicininės inžinerijos institute. Mokslinius tyrimus rėmė Lietuvos mokslo taryba.

Mokslinis vadovas:

Dr. Andrius PETRĖNAS (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija, T 001)

Redagavo: anglų kalbos redaktorius dr. Armandas Rumšas (leidykla „Technologija“), lietuvių kalbos redaktorė Aurelija Gražina Rukšaitė (leidykla „Technologija“).

Elektros ir elektronikos inžinerijos mokslo krypties disertacijos gynimo taryba:

Prof. dr. Arminas RAGAUSKAS (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija, T 001) – **pirmininkas**;

Prof. dr. Tomas KAZAKEVIČIUS (Lietuvos sveikatos mokslų universitetas, biomedicinos mokslai, medicina, M 001);

Prof. habil. dr. Arūnas LUKOŠEVIČIUS (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija, T 001);

Dr. Michela MASÈ (Trento universitetas, Italija, biomedicinos mokslai, biofizika – N 011);

Prof. dr. Algimantas VALINEVIČIUS (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija, T 001).

Disertacija bus ginama viešame Elektros ir elektronikos inžinerijos mokslo krypties disertacijos gynimo tarybos posėdyje 2022 m. rugsėjo 2 d. 10 val. Kauno technologijos universiteto Studentų miestelio bibliotekoje, M7 salėje.

Adresas: Studentų g. 48, 51367 Kaunas, Lietuva.

Tel. (+370) 37 30 00 42; faks. (+370) 37 32 41 44; el. paštas doktorantura@ktu.lt

Disertacija išsiųsta 2022 m. rugpjūčio 2 d.

Su disertacija galima susipažinti interneto svetainėje <http://ktu.edu> ir Kauno technologijos universiteto bibliotekoje (K. Donelaičio g. 20, 44239 Kaunas, Lietuva).

Contents

INTRODUCTION	8
1. OVERVIEW OF APPROACHES TO CHARACTERIZING ATRIAL FIBRILLATION PATTERN	14
1.1. Atrial fibrillation background	14
1.2. Introduction to atrial fibrillation patterns	16
1.2.1. Medical background on atrial fibrillation pattern	17
1.2.2. Technologies for collection of atrial fibrillation pattern	18
1.2.3. Performance evaluation of atrial fibrillation detectors	23
1.2.4. Existing approaches to characterizing atrial fibrillation pattern	26
1.3. Conclusions of the chapter	28
2. DETECTION AND CHARACTERIZATION OF PAROXYSMAL ATRIAL FIBRILLATION PATTERNS	29
2.1. Atrial fibrillation detectors	29
2.1.1. Rhythm-based detector	29
2.1.2. Rhythm- and morphology-based detector	30
2.1.3. Deep learning-based detector	31
2.2. Characterization of atrial fibrillation patterns	32
2.2.1. Distribution-based pattern characterization	32
2.2.2. Model-based pattern characterization	36
2.2.3. Parameter-based pattern characterization	40
2.3. Conclusions of the chapter	42
3. ATRIAL FIBRILLATION PATTERN DATABASES AND PERFORMANCE EVALUATION	44
3.1. Clinical signals	44
3.2. Simulated signals	45
3.3. Database subsets used for investigation	47
3.4. Performance evaluation of atrial fibrillation detectors	49
3.4.1. Annotation comparison approaches	49

3.4.2. Performance evaluation measures	51
3.5. Conclusions of the chapter	52
4. RESULTS	53
4.1. Investigation of atrial fibrillation detectors	53
4.1.1. Analysis of performance measures	53
4.1.2. Analysis of annotation comparison approaches	56
4.1.3. Investigation of detection influencing factors	58
4.1.4. Comparison of detector performance	62
4.2. Characterization of atrial fibrillation patterns	64
4.2.1. Distribution-based pattern characterization	64
4.2.2. Model-based pattern characterization	66
4.2.3. Parameter-based pattern characterization	68
4.3. Analysis of atrial fibrillation patterns	71
4.3.1. Circadian analysis of atrial fibrillation	71
4.3.2. Investigation of atrial fibrillation pattern reconstruction .	73
4.4. Atrial fibrillation pattern relationship with atrial echocardiographic parameters	75
4.5. Conclusions of the chapter	77
5. CONCLUSIONS	79
SUMMARY	80
REFERENCES	105
LIST OF PUBLICATIONS ON THE SUBJECT OF THE DOC- TORAL THESIS	119

List of terms and abbreviations

\mathcal{A}	Aggregation parameter describing the distribution of episodes
Acc	Detection accuracy
Acc_B	Balanced detection accuracy
AF	Atrial fibrillation
AF pattern	Temporal distribution of atrial fibrillation episodes
AFDB	MIT-BIH atrial fibrillation database from <i>PhysioNet</i>
AIC	Akaike information criterion
APB	Atrial premature beats
AUC	Area under the receiver operating characteristic
β	AF burden describing the total time spend in atrial fibrillation
BIC	Bayesian information criterion
CI	Confidence interval
DL	Deep learning
ECG	Electrocardiogram
F_1	F_1 score
FN	False negative
FP	False positive
\mathcal{G}	Gini coefficient describing the difference in episode duration
LA	Left atrial
LTAfDB	Long-term atrial fibrillation database from <i>PhysioNet</i>
Mcc	Matthews correlation coefficient
non-AF	Other rhythm than atrial fibrillation
NPV	Negative predictive value
PDF	Probability density function
PPG	Photoplethysmogram
PPV	Positive predictive value
RMS	Root mean square
ROC	Receiver operating characteristic
RR interval	Time interval between two contractions of the ventricles
Se	Sensitivity
Sp	Specificity
SPAFDB	St. Petersburg atrial fibrillation database
SR	Sinus rhythm
TN	True negative
TP	True positive
β_1	Model-based parameter related to episode clustering
μ	Model-based parameter related to rhythm dominance

INTRODUCTION

Relevance of the research

Atrial fibrillation (AF) is the most common cardiac arrhythmia worldwide. According to 2010 data, in the European Union, there were 8.8 million individuals older than 55 years with diagnosed AF [1]. Unfortunately, AF prevalence has been rising due to the aging of the global population [2], and it is expected to increase more than 2-fold by 2060 [1]. AF is related to the increased risk of various comorbidities. For example, AF patients have a 5-fold increased risk of stroke, 3-fold increased risk of heart failure, and 1.5–3.5-fold increased risk of general mortality [3]. AF-related stroke or heart failure impose a higher mortality rate comparing to either condition alone [4]. AF patients also suffer from an increased number of hospital admissions, i.e., 30% of AF patients have one and 10% have two or more hospital admissions annually. Therefore, 16–20% AF patients suffer from depression and more than 60% of AF patients endure significantly impaired quality of life [3].

AF is a progressive disease often initially manifested by self-terminating paroxysmal AF episodes. Since paroxysmal AF episodes are usually brief, rarely occurring, and asymptomatic [2], AF may be diagnosed when AF has already progressed to a more sustained form. Progression to persistent or permanent AF is associated with an increased incidence of various comorbidities, hospital admissions, and even death [3,5]. It is important to diagnose AF in the initial stage since the success of the treatment depends on what stage AF is diagnosed at.

Recent advancements in sensor technologies allow long-term continuous patient monitoring using various wearable devices, e.g., smartwatches. Long-term monitoring may be useful to understand AF progression at the individual patient level [6]. In addition, long-term monitoring enables characterization of the paroxysmal AF episode pattern, including the analysis of temporal distribution or clustering of AF episodes [7]. The need for AF episode pattern analysis, complementing the commonly used AF burden, has been emphasized in recent clinical guidelines [3]. However, so far, little is known about the role of AF patterns in AF progression and the development of complications. It might be, that the AF pattern is related to the risk of thrombus formation. Since the flow velocity in the left atrial appendage decreases during AF [8], it is assumed that the risk of thrombus formation is higher when episodes are aggregated in time. Therefore, the understanding of AF patterns may have implications on patient-specific therapy management and the prediction of the health outcome (e.g., stroke).

AF patterns can vary considerably [7] with respect to the AF burden, the number of episodes, the episode duration, and the temporal distribution of episodes. However, there is still a lack of methods to characterize the AF pattern. Also, in order to take a further step in AF pattern characterization, it is essential to understand how

well episode patterns can be captured when using different AF detectors (i.e., rhythm-based, rhythm- and morphology-based, or deep learning-based). The recent, rapid progress in AF detector development comes with the challenge of adequately evaluating and comparing performance relative to the published detectors. The use of performance measures should be accompanied by the investigation of the reliability of AF pattern reconstruction. For example, it is unclear how AF detectors perform on different AF patterns, i.e., whether detection performance is the same in a pattern with a few brief episodes and a pattern dominated by long episodes. The currently available studies on AF detection offer very little insight on how well episode patterns are captured; therefore, there is a need to investigate the influence of the AF pattern reconstruction on the pattern characterizing parameters.

Scientific-technological problem and working hypothesis

The current paradigm provides a scientific basis to assume that the temporal AF episode pattern is associated to the risk of thrombus formation. That is, the risk may potentially be higher when AF episodes are aggregated in time since the flow velocity in the left atrial appendage decreases during AF. While information about AF patterns is lacking, the emerging non-invasive technologies for long-term monitoring are expected to fill in this gap in knowledge. However, in order to properly tackle this problem, novel approaches to characterizing the variety of types of patterns are needed.

Scientific-technological problem: By which means can temporal distribution of paroxysmal AF episodes be characterized so that to distinguish different pattern types which are essential to better understanding of pattern relationship with the risk of thrombus formation?

The working hypothesis: Paroxysmal AF patterns can be characterized in terms of the AF duration, temporal distribution and clustering of AF episodes by using distribution-based, model-based, and parameter-based approaches.

Research object

The research is based on the development and investigation of the signal processing algorithms for the characterization of temporal AF episode patterns.

The aim of the research

This doctoral thesis aims to develop and investigate algorithms for the characterization of temporal AF episode patterns.

The objectives of the research

1. To develop a phenomenological model of generating temporal AF episode patterns.
2. To develop algorithms for the characterization of temporal AF episode patterns.
3. To investigate the influence of AF detector properties on AF pattern characterizing parameters.

Scientific novelty

In this doctoral thesis, performance evaluation of AF detectors is considered to shed needed light on the aspects crucial to reliable reconstruction of the AF pattern. The overall strengths and weaknesses of various types of the AF detector (i.e., rhythm-based, rhythm- and morphology-based, and deep learning-based) are demonstrated. Additionally, challenges in AF pattern reconstruction are highlighted, and recommendations are proposed on how to handle AF data and evaluate AF detection performance.

Three different approaches to characterize AF patterns have been proposed and investigated. One of them is based on the statistical distribution analysis which rests on the assumption that episodes are statistically independent. This assumption may be questioned since AF episodes tend to cluster. Also, this approach is less suitable for the characterization of short AF patterns (i.e., day-long) due to the small number of episodes. Another approach is to use parameters from the AF pattern model which is based on an alternating, bivariate Hawkes process. The model-based parameters provide information on AF episode clustering and rhythm dominance (i.e., AF versus non-AF). The final approach is to use various parameters, i.e., AF burden, aggregation, and the Gini coefficient. AF burden is a well-known parameter in AF studies, however, it provides information only about the time spent in AF. Meanwhile, aggregation provides information on the temporal distribution of AF episodes, and the Gini coefficient, a well-known parameter in economics, provides information on the episode duration inequality. Therefore, a combination of model-based and parameter-based characterization offers a solution to analyze different types of patterns.

Practical significance

1. The investigation of the performance of AF detectors and the proposed solutions for characterization of the paroxysmal AF pattern can be used in the following applications:
 - (a) The investigation of AF detectors highlights the remaining challenges in developing AF detectors.

- (b) Recommendations have been proposed on how to handle AF data and characterize AF detection performance.
 - (c) The proposed model for generating a temporal AF pattern is capable of estimating AF pattern characterizing parameters, thus the model can be useful to characterize AF patterns in terms of the AF episode clustering and rhythm dominance.
 - (d) The proposed algorithms to characterize AF patterns can improve the understanding of arrhythmia progression.
 - (e) The proposed algorithms to characterize AF patterns can be useful for understanding the relation between the AF pattern and risk of complications (e.g., risk of thrombus formation, or ischemic stroke).
 - (f) The proposed algorithms to characterize AF patterns can be useful for identifying potential AF triggers.
2. The methods provided in this thesis have been developed in the framework of the project *Methods for long-term unobtrusive monitoring of atrial arrhythmias in post-stroke patients – AFterStroke* funded by the Research Council of Lithuania (S-MIP-17/81), 2017–2019.
 3. Currently, the developed methods are being used in the following project *Wearable technology for personalized identification and management of paroxysmal atrial fibrillation triggers – TriggersAF* funded by the European Regional Development Fund (01.2.2-LMT-K-718-03-0027) under grant agreement with the Research Council of Lithuania, 2020–2023.

Approval of the results

The doctoral thesis relies on two main papers published in the international scientific journals with the impact factor referred in the *Clarivate Analytics Web of Science* database, while, in total, the results have been published in seven scientific papers. The essential results have been presented in three international worldwide recognized conferences: 45th, 47th, and 48th conferences *Computing in Cardiology*.

The research was nominated as a semi-finalist at the Rosanna Degani Young Investigators' Award competition at the 48th conference *Computing in Cardiology* (Brno, Czech Republic, 2021). In 2019, 2020, and 2022, promotion scholarships for academic research, granted by the Research Council of Lithuania, were received. In 2018, 2019, and 2021, the awards of the most active PhD student in the field of Electrical and Electronic Engineering, granted by Kaunas University of Technology, were received.

The statements presented for defense

1. Detection performance differs depending on the approach taken to compare the detector output to annotation and depending on the selected performance measures. To characterize AF episode patterns, the episode-to-episode comparison should be used, while performance should be evaluated in terms of the Matthews correlation coefficient Mcc , rather than accuracy Acc .
2. The reliability of AF pattern reconstruction is affected by the ECG signal properties (i.e., the ECG morphology, the number of atrial premature beats, the noise level) as well as the AF pattern properties (i.e., the AF burden, the episode duration). Depending on the structure of the AF detector, the reliability of AF pattern reconstruction differs, i.e., the deep learning-based detector tends to merge a few consecutive episodes into a single one, while the rhythm-based and the rhythm- and morphology-based detectors tend to split a single episode into a cluster.
3. The temporal AF episode pattern can be characterized by using distribution-based, model-based, and parameter-based approaches. Distribution-based pattern characterization is less suitable for day-long pattern characterization; also it rests on the assumption that episodes are statistically independent. A combination of model-based and parameter-based characterization offers a solution to analyze different types of AF patterns, since it provides complementary information about the AF pattern, i.e., model-based parameters account to episode clustering and rhythm dominance, while parameter-based characterization provides information about the total time spent in AF, the temporal distribution of AF episodes, and the AF episode duration.

Structure of doctoral thesis

The doctoral thesis is organized as follows. Section 1 is devoted to the analysis of the relevant scientific literature with respect to the clinical significance of AF pattern characterization, the currently available technologies suitable to collect long-term AF patterns, and the currently existing approaches used for AF pattern characterization. Section 2 presents different types of AF detectors and three approaches to characterize the AF pattern, i.e., distribution-based, model-based, and parameter-based. Section 3 describes databases used for investigation as well as performance evaluation. Section 4 presents the results obtained from the investigation of AF detectors and the AF pattern characterization. In the same section, the analysis of AF patterns, the reliability of AF pattern reconstruction, and the investigation of the association between the pattern characterizing parameters and clinical measures are proposed. The doctoral thesis is finished with general conclusions provided in Sec. 5.

Parts of Sections 2–4 have been quoted verbatim from the previously published articles: [9, 10, 11, 12, 13].

The thesis consists of 122 pages, 42 figures, 11 tables, and 142 references.

Work done in collaboration

The proposed model for generating the temporal AF pattern and estimating the AF pattern characterizing parameters is developed in collaboration with Lund University (Lund, Sweden). For this purpose, three-week (from February 25 to March 16, 2019) internship in Lund University was performed, during which, the concept of the model was discussed.

The AF detectors used for the investigation were developed by Andrius Petrėnas (the rhythm-based detector and the rhythm- and morphology-based detector) and Andrius Sološenko (the deep learning-based detector).

1. OVERVIEW OF APPROACHES TO CHARACTERIZING ATRIAL FIBRILLATION PATTERN

1.1. Atrial fibrillation background

Atrial fibrillation (AF) is diagnosed from the electrocardiogram (ECG) signal. During AF, the atrial activity becomes irregular and very rapid, i.e., atria may beat at a rate of 300–600 times per minute. Fortunately, the atrioventricular node blocks impulses arriving from the atria, thus only a part of the electrical impulses reach the ventricles. Yet, the activation of the ventricles becomes irregular nevertheless, i.e., the time between two consecutive contractions of ventricles (RR intervals) becomes irregular. Also, P-waves characterizing the normal atrial activity are replaced by continuous rapid f-waves. Irregular rhythm, the absence of P-waves, and the presence of f-waves are the main features used to diagnose AF in the ECG signal (Fig. 1.1). The temporal distribution of AF episodes is called the AF pattern. For example, the AF pattern may consist of several episodes aggregated in a short time interval, thereby resulting in an AF cluster (see Fig. 1.1).

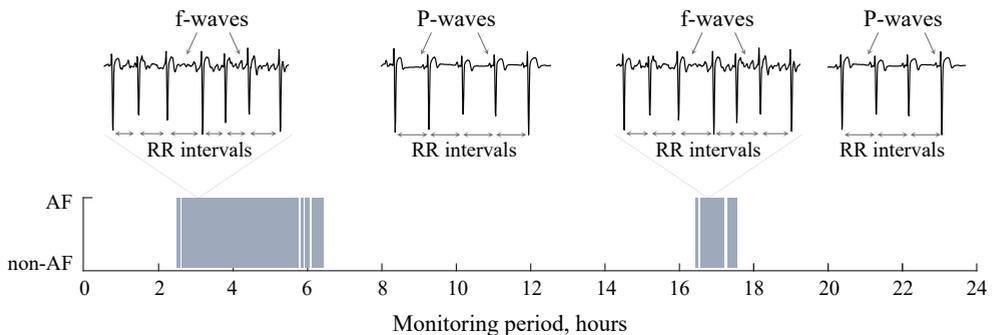


Figure 1.1. Temporal AF episode pattern with two episode clusters. High level in the AF pattern indicates AF, while low level indicates non-AF, e.g., sinus rhythm. During AF, P-waves in ECG signal are replaced by irregular, continuous f-waves, and RR intervals become irregular

According to the guidelines for the management of patients with AF [3], five stages of AF can be distinguished depending on the AF episode duration and the termination of episodes, i.e., whether it is self-terminating or not (see Table 1.1). All patients with AF diagnosed for the first time are assigned to the stage ‘new onset AF’. AF is named ‘paroxysmal’ if the AF episode self-terminates spontaneously in less than 7 days. Paroxysmal AF episodes can last from 30 s to 7 days. Brief AF episodes (<30 s) are not currently considered as paroxysmal, although they are gaining increasing interest in studies analyzing the association with an increased risk of thrombus formation [14, 15]. AF lasting longer than 7 days is classified as persistent or long-standing persistent. Since, at this stage, AF is non-self-terminating, treatment

can be applied to recover the sinus rhythm (SR). However, when AF cannot be terminated, it is assigned to permanent AF. A limitation of the currently used clinical AF classification is that it does not account to the time spent in AF (e.g., AF burden) [3] and poorly reflects the temporal AF episode pattern.

Table 1.1. Classification of AF based on [3]

AF stages	Definition
New onset	Occurrence of the first diagnosed AF episode; episode duration or severity of AF-related symptoms is not important
Paroxysmal	AF is recurrent and self-terminates within 7 days
Persistent	AF fails to self-terminate within 7 days
Permanent	Continuous AF; no further attempts to recover/maintain SR will be undertaken

Usually AF is a progressive disease [16], i.e., a patient with paroxysmal self-terminating AF episodes may progress to persistent AF with non-self-terminating episodes within months [17] or years [18, 19]. On the other hand, other patients may stay in paroxysmal AF for decades or even ‘forever’ [17]. Also, the AF stage may be reversible, i.e., some patients with persistent AF may revert to paroxysmal AF [20,21]. However, the heterogeneity of the AF behavior and its progression or regression is not well understood yet [6].

Long-term continuous monitoring using wearable devices has the potential to change the categorical AF classification in four stages to continuous evaluation accounting to the time spent in AF. This would open new possibilities in AF management [22]. There are a few studies trying to improve characterization of paroxysmal AF episodes. One of the studies proposed the ‘staccato’ (i.e., frequent and short AF episodes) and the ‘legato’ (i.e., infrequent and long AF episodes) paroxysmal AF subtypes [23]; however, the exact threshold values for this classification were not provided. More importantly, the proposed AF subtypes were not analyzed with respect to clinical characteristics. Another study proposed three subtypes of paroxysmal AF based on the duration of the longest AF episode during the monitoring period and the AF burden, i.e., short episodes (<6 hours) and low AF burden (0–0.5%); intermediate episodes (6–12 hours) and intermediate burden (0.5–2.5%); long episodes (>12 hours) and high burden (>2.5%). However, these cut-off values require validation in future studies [7]. The limitation of this study is that only the AF burden is analyzed, while the AF pattern is defined by the number and duration of AF episodes without analysis of the temporal distribution of AF episodes.

Recently, the 4S-AF scheme was proposed in the guidelines for the management of patients with AF [3] which has the potential to improve the single-domain AF classification (the one provided in Table 1.1). The suggested 4S-AF characterization scheme includes four AF-related domains: stroke risk, symptoms, severity of the AF

burden, and the substrate severity (Fig 1.2) [3, 24]. Stroke risk is evaluated by using CHA₂DS₂-VASc score: 1 point for congestive heart failure, 1 point – hypertension, 2 points – ≥ 75 years of age, 1 point – diabetes mellitus, 2 points – previous stroke, 1 point – vascular disease, 1 point – age between 64–74 years of age, 1 point – the gender category (female). The symptom severity is assessed by a physician using the European Heart Rhythm Association (EHRA) symptom classification system. The severity of the AF burden is evaluated based on the time spent in AF and the density of AF episodes per unit of time (e.g., the annual number of episodes). This domain also includes termination of AF episodes (spontaneously terminating or not), which can be an indicator of the progression of AF. The substrate for the AF domain evaluates the complexity of AF pathophysiology. It includes clinical characteristics, such as age, cardiovascular risk factors (e.g., obesity), comorbidities (e.g., stroke), and atrial cardiomyopathy (e.g., atrial enlargement). Even that 4S-AF scheme has the potential to facilitate the management of AF, yet it still needs to be validated [3].

	Stroke risk	Symptoms	Severity of AF burden	Substrate
Description	Truly low risk of stroke (yes/no)	Asymptomatic Moderate Severe	Mode of termination AF duration and density per unit of time	Comorbidities, cardiovascular risk factors Atrial cardiomyopathy
Assessment	CHA ₂ DS ₂ -VASc	EHRA symptom score	AF stage (paroxysmal, persistent, permanent) AF burden (total time in AF, the longest episode, number of episodes, etc.)	Clinical assessment (AF risk or progression scores) Imaging information

Figure 1.2. The 4S-AF scheme for structured characterization of patient with AF. Illustration is adapted from [3, 24]

The understanding of AF and its progression is constantly evolving; however, recommendations for the AF management are still based on AF stages, i.e., paroxysmal, persistent, etc. [24]. Since AF patients classified to the same AF stage may differ in terms of the temporal AF pattern and the AF burden [24], AF management should be personalized. Also, it is hypothesized that the risk profile of the patient should be regularly reevaluated, thus making AF management become dynamic [24]. Therefore, it is essential to investigate capabilities to obtain the long-term AF pattern and to develop algorithms for AF pattern characterization.

1.2. Introduction to atrial fibrillation patterns

Paroxysmal AF has very heterogeneous temporal patterns [7]. For example, the AF burden is 74%, 24%, and 9% in the AF patterns provided in Fig. 1.3 a)–c), respectively. The number of episodes and the episode duration differ as well. The episode can last from a few seconds to several days. Here, it should be noted that the duration of an AF

episode can be less than 30 s; however, based on the current guidelines, only episodes longer than 30 s are considered as paroxysmal [2,3]. Finally, the temporal distribution of AF episodes is different, i.e., episodes can be aggregated in clusters or spread over time.

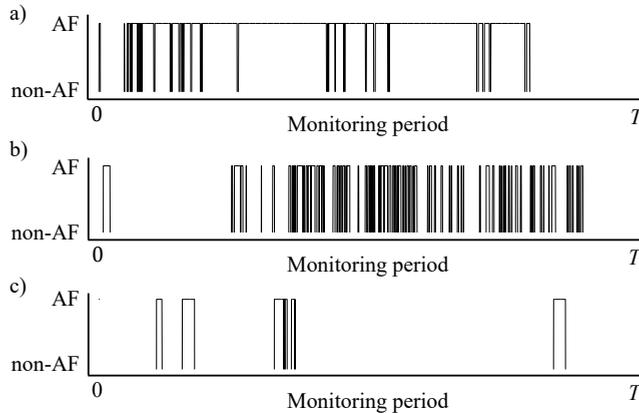


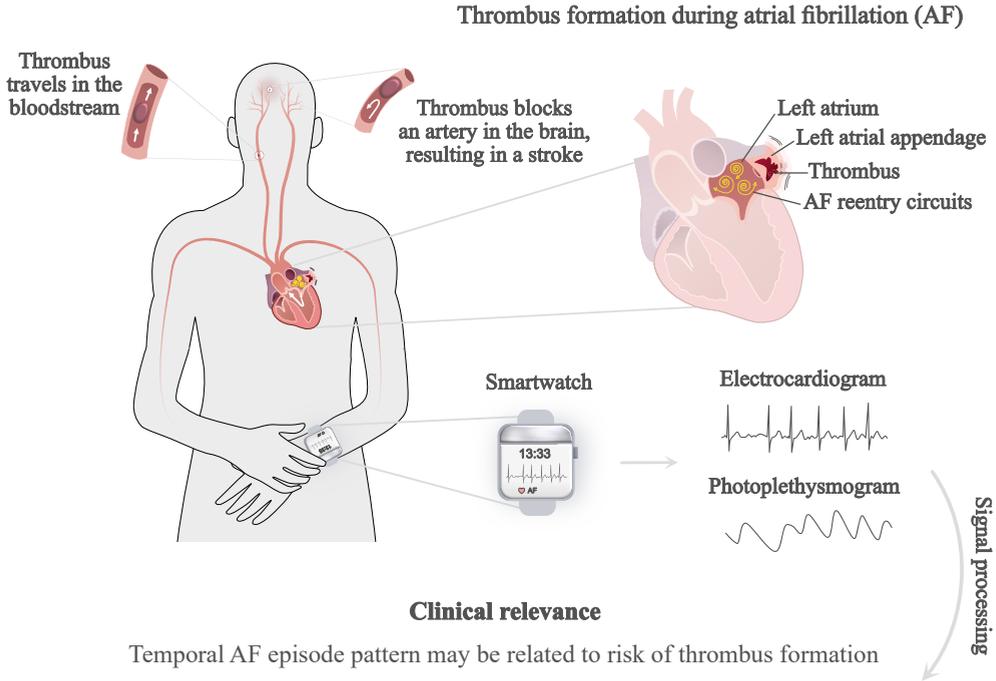
Figure 1.3. Examples of AF patterns: a) AF pattern with high AF burden (74%) and 36 episodes, duration of the longest episode is $>9\,000$ beats; b) AF pattern with low AF burden (24%) and 82 brief episodes; c) AF pattern with low AF burden (9%) and 8 episodes spread in time

1.2.1. Medical background on atrial fibrillation pattern

AF is a major predictor of ischemic stroke [4] occurring when thrombus travels in the bloodstream and blocks an artery in the brain (see Fig. 1.4). During AF, usually, thrombus are formed inside the left atrial appendage [25]. For example, more than 90% of embolic strokes are caused by thrombus formed in the appendage [26]. Reduction of the flow velocity in the left atrial appendage is associated with an increased risk of thrombus formation in patients with paroxysmal AF [8]. During SR, the flow pattern of filling and emptying in the left atrial appendage is regular, while, during AF, the flow pattern becomes ‘sawtooth’ with no identifiable flow waves [8]. Since the flow velocity decreases as AF progresses to longer episodes, the analysis of the temporal episode pattern may provide risk information [25]. It is likely that the risk of thrombus formation is higher when AF episodes are aggregated in time (see Fig. 1.4). However, this question remains unanswered since limited information is available in AF pattern analysis.

The hypothesis that the temporal AF pattern may provide relevant information in AF progression and can be related to various comorbidities may be supported by a few recent studies [7,24]. In [7], it was found that longer AF episodes and a higher AF burden were associated with the AF progression within 1-year, also, those patients had

Risk of thrombus formation



Clinical relevance

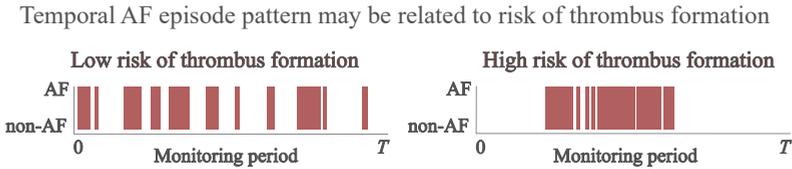


Figure 1.4. AF relation to risk of thrombus formation, collection of long-term continuous AF patterns, and clinical relevance of AF pattern analysis and characterization

more severe comorbidities. In addition, the recently proposed 4S-AF scheme suggests that long and very frequent AF episodes are related to a higher risk comparing to short and infrequent episodes [24]. Fortunately, long-term continuous monitoring allows to obtain the temporal AF pattern, and it enables an improved characterization of AF, which may help to understand the differences in AF progression and the AF pattern relation to clinical outcomes [7].

1.2.2. Technologies for collection of atrial fibrillation pattern

The gold standard for AF detection is the standard 12-lead ECG or Holter monitoring. However, 12-lead ECG recording is not suitable for AF pattern collection since recordings last as little as several seconds. Nevertheless, rapid advancement of sensor technologies opens an opportunity to collect long-term AF patterns. For example, wearable technologies are becoming increasingly popular and convenient to use in daily living (Fig. 1.5, non-invasive devices). In this section, ECG-based devices as

well as other modalities for AF screening are discussed according to their suitability for AF pattern collection. It should be noted that technologies with intermittent monitoring, such as hand-held ECG devices, smartphone camera, etc., are not included in the analysis since it cannot be used to obtain a continuous long-term AF pattern.

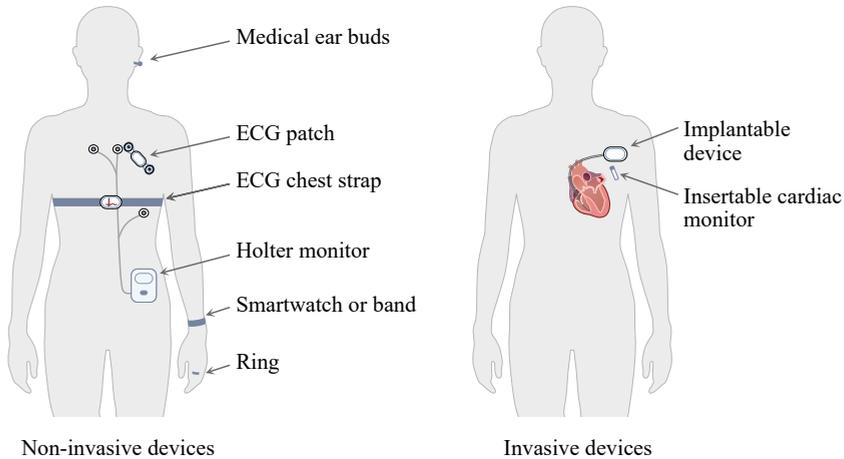


Figure 1.5. Non-invasive and invasive technologies for long-term monitoring of AF pattern. Note, the positioning/placement of electrodes and devices is only for visualization purposes, thus they may be incorrect

Non-invasive ECG-based devices. The Holter monitor is widely used for ambulatory monitoring since it is capable of recording ECG continuously for 1–2 days. Usually, 3-lead ECGs are acquired. After monitoring, recordings are analyzed offline by using some dedicated software. That is, there is no indication about rhythm abnormality during monitoring. Since the Holter monitor provides continuous recording, this device is suitable for collecting AF patterns and can be considered as the gold standard (reference). However, adhesive electrodes and connecting wires are uncomfortable for long-term monitoring (see Fig. 1.5). This is one of the main reasons leading to the termination of monitoring. Therefore, monitoring by using Holter monitors is limited in time.

As an alternative to the Holter monitor, there are several smaller devices which are more patient-friendly [27]. For example, *Bittium Faros* (Bittium, Finland) can be used to record single-lead or 3-lead ECG depending on the series of the device. In addition, the physical activity (accelerometer) and the temperature can be monitored as well. The device is compact and convenient to wear, and it allows monitoring the patient for up to 7 days. As in the case of the Holter monitor, there is no embedded algorithm for rhythm detection. Raw data is saved in an EDF file (the European Data Format) and can be further processed offline, i.e., to detect AF episodes.

The ECG patch is an ECG-based modality which is attached to the body directly without additional connecting wires (see Fig. 1.5). The most popular in cardiovascular clinical trials [22] is the *Zio-Patch* (*iRhythm Technologies*, US) allowing to monitor continuous single-lead ECG up to 14 days. The accuracy of the *Zio-Patch* is comparable with the accuracy of the implanted pacemaker [28]. Another example is *BodyGuardian* (*Preventice Solutions*, US) allowing to record continuous single-lead ECG for up to 24 hours. The data is transferred to the smartphone via Bluetooth. There are also a lot of other patches suitable to acquire ECG, e.g., *Nuvant MCT* (*Corventis*, US), *BardyDx CAM* (*Bardy Dx*, US), *BioTel Heart* (*BioTelemetry*, US), *MediBioSense MBS HealthStream* (*MediBioSense*, UK) [22]. All of them have been approved by the U.S. Food and Drug Administration. ECG patches are suitable for AF pattern collection, while the monitoring time depends on a specific ECG patch.

ECG patches are attached to the body by using adhesive electrodes, which may cause skin irritation. As an alternative, ECG chest straps (wearable belts) can be used instead. For example, *Polar Electro OY* company (Finland) is offering the *Polar* heart rate sensor which can be attached to the body by using chest strap. The battery can last for 400 hours. This model is more popular during physical activity tracking; however, it can be used to obtain AF patterns as well.

To conclude, ECG patches can be used for continuous monitoring of the AF pattern (Table 1.2). However, patient monitoring is limited due to long-term inconvenience [22]. Also, none of them has an embedded algorithm for rhythm detection so far, thus AF episodes should be detected offline by using AF detectors.

Non-invasive PPG-based devices. The modern sensor technology has helped form a new paradigm of the long-term AF monitoring relying on the analysis of the photoplethysmogram (PPG) signal. In the PPG signal, AF is detected based on the irregularity of pulse-to-pulse intervals and PPG morphology [30]. However, AF detection in the PPG signal can be challenging due to noise and artifacts. Also, PPG application in AF diagnosis is limited since there are no guidelines on the interpretation of the PPG signal so far. Therefore, even if AF is already detected by using PPG-based automatic detectors, additional ECG recording (e.g., a 12-lead ECG, Holter monitoring, etc.) has to be obtained to confirm the diagnosis.

Smartwatches are becoming increasingly popular in AF studies [22,31]. A comparison of the *Apple Watch 3* (*Apple*, USA) and the *Fitbit Charge 2* (*Fitbit*, USA) showed that these smartwatches are capable to monitor the heart rate during daily living (e.g., during sitting, walking, or running). The accuracy of the *Apple Watch 3* and the *Fitbit Charge 2* were compared with the gold standard ECG, the mean agreement was 95% and 91%, respectively [22]. However, the PPG signal from wrist-worn devices during physical activity can be very noisy. In order to ensure accuracy during physical activity, it is recommended to use chest-strap monitors [22].

Table 1.2. Comparison of non-invasive and invasive technologies for long-term AF monitoring and their suitability for AF pattern collection

Device type	Accuracy*	Suitability for AF pattern collection
Non-invasive ECG-based devices		
Holter monitor	Reference	Can be used to collect patterns lasting 1–2 days
Single-lead ECG (patches or chest strap)	Sensitivity 94–98% Specificity 76–95%	Can be used to collect continuous patterns, monitoring time depends on specific model (1–2 week)
Non-invasive PPG-based devices		
Smartwatch or band	Sensitivity 97–99% Specificity 83–94%	Well-suited for continuous AF pattern monitoring in patient-friendly way
Medical ear buds		Not investigated in AF studies
Ring		Not investigated in AF studies
Invasive devices		
Insertable cardiac monitor	Sensitivity 88–95%**	Can only be used when it is enough to have information about the onset and end of AF episodes
Implantable device	Reference for AF burden	Well-suited for continuous AF pattern monitoring in a particular patient group

*Accuracy is based on recent guidelines for AF management [3]

** Accuracy is based on [29]

It is important to highlight that some smartwatches have embedded electrodes to record single-lead short-term ECGs. This novelty is very important for confirmation purposes, i.e., when the embedded algorithm detects an AF episode in the PPG signal. A short-term ECG can be recorded by placing a finger on the embedded electrode on the watch (e.g., on the crown as in *Apple Watch*), while the back of the watch contacting with the inner part of the wrist serves as the other electrode [22]. Recently, a wrist-worn device capable of recording continuous PPG and short-term 6-lead ECG for rhythm confirmation was presented [32]. The ECG can be acquired without additional wires or electrodes attached to the body since the device has three embedded electrodes: one in the inner part and two on the outer part of the device. To record ECG, one of the electrodes embedded in the outer part of the device should be touched by the finger, while another one should be held on the left upper abdomen under the rib cage. In this way, lead I and lead II are recorded, while others (lead III, aVR, aVL, and aVF) are calculated from the recorded leads. However, only the *Apple* smartwatch has the U.S. Food and Drug Administration clearance for its ECG tracking functionality [31].

Similar to smartwatches, various wrist or arm bands are also commercially available and can be used to record the PPG signal. For example, *Polar* has several series of smartwatches, and also produces an arm-band (*Polar Verity Sense*) which is more convenient to use comparing to smartwatches, especially during physical activ-

ity. However, *Polar* devices, especially the arm-band, are commonly used for monitoring physical activity. Another example is the *Biostrap* wearable device (*Biostrap*, Los Angeles), i.e., a wristband strap allowing to monitor the PPG signal. It allows exporting raw or processed PPG, gyroscope, and accelerometer data for further processing. However, there is no embedded AF detector. In the official web-page, it is mentioned that *Biostrap* was validated and compared with the gold standard, and the device was used in 22 clinical studies. However, the official web-page do not provide reference to the publications.

Other modern PPG-based modalities are medical ear buds or rings. For example, *Cosinuss° two* (*Cosinuss GmbH*, Germany) is an in-ear wearable sensor allowing to obtain continuous PPG and acceleration signals. The device automatically provides the heart rate, the RR intervals, the signal quality, the body temperature, the blood oxygen saturation, and the perfusion index. Another modality is the smart earring, e.g., *Joule Earrings* (*BioSensive Technologies*, India) which provides the heart rate and physical activity information. Lastly, the heart rate as well as the physical activity and sleep information can be acquired by using rings, e.g., *Oura Ring* (*Oura*, Finland). So far, none of these modalities has been used in AF studies. However, they have potential to acquire long-term AF patterns in a convenient and unobtrusive way, and the main question is the signal quality.

To conclude, wearable devices can provide continuous monitoring of patients in a convenient way (Table 1.2). However, most of the wearable devices activate continuous PPG monitoring only during physical activity. In order to save the battery life, PPG monitoring during rest or sleep is not continuous, and PPG monitoring occurs at constant time intervals [22]. In order to obtain the AF pattern, continuous PPG monitoring should be ensured.

Invasive devices. Long-term monitoring of AF in a convenient way can be ensured by using invasive devices, such as insertable cardiac monitors or implantable devices (e.g., pacemakers or cardioverters-defibrillators), see Fig. 1.5. However, it requires surgical intervention.

Insertable cardiac monitors are small, very thin, light, and wireless devices which are inserted under the skin in the anterior chest. This type of devices contains two built-in electrodes and is suitable for recording a single-lead ECG. However, due to its small memory, continuous ECG cannot be saved. The device has a looping memory, which means that ECG is recorded continuously; however, after 10 min, the recorded interval is deleted. If the embedded algorithm detects rhythm abnormality or if the user activates the device, the ECG interval is saved. There are numerous insertable cardiac monitors or loop recorders; however, only a few are suited for AF monitoring, e.g., *Reveal XT* (*Medtronic*, US). The suitability for AF pattern collection is limited since only the information about the onset and end of AF episodes is

available. This means that the embedded algorithm should ensure that all episodes are detected. On the other hand, by increasing the sensitivity, the number of false alarms increases as well, therefore, AF episodes have to be manually reviewed.

Implanted devices (i.e., pacemakers or cardioverters-defibrillators) are used to regulate the heart rhythm, however, they can also be used to acquire the AF pattern. One of the examples of pacemakers can be the devices from the *Biotronik* company (Germany). Implanted devices can record the intra-atrial ECG signal which is obtained directly from the heart by using implanted lead in the atria. It provides continuous monitoring of the patient by providing the date and time of an AF episode onset and the AF episode duration [33]. It should be noted that obtained intra-atrial ECG differs from the surface ECG, thus it additionally allows analyzing AF episodes with respect to the atrial rate. Newer technologies of single-ventricular-chamber devices used for the prevention of the sudden cardiac death feature ventricular leads; therefore it allows AF detection based on RR intervals [34]. However, these devices offer lower AF detection performance comparing to the implanted devices with the atrial lead [35].

Implanted devices can ensure the long-term monitoring of the AF pattern for a particular patient group, i.e., the ones who already have the implanted device (e.g., for bradycardia or sudden death prevention). However, device-detected atrial high-rate episodes, defined as atrial tachycardias above a predefined atrial rate threshold (e.g., 190 beats per minute), should be interpreted with caution since it can be assigned to AF [33]. Therefore, manual annotation is needed to confirm the diagnosis of atrial tachycardia or AF. Another limitation is that very brief or slow AF episodes can be missed. Diagnostic accuracy becomes reliable when an AF episode is longer than 5–6 min [33]. However, in this case, brief AF episodes are missed, which are important in the paroxysmal AF pattern collection.

1.2.3. Performance evaluation of atrial fibrillation detectors

The performance of AF detectors is very important in order to capture AF patterns properly. The recent interest in deep learning (DL) has led to an avalanche of AF detectors, e.g., [36]–[52]. As a consequence, the problem of how to evaluate and compare performance between different detectors, whether based on DL or on expert-crafted features (e.g., rhythm or morphology features), has been brought into focus. To outline a framework for evaluation that not only ensures a fair comparison but also goes beyond reporting the overall performance measures is therefore essential.

While public databases facilitate the comparison of detector performance, conclusions should be made with caution for a number of reasons. Rather than using the entire database, certain detectors have been evaluated on a subset, e.g., by excluding poor-quality signal segments, or by omitting segments for the purpose of balancing the

datasets. Also, depending on the approach taken to comparing the detector output to database annotations, i.e., beat-to-beat [53]–[59], segment-to-segment [51, 57], [60]–[64], or episode-to-episode comparison [59, 63, 65, 66], the performance can differ considerably. Although only results computed by using the same approach must be compared, this is not always the case.

Expression of performance in terms of statistical measures, e.g., sensitivity and specificity, is common practice. However, the use of performance measures should be accompanied by results uncovering the detector properties. For example, by investigating what signal scenarios cause frequent false alarms, weaknesses in the detector design can be addressed more efficiently. Such understanding can be gained by means of simulated ECG signals which, in contrast to real signals, offer control of the principal quantities, such as the type and level of noise, the rate of atrial premature beats (APBs), and the AF burden [67, 68]. The interest in brief AF episodes (<30 s) and their association with the future risk of stroke [14, 15] motivates the simulation of signals with varying episode lengths to enrich the understanding of performance.

A meaningful comparison of detection performance requires that the datasets for training and testing should be handled in the same way across studies. Firstly, all records of the database should be used, meaning that no records should be excluded due to poor signal quality or as a means to obtain balanced datasets [69]. Secondly, testing should be done on a database different from the one used for training so that the performance can be established on unseen data. Thirdly, the same patient should not appear in both the training and the test datasets. Though not critical to a comparison, it is highly desirable to provide insight on what particular problem situations cause the performance to deteriorate, e.g., by presenting examples of motion artefacts and non-AF arrhythmias.

Tables 1.3 and 1.4 show to what extent DL-based and expert-crafted detectors, respectively, comply with the above-mentioned requirements; the listed detectors were all evaluated on the MIT–BIH Atrial Fibrillation Database (AFDB). It is obvious that a comparison of performance can be highly misleading as data handling differs among the studies. Only 7 out of 14 (50%) of the DL-based detectors were tested on all records of AFDB, in comparison to 10 out of 13 (77%) of the expert-crafted detectors under analysis; the records excluded in [48, 61, 64] were motivated by incorrect annotations. Similarly, as few as 4 (29%) of the DL-based detectors used different patients in the training and the test sets, whereas 10 (77%) of the expert-crafted detectors did so. The effect of using different patients in the training and the test sets was illustrated by a recent study which reported excellent performance of the proposed DL model for AF detection ($Se = 99.1\%$, $Sp = 98.5\%$) when the same patient appeared in both sets [52]; however, when the sets contained different patients, the performance was mediocre ($Se = 90.5\%$, $Sp = 79.7\%$). Concerning testing on a database other than that used for training, only 1 DL-based detector (7%) complied with this re-

Table 1.3. Comparison of DL-based AF detectors

Authors	All records of AFDB used	Different patients in training & test sets	Testing on other databases	Plots of problem signals
Xia et al., 2018 [50]	No	No	No	No
Faust et al., 2018 [49]	Yes	Yes	No	No
He et al., 2018 [48]	No	Yes	No	No
Andersen et al., 2019 [46]	Yes	No	Yes	Yes
Lai et al., 2019 [45]	Yes	No	No	No
Dang et al., 2019 [44]	Yes	No	No	No
Fujita et al., 2019 [43]	No	No	No	No
Wang, 2020 [42]	Yes	No	No	No
Jin et al., 2020 [41]	No	Yes	No	No
Huang et al., 2020 [40]	No	No	No	No
Zhang et al., 2020 [39]	Yes	No	No	No
Ghosh et al., 2020 [38]	No	No	No	No
Shi et al., 2020 [37]	No	No	No	No
Mousavi et al., 2020 [52]	Yes	Yes	No	No
Percentage ‘Yes’	50%	29%	7%	7%

Table 1.4. Comparison of expert-crafted AF detectors

Authors	All records of AFDB used	Different patients in training & test sets	Testing on other databases	Plots of problem signals
Dash et al., 2009 [64]	No	Yes	Yes	Yes
Babaeizadeh et al., 2009 [63]	Yes	Yes	Yes	Yes
Lake et al., 2011 [70]	Yes	Yes	Yes	Yes
Lian et al., 2011 [62]	Yes	Yes	Yes	Yes
Huang et al., 2011 [71]	Yes	Yes	Yes	No
Shouldice et al., 2012 [72]	Yes	Yes	Yes	No
Carvalho et al., 2012 [73]	Yes	No	No	No
Lee et al., 2013 [61]	No	Yes	Yes	Yes
Zhou et al., 2014 [74]	Yes	Yes	Yes	No
Ródenas et al., 2015 [55]	No	No	No	No
Asgari et al., 2015 [60]	Yes	No	No	Yes
Petrénas et al., 2015 [66]	Yes	Yes	Yes	Yes
Zhou et al., 2015 [75]	Yes	Yes	Yes	No
Percentage ‘Yes’	77%	77%	77%	54%

quirement, whereas 10 (77%) of the expert-crafted detectors did so. Interestingly, the performance of that particular DL-based detector was found to drop dramatically when tested on another database (Se remained unchanged at 99.0% while Sp dropped from 97.0% to 86.0% [46]), thus offering a possible reason why different training and test databases have been shunned in the literature. Concerning the plots of problem

signals, again only 1 study (7%) on DL-based detection provided such information, whereas 7 (54%) of the studies on expert-crafted detectors.

Lastly, it is important to highlight that the currently available studies on AF detection offer very little insight on how well the episode patterns are captured. For example, it is unclear whether an AF pattern with a few brief episodes is reconstructed as reliably as a pattern dominated by long episodes. Therefore, there is a need to investigate the reliability of AF pattern reconstruction and the influence of the missed and falsely detected episodes on the pattern characterizing parameters.

This subsection has been quoted verbatim from the previously published article: [9].

1.2.4. Existing approaches to characterizing atrial fibrillation pattern

AF burden, being one of the most popular parameters in AF studies, is defined as the part of the time the patient spends in AF during the monitoring period (e.g., 1 day) [3]. Numerous studies analyzed the relation between the AF burden and an increased risk of thrombus formation or stroke (see Table 1.5); also, there is a recently published systematic review of these studies [76]. Usually, the relation between the AF burden and the risk of thrombus formation is considered as linear, i.e., larger AF burden values are related to a higher risk. However, the threshold above which the risk of thrombus formation increases differs among studies [33]. It is unclear whether the risk of thrombus formation is continuous, or maybe a threshold could be defined at which the risk increases significantly [24]. The guidelines for AF management claim that there is not enough evidence on the relation between the AF burden and health outcomes (i.e., the risk of stroke) [3]. However, advancements in wearable devices allow monitoring the AF burden for a prolonged time, therefore, it might be that this knowledge gap will get filled in the near future [24].

So far, little is known about the role of temporal AF episode patterns in AF progression and the development of complications [11, 91]. The need for episode pattern analysis, complementing the AF burden, has been emphasized in recent clinical guidelines, e.g., by determining the density of episodes per unit of time [3, 24]. Further investigation of the AF pattern is mandatory since the temporal distribution of AF episodes may be different even if the AF burden is the same [33]. Also, discordant results regarding the threshold of the AF burden already presented in Table 1.5 may suggest that the AF pattern itself is important as well. However, there is a lack of approaches to evaluate and characterize the AF pattern.

The problem of how to characterize AF episode patterns received certain attention around the turn of the millennium. At the time, emphasis was put on univariate statistical analysis of either the intervals between consecutive AF episodes (interepisode intervals) [92, 93, 94], or on the intervals between the onsets of consecutive

Table 1.5. Comparison of studies analyzing the relationship between AF burden and risk of stroke

Study	AF burden	Relation to risk of stroke
MOSTT, 2003 [77]	≥ 5 min episode	> 2-fold increased risk of stroke
Capucci et al., 2005 [78]	≥ 24 h episode	3.1-fold increased risk of thrombus formation
Botto et al., 2008 [79]	> 5 min episode	Increased risk of stroke
TRENDS, 2009 [80]	≥ 5.5 h daily burden	2-fold increased risk of thrombus formation
Ziegler et al., 2010 [81]	≥ 5 min daily burden	Increased risk of stroke
Boriani et al., 2011 [82]	> 5 min episode	Increased risk of stroke
Shanmugam et al., 2012 [83]	≥ 3.8 h daily burden	9-fold increased risk of thrombus formation
ASSERT, 2012 [84]	> 6 min episode	2.5-fold increased risk of thrombus formation
SOS AF, 2013 [85]	≥ 1 h daily burden	2-fold increased risk of stroke
IMPACT, 2015 [86]	> 5.5 h daily burden	3-fold increased risk of thrombus formation
Turakhia et al., 2015 [87]	5.5 h daily burden	4-fold increased risk of stroke
ASSERT, 2017 [88]	> 24 h episode	3.2-fold increased risk of thrombus formation
Kaplan et al., [89]	> 6 min episode	Increased risk of stroke
Chu et al., 2020 [90]	> 6 min episode	6.8-fold increased risk of thrombus formation

AF episodes (inter-detection intervals) [95]. While it was speculated that information on episode patterns can be useful to predict the outcome [96], e.g., by relating the degree of episode clustering to antiarrhythmic therapy [93], the clinical significance was never investigated.

Most of the ‘millennial studies’ were based on the series of RR intervals produced by the AF detector in an implantable device. This approach offered continuous operation for a year or more with the potential to characterize the AF progression. When using a small dataset, the initial results suggested that inter-episode intervals could be described by a homogenous Poisson model, thus implying that inter-episode intervals follow an exponential probability density function (PDF) [92]. This PDF model was, however, later discarded in favor of the Weibull PDF [93,94] or the power law PDF [95] as the latter two PDFs were found to be more adequate for the modeling of inter-episode intervals. The above-mentioned PDF-based approach to characterizing the AF inter-episode or inter-detection intervals rests on the assumption that episodes are statistically independent – which is an assumption that may be questioned since AF episodes tend to cluster [93].

An important disadvantage of the PDF-based approach is that it is less effective for short monitoring periods due to an insufficient number of AF episodes. This restriction does not apply to the parameter called ‘AF density’ which depends on the temporal dispersion of AF episodes over the monitoring period [97,98]. AF density is one of the very few parameters proposed for characterizing the aggregation of the AF burden in patients subjected to year-long monitoring using an implantable device [98]. The parameter is defined on the interval $[0, 1]$ where a value close to 0 indicates that the AF burden is evenly distributed on the day-to-day basis throughout the monitored

period, and a value close to 1 shows that the AF burden is confined to an interval much shorter than the monitored period. However, it is still unclear whether different types of patterns can be distinguished, and the parameter, as well as in the PDF-based approach, was not investigated in relation to the patient outcome.

Parts of this subsection have been quoted verbatim from the previously published articles: [10, 13].

1.3. Conclusions of the chapter

1. AF is classified in five stages, i.e., new onset, paroxysmal, persistent, long-standing persistent, and permanent. However, the currently used clinical classification does not correspond to the AF burden and poorly reflects the temporal AF episode pattern. Long-term continuous monitoring offers potential to change the categorical AF classification to continuous evaluation. This would open new possibilities in AF management.
2. During AF, thrombus formation most often occurs in the left atrial appendage. Reduction of the flow velocity in the left atrial appendage is associated with an increased risk of thrombus formation. Since the flow velocity decreases as AF progresses, analysis of the temporal episode pattern may provide risk-related information.
3. Recent advancements in sensor technologies allow long-term continuous patient monitoring in a convenient and unobtrusive way by using various wearable devices, e.g., smartwatches, bands, chest straps, medical ear buds, or rings.
4. The currently available studies on AF detection offer very little insight on how well episode patterns are captured. To take a further step in the AF pattern characterization, it is essential to fill this knowledge gap by investigating the reliability of AF pattern reconstruction when using different automatic AF detectors.
5. AF patterns can be widely different with respect to the AF burden, the number of episodes, the episode duration, and the temporal distribution of episodes. However, little is known about the role of temporal AF episode patterns in the AF progression and development of complications.
6. Long-term continuous patient monitoring enables the characterization of paroxysmal AF episode patterns, however, there is a lack of approaches aiming to characterize an AF pattern.
7. There is a need to investigate the influence of missed and falsely detected episodes on the pattern characterizing parameters.

2. DETECTION AND CHARACTERIZATION OF PAROXYSMAL ATRIAL FIBRILLATION PATTERNS

2.1. Atrial fibrillation detectors

Three types of AF detectors can be discerned in the literature, namely, those exploring the rhythm only, both the rhythm and the morphology, and those using segments of ECG samples as the input. The first two types are expert-crafted detectors, and they require prior QRS detection, here accomplished by the wavelet-based detector described in [99]. Meanwhile, the third type is a deep learning-based (DL-based) detector which does not require prior QRS detection or feature extraction. In the following subsections, one representative of each detector type is considered.

Parts of Sec. 2.1 have been quoted verbatim from the previously published article: [9].

2.1.1. Rhythm-based detector

Rhythm-based AF detection relies on the observation that AF episodes are manifested by irregular RR intervals which are often associated with the increased heart rate. The detector includes blocks for ectopic beat filtering, bigeminal suppression, the characterization of RR interval irregularity, and signal fusion (see Fig. 2.1) [56]. In the part of ectopic beat filtering, a 3-point median filter is used in order to eliminate ectopic beats which can be misclassified as AF when short time RR interval series are under analysis. The second part of preprocessing is the estimation of the RR trend accomplished by using the exponential averager. It is known that the heart rate usually increases during AF, therefore an estimate of the mean RR interval is used as a feature in the AF detection. During the AF detection stage, RR irregularity is estimated in a sliding detection window, and the number of all pairwise RR interval combinations differing more than the predefined threshold is determined. Also, additional measures of RR irregularity are used [56] in order to avoid false detections due to bigeminy. At last, signal fusion is used to produce a detection output (i.e., a value between 0 and 1) reflecting the likelihood of AF being present in the sliding detection window.

The detector is designed to detect brief AF episodes, and it requires only RR interval series as an input. An important feature is that this detector can be implemented with just a few arithmetic operations [56]. Therefore, this low complexity detector is well suited for implementation in wearable devices and can ensure continuous long-term monitoring.

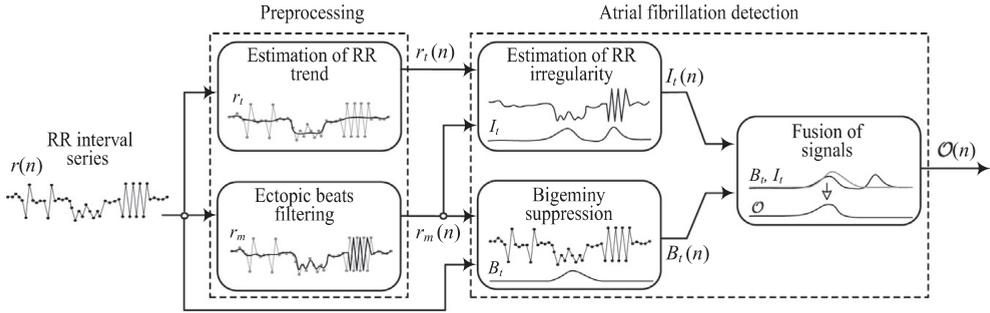


Figure 2.1. Block diagram of the rhythm-based detector [56]

2.1.2. Rhythm- and morphology-based detector

The rhythm- and morphology-based detector is designed to detect brief (<30 s) episodes of paroxysmal AF. Four parameters serve as an input to the detector, capable of detecting AF episodes as short as 8-beats [66]: 1. rhythm irregularity, quantified by the rhythm-based detector described in Sec. 2.1.1.; 2. P-wave absence, quantified by computing the normalized ratio of the rectified signal in the PQ interval to that of the TQ interval; 3. f-wave presence, quantified by the squared and summed error between different PR intervals; and 4. noise level, quantified by the spectral entropy ratio-weighted root mean square value of the extracted f-wave signal. The latter three parameters are determined from an f-wave signal extracted by using an echo state network [100]. The parameters are fed to a fuzzy logic classifier producing a fuzzy output, i.e., a value between 0 and 1, reflecting the likelihood of AF being present in the sliding detection window. The detector requires at least two ECG leads, i.e., one with negligible atrial activity (e.g., lead V_6) and the other containing atrial activity (e.g., lead V_1). The block diagram of the rhythm- and morphology-based detector is presented in Fig. 2.2.

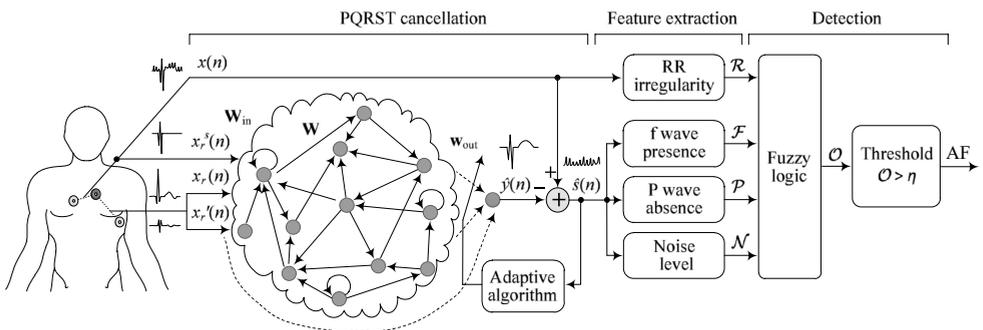


Figure 2.2. Block diagram of the rhythm- and morphology-based detector [66]

2.1.3. Deep learning-based detector

The deep learning-based detector described in [9] uses a 1D convolutional neural network to process 30-s non-overlapping ECG segments. The ECG signal is preprocessed by using a band-pass filter (0.5–40 Hz) to remove baseline wander and high-frequency noise. The convolutional neural network is composed of two convolutional layers and one fully connected layer. Both convolutional layers rely on 128 kernels with a stride of one, followed by a 1×32 average-pooling layer with a stride of 32. The fully connected layer consists of 256 neurons with a rectified linear unit activation function and two output neurons with a softmax activation function. To mitigate the risk of overfitting, all layers are followed by dropout layers with probabilities of 0.5. The outputs of the convolutional layers are batch-normalized. The DL-based detector is trained by using the gradient-based Adam optimizer [101] with a learning rate of 0.001 and a batch size of 128. The block diagram of the DL-based detector is presented in Fig. 2.3.

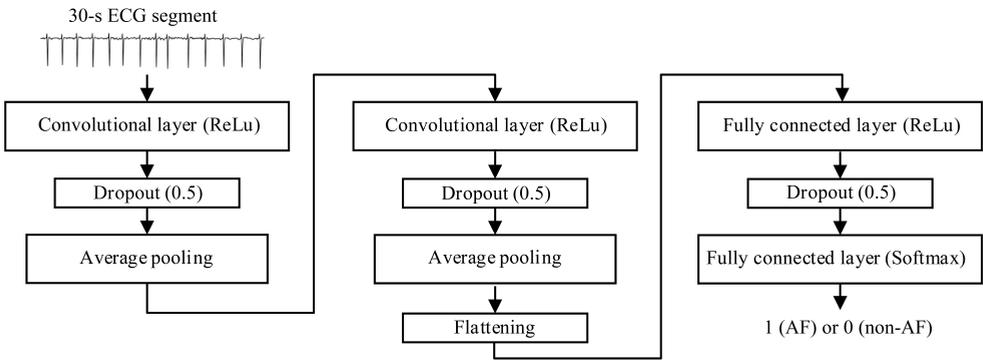


Figure 2.3. Block diagram of the DL-based detector

The detector was trained on two-thirds of the MIT–BIH Atrial Fibrillation Database (AFDB) and validated on the remaining one-third, using the lead with the most negative S-wave which reasonably well mimics the V_1 of the test databases. In order to increase the number of segments for training, each signal was divided into 30-s segments with 50% overlap. Poor-quality segments were eliminated based on sample skewness and kurtosis as proposed in [102]. In total, 358 segments out of 59,185 were eliminated due to poor quality. The resulting training dataset consists of 25,169 segments assigned to AF and 33,658 segments assigned to non-AF. To equalize the signal amplitude across the recording, the modulus of each segment was taken and normalized to the interval $[0, 1]$. The detector was tested on other databases (lead V_1), again divided into 30 s segments but without any overlap.

2.2. Characterization of atrial fibrillation patterns

A temporal AF episode pattern can be characterized by using three different approaches. The first approach relies on statistical distribution analysis, and it was used in several studies performed 20 years ago [92]–[95]. Another approach is to use the model for generating AF patterns which is capable of estimating the AF pattern characterizing parameters (the model was proposed in [10]), and the third approach is to use various parameters. All these approaches are described below in detail.

2.2.1. Distribution-based pattern characterization

Distribution-based pattern characterization is based on the fitting of a probability distribution to a series of data. There are many probability distributions, of which some can be fitted more closely to the observed data. However, in AF pattern analysis, usually only normal, lognormal, exponential, and Weibull distributions are analyzed [93, 94].

In this work, AF episode intervals, inter-episode intervals, and inter-detection intervals are characterized based on distribution analysis. All these intervals can be obtained from the AF pattern, see Fig. 2.4. The definition of an AF episode interval is intuitive; it is the time interval from the onset of an AF episode till the end of that episode, it can also be called the episode duration. The time interval between two consecutive AF episodes is the inter-episode interval, whereas the inter-detection interval is the time interval between the onsets of consecutive AF episodes. It should be noted that, in previous studies [92]–[95], only inter-episode and inter-detection intervals were analyzed.

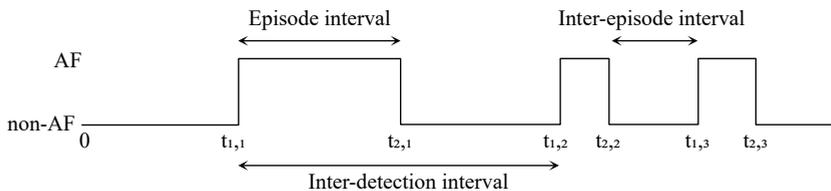


Figure 2.4. Visual explanation of AF episode interval, inter-episode interval, and inter-detection interval. Times $t_{1,1}, t_{1,2}, t_{1,3}, \dots$ show the onset of AF episodes (transitions from non-AF to AF), while times $t_{2,1}, t_{2,2}, t_{2,3}, \dots$ show the end of episodes (transitions to non-AF)

In order to characterize an AF pattern, the histograms of AF episode intervals, inter-episode intervals, and inter-detection intervals are fitted to each of the four distributions under analysis (Fig. 2.5). The distributions under analysis, parameter estimation, and goodness-of-fit evaluation are described below.

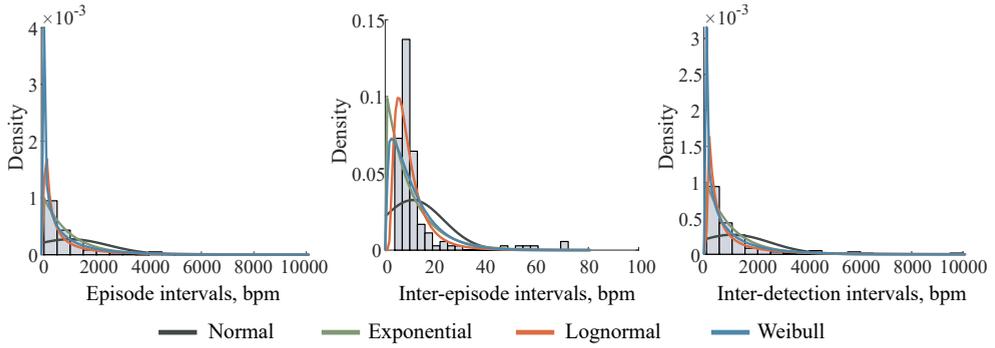


Figure 2.5. Example of distribution-based characterization of AF pattern. The histograms of episode intervals, inter-episode intervals, and inter-detection intervals are fitted to each of four distributions under analysis, i.e., normal, exponential, lognormal, and Weibull distribution. Bars show histograms, while colored lines show fitted distributions

Description of distributions under analysis. The normal distribution is a continuous probability distribution. The probability density function (PDF) of the normal distribution has a bell-shape which is symmetric around average μ . The PDF shape is determined by two parameters, i.e., the average μ and the standard deviation σ :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \quad (2.1)$$

The average μ shows the location of the normal distribution, while the standard deviation σ shows how wide the distribution is. Figure 2.6 shows the influence of parameter σ on the PDF, when μ is constant. It should be noted that 68% of the data in the normal distribution is within $\mu \pm \sigma$, 95% of the data is within $\mu \pm 2\sigma$, and 99.7% of the data is within $\mu \pm 3\sigma$.

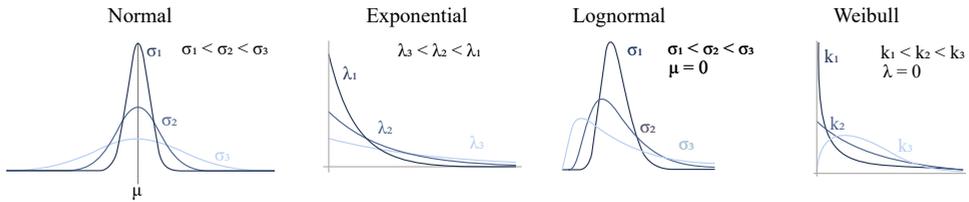


Figure 2.6. Examples of four distributions under analysis. Normal distribution is defined by two parameters (average μ and standard deviation σ), exponential distribution by one (rate λ), lognormal distribution by two (average μ and standard deviation σ), and Weibull distribution by two (shape k and scale λ). In each case, parameter influence on the PDF is showed

Exponential distribution is a continuous probability distribution which is used to model the time points between events in a Poisson point process in which events occur independently. The PDF shape of the exponential distribution is determined by

one parameter – rate λ :

$$f(x) = \lambda e^{-\lambda x}. \quad (2.2)$$

The rate parameter λ is a scale parameter which indicates the decay of the exponential function. Figure 2.6 shows the influence of the rate parameter λ on the PDF.

Lognormal distribution is a continuous probability distribution which is related to the normal distribution. That is, if the variable x is lognormally distributed, then the logarithm function of x will give the normal distribution. On the contrary, if x is normally distributed, then the exponential function of x will give the lognormal distribution. The PDF shape of the lognormal distribution is determined by two parameters, namely, the average μ and the standard deviation σ :

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}. \quad (2.3)$$

The average μ is a scale parameter which shrinks or stretches the PDF, while the standard deviation σ is a shape parameter which affects the overall shape of the PDF. Figure 2.6 shows the influence of parameter σ on the PDF when $\mu = 0$.

The Weibull distribution is a continuous probability distribution, and its PDF shape is determined by two parameters, specifically, the shape parameter k and the scale parameter λ :

$$f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}. \quad (2.4)$$

The shape parameter k shows the slope of the PDF, whereas the scale parameter λ has the effect on the distribution regarding the abscissa scale. That is, if the shape parameter k is constant, increasing the scale parameter λ will stretch out the PDF. Figure 2.6 shows the influence of the shape parameter k on the PDF when $\lambda = 0$.

Estimation of distribution parameters. To sum up, there are four parameters which can be used to define the distribution: the location parameter indicates where the distribution lies along the x -axis; the scale parameter determines the spread of the distribution; the shape parameter allows the distribution to take different shapes, where the larger parameter values indicate that the distribution tends to be skewed to the left; and the threshold parameter defines the minimum value of the distribution along the x -axis. The number of parameters used to define distribution depends on its type, e.g., normal and lognormal distributions are described by the location and scale parameters, the Weibull distribution by the shape and scale parameters, while the exponential distribution is described by only one parameter – the scale.

A number of statistical techniques can be used to estimate the parameters defining the distribution. The most popular technique is the maximum likelihood estimation, where the likelihood function is defined as [103]:

$$L = \prod_{i=1}^n f(x_i), \quad (2.5)$$

here $f(x_i)$ is the PDF of the particular distribution, e.g., the PDF of the normal distribution defined in (2.1) Equation. The goal of the maximum likelihood estimation is to find the values of the parameters describing a distribution that maximizes the likelihood function L . However, it is easier to minimize the negative log likelihood, since the negative log likelihood converts the product of PDFs to the sum of PDFs. Therefore, the negative log of (2.5) Equation is taken:

$$-\ln L = -\ln \left(\prod_{i=1}^n f(x_i) \right) = -\sum_{i=1}^n \ln(f(x_i)). \quad (2.6)$$

The next step is to find parameter values minimizing the function defined in (2.6) Equation. For this purpose, the derivative of $-\ln L$ with respect of the specific distribution parameter x (i.e., can be either the location, scale, shape, or the threshold parameter) is taken and set equal to 0:

$$\frac{d(-\ln L)}{dx} = 0. \quad (2.7)$$

Solving the above described Equation by the specific distribution parameter x gives an x value which minimizes the log likelihood function. The calculations become more difficult for distributions defined by a few parameters. The log likelihood function has to be differentiated with respect to each parameter used to define the distribution.

Goodness-of-fit analysis. When having estimated parameters, the final part is to determine which distribution fits the data best. A few methods can be used for the goodness-of-fit analysis: the Anderson-Darling test [104], the Akaike information criterion (AIC), and the Bayesian information criterion (BIC) [105].

The Anderson-Darling test is basically a hypothesis test that determines whether the data (in this case, AF episode intervals, inter-episode intervals, and inter-detection intervals) was drawn from a population that follows a hypothesized probability distribution. As in the statistical hypothesis test, the null hypothesis is used, and it claims that the data follows the hypothesized distribution. Here, the formulation of the null hypothesis differs from the one used in the statistical theory claiming that there is no difference among the groups. The result of the Anderson-Darling test is the Anderson-Darling statistic which is used to calculate the p -value. The small p -value indicates that the null hypothesis can be rejected, and a conclusion can be made that the data was not drawn from the hypothesized distribution. Meanwhile, the distribution with the largest p -value is selected as the best model. In such a case, when the data is not

well-modeled by any distribution ($p < 0.05$, significance level set to 0.05), such data is considered as non-belonging to any of these distributions. The Anderson-Darling test can only be used for normal, exponential, lognormal, and Weibull distributions.

Both the Akaike and the Bayesian information criteria are based on the maximum log-likelihood function; therefore, these two methods are closely related and defined by:

$$AIC = 2k - \ln(L), \quad (2.8)$$

$$BIC = k \ln(n) - 2 \ln(L), \quad (2.9)$$

where k is the number of estimated parameters, L is the maximum value of the log-likelihood function, and n is the sample size. In both cases, the one distribution with the minimum value of AIC or BIC is selected as the best fit to the data. The strength of evidence can be evaluated by determining the difference in values, and if the minimum BIC value differs from others by less than 2, the evidence is poor, and selected distribution can be rejected; otherwise, if the difference is more than 2, the distribution can be selected [105]. The same applies for AIC . The AIC and BIC are not useful for a single distribution, since these methods are not a statistical test like the Anderson-Darling test. Both AIC and BIC are used to compare the relative quality of several distributions.

2.2.2. Model-based pattern characterization

Another approach characterizing AF episode patterns is through history-dependent point process modeling when using an alternating, bivariate Hawkes self-exciting model. This model and the statistical framework to characterize AF episode patterns are described in detail in [10] and shortly in this section below. This part of the thesis was done in collaboration with Lund University (Lund, Sweden).

Parts of Sec. 2.2.2 have been quoted verbatim from the previously published articles: [10, 12].

Model description. The temporal pattern of AF episodes is modeled by two point processes: the first one $N_1(t)$ accounts for the transitions from non-AF to AF occurring at times $t_{1,1}, t_{1,2}, \dots$, while the second one $N_2(t)$ for the transitions from AF to non-AF occurring at times $t_{2,1}, t_{2,2}, \dots$. This bivariate point process is completely characterized by the two conditional intensity functions $\lambda_1(t)$ and $\lambda_2(t)$ defined by [106]:

$$\lambda_m(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(N_m(t + \Delta t) - N_m(t) = 1 | H_t)}{\Delta t}, \quad (2.10)$$

where the numerator is the conditional probability of a transition occurring in the interval $[t, t + \Delta t]$, and H_t is the history of the bivariate point process, i.e., the transition times $t_{1,1}, t_{2,1}, t_{1,2}, \dots$ that have occurred up to but not including t .

When using the bivariate Hawkes model [107], these two counting processes $N_1(t)$ and $N_2(t)$ with conditional intensity functions $\lambda_1(t)$ and $\lambda_2(t)$, respectively, take the following expression:

$$\lambda_m(t) = \mu_m \sum_{n=1}^2 \sum_{\{k:t>t_{n,k}\}} \alpha_{m,n} e^{-\beta_{m,n}(t-t_{n,k})}, \quad (2.11)$$

where $\mu_m > 0$, $\alpha_{m,n} \geq 0$, $\beta_{m,n} \geq 0$ for $m, n = 1, 2$, and $t_{n,k}$ the occurrence times.

The principal characteristic of the Hawkes model is that each time a new point arrives, the conditional intensity $\lambda_1(t)$ immediately increases after the transition by a factor of $\alpha_{1,1}$ (self-excitation property), and then decreases exponentially (defined by the decay parameter $\beta_{1,1}$) towards the base intensity μ_1 . The same characteristic applies to $\lambda_2(t)$ but then defined by $\alpha_{2,2}$, $\beta_{2,2}$, and μ_2 . Since a transition increases the probability of getting other points immediately after, this model can be used to obtain clustered patterns (groups of AF episodes appearing close in time). In addition to self-excitation, $\lambda_1(t)$ contains another term, defined by $\alpha_{1,2}$ and $\beta_{1,2}$, which lets $N_2(t)$ influence $N_1(t)$ (cross-excitation property); $\lambda_2(t)$ is defined in the same way as $\lambda_1(t)$, but with parameters $\alpha_{2,1}$ and $\beta_{2,1}$.

A disadvantage of the bivariate Hawkes model in its original form is that it does not impose alternation between transitions, i.e., a transition from non-AF to AF (the onset of the episode) is not necessarily followed by a transition from AF to non-AF (the end of the episode) to ensure that $t_{1,1} < t_{2,1} < t_{1,2} < t_{2,2} < \dots$ (see Fig. 2.7 a). Thus, the bivariate Hawkes model in its original form is not meaningful to use for simulating AF episode patterns. To overcome this limitation, both $\lambda_1(t)$ and $\lambda_2(t)$ are multiplied by the corresponding ‘occurrence’ function, thereby ensuring that AF occurs after non-AF,

$$o_1(t) = \begin{cases} 1, & N_1(t) = N_2(t - d_2), \\ 0, & \text{otherwise,} \end{cases} \quad (2.12)$$

and non-AF occurs after AF,

$$o_2(t) = \begin{cases} 1, & N_2(t) \neq N_1(t - d_2), \\ 0, & \text{otherwise,} \end{cases} \quad (2.13)$$

where d_1 and d_2 are the minimum duration of AF and non-AF episodes, respectively. Therefore, the alternating bivariate Hawkes model ensures that a transition from non-AF to AF (the onset of an AF episode) must, once a certain time d_1 has elapsed, be

followed by a transition from AF to non-AF (the end of an AF episode), and so on (see Fig. 2.7 b). The choice of d_1 and d_2 has implications on the number of episodes contained in the AF pattern, i.e., small values of d_1 and d_2 typically imply more episodes than do large values. Though d_1 may be set according to clinical guidelines (30 s for paroxysmal AF) [2, 3], the recent interest in brief AF episodes [108, 109] motivates the use of a small value of d_1 and, therefore, it is set to 3 s. Since d_2 has not received much clinical attention, it is set identical to d_1 .

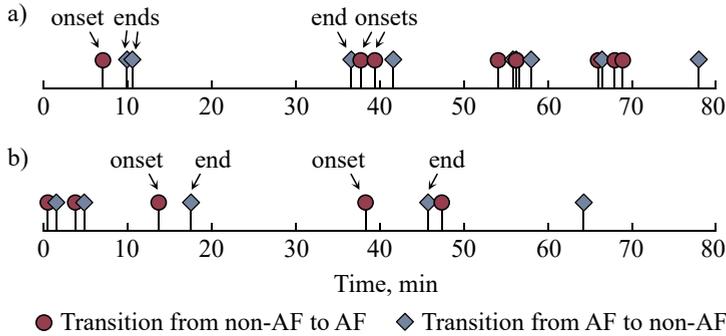


Figure 2.7. Comparison of a) simulated realization of the bivariate Hawkes point process and b) realization of the alternating, bivariate Hawkes point process. The marks ‘o’ and ‘◊’ indicate transitions from non-AF to AF (onset of AF episode) and from AF to non-AF (end of AF episode), respectively

Finally, the conditional intensity functions describing the alternating, bivariate Hawkes process are given by:

$$\Lambda_m(t) = \lambda_m(t)o_m(t), m = 1, 2. \quad (2.14)$$

The structure of $\Lambda_m(t)$ is identical to the bivariate Hawkes process in (2.2), except that a transition from non-AF to AF must, once a certain time d_1 has elapsed, be followed by a transition from AF to non-AF, and so on. A realization of the alternating, bivariate Hawkes point process with the associated $\lambda_1(t)$ and $\lambda_2(t)$ is illustrated in Fig. 2.8.

Model parameters. Assuming that $\beta_{1,1} = \beta_{1,2} = \beta_1$ and $\beta_{2,1} = \beta_{2,2} = \beta_2$ for simplicity, the model is defined by a total of eight parameters (see Table 2.1): $\mu_1, \alpha_{1,1}, \alpha_{1,2}, \beta_1$ describing $\Lambda_1(t)$ and $\mu_2, \alpha_{2,1}, \alpha_{2,2}, \beta_2$ describing $\Lambda_2(t)$. To find the model parameters from real data, the maximum likelihood estimator is derived and used. It requires a certain minimum number of AF episodes to obtain reasonably reliable parameter estimates. In this case, the maximum likelihood estimation can be performed in ECG recordings with at least 10 episodes, i.e., 20 transitions. To characterize an AF episode pattern, two main model parameters are selected and described in detail below.

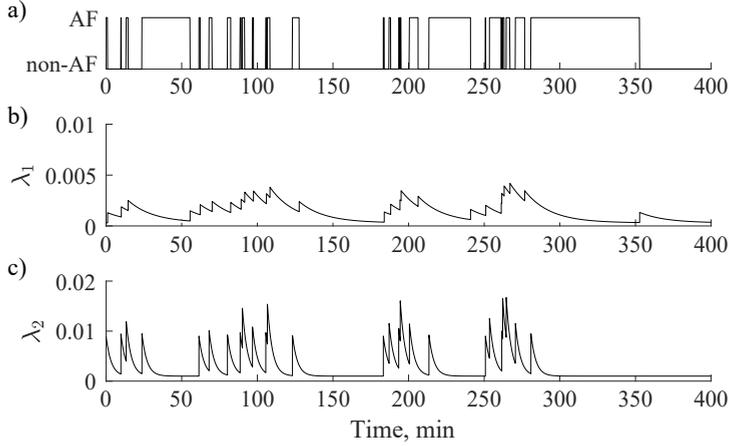


Figure 2.8. A realization of the alternating, bivariate Hawkes process: a) temporal AF episode pattern; b) and c) conditional intensity function $\lambda_1(t)$ and $\lambda_2(t)$, respectively

Table 2.1. Parameters used in the alternating, bivariate Hawkes model

Parameter	Significance
μ_1	Base intensity of λ_1
μ_2	Base intensity of λ_2
$\alpha_{1,1}$	Self-excitation of λ_1
$\alpha_{2,2}$	Self-excitation of λ_2
$\alpha_{1,2}$	Cross-excitation of $N_2(t)$ on $N_1(t)$
$\alpha_{2,1}$	Cross-excitation of $N_1(t)$ on $N_2(t)$
β_1	Intensity decay rate of λ_1
β_2	Intensity decay rate of λ_2

The parameter β_1 may be related to AF episode clustering since a slow exponential decay (i.e., a low value of β_1) increases the likelihood that an AF episode is followed by additional AF episodes. On the contrary, as β_1 increases, the decay of the intensity function towards the base intensity is faster, thus leading to AF episodes appearing more spread in time.

The base intensities μ_1 and μ_2 reflect the mean rates of transitions from non-AF to AF and from AF to non-AF, respectively. The ratio of the base intensities μ_1 and μ_2 , defined as μ ,

$$\mu = \frac{\mu_1}{\mu_2}, \quad (2.15)$$

indicates the dominance of AF or non-AF states, depending on whether $\mu > 1$ or $\mu < 1$, respectively. Thus, a few long AF episodes occurring closely in time are characterized by $\mu > 1$, whereas many short, evenly distributed AF episodes are characterized by $\mu < 1$.

2.2.3. Parameter-based pattern characterization

Parts of Sec. 2.2.3 have been quoted verbatim from the previously published article: [13].

Burden. The AF burden is defined as part of the time a patient spends in AF during the monitoring period (e.g., 1 day):

$$\mathcal{B} = \frac{T_{AF}}{T}, \quad (2.16)$$

where T_{AF} is the total time the patient spends in AF, and T is the total monitoring period. The AF burden takes values between 0 and 1, where 0 means that no AF was observed during the monitoring period, whereas 1 means that AF takes the whole monitoring period. It is important to note that the monitoring period should be provided when reporting an AF burden. Usually AF burden is reported over a 24-h period, however, the optimal monitoring period is yet to be determined [3]. Also, it should be noted that the AF burden only provides information about the time spent in AF, and it does not provide information about the temporal AF behavior.

Aggregation. The temporal AF episode pattern can be characterized by the parameter named aggregation \mathcal{A} which quantifies the deviation between an observed AF pattern and a hypothesized uniformly distributed pattern. Its definition is inspired by the parameter under the name of AF density [97, 98, 110]. AF density was modified to account for episode duration and to become suitable for the characterization of day-long recordings on the RR interval basis.

The aggregation \mathcal{A} uses the RR interval sequence as its starting point and provides detailed information of the temporal distribution of AF episodes. The sequence i_n , $n = 1, \dots, N_{RR}$ indicates whether the n :th RR interval is contained in an AF episode,

$$i_n = \begin{cases} 1, & \text{RR}_n \in \text{AF}, \\ 0, & \text{otherwise,} \end{cases} \quad (2.17)$$

where N_{RR} is the total number of the analyzed RR intervals, and $N_{AF} = \sum_{n=1}^{N_{RR}} i_n$ is the total number of RR intervals in AF.

In order to characterize the temporal distribution of AF episodes, the actual a and the reference uniform u cumulative distributions have to be identified. The actual a distribution is obtained by moving a sliding window throughout the entire binary sequence i_n (the step size is equal to one RR interval), and finding the maximal number of RR intervals assigned to AF. The window length is selected from 1 to the maximal number of RR intervals. When the window length is the same as the length of i_n , the

window embraces the entire AF pattern. The uniform u cumulative distribution represents evenly spread AF episodes throughout the entire monitoring period, and serves as a reference for finding the difference between the a and u cumulative distributions.

When having the actual a and the reference uniform u cumulative distributions, the aggregation \mathcal{A} is defined by:

$$\mathcal{A} = \frac{2}{N_{RR}N_{AF}} \sum_{n=1}^{N_{RR}} |a_n - u_n|. \quad (2.18)$$

Graphically, \mathcal{A} is defined as the ratio between the area that lies between the a and u cumulative distributions (the dark blue area in Fig. 2.9 b) and the total area above u (the dark and light blue areas in Fig. 2.9 b). The aggregation \mathcal{A} takes values between 0 and 1. Values close to 1 indicate high temporal aggregation which are inherent for patterns with a single short AF episode or a cluster, while values close to 0 indicate low aggregation; this applies to AF patterns with episodes evenly spread over the entire monitoring period or AF patterns with a single episode taking almost entire monitoring period.

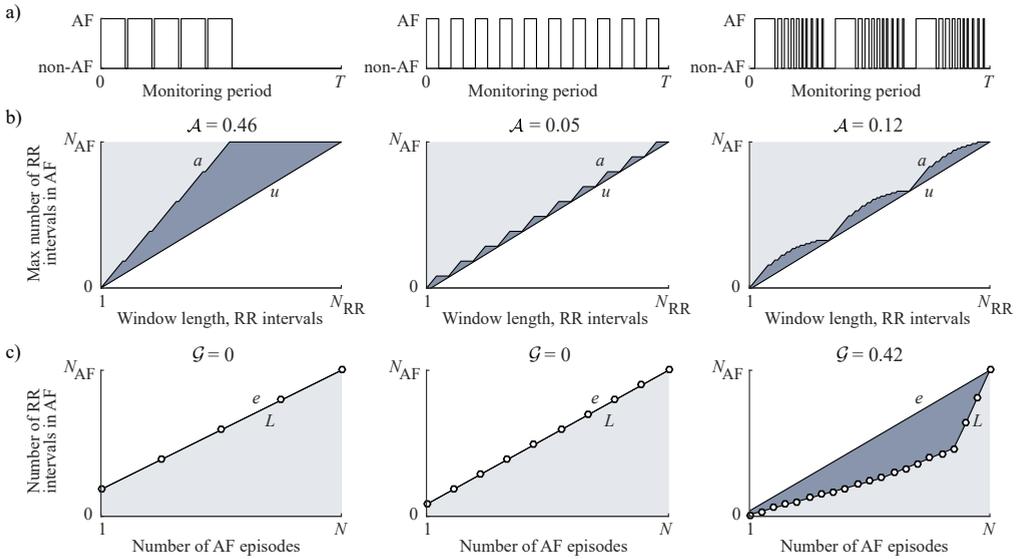


Figure 2.9. a) Different AF patterns: a single cluster with equal episode duration (left column), spread episodes (middle column), and multiple clusters with varying episode duration (right column). Graphical illustration of b) aggregation \mathcal{A} and c) Gini coefficient \mathcal{G} . \mathcal{A} is obtained by dividing the dark blue area by the total area above the uniform cumulative distribution u . Similarly, \mathcal{G} is obtained by dividing the dark blue area by the total area under the line of equality e . Note, AF burden is the same for all three patterns ($\mathcal{B} = 0.5$)

In contrast to the original implementation reported in [97], in which a single AF episode, shorter than the total monitoring period, is always assigned to a maximal aggregation, the aggregation \mathcal{A} is modified to become duration-dependent. That is, the aggregation increases when the duration of a continuous AF episode decreases. This update is motivated by the rationale that there is a major difference between a single very short episode (e.g., 30 s) and a long one (e.g., 5 h). Therefore, it is incorrect to assign such diverse AF patterns to the same aggregation value.

Gini coefficient. The AF pattern can also be characterized with respect to AF episode duration since an AF episode can take from a few seconds to several hours or days. Differences in the AF episode duration are evaluated by the Gini coefficient which is commonly used in economics to evaluate the income inequality [111]. The Gini coefficient \mathcal{G} uses the number of AF episodes N as its starting point. To evaluate the Gini coefficient, the line of equality e and the Lorenz curve L have to be identified. The line of equality e represents an AF pattern with AF episodes of the same duration, while the Lorenz L curve represents the cumulative sum of episode durations sorted in the ascending order.

When having the line of equality e and the Lorenz curve L , the Gini coefficient \mathcal{G} is defined by:

$$\mathcal{G} = \frac{2}{NN_{AF}} \sum_{i=1}^N |e_i - L_i|. \quad (2.19)$$

Graphically, \mathcal{G} is defined as the ratio between the area that lies between the line of equality e and the Lorenz curve L (the dark blue area in Fig. 2.9 c), and the total area under the line of equality e (the dark and light blue areas in Fig. 2.9 c). The Gini coefficient \mathcal{G} takes values from 0 to 1. Values close to 0 are obtained when the Lorenz curve is similar to the line of equality e . It is inherent for AF patterns with all episodes of the same duration. Meanwhile, values close to 1 are inherent for AF patterns with widely different episode durations.

2.3. Conclusions of the chapter

1. Three types of the AF detector, using either information on rhythm, rhythm and morphology, or ECG segments, have been described. The first two are expert-crafted detectors, while the third one is a deep learning-based detector. These detectors will be used to investigate the influence of the ECG signal and AF pattern properties on the detection performance and to investigate the reliability of AF pattern reconstruction.

2. A model for simulating temporal AF episode patterns has been developed which is based on the alternating, bivariate Hawkes process. The proposed model is capable of estimating the AF pattern characterizing parameters; thus the model can also be useful for AF pattern characterization.
3. Three different approaches to characterize AF patterns have been suggested. The first, distribution-based approach, relies on the fitting of probability distributions (i.e., normal, exponential, lognormal, and Weibull) to histograms of episode intervals, inter-episode intervals, and inter-detection intervals. The second approach is model-based characterization which uses two parameters from the model for simulating AF patterns: parameter β_1 accounts to episode clustering, and parameter μ accounts for rhythm dominance. Meanwhile, the third approach includes three parameters which evaluate the time spend in AF, the temporal distribution of AF episodes, and the differences in AF episode durations; these parameters are called AF burden \mathcal{B} , aggregation \mathcal{A} , and Gini coefficient \mathcal{G} , respectively.

3. ATRIAL FIBRILLATION PATTERN DATABASES AND PERFORMANCE EVALUATION

Both clinical and simulated signals have been used in this work. Clinical signals were used to investigate AF detectors (i.e., to compare annotation comparison approaches and performance measures), the reliability of AF pattern reconstruction, and for the analysis and characterization of AF patterns. Whereas simulated signals were involved for the investigation of AF detectors in order to evaluate the influence of the ECG signal and the AF pattern properties on the detection performance.

3.1. Clinical signals

Three clinical databases containing ECG signals with AF have been used in this work. Two of them are from publicly available *PhysioNet* database [112], while the third one was obtained in collaboration with Lund University (Lund, Sweden). A brief description of each clinical database is provided below.

***PhysioNet* database.** The MIT–BIH Atrial Fibrillation Database (AFDB) consists of 25 10-h, two-lead ambulatory ECG recordings from patients with AF, mostly paroxysmal [112]. Two out of 25 recordings are with a continuous AF episode. In total, AFDB consists of 297 manually annotated AF episodes which account for 43% of the total monitoring time. Figure 3.1 a) shows that AF patterns in AFDB are dominated by the small number of AF episodes, the median is 6 episodes. There is only one AF pattern consisting of more than 80 episodes, while others consist of less than 40 episodes. The duration of AF episodes varies from a few beats to more than 60 thousand of beats, the median is 171 beats (see Fig. 3.2 a).

The Long-Term AF Database (LTAfDB) consists of 84 24-h two-lead ambulatory ECG recordings acquired from patients with paroxysmal or persistent AF [112]. Twelve recordings are with continuous AF, and one recording is without AF. In total, LTAfDB consists of 7329 manually annotated AF episodes which account for 59% of the total monitoring time. Figure 3.1 b) shows that AF patterns in LTAfDB are dominated by the small number of AF episodes, the median is 12 episodes. There are only eight AF patterns consisting of more than 200 episodes. The duration of AF episodes varies from a few beats to more than 145 thousand of beats, the median is 18 beats (see Fig. 3.2 b).

Saint Petersburg database. The Saint Petersburg Atrial Fibrillation Database (SPAFDB) consists of 36 three-lead ambulatory recordings lasting from 1 to 7 days and amounting to a total of 158 days of monitoring [10]. AF episodes were annotated manually by an expert on AF analysis who consulted other experts in doubtful cases.

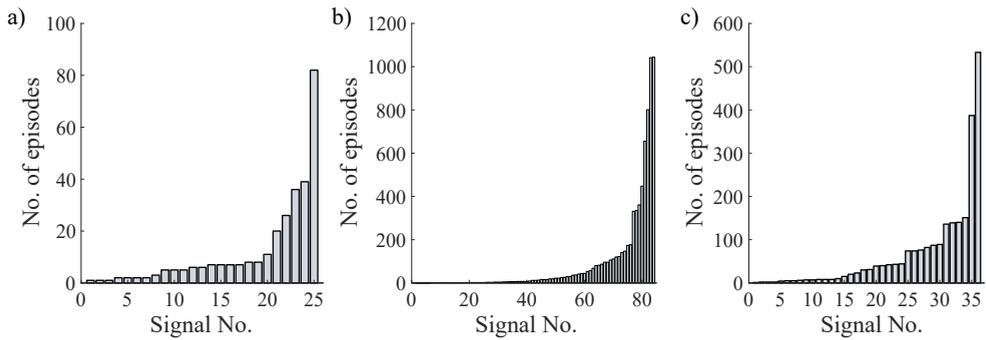


Figure 3.1. Number of AF episodes in a) AFDB, b) LTAfDB, and c) SPAfDB. Note, signals are sorted in an increasing order of AF episodes

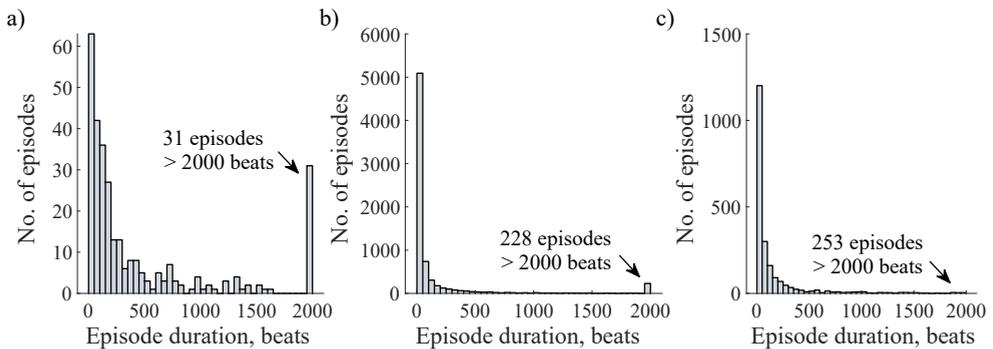


Figure 3.2. Histograms of AF episode duration in a) AFDB, b) LTAfDB, and c) SPAfDB

In total, the SPAfDB consists of 2370 AF episodes which take 19% of the total monitoring time. Figure 3.1 c) shows that AF patterns in SPAfDB are dominated by a larger number of AF episodes comparing to AFDB and LTAfDB, the median is 30 episodes. However, there are only two AF patterns consisting of more than 200 episodes. The duration of AF episodes varies from a few beats to more than 230 thousand of beats, the median is 47 beats (see Fig. 3.2 c).

Additional clinical information is available for SPAfDB, including the gender, age, and anthropometric measures together with the atrial echocardiogram test in a subgroup of 14 patients. For those patients, several atrial echocardiogram measurements, such as the left atrial (LA) volume and the LA strain are available (see Table 3.1). This study was approved by the local ethical review board.

3.2. Simulated signals

To investigate the influence of the ECG signal properties (i.e., the ECG morphology, the number of APBs, and the noise level) and the AF pattern properties (i.e., the AF burden and the median AF episode length) on the detector performance, simulated

Table 3.1. Characteristics of subgroup of patients of SPAFDB with available additional clinical information including echo-derived measures. Data is presented as mean \pm standard deviation and as absolute frequencies (percentages)

Variable	Subgroup of patients ($n = 14$)
Age (years)	62 ± 4.9
Gender (males)	8(57.1%)
Height (cm)	171 ± 8
Weight (kg)	82 ± 12.5
Left atrial volume (ml)	60 ± 16
Left atrial strain (%)	27.8 ± 12.3

ECGs with paroxysmal AF are used. The model used for simulating ECGs is described in detail in [67] and is shortly presented in the paragraph below.

The model produces 12-lead ECGs composed of real signal components randomly selected from three datasets, each consisting of the ventricular rhythm, atrial activity (f-waves or P-waves), and QRST complexes. Accounting for the switching between the sinus rhythm and AF, these components together with noise are added to produce 12-lead ECGs. Three types of noise (baseline wander, muscle noise, and electrode movement artifacts) are scaled to the desired root mean square (RMS) value and added to each lead. The block diagram of the model is shown in Fig. 3.3. A number of model parameters are set based on the characteristics in AFDB, namely, the AF burden is set to 0.37, the median episode length to 167 beats, the APB rate to 0.05, and the noise RMS level to 0.02 mV [67]. The remaining model parameters have default values [67].

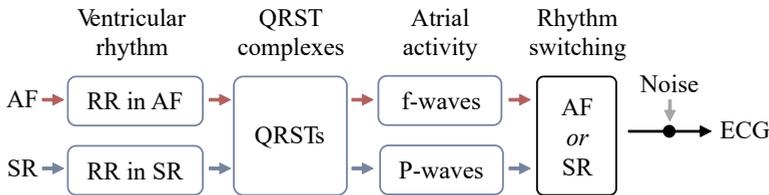


Figure 3.3. Block diagram of the model used to simulate 12-lead ECGs with paroxysmal AF

To investigate the influence of ECG signal properties, three different datasets were simulated. The dataset used for the investigation of ECG morphology contains 100 1-h simulated ECGs in leads I , II , III , V_1 , V_2 , and V_3 . The dataset used for the investigation of APBs contains 100 1-h simulated ECGs with the APB rate set to 0, 0.1, 0.2, 0.3, 0.4, and 0.5, resulting in a total of 600 ECGs. The dataset used for the investigation of noise contains 100 1-h simulated ECGs with the noise level set to 0, 0.05, 0.1, 0.15, 0.2, 0.25, and 0.3, resulting in a total of 700 ECGs.

To investigate the influence of AF pattern properties, two datasets with the median AF episode lengths of 30 and 167 beats were produced. The first dataset consists of AF patterns with brief episodes, while the second dataset consists of AF patterns similar to those in AFDB. Each dataset contains 100 1-h simulated ECGs with the AF burden set to 0.2, 0.5, and 0.8, resulting in a total of 300 ECGs.

3.3. Database subsets used for investigation

Database subsets used for the investigation of AF detectors (Sec. 4.1), AF pattern characterization (Sec. 4.2), AF pattern analysis (Sec. 4.3), and clinical association (Sec. 4.4) are provided in Fig. 3.4 and described in detail below.

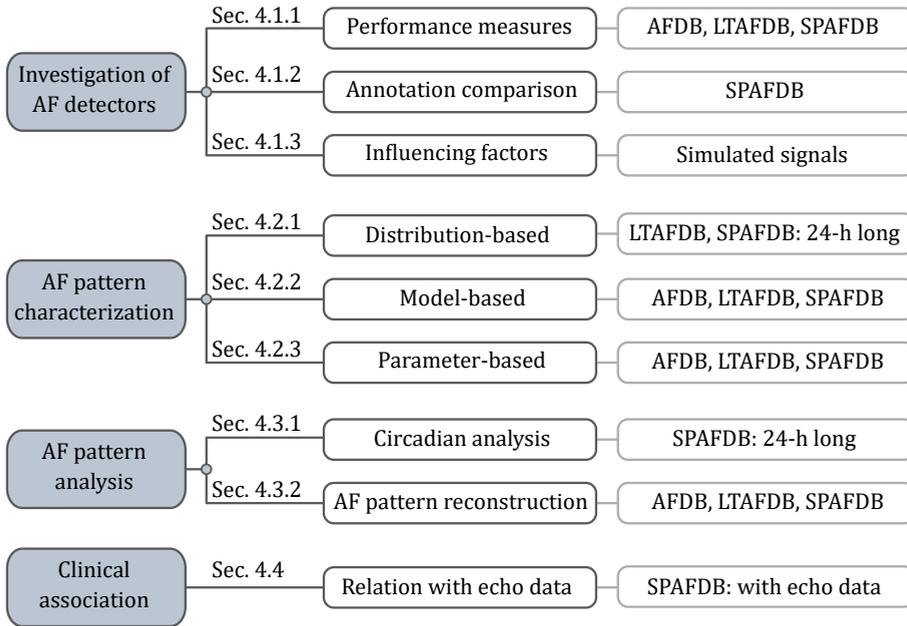


Figure 3.4. Database subsets used for investigation of AF detectors (Sec. 4.1), AF pattern characterization (Sec. 4.2), AF pattern analysis (Sec. 4.3), and clinical association (Sec. 4.4)

In Sec. 4.1, the investigation of AF detectors is performed by using both clinical and simulated signals. Analysis of performance measures is performed by using the rhythm-based detector on three clinical databases (SPAFDB, AFDB, and LTAFDB), while the analysis of annotation comparison is performed by using only SPAFDB since the investigation of all three detector types is provided. The rhythm-based detector requires only one lead, thus all three clinical databases can be used for investigation, while the rhythm- and morphology-based detector requires at least two ECG leads (one with negligible atrial activity, and the other containing atrial activity) which are only available in SPAFDB. To investigate the factors influencing detection

performance, simulated databases are used since the model for 12-lead ECGs allows to control different factors (e.g., the APB rate, the noise level, etc.).

In Sec. 4.2, the AF pattern characterizing approaches are investigated by using only clinical signals. For distribution-based pattern characterization (Sec. 4.2.1), only 24-h long AF patterns from LTAFDB and SPAFDB are used. Recordings from SPAFDB lasting a few days are divided to 24-h patterns, and only those starting at 00:00 and ending at 23:59 are taken into consideration. To sum up, there are 84 and 66 24-h patterns in LTAFDB and SPAFDB, respectively. However, distribution fitting requires at least four episodes; therefore, only those patterns consisting of more than four episodes are analyzed, resulting in total of 78 AF patterns (53 and 25 AF patterns from LTAFDB and SPAFDB, respectively). Note that for the inter-episode interval analysis, 80 patterns (53 and 27 AF patterns from LTAFDB and SPAFDB, respectively) fulfill the requirement of at least four inter-episodes per pattern.

To investigate the model-based and parameter-based pattern characterization (Secs. 4.2.2 and 4.2.3), SPAFDB, AFDB, and LTAFDB are used. Three types of AF pattern were manually defined according to the temporal characteristics: a single episode or a congregation of several episodes in a single cluster, a congregation of several episodes into multiple clusters, and episodes dispersed over the monitoring period, resulting in 33, 43, and 54 AF patterns, respectively (Fig. 3.5). Recordings without episodes (one from LTAFDB) or entirely in AF (two from AFDB and 12 from LTAFDB) were excluded from the investigation of the AF pattern characterization since these recordings cannot be assigned to any of the defined AF pattern types.

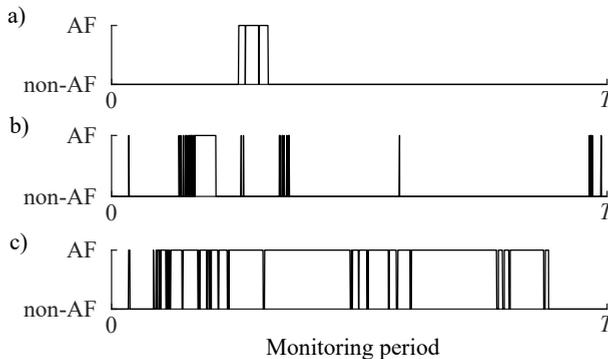


Figure 3.5. Illustration of different types of AF patterns: a) single episode or cluster, b) multiple clusters, and c) episodes dispersed over the monitoring period

In Sec. 4.3, only clinical signals are used. SPAFDB is used for the analysis of the circadian variation of AF episodes (Sec. 4.3.1). Since only 22 out of 36 recordings in SPAFDB have time when AF monitoring starts, only those can be used for the circadian pattern analysis. However, three recordings of those 22 are two days long; therefore, there is no continuous day-long pattern, i.e., from 00:00 to 23:59. Thus,

in total, 66 day-long patterns from 19 patients were analyzed. To investigate the reliability of the AF pattern reconstruction (Sec. 4.3.2), all three clinical databases are used.

In Sec. 4.4, the investigation of the relationship between the AF pattern characterizing parameters and the atrial echocardiographic parameters is performed by using SPAFDB with the available echo data. Only 14 patients were being made the LA volume and LA strain measurements. However, to evaluate model-based parameters, the AF pattern should consist of at least 10 episodes, resulting in 8 patients with the available LA volume and LA strain measurements.

3.4. Performance evaluation of atrial fibrillation detectors

Ideally, the detector-based pattern should be exactly the same as the reference pattern. However, misdetections and falsely detected episodes (false alarms) might distort the pattern as it is shown in Fig. 3.6. In this section, annotation comparison approaches as well as measures used to evaluate detection performance are presented.

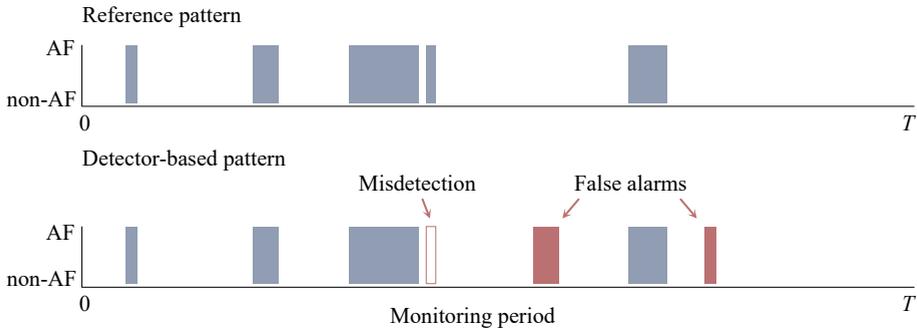


Figure 3.6. Influence of misdetections and false alarms on AF pattern. Note, the detector-based pattern should be exactly the same as reference

3.4.1. Annotation comparison approaches

The predominant approach to processing annotations is to compare the detector output (the detector-based pattern) to the annotations (the reference pattern) on the beat-to-beat basis since each beat is assigned to either AF or non-AF (Fig. 3.7 a). Another approach is to compare L -beat segments, where L is the number of beats in the segment (e.g., 8-beat, 64-beat, 128-beat, etc.); in other words, it defines the segment length. A segment is assigned to AF when at least 50% of the L detected beats are in the AF (Fig. 3.7 b). The output of the DL-based detector is segment-to-segment which means that each segment is assigned to ‘AF’ or ‘non-AF’. Yet another approach is to count the number of the correctly detected AF episodes: an episode is considered correctly

detected when the overlap between the detector output and the episode annotation exceeds a predefined threshold, e.g., 50% (Fig. 3.7 c).

In the following text, these three approaches to processing annotations are referred to as the beat-to-beat, segment-to-segment, and episode-to-episode comparison, respectively.

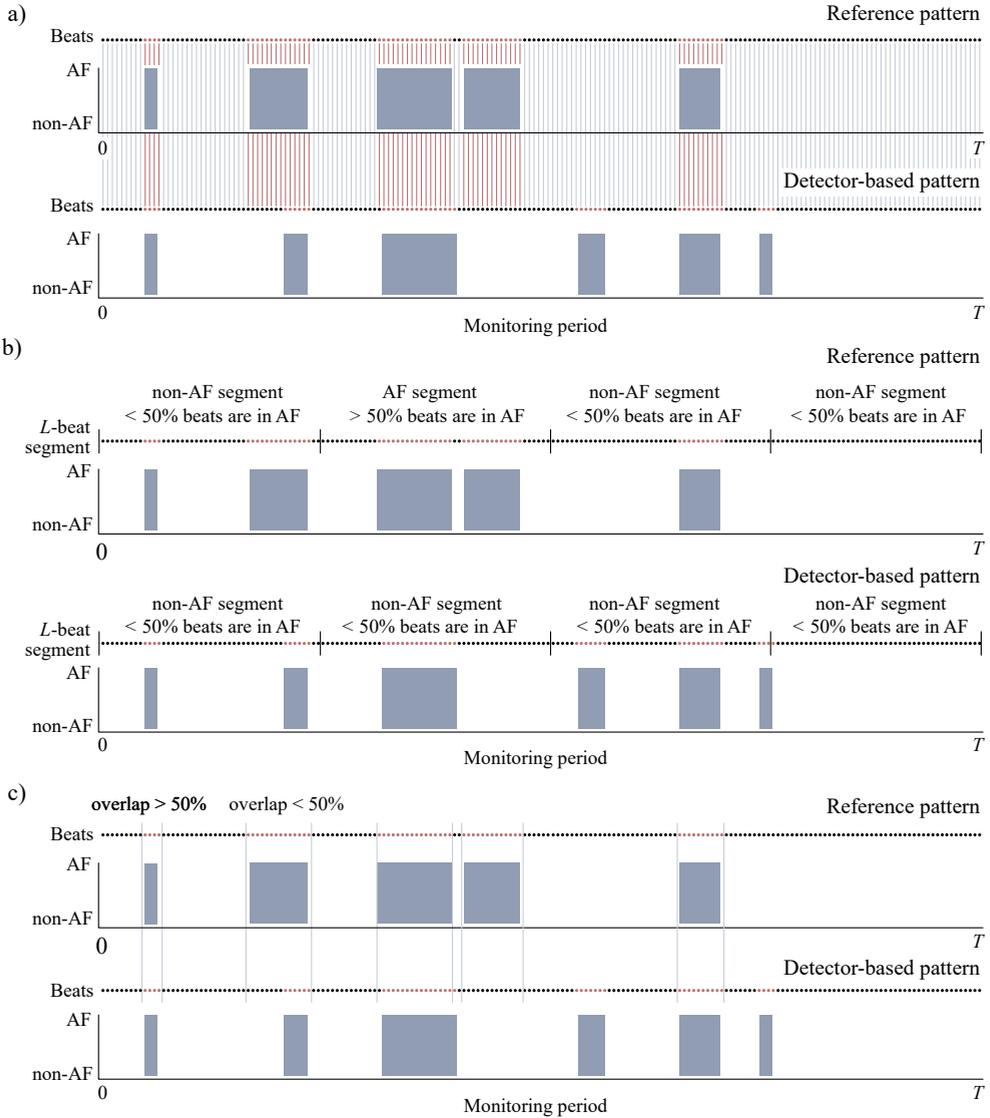


Figure 3.7. Visualization of annotation comparison approaches: a) beat-to-beat, b) segment-to-segment, and c) episode-to-episode comparison

3.4.2. Performance evaluation measures

Detection performance is evaluated by determining the number of correctly detected AF cases (true positive, TP), the number of correctly detected non-AF cases (true negative, TN), the number of falsely detected AF cases (false positive, FP), and the number of missed AF cases (false negative, FN). Depending on the annotation comparison approach, ‘case’ refers to a beat, a segment, or an episode. From these four counts, performance measures can be computed.

Sensitivity (Se) is defined by the number of correctly detected AF beats (TP) divided by the total number of AF beats (the sum of TP and FN), whereas specificity (Sp) is defined by the number of correctly detected non-AF beats (TN) divided by the total number of non-AF beats (the sum of TN and FP). Se and Sp take values in the interval $[0, 1]$, where 1 shows the perfect performance. Another widely used measure pair is the positive predictive value (PPV) and the negative predictive value (NPV). PPV is defined by the number of correctly detected AF beats (TP) divided by the total number of detected AF beats (the sum of TP and FP), whereas NPV is defined by the number of correctly detected non-AF beats (TN) divided by the total number of detected non-AF beats (the sum of TN and FN). PPV and NPV also take values in the interval $[0, 1]$, where 1 means that it is 100% true that there is an AF or non-AF if the result is positive or negative, respectively. It should be noted that the number of beats can be changed to the number of segments or episodes depending on the annotation comparison approach.

The overall detection performance can be evaluated by using accuracy (Acc) which is defined by:

$$Acc = \frac{TP + TN}{TP + FN + FP + TN}. \quad (3.1)$$

Acc takes values in the interval $[0, 1]$, where 1 means the perfect detection, and 0.5 means random detection. However, the Acc should only be used when both classes (the duration in AF and non-AF) have approximately the same size; otherwise, Acc can show overoptimistic inflated results. In imbalanced datasets, the balanced accuracy (Acc_B), the F_1 score, or the Matthews correlation coefficient (Mcc) may be a better choice [16, 113, 114]. These measures are defined by:

$$Acc_B = \frac{1}{2}(Se + Sp), \quad (3.2)$$

$$F_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}, \quad (3.3)$$

$$Mcc = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (3.4)$$

Acc_B and F_1 both take values in the interval $[0, 1]$, where 1 means the perfect detection, and 0.5 means random detection. In its original definition, Mcc takes values in the interval $[-1, 1]$; however, to facilitate a comparison, Mcc is normalized to take values in the interval $[0, 1]$ [113].

3.5. Conclusions of the chapter

1. Three clinical databases with ECG signals containing paroxysmal AF episodes were selected for the analysis of AF patterns and its characterization. However, real clinical signals do not allow investigating specific AF detector properties; therefore simulated ECG signals with paroxysmal AF episodes were used as well. The selected model allows controlling the essential properties of the simulated signals, i.e., the median AF episode length, the AF burden, the number of atrial premature beats, and the noise level.
2. Since no consensus has been established on what annotation comparison approach and performance measures should be used to report on detection performance, an important aim is to facilitate such a consensus by highlighting the differences in the annotation comparison approaches and various properties of performance measures commonly used in the literature.

4. RESULTS

4.1. Investigation of atrial fibrillation detectors

4.1.1. Analysis of performance measures

Figure 4.1 a) shows an annotated AF pattern from SPAFDB which is composed of just a few AF episodes, together with the output of the rhythm-based detector composed of numerous false detections making up for 2% of the total number of beats. By using beat-to-beat comparison, the receiver operating characteristic (ROC) shown in Fig. 4.1 b) suggests near-perfect performance. However, due to the huge imbalance between non-AF and AF beats (96.7% of beats are non-AF), such a conclusion is misleading. Since 98.1% of the AF beats and 98.0% of the non-AF beats are correctly detected, the false AF detections have negligible influence on the ROC. In terms of performance measures, Acc and Acc_B are insensitive to data imbalance and therefore indicate high performance (98.0%), whereas F_1 and Mcc are sensitive and therefore indicate lower performance (76.4% and 88.7%, respectively), see Fig. 4.1 a). Also, PPV suggests that it is only 62.6% true that there is an AF if the result is positive, while NPV suggests that it is 99.9% true that there is no AF if the result is negative.

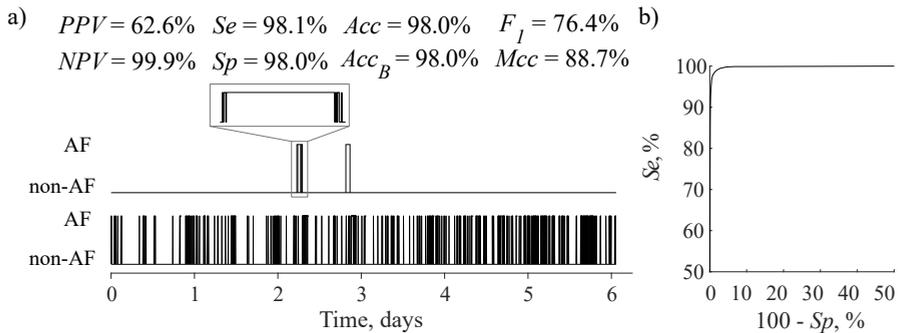


Figure 4.1. a) Annotated AF pattern from SPAFDB (the upper panel), output of the rhythm-based detector (the lower panel), and b) corresponding ROC. Performance measures are calculated by using the beat-to-beat comparison. The annotated pattern consists of 8 episodes with a median episode length of 113 beats, while the detector-produced pattern consists of 518 episodes with a median episode length of 15 beats

In order to shed further light on data imbalance, the performance of the rhythm-based detector is studied on 103 recordings from SPAFDB, AFDB, and LTAfDB; 40 recordings with AF burden < 0.01 and > 0.99 were excluded. Figure 4.2 shows that imbalance has only a minor effect on Acc , Acc_B , F_1 , and Mcc when the AF burden is between 0.1 (negative imbalance) to 0.8 (positive imbalance). Interestingly, only F_1 and Mcc are influenced by a negative imbalance of 0.01–0.1, while Acc and Acc_B remain essentially unchanged. Since the sectors 0.8–0.9 and 0.9–0.99 contain very few values, no meaningful observations can be made.

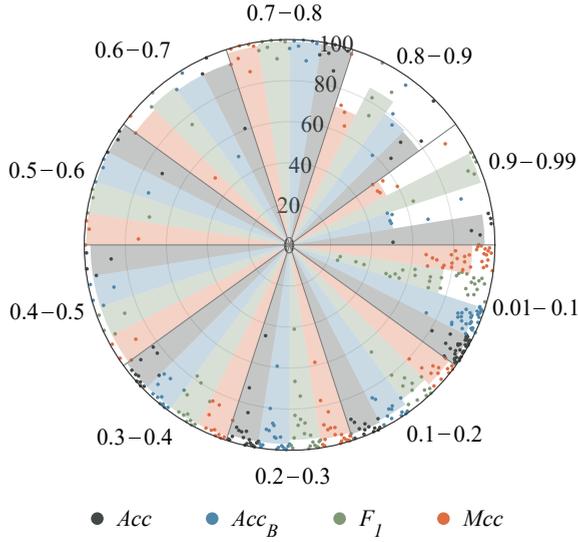


Figure 4.2. The effect of data imbalance, expressed as AF burden, on different performance measures. The arc of the circle indicates AF burden. The dots in each colored sector show the values of a performance measure obtained for different AF patterns by using the rhythm-based detector and the beat-to-beat comparison. The radius of the colored sector represents the median of the values of a performance measure. The results are obtained by using SPAFDB, AFDB, and LTAFDB

The information carried by the different performance measures is investigated by correlation analysis, again by using the rhythm-based detector and the beat-to-beat comparison on 103 recordings. Figure 4.3 shows that Sp , NPV , Acc_B , and Mcc are strongly correlated ($r > 0.8$) with each other, while PPV and F_1 do not correlate with Sp , NPV , Acc_B , or Mcc . The measure PPV correlates strongly with F_1 since both are determined by the number of false positives. On the other hand, Sp and NPV are strongly correlated due to the fact that missed AF beats are uncommon in rhythm-based AF detection, thus reducing NPV .

Discussion. Since AF detection represents a binary problem, it is intimately associated with the 2×2 confusion matrix defined by the counts TP , FN , TN , and FP , cf. Sec. 3.4.2. Combinations of these four counts have been used to define performance measures, with Se , Sp , PPV , and NPV as the most popular [56, 63, 71, 115]. None of these measures can, however, be considered fully informative as their respective definitions involve only two counts of the confusion matrix [113]. The joint use of all four measures provides richer information on performance, but also renders a comparison of performance more complicated. Therefore, it is understandable that the use of a single overall performance measure, e.g., Acc , F_1 , Mcc , has become popular [116, 117]. However, a single overall performance measure hides important properties. It is well-

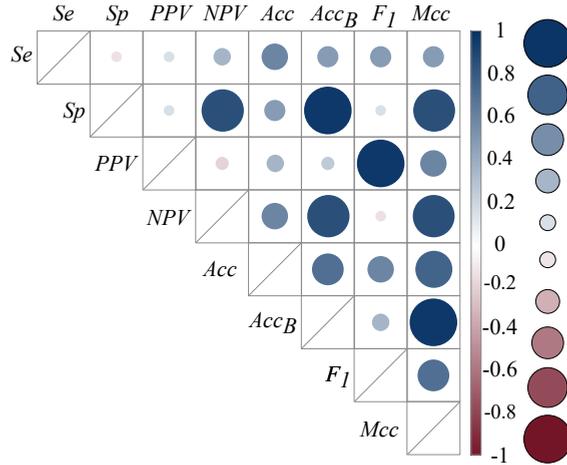


Figure 4.3. Pearson correlation coefficient for different performance measures, obtained by using the rhythm-based detector and beat-to-beat comparison. The results are obtained by using SPAFDB, AFDB, and LTAFDB

known that *Acc*, being a popular measure in AF detection, tends to inflate performance for imbalanced datasets [113, 118, 119], cf. Fig. 4.2. By comparing F_1 with *Mcc*, it should be highlighted that *Mcc* depends on the number of samples correctly classified as true negatives, while F_1 does not. Since *Mcc* indicates good performance only when most AF episodes and most non-AF ‘episode’ are correctly detected, it is recommended to use *Mcc* instead of F_1 or *Acc* when evaluating the overall performance.

The area under the ROC, known as the area-under-the-curve (AUC), is another single overall performance measure popular in many studies, e.g., [47, 56, 57, 60, 74, 120]. Unfortunately, the AUC results from integrating *Se* and *Sp* not only in the regions of operational interest, but also in regions of no clinical interest [113, 121, 122]. Hence, it is recommended to disregard AUC when reporting on performance, while it may be used to provide better understanding of how different parameter settings influence performance [56, 60, 74].

Expression of performance in terms of statistical measures (e.g., *Se*, *Sp*, etc.) is common practice. However, the use of performance measures should be accompanied by the results uncovering detector properties. For example, by investigating what signal scenarios cause frequent false alarms, weaknesses in the detector design can be more efficiently addressed. Such understanding can be gained by means of simulated ECG signals which, in contrast to real signals, offer control of the principal quantities such as the type and level of noise, the rate of APBs, the AF burden, etc. [67, 68]. The interest in brief AF episodes (<30 s) and their association with the future risk of stroke [14, 15] motivates the simulation of signals with varying episode lengths to enrich the understanding of detection performance.

4.1.2. Analysis of annotation comparison approaches

Figure 4.4 shows how the type of comparison, i.e., beat-to-beat, episode-to-episode, and segment-to-segment, influences detection performance. The results are obtained by using only SPAFDB since all three detectors are included in the analysis. For the rhythm-based and the rhythm- and morphology-based detectors, episode-to-episode comparison indicates much lower performance for all measures than do the other two types of comparison. However, for the rhythm- and morphology-based detector, the difference in performance is less pronounced which can be explained by the comparison taken to the detector design. While the segment-to-segment comparison indicates the best performance, the difference relative to the beat-to-beat comparison is negligible.

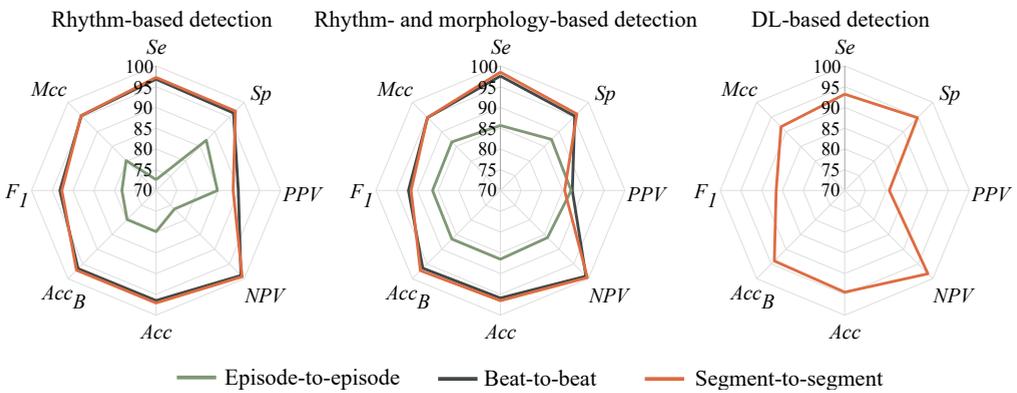


Figure 4.4. Detector performance using episode-to-episode (50% overlap), beat-to-beat, and segment-to-segment (30 s) annotation comparison. Note that only segment-to-segment comparison can be used to describe the performance of the DL-based detector, since the detector structure does not lend itself to the other two types of comparison. The results are obtained by using SPAFDB

Figure 4.5 shows how the segment length influences performance when using segment-to-segment comparison. As expected, performance deteriorates as the length shortens; due to that, shorter manifestations of noise and sporadic ectopic beats cause more false detections.

Figure 4.6 shows how the overlap percentage between the detected and annotated episodes influences performance when using episode-to-episode comparison. As expected, Se and NPV decrease and Sp and PPV increase as the overlap percentage increases since fewer episodes are detected. However, the intersection point between the Se/NPV and Sp/PPV curves differs considerably for the two types of detector, with the value being 15% for the rhythm-based and 48% for the rhythm- and morphology-based detectors. The latter detector was developed with reference to episode-to-episode comparison with 50% overlap, thus explaining the 48% intersection point.

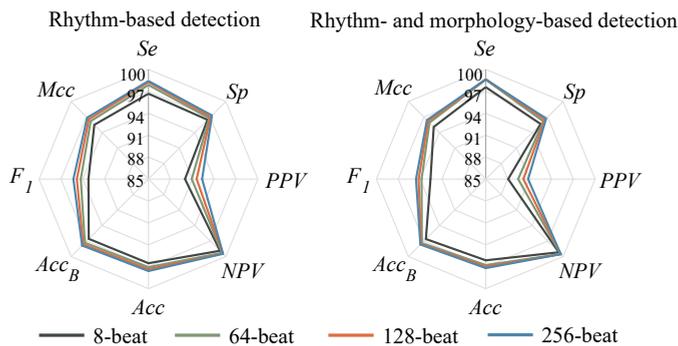


Figure 4.5. Influence of segment length on detector performance by using segment-to-segment comparison. Note, the DL-based detector is not included since it was trained to process 30-s segments. The results are obtained by using SPAFDB

These intersection points are also related to changes in the overall performance measures (i.e., Acc , Acc_B , F_1 , and Mcc) which have the inverted U-shape depending on the episode overlap.

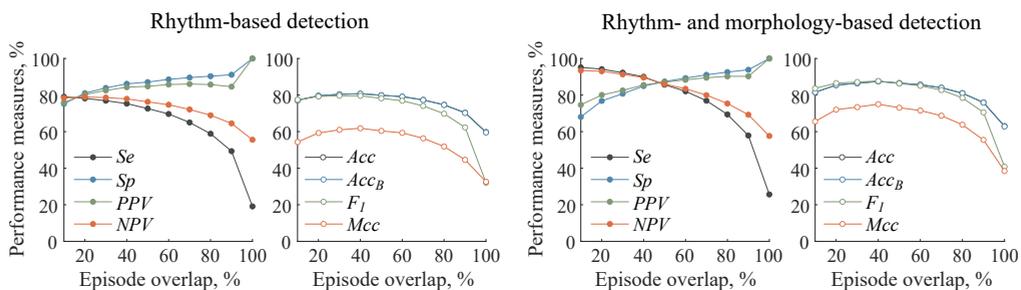


Figure 4.6. Influence of episode overlap on detector performance by using episode-to-episode comparison. The results are obtained by using SPAFDB. Note, results from Acc and Acc_B differ only slightly; therefore, Acc is covered by the Acc_B line

Discussion. Depending on the approach taken to comparing the detector output to database annotations, i.e., beat-to-beat [53]–[59], segment-to-segment [51, 57], [60]–[64], or episode-to-episode comparison [59, 63, 65, 66], the performance can differ considerably as it was shown in this study. The results show that performance depends on the selected type, notably, that episode-to-episode comparison indicates much poorer performance than do the other two types of comparison (Fig. 4.4). Even when the same type is used, the segment length (Fig. 4.5) and the episode overlap (Fig. 4.6) influence performance, e.g., it is shown to be increasingly poorer when shorter segments are used. Therefore, a meaningful comparison can only be made when these aspects are taken into consideration. If not, conclusions on detector superiority, which tend to

be common in the literature, cannot and should not be drawn.

Another aspect which deserves consideration is that neither beat-to-beat nor segment-to-segment comparison indicates the number of the detected AF episodes. Obviously, the episode-to-episode comparison is more appropriate to use, especially when temporal paroxysmal AF patterns are the subject of analysis; however, this approach is rarely used. A likely reason is that the episode-to-episode approach results in lower performance figures (Fig. 4.4). For DL-based detectors, segment-to-segment comparison is the preferred choice since deep neural networks typically do not rely on heartbeat timing; the segment length is usually related to what is deemed the shortest detectable episode. DL-based detectors requiring heartbeat timing include [46, 52, 123].

4.1.3. Investigation of detection influencing factors

Since simulated AF patterns used for the investigation of detection influencing factors are balanced (the AF burden set to 0.37), the detection performance in Figs. 4.7, 4.8, 4.9, and 4.10 is evaluated in terms of accuracy.

Lead selection. Detection accuracy as a function of the processed ECG lead is presented in Fig. 4.7. The performance of the DL-based detector depends heavily on the lead, with the best performance obtained for lead V_1 (i.e., the one used for training), then dropping dramatically for the other leads with a lower f-wave amplitude. The performance of the expert-crafted detectors (i.e., the rhythm-based and the rhythm- and morphology-based detector) is largely independent of the selected lead. The lead dependence of the rhythm-based detector is due to the fact that the performance of the QRS detector is lead-dependent.

APB rate. Detection accuracy as a function of the APB rate is presented in Fig. 4.8. The performance of the rhythm-based and the DL-based detectors drops rapidly as the APB rate increases, whereas the rhythm- and morphology-based detector performs well even at high APB rates thanks to the inclusion of morphological information. For the rhythm-based detector, the drop in performance is expected as this type of detector is known to poorly discriminate AF from irregular rhythms with APBs.

Noise level. Detection accuracy as a function of the noise level is presented in Fig. 4.9. The performance of the rhythm-based and the rhythm- and morphology-based detectors drops rapidly when the noise level exceeds 0.15 mV, which is largely attributed to the drop in performance of the QRS detector. Although less dependent on the noise level, the DL-based detector performs considerably worse at the lower noise levels

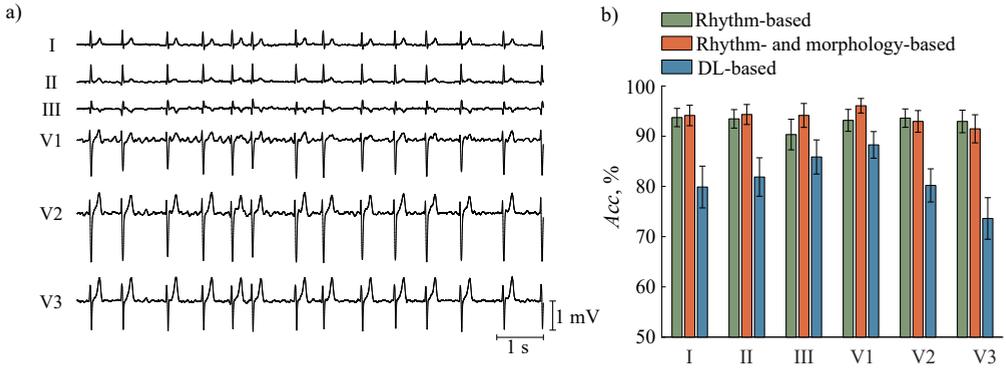


Figure 4.7. a) Simulated multi-lead ECG during AF and b) detection accuracy (Acc) as a function of lead selection. The results are based on 100 simulated 1-h ECGs and presented as mean \pm confidence interval (CI) (95%). Note, lead V_1 was used for training of the DL-based detector

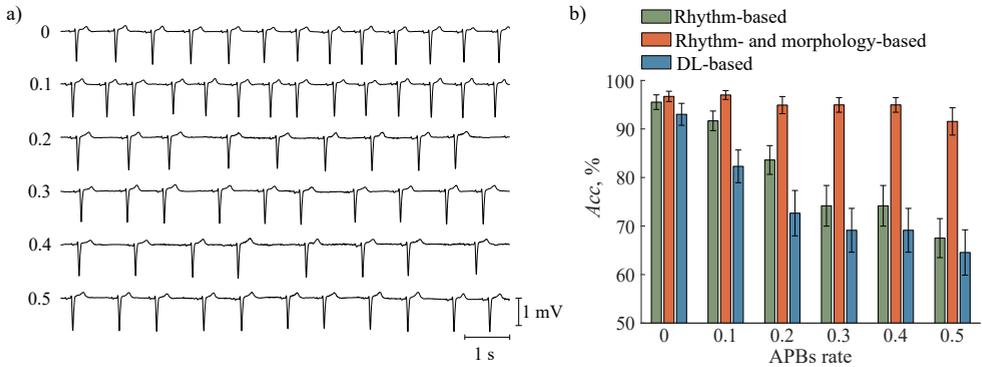


Figure 4.8. a) Simulated ECGs (lead V_1) with different APB rates and b) detection accuracy (Acc) as a function of APB rate. The results are based on 100 simulated 1-h ECGs and presented as mean \pm CI (95%)

than the other two detectors. It should be emphasized that the performance of the DL-based detector does not depend on the QRS detection performance since it does not require prior QRS detection.

AF pattern properties. Detection performance as a function of the AF burden is presented in Fig. 4.10. For a high AF burden (0.8), Se increases only slightly independently to the detector used; however, Sp is considerably influenced by the AF burden (see Fig. 4.10 b). That is, Sp drops from 99.5% to 93.3% for a high AF burden (0.8) compared to a low (0.2) when using the rhythm- and morphology-based detector, from 99.0% to 90.6% for the rhythm-based detector, and from 88.1% to 70.7% for the DL-based detector when the median AF episode length is set to 167 beats. The influence of the AF burden on Sp becomes even more evident for AF patterns

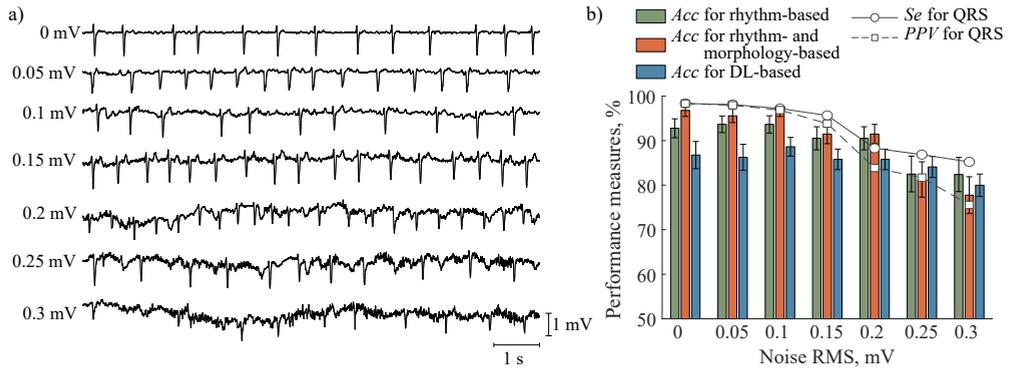


Figure 4.9. a) Simulated ECGs (lead V_1) during AF with different noise levels (RMS) and b) detection accuracy (Acc) and QRS detection performance (Se and PPV) as a function of noise level. The results are based on 100 simulated 1-h ECGs and presented as mean \pm CI (95%)

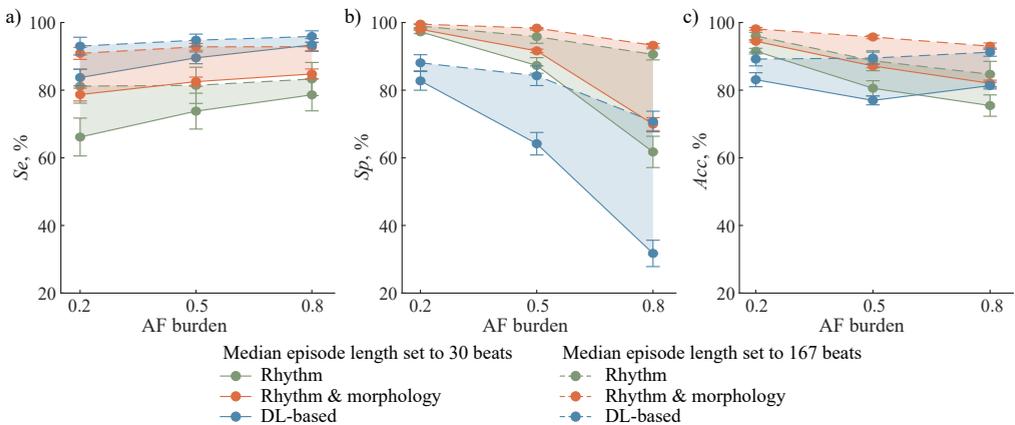


Figure 4.10. Detection performance as a function of AF burden for three types of AF detectors: a) sensitivity Se , b) specificity Sp , and c) accuracy Acc . Shaded area shows the difference in detection performance depending on the median AF episode length

with brief episodes, i.e., Sp drops from 98.0% to 70.0% when using the rhythm- and morphology-based detector, from 97.2% to 61.7% for the rhythm-based detector, and from 82.7% to 31.7% for the DL-based detector.

The detection accuracy decreases only slightly for a high AF burden compared to a low AF burden when using the rhythm- and morphology-based detector, while Acc of the rhythm-based detector is more influenced by the change in the AF burden (Fig. 4.10 c). On the contrary, Acc of the DL-based detector increases from 89.2% to 91.2% for a high AF burden when the median AF episode length is set to 167 beats.

The performance of all detector types decreases when AF patterns with brief episodes are processed (Fig. 4.10). For a low AF burden, the detection Sp drops only

slightly. However, Se decreases from 90.9% to 78.7% for patterns with brief episodes when using the rhythm- and morphology-based detector, from 81.2% to 66.1% for the rhythm-based detector, and from 93.0% to 83.7% for the DL-based detector. While for a high AF burden, Sp is substantially influenced by the change in the median episode length, i.e., Sp decreases from 93.3% to 70.0% when using the rhythm- and morphology-based detector, from 90.6% to 61.7% for the rhythm-based detector, and from 70.7% to 31.7% for the DL-based detector.

Discussion. The influence of various physiological and technical factors as well as AF pattern properties on the detection performance is rarely investigated in the literature despite the fact that such an investigation can provide essential information on the detector properties. For example, those situations in which detection performance degrades may receive particular attention.

Since the performance of the DL-based detector depends heavily on the lead selected for processing (Fig. 4.7), the datasets used for training and testing should consist of recordings from the same lead to achieve the best performance. In the present study, lead V_1 was used since its f-waves are more prominent than in any of the other leads of the standard 12-lead ECG. When using different databases for training and testing, it is not only important to use a similar lead, but also to avoid differences in measurement equipment and a large variation in the signal quality. These observations are the probable reasons why nearly all DL-based detectors in the literature have been tested by using cross-validation on the training database, see Sec. 1.2.3, Table 1.4. As for the two expert-crafted detectors, their performance is not nearly as sensitive to lead selection (Fig. 4.7).

Since P- and f-wave information is part of the decision process, the performance of the rhythm- and morphology-based detector remains largely unchanged for an increasing APB rate [66], cf. Fig. 4.8. On the other hand, as expected, the performance of the rhythm-based detector deteriorates considerably since decisions are based on RR interval information. More surprising is that the performance of the DL-based detector also deteriorates considerably. This behavior can likely be explained by a training process that identifies rhythm irregularity as a prominent AF feature; however, a representative training database with a greater variety of cardiac rhythms than that of AFDB should help improving the performance [30]. In addition to investigating the performance as a function of the APB rate, other AF-masquerading arrhythmias, e.g., bigeminy, trigeminy, atrial tachycardia, and atrial flutter, deserve to be investigated as well.

The influence of missed and falsely detected QRS complexes on the AF detector performance is rarely reported in the literature. In many studies, QRS detection is assumed to be perfect simply because the database annotations on QRS occurrence times serve as the starting point for AF detection [62, 64, 70, 71, 74]. However, in practice,

ECGs are often noisy, e.g., when recorded under ambulatory conditions, and, therefore, QRS detection is far from perfect. As evidenced by Fig. 4.9, the performance of the expert-crafted AF detectors deteriorates at higher noise levels because of the deteriorating QRS detector performance. In addition, for the rhythm- and morphology-based detectors, the noise enters through P- and f-wave features, thus calling for their careful use at higher noise levels.

The AF detector performance as a function of different AF episode lengths deserves attention since the performance will deteriorate as the length becomes increasingly shorter. For example, when the median episode length of simulated signals decreased from 120 to 30 beats, the accuracy of the rhythm-based and the rhythm- and morphology-based detectors decreased from 84% to 65% and from 92% to 80%, respectively [67]. This is in agreement with the results shown in Fig. 4.10, where not only the influence of the median episode length is presented, but also the influence of the AF burden.

The sensitivity of the DL-based detector is less influenced by the AF pattern properties than are the other detector types. However, the specificity of the DL-based detector decreases considerably when processing AF patterns with brief episodes. The reason behind is that the DL-based detector uses quite long 30-s segments, therefore, it is still going to detect AF if, e.g., only a half of that segment contains AF. Another reason is the comparison of different AF detection approaches, where one is based on the analysis on ECG segments, while the two others process ECG on the beat-to-beat basis.

As it is shown in this study, the AF pattern properties, such as the median AF episode length and the AF burden, also have influence on the detection performance. This may be a challenge aiming at the characterization of temporal AF patterns, e.g., the analysis of episode clustering [10], or the temporal distribution of AF episodes [13]. Therefore, it is important to understand the reliability of the AF pattern reconstruction by using automatic AF detectors. For this purpose, investigation of the influence of the missed and falsely detected AF episodes in terms of the pattern characterizing parameters is essential.

4.1.4. Comparison of detector performance

A summary of the strengths and weaknesses of rhythm-based, rhythm- and morphology-based, and DL-based detectors is presented in Table 4.1.

Limitations. One detector representative for each of the three main types found in the scientific literature, i.e., rhythm-based, rhythm- and morphology-based, and DL-based, have been studied. Another type of detectors is the one relying solely on atrial information; however, this type was not considered as it is known to perform poorly

Table 4.1. Strengths and weaknesses of rhythm-based, rhythm- and morphology-based, and DL-based detectors

Detector type	Strengths	Weaknesses
Rhythm-based	Low computational demands Well-suited for implementation in wearable devices	Depends on QRS detection performance Difficult to distinguish AF from other irregular rhythms
Rhythm- and morphology-based	Well-suited for distinguishing AF from other irregular rhythms	P- and f-wave features are sensitive to noise Depends on QRS detection performance
DL-based	No need for expert-crafted features No need for QRS detection Performance can be improved using training dataset with non-AF arrhythmias	Sensitive to changes in ECG morphology Unclear how the detector generalizes on unseen data Detection is data-driven and thus lacks interpretability Complex networks are computationally demanding

in noisy signals [16]. While other representatives could have been chosen, the aim of this work is to identify structure-dependent aspects on performance evaluation, but not to grade the performance of different detectors, therefore making the choice of representatives less critical.

In certain applications, e.g., wearable devices, the computational complexity needs to be considered when evaluating the performance. Since complexity is detector-specific, such considerations have been left out of the present study. Nonetheless, the structure of a rhythm-based detector is typically less complex than that of a DL-based. For example, the rhythm-based detector in [56] requires 8 multiplications per RR interval, whereas the DL-based detector in [46], with its 159,841 trainable parameters, evidently requires many more multiplications as well as dramatically more memory.

The DL-based detector was trained on ECG segments whose quality was determined from sample skewness and kurtosis [102]. Recently, other approaches towards quality assessment have been proposed to be designed specifically for use in AF detection [124]–[127]. These approaches may lead to better training results, and, ultimately, better performance.

The modern sensor technology has helped form a new paradigm of long-term AF monitoring relying on the analysis of PPG signals. As a result, a large number of PPG-based AF detectors have been published, e.g., [30, 54, 58, 65], [128]–[133]. While PPG-based detection was not addressed in this work, the considerations made on the performance evaluation of ECG-based detectors apply to PPG-based detectors as well.

Recommendations. From the implications of the results presented in Sec. 4.1 as well as from reviewing a large number of recent, peer-reviewed papers (cf. Sec. 1.2.3, Tables 1.3 and 1.4), the present work leads up to the following five recommendations on evaluating detector performance:

1. To use different datasets for training and testing, and to ensure that the two datasets do not contain signals from the same patient.
2. To substantiate the approach taken to annotation comparison, and, if applicable, to report the segment length and episode overlap.
3. To evaluate performance in terms of Mcc , rather than Acc or F_1 , and to include Se , Sp , and PPV so as to facilitate a comparison of the many published detectors; AUC should be left out when reporting performance.
4. To evaluate the influence of the physiological and technical factors on performance, including lead selection, APB rate, noise level, and AF-masquerading arrhythmias.
5. To pay special attention to detection performance when the aim is to characterize AF episode patterns.

Parts of Sec. 4.1 have been quoted verbatim from the previously published articles: [9, 11].

4.2. Characterization of atrial fibrillation patterns

4.2.1. Distribution-based pattern characterization

Episode intervals. The goodness-of-fit of episode intervals is provided in Table 4.2. The results obtained by using AIC and BIC are the same except for the two patterns assigned to exponential distribution based on BIC. By using these methods, most of the cases are assigned to lognormal distribution, i.e., 54 out of 78 patterns (69%). AIC and BIC rejected 22 (28%) and 20 (26%) patterns, respectively, while the Anderson-Darling test rejected 40 (51%) patterns. Coincidence cases among these three methods used for the evaluation of goodness-of-fit are distributed as follows: 10 (13%) patterns are fitted to lognormal distribution, and 1 (1%) to Weibull distribution.

Inter-episode intervals. The goodness-of-fit of inter-episode intervals is provided in Table 4.2. The results obtained by using AIC and BIC are the same: 63 (79%) patterns are assigned to lognormal distribution, and 1 (1%) pattern is assigned to Weibull distribution. AIC and BIC rejected 16 (20%) patterns out of 80, while the Anderson-Darling test rejected 53 (66%) patterns. Coincidence cases among these three methods used for the evaluation of goodness-of-fit are distributed as follows: 18 patterns are fitted to lognormal distribution.

Inter-detection intervals. The goodness-of-fit of inter-detection intervals is provided in Table 4.2. The results obtained by using AIC and BIC are the same, i.e., all 78 patterns are fitted to lognormal distribution. However, the Anderson-Darling test showed that 73 out of 78 patterns (94%) do not fit to any of the four distributions. Only 5 (6%) patterns are fitted to lognormal distribution, which is in agreement with the results of AIC and BIC for those patterns.

Table 4.2. Goodness-of-fit based on three criteria: p -value from the Anderson-Darling test, the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). The results are obtained by using 24-h patterns from SPAFDB and LTAADB

Distribution	No. of fitted patterns								
	Episode intervals			Inter-episode intervals			Inter-detection intervals		
	p -value	AIC	BIC	p -value	AIC	BIC	p -value	AIC	BIC
Normal	0	0	0	0	0	0	0	0	0
Exponential	4	0	2	1	0	0	0	0	0
Lognormal	21	54	54	19	63	63	5	78	78
Weibull	13	2	2	7	1	1	0	0	0
None	40	22	20	53	16	16	73	0	0

Discussion. Since both, AIC and BIC, are based on the maximum log-likelihood function, the results from these two criteria match in almost all cases. However, the results differ considerably from the Anderson-Darling test. One of the reasons is that the Anderson-Darling test is a statistical test, and it is based on the p -value (significance level set to 0.05). While BIC and AIC select the one with the minimum likelihood value, and the evidence strength is evaluated as the difference between the selected distribution and the remaining distributions.

Based on AIC and BIC, the analysis of daily patterns showed that the majority of patterns of episode intervals (69%), inter-episode intervals (79%), and inter-detection intervals (100%) can be described by lognormal distribution. However, the Anderson-Darling test showed that the majority of patterns of episode intervals (51%), inter-episode intervals (66%), and inter-detection intervals (94%) cannot be fitted to the normal, exponential, lognormal, or Weibull distributions. These results differ from other studies, e.g., it was found that inter-episode intervals follow exponential distribution [92] or Weibull distribution [93,94]. Thus, the conclusion should be made with caution.

The reason why the majority of cases were not fitted to one of the distributions can be explained by the observation that AF episodes tend to cluster [93], while distribution-based characterization rests on the assumption that episodes are statistically independent. Also, a day-long AF pattern is likely to be too short to apply distribution-based analysis. Further investigation should be based on longer AF pat-

terns, e.g., several weeks. Lastly, in one of the recently published studies, it was found that the distribution of the AF burden among patients is bimodal, which means that the distribution is skewed to the extreme values (i.e., a low and high AF burden) [134].

4.2.2. Model-based pattern characterization

The maximum likelihood estimator was derived to find the model parameters from the real data by using LTAfDB, AFDB, and SPAfDB (see [10] for details). However, only recordings with at least 10 episodes are included in the analysis. When using the maximum likelihood method for parameter estimation, the goodness-of-fit analysis demonstrated that the alternating, bivariate Hawkes model fitted the data in the vast majority of recordings. In quantitative terms, Λ_1 and Λ_2 fit the data in 68% and 97% recordings, respectively. This result implies that a wide range of episode patterns can be modeled by using the proposed alternating, bivariate Hawkes model.

Figure 4.11 shows three different modeled AF patterns by changing the parameter μ which accounts for rhythm dominance (AF or non-AF) and the parameter β_1 which accounts for the degree of episode clustering. AF patterns with $\mu > 1$ are dominated by AF (see Fig. 4.11 c), while patterns with $\mu < 1$ are dominated by non-AF (see Fig. 4.11 a and b). AF episode clustering is related to β_1 , i.e., a low value of β_1 increases the likelihood that an AF episode is followed by additional AF episodes, while AF episodes appear to be more spread out in time as β_1 increases. This effect is illustrated in Fig. 4.11 a) and b).

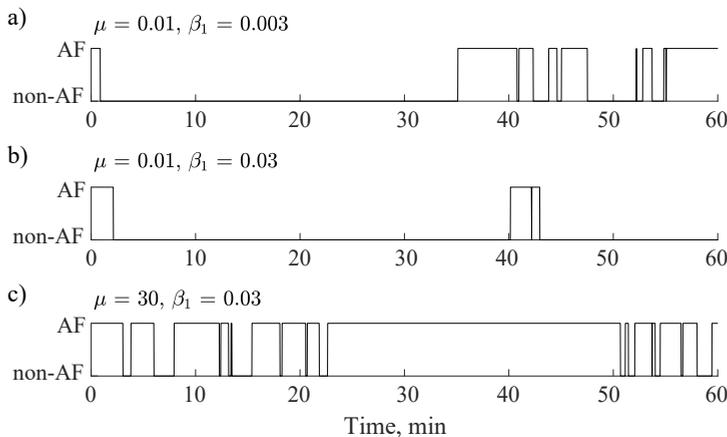


Figure 4.11. AF episode patterns generated by using the alternating, bivariate Hawkes model: a) clustered AF episode pattern dominated by non-AF; b) AF pattern dominated by non-AF and increasing β_1 from 0.003 to 0.03, leading to less clustering of AF episodes; and c) AF pattern dominated by AF, increasing μ from 0.01 to 30

The results presented in Table 4.3 show the model-based parameter, β_1 and μ , values for different AF pattern types. It is of interest that AF patterns with dispersed episodes and multiple clusters take much lower β_1 values comparing to the patterns with a single cluster. Also, AF patterns with dispersed episodes have highest μ values (the mean is 4.19) which means that these patterns are dominated by AF, i.e., $\mu > 1$ is inherent for patterns with longer episodes. The other two types of patterns are dominated by non-AF (the mean is 0.70 and 0.16 for patterns with single and multiple clusters, respectively), i.e., $\mu < 1$ is inherent for patterns with many short episodes.

Table 4.3. Model-based parameters β_1 and μ for different AF pattern types. The results are obtained by using SPAFDB, AFDB, and LTAfDB and are shown as mean and CI (95%). AF patterns with less than 10 episodes were excluded. The outliers, defined as three scaled median absolute deviations away from the median, were excluded as well, which resulted in a different number of pattern types when reporting β_1 and μ parameters

Pattern type	No.	β_1	No.	μ
Single cluster	19	0.043 [0.015–0.071]	7	0.702 [0.125–1.278]
Multiple clusters	26	0.008 [0.005–0.011]	26	0.163 [0.103–0.223]
Dispersed episodes	10	0.006 [0.002–0.009]	17	4.191 [0.670–7.713]
Entire database	52	0.009 [0.006–0.011]	50	0.375 [0.234–0.516]

A desirable property of the model-based characterization is to convey information complementary to the AF burden. Here, the main model parameters (β_1 and μ) are analyzed with the aim to determine whether they are strongly correlated with the AF burden or provide different information about the AF pattern. Figure 4.12 presents scatter plots for the AF burden, β_1 , and μ . As it is shown, β_1 is negatively weakly correlated with the AF burden ($r = -0.16$), whereas μ is positively weakly correlated ($r = 0.24$).

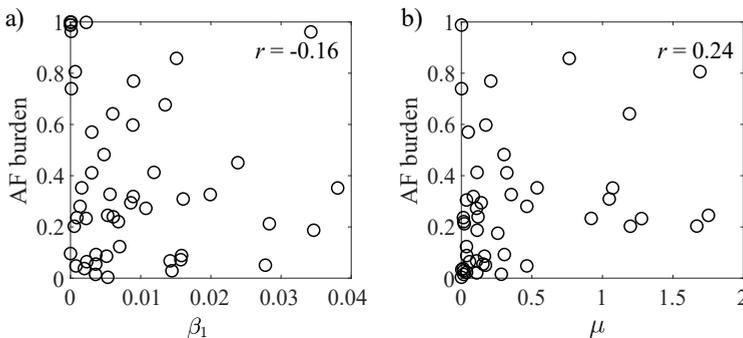


Figure 4.12. Scatter plots of the AF burden and model-based parameters: a) clustering parameter β_1 and b) rhythm dominance parameter μ . The sample Pearson cross-correlation coefficient r is given in each plot. The results are obtained by using SPAFDB, AFDB, and LTAfDB. AF patterns with less than 10 episodes as well as outliers were excluded

Discussion. Goodness-of-fit analysis showed that the alternating, bivariate Hawkes model is appropriate to model a wide range of AF patterns. The correlation analysis in Fig. 4.12 shows that the clustering parameter β_1 and the rhythm dominance parameter μ are only weakly correlated with the AF burden and, therefore, may provide complementary information, which may be useful for risk assessment of ischemic stroke and as well as for better understanding of AF progression. However, correlation analysis cannot say anything about the clinical significance of a certain parameter, thus necessitating a study which investigates the significance of the model parameters for the purpose of, e.g., predicting the risk of stroke. Indeed, the clinical evaluation of methods for episode pattern characterization has turned out to be a major challenge as neither the millennial studies [93, 94, 95] nor the study introducing the AF density [97] investigated the relationship between the pattern characteristics and the patient outcome, although all of these were published in clinical journals.

The alternating, bivariate Hawkes model requires a certain minimum number of episodes to produce reliable results. By setting this number to 10, i.e., 20 transitions, a trade-off was made between the risk of model overfitting and the wish to include as many recordings as possible. In a future clinical study, this choice may very well be the subject of further investigation.

4.2.3. Parameter-based pattern characterization

The results presented in Table 4.4 show the AF burden \mathcal{B} , aggregation \mathcal{A} , and Gini coefficient \mathcal{G} values for different AF pattern types. AF patterns with single or multiple clusters take low AF burden values compared to AF patterns with dispersed episodes (the mean is 0.19, 0.20, and 0.64, respectively). These three types of AF patterns can be distinguished by using the aggregation parameter, i.e., AF patterns with a single cluster take large \mathcal{A} values (the mean is 0.79), slightly smaller values for AF patterns with multiple clusters (the mean is 0.59), and patterns with episodes dispersed over the monitoring period take much smaller values (the mean is 0.22). On the other hand, the Gini coefficient \mathcal{G} is similar for all types of pattern, with the larger values for multiple clusters and the smaller values for a single cluster.

Table 4.4. AF burden \mathcal{B} , aggregation \mathcal{A} , and Gini coefficient \mathcal{G} for different AF pattern types. The results are obtained by using SPAFDB, AFDB, and LTAfDB and shown as mean and CI (95%)

Pattern type	No.	\mathcal{B}	\mathcal{A}	\mathcal{G}
Single cluster	33	0.192 [0.126–0.258]	0.788 [0.726–0.851]	0.467 [0.361–0.572]
Multiple clusters	43	0.195 [0.144–0.247]	0.592 [0.517–0.667]	0.732 [0.681–0.782]
Dispersed episodes	54	0.678 [0.583–0.773]	0.218 [0.151–0.285]	0.629 [0.572–0.685]
Entire database	130	0.395 [0.333–0.456]	0.486 [0.429–0.544]	0.622 [0.579–0.664]

Figure 4.13 clearly illustrates that aggregation \mathcal{A} depends on the temporal distribution of AF episodes, while the Gini coefficient \mathcal{G} depends on the episode duration. That is, \mathcal{A} differs considerably between patterns with highly aggregated episodes and those with dispersed episodes, see Fig. 4.13 a) comparing to b) and c). Meanwhile, \mathcal{G} is similar due to the presence of episodes with a widely varying duration, see Fig. 4.13 a) and b). On the other hand, the patterns with a high burden \mathcal{B} are better reflected by the \mathcal{G} values, see Fig. 4.13 b) and c). Therefore, it is obvious that the presented parameters provide different information about AF pattern.

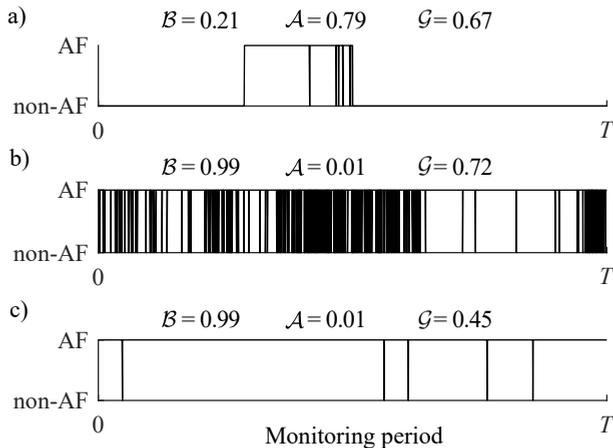


Figure 4.13. AF burden \mathcal{B} , aggregation \mathcal{A} , and Gini coefficient \mathcal{G} for different AF patterns: a) AF pattern with a single cluster (AFDB), b) AF pattern with 331 short episodes (LTAfDB), and c) AF pattern with 6 long episodes (LTAfDB)

Analyzing AF patterns from SPAfDB, AFDB, and LTAfDB, the aggregation \mathcal{A} and the Gini coefficient \mathcal{G} are analyzed with the aim to determine whether they are strongly correlated with the AF burden \mathcal{B} . Figure 4.14 a) shows that \mathcal{A} is negatively strongly correlated with the AF burden \mathcal{B} ($r = -0.90$), and the correlation increases for an increasing \mathcal{B} . For example, by analyzing only AF patterns with the AF burden less than 0.5 (87 patterns), the correlation between \mathcal{B} and \mathcal{A} is -0.70 , while analyzing AF patterns with the AF burden less than 0.4 (81 patterns), 0.3 (98 patterns), 0.2 (49 patterns), and 0.1 (41 patterns), the correlation decrease to -0.59 , -0.50 , -0.28 , and -0.20 , respectively. This means that the aggregation parameter is particularly important in the AF pattern with the AF burden less than 0.5 (see Fig. 4.15).

On the other hand, the Gini coefficient \mathcal{G} is independent of the AF burden \mathcal{B} ($r = 0.07$, Fig. 4.14 b). Even for \mathcal{B} values close to 1, \mathcal{G} takes different values since the episode duration is highly variable among the distinct AF patterns.

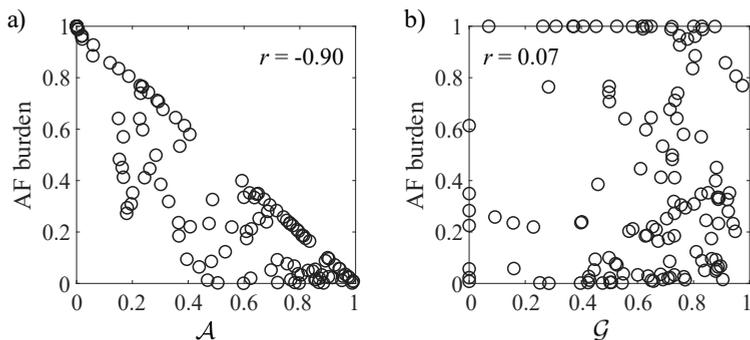


Figure 4.14. Scatter plots of AF burden \mathcal{B} and a) aggregation \mathcal{A} and b) Gini coefficient \mathcal{G} . The sample Pearson cross-correlation coefficient r is given in each plot. The results are obtained by using SPAFDB, AFDB, and LTAfDB

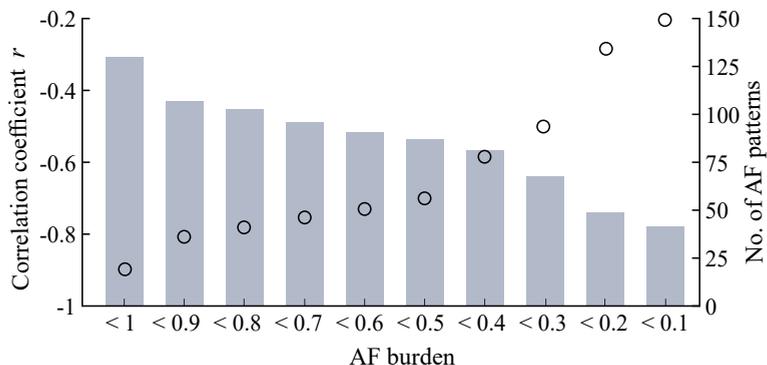


Figure 4.15. Influence of AF burden on the Pearson cross-correlation coefficient r between AF burden \mathcal{B} and aggregation \mathcal{A} . The circles represent correlation coefficient value, while the bars represent the number of AF patterns with AF burden $< 1, < 0.9, \dots < 0.1$. The results are obtained by using SPAFDB, AFDB, and LTAfDB

Discussion. Several studies have shown a link between the AF burden and the increased risk of stroke, however, the threshold of the AF burden is unclear and varies considerably among studies [135, 136]. Therefore, there is a good reason to presume that the pattern itself may have a role on thrombus formation. This investigation shows that the aggregation is capable of differentiating patterns even when only day-long or a few-days long recordings are available. This is in contrast to the distribution-based characterization which requires a large number of AF episodes, and, therefore, day-long recordings are usually insufficient. The results showed that the aggregation is useful for AF pattern analysis with a low AF burden (e.g., < 0.50), while the Gini coefficient is useful for discriminating patterns with a high AF burden, and can thus be used as a complementary to the aggregation.

The limitation of the present investigation of the model-based and parameter-based characterization is that, due to the lack of knowledge about the AF pattern, each observed pattern was assigned to one of the three types of pattern heuristically defined based on manual inspection. Some patterns could not be easily assigned to a specific pattern type, thus the results should be interpreted with caution. In contrast to the commonly used AF burden, the investigated parameters are not as easily interpreted, and the ranges characterizing different patient groups should be drawn with caution.

Parts of Sec. 4.2 have been quoted verbatim from the previously published articles: [10, 13].

4.3. Analysis of atrial fibrillation patterns

4.3.1. Circadian analysis of atrial fibrillation

The AF pattern may vary considerably among patients. Figure 4.16 shows 4-days long AF patterns from three different patients from SPAFDB. In the first example (Fig. 4.16 a), AF takes 59.39% of the monitoring time and consists of only 20 episodes, while the AF patterns in Fig. 4.16 b) and c) consist of 533 and 151 episodes, respectively. However, the AF burden is 4–11 times smaller in these patterns comparing to the first one. Also, from Fig. 4.16 b) and c), it can be clearly seen that AF episodes tend to cluster in time.

Figure 4.17 shows the circadian variation of AF expressed as the AF burden per hour. The results showed the clear dominance of AF at night and a peak around 19:00 o'clock (the mean AF burden is more than 0.2). Meanwhile, during the day-time, i.e., between 11:00–17:00 o'clock, the AF burden is lower.

Discussion. The circadian rhythm helps to regulate the process in the body during the 24-hour period. It is inherent for cardiovascular diseases as well [137]. AF episodes do not occur randomly; for example, some patients report that AF episodes occur at about the same time of the day, usually in the evening or at night [138, 139]. This is in agreement with the present investigation showing that AF episodes tend to occur during the night-time (cf. Fig. 4.17). However, other studies also showed that AF episodes are more common during the morning hours [140, 141] or in the daytime [140].

Even though results from different studies are not conclusive, they still suggest that the AF pattern might change depending on the circadian rhythm. It might be that the differences among studies are related to patient characteristics, e.g., age. A few studies reported that, for younger patients, AF episodes occur at night, while, for older patients, AF episodes occur during the day [138, 142]. In future studies, circadian analysis could be extended to the weekly or yearly AF pattern analysis [141]. However, it requires continuous long-term monitoring of patients.

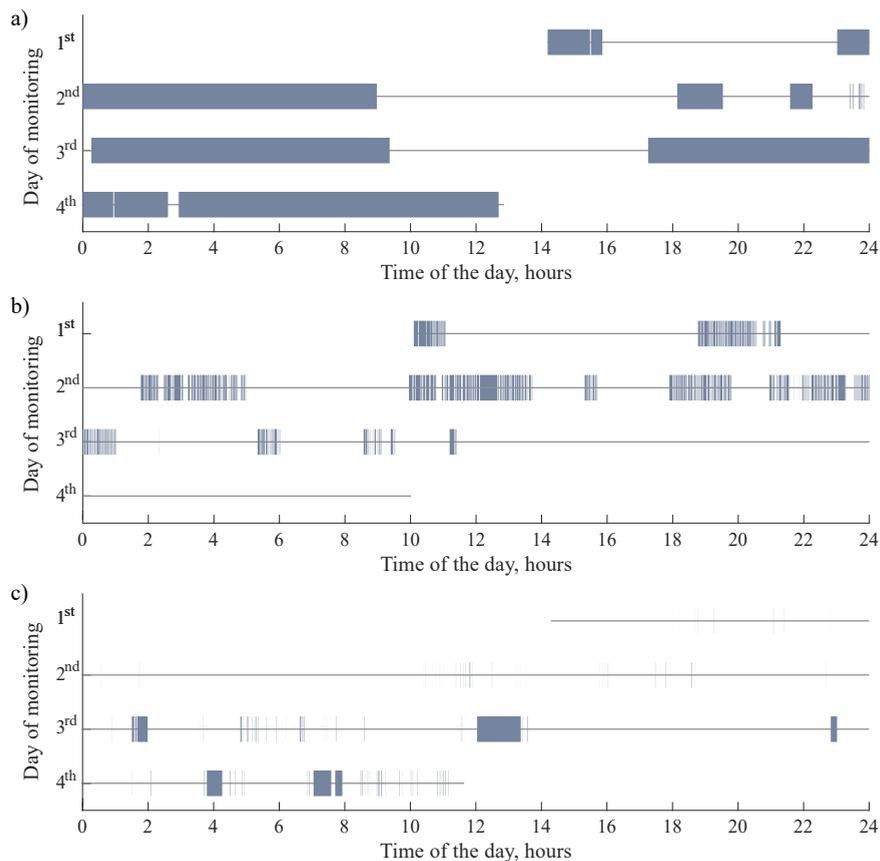


Figure 4.16. AF patterns from SPAFDB consisting of different number of AF episodes: a) 20 episodes which take 59.39% of the total monitoring time, b) 533 episodes which take 15.19% of the total monitoring time, and c) 151 episodes which take 5.51% of the total monitoring time. The blue lines show AF episodes

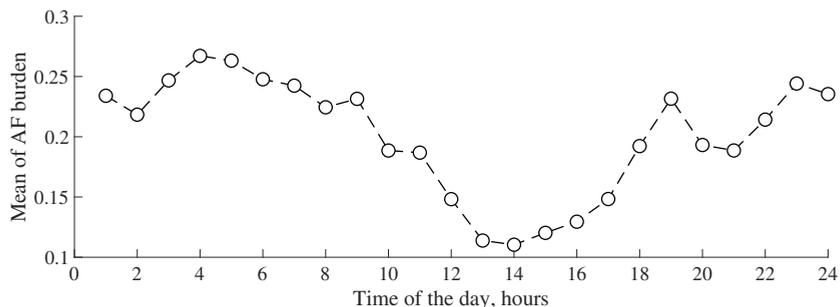


Figure 4.17. Circadian variation of AF burden. Results are obtained from 66 day-long AF patterns from 19 patients from SPAFDB. Note, only these AF patterns which take the whole day, i.e., from 00:00 to 23:59, are included in the analysis

4.3.2. Investigation of atrial fibrillation pattern reconstruction

Since the AF pattern can vary considerably (e.g., see Fig. 4.16), it is essential to understand how well AF episode patterns can be captured by using different AF detectors and how false detection influences the AF pattern characterizing parameters.

Figure 4.18 shows two examples of AF patterns, one with brief AF episodes, and the other one with longer episodes. The AF pattern with brief AF episodes (Fig. 4.18, left column) is best captured by the rhythm- and morphology-based detector. The DL-based detector has the highest sensitivity, however, the detector-produced pattern differs from the reference pattern since a few consecutive episodes are merged into a single episode. Therefore, the AF pattern cannot be captured properly, i.e., the AF burden for the reference AF pattern is 0.56, while the AF burden resulting from the DL-based detector is 0.81. For comparison, the AF burden resulting from the rhythm-based and the rhythm- and morphology-based detectors is 0.57 and 0.53, respectively. Also, in this case, the information about AF episode clustering is lost.

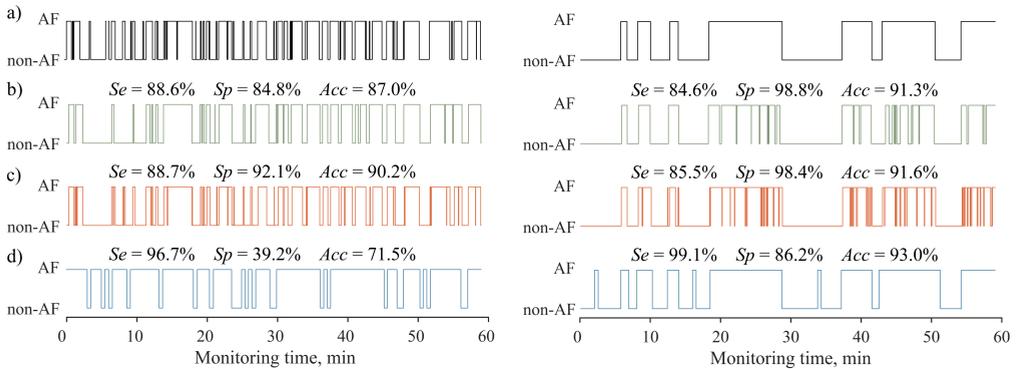


Figure 4.18. a) Reference AF pattern with brief (left column) and longer (right column) episodes and detector-produced patterns by using b) rhythm-based, c) rhythm- and morphology-based, and d) DL-based detector

On the contrary, when processing the AF pattern with longer episodes (Fig. 4.18, right column), both the rhythm-based and the rhythm- and morphology-based detectors tend to split a single AF episode into a cluster, while the DL-based detector does not do so since it processes ECG segments. This result also influences the AF pattern, i.e., the AF burden for the reference AF pattern is 0.53, while the AF burden for the detector-produced patterns by using the rhythm-based, the rhythm- and morphology-based, and the DL-based detectors are 0.45, 0.46, and 0.59, respectively.

Detector-based patterns consisting of short false alarms may have a different clinical meaning and may carry different information on AF progression. In the given example in Fig. 4.19, the annotated pattern is contaminated with false alarms resulting in a slightly larger AF burden \mathcal{B} , while the influence of false alarms on the parameters

used for pattern characterization (i.e., β_1 , μ , and \mathcal{A}) is considerably larger. Since AF episodes as well as false alarms are short, the influence on the Gini coefficient is small in this particular example.

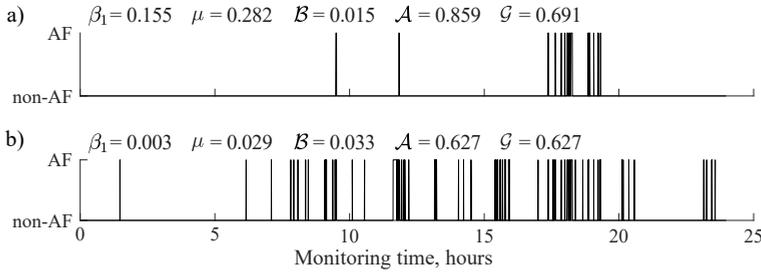


Figure 4.19. Influence of false alarms on parameters characterizing AF pattern: a) annotated AF pattern from LTAfDB and b) detector-based pattern

To shed further light on the problems regarding the reconstruction of the AF pattern, Fig. 4.20 shows differences in the clustering parameter β_1 , the rhythm dominance parameter μ , the AF burden \mathcal{B} , the aggregation \mathcal{A} , and the Gini coefficient \mathcal{G} obtained from the reference (annotated) and detector-based patterns. In most cases, the AF burden of the detector-based pattern is similar to that of the annotated pattern ($r = 0.97$). Conversely, the aggregation values computed for detector-based patterns differ from the annotated one ($r = 0.89$), which means that false alarms, undetected episodes, and fractured episodes have a large influence on the reconstructed pattern. The most challenging are the AF patterns with highly aggregated episodes since undetected and falsely detected episodes distort the pattern to the extent that it becomes substantially different from the annotated one. An ever worse situation is with the β_1 parameter, the μ parameter, and the Gini coefficient, as the influence of falsely detected episodes is huge ($r = 0.14$, $r = 0.64$, and $r = 0.39$, respectively).

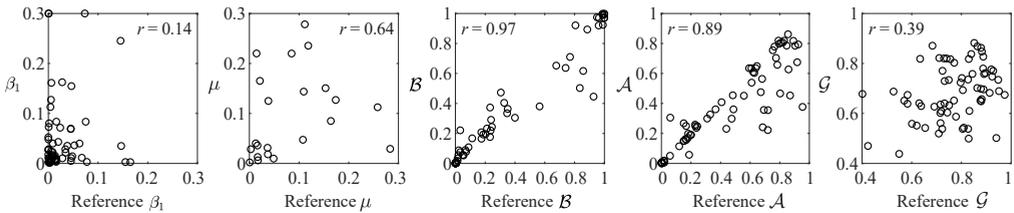


Figure 4.20. Detector-based reconstruction of AF patterns expressed as episode clustering parameter β_1 , rhythm dominance parameter μ , AF burden \mathcal{B} , aggregation \mathcal{A} , and Gini coefficient \mathcal{G} . The sample Pearson cross-correlation coefficient r is given in each plot. The results are obtained by using the rhythm-based detector on SPAfDB, AFDB, and LTAfDB databases. AF patterns with less than 10 episodes were excluded, resulting in a total of 69 patterns

Discussion. In the current clinical practice, AF has to be documented before the treatment, e.g., anticoagulation, can be started. Thanks to the advancements in the sensor technology, the ECG can today be registered over an extended time period and analyzed for the purpose of understanding the AF progression. The analysis should expand beyond the AF burden and provide more detailed characterization of the temporal AF episode patterns, e.g., whether the episodes are clustered or distributed evenly across the monitored period. Such information may be used to understand the significance of AF triggers and the development of such complications as stroke. However, more sophisticated analysis of episode patterns implies higher demands on AF detection performance. As illustrated by Fig. 4.19, annotated patterns differ considerably from those produced by the detector, and, as a result, β_1 , μ , \mathcal{A} , and \mathcal{G} differ considerably as well (Fig. 4.20) since these parameters account for the AF pattern (i.e., clustering of episodes, temporal distribution of episodes, and episode duration). In contrast, the AF burden \mathcal{B} obtained from the detector-produced patterns highly correlate with the reference ($r = 0.97$), however, it only evaluates the time spent in AF.

Unfortunately, episode analysis is made difficult in recordings containing noisy segments as AF detection becomes unreliable. Rather than simply discarding such segments from further analysis, as it is often done, future research should focus on improving the electrode technology and algorithms for signal processing and machine learning so that to ensure more reliable reconstruction of the AF pattern which will lead to the more sophisticated characterization of the episode patterns. So far, information about the reliability of the AF pattern reconstruction is still lacking.

Parts of Sec. 4.3.2 have been quoted verbatim from the previously published articles: [9, 11].

4.4. Atrial fibrillation pattern relationship with atrial echocardiographic parameters

Here, by using SPAFDB, the main model parameters (the parameter accounting for episode clustering β_1 and rhythm dominance μ), the AF burden \mathcal{B} , the aggregation \mathcal{A} , and the Gini coefficient \mathcal{G} are analyzed with the aim to determine whether they are correlated with echocardiographic measurements, such as the LA volume and the LA strain, reflecting the mechanical atrial performance.

Figure 4.21 presents scatter plots for the AF pattern characterizing parameters and echocardiographic measurements. As it is shown, β_1 , μ , and the Gini coefficient are weakly correlated with the LA volume ($r = 0.34$, $r = 0.19$, and $r = 0.09$, respectively), whereas the AF burden and the aggregation is somewhat more correlated ($r = 0.46$ and $r = -0.50$, respectively). Meanwhile, the AF strain is somewhat more correlated with μ and the Gini coefficient ($r = -0.48$ and $r = -0.57$, respectively).

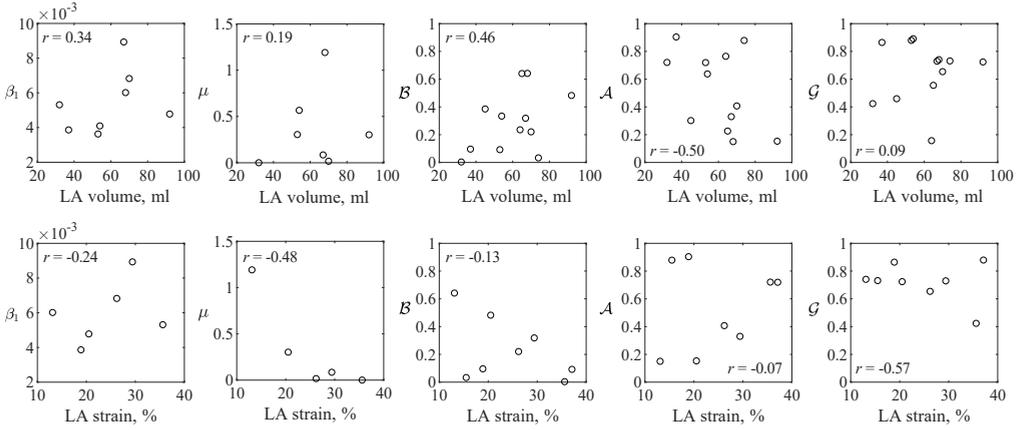


Figure 4.21. Association between AF pattern characterizing parameters and echocardiographic measurements (i.e., LA volume and LA strain). β_1 and μ are model-based parameters related to episode clustering and rhythm dominance, respectively; while B is burden, A – aggregation, and G – Gini coefficient. The sample Pearson cross-correlation coefficient r is given in each plot. The results are obtained by using SPAFDB

Discussion. Although the results are not conclusive due to the small sample size, they still shed some light on this research line. Atrial structural remodeling is associated with the changes in AF characteristics, often manifested as episodes of an increasing duration (i.e., the increasing μ and the AF burden B). Therefore, μ and the AF burden B may reflect the degree of atrial electrical and structural remodeling. Moreover, the temporal distribution of AF episodes, evaluated by aggregation A , may be related to the reduction of flow velocity in the left atrial appendage which is associated with an increased risk of thrombus formation [26]. However, this relation was not investigated since the flow velocity of the left atrial appendage was not available in the database.

Nevertheless, further investigation in larger databases is still required in order to confirm the clinical value of the AF pattern characterizing parameters (i.e., β_1 , μ , A , and G). A prospective study of the association between the AF pattern, not limiting to the AF burden or the AF stages (i.e., paroxysmal, persistent, permanent), and stroke would help in the identification of the potential risk factors in AF. Due to the aging population and the increasing prevalence of AF, it is one of the main challenges that public health systems are facing today.

Parts of Sec. 4.4 have been quoted verbatim from the previously published article: [12].

4.5. Conclusions of the chapter

1. Three clinical databases were used to investigate three types of AF detectors. Performed investigation revealed the problem related to data imbalance, i.e., *Acc* being a popular measure in AF detection tends to inflate the performance for imbalance data. Thus, *Mcc* should be considered instead. Also, the detection performance can differ depending on the approach taken to compare the detector output to annotations. In order to characterize AF patterns, episode-to-episode comparison should be used since it provides information on how well the episodes are captured.
2. Simulated recordings helped to investigate the influence of the ECG signal and the AF pattern properties on detection performance. The influence of the ECG signal properties depends on the detector structure: i.e., the detection accuracy of the DL-based detector depends on the ECG morphology; only the rhythm- and morphology-based detector does not depend on the APB rate; also, the DL-based detector is less dependent on noise. Meanwhile, the investigation of the AF pattern properties showed that AF patterns with a high AF burden and brief episodes imply higher demands on the detection performance.
3. The results of the distribution-based pattern characterization suggest that this type of method is less suitable for the day-long pattern characterization. Even if AIC and BIC assigned the majority of patterns to lognormal distribution, the Anderson-Darling test showed that 51%, 66%, and 94% of the episode, inter-episode, and inter-detection intervals, respectively, do not fit to any of the distributions under analysis.
4. The alternating, bivariate Hawkes model is well suited to produce a wide range of the AF patterns accounting to AF episode clustering. The main model parameters (i.e., β_1 accounting to episode clustering and μ accounting to rhythm dominance) can be used for AF pattern characterization. These parameters convey information complementary to the AF burden ($r = -0.16$ and $r = 0.24$, respectively).
5. A combination of the parameters suggested for AF pattern characterization (i.e., aggregation and the Gini coefficient) allows to analyze different types of AF patterns and can be used for characterization. The aggregation is negatively correlated with the AF burden ($r = -0.90$), however, the correlation decreases for the decreasing AF burden, e.g., the correlation decreases to -0.70 , -0.59 , -0.50 , -0.28 , and -0.20 for AF patterns with the AF burden less than 0.5, 0.4, 0.3, 0.2, and 0.1, respectively. The Gini coefficient is independent of the AF burden ($r = 0.07$).

6. The circadian analysis of AF patterns showed that AF is dominated at night, and there is a peak at around 19:00 o'clock.
7. AF pattern capturing is challenging since false alarms and misdetected episodes distort AF patterns, and this exerts influence on the pattern characterizing parameters. The correlation among the parameters obtained from the annotated and detector-based patterns is 0.14, 0.64, 0.97, 0.89, 0.39 for β_1 , μ , the AF burden, the aggregation, and the Gini coefficient, respectively. Thus, more sophisticated analysis of episode patterns implies higher demands on the AF detection performance. Further research should focus on ensuring reliable recognition of AF patterns.
8. Despite the small database, the AF pattern characterizing parameters were investigated together with echocardiographic measurements. However, only the AF burden and the aggregation is somewhat more correlated with the LA volume ($r = 0.46$ and $r = -0.50$, respectively), while the AF strain is somewhat more correlated with μ and the Gini coefficient ($r = -0.48$ and $r = -0.57$, respectively).

5. CONCLUSIONS

1. The model of the temporal distribution of paroxysmal AF episodes has been developed which accounts to AF episode clustering. The proposed model is based on the alternating, bivariate Hawkes process and can be used to model a wide range of episode patterns. An important feature of the model is that it is capable of estimating AF pattern characterizing parameters, i.e., the parameter β_1 is related to AF episode clustering, while the parameter μ indicates the dominance of AF vs. non-AF rhythm.
2. Three different approaches towards the characterization of the temporal distribution of paroxysmal AF episodes have been proposed. The distribution-based pattern characterization is less suitable to characterize AF patterns since 51%, 66%, and 94% of episode, inter-episode, and inter-detection intervals, respectively, do not fit to any of the distributions under investigation. The model-based parameters convey information complementary to the AF burden (the correlation between the AF burden and β_1 and μ is -0.16 and 0.24 , respectively). Even if the aggregation is negatively correlated with the AF burden ($r = -0.90$), a combination of the AF burden, aggregation, and the Gini coefficient allows to distinguish different types of the pattern (i.e., a single cluster, multiple clusters, dispersed episodes). The aggregation is better suited for discriminating the temporal AF patterns with a low AF burden (< 0.5), while the Gini coefficient is useful for discriminating patterns with a high AF burden and can thus be used as a complementary parameter to the aggregation.
3. Detection performance is influenced by the ECG signal properties (i.e., the ECG morphology, the number of atrial premature beats, the noise level) as well as the AF pattern properties (i.e., the AF burden and the episode length). For the AF pattern characterization, the detector-based pattern should preferably be compared with annotation by using the episode-to-episode comparison approach, while the detection performance should be evaluated in terms of Mcc , rather than Acc or F_1 , and Se , Sp , and PPV should be provided as well. More sophisticated analysis of episode patterns implies higher demands on the AF detection performance. The correlation among the parameters obtained from annotated and detector-based patterns is 0.14 , 0.64 , 0.97 , 0.89 , 0.39 for β_1 , μ , the AF burden, aggregation, and the Gini coefficient, respectively. Thus, future research should focus on the quality of capturing episode patterns.

SANTRAUKA

IVADAS

Tyrimo aktualumas

Prieširdžių virpėjimas (PV) – labiausiai paplitusi širdies aritmija visame pasaulyje. PV yra vyresnio amžiaus žmonių liga, todėl dėl sparčiai senstančios populiacijos PV paplitimas didėja [2]. 2010 m. duomenimis, Europos Sąjungoje PV sirgo 8,8 mln. vyresnių nei 55 metų asmenų, o iki 2060 m. šis skaičius gali padidėti iki 17,9 mln. PV sergantys asmenys yra labiau linkę sirgti gretutinėmis ligomis. Šiems asmenims yra 5 kartus padidėjusi insulto rizika, 3 kartus padidėjusi širdies nepakankamumo rizika ir 1,5–3,5 karto padidėjusi bendro mirtingumo rizika [3]. Su PV susijęs insultas arba širdies nepakankamumas lemia didesnę mirtingumą nei šios ligos atskirai [4]. Maždaug 30 % PV sergančių pacientų kasmet bent vieną kartą patenka į ligoninę, o 10 % – daugiau nei du kartus. Todėl 16–20 % PV pacientų kenčia nuo depresijos, o gyvenimo kokybė pablogėja daugiau nei 60 % pacientų.

PV yra progresuojanti liga, pradžioje pasireiškianti savaime nutrūkstančiais paroksizminiais PV epizodais. Paroksizminiai PV epizodai yra reti, trumpi ir dažnai besimptomiai, todėl neretai PV yra diagnozuojamas vėlesnių stadijų, ligai progresavus. PV progresavimas į nuolatinę formą (persistentinį arba permanentinį PV) yra siejamas su išaugusia širdies ir kraujagyslių ligų, hospitalizacijos ir mirties rizika [3, 5]. Gydomo sėkmė labai priklauso nuo to, kurioje aritmijos vystymosi stadijoje diagnozuojamas PV, todėl svarbu diagnozuoti PV dar pradinėje stadijoje, o tam reikalingos technologijos, kurios užtikrintų ilgalaikę stebėseną.

Naujausia technologijų pažanga užtikrina ilgalaikę pacientų stebėseną panaudojant įvairius dėvimus prietaisus, pvz., išmaniuosius laikrodžius. Ilgalaikis pacientų stebėjimas yra svarbus norint individualizuoti PV būsenos vertinimą [6]. Be to, ilgalaikis stebėjimas atveria galimybę charakterizuoti PV profilį atsižvelgiant į epizodų pasiskirstymą laike ir jų susigrupavimą į klasterius [7]. Naujausiose klinikinėse gairėse [3] pabrėžiamas PV profilių analizės poreikis, taip papildant dažnai naudojamą bendros PV trukmės įvertį. Tačiau iki šiol mažai žinoma apie PV profilių vaidmenį aritmijos progresavimui ir komplikacijų vystymuisi. Manoma, kad PV profilis gali būti susijęs su krešulių susidarymo rizika. Kadangi kraujo tėkmės greitis kairiojo prieširdžio ausytėje mažėja PV metu [8], daroma prielaida, kad krešulių susidarymo rizika yra didesnė, kai PV epizodai yra susigrupavę laike. Todėl PV profilių supratimas gali turėti įtakos individualizuoto gydymo valdymui ir komplikacijų (pvz., insulto) prognozavimui bei rizikos įvertinimui.

Atsižvelgiant į bendrą PV trukmę, epizodų skaičių, epizodų trukmę ir epizodų pasiskirstymą laike, PV profiliai yra labai skirtingi [7]. Tačiau trūksta metodų, kurie leistų visapusiškai charakterizuoti PV profilį. Taip pat, siekiant charakterizuoti PV

profilus, yra svarbu įvertinti, kaip tiksliai galima atkurti PV profilus naudojant automatinius atpažinimo algoritmus. Pastaruoju metu sparčiai kuriami įvairūs PV atpažinimo algoritmai, pvz., [36]–[52], dėl to kyla iššūkis, kaip tinkamai įvertinti ir palyginti esamų ir kuriamų detektorių patikimumą. Detektorių patikimumą vertinantys įverčiai turėtų būti papildomi tyrimais, vertinančiais PV profilio atkūrimo patikimumą. Pavyzdžiui, neaišku, kaip patikimai veikia PV detektoriai esant skirtingiems PV profilams, t. y. ar profilis su keliais trumpais PV epizodais yra atpažįstamas taip pat patikimai, kaip ir profilis, kuriame dominuoja ilgi epizodai. Vis dėlto esami tyrimai su PV detektoriais suteikia labai mažai informacijos apie PV profilio atkūrimo patikimumą.

Mokslinė ir technologinė problema bei darbinė hipotezė

Laikinis PV epizodų profilis gali būti susijęs su krešulių susidarymo rizika. Daroma prielaida, kad rizika yra didesnė, kai PV epizodai yra susigrupavę laike, nes PV metu kairiojo prieširdžio ausytėje kraujo srauto greitis lėtėja. Nors informacijos apie PV profilį trūksta, tikimasi, kad atsirandančios neinvazinės ilgalaikio stebėjimo technologijos padės užpildyti šią žinių spragą. Tačiau, norint iki galo išspręsti šią problemą, reikia metodų, skirtų PV profiliui charakterizuoti.

Mokslinė ir technologinė problema: Kokiais būdais galima charakterizuoti paroksizminio PV epizodų pasiskirstymą laike, kad būtų galima atskirti skirtingus PV profilų tipus, kurie yra svarbūs norint geriau suprasti PV profilio ryšį su krešulių susidarymo rizika?

Darbinė hipotezė: Paroksizminio PV profilus galima charakterizuoti atsižvelgiant į PV trukmę, PV epizodų pasiskirstymą laike ir susigrupavimą į klasterius, panaudojant statistinių skirstinių analize pagrįstą būdą, modeliu pagrįstą būdą ir parametrais pagrįstą būdą.

Tyrimo objektas – šiame darbe vystomi ir tiriama signalų apdorojimo algoritmai, skirti PV profiliui charakterizuoti.

Tyrimo tikslas – sukurti ir iširti prieširdžių virpėjimo profilų charakterizavimo algoritmus.

Tyrimo uždaviniai

1. Sukurti prieširdžių virpėjimo epizodų pasiskirstymo fenomenologinį modelį.
2. Sukurti prieširdžių virpėjimo epizodų pasiskirstymo laike charakterizavimo algoritmus.
3. Iširti prieširdžių virpėjimo atpažinimo algoritmų savybių įtaką prieširdžių virpėjimo profilį charakterizuojantiems parametrams.

Mokslinis naujumas

Šioje daktaro disertacijoje pristatomi PV atpažinimo algoritmų tyrimų rezultatai, kurie yra būtini siekiant atskleisti aspektus, susijusius su patikimu PV profilių atkūrimu. Darbe pateikiamos skirtingų PV detektorių tipų (t. y. ritmo analize, ritmo ir morfologijos analize pagrįstų detektorių ir giliojo mokymosi detektoriaus) silpnybės ir stiprybės. Taip pat pabrėžiami PV profilio atkūrimo iššūkiai ir pateikiamos rekomendacijos, kaip tvarkyti PV duomenis ir įvertinti PV detektorių patikimumą.

Pasiūlyti ir ištirti trys būdai PV profiliui charakterizuoti. Vienas iš jų yra pagrįstas statistinių skirstinių analize, kuri remiasi prielaida, kad epizodai yra statistiškai nepriklausomi. Kadangi PV epizodai yra linkę susigrupuoti laike į klasterius, ši prielaida kelia abejonių dėl šio būdo tinkamumo įvertinti PV profilius. Taip pat šis būdas yra mažiau tinkamas trumpiems (pvz., dienos trukmės) PV profiliams analizuoti dėl nedidelio epizodų skaičiaus. Kitas būdas – naudoti parametrus, gautus iš PV profilių modelio, kuris pagrįstas kintamu dvimačiu Hawkeso procesu. Modelio parametrai suteikia informaciją apie PV epizodų klasterius ir ritmo dominavimą (t. y. PV ar ne). Paskutinis būdas – panaudoti įvairius parametrus, tokius kaip santykinė bendra PV trukmė, agregacija ir Gini koeficientas. Santykinė bendra PV trukmė yra gerai žinomas parametras, naudojamas PV tyrimuose, tačiau jis suteikia informaciją tik apie bendrą PV trukmę. Agregacija įvertina PV epizodų pasiskirstymą laike, o Gini koeficientas, gerai žinomas ekonomikoje, suteikia informaciją apie epizodų trukmių nelygę. Kombinuojant modelio parametrus su pasiūlytais papildomais parametrais (santykine bendra PV trukme, agregacija ir Gini koeficientu), galima analizuoti skirtingus PV profilių tipus – sudarytus iš vieno klasterio, iš kelių klasterių ar epizodų, pasiskirsčiusių per visą stebėjimo laikotarpį.

Praktinė reikšmė

1. Atlikti PV detektorių tyrimai bei pasiūlyti PV profilių charakterizavimo spendimai yra svarbūs:
 - (a) Atskleisti iššūkiai susiję su PV atpažinimo algoritmų kūrimu.
 - (b) Pasiūlytos rekomendacijos, kaip tvarkyti PV duomenis ir įvertinti PV detektorių patikimumą.
 - (c) Pasiūlytas modelis, skirtas PV profiliams modeliuoti, yra tinkamas charakterizuoti PV profilį panaudojant modelio parametrus, kurie suteikia informaciją apie epizodų susigrupavimą į klasterius ir ritmo dominavimą.
 - (d) Siūlomi PV profilių charakterizavimo algoritmai gali pagerinti aritmijos progresavimo supratimą.
 - (e) Siūlomi PV profilių charakterizavimo algoritmai gali būti panaudojami norint suprasti ryšį tarp PV profilio ir komplikacijų rizikos (pvz., krešulių susidarymo ar išeminio insulto rizikos).

(f) Siūdomi PV profilių charakterizavimo algoritmai gali būti panaudojami identifikuojant potencialius PV trigerius.

2. Šioje daktaro disertacijoje pateikiamiems metodams plėtoti galimybes sudarė Lietuvos mokslo tarybos mokslininkų grupių projektas „Poinsoulinės būklės pacientų prieširdžių aritmijų ilgalaikės netrukdančios stebėsenos metodai – AFterStroke“ (S-MIP-17/81), 2017–2019 m.
3. Šiuo metu sukurti metodai taikomi Europos Sąjungos struktūrinių fondų finansuojamame projekte „Personalizuotas paroksizminio prieširdžių virpėjimo trigerių atpažinimas ir valdymas naudojant dėvimas technologijas – TriggersAF“ (01.2.2-LMT-K-718-03-0027), 2020–2023 m.

Tyrimo rezultatų aprobavimas

Daktaro disertacija remiasi dviem pagrindiniais straipsniais, publikuotais tarptautiniuose moksliniuose žurnaluose, turinčiuose cituojamumo rodiklį „Clarivate Analytics Web of Science“ duomenų bazėje. Iš viso rezultatai publikuoti septyniuose moksliniuose straipsniuose. Sukurti sprendimai buvo pristatyti trijose tarptautiniuose pripažintose konferencijose: 45-oje, 47-oje ir 48-oje konferencijoje „Computing in Cardiology“.

Atlikti tyrimai buvo teigiamai įvertinti 48-oje konferencijoje „Computing in Cardiology 2021“ (Brno, Čekija), pristatytas tyrimas pateko tarp pusfinalio dalyvių Rosanna Degani jaunuųjų tyrėjų apdovanojimo konkurse. 2019 m., 2020 m. ir 2022 m. gauta Lietuvos mokslo tarybos stipendija už studijų rezultatus. 2018 m., 2019 m. ir 2021 m. gauta aktyviausio KTU doktoranto stipendija Elektros ir elektronikos inžinerijos studijų kryptyje.

Ginti teikiami teiginiai

1. PV atpažinimo algoritmų patikimumo įvertinimas priklauso nuo metodo, pasirinkto siekiant palyginti detektoriaus išėjimą su anotacijomis, bei nuo pasirinktų parametru, įvertinančių detektoriaus veikimą. Norint charakterizuoti PV profilį, reikėtų naudoti epizodų palyginimo metodą, o detektorius veikimas turėtų būti įvertinamas naudojant Matthewso koreliacijos koeficientą Mcc , o ne diagnostinį tikslumą Acc .
2. PV profilio atkūrimo patikimumas priklauso nuo elektrokardiogramos (EKG) signalo savybių (t. y. EKG morfologijos, priešlaikinių prieširdžių susitraukimų skaičiaus, triukšmo lygio) bei PV profilio savybių (t. y. bendros PV trukmės, epizodų ilgio). Taip pat PV profilio atkūrimo patikimumas skiriasi priklausomai nuo PV detektoriaus struktūros, t. y. giliojo mokymosi detektorius linkęs sujungti keletą trumpų greta esančių epizodų į vieną, o ritmo analize ir ritmo

bei morfologijos analize pagrįsti detektoriai suskaito vieną epizodą į epizodų klasterį.

3. PV profiliai gali būti charakterizuojami trimis būdais: panaudojant statistinius skirstinius, PV profilių modelį ir pasiūlytus parametrus. Statistinių skirstinių analizė yra mažiau tinkamas būdas dienos trukmės profiliams charakterizuoti, taip pat šis būdas remiasi prielaida, kad epizodai yra statistiškai nepriklausomi. Modelio parametrų derinys su papildomais pasiūlytais parametrais leidžia analizuoti skirtingus PV profilių tipus, nes šie būdai suteikia papildomos informacijos apie PV profilį, t. y. modeliu pagrįsti parametrai suteikia informacijos apie epizodų klasterius ir ritmo dominavimą, o parametrais pagrįstas charakterizavimas suteikia informacijos apie bendrą PV trukmę, PV epizodų pasiskirstymą laike ir epizodų trukmių netolygumą.

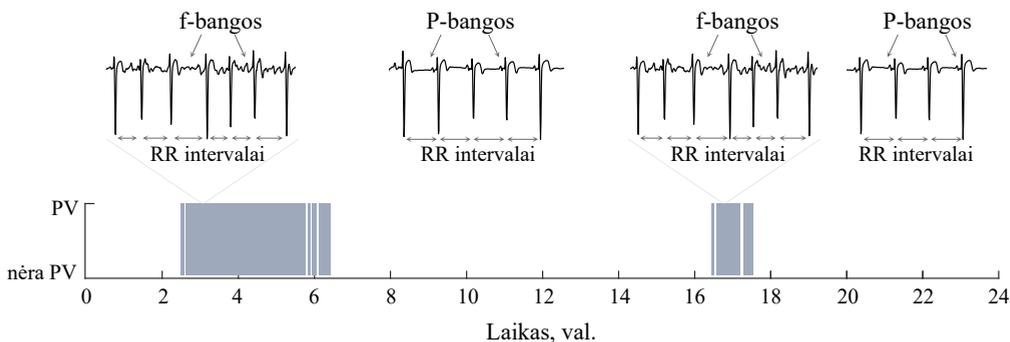
Bendradarbiavimas

Pasiūlytas modelis, skirtas PV profiliams modeliuoti ir PV profilį charakterizuojantiems parametrams įvertinti, yra sukurtas bendradarbiaujant su Lundo universitetu (Lundas, Švedija). Šiam tikslui pasiekti buvo atlikta trijų savaičių stažuotė Lundo universitete (nuo vasario 25 d. iki kovo 16 d., 2019 m.), kurios metu aptarta modelio koncepcija ir įgyvendinimo galimybės.

Tyrimams atlikti naudojami PV atpažinimo algoritmai yra sukurti Andriaus Petreno (ritmo analize pagrįstas detektorius ir ritmo ir morfologijos analize pagrįstas detektorius) ir Andriaus Sološenko (giliojo mokymosi detektorius).

1. PRIEŠIRDŽIŲ VIRPĖJIMO PROFILIŲ CHARAKTERIZAVIMO APŽVALGA

PV įprastai diagnozuojamas iš EKG signalo, kuriame matoma nereguliari skilvelių veikla (RR intervalai), o normaliam sinusiniam širdies ritmui būdingas P-bangas pakeičia nepertraukiamos chaotiškos virpėjimą charakterizuojančios f-bangos (1.1 pav.). Savaimė nutrukstančių paroksizminio PV epizodų pasiskirstymas laike vadinamas PV profiliu, kuris gali būti labai įvairus. PV profiliai skiriasi atsižvelgiant į bendrą PV trukmę, epizodų skaičių, epizodų trukmę ir epizodų pasiskirstymą laike. Pavyzdžiui, PV profilis gali susidaryti iš kelių laike susikoncentravusių epizodų, taip sudarydamas klasterius (1.1 pav.).



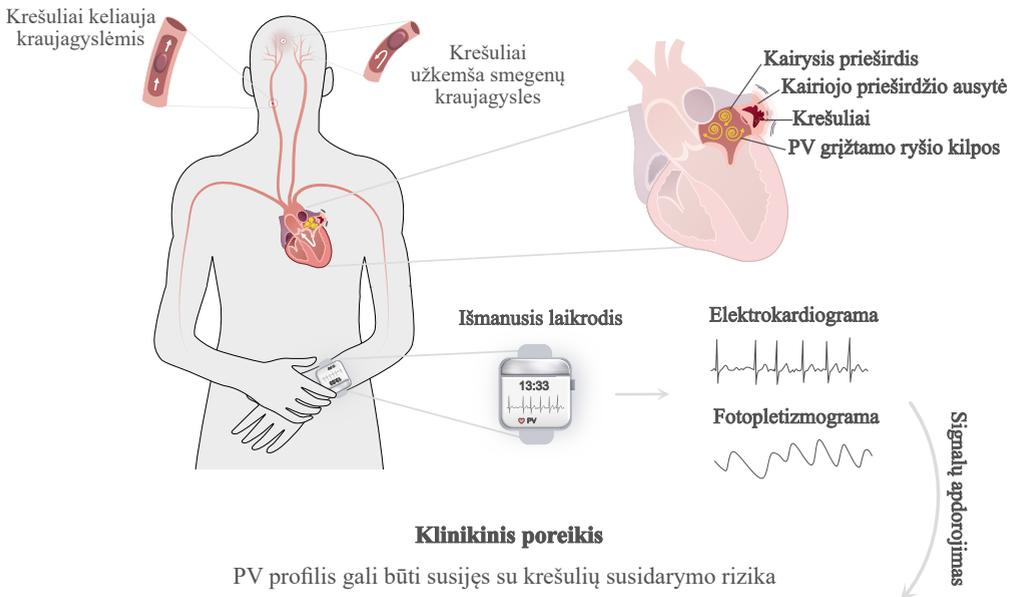
1.1 pav. PV profilis su dviem epizodų klasteriais. Aukštas lygis žymi PV, o žemas lygis žymi, kad nėra PV (sinusinis ritmas ar kito tipo aritmijos). PV metu EKG signalė P-bangos pakeičia nepertraukiamos virpėjimą charakterizuojančios f-bangos, o RR intervalai tampa nereguliarūs

PV yra siejamas su padidėjusia išeminio insulto rizika [4], žr. 1.2 pav. Prieširdžio kairioji ausytė yra viena pagrindinių širdies dalių, kurioje formuojasi krešuliai PV metu [25]. Daugiau nei 90 % išeminių insultų kyla dėl krešulių, susiformavusių ausytėje [26]. Sinusinio ritmo metu kraujas reguliariai pasišalina iš ausytės, tačiau PV metu kraujo srautas prieširdžio kairiojoje ausytėje tampa nereguliarus, sulėtėja ausytės išsivalymo srauto greitis, o būtent tai ir didina krešulių susidarymo riziką [8]. Progresuojant PV, ausytės išsivalymo srauto greitis lėtėja. Keliama klinikinė hipotezė, kad PV profilis gali būti susijęs su krešulių susidarymo rizika [25], t. y. profilis su epizodais, susikoncentravusiais laike, yra siejamas su didesne rizika (1.2 pav.). Vis dėlto šiandien trūksta informacijos apie PV profilį ir metodų, kurie leistų charakterizuoti PV profilį.

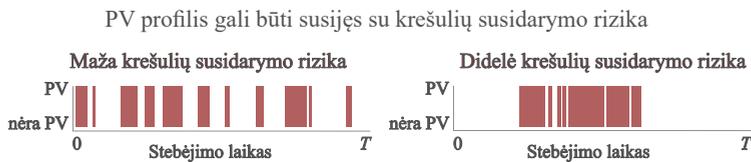
PV profilio charakterizavimo problema buvo iškelta dar praeitame šimtmetyje, tuo tikslu buvo atlikti keli tyrimai analizuojant statistinius skirstinius [92]–[95]. Pastaruosius 10 metų ypač suaktyvėjo tyrimai, analizuojant sąsają tarp bendros PV trukmės ir krešulių susidarymo rizikos [76]. Tačiau, nepaisant didelio mokslininkų susido-

mėjimo, bendros PV trukmės ir krešulių susidarymo sąsaja šiuo metu yra vertinama prieštaringai ir nėra bendrai priimto sprendimo, kokia bendra PV trukmė ar ilgiausio epizodo trukmė yra reikšminga [24,33]. Naujausiose klinikinėse gairėse pabrėžiamas poreikis analizuoti ne tik bendrą PV trukmę, bet ir patį PV profilį, pvz., įvertinant epizodų susikoncentravimą tam tikrame laiko intervale [3].

Krešulių susidarymo rizika



Klinikinis poreikis



1.2 pav. PV ryšys su krešulių susidarymo rizika, ilgalaikių PV profilių rinkimas ir PV profilių analizės klinikinis poreikis

2. PRIEŠIRDŽIŲ VIRPĖJIMO PROFILIŲ ATPAŽINIMAS IR CHARAKTERIZAVIMAS

2.1 Prieširdžių virpėjimo atpažinimo algoritmai

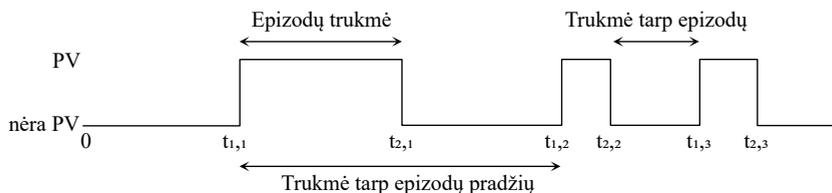
Šiame darbe naudojami trijų tipų PV atpažinimo algoritmai. Rimto analize pagrįstas detektorius grindžiamas prielaida, kad PV metu sutrinka reguliarus sinusinis širdies ritmas ir padidėja vidutinis širdies susitraukimų dažnis. Algoritmas analizuoja tik trukmių tarp gretimų skilvelių susitraukimo sekas (RR intervalus) [56]. R danteliams atpažinti naudojamas [99] detektorius. Antrasis yra rimto ir morfologijos analize pagrįstas detektorius, sukurtas trumpiems (< 30 sek.) PV epizodams atpažinti. Jo įėjimą sudaro keturi parametrai: RR nereguliarumas įvertinamas prieš tai aprašytu detektoriumi, P-bangų išnykimas, f-bangų atsiradimas ir triukšmo lygis [66]. Trečiasis detektorius, skirtingai nei pirmieji du, nereikalauja iš anksto apibrėžtų požymių. Tai giliojo mokymosi detektorius, kuris naudoja 1D konvoliucinį neuroninį tinklą 30 sek. EKG segmentams apdoroti (V_1 derivacija) [9].

2.2 Prieširdžių virpėjimo profilių charakterizavimas

PV profilis gali būti charakterizuojamas trimis skirtingais būdais. Pirmasis būdas yra pagrįstas statistinių skirstinių analize. Antrasis būdas yra pagrįstas modeliu, skirtu PV profiliui modeliuoti, o trečiasis būdas yra panaudoti įvairius parametrus.

2.2.1 Skirstinių analizė

Skirstinių analize pagrįstas charakterizavimo būdas remiasi statistinių skirstinių pritaikymu prie duomenų. PV profilių analizėje dažniausiai naudojami keturi skirstiniai: normalusis, lognormalusis, eksponentinis ir Weibullio [93, 94]. Šiame darbe analizuojamos PV epizodų trukmės, trukmės tarp gretimų epizodų ir trukmės tarp gretimų epizodų pradžių, kurios gaunamos iš PV profilio (2.1 pav.). Šių trukmių histogramos yra pritaikomos prie kiekvieno iš skirstinių, ištraukiami parametrai, apibūdinantys tikimybių tankio funkciją, ir įvertinama skirstinio pritaikymo kokybė.



2.1 pav. PV profilis su pažymėta epizodų trukme, trukme tarp gretimų epizodų ir trukme tarp gretimų epizodų pradžių. Laikai $t_{1,1}, t_{1,2}, t_{1,3}, \dots$ žymi PV epizodo pradžią, o $t_{2,1}, t_{2,2}, t_{2,3}, \dots$ žymi PV epizodo pabaigą

Skirstinių tikimybių tankio funkcija gali būti aprašoma keturiais parametrais (vietos, mastelio, formos ir slenksčio), kurių skaičius kinta priklausomai nuo skirstinio tipo: normalusis ir lognormalusis skirstinys aprašomas dviem parametrais, vidurkiu ir standartiniu nuokrypiu; eksponentinis skirstinys aprašomas mastelio parametru, nurodančiu eksponentės slopimo greitį; Weibullio skirstinys aprašomas formos ir mastelio parametrais. Parametrus ištraukti naudojamas didžiausio tikėtino metodo, kurio tikslas yra rasti tokius parametrus, su kuriais tikėtino funkcija yra didžiausia.

Siekiant didesnio rezultatų patikimumo, skirstinių prisitaikymas prie trukmių histogramų įvertinamas naudojant tris metodus: Andersono-Darlingo testą [104], Akaike'ės ir Bayesian'o informacijos kriterijus [105]. Andersono-Darlingo testas – statistinis testas, kurio metu keliami nulinė hipotezė, kad PV profilių trukmių histograma yra pasiskirsčiusi pagal nurodytą skirstinį. Lyginant skirstinius tarpusavyje, pasirenkamas tas, kuris turi didžiausią p vertę, o skirstiniai, kurių $p \leq 0,05$ (reikšmingumo lygmuo 0,05), yra atmetami. Akaike'ės ir Bayesian'o informacijos kriterijai remiasi didžiausio tikėtino funkcija, šiais atvejais parenkamas tas skirstinys, kuris turi mažiausią įvertį. Kitų skirstinių atmetimo reikšmingumas gali būti įvertinamas remiantis prielaida: jei skirtumas tarp išskirto skirstinio informacijos kriterijaus vertės ir verčių iš kitų skirstinių yra < 2 , tada nėra pagrindo išskirti vieno skirstinio, pagal kurį pasiskirsčiusios trukmių histogramos.

2.2.1 Prieširdžių virpėjimo profilių modelio panaudojimas

PV profiliai gali būti charakterizuojami panaudojant PV profilių modelį [10]. Modelis remiasi kintamu dvimačiu Hawkeso procesu ir yra trumpai aprašytas žemiau.

Modelio aprašymas. PV profilis yra modeliuojamas panaudojant du taškinis procesus: $N_1(t)$ atsižvelgia į perėjimą iš sinusinio ritmo (SR) į PV laiko momentais $t_{1,1}, t_{1,2}, \dots$, o $N_2(t)$ atsižvelgia į perėjimą iš PV į SR laiko momentais $t_{2,1}, t_{2,2}, \dots$. Šis dvimatis taškinis procesas yra aprašomas sąlyginėmis intensyvumo funkcijomis $\lambda_1(t)$ ir $\lambda_2(t)$ [106]:

$$\lambda_m(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(N_m(t + \Delta t) - N_m(t) = 1 | H_t)}{\Delta t}, \quad (2.1)$$

čia skaitiklis yra sąlyginė tikimybė, kad perėjimas atsiras laiko intervale $[t, t + \Delta t]$, ir H_t yra dvimačio taškinio proceso istorija, t. y. perėjimų laiko momentai $t_{1,1}, t_{2,1}, t_{1,2}, \dots$, kurie atsirado iki laiko t .

Panaudojant dvimatį Hawkeso procesą [107], šie du taškiniai procesai $N_1(t)$ ir $N_2(t)$ su sąlyginėmis intensyvumo funkcijomis $\lambda_1(t)$ ir $\lambda_2(t)$ yra aprašomi:

$$\lambda_m(t) = \mu_m \sum_{n=1}^2 \sum_{\{k:t>t_{n,k}\}} \alpha_{m,n} e^{-\beta_{m,n}(t-t_{n,k})}, \quad (2.2)$$

čia $\mu_m > 0$, $\alpha_{m,n} \geq 0$, $\beta_{m,n} \geq 0$, kai $m, n = 1, 2$, o $t_{n,k}$ prasidejimo laikai. Pagrindinė Hawkeso proceso savybė yra ta, kad kiekvienu laiko momentu, kai atsiranda naujas taškas (būsenos perėjimas), sąlyginė intensyvumo funkcija $\lambda_1(t)$ iškart padidėja, priklausomai nuo $\alpha_{1,1}$ (savaiminio sužadavimo savybė), ir tada mažėja eksponentiškai su parametru $\beta_{1,1}$ iki bazinio intensyvumo lygio μ_1 . Tas pats galioja ir $\lambda_2(t)$, tik šiuo atveju turime $\alpha_{2,2}$, $\beta_{2,2}$ ir μ_2 parametrus. Kadangi perėjimo tikimybė, kad atsiras kitas taškas, padidėja iškart, kai įvyksta būsenos perėjimas, šis modelis gali būti naudojamas modeliuojant PV epizodų klasterius. Modelyje taip pat yra kryžminio sužadavimo savybė, kuri leidžia $N_2(t)$ procesui paveikti $N_1(t)$ procesą, t. y. $\lambda_1(t)$ turi parametrus $\alpha_{1,2}$ ir $\beta_{1,2}$, o $\lambda_2(t) - \alpha_{2,1}$ ir $\beta_{2,1}$.

Dvimačio Hawkeso proceso trūkumas yra tas, kad jis neatsižvelgia į būsenos perėjimo seką, t. y. perėjimas iš SR į PV (PV epizodo pradžia) ne visada baigiasi perėjimu iš PV į SR (PV epizodo pabaiga). Tai iš principo netinka PV profiliams modeliuoti. Dėl šios priežasties $\lambda_1(t)$ ir $\lambda_2(t)$ yra padauginamos iš „įvykimo“ funkcijos, taip užtikrinant, kad PV atsiranda po SR, t. y. $t_{1,1} < t_{2,1} < t_{1,2} < t_{2,2} < \dots$:

$$o_1(t) = \begin{cases} 1, & N_1(t) = N_2(t - d_2), \\ 0, & \text{kitu atveju} \end{cases} \quad (2.3)$$

ir SR atsiranda po PV:

$$o_2(t) = \begin{cases} 1, & N_2(t) \neq N_1(t - d_2), \\ 0, & \text{kitu atveju,} \end{cases} \quad (2.4)$$

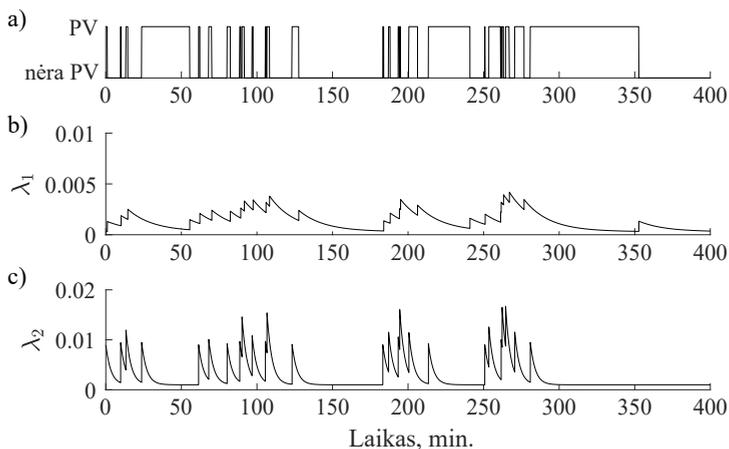
čia d_1 ir d_2 yra minimalios PV epizodo ir SR „epizodo“ trukmės. Šios trukmės užtikrina, kad, perėjus iš SR į PV, perėjimas iš PV į SR atsirastų tik tada, kai praeina laikas d_1 , ir t. t.

Apibendrinant, kintamo dvimačio Hawkeso proceso vizualizacija pateikta 2.1 pav., o sąlyginė intensyvumo funkcija aprašoma:

$$\Lambda_m(t) = \lambda_m(t)o_m(t), m = 1, 2. \quad (2.5)$$

Modelio parametrai. Priimant, kad $\beta_{1,1} = \beta_{1,2} = \beta_1$ ir $\beta_{2,1} = \beta_{2,2} = \beta_2$, modelis aprašomas 8 parametrais: μ_1 , $\alpha_{1,1}$, $\alpha_{1,2}$, β_1 aprašo $\Lambda_1(t)$ ir μ_2 , $\alpha_{2,1}$, $\alpha_{2,2}$, β_2 aprašo $\Lambda_2(t)$. Parametrams ištraukti iš realių duomenų yra naudojamas didžiausio tikėtimumo metodas. Jis reikalauja minimalaus epizodų skaičiaus, siekiant užtikrinti ištrauktų parametrų patikimumą. Šiuo atveju pasirinkta, kad profilis turi būti sudarytas bent iš 10 epizodų.

PV profiliui charakterizuoti pasirinkti du modelio parametrai. Pirmasis β_1 yra susijęs su epizodų susigrupavimu į klasterius, kadangi lėtas eksponentės mažėjimas (maža β_1 vertė) padidina tikėtimumą, kad po PV epizodo seks kitas PV epizodas.



2.1 pav. Kintamo dvimačio Hawkeso proceso realizacija: a) PV profilis; b) ir c) sąlyginės intensyvumo funkcijos $\lambda_1(t)$ ir $\lambda_2(t)$

Priešingai, didėjant β_1 parametro vertei, intensyvumo funkcija mažėja sparčiau, kas lemia, kad epizodai pasiskirsto laike. Antrasis parametras yra susijęs su bazinio intensyvumo μ_1 ir μ_2 parametrais, kurie nurodo vidutinį perėjimų dažnį. O μ_1 ir μ_2 santykis suteikia informaciją apie ritmo dominavimą, ar tai PV ($\mu > 1$), ar SR ($\mu < 1$).

2.2.1 Parametrai, charakterizuojantys prieširdžių virpėjimo profilį

Santykinė bendra PV trukmė. Santykinė bendra PV trukmė \mathcal{B} parodo, kokią laiko dalį visame įrašė (pvz., 24 val.) truko suminis virpėjimas, ir yra apskaičiuojama pagal:

$$\mathcal{B} = \frac{T_{AF}}{T}, \quad (2.6)$$

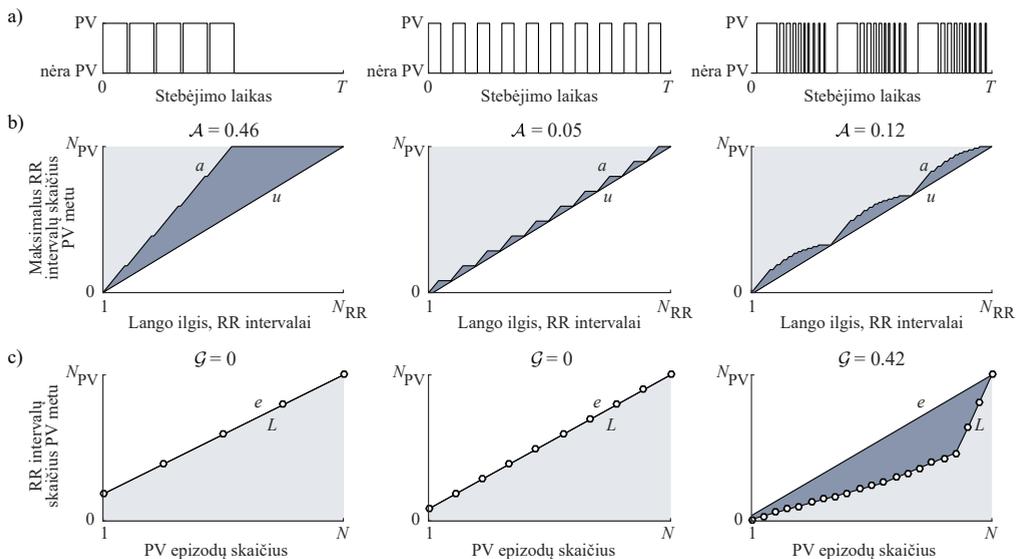
čia T_{AF} – suminė bendra PV epizodų trukmė, T – visas stebėjimo laikas. Priklausomai nuo virpėjimo trukmės, \mathcal{B} vertė gali kisti nuo 0 iki 1. PV profiliai, kuriuose epizodai užima beveik visą stebėjimo laikotarpį, nesvarbu, ar buvo daug trumpų epizodų, ar keli ilgi, įgyja parametro vertes, artimas 1, o PV profiliai, kuriuose vienas ar keli trumpi epizodai užima tik mažą dalį viso stebėjimo laikotarpio, įgyja vertes, artimas 0. Vis dėlto \mathcal{B} suteikia informaciją tik apie laiką, praleistą PV būsenoje, tačiau neatšivėlgia į laikinį epizodų pasiskirstymą.

Agregacija. Laikinis epizodų pasiskirstymas gali būti charakterizuojamas agregacijos parametru \mathcal{A} , kuris kiekybiškai įvertina nuokrypį tarp stebimo PV profilio bei teorinio tolygiai pasiskirsčiusio profilio. Agregacijai \mathcal{A} apskaičiuoti reikia RR intervalų sekos, kur kiekvienas RR intervalas priskiriamas PV (1) arba SR (0). Taip pat reikia apibrėžti realų a ir teorinį tolygų u PV epizodų trukmės pasiskirstymą laiko interva-

luose. Realus pasiskirstymas a gaunamas slenkant langą per dvejetainę seką (žingsnis lygus vienam RR intervalui) ir randant maksimalų RR intervalų skaičių, priskirtą PV. Lango ilgis pasirenkamas nuo 1 iki RR intervalų skaičiaus. Teorinis tolygus pasiskirstymas u atspindi PV profilį su tolygiai pasiskirsčiusiu PV per visą stebėjimo laikotarpį ir yra laikomas atraminiu. Toliau agregacija \mathcal{A} yra apskaičiuojama pagal:

$$\mathcal{A} = \frac{2}{N_{RR}N_{AF}} \sum_{n=1}^{N_{RR}} |a_n - u_n|, \quad (2.7)$$

čia N_{RR} – bendras RR intervalų skaičius, N_{AF} – RR intervalų skaičius PV metu. Žiūrint grafiškai, \mathcal{A} yra apibrėžiama kaip santykis tarp ploto, apriboto a ir u kreivėmis (tamsiai mėlynas plotas 2.2 pav., b), ir ploto virš kreivės u (tamsiai ir šviesiai mėlynas plotas 2.2 pav., b). Agregacijos \mathcal{A} vertė gali kisti nuo 0 iki 1. Vertės, artimos 1, rodo didelę agregaciją, būdingą PV profiliams, kuriuose yra vienas trumpas PV epizodas arba epizodų klasteris. Ir, atvirkščiai, vertės, artimos 0, rodo mažą agregaciją, būdingą PV profiliams, kuriuose epizodai pasiskirstę tolygiai per visą stebėjimo laikotarpį, arba PV profiliams, kuriuose yra vienas epizodas, trunkantis beveik visą stebėjimo laikotarpį.



2.2 pav. a) Skirtingi PV profiliai: vienas klasteris sudarytas iš vienodos trukmės epizodų (kairėje), epizodai pasiskirstę tolygiai (viduryje) ir keli klasteriai su skirtingomis epizodų trukmėmis (dešinėje). Grafinė iliustracija b) agregacijos \mathcal{A} ir c) Gini koeficiento \mathcal{G} . Pastaba: santykinė bendra PV trukmė yra vienoda visuose profiliuose ($\mathcal{B} = 0,5$)

Gini koeficientas. PV profilis gali būti charakterizuojamas atsižvelgiant į PV epizodų trukmę, nes epizodai gali trukti nuo kelių sekundžių iki kelių valandų ar dienų. Epizodų trukmių skirtumas yra įvertinamas Gini koeficientu \mathcal{G} , kuris yra gerai žinomas parametras ekonomikoje, įvertinantis pajamų nelygybę [111]. Gini koeficientui apskaičiuoti reikia epizodų skaičiaus N ir jų trukmių, kurias panaudojant randamos absoliučios lygybės e ir Lorenco L kreivės. Absoliučios lygybės kreivė e atspindi PV profilį, sudarytą iš vienodos trukmės epizodų, o Lorenco kreivė L atspindi kaupiamąją epizodų trukmės sumą, kai epizodai išrikiuoti didėjančia tvarka.

Turint absoliučios lygybės e ir Lorenco L kreives, Gini koeficientas \mathcal{G} apskaičiuojamas pagal:

$$\mathcal{G} = \frac{2}{NN_{AF}} \sum_{i=1}^N |e_i - L_i|. \quad (2.8)$$

Žiūrint grafiškai, \mathcal{G} yra apibrėžiama kaip santykis tarp ploto, apriboto e ir L kreivėmis (tamsiai mėlynas plotas 2.2 pav., c), ir ploto žemiau kreivės e (tamsiai ir šviesiai mėlynas plotas 2.2 pav., c). Gini koeficiento \mathcal{G} vertė gali kisti nuo 0 iki 1. Vertės, artimos 1, rodo epizodų trukmių pasiskirstymo nelygybę. Ir, atvirkščiai, vertės, artimos 0 rodo, kad per visą laikotarpį PV epizodų trukmės yra lygiai pasiskirsčiusios, t. y. Lorenco kreivė L artima absoliučios lygybės tiesei e .

3. PRIEŠIRDŽIŲ VIRPĖJIMO PROFILIŲ DUOMENŲ BAZĖS IR DETEKTORIŲ PATIKIMUMĄ ĮVERTINANTYS METODAI

3.1 Prieširdžių virpėjimo profilių duomenų bazės

Šiame darbe naudojami tiek realūs, tiek simuliuoti signalai. Realūs signalai yra gauti iš laisvai prieinamos fiziologinių signalų duomenų bazės PhysioNet [112] bei bendradarbiaujant su Lundo universitetu gauta duomenų bazė iš Sankt Peterburgo universiteto. MIT-BIH prieširdžių virpėjimo duomenų bazė (AFDB) yra sudaryta iš 25 įrašų, kurių kiekvieno trukmė yra 10 valandų. Ilgai besitęsiančio prieširdžių virpėjimo duomenų bazė (LTAfDB) sudaryta iš 84 įrašų, kurių kiekvieno trukmė yra 24–25 valandos. Sankt Peterburgo universiteto duomenų bazė (SPAFDB) yra sudaryta iš 36 įrašų, kurių trukmė kinta nuo 1 iki 7 dienų (iš viso 158 dienos). Svarbu paminėti, kad 14 įrašų iš SPAFDB turi papildomus echokardiogramos duomenis (kairiojo prieširdžio tūrį ir įtempimą). Šios trys duomenų bazės yra naudojamos norint ištirti detektorių patikimumo įverčius, ištirti pasiūlytus būdus charakterizuoti PV profilį, įvertinti PV profilių atkūrimo patikimumą ir ištirti PV profilį charakterizuojančių parametrų sąsają su echokardiogramos parametrais.

Simuliuotiems signalams gauti naudojamas modelis, kuris yra skirtas 12 derivacijų EKG signalui su PV epizodais modeliuoti [67]. Modelis leidžia kontroliuoti santykinę bendrą PV trukmę, PV epizodų trukmės medianą, triukšmo lygį ir priešlaikinių prieširdžių susitraukimų skaičių. Iš viso buvo sumodeliuotos penkios duomenų bazės, skirtos EKG savybių (morfologijos, priešlaikinių prieširdžių susitraukimų ir triukšmo) bei PV profilio savybių (epizodų trukmės ir santykinės bendros PV trukmės) įtakai detektorių veikimui ištirti. Kiekvienu atveju sumodeliuota po 100 1 val. trukmės EKG įrašų.

3.2 Detektorių patikimumo įvertinimas

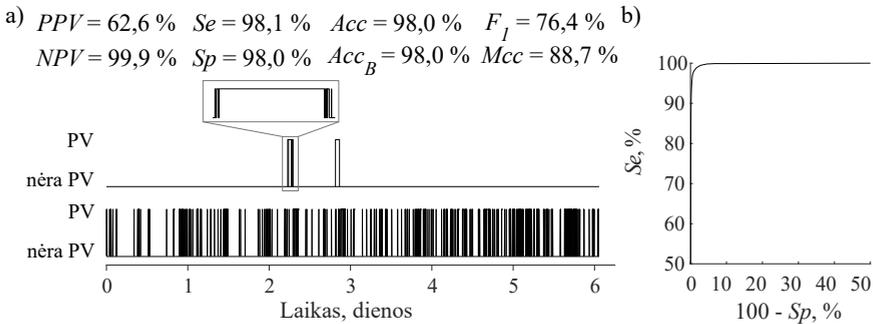
Idealiu atveju PV profilis, gautas iš detektoriaus, turėtų būti lygiai toks pat, kaip atraminis (anototas), tačiau reali situacija nėra tokia. PV profilis, gautas iš detektorių, neretai būna užterštas klaidingai atpažintais ar neatpažintais epizodais. Atraminis PV profilis gali būti palyginamas su profilium, gautu iš detektoriaus, naudojant tris būdus. Labiausiai paplitęs būdas yra kiekvieno širdies dūžio palyginimas (ar jis priskirtas PV, ar ne). Kitas būdas – palyginti pasirinkto ilgio segmentus, priimant, kad segmentas priskiriamas PV, jei $>50\%$ dūžių yra PV. Paskutinis būdas – palyginti teisingai atpažintus PV epizodus, vertinant, kad epizodas atpažįstamas, jei persidengimas viršija numatytą slenkstį, pvz., 50% .

Detektorių patikimumas įvertinamas šiais įverčiais: jautrumas (Se), specifiskumas (Sp), teigiamo testo prognostinė vertė (PPV), neigiamo testo prognostinė vertė (NPV), diagnostinis tikslumas (Acc), balansuotas diagnostinis tikslumas (Acc_B), F_1 įvertis (F_1) ir Mathewso koreliacijos koeficientas (Mcc).

4. REZULTATAI

4.1 Prieširdžių virpėjimo detektorių patikimumo tyrimas

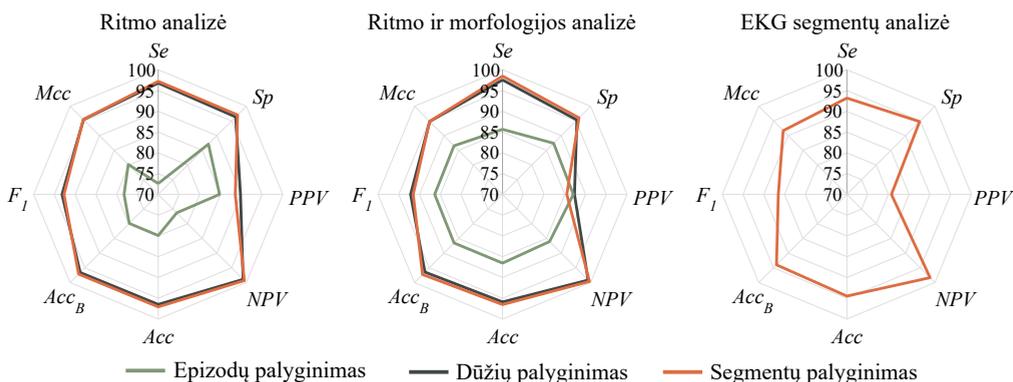
Detektorių patikimumo įverčiai. Nepaisant to, kad pasiūlyta įvairių PV detektorių patikimumo įverčių, nėra iširta, kokią įtaką jiems daro PV profilių savybės. Pavyzdžiui, 4.1 pav. pateiktas anotuotas PV profilis, kurį sudaro keli PV epizodai, tačiau PV profilis, gautas naudojant PV detektorių, labai skiriasi dėl klaidingai atpažintų trumpų PV epizodų. Nepaisant to, operatoriaus charakteringa kreivė (4.1 pav. b) ir diagnostinis tikslumas (Acc) neatspindi šios problemos. Patikimumo įverčių nejautrumas klaidingiems aliarmams yra susijęs su duomenų klasių disbalansu, t. y. pateiktame pavyzdyje 96,7 % širdies dūžių priklauso SR, o likę 3,3 % dūžių – PV. Parametrai F_1 ir Mcc yra jautrūs duomenų disbalansui ir įgyja daug mažesnes vertes (atitinkamai 76,4 % ir 88,7 %).



4.1 pav. a) Anotuotas PV profilis iš SPAFDB (viršuje), kurį sudaro 8 epizodai (medianinė trukmė 113 dūžių), ir PV detektoriumi gautas PV profilis (apačioje), kurį sudaro 518 epizodų (medianinė trukmė 15 dūžių); b) operatoriaus charakteringa kreivė ROC

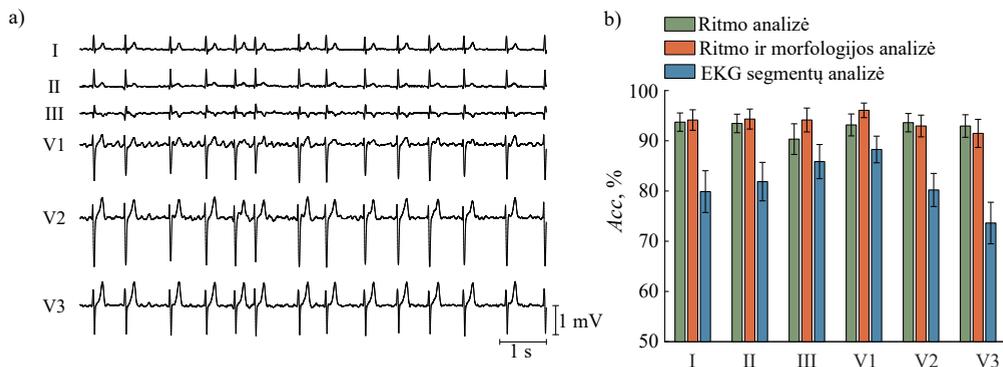
Anotuoti PV profiliai palyginti su skirtingais PV detektoriais gautais profiliais, taikant tris skirtingus profilių palyginimo būdus. Tyrimas parodė, kad patikimumo įverčių vertės daug mažesnės naudojant epizodų palyginimo būdą nei taikant dūžių arba segmentų palyginimo būdus (4.2 pav.). Kadangi nei dūžių, nei segmentų palyginimo metodas neatsižvelgia į atpažintus epizodus, epizodų palyginimo metodas yra labiau tinkamas detektorių patikimumui vertinti, kai tolimesnis tikslas yra PV profilių charakterizavimas.

Veiksniai, darantys įtaką detektorių patikimumui. PV detektorių ir kartu PV profilių atkūrimo patikimumui didelę įtaką daro įvairūs fiziologiniai veiksniai ir paties PV profilio savybės. Tyrimo rezultatai su šių veiksnių įtaka detektorių patikimumui pristatyti šiame poskyryje.



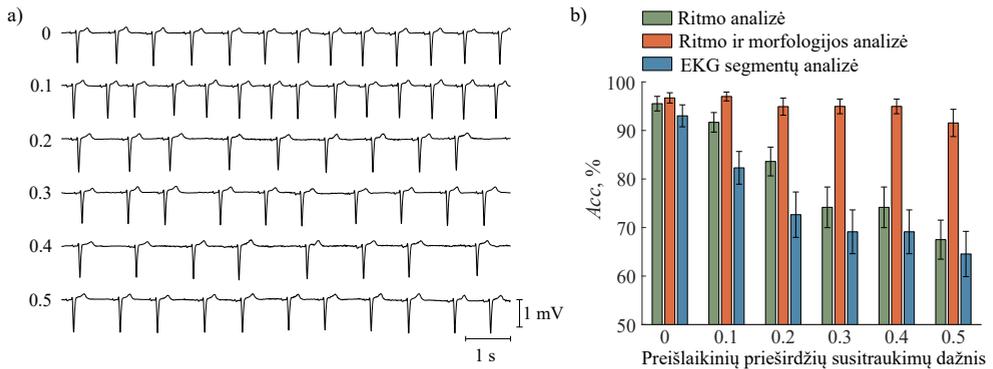
4.2 pav. Trijų tipų PV detektorių patikimumas taikant epizodų, dūžių ir segmentinį palyginimo būdus. Dėl algoritmo specifikos giliojo mokymosi detektoriaus patikimumas gali būti įvertintas tik taikant segmentinį palyginimo būdą. Rezultatai gauti naudojant SPAFDB

Priklausomai nuo EKG derivacijos, naudojamos PV atpažinti, kinta PV profilio atkūrimo diagnostinis tikslumas (4.3 pav.). EKG derivacija didžiausią įtaką daro giliojo mokymosi PV detektoriui, t. y. didžiausias diagnostinis tikslumas gautas V₁ derivacijos, kuri ir buvo naudojama apmokant detektorių, tačiau smarkiai sumažėjo analizuojant kitas derivacijas. O ritmo bei ritmo ir morfologijos analize pagrįsti detektoriai mažiau jautrūs EKG derivacijos parinkimui.



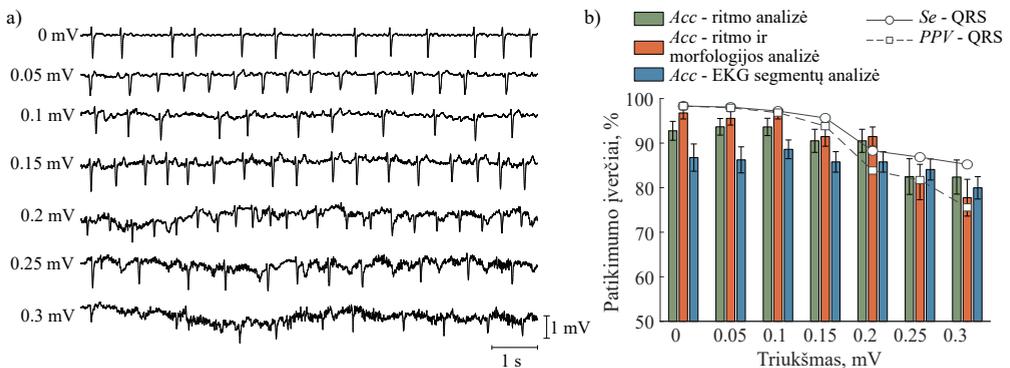
4.3 pav. a) EKG derivacijų PV metu pavyzdžiai ir b) diagnostinio tikslumo (Acc) priklausomybė nuo EKG derivacijos, taikant skirtingą PV detektorių. Rezultatai pateikti vidurkiu ir 95 % pasikliautinoju intervalu

Ritmo analizės ir giliojo mokymosi detektorių patikimumą labai blogina priešlaininiai prieširdžių susitraukimai. Didėjant jų skaičiui, šių PV detektorių diagnostinis tikslumas mažėja (4.4 pav.). O ritmo ir morfologijos analize pagrįsto PV detektoriaus patikimumas mažiau priklauso nuo priešlaikinių prieširdžių susitraukimų skaičiaus, nes jis atsižvelgia į EKG morfologiją.



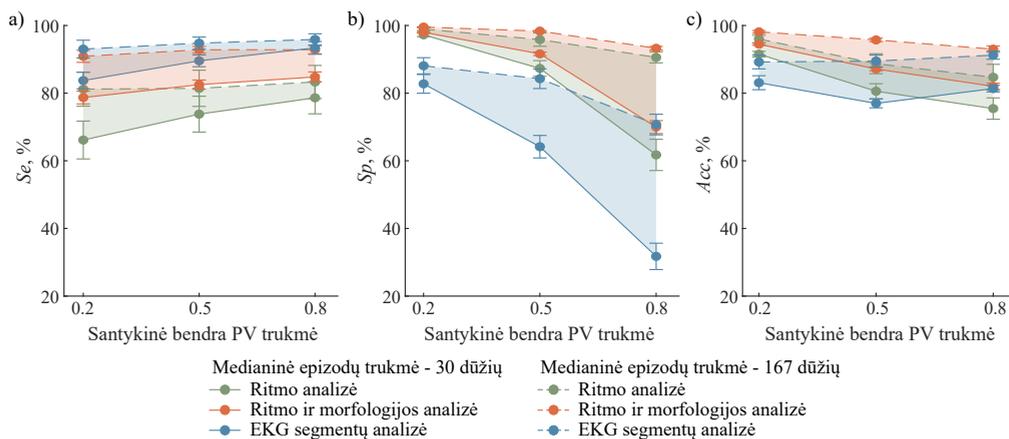
4.4 pav. a) EKG su skirtingu priešlaikinių prieširdžių susitraukimų skaičiumi ir b) diagnostinio tikslumo (*Acc*) priklausomybė nuo priešlaikinių susitraukimų skaičiaus. Rezultatai pateikti vidurkiu ir 95 % pasikliautinoju intervalu

Ištyrus PV detektorių patikimumą keičiant triukšmo lygį, įvertinta, kad ritmo analize bei ritmo ir morfologijos analize pagrįstų detektorių patikimumas pradeda mažėti, kai triukšmo lygis viršija 0,15 mV (4.5 pav.). Didžiąja dalimi tai priklauso nuo QRS kompleksų atpažinimo tikslumo (ilustracijoje pateiktas QRS kompleksų atpažinimo jautrumas ir teigiamo testo prognostinė vertė). Giliojo mokymo detektorius mažiausiai priklauso nuo triukšmo lygio, tačiau, net ir esant mažam triukšmo lygiui, diagnostinis tikslumas mažesnis, palyginti su kitais dviem detektoriais.



4.5 pav. a) EKG V_1 derivacija PV metu su skirtingu triukšmo lygiu ir b) PV detektorių bei QRS kompleksų atpažinimo algoritmo patikimumo priklausomybė nuo triukšmo lygio. Rezultatai pateikti vidurkiu ir 95 % pasikliautinoju intervalu

Atliktas tyrimas, siekiant įvertinti PV profilio savybių įtaką profilių atkūrimo patikimumui. Tyrimas parodė, kad PV detektorių specifiskumas mažėja, didėjant santykinei bendrai PV trukmei bei trumpėjant epizodų trukmei (4.6 pav.). Tai leidžia daryti išvadą, kad PV profilius su didele santykine bendra PV trukme bei trumpais epizodais atkurti naudojant PV detektorius yra sudėtingiausia.



4.6 pav. PV detektorių patikimumo priklausomybė nuo santykinės bendros PV trukmės: a) jautrumas, b) specifiškumas, c) diagnostinis tikslumas

Detektorių patikimumo palyginimas. Siekiant palyginti detektorių patikimumą tarp skirtingų studijų, reikia, kad apmokymo ir testavimo duomenys būtų apdorojami ir tvarkomi vienodai. Pirmiausia turėtų būti naudojami visi įrašai iš duomenų bazės, įrašai neturėtų būti atmetami dėl prastos signalo kokybės ar siekiant gauti subalansuotus duomenų rinkinius [69]. Antra, detektorių testavimas turėtų būti atliekamas su kita duomenų baze nei ta, kuri buvo naudojama apmokyti. Trečia, tas pats pacientas neturėtų būti įtrauktas į apmokymo ir testavimo duomenų rinkinius. Ketvirta, pageidautina pateikti įžvalgų, dėl kokių konkrečių probleminių situacijų pablogėja detektorių patikimumas.

Atliktas tyrimas, kurio metu išanalizuota, kaip giliojo mokymo detektoriai ir detektoriai, sukurti naudojant ekspertų iš anksto apibrėžtus požymius (ritmo ir / arba morfologijos informacija), atitinka apibrėžtus kriterijus. Į analizę įtraukta 14 giliojo mokymo detektorių [37]–[46], [48, 49, 50, 52] ir 13 ekspertinių detektorių [55], [60]–[64], [66], [70]–[75]. Tyrimas parodė, kad šių detektorių palyginimas gali būti klaidinantis, nes buvo taikyti skirtingi duomenų apdorojimo metodai. Pavydžiui, tik 50 % giliojo mokymo ir 70 % ekspertinių detektorių buvo ištestuoti naudojant visus įrašus iš AFDB. Skirtingų pacientų duomenys apmokymo ir testavimo duomenų rinkiniuose buvo naudojami tik 29 % giliojo mokymo ir 77 % ekspertinių detektorių. Vos vienas (7 %) giliojo mokymo detektorius buvo ištestuotas su kita duomenų baze, o tai buvo padaryta su 77 % ekspertinių detektorių. Probleminiai atvejai buvo atskleisti tik su vienu (7 %) giliojo mokymo ir su 7 (54 %) ekspertiniais detektoriais.

Rekomendacijos. Atsižvelgiant į atliktų tyrimų rezultatus ir atliktą tyrimų analizę, išskirtos penkios rekomendacijos, kaip reikėtų įvertinti detektorių patikimumą:

1. Naudoti skirtingus duomenų rinkinius apmokymo ir testavimo fazėse, užtikrinant, kad tas pats pacientas neatsirastų šiuose dviejuose duomenų rinkiniuose.
2. Pateikti anotacijų palyginimo metodą ir pagrįsti jo pasirinkimą. Jei reikia, pateikti segmento ilgį bei epizodų persidengimo slenkstį.
3. Detektorių patikimumui įvertinti naudoti Mcc , o ne Acc ar F_1 . Pateikti Se , Sp ir PPV , kad būtų lengviau palyginti su jau publikuotais PV detektoriais. Plotas po ROC kreive neturi būti naudojamas kaip įvertis, vertinantis detektorių patikimumą.
4. Įvertinti fiziologinių ir techninių veiksnių įtaką detektorių patikimumui, įskaitant derivacijos pasirinkimą, prieširdžių priešlaikinių susitraukimų dažnį, triukšmo lygį, kitas aritmijas.
5. Skirti ypatingą dėmesį detektorių patikimumui, jei siekiama charakterizuoti PV profilius.

4.2 Prieširdžių virpėjimo profilių charakterizavimas

4.2.1 Skirstinių analizė

Epizodų trukmių, trukmių tarp epizodų ir trukmių tarp epizodų pradžių prisitaikymo prie skirstinių rezultatai pateikti 4.1 lentelėje. Rezultatai, gauti iš AIC ir BIC, yra identiški beveik visais atvejais, didžioji dalis profilių priskirti lognormaliam skirstiniui: 69 % analizuojant epizodų trukmes, 79 % analizuojant trukmes tarp epizodų ir 100 % analizuojant trukmes tarp epizodų pradžių. Vis dėlto rezultatai, gauti panaudojant Andersono-Darlingo testą, yra prieštaringi. Šiuo atveju didžioji dalis profilių buvo nepriskirta nė vienam skirstiniui: 51 %, 66 %, 94 % profilių buvo atmesti analizuojant atitinkamai epizodų trukmes, trukmes tarp epizodų ir trukmes tarp epizodų pradžių.

Remiantis tyrimo rezultatais, galima daryti išvadą, kad šis metodas yra mažiau tinkamas analizuojant dienos trukmės PV profilius dėl mažo epizodų skaičiaus. Taip pat svarbu ir tai, kad statistiniai skirstiniai remiasi prielaida, jog epizodai yra statistiškai nepriklausomi, tačiau žinoma, kad PV epizodai turi tendenciją sudaryti klasterius [96].

4.1 lentelė. Skirstinių pritaikymas remiantis Andersono-Darlingo testu (p -vertė), Akaike'ės (AIC) ir Bayesian'o (BIC) informacijos kriterijais. Rezultatai gauti iš 24 val. trukmės profilių iš SPAFDB ir LTAfDB. Pastaba, epizodas trumpinamas "epi."

Skirstinys	Pritaikytų profilių skaičius								
	Epi. trukmė			Trukmė tarp epi.			Trukmė tarp epi. pradžių		
	p -vertė	AIC	BIC	p -vertė	AIC	BIC	p -vertė	AIC	BIC
Normalusis	0	0	0	0	0	0	0	0	0
Ekspontentinis	4	0	2	1	0	0	0	0	0
Lognormalusis	21	54	54	19	63	63	5	78	78
Weibullio	13	2	2	7	1	1	0	0	0
Nė vienas	40	22	20	53	16	16	73	0	0

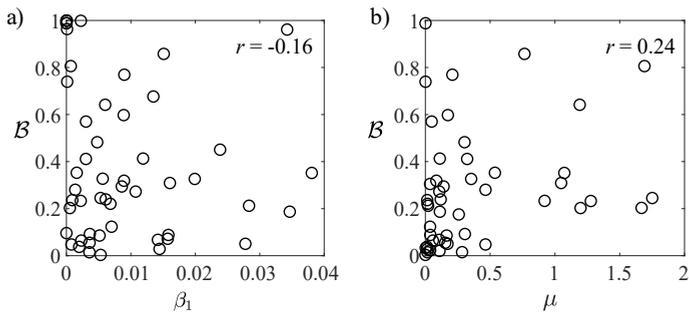
4.2.2 Prieširdžių virpėjimo modelio panaudojimas

PV profiliai gali būti suskirstyti į tris tipus: profiliai su vienu klasteriu, profiliai, sudaryti iš kelių klasterių, ir profiliai, kuriuose epizodai pasiskirstę per visą stebėjimo laikotarpį. Rezultatai, pateikti 4.2 lentelėje, rodo modelio parametru β_1 ir μ vertes šių trijų profilių tipų. Įdomu tai, kad β_1 parametro vertė yra daug mažesnė PV profiluose su pasiskirsčiais epizodais, palyginti su profilių tipais, kuriuose yra epizodų klasteriai. PV profiliai su pasiskirsčiais epizodais įgyja didesnes μ vertes (vidurkis – 4,19), tai reiškia, kad šiuose profiluose dominuoja PV.

4.2 lentelė. Skirtingų PV profilių tipų PV profilių modelio parametru β_1 ir μ vertės. Rezultatai gauti naudojant SPAFDB, AFDB ir LTAfDB ir atvaizduoti vidurkiu ir 95 % pasikliautinoju intervalu. PV profiliai su mažiau nei 10 epizodų buvo pašalinti iš tyrimo bei pašalintos parametru išskirtys. Dėl išskirčių pašalinimo profilių skaičius (kiekis) kinta priklausomai nuo parametro

Profilio tipas	Kiekis	β_1	Kiekis	μ
Vienas klasteris	19	0,043 [0,015–0,071]	7	0,702 [0,125–1,278]
Keli klasteriai	26	0,008 [0,005–0,011]	26	0,163 [0,103–0,223]
Pasiskirstę epizodai	10	0,006 [0,002–0,009]	17	4,191 [0,670–7,713]
Visi profiliai	52	0,009 [0,006–0,011]	50	0,375 [0,234–0,516]

Atliktas tyrimas, siekiant įvertinti, ar modelio parametrai (β_1 ir μ) suteikia papildomos informacijos apie PV profilį. Tuo tikslu, 4.7 pav. pateikta koreliacija tarp santykinės bendros PV trukmės \mathcal{B} ir modelio parametru. Rezultatai parodė silpną neigiamą koreliaciją tarp \mathcal{B} ir β_1 parametro ($r = -0,16$) ir silpną koreliaciją tarp \mathcal{B} ir μ ($r = 0,24$). Taigi, galima daryti išvadą, kad modelio parametrai suteikia papildomos informacijos apie PV profilį, β_1 įvertina epizodų susigrupavimą į klasterius, o μ įvertina ritmo dominavimą.



4.7 pav. Koreliacija tarp santykinės bendros PV trukmės \mathcal{B} ir modelio parametru: a) β_1 ir b) μ . Kiekvienu atveju pateiktas Pirono koreliacijos koeficientas r . Rezultatai gauti naudojant SPAFDB, AFDB ir LTAfDB. PV profiliai su mažiau nei 10 epizodų buvo pašalinti iš tyrimo ir pašalintos parametru išskirtys

4.2.3 Parametrai, charakterizuojantys prieširdžių virpėjimo profilį

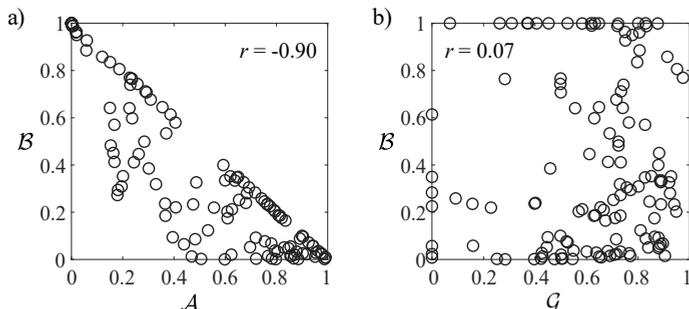
Rezultatai, pateikti 4.3 lentelėje, rodo skirtingų PV profilių tipų santykinės bendros PV trukmės \mathcal{B} , agregacijos \mathcal{A} ir Gini koeficiento \mathcal{G} vertes. PV profiliai su vienu klasteriu ar keliais klasteriais įgyja mažas \mathcal{B} vertes, palyginti su PV profiliais, kuriuose epizodai yra pasiskirstę laike. Šie trys profilių tipai gali būti atskiriami panaudojant agregacijos \mathcal{A} parametru: PV profiliai su vienu klasteriu įgyja dideles \mathcal{A} vertes (vidurkis – 0,79), profiliai su keliais klasteriais įgyja vidutines \mathcal{A} vertes (vidurkis – 0,59), o profiliai su pasiskirsčiusiais epizodais įgyja mažas \mathcal{A} vertes (vidurkis – 0,22). Gini koeficientas \mathcal{G} įgyja panašias vertes visais atvejais.

4.3 lentelė. Skirtingų PV profilių tipų santykinės bendros PV trukmės \mathcal{B} , agregacijos \mathcal{A} ir Gini koeficiento \mathcal{G} vertės. Rezultatai gauti naudojant SPAFDB, AFDB ir LTAfDB ir atvaizduoti vidurkiu ir 95 % pasikliautiniu intervalu

Profilio tipas	Kiekis	\mathcal{B}	\mathcal{A}	\mathcal{G}
Vienas klasteris	33	0,192 [0,126–0,258]	0,788 [0,726–0,851]	0,467 [0,361–0,572]
Keli klasteriai	43	0,195 [0,144–0,247]	0,592 [0,517–0,667]	0,732 [0,681–0,782]
Pasiskirstę epizodai	54	0,678 [0,583–0,773]	0,218 [0,151–0,285]	0,629 [0,572–0,685]
Visi profiliai	130	0,395 [0,333–0,456]	0,486 [0,429–0,544]	0,622 [0,579–0,664]

Analogiškai, kaip ir su modelio parametrais, atliktas tyrimas, siekiant įvertinti, ar agregacija \mathcal{A} ir Gini koeficientas \mathcal{G} suteikia papildomos informacijos apie PV profilį, palyginti su santykinė bendra PV trukme \mathcal{B} . Koreliacija tarp santykinės bendros PV trukmės \mathcal{B} ir agregacijos \mathcal{A} bei Gini koeficiento \mathcal{G} pateikta 4.8 pav. Rezultatai parodė stiprią neigiamą koreliaciją tarp \mathcal{B} ir \mathcal{A} ($r = -0,90$), tačiau koreliacija mažėja, mažėjant \mathcal{B} vertei. Agregacijos parametras \mathcal{A} yra svarbus analizuojant profilius, kurių \mathcal{B} yra mažesnė nei 0,5 ($r = -0,70$). Gini koeficientas \mathcal{G} nepriklauso nuo \mathcal{B} ($r = 0,07$). Net ir tada, kai \mathcal{B} vertės yra artimos 1, \mathcal{G} įgyja skirtingas vertes, nes epizodų trukmės

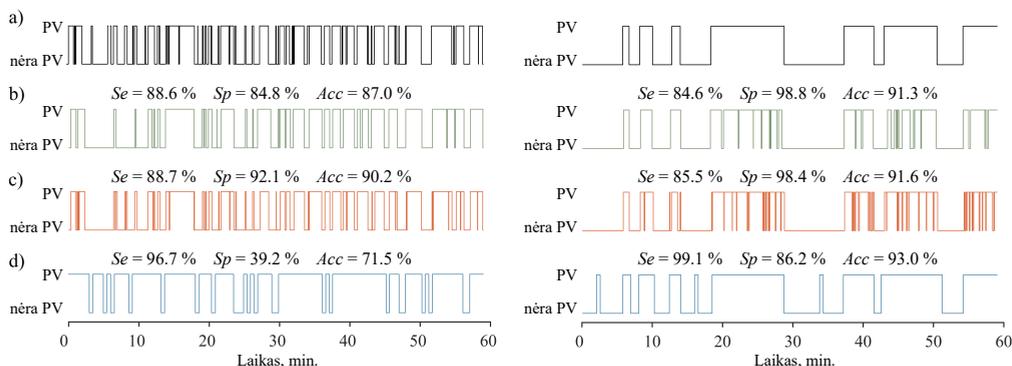
skiriasi. Tai parodo, kad Gini koeficientas \mathcal{G} yra svarbus analizuojant profilius su didele santykinė bendra PV trukme.



4.8 pav. Koreliacija tarp santykinės bendros PV trukmės \mathcal{B} ir a) agregacijos \mathcal{A} ir b) Gini koeficiento \mathcal{G} . Kiekvienu atveju pateiktas Pironso koreliacijos koeficientas r . Rezultatai gauti naudojant SPAFDB, AFDB ir LTAfDB

4.3 Prieširdžių virpėjimo profilių atkūrimo patikimumo tyrimas

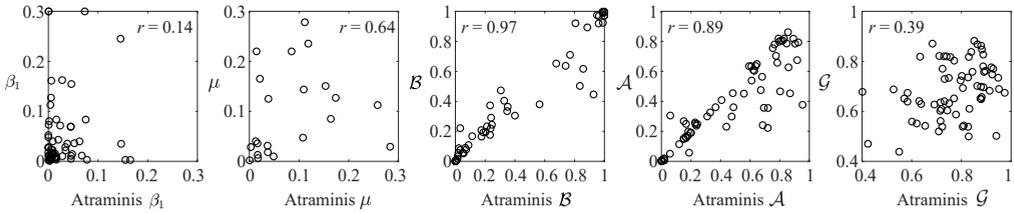
Pagrindinės problemos, susijusios su PV profilių atkūrimu, iliustruojamos 4.9 pav. Giliojo mokymosi PV detektorius, kuris analizuoja EKG segmentus, tendencingai jungia trumpus PV epizodus į ilgesnius, taip prarandama informacija apie epizodų klasterius. Ritmo bei ritmo ir morfologijos analize pagrįsti PV detektoriai elgiasi priešingai – suskaido ilgesnius PV epizodus į klasterius.



4.9 pav. a) Anotuoti PV profiliai su ypač trumpais (kairėje) ir trumpais (dešinėje) PV epizodais. PV profiliai, gauti naudojant: b) ritmo analize pagrįstą detektorius, c) ritmo ir morfologijos analize pagrįstą detektorius ir d) giliojo mokymosi detektorius

PV profilių atkūrimo problema išsamiau iliustruojama 4.10 pav., kuriame pateiktos modelio parametrų β_1 ir μ , santykinės bendros PV trukmės \mathcal{B} , agregacijos \mathcal{A} ir Gini koeficiento \mathcal{G} vertės, gautos iš atraminio profilio (anotuoto) ir profilio, gauto iš detektorius. Dažniausiai \mathcal{B} vertės atitinka abiem atvejais ($r = 0,97$). Tačiau agregacijos \mathcal{A} vertės skiriasi nuo atraminų ($r = 0,89$), tai reiškia, kad klaidingai atpažinti

epizodai ar suskaidyti epizodai turi didelę įtaką epizodų pasiskirstymui laike. Dar sudėtingesnė situacija yra su modelio parametrais (β_1 ir μ) ir Gini koeficientu \mathcal{G} , šiuo atveju klaidingai atpažintų epizodų įtaka yra didžiulė.



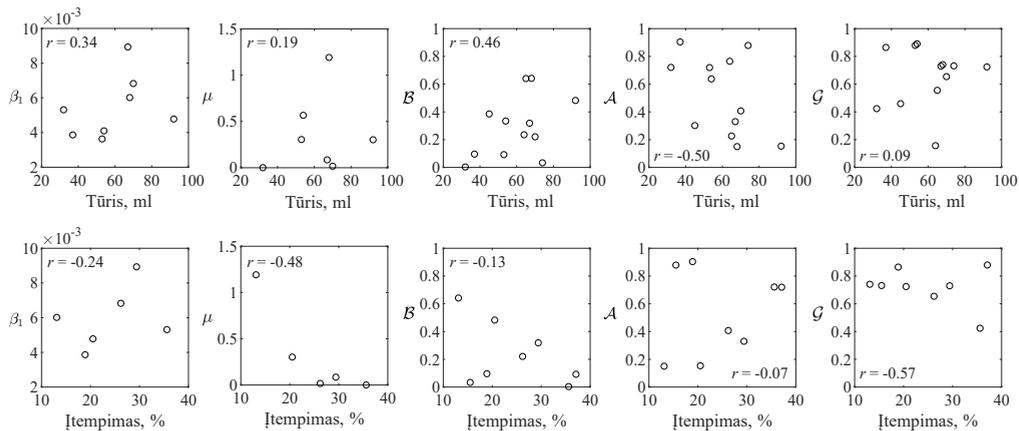
4.10 pav. PV profilio atkūrimo vertinimas pagal profilį charakterizuojančius parametrus: modelio parametrus (β_1 ir μ), santykinę bendrą PV trukmę \mathcal{B} , agregaciją \mathcal{A} ir Gini koeficientą \mathcal{G} . Kiekvienu atveju pateiktas Pironso koreliacijos koeficientas r . Rezultatai gauti naudojant SPAFDB, AFDB ir LTAfDB. PV profiliai su mažiau nei 10 epizodų buvo pašalinti iš tyrimo

Šis tyrimas atskleidė, kad, nepaisant aukštų detektoriaus patikimumo įverčių, PV profilių atkūrimas gali būti nepatikimas. Klaidingai atpažinti epizodai ar neatpažinti epizodai pakeičia PV profilį, ir tai atspindi PV profilį charakterizuojančių parametrų vertės. Tolimesni tyrimai turi atsižvelgti ne tik į detektorių patikimumą, bet ir į PV profilio atkūrimo patikimumą ir jį charakterizuojančius parametrus.

4.4 Prieširdžių virpėjimo profilio ryšys su echokardiogramos parametrais

Panaudojant PV profilius iš SPAFDB, kurie turi echokardiogramos parametrus, atliktas tyrimas, siekiant įvertinti kairiojo prieširdžio tūrio ir įtempimo ryšį su modelio parametrais (β_1 ir μ), santykinę bendrą PV trukmę \mathcal{B} , agregaciją \mathcal{A} ir Gini koeficientu \mathcal{G} . Tyrimo rezultatai, pateikti 4.11 pav., parodė, kad β_1 , μ ir Gini koeficientas \mathcal{G} yra silpnai koreliuoti su kairiojo prieširdžio tūriu (atitinkamai $r = 0,34$, $r = 0,19$, $r = 0,09$). O santykinė bendra PV trukmė \mathcal{B} ir agregacija \mathcal{A} yra vidutiniškai koreliuoti su kairiojo prieširdžio tūriu (atitinkamai $r = 0,46$ ir $r = -0,50$). Kairiojo prieširdžio įtempimas yra vidutiniškai koreliuotas su μ ir Gini koeficientu \mathcal{G} (atitinkamai $r = -0,48$ ir $r = -0,57$).

Nors gautiems rezultatams trūksta patikimumo dėl mažos imties, tačiau tyrimas atskleidžia šios tyrimo krypties svarbą. Prieširdžių struktūrinis pakitimas gali būti susijęs su PV profilio charakteristikų pokyčiais, dažnai pasireiškiančiais ilgėjančiais PV epizodais (t. y. didėja μ ir \mathcal{B}). Todėl reikia atlikti tolimesnius tyrimus su didesne duomenų baze, kad būtų patvirtinta PV profilį charakterizuojančių parametrų klinikinė reikšmė.



4.11 pav. Kaičiojo prieširdžio tūrio ir įtempimo ryšys su modelio parametrais (β_1 ir μ), santykinė bendra PV trukmė \mathcal{B} , agregacija \mathcal{A} ir Gini koeficientu \mathcal{G} . Kiekvienu atveju pateiktas Pirsono koreliacijos koeficientas r . Rezultatai gauti naudojant SPAFDB

IŠVADOS

1. Sukurtas paroksizminių PV epizodų pasiskirstymo laike modelis, kuris atsižvelgia į epizodų susigrupavimą į klasterius. Siūlomas modelis yra pagrįstas kintamu dvimačiu Hawkeso procesu ir gali būti naudojamas modeliuojant įvairius PV profilius. Svarbi modelio savybė yra ta, kad jis gali būti naudojamas ištraukiant modelio parametrus iš realių signalų, o šie parametrai gali būti naudojami charakterizuoti PV profiliui, t. y. parametras β_1 yra susijęs su PV epizodų susigrupavimu į klasterius, o parametras μ nurodo širdies ritmo dominavimą (PV ar sinusinis ritmas).
2. Pasiūlyti trys būdai PV epizodų pasiskirstymui laike charakterizuoti. Skirstinių analize pagrįstas būdas yra mažiau tinkamas PV profiliams vertinti, nes 51 % histogramų su epizodų trukmėmis, 66 % su trukmėmis tarp epizodų ir 94 % su trukmėmis tarp epizodų pradžių nebuvo priskirtos nė vienam iš skirstinių. Modelio parametrai suteikia papildomos informacijos apie PV profilį, koreliacija tarp santykinės bendros PV trukmės ir β_1 bei μ atitinkamai yra $-0,16$ ir $0,24$. O agregacija yra neigiamai stipriai koreliuota su santykinę bendra PV trukme ($r = -0,90$). Agregacija labiau tinka charakterizuoti PV profilius su maža santykinę bendra PV trukme ($< 0,5$), o Gini koeficientas, atvirkščiai – naudingas atskiriant PV profilius, turinčius didelę santykinę bendrą PV trukmę. Taigi, parametų derinys leidžia analizuoti skirtingus PV profilių tipus (t. y. profilius, sudarytus iš vieno klasterio, iš kelių klasterių, ir profilius, kuriuose epizodai pasiskirstę laike).
3. EKG signalo savybės (t. y. EKG morfologija, prieširdžių priešlaikinių susitraukimų skaičius, triukšmo lygis) bei PV profilio savybės (t. y. santykinę bendrą PV trukmę ir epizodo trukmę) daro įtaką detektorių patikimumui. Siekiant charakterizuoti PV profilius, rekomenduojama palyginti gautą profilį iš detektorių su anotuotu profiliumi taikant epizodų palyginimo metodą, o detektorių patikimumui įvertinti naudoti Mcc įvertį, o ne Acc ar F_1 , šalia taip pat pateikti Se , Sp ir PPV . PV profilių charakterizavimas reikalauja didesnio detektorių patikimumo. Pavyzdžiui, koreliacija tarp parametų, gautų iš profilių naudojant detektorius ir anotuotų profilių, yra $0,14$, $0,64$, $0,97$, $0,89$, $0,39$ atitinkamai β_1 , μ , santykinės bendros PV trukmės, agregacijos ir Gini koeficiento. Taigi, tolimesni tyrimai turėtų būti sutelkti į PV profilių atkūrimo patikimumą ir jo užtikrinimą.

REFERENCES

1. KRIJTHE, B. P., KUNST, A., BENJAMIN, E. J., LIP, G. Y., FRANCO, O. H., et al. Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. *European Heart Journal*. 2013, 34(35), 2746–2751.
2. KIRCHHOF, P., BENUSSI, S., KOTECHEA, D., AHLSSON, A., ATAR, D., et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *European Journal Of Cardio-Thoracic Surgery*. 2016, 50(5), E1–E88.
3. HINDRICKS, G., POTPARA, T., DAGRES, N., ARBELO, E., BAX, J. J., et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association of Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *European Heart Journal*. 2020, 1–138.
4. CHEN-SCARABELLI, C., SCARABELLI, T. M., ELLENBOGEN, K. A., and HALPERIN J. L. Device detected atrial fibrillation: what to do with asymptomatic patients? *Journal of the American College of Cardiology*. 2015, 65(3), 281–294.
5. BAROUTIDOU, A., KARTAS, A., SAMARAS, A., PAPAZOGLU, A. S., VRANA, E., et al. Associations of atrial fibrillation patterns with mortality and cardiovascular events: Implications of the MISOAC-AF trial. *Journal of Cardiovascular Pharmacology and Therapeutics*. 2022, 27, 10742484211069422.
6. BORIANI, G., VITOLO, M., DIEMBERGER, I., PROIETTI, M., VALENTI, A. C., et al. Optimizing indices of atrial fibrillation susceptibility and burden to evaluate atrial fibrillation severity, risk and outcomes. *Cardiovascular Research*. 2021, 117(7), 1–21.
7. DE WITH, R. R., ERKUNER, O., RIENSTRA, M., NGUYEN, B.-O., KORVER, F. W., et al. Temporal patterns and short-term progression of paroxysmal atrial fibrillation: data from RACE V. *EP Europace*. 2020, 22(8), 1162–1172.
8. AL-SAADY, N., OBEL, O., and CAMM, A. Left atrial appendage: structure, function, and role in thromboembolism. *Heart*. 1999, 82(5), 547–554.
9. BUTKUVIENĖ, M., PETRĖNAS, A., SOLOŠENKO, A., MARTIN-YEBRA, A., MAROZAS, V., and SÖRNMO, L. Considerations on performance evaluation of atrial fibrillation detectors. *IEEE Transactions on Biomedical Engineering*. 2021, 68(11), 3250–3260.

10. HENRIKSSON, M., MARTIN-YEBRA, A., BUTKUVIENĖ, M., RASMUSSEN, J., MAROZAS, V., et al. Modeling and estimation of temporal episode patterns in paroxysmal atrial fibrillation. *IEEE Transactions on Biomedical Engineering*. 2020, 68(1), 319–329.
11. BUTKUVIENĖ, M., PETRĖNAS, A., SOLOŠENKO, A., MARTIN-YEBRA, A., MAROZAS, V., and SÖRNMO, L. Atrial fibrillation episode patterns and their influence on detection performance. *In 2021 Computing in Cardiology (CinC)*. 2021, 48, 1–4.
12. MARTIN-YEBRA, A., HENRIKSSON, M., BUTKUVIENĖ, M., MAROZAS, V., PETRĖNAS, A., et al. Model-based characterization of atrial fibrillation episodes and its clinical association. *In 2020 Computing in Cardiology (CinC)*. 2020, 62, 4–9.
13. ŠIMAITYTĖ, M., PETRĖNAS, A., MAROZAS, V., HENRIKSSON, M., BACEVIČIUS, J., et al. Quantitative evaluation of temporal episode patterns in paroxysmal atrial fibrillation. *In 2018 Computing in Cardiology (CinC)*. 2018, 45, 1–4.
14. SEET, R. C., FRIEDMAN, P. A., and RABINSTEIN, A. A. Prolonged rhythm monitoring for the detection of occult paroxysmal atrial fibrillation in ischemic stroke of unknown cause. *Circulation*. 2011, 124(4), 477–486.
15. KISHORE A., VAIL, A., MAJID, A., DAWSON, J., LEES, K. R., et al. Detection of atrial fibrillation after ischemic stroke or transient ischemic attack: a systematic review and meta-analysis. *Stroke*. 2014, 45(2), 520–526.
16. SÖRNMO, L. Atrial fibrillation from an engineering perspective. *Springer*. 2018.
17. KOTTKAMP, H. Human atrial fibrillation substrate: towards a specific fibrotic atrial cardiomyopathy. *European Heart Journal*. 2013, 34(35), 2731–2738.
18. DE VOS, C. B., PISTERS, R., NIEUWLAAT, R., PRINS, M. H., TIELEMAN, R. G., et al. Progression from paroxysmal to persistent atrial fibrillation: clinical correlates and prognosis. *Journal of the American College of Cardiology*. 2010, 55(8), 725–731.
19. SCHNABEL, R. B., PECEN, L., ENGLER, D., LUCERNA, M., SELLAL, J. M., et al. Atrial fibrillation patterns are associated with arrhythmia progression and clinical outcomes. *Heart*. 2018, 104(19), 1608–1614.
20. SUGIHARA, C., VEASEY, R., FREEMANTLE, N., PODD, S., FURNISS, S., and SULKE, N. The development of AF over time in patients with permanent pacemakers: objective assessment with pacemaker diagnostics demonstrates distinct patterns of AF. *Europace*. 2015, 17(6), 864–870.

21. VEASEY, R., SUGIHARA, C., SANDHU, K., DHILLON, G., FREEMANTLE, N., et al. The natural history of atrial fibrillation in patients with permanent pacemakers: is atrial fibrillation a progressive disease? *Journal of Interventional Cardiac Electrophysiology*. 2015, 44(1), 23–30.
22. BAYOUMY, K., GABER, M., ELSHAFFEEY, A., MHAIMEED, O., DINEEN, E. H., et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. *Nature Reviews Cardiology*. 2021, 18(8), 581–599.
23. WINEINGER, N. E., BARRETT, P. M., ZHANG, Y., IRFANULLAH, I., MUSE, E. D., et al. Identification of paroxysmal atrial fibrillation subtypes in over 13,000 individuals. *Heart Rhythm*. 2019, 16(1), 26–30.
24. POTPARA, T. S., LIP, G. Y., BLOMSTROM-LUNDQVIST, C., BORIANI, G., VAN GELDER, I. C., et al. The 4S-AF scheme (stroke risk; symptoms; severity of burden; substrate): A novel approach to in-depth characterization (rather than classification) of atrial fibrillation. *Thrombosis and Haemostasis*. 2021, 121(3), 270–278.
25. PETERSEN, M., ROEHRICH, A., BALZER, J., SHIN, D.-I., MEYER, C., et al. Left atrial appendage morphology is closely associated with specific echocardiographic flow pattern in patients with atrial fibrillation. *Europace*. 2014, 17(4), 539–545.
26. FUKUSHIMA, K., FUKUSHIMA, N., KATO, K., EJIMA, K., SATO, H., et al. Correlation between left atrial appendage morphology and flow velocity in patients with paroxysmal atrial fibrillation. *European Heart Journal-Cardiovascular Imaging*. 2015, 17(1), 59–66.
27. WALSH III, J. A., TOPOL, E. J., and STEINHUBL, S. R. Novel wireless devices for cardiac monitoring. *Circulation*. 2014, 130(7), 573–581.
28. EYSENCK W., FREEMANTLE, N., and SULKE, N. A randomized trial evaluating the accuracy of af detection by four external ambulatory ECG monitors compared to permanent pacemaker af detection. *Journal of Interventional Cardiac Electrophysiology*. 2020, 57(3), 361–369.
29. CICONTE, G., GIACOPELLI, D., and PAPPONE, C. The role of implantable cardiac monitors in atrial fibrillation management. *Journal of Atrial Fibrillation*. 2017, 10(2), 1590.
30. PEREIRA, T., TRAN, N., GADHOUMI, K., PELTER, M. M., DO, D. H., et al. Photoplethysmography based atrial fibrillation detection: a review. *NPJ Digital Medicine*. 2020, 3(1), 1–12.

31. NAZARIAN, S., LAM, K., DARZI, A., and ASHRAFIAN, H. Diagnostic accuracy of smartwatches for the detection of cardiac arrhythmia: systematic review and meta-analysis. *Journal of Medical Internet Research*. 2021, 23(8), e28974.
32. BACEVIČIUS, J., ABRAMIKAS, Ž., DVINELIS E., AUDZIJONIENĖ, D., PETRYLAITĖ, M., et al. High specificity wearable device with photoplethysmography and six-lead electrocardiography for atrial fibrillation detection challenged by frequent premature contractions: DoubleCheck-AF. *Frontiers in Cardiovascular Medicine*. 2022, 9, 1–11.
33. BOTTO, G. L., TORTORA, G., CASALE, M. C., CANEVESE, F. L., and BRASCA, F. A. M. Impact of the pattern of atrial fibrillation on stroke risk and mortality. *Arrhythmia and Electrophysiology Review*. 2021, 10(2), 68.
34. DESHMUKH, A., BROWN, M. L., HIGGINS, E., SCHOUSEK, B., ABEYRATNE, A., et al. Performance of atrial fibrillation detection in a new single-chamber ICD. *Pacing and Clinical Electrophysiology*. 2016, 39(10), 1031–1037.
35. PODD, S. J., SUGIHARA, C., FURNISS, S. S., and SULKE, N. Are implantable cardiac monitors the ‘gold standard’ for atrial fibrillation detection? A prospective randomized trial comparing atrial fibrillation monitoring using implantable cardiac monitors and DDDR permanent pacemakers in post atrial fibrillation ablation patients. *EP Europace*. 2016, 18(7), 1000–1005.
36. CAI, W., CHEN, Y., GUO, J., HAN, B., SHI, Y., et al. Accurate detection of atrial fibrillation from 12-lead ECG using deep neural network. *Computers in Biology and Medicine*. 2020, 116, 103378.
37. SHI, H., WANG, H., QIN, C., ZHAO, L., and LIU, C. An incremental learning system for atrial fibrillation detection based on transfer learning and active learning. *Computer Methods and Programs in Biomedicine*. 2020, 187, 105219.
38. GHOSH, S. K., TRIPATHY, R. K., PATERNINA, M. R. A., ARRIETA, J. J., ZAMORA-MENDEZ, A., and NAIK, G. R. Detection of atrial fibrillation from single lead ECG signal using multirate cosine filter bank and deep neural network. *Journal of Medical Systems*. 2020, 44(6), 114.
39. ZHANG, H., HE, R., DAI, H., XU, M., and WANG, Z. SS-SWT and SI-CNN: An atrial fibrillation detection framework for time-frequency ECG signal. *Journal of Healthcare Engineering*. 2020, 2020.
40. HUANG, M.-L., and WU, Y.-S. Classification of atrial fibrillation and normal sinus rhythm based on convolutional neural network. *Biomedical Engineering Letters*. 2020, 1–11.

41. JIN, Y., QIN, C., HUANG, Y., ZHAO, W., and LIU, C. Multi-domain modeling of atrial fibrillation detection with twin attentional convolutional long short-term memory neural networks. *Knowledge-Based Systems*. 2020, 105460.
42. WANG, J. A deep learning approach for atrial fibrillation signals classification based on convolutional and modified elman neural network. *Future Generation Computer Systems*. 2020, 102, 670–679.
43. FUJITA, H. and CIMR, D. Computer aided detection for fibrillations and flutters using deep convolutional neural network. *Information Sciences*. 2019, 486, 231–239.
44. DANG, H., SUN, M., ZHANG, G., QI, X., ZHOU, X., and CHANG, Q. A novel deep arrhythmia-diagnosis network for atrial fibrillation classification using electrocardiogram signals. *IEEE Access*. 2019, 7, 75577–75590.
45. LAI, D., ZHANG, X., BU, Y., SU, Y., and MA, C.-S. An automatic system for real-time identifying atrial fibrillation by using a lightweight convolutional neural network. *IEEE Access*. 2019, 7, 130074–130084.
46. ANDERSEN, R. S., PEIMANKAR, A., and PUTHUSSERYPADY, S. A deep learning approach for real-time detection of atrial fibrillation. *Expert Systems with Applications*. 2019, 115, 465–473.
47. HANNUN, A. Y., RAJPURKAR, P., HAGHPANAHI, M., TISON, G. H., BOURN, C., et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*. 2019, 25(1), 65.
48. HE, R., WANG, K., ZHAO, N., LIU, Y., YUAN, Y., et al. Automatic detection of atrial fibrillation based on continuous wavelet transform and 2D convolutional neural networks. *Frontiers in Physiology*. 2018, 9, 1206.
49. FAUST, O., SHENFIELD, A., KAREEM, M., SAN, T. R., FUJITA, H., and ACHARYA, U. R. Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. *Computers in Biology and Medicine*. 2018, 102, 327–335.
50. XIA, Y., WULAN, N., WANG, K., and ZHANG, H. Detecting atrial fibrillation by deep convolutional neural networks. *Computers in Biology and Medicine*. 2018, 93, 84–92.
51. POURBABAEE, B., ROSHTKHARI, M. J., and KHORASANI, K. Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*. 2018, 48(12), 2095–2104.

52. MOUSAVI, S., AFGHAH, F., and ACHARYA, U. R. HAN-ECG: An interpretable atrial fibrillation detection model using hierarchical attention networks. *Computers in Biology and Medicine*. 2020, 127, 104057.
53. MARSILI, I. A., BIASIOLLI, L., MASE, M., ADAMI, A., ANDRIGHETTI, A. O., et al. Implementation and validation of real-time algorithms for atrial fibrillation detection on a wearable ECG device. *Computers in Biology and Medicine*. 2020, 116, 103540.
54. SOLOŠENKO, A., PETRĚNAS, A., PALIAKAITĖ, B., SÖRNMO, L., and MAROZAS, V. Detection of atrial fibrillation using a wrist-worn device. *Physiological Measurement*. 2019, 40(2), 025003.
55. RODENAS, J., GARCIA, M., ALCARAZ, R., and RIETA, J. J. Wavelet entropy automatically detects episodes of atrial fibrillation from single-lead electrocardiograms. *Entropy*. 2015, 17(9), 6179–6199.
56. PETRĚNAS, A., MAROZAS, V., and SÖRNMO, L. Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Computers in Biology and Medicine*. 2015, 65, 184–191.
57. LADAVICH, S. and GHORAANI, B. Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity. *Biomedical Signal Processing and Control*. 2015, 18, 274–281.
58. LEE, J., REYES, B. A., MCMANUS, D. D., MAITAS, O., and CHON, K. H. Atrial fibrillation detection using an iPhone 4S. *IEEE Transactions on Biomedical Engineering*. 2013, 60(1), 203–206.
59. JIANG, K., HUANG, C., YE, S.-M., and CHEN, H. High accuracy in automatic detection of atrial fibrillation for Holter monitoring. *Journal of Zhejiang University-SCIENCE B*. 2012, 13(9), 751–756.
60. ASGARI, S., MEHRNIA, A., and MOUSSAVI, M. Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine. *Computers in Biology and Medicine*. 2015, 60, 132–142.
61. LEE, J., NAM, Y., MCMANUS, D. D., and CHON, K. H. Time-varying coherence function for atrial fibrillation detection. *IEEE Transactions on Biomedical Engineering*. 2013, 60(10), 2783–2793.
62. LIAN, J., WANG, L., and MUESSIG, D. A simple method to detect atrial fibrillation using RR intervals. *American Journal of Cardiology*. 2011, 107(10), 1494–1497.

63. BABAEIZADEH, S., GREGG, R. E., HELFENBEIN, E. D., LINDAUER, J. M., and ZHOU, S. H. Improvements in atrial fibrillation detection for real-time monitoring. *Journal of Electrocardiology*. 2009, 42(6), 522–526.
64. DASH, S., CHON, K., LU, S., and RAEDER, E. Automatic real time detection of atrial fibrillation. *Annals of Biomedical Engineering*. 2009, 37(9), 1701–1709.
65. WASSERLAUF, J., YOU, C., PATEL, R., VALYS, A., ALBERT, D., and PASSMAN, R. Smartwatch performance for the detection and quantification of atrial fibrillation. *Circulation: Arrhythmia and Electrophysiology*. 2019, 12(6), e006834.
66. PETRĖNAS, A., SÖRNMO, L., LUKOŠEVIČIUS, A., and MAROZAS, V. Detection of occult paroxysmal atrial fibrillation. *Medical and Biological Engineering and Computing*. 2015, 53(4), 287–297.
67. PETRĖNAS, A., MAROZAS, V., SOLOŠENKO, A., KUBILIUS, R., SKIBARKIENĖ, J., et al. Electrocardiogram modeling during paroxysmal atrial fibrillation: application to the detection of brief episodes. *Physiological Measurement*. 2017, 38(11), 2058.
68. SOLOŠENKO, A., PETRĖNAS, A., MAROZAS, V., and SÖRNMO, L. Modeling of the photoplethysmogram during atrial fibrillation. *Computers in Biology and Medicine*. 2017, 81, 130–138.
69. SÖRNMO, L., PETRĖNAS, A., HENRIKSSON, M., and MAROZAS, V. Letter regarding the article “Detecting atrial fibrillation by deep convolutional neural networks” by Xia et al. *Computers in Biology and Medicine*. 2018, 100, 41–42.
70. LAKE, D. E. and MOORMAN, J. R. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *The American Journal of Physiology-Heart and Circulatory Physiology*. 2011, 300(1), H319–25.
71. HUANG, C., YE, S., CHEN, H., LI, D., HE, F., and TU, Y. A novel method for detection of the transition between atrial fibrillation and sinus rhythm. *IEEE Transactions on Biomedical Engineering*. 2011, 58(4), 1113–1119.
72. SHOULDICE, R. B., HENEGHAN, C., and DE CHAZAL, P. Automatic detection of paroxysmal atrial fibrillation. *In Atrial Fibrillation-Basic Research and Clinical Applications*, IntechOpen, 2012.
73. DE CARVALHO, P., HENRIQUES, J., COUCEIRO, R., HARRIS, M., ANTUNES, M., and HABETHA, J. Model-based atrial fibrillation detection. *In ECG Signal Processing, Classification and Interpretation*, Springer, 2012, 99–133.

74. ZHOU, X., DING, H., UNG, B., PICKWELL-MACPHERSON, E., and ZHANG, Y. Automatic online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy. *BioMedical Engineering OnLine*. 2014, 13(1), 18.
75. ZHOU, X., DING, H., WU, W., and ZHANG, Y. A real-time atrial fibrillation detection algorithm based on the instantaneous state of heart rate. *PloS One*. 2015, 10(9).
76. YANG, S.-Y., HUANG, M., WANG, A.-L., GE, G., MA, M., et al. Atrial fibrillation burden and the risk of stroke: A systematic review and dose-response meta-analysis. *World Journal of Clinical Cases*. 2022, 10(3), 939.
77. GLOTZER, T. V., HELLKAMP, A. S., ZIMMERMAN, J., SWEENEY, M. O., YEE, R., et al. Atrial high rate episodes detected by pacemaker diagnostics predict death and stroke: report of the atrial diagnostics ancillary study of the mode selection trial (MOST). *Circulation*. 2003, 107(12), 1614–1619.
78. CAPUCCI, A., SANTINI, M., PADELETTI, L., GULIZIA, M., BOTTO, G., et al. Monitored atrial fibrillation duration predicts arterial embolic events in patients suffering from bradycardia and atrial fibrillation implanted with anti-tachycardia pacemakers. *Journal of the American College of Cardiology*. 2005, 46(10), 1913–1920.
79. BOTTO, G. L., PADELETTI, L., SANTINI, M., CAPUCCI, A., GULIZIA, M., et al. Presence and duration of atrial fibrillation detected by continuous monitoring: crucial implications for the risk of thromboembolic events. *Journal of Cardiovascular Electrophysiology*. 2009, 20(3), 241–248.
80. GLOTZER, T. V., DAOUD, E. G., WYSE, D. G., SINGER, D. E., EZEKOWITZ, M. D., et al. The relationship between daily atrial tachyarrhythmia burden from implantable device diagnostics and stroke risk: the TRENDS study. *Circulation: Arrhythmia and Electrophysiology*. 2009, 2(5), 474–480.
81. ZIEGLER, P. D., GLOTZER, T. V., DAOUD, E. G., WYSE, D. G., SINGER, D. E., et al. Incidence of newly detected atrial arrhythmias via implantable devices in patients with a history of thromboembolic events. *Stroke*. 2010, 41(2), 256–260.
82. BORIANI, G., BOTTO, G. L., PADELETTI, L., SANTINI, M., CAPUCCI, A., et al. Improving stroke risk stratification using the CHADS2 and CHA2DS2-VASc risk scores in patients with paroxysmal atrial fibrillation by continuous arrhythmia burden monitoring. *Stroke*. 2011, 42(6), 1768–1770.
83. SHANMUGAM, N., BOERDLEIN, A., PROFF, J., ONG, P., VALENCIA, O., et al. Detection of atrial high-rate events by continuous home monitoring: clinical significance in the heart failure–cardiac resynchronization therapy population. *Europace*. 2011, 14(2), 230–237.

84. HEALEY, J. S., CONNOLLY, S. J., GOLD, M. R., ISRAEL, C. W., VAN GELDER, I. C., et al. Subclinical atrial fibrillation and the risk of stroke. *New England Journal of Medicine*. 2012, 366(2), 120–129.
85. BORIANI, G., GLOTZER, T. V., SANTINI, M., WEST, T. M., DE MELIS, M., et al. Device-detected atrial fibrillation and risk for stroke: an analysis of > 10 000 patients from the SOS AF project (stroke prevention strategies based on atrial fibrillation information from implanted devices). *European Heart Journal*. 2013, 35(8), 508–516.
86. MARTIN, D. T., BERSOHN, M. M., WALDO, A. L., WATHEN, M. S., CHOU-CAIR, W. K., et al. Randomized trial of atrial arrhythmia monitoring to guide anticoagulation in patients with implanted defibrillator and cardiac resynchronization devices. *European Heart Journal*. 2015, 36(26), 1660–1668.
87. TURAKHIA, M. P., ZIEGLER, P. D., SCHMITT, S. K., CHANG, Y., FAN, J., et al. Atrial fibrillation burden and short-term risk of stroke: case-crossover analysis of continuously recorded heart rhythm from cardiac electronic implanted devices. *Circulation: Arrhythmia and Electrophysiology*. 2015, 8(5), 1040–1047.
88. VAN GELDER, I. C., HEALEY, J. S., CRIJNS, H. J., WANG, J., HOHN-LOSER, S. H., et al. Duration of device detected subclinical atrial fibrillation and occurrence of stroke in ASSERT. *European Heart Journal*. 2017, 38(17), 1339–1344.
89. KAPLAN, R. M., KOEHLER, J., ZIEGLER, P. D., SARKAR, S., ZWEIBEL, S., and PASSMAN, R. S. Stroke risk as a function of atrial fibrillation duration and CHA2DS2-VASc score. *Circulation*. 2019, 140(20), 1639–1646.
90. CHU, S.-Y., JIANG, J., WANG, Y.-L., SHENG, Q.-H., ZHOU, J., and DING, Y.-S. Pacemaker detected atrial fibrillation burden and risk of ischemic stroke or thromboembolic events – a cohort study. *Heart and Lung*. 2020, 49(1), 66–72.
91. MONT, L. and GUASCH, E. Atrial fibrillation progression: How sick is the atrium? *Heart Rhythm*. 2017, 14(6), 808–809.
92. GREER, G. S., WILKINSON, W. E., MCCARTHY, E. A., and PRITCHETT, E. L. Random and nonrandom behavior of symptomatic paroxysmal atrial fibrillation. *The American Journal of Cardiology*. 1989, 64(5), 339–342.
93. GILLIS, A. M. and ROSE, M. S. Temporal patterns of paroxysmal atrial fibrillation following DDDR pacemaker implantation. *American Journal of Cardiology*. 2000, 85(12), 1445–1450.

94. KAEMMERER, W. F., ROSE, M. S., and MEHRA, R. Distribution of patients? Paroxysmal atrial tachyarrhythmia episodes: implications for detection of treatment efficacy. *Journal of Cardiovascular Electrophysiology*. 2001, 12(2), 121–130.
95. SHEHADEH, L. A., LIEBOVITCH, L. S., and WOOD, M. A. Temporal patterns of atrial arrhythmia recurrences in patients with implantable defibrillators: implications for assessing antiarrhythmic therapies. *Journal of Cardiovascular Electrophysiology*. 2002, 13(4), 303–309.
96. GILLIS, A. M. Modeling temporal patterns of atrial tachyarrhythmias: a new surrogate outcome measure for clinical studies. *Journal of Cardiovascular Electrophysiology*. 2002, 13, 310–311.
97. CHARITOS, E. I., STIERLE, U., ZIEGLER, P. D., BALDEWIG, M., ROBINSON, D. R., et al. A comprehensive evaluation of rhythm monitoring strategies for the detection of atrial fibrillation recurrence: insights from 647 continuously monitored patients and implications for monitoring after therapeutic interventions. *Circulation*. 2012, 126(7), 806–814.
98. CHARITOS, E. I., ZIEGLER, P. D., STIERLE, U., SIEVERS, H.-H., PAARMANN, H., and HANKE, T. Atrial fibrillation density: a novel measure of atrial fibrillation temporal aggregation for the characterization of atrial fibrillation recurrence pattern. *Applied Cardiopulmonary Pathophysiology*. 2013, 17, 3–10.
99. MARTINEZ, J. P., ALMEIDA, R., OLMOS, S., ROCHA, A. P., and LAGUNA, P. A wavelet based ECG delineator: evaluation on standard databases. *IEEE Transactions on Biomedical Engineering*. 2004, 51(4), 570–581.
100. PETRĚNAS, A., MAROZAS, V., SÖRNMO, L., and LUKOŠEVIČIUS, L. An echo state neural network for QRST cancellation during atrial fibrillation. *IEEE Transactions on Biomedical Engineering*. 2012, 59(10), 2950–2957.
101. KINGMA, D. and BA, J. Adam: A method for stochastic optimization. 2014. Available: <https://arxiv.org/abs/1412.6980>.
102. GHAFFARI, A., HOMAINEZHAD, M., KHAZRAEE, M., and DAEVAEIHA, M. Segmentation of Holter ECG waves via analysis of a discrete wavelet-derived multiple skewness-kurtosis based metric. *Annals of Biomedical Engineering*. 2010, 38(4), 1497–1510.
103. THERRIEN, C. W. Discrete random signals and statistical signal processing. *Prentice Hall PTR*. 1992.
104. RAZALI, N. M. and WAH, Y. B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*. 2011, 2(1), 21–33.

105. NEATH, A. A. and CAVANAUGH, J. E. The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2012, 4(2), 199–203.
106. DALEY, D. J. and VERE-JONES, D. An introduction to the theory of point processes: volume I: elementary theory and methods. *Springer*, 2003.
107. HAWKES, A. G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*. 1971, 58(1), 83–90.
108. FLINT, A. C., BANKI, N. M., REN, X., RAO, V. A., and GO, A. S. Detection of paroxysmal atrial fibrillation by 30-day event monitoring in cryptogenic ischemic stroke: the stroke and monitoring for PAF in real time (SMART) registry. *Stroke*. 2012, 43(10), 2788–2790.
109. CERASUOLO, J. O., CIPRIANO, L. E., and SPOSATO, L. A. The complexity of atrial fibrillation newly diagnosed after ischemic stroke and transient ischemic attack: advances and uncertainties. *Current Opinion in Neurology*. 2017, 30(1), 28.
110. EATOCK, J., LIN, Y. T., CHANG, E. T., GALLA, T., and CLAYTON, R. H. Assessing measures of atrial fibrillation clustering via stochastic models of episode recurrence and disease progression. *Computing in Cardiology Conference*. 2015, 265–268.
111. DE MAIO, F. G. Income inequality measures. *Journal of Epidemiology and Community Health*. 2007, 61(10), 849–852.
112. GOLDBERGER, A. L., AMARAL, L. A., GLASS, L., HAUSDORFF, J. M., IVANOV, P. C., et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*. 2000, 101(23), e215–e220.
113. CHICCO, D. and JURMAN, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020, 21(1), 6.
114. THARWAT, A. Classification assessment methods. *Applied Computing and Informatics*. 2018.
115. ANDERSSON, O., CHON, K. H., SÖRNMO, L., and RODRIGUES, J. N. A 290 mV Sub-VT ASIC for real-time atrial fibrillation detection. *IEEE Transactions on Biomedical Circuits and Systems*. 2014, 9(3), 377–386.
116. RIZWAN, M., WHITAKER, B. M., and ANDERSON, D. V. AF detection from ECG recordings using feature selection, sparse coding, and ensemble learning. *Physiological Measurement*. 2018, 39(12), 124007.

117. SHAO, M., BIN, G., WU, S., BIN, G., HUANG, J., and ZHOU, Z. Detection of atrial fibrillation from ECG recordings using decision tree ensemble with multi-level features. *Physiological Measurement*. 2018, 39(9), 094008.
118. GU, Q., ZHU, L., and CAI, Z. Evaluation measures of the classification performance of imbalanced data sets. *International Symposium on Intelligence Computation and Applications*. 2009, 461–471.
119. BEKKAR, M., DJEMAA, H. K., and ALITOUICHE, T. A. Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*. 2013, 3(10), 27–38.
120. LANGLEY, P., DEWHURST, M., DI MARCO, L., ADAMS, P., DEWHURST, F., et al. Accuracy of algorithms for detection of atrial fibrillation from short duration beat interval recordings. *Medical Engineering and Physics*. 2012, 34(10), 1441–1447.
121. LOBO, J. M., JIMENEZ-VALVERDE, A., and REAL, R. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*. 2008, 17(2), 145–151.
122. HANCZAR, B., HUA, J., SIMA, C., WEINSTEIN, J., BITTNER, M., and DOUGHERTY, E. R. Small sample precision of ROC-related estimates. *Bioinformatics*. 2010, 26(6), 822–830.
123. BAALMAN, S. W., SCHROEVERS, F. E., OAKLEY, A. J., BROUWER, T. F., VAN DER STUIJT, W., et al. A morphology based deep learning model for atrial fibrillation detection using single cycle electrocardiographic samples. *International Journal of Cardiology*. 2020, 316, 130–136.
124. OSTER, J. and CLIFFORD, G. D. Impact of the presence of noise on RR interval-based atrial fibrillation detection. *Journal of Electrocardiology*. 2015, 48(6), 947–951.
125. TAJI, B., CHAN, A. D., and SHIRMOHAMMADI, S. False alarm reduction in atrial fibrillation detection using deep belief networks. *IEEE Transactions on Instrumentation and Measurement*. 2017, 67(5), 1124–1131.
126. HENRIKSSON, M., PETRÉNAS, A., MAROZAS, V., SANDBERG, F., and SÖRNMO, L. Model-based assessment of f-wave signal quality in patients with atrial fibrillation. *IEEE Transactions on Biomedical Engineering*. 2018, 65(11), 2600–2611.
127. BASHAR, S. K., HAN, D., HAJEB-MOHAMMADALIPOUR, S., DING, E., WHITCOMB, C., et al. Atrial fibrillation detection from wrist photoplethysmography signals using smartwatches. *Scientific Reports*. 2019, 9(1), 1–10.

128. FALLET, S., LEMAY, M., RENEVEY, P., LEUPI, C., PRUVOT, E., and VESIN, J.-M. Can one detect atrial fibrillation using a wrist-type photoplethysmographic device? *Medical and Biological Engineering and Computing*. 2019, 57(2), 477–487.
129. KWON, S., HONG, J., CHOI, E.-K., LEE, E., HOSTALLERO, D. E., et al. Deep learning approaches to detect atrial fibrillation using photoplethysmographic signals: Algorithms development study. *JMIR mHealth and uHealth*. 2019, 7(6), e12770.
130. TISON, G. H., SANCHEZ, J. M., BALLINGER, B., SINGH, A., OLGIN, J. E., et al. Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiology*. 2018, 3(5), 409–416.
131. BONOMI, A. G., SCHIPPER, F., EERIKAINEN, L. M., MARGARITO, J., VAN DINTHER, R., et al. Atrial fibrillation detection using a novel cardiac ambulatory monitor based on photoplethysmography at the wrist. *Journal of the American Heart Association*. 2018, 7(15), e009351.
132. EERIKAINEN, L. M., BONOMI, A. G., SCHIPPER, F., DEKKER, L. R., VULLINGS, R., et al. Comparison between electrocardiogram-and photoplethysmogram-derived features for atrial fibrillation detection in free-living conditions. *Physiological Measurement*. 2018, 39(8), 084001.
133. CORINO, V. D., LAUREANTI, R., FERRANTI, L., SCARPINI, G., LOMBARDI, F., and MAINARDI, L. T. Detection of atrial fibrillation episodes using a wristband device. *Physiological Measurement*. 2017, 38(5), 787–799.
134. STEINBERG, B. A., LI, Z., SHRADER, P., CHEW, D. S., BUNCH, T. J., et al. Bimodal distribution of atrial fibrillation burden in 3 distinct cohorts: What is ‘paroxysmal’ atrial fibrillation? *American Heart Journal*. 2022, 244, 149–156.
135. BORIANI, G., DIEMBERGER, I., ZIACCHI, M., VALZANIA, C., GARDINI, B., et al. AF burden is important—fact or fiction? *International Journal of Clinical Practice*. 2014, 68(4), 444–452.
136. MAHAJAN, R., PERERA, T., ELLIOTT, A. D., TWOMEY, D. J., KUMAR, S., et al. Subclinical device-detected atrial fibrillation and stroke risk: a systematic review and meta-analysis. *European Heart Journal*. 2018, 39(16), 1407–1415.
137. BUURMA, M., VAN DIEMEN, J. J., THIJS, A., NUMANS, M. E., and BONTEN, T. N. Circadian rhythm of cardiovascular disease: the potential of chronotherapy with aspirin. *Frontiers in Cardiovascular Medicine*. 2019, 6, 84.
138. HANSSON, A., MADSEN-HARDIG, B., and OLSSON, S. B. Arrhythmia-provoking factors and symptoms at the onset of paroxysmal atrial fibrillation:

a study based on interviews with 100 patients seeking hospital assistance. *BMC Cardiovascular Disorders*. 2004, 4(1), 1–9.

139. MITCHELL, A. R. J., SPURRELL, P. A. R., and SULKE, N. Circadian variation of arrhythmia onset patterns in patients with persistent atrial fibrillation. *American Heart Journal*. 2003, 146(5), 902–907.
140. CAPUCCI, A., CALCAGNINI, G., MATTEI, E., TRIVENTI, M., BARTOLINI, P., et al. Daily distribution of atrial arrhythmic episodes in sick sinus syndrome patients: implications for atrial arrhythmia monitoring. *Europace*. 2012, 14(8), 1117–1124.
141. VISKIN, S., GOLOVNER, M., MALOV, N., FISH, R., ALROY, I., et al. Circadian variation of symptomatic paroxysmal atrial fibrillation. Data from almost 10000 episodes. *European Heart Journal*. 1999, 20(19), 1429–1434.
142. YAMASHITA, T., MURAKAWA, Y., HAYAMI, N., SEZAKI, K., INOUE, M., et al. Relation between aging and circadian variation of paroxysmal atrial fibrillation. *The American Journal of Cardiology*. 1998, 82(11), 1364–1367.

LIST OF PUBLICATIONS ON THE SUBJECT OF THE DOCTORAL THESIS

Publications in the journals referred in the *Clarivate Analytics Web of Science* database with impact factor

1. **Butkuvienė, Monika**; Petrėnas, Andrius; Sološenko, Andrius; Martín-Yebra, Alba; Marozas, Vaidotas; Sörnmo, Leif. Considerations on Performance Evaluation of Paroxysmal Atrial Fibrillation Detectors. *IEEE Transactions on Biomedical Engineering*. 2021, vol. 68, iss. 11, p. 3250–3260. [IF 4.756, Q2, 2021].
2. Henriksson, Mikael; Martín-Yebra, Alba; **Butkuvienė, Monika**; Rasmussen, Jakob G.; Marozas, Vaidotas; Petrėnas, Andrius; Savelev, Aleksei; Platonov, Pyotr G.; Sörnmo, Leif. Modeling and Estimation of Temporal Episode Patterns in Paroxysmal Atrial Fibrillation. *IEEE Transactions on Biomedical Engineering*. 2021, vol. 68, iss. 1, p. 319–329. [IF 4.756, Q2, 2021].
3. **Butkuvienė, Monika**; Petrėnas, Andrius; Kravčenko, Julija; Kaldoudi, Eleni; Marozas, Vaidotas. Objective Evaluation of Physical Activity Pattern using Smart Devices. *Scientific Reports, Nature Research*. 2019, vol. 9, iss. 1, p. 1–9. [IF 3.998, Q1, 2019].
4. Bacevičius, Justinas; Abramikas, Žygimantas; Dvinelis, Ernestas; Audzijonienė, Deimilė; Petrylaitė, Marija; Marinskienė, Julija; Staigytė, Justina; Karuzas, Albinas; Juknevičius, Vytautas; Jakaitė, Rusnė; Basytė-Bacevičė, Viktorija; Bileišienė, Neringa; Sološenko, Andrius; Sokas, Daivaras; Petrėnas, Andrius; **Butkuvienė, Monika**; Paliakaitė, Birutė; Daukantas, Saulius; Rapalis, Andrius; Marinskis, Germanas; Jasiūnas, Eugenijus; Darma, Angeliki; Marozas, Vaidotas; Aidietis, Audrius. High Specificity Wearable Device with Photoplethysmography and Six-Lead Electrocardiography for Atrial Fibrillation Detection Challenged by Frequent Premature Contractions: DoubleCheck-AF. *Frontiers in Cardiovascular Medicine*. 2022, vol. 9, p. 1–11. [IF 5.846, Q1, 2021].

Publications referred in the *Clarivate Analytics Web of Science* database without impact factor

1. **Butkuvienė, Monika**; Petrėnas, Andrius; Sološenko, Andrius; Martín-Yebra, Alba; Marozas, Vaidotas; Sörnmo, Leif. Atrial Fibrillation Episode Patterns and Their Influence on Detection Performance. *Computing in Cardiology (CinC)*: September 12–15, 2021, Brno, Czech Republic. 2021, vol. 48, p. 1–4.
2. Martín-Yebra, Alba; Henriksson, Mikael; **Butkuvienė, Monika**; Marozas, Vaidotas; Petrėnas, Andrius; Savelev, Aleksei; Platonov, Pyotr G.; Sörnmo, Leif. Model-Based Characterization of Atrial Fibrillation Episodes and its Clinical

Association. *Computing in Cardiology* (CinC): September 13–16, 2020, Rimini, Italy. 2020, vol. 47, p. 1–4.

3. **Butkuvienė, Monika**; Petrėnas, Andrius; Marozas, Vaidotas; Henriksson, Mikael; Bacevičius, Justinas; Aidietis, Audrius; Sörnmo, Leif. Quantitative Evaluation of Temporal Occurrence Patterns of Paroxysmal Atrial Fibrillation. *Computing in Cardiology* (CinC): September 23–26, 2018, Maastricht, Netherlands. 2018, vol. 45, p. 1–4.

List of attended conferences

1. **Butkuvienė, Monika**; Petrėnas, Andrius; Sološenko, Andrius; Martín-Yebra, Alba; Marozas, Vaidotas; Sörnmo, Leif. Atrial Fibrillation Episode Patterns and Their Influence on Detection Performance. *Computing in Cardiology* (CinC): September 12–15, 2021, Brno, Czech Republic. Award: semi-finalist of the Rosanna Degani Young Investigator Award (YIA).
2. **Butkuvienė, Monika**; Petrėnas, Andrius; Marozas, Vaidotas; Henriksson, Mikael; Bacevičius, Justinas; Aidietis, Audrius; Sörnmo, Leif. Quantitative Evaluation of Temporal Occurrence Patterns of Paroxysmal Atrial Fibrillation. *Computing in Cardiology* (CinC): September 23–26, 2018, Maastricht, Netherlands.

Patents registered by the State Patent Bureau of the Republic of Lithuania

1. Poinšultinės būklės pacientų netrukdančios ilgalaikės prieširdžių aritmijų stebėsenos ir charakterizavimo sistema ir būdas / išradėjai: V. Marozas, A. Petrėnas, A. Sološenko, S. Daukantas, **M. Butkuvienė**, A. Rapalis, J. Bacevičius, A. Aidietis; savininkai: Kauno technologijos universitetas, Vilniaus universitetas. LT 6743 B. 2020-07-10. [Lietuvos Respublikos patentų duomenų bazė; Espacenet]
2. Būdas ir tą būdą įgyvendinanti biomedicininė elektroninė įranga stebėti žmogaus būseną po insulto / išradėjai: D. Jegelevičius, A. Lukoševičius, V. Marozas, A. Petrėnas, S. Daukantas, **M. Butkuvienė**, D. Rastenytė, V. Matijošaitis, K. Laučkaitė, R. Makauskienė, M. Mikulėnas, F. Boero, A. Sansalone; savininkai: Kauno technologijos universitetas, Lietuvos sveikatos mokslų universitetas, UAB Gruppo FOS Lithuania. LT 6729 B. 2020-04-10. [Lietuvos Respublikos patentų duomenų bazė; Espacenet]

Patent applications

1. A long-term non-invasive system for monitoring and characterizing atrial arrhythmias in patients with post-stroke conditions : United States patent application / inventors: V. Marozas, A. Petrėnas, A. Sološenko, S. Daukantas,

- M. Butkuvienė**, A. Rapalis, J. Bacevičius, A. Aidietis; applicant: Kaunas University of Technology, Vilnius University. US 2022015683 A1. 2022-01-20. [USPTO Patent Full-text and Image Database; Espacenet]
2. A long-term non-invasive system for monitoring and characterizing atrial arrhythmias in patients with post-stroke conditions : Japan patent application / inventors: V. Marozas, A. Petrėnas, A. Sološenko, S. Daukantas, **M. Butkuvienė**, A. Rapalis, J. Bacevičius, A. Aidietis; applicant: Kaunas University of Technology. JP 2022508124 A. 2022-01-19. [Espacenet]
 3. A long-term non-invasive system for monitoring and characterizing atrial arrhythmias in patients with post-stroke conditions : European patent application / inventors: V. Marozas, A. Petrėnas, A. Sološenko, S. Daukantas, **M. Butkuvienė**, A. Rapalis, J. Bacevičius, A. Aidietis; applicant: Kaunas University of Technology, Vilnius University. EP 3883463 A2. 2021-09-29. [Espacenet]
 4. Method and biomedical electronic equipment for monitoring patient's condition after a stroke : United States patent application / inventors: D. Jegelevičius, A. Lukoševičius, V. Marozas; A. Petrėnas, S. Daukantas, **M. Butkuvienė**, D. Rastenytė, V. Matijošaitis, K. Laučkaitė, R. Makauskienė, M. Mikulėnas, F. Boero, A. Sansalone; applicants: Kaunas University of Technology, UAB Gruppo FOS LITHUANIA, Lithuanian University of Health Sciences. US 2021251505 A1. 2021-08-19. [USPTO Patent Full-text and Image Database; Espacenet]
 5. Method and biomedical electronic equipment for monitoring patient's condition after a stroke : European patent specification / inventors: D. Jegelevičius, A. Lukoševičius, V. Marozas; A. Petrėnas, S. Daukantas, **M. Butkuvienė**, D. Rastenytė, V. Matijošaitis, K. Laučkaitė, R. Makauskienė, M. Mikulėnas, F. Boero, A. Sansalone; applicants: Kaunas University of Technology, Lithuanian University of Health Sciences, UAB Gruppo FOS LITHUANIA. EP 3833248 A1. 2021-06-16. [Espacenet]
 6. A long-term non-invasive system for monitoring and characterizing atrial arrhythmias in patients with post-stroke conditions : international patent application / inventors: V. Marozas, A. Petrėnas, A. Sološenko, S. Daukantas, **M. Butkuvienė**, A. Rapalis, J. Bacevičius, A. Aidietis; applicant: Kaunas University of Technology. WO 2020104986 A2. 2020-05-28. [Espacenet]
 7. Method and biomedical electronic equipment for monitoring patient's condition after a stroke : international patent application / inventors: D. Jegelevičius, A. Lukoševičius, V. Marozas; A. Petrėnas, S. Daukantas, **M. Butkuvienė**, D. Rastenytė, V. Matijošaitis, K. Laučkaitė, R. Makauskienė, M. Mikulėnas, F. Boero,

A. Sansalone; applicants: Kaunas University of Technology, Lithuanian University of Health Sciences, UAB Gruppo FOS Lithuania. WO 2020031104 A1. 2020-02-13. [Espacenet]

UDK 004.421.2+616.12-073.7](043.3)

SL 344. 2022-07-14, 15,25 leidyb. apsk. 1. Tiražas 14 egz. Užsakymas 132.

Išleido Kauno technologijos universitetas, K. Donelaičio g. 73, 44249 Kaunas
Spausdino leidyklos „Technologija“ spaustuvė, Studentų g. 54, 51424 Kaunas

