CARDIAC ARRHYTHMIA MONITORING FROM CLINICAL SETTING TO DAILY LIFE

The research presented in this thesis was performed within the framework of the Eindhoven MedTech Innovation Center (e/MTIC). Parts of this thesis were funded by ITEA3 project, 15032, eWatch.

Cover design by By your side creative
Printed by Gildeprint

A catalogue record is available from the Eindhoven University of Technology Library
ISBN: 978-90-386-5078-4

CARDIAC ARRHYTHMIA MONITORING FROM CLINICAL SETTING TO DAILY LIFE

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof. dr. ir. F. P. T. Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op woensdag 14 oktober 2020 om 13:30 uur

door

Linda Maria Eerikäinen

geboren te Helsinki, Finland

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter	prof.dr.ir. A.B. Smolders
1^{e} promotor:	prof.dr. R.M. Aarts
2^{e} promotor:	prof.dr. L.R.C. Dekker
copromotor:	dr.ir. R. Vullings
leden:	prof.dr.ing. S. Leonhardt (RTWH Aachen)
	prof.dr. I.C. van Gelder (UMC Groningen)
	prof.dr.ir. G. de Haan
adviseur:	dr. A.G. Bonomi (Philips Research)

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

SUMMARY

CARDIAC ARRHYTHMIA MONITORING FROM CLINICAL SETTING TO DAILY LIFE

Cardiac arrhythmias are conditions in which the heart rate or rhythm is abnormal – the heart is beating too fast, too slow, and/or irregularly. Arrhythmias can vary from life-threatening to benign, and when requiring treatment, their detection and diagnosis are important. This can be enabled through monitoring. Because cardiac arrhythmias can occur in different settings, there is also a range of different measurement devices for their monitoring, e.g. 12-lead electrocardiography (ECG), Holter monitors, and implantable loop recorders. The applications vary from hospital to ambulatory setting, from intensive monitoring to intermittent, from invasive to non-invasive, and from short-term to long-term. ECG is the gold standard for diagnosing and monitoring arrhythmias and can be measured with surface electrodes. Similar information can be also obtained invasively, e.g. with implantable devices that can be used for long-term.

Moreover, indications for monitoring are different between clinical and ambulatory settings. In a clinical setting they are defined mainly by factors such as reason for hospital admission, type of operation, and presence or lack of symptoms or complications. For ambulatory monitoring, symptoms and their frequency are the key indicators. In ambulatory monitoring, especially intermittent episodes of atrial fibrillation (AF), which is an arrhythmia increasing risk of stroke and heart failure, are difficult to detect and the arrhythmia can easily remain undiagnosed. Studies have shown that prolonged continuous monitoring of populations at risk with implantable devices increases the percentage of subjects detected with AF, and therefore prolonged monitoring could help in diagnosing AF and starting of appropriate therapy earlier. In order to monitor patients for long-term unobtrusively, alternatives for ECG are needed.

In this thesis, we explore possibilities of using additional measurement modalities to ECG in cardiac monitoring both in a clinical setting and in daily life. More precisely the focus is on pulsatile waveforms: arterial blood pressure (ABP) and photoplethysmography (PPG). For daily life, PPG provides an unobtrusive, low-cost, and easy-to-use measurement modality often used in wrist-worn applications. Compared to ECG, which measures the electrical activity of the heart, pulse waveforms reflect the pressure and volume changes adding therefore complementary information. When considered alone, they can be used for extracting heart beats and deriving information about the heart rate and rhythm.

Part I of the thesis focuses on monitoring in clinical setting. In intensive care, false alarms for life-threatening arrhythmias can be caused, e.g. by a noisy ECG signal due to movement or a detached electrode. In Chapter 2, information from available ABP and/or PPG was used in addition to ECG to reduce false alarms for asystole, extreme bradycardia, extreme tachycardia, ventricular tachycardia and ventricular fibrillation/flutter. After signal quality assessment and selection of best quality signals, alarms were classified by using machine learning, leading to a significant reduction of the false alarms for each arrhythmia. In Chapter 3, accuracy of AF detection with wrist-worn PPG data measured in a clinical setting was compared to AF detection in daily life. It is shown that when using the same methods, the detection accuracy varies between clinical setting and daily life due to different measurement settings and patient characteristics.

Part II of the thesis focuses on daily life settings, in particular on the use of PPG in wrist-worn applications in order to detect AF. In Chapter 4, a set of features was extracted from inter-beat intervals and used for AF detection. The classification performance was compared between ECG and PPG measurements of 24 hours in daily life. When assessing the PPG data quality by using information from simultaneously measured acceleration of the wrist, similar detection performance of AF was obtained with PPG compared to ECG when data during movement was not included in the analysis of PPG data. In Chapter 5, we studied if detecting another arrhythmia that also increases stroke risk, atrial flutter (AFL), could be detected separately from the PPG data, and if adding an AFL detection next to AF detection would improve the accuracy of AF detection. In addition, more information was extracted from the measured data compared to Chapter 4, such as heart rate variability, PPG waveform, and acceleration information, which were combined with machine learning. The results showed that when combining this information, AFL could be detected separately, improving also accuracy of AF detection. In Chapter 6, a new feature from the PPG signal was studied which is present during AF but not during regular rhythms and could be related to the contractile force-interval properties of the heart.

For building a comprehensive view about the state-of-the-art of PPG monitoring of AF and understanding the current knowledge gaps and future directions, the literature is reviewed and discussed in the final chapter along with the contributions of this thesis. To conclude, the information given by pulse waveform measurements, i.e. ABP and PPG, can be useful in cardiac arrhythmia monitoring. In the hospital, where these measurements are often already present, including information from them to the detection of life-threatening arrhythmias can improve the accuracy and reduce the burden of false alarms. For screening AF, PPG provides a promising solution which could be used when the data quality is sufficient. Future research should focus on further devel-

opment of AF detection with PPG in daily life, improving the measurement coverage and accuracy during movement and in the presence of other cardiac arrhythmias. Ultimately, clinical studies are needed to prove the clinical added value of this AF-screening technology.

CONTENTS

1 Introduction **1**
- 1.1 Cardiac arrhythmias . . . 3
  - 1.1.1 Normal conduction in the heart . . . 3
  - 1.1.2 Atrial fibrillation . . . 3
  - 1.1.3 Atrial tachycardias . . . 5
  - 1.1.4 Ventricular arrhythmias . . . 6
- 1.2 Arrhythmia monitoring in clinical setting . . . 7
  - 1.2.1 Indications for monitoring . . . 7
  - 1.2.2 Challenges in hospital monitoring . . . 8
- 1.3 Arrhythmia monitoring in daily life . . . 9
  - 1.3.1 Ambulatory monitoring tools . . . 9
  - 1.3.2 Indications for monitoring . . . 10
  - 1.3.3 Challenges with current tools . . . 11
  - 1.3.4 From electrocardiography to photoplethysmography . . . 11
- 1.4 Objective . . . 14
- 1.5 Outline of the thesis . . . 14

I Monitoring in clinical setting **17**

2 Reducing false alarms **19**
- 2.1 Introduction . . . 20
- 2.2 Methods . . . 21
  - 2.2.1 Beat detection . . . 21
  - 2.2.2 Signal selection . . . 24
  - 2.2.3 Feature computation . . . 27
  - 2.2.4 Classification . . . 28
  - 2.2.5 Performance evaluation . . . 29
- 2.3 Results . . . 30
- 2.4 Discussion and conclusion . . . 31

3 Comparison between hospital setting and daily life **35**
- 3.1 Introduction . . . 36
- 3.2 Data . . . 36
  - 3.2.1 Measurements in hospital conditions . . . 37
  - 3.2.2 Measurements in free-living conditions . . . 37
- 3.3 Methods . . . 39
  - 3.3.1 Features . . . 39
  - 3.3.2 Performance metrics . . . 40
  - 3.3.3 Cross-validation . . . 40

3.4 Results . . . 41
3.5 Discussion . . . 42
3.6 Conclusion . . . 43

II Monitoring in daily life 45

4 Comparison between ECG and PPG 47
4.1 Introduction . . . 48
4.2 Methods . . . 49
4.2.1 Data . . . 49
4.2.2 Preprocessing and data synchronization . . . 50
4.2.3 Features . . . 52
4.2.4 Movement intensity . . . 54
4.2.5 Performance metrics . . . 54
4.2.6 Cross-validation . . . 55
4.3 Results . . . 56
4.4 Discussion . . . 60
4.5 Conclusion . . . 62

5 Arrhythmia detection 63
5.1 Introduction . . . 64
5.2 Methods . . . 65
5.2.1 Data . . . 65
5.2.2 Preprocessing and pulse detection . . . 67
5.2.3 Modeling architecture . . . 68
5.2.4 Features . . . 70
5.2.5 Feature selection . . . 73
5.2.6 Classifiers . . . 73
5.2.7 Cross-validation . . . 74
5.3 Results . . . 77
5.3.1 Training set . . . 77
5.3.2 Test set . . . 77
5.4 Discussion . . . 79
5.5 Conclusion . . . 82

6 Beyond arrhythmia detection 83
6.1 Introduction . . . 84
6.2 Methods . . . 85
6.2.1 Data . . . 85
6.2.2 Features . . . 85
6.2.3 Statistical analysis . . . 87
6.3 Results . . . 87

6.4 Discussion 90
6.5 Conclusion 91

Epilogue **93**

7 Discussion and future prospects **95**
7.1 General conclusions and discussion 96
7.1.1 Reduction of false positive cardiac arrhythmia detections 97
7.1.2 PPG-based atrial fibrillation detection in daily life . . . 100
7.2 Future prospects 108

Bibliography **111**

8 List of author's publications **127**

9 Acknowledgements **131**

10 About the author **135**

Appendices **137**

LIST OF ABBREVIATIONS

ABP	Arterial blood pressure
AC	Alternating current
Acc	Accuracy
ACS	Acute coronary syndrome
AF	Atrial fibrillation
AFL	Atrial flutter
ASY	Asystole
AT	Atrial tachycardia
AUC	Area under curve
AV	Atrioventricular
CoSEn	Coefficient of sample entropy
DC	Direct current
EBR	Extreme bradycardia
ETC	Extreme tachycardia
ECG	Electrocardiography
FIR	Force-interval relationship
FN	False negative
FP	False positive
FPR	False positive rate
HF	Heart failure
HR	Heart rate
IBI	Inter-beat interval
ICD	Implantable cardiac device
ICU	Intensive care unit
ILR	Implantable loop recorder
IPI	Inter-pulse interval

LA	Left atrium
LED	Light emitting diode
LV	Left ventricle
nRMSSD	Normalized root mean square of successive differences
NSR	Normal sinus rhythm
PAC	Premature atrial contraction
PESP	Post-extrasystolic potentiation
PD	Photodetector
PPG	Photoplethysmography
PPPI	Pulse-to-pulse-to-pulse interval
PPV	Positive predictive value
PVC	Premature ventricular contraction
RA	Right atrium
RF	Random Forest
RMSSD	Root mean square of successive differences
ROC	Receiver operating characteristics
RV	Right ventricle
SA	Sinoatrial
SampEn	Sample entropy
SE	Spectral entropy
Sens	Sensitivity
ShE	Shannon entropy
SNR	Signal-to-noise ratio
Spec	Specificity
SPI	Spectral purity index
SQI	Signal quality index
SVPB	Supraventricular premature beat
SR	Sinus rhythm

TN	True negative
TNR	True negative rate
TP	True positive
TPR	True positive rate
VF/VFB	Ventricular fibrillation/flutter
VPB	Ventricular premature beat
VT/VTA	Ventricular tachycardia

1

INTRODUCTION

This first chapter introduces the current practice and challenges in cardiac arrhythmia monitoring, and highlights the motivation for the research conducted in the thesis.

When the pulse is irregular and tremulous and the beats occur at intervals, then the impulse of life fades; when the pulse is slender (smaller than feeble, but still perceptible, thin like a silk thread) then the pulse of life is small.

Huang Huang Ti Nei Ching Su Wen, 2000 BC

THIS observation is from *The Yellow Emperor's Classic of Internal Medicine*, and it is probably the earliest notation of atrial fibrillation (AF), the most common sustained cardiac arrhythmia today. Already in ancient times, physicians associated chaotic irregular heart rhythms, later discovered to be AF, with poor prognosis. Centuries later, in 1928, the relation between cardiac and cerebrovascular death was investigated in detail in the Framingham study.[1] Today, it is well established that AF increases the risk of stroke and other serious health related consequences, such as heart failure, impaired quality of life, hospitalization, and death.

AF is not the only cardiac arrhythmia that can have serious negative consequences to health. Some arrhythmias originating from the ventricles, i.e. from the two lower chambers of the heart, can be life-threatening. The discovery of ventricular fibrillation was made in 1850 by Ludwig and Hoffa and named *"mouvement fibrillaire"* by Edmé Félix Alfred Vulpian in 1874.[2] Initially this rhythm disturbance, which had been initiated by giving the heart electrical shocks, was thought to be an experimental curiosity. However, in 1889 a Scottish physiologist John Alexander MacWilliam advanced the hypothesis that ventricular fibrillation was the primary mechanism of sudden death in humans.[2]

An important step forward in studying cardiac arrhythmias was the invention of electrocardiography (ECG) in 1900 by Dutch physician and physiologist Willem Einthoven,[1] who was awarded with the Nobel prize in 1924. The first ECG recording of AF in a patient was published in 1906 by Einthoven which he describes as *"pulsus inaequalis et irregularis"*.[3] In 1912, the first ECG of ventricular fibrillation was published by August Hoffman. To this date, ECG is still the most important tool for monitoring cardiac arrhythmias.

ECG monitoring can be used both in clinical setting for hospitalized patients and in daily life for ambulatory monitoring. It has an important role in diagnosis and management of arrhythmias, but in both monitoring settings unresolved challenges, such as skin irritation caused by the electrodes in long-term monitoring, still remain. Novel technologies have been developed for heart rate monitoring, such as photoplethysmography (PPG), which is used, e.g. in smartwatches. PPG is also measured in clinical setting, usually for derivation of oxygen saturation of blood. This thesis investigates whether PPG, and in a clinical setting other continuous pulse measurements, can be used as a complementary measurement modality to ECG to overcome some of the

current challenges in arrhythmia monitoring.

1.1 CARDIAC ARRHYTHMIAS

During cardiac arrhythmia the rhythm of the heart is different from normal, i.e. too slow, too fast, or irregular. Arrhythmias differ in origin, symptoms and severity. Alteration of the heart rhythm is due to a disturbance in the normal electrical conduction in the heart. This section first describes the normal conduction in the heart, i.e. during sinus rhythm, followed by a description of different types of arrhythmias.

1.1.1 Normal conduction in the heart

The heart consists of four chambers, i.e. right atrium (RA), left atrium (LA), right ventricle (RV) and left ventricle (LV) (Fig. 1.1a). The electrical impulse, which makes the atria and ventricles contract, originates from the sinoatrial (SA) node that is the natural pacemaker of the heart, located in the upper wall of the right atrium, and it defines the rate of beating. The impulse propagates through the atria and ultimately reaches the atrioventricular (AV) node (Fig. 1.1b). On an ECG, atrial depolarization and contraction is visible as the P-wave. The AV node regulates the delay in passing the impulse from the atria to the ventricles, which permits the optimal filling of the ventricles during atrial contraction. After filling of the ventricles, the impulse propagates over the ventricles initiating a contraction (Fig. 1.1c). On an ECG, ventricular depolarization is denoted by the QRS complex and ventricular repolarization by the T-wave.[4]

1.1.2 Atrial fibrillation

Atrial fibrillation (AF) is a chaotic electrical activation of the atria, causing the ventricles to beat irregularly and at a faster rate. In contrast to atrial activation during sinus rhythm (Fig 1.2a), during AF the activation is irregular and very rapid causing an uncoordinated atrial contraction, which is inefficient (Fig 1.2b). This in turn leads to an irregular ventricular response that is often also very rapid, depending on how the AV node filters the impulses coming from the atrium, and the contractions are weaker than during sinus rhythm.[4]

The rapid and irregular rhythm during AF can be maintained by three different mechanisms. First mechanism is by presence of one or more atrial ectopic foci with irregular conduction towards the rest of the atria that cause an irregular fibrillatory activity.[4] In the second mechanism, one or a small number of

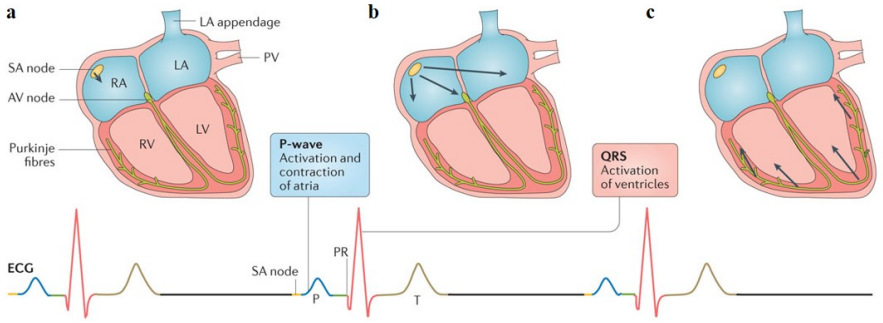

Figure 1.1: Electrical conduction during sinus rhythm: **a** initiation of the contraction from sinoatrial (SA) node, **b** activation and contraction of atria, followed by **c** activation of ventricles ((Reprinted by permission from Springer Nature: Lip et al.,[4]© 2016). RA = right atrium, LA = left atrium, RV = right ventricle, LV = left ventricle, AV = atrioventricular, PV = pulmonary vein.

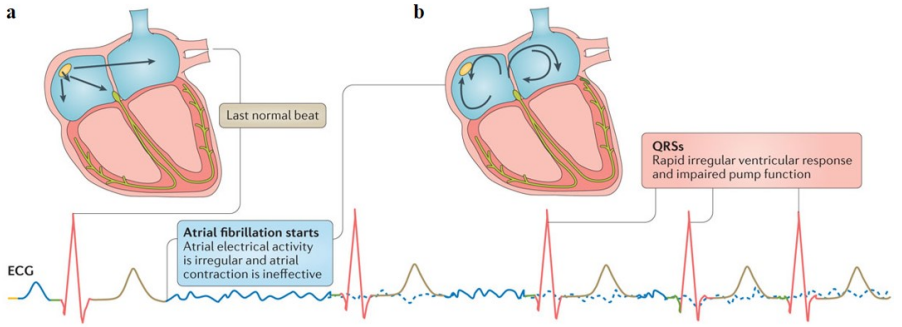

Figure 1.2: Onset of atrial fibrillation (AF): **a** last normal beat before onset of AF, **b** rapid and uncoordinated atrial activity during AF (Reprinted by permission from Springer Nature: Lip et al.,[4]© 2016).

rotors may produce a local activation with fibrillatory conduction causing AF.[4] The third mechanism is by many functional reentry waves.[4] Reentry means that there is a continuous circulating electricity and the impulse, instead of terminating at the end of the contraction, reenters and continues to repetitively excite a region of the heart. In AF, the reentry waves have irregular patterns and no clear activation pattern.[4]

The electromechanical consequences of AF have important clinical implications, such as increasing the risk of stroke and heart failure. The risk of blood coagulation and thrombosis are caused by the absence of effective atrial contraction. The efficiency of ventricular contraction can be also reduced by AF due to the rapid and irregular rhythm, leading to worsening of existing heart

failure or even causing heart failure. Therefore, AF contributes to increased morbidity and mortality as well as impaired quality of life.[4] For preventing stroke, treatment of AF includes administering anticoagulants if the stroke risk is considered sufficiently high. Rate control is intended for maintaining heart rate within desired limits with medication whereas rhythm control is for restoring sinus rhythm and can be achieved either with antiarrhythmic drugs or with electrical cardioversion. Cathether ablation, in which pulmonary veins on an atrial level are isolated, is also an effective way to restore and maintain sinus rhythm, and is usually a treatment option after failure of or intolerance to antiarrhythmic treatment.[5]

AF is a progressive disorder and is classified as paroxysmal, persistent, long-standing persistent or permanent, depending on the duration of the arrhythmia. Definitions of these classes are commonly that paroxysmal AF self-terminates within 7 days, persistent AF lasts continuously more than 7 days, long-standing persistent AF is continuously present more than a year, and eventually AF can be permanent or chronic.[4]

AF is the most commonly experienced sustained rhythm disorder, being a growing epidemic and a major public health issue. Prevalence of AF is estimated to be approximately 3% in adults over 20 years old[5] and can be present in 3-6% in patients admitted in the hospital with acute conditions.[4] In addition, the prevalence is higher in older populations and in patients with conditions such as obesity, hypertension, diabetes, heart failure, coronary artery disease, valvular heart disease, or chronic kidney disease.[5] These conditions are considered as risk factors for developing AF and other risk factors include male sex, cigarette smoking, alcohol consumption, obstructive sleep apnea, chronic obstructive pulmonary disease, hyperthyroidism and family history.[6]

1.1.3 Atrial tachycardias

Atrial tachycardias (AT) are regular atrial rhythms at a constant rate that exceeds 100 beats per minute (bpm) and originate outside of the SA node region.[7] The division of ATs is made by their mechanism to focal and macroreentrant, excluding some tachycardias that cannot be classified by this division. Sustained, focal ATs usually arise with absence of structural heart disease, with greater incidence in middle-aged individuals, but can arise at any age.[8] Focal ATs occur due to an automatic, triggered, or microreentrant mechanism.[7] In macroreentry, the reentry is through a relatively large, and potentially well-characterized circuit. The division of ATs by mechanism is important in planning of ablation therapy[8] in which the atrium is scarred with radiofrequency electrical current to terminate the tachycardia.

Atrial flutter (AFL) is a cardiac arrhythmia that often occurs with structural

heart disease and may also be present in context of a disease, such as sepsis or myocardial infarction.[9] It is less common than AF, occurring <1/10 times compared to AF.[10] These two arrhythmias have a close pathophysiological relation and because of the dependence of flutter on fibrillation, even up to 75% of patients with AFL have been documented to have also AF.[9]

The mechanism of AFL is reentry,[9] and typical atrial flutter (AFL) belongs to macroreentrant ATs. AFLs are divided into typical or atypical based on whether it is dependent on cavotricuspid isthmus, a region in the right atrium. Of the two types, typical AFL is more common.[9]

ATs are associated with stroke risk[11] and the stroke risk for AFL is not much different from AF.[5] The guidelines for administering anticoagulants for AFL are similar to AF.[5]

1.1.4 Ventricular arrhythmias

Ventricular arrhythmias range from premature ventricular contractions (PVC) to ventricular fibrillation (VF) and the clinical manifestations vary from lack of symptoms to cardiac arrest. The risk of ventricular arrhythmia and sudden cardiac death vary in populations with different underlying cardiac conditions, family history, and genetic variants. Most life-threatening ventricular arrhythmias occur with ischemic heart disease, especially in older patients.[12]

Ventricular arrhythmias can be divided in non-sustained and sustained arrhythmias. Non-sustained ventricular arrhythmias include premature ventricular contractions (PVCs) and non-sustained ventricular tachycardia (VT), which is defined as maximum three consecutive contractions originating from the ventricles at rate 100 bpm. PVCs are common and increase in frequency with age. Frequent PVCs, which are defined as 1 PVC in 10 seconds or more than 30 PVCs per hour, are associated to an increased cardiovascular risk and increased mortality in the general population. Because some studies have shown this association to adverse events, the detection of PVCs, especially when frequent and multifocal, leads to an evaluation whether the patient may have an underlying cardiac disease. In patients with cardiovascular disease, PVCs and non-sustained VT are common and are associated to adverse events.[12]

VT and ventricular fibrillation/flutter (VF) can occur during acute coronary syndromes (ACS), i.e. the blood flow in the vessels of the heart is reduced. In patients with ACS and VT or VF, VT and VF are associated to substantially higher risk of death within 90 days from occurring. Sustained VT and VF can also occur in patients without ACS and then they are most often associated with structural heart disease. The risk and predictors of VT in this patient group depend on the type, severity, and duration of structural heart disease.[12]

1.2 Arrhythmia monitoring in clinical setting

As described in the previous section, many cardiac arrhythmias can arise from a structural heart disease, or have an impact on prognosis either in short- or long-term. Therefore, knowing about the presence of an arrhythmia and the type of arrhythmia is important and can be achieved by monitoring.

The standard method for cardiac arrhythmia monitoring is electrocardiography (ECG), which measures the electrical activity of the heart from the body surface with an attached set of electrodes. Figure 1.3 shows examples of typical ECGs during normal sinus rhythm (a), AF (b,c), and AFL (d).

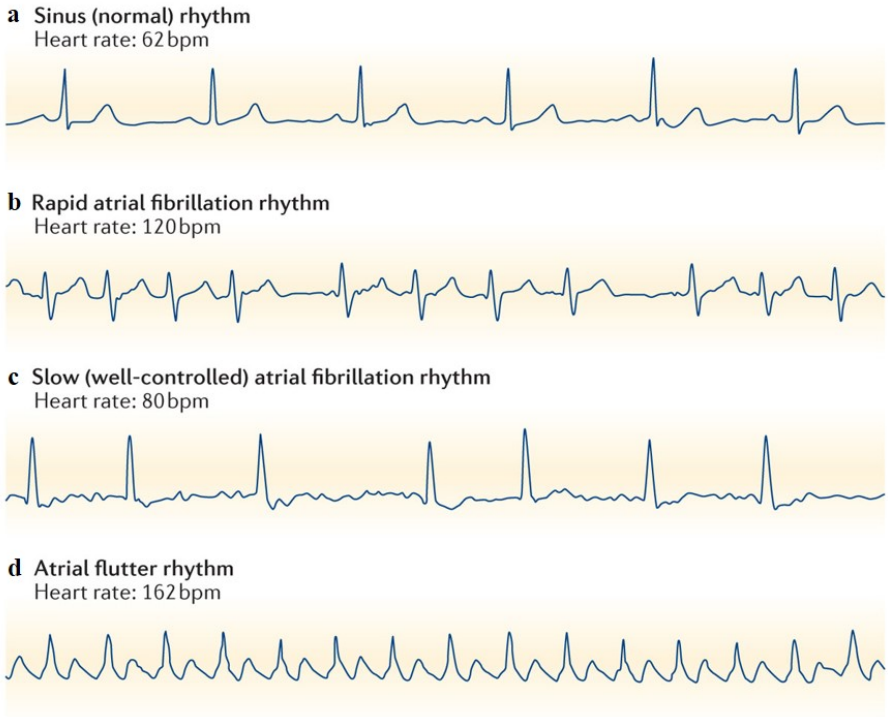

Figure 1.3: Examples of ECG signals during different rhythms (Reprinted by permission from Springer Nature: Lip et al.,[4] © 2016).

1.2.1 Indications for monitoring

The need for and the duration of cardiac arrhythmia monitoring in clinical settings depends on different factors, such as reason for hospital admission, type of operation, and presence or lack of symptoms or complications. According

to practice standards for ECG monitoring in hospital settings,[13] there are four rationales for arrhythmia monitoring:

1. Immediate recognition of sudden cardiac arrest to improve time to defibrillation.
2. To recognize deteriorating conditions that may lead to a life-threatening, sustained arrhythmia.
3. To facilitate management of arrhythmias even if not immediately life-threatening.
4. To facilitate diagnosis of arrhythmias or cause of symptoms and subsequently guide appropriate management.

The practice standards[13] address mainly continuous ECG monitoring. Monitoring should be performed for patients resuscitated from cardiac arrest or unstable VT because of high risk of recurrent arrhythmia. The monitoring should last until the implantation of an implantable cardiac defibrillator (ICD). Patients who have already an ICD and have ICD shocks that require hospital admission should have continuous ECG monitoring during their hospitalization. Continuous ECG monitoring can be considered also for patients with PVCs and non-sustained VT who do not have other indications for monitoring.

Continuous ECG will aid in diagnosis and management of unidentified atrial arrhythmias. Patients should be on continuous ECG monitoring if they are admitted for new onset or recurrent atrial arrhythmias, hemodynamically unstable or symptomatic atrial arrhythmias, or if rate control for arrhythmias is deemed necessary. Hemodynamically unstable arrhythmic conditions should be also monitored until appropriate treatment. It is recommended that patients with transcatheter interventions are monitored with continuous ECG to detect arrhythmia, bundle branch block, and AV block.

Syncope, i.e. loss of consciousness and muscle strength, and stroke also cause a potential need for ECG monitoring. Patients with syncope of unknown cause and meeting admission criteria are recommended to have continuous ECG monitoring. Patients with stroke, when there is suspicion of unknown cause for stroke or occult AF, may have 24-48 hours or longer periods of monitoring. Continuous ECG monitoring should be considered also for patients following non-cardiac surgery due to the risk of AF.

1.2.2 Challenges in hospital monitoring

Many types of patients require arrhythmia monitoring in the hospital and automated detection methods for cardiac arrhythmias have been developed and incorporated in the monitors. In order not to miss any arrhythmia events,

the algorithms in the monitors have very high sensitivity. This in turn leads to a high rate of false positive alarms. A study showed that during a 31-day period there were in total 12 671 audible alarms for six types of arrhythmia and 88.8% of the alarms were false, the false positive rate per arrhythmia type varying from 32.3% to 96.7%.[14] On average the entire alarm burden of all alarms was 187 alarms per bed per day.[14] In another study, from 5386 critical ECG arrhythmia alarms reviewed by multiple experts, 42.7% were found to be false.[15] For the five different arrhythmia types included in the study, the false positive rates varied from 23.1% to 90.7%.[15] In both of the studies, not all the available ECG leads were used for arrhythmia analysis by the monitors in clinical practice.[14,15]

The implications of a large number of false alarms is that the high demand of attention from nurses may delay the interventions when trying to recognize the relevant alarms.[16] In addition, the noise level can exceed 80 dB[17] which causes stress and fatigue.[18] Although policy makers and clinicians bear a significant responsibility in reducing alarm hazard, in a Scientific Statement from American Heart Association technical development is called for from monitor manufacturers to improve the situation.[13]

1.3 Arrhythmia monitoring in daily life

Remote telemetry monitoring was developed to enable ECG monitoring from home for patients suspected with cardiac arrhythmia. Today, there is an availability of various devices, and the development over time has decreased the size of devices and lead to a possibility for automatic arrhythmia detection and wireless data transmission. These developments improve diagnostic yield and efficacy and ease-of-use of the technologies.

1.3.1 Ambulatory monitoring tools

The concept of remote ECG monitoring and the first device for it was introduced by Norman J. Holter, an American biophysicist, in the 1940s. The original Holter monitor could record a single-lead ECG for several hours and provided the method to record and analyze ambulatory ECG data outside standard clinical setting. The system had the size of a backpack and had a reel-to-reel frequency modulated tape recorder, analog patient interface electronics, and large batteries.[19]

The modern Holter monitors are lightweight (200-300 grams) and small in size. They use soft wire patient cables and standard wet gel electrodes that are worn continuously, the recording period being traditionally 24 or

48 hours, but some newer devices can record up to 30 days.[20] The lead configurations in Holters vary between 2-leads, 3-leads and 12-leads. There can be several options for monitoring strategies with Holter monitor, e.g. 24h monitoring every month, 24h every six months, or 24h monitoring in patients reporting palpitations.[21] Holter monitoring traditionally requires active patient participation by keeping a diary about symptoms or recording them by pressing a button.[20]

Wearable on-skin adhesive monitors, so called patch ECG monitors, have embedded electrodes and come with wireless data transfer. They can record continuously single- or 2-lead ECG up to 14 days. For recording symptoms, patients can press a button.[20]

Event recorders are small, leadless devices that patients carry with them and activate the measurement when they experience symptoms. The device is applied to the chest and the electrodes on the back of the device record a brief single-lead ECG recording, which is typically 90 seconds.[22]

Implantable loop recorders (ILRs) are devices that are subcutaneously inserted and intended for arrhythmia monitoring. They are leadless and record single-lead ECG through 2 electrodes within the device and allow up to three years of continuous monitoring. The device can either be triggered automatically or by activation by the patient.[19]

1.3.2 Indications for monitoring

Ambulatory ECG monitoring is typically aimed at determining the cause of symptoms such as dizziness, syncope, chest pain, palpitations, or shortness of breath, because these may correlate with intermittent arrhythmias.[20] To a lesser extent, ambulatory monitoring can be used in patients at potential risk of sudden cardiac death to identify PVCs or nonsustained VT.[19] In addition, monitoring is used to evaluate response of a patient to initiation, revision, or discontinuation of arrhythmic drug therapy and to assess prognosis in specific clinical contexts.[20] In patients with suspected or known AF, it is important to identify the occurrence of AF and a standard 12-lead ECG, which lasts 10 seconds, is not often sufficient to guide clinical management. Symptoms for AF are frequently absent, especially in elderly patients, and the diagnosis needs to be made based on ECG.[4] Because AF can be paroxysmal, monitoring may occur between arrhythmia episodes, and therefore prolonged monitoring can be beneficial.

The duration and device chosen for monitoring depends largely on the frequency of symptoms. For more infrequent events, longer term ECG monitoring is needed. Understanding the correlation between symptoms and arrhythmias, or the lack of it, is key. A typical clinical workflow may include ambulatory

ECG monitoring from 24h up to 7 days, and if that is not successful, followed by intermittent external loop recording that can last from weeks to months. ILRs may be necessary for patients that remain undiagnosed after prolonged long-term monitoring.[20]

1.3.3 Challenges with current tools

In asymptomatic patients or when arrhythmic events occur infrequently, prolonged monitoring up to several weeks or even months may be needed. Especially in the case of AF, several studies have shown that prolonged monitoring increases the detection of AF.[23–25] However, with external devices when used for a longer period of time, poor tolerability to wire-electrode system and adverse skin reactions cause challenges to patient compliance.[20] Prolonged Holter monitoring of 28 days was poorly tolerated in a pilot study.[26]

Patient activated event or loop recorders can be used for several weeks at a time, which increases the monitoring duration compared to a Holter monitor or an ECG patch, but by design the recording is intermittent and they do not record asymptomatic events. In addition, the initiation of the arrhythmia and short arrhythmias that terminate before the recording starts are missed.

For long-term monitoring, ILRs are well tolerated and require no cooperation from the patient when the data transferring from home is automated. However, they have a high initial cost and require significant resources to process the recordings.[21]

1.3.4 From electrocardiography to photoplethysmography

In current ambulatory ECG monitoring the key points are 1) long-term monitoring capabilities, 2) instant feedback, 3) simplicity, 4) every-day use, and 5) non-invasive nature.[27] Continuous long-term monitoring increases the diagnostic yield, but as described in the previous section, electrodes can cause skin irritation in long-term use which reduces patient compliance of external monitors. Moreover, the implantable devices have a significant initial cost and are semi-invasive which prevents them to be the first choice as monitoring strategy.

A tool for continuous heart rate monitoring, which has been used in fitness and activity monitoring, are wrist-worn wearables, such as smartwatches. These are non-invasive, can give instant feedback, are simple to use on a daily basis and therefore can be worn for long periods of time. Heart rate monitoring with smartwatches is usually based on an unobtrusive measurement modality called photoplethysmography (PPG).[28–30] PPG is based on emitting light to the skin with an light emitting diode (LED) and measuring the light

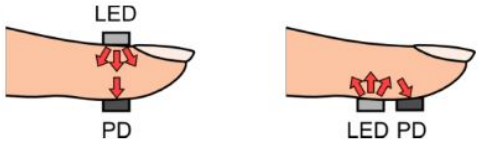

Figure 1.4: Light-emitting diode (LED) and photodetector (PD) placement for transmission- and reflectance-mode photoplethysmography (PPG); from Tamura et al.[29]

that is transmitted through or reflected back with a photodetector (Figure 1.4). The principle of PPG is based on the different absorbtion rates by different substances, such as pigments in the skin, bone, and arterial and venous blood, when the light travels through biological tissue.[29]

Figure 1.5 shows an example of a PPG waveform with alternating current (AC) and direct current (DC) components.[29] The DC component represents the reflected or transmitted optical signal from the tissue, which is influenced by the structure of tissue and the average of arterial and venous blood, and varies with respiration.[29] Changes in the AC component depend on the blood volume changes in the tissue between systolic and diastolic, i.e. contracting and relaxation, phase in the cardiac cycle.[29] Every pulse reflects a heart beat and the fundamental frequency of the AC component, which is superimposed onto the DC component, reflects heart rate.[29] As individual heart beats can be recognized, it is possible to calculate inter-beat intervals (IBI) from the PPG signal that correspond to RR intervals obtained from ECG.

Especially for the detection of AF, which can be without symptoms and paroxysmal, easy-to-use low-cost technologies could possibly increase diagnostic yield. Diagnosis of AF from ECG is based on complete irregularity of the RR intervals and absence of P-waves, and fibrillation waves in the baseline of ECG.[4] Many automated methods for AF detection from ECG have been developed over the years. These methods include assessment of RR interval irregularity, assessment of RR interval irregularity and absence of P-waves, and deep learning on ECG signals.

Irregularity of RR intervals for AF detection has been assessed by modeling the RR interval series as a Markov process,[31,32] calculating coefficient of variation,[33] by a rule based approach using deterministic automation,[34] analyzing Poincaré plots,[35–38] and by extracting statistical, heart rate variability, and/or entropy features.[39–42] As the diagnosis of AF from ECG is based on both RR irregularity and absence of P-waves, some methods have incorporated an additional P-wave analysis to the irregularity assessment.[32,43,44] This has been especially beneficial for reducing false positive detections. The recent

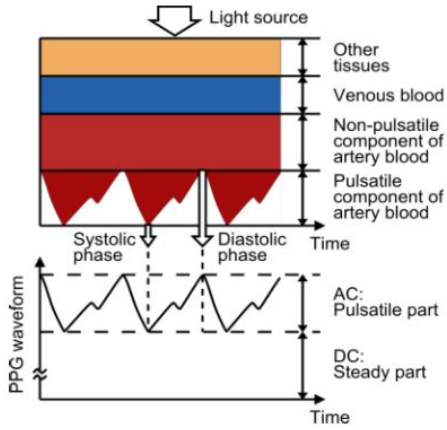

Figure 1.5: Variation light attenuation by tissue; from Tamura et al.[29]

developments towards approaches with deep learning have lead to AF detection methods using deep neural networks with inputs, such as RR interval series,[45] spectrograms,[46] and raw ECG signals.[47,48]

As ECG based methods assessing RR interval irregularity have been successful in AF detection, and extracting information corresponding to RR intervals is possible from PPG, exploring detection of AF from PPG creates an interesting topic for research. At the time of starting this thesis work, promising results in AF detection with PPG had been shown with smartphone-based applications that can measure PPG with a camera of a phone.[49–51] However, these applications only provide a snapshot and do not address the need for long-term continuous monitoring solutions. In contrast, wrist-worn wearables can monitor continuously and long-term. Over the years 2016-2019, there has been a growing interest towards detecting AF with wrist-worn wearables measuring PPG, and in addition to the work presented in this thesis, the final chapter summarizes this development. Figure 1.6 illustrates a wrist-worn wearable and PPG signals measured during different cardiac rhythms and movement artifacts.

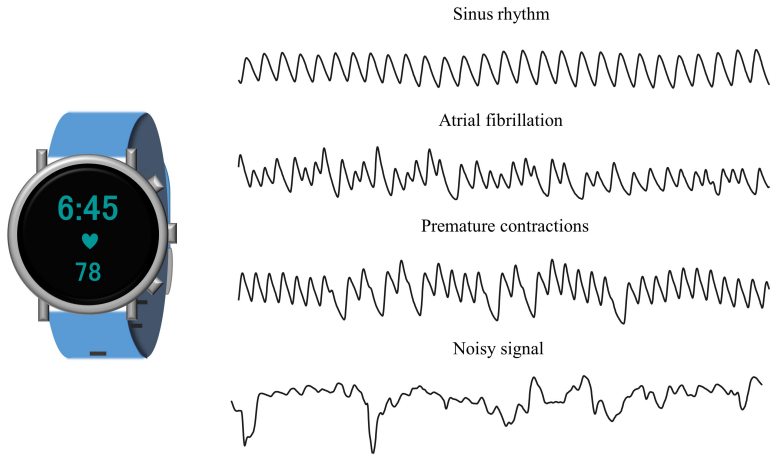

Figure 1.6: Wrist-worn wearable and signals during different cardiac rhythms.

1.4 Objective

The challenges in arrhythmia monitoring in clinical setting are from a technological perspective related to the high sensitivity of automated detection methods for arrhythmias which cause a large number of false alarms. In daily life, the challenges in long-term monitoring are related to the current devices available for continuous monitoring: cost and invasiveness. The research work in this thesis aims to study (1) whether the accuracy of arrhythmia detection in clinical setting can be improved by adding information from pulsatile waveforms to ECG, and (2) whether in daily life photoplethysmography can be used for arrhythmia monitoring.

1.5 Outline of the thesis

This thesis is subdivided into two parts: *Part I - Monitoring in clinical setting* and *Part II - Monitoring in daily life*.

Part I The focus in this part is on the monitoring and analysis of data recorded in the hospital. Chapter 2 addresses the issue of excessive number of false alarms for life-threatening arrhythmias in intensive care. An approach is presented that utilizes both ECG data and available measurements of arterial blood pressure and/or PPG to select the best quality signal for arrhythmia detection. The selected measurements are then used as input to a machine

learning algorithm to classify occurred cardiac arrhythmia alarms as either true or false, leading to reduction of false alarms and still maintaining high sensitivity.

The following chapter (Chapter 3) proceeds from intensive care to measuring patients undergoing elective cardioversion as treatment for AF. With the aim of AF detection from wrist-worn PPG, this data was used to extract IBI features for AF detection and the results of this method were assessed also with data coming from daily life. This chapter shows that methods developed in relatively controlled hospital settings are not necessarily generalizable to a daily life setting.

Part II After the comparison between hospital setting and daily life, *Part II* focuses solely on measurements in daily life. First in Chapter 4, the challenge of movement artifacts in wrist-worn PPG is approached by using accelerometer data for defining which data segments of PPG are used for AF detection. The accuracy of AF detection between ECG and PPG is compared by varying the threshold for the movement level assessed with the accelerometer data. The results show that with the same detection method, AF detection with PPG obtains similar results to ECG when enough data affected by movement are discarded.

The aim in Chapter 5 is to study if by extracting information from IBIs, PPG waveform and accelerometer, the detection could be extended to multiple rhythms and less data could be discarded compared to Chapter 4 while still maintaining similar accuracy. A multi-rhythm classifier is presented that classifies AF, AFL, and sinus rhythm accompanied with or without premature contractions with high accuracy in daily life.

In Chapter 6, the attention is beyond arrhythmia detection. A novel feature to assess force-interval relationship, a phenomenon present during irregularity and reflecting contractility of the heart, from PPG data is presented. This relationship is altered in failing hearts compared to non-failing hearts, and new biomarkers from wrist-worn PPG could be valuable, especially when AF and heart failure coexist.

The thesis is concluded with Chapter 7 that provides an overview of the development during the years of the thesis work in using wrist-worn wearables measuring PPG for detecting AF and discusses the research and contributions of this thesis work. In addition, the future prospects in arrhythmia monitoring are discussed.

Part I

MONITORING IN CLINICAL SETTING

2

REDUCING FALSE ALARMS

Abstract - *In this chapter, an algorithm is proposed that classifies whether a generated cardiac arrhythmia alarm is true or false. The large number of false alarms in intensive care is a severe issue. The noise peaks caused by alarms can be high and in a noisy environment nurses can experience stress and fatigue. In addition, patient safety is compromised because reaction time of the caregivers to true alarms is reduced. The data for the algorithm development consisted of records of electrocardiogram (ECG), arterial blood pressure (ABP), and photoplethysmogram (PPG) signals in which an alarm for either asystole, extreme bradycardia, extreme tachycardia, ventricular fibrillation or flutter, or ventricular tachycardia occurs. First, heart beats are extracted from every signal. Next, the algorithm selects the most reliable signal pair from the available signals by comparing how well the detected beats match between different signals based on* F_1*-score and selecting the best match. From the selected signal pair, arrhythmia specific features, such as heart rate features and Spectral Purity Index (SPI) are computed for the alarm classification. The classification is performed with five separate Random Forest models. In addition, information on the local noise level of the selected ECG lead is added to the classification. The algorithm was trained and evaluated with the PhysioNet/Computing in Cardiology Challenge 2015 data set. In the test set the overall true positive rates were 93 and 95 % and true negative rates 80 and 83 %, respectively for events with no information and events with information after the alarm. The overall challenge scores were 77.39 and 81.58.*

Based on: L.M Eerikäinen, J. Vanschoren, M.J. Rooijakkers, R. Vullings, H.M. de Morree, R.M. Aarts,"Reduction of false arrhythmia alarms using signal selection and machine learning", *Physiol. Meas.*, vol. 37, No. 8, 2016, ©Institute of Physics and Engineering in Medicine, 2016.

2.1 INTRODUCTION

THE number of false alarms in intensive care units (ICUs) have been reported to be 40–86 %.[5,25,35,4] The noise level in intensive care is high and alarms may reach a noise peak that exceeds 80 dB.[17] Noise can cause stress and fatigue to the nursing staff.[18] In addition, patient safety is compromised because the excessive number of alarms affects the reaction time of caregivers to respond.[55]

The problem of false arrhythmia alarm reduction has been approached with various strategies. These approaches can be roughly divided into filtering methods, signal quality assessment, multi-parametric analysis, machine learning approaches, or combinations of the previous.

The filtering methods aim to suppress the variation and outliers in the signal that may cause the false alarms. Proposed methods include median filtering,[56] statistical signal filtering,[57] and model-based filtering.[58]

A signal quality assessment of electrocardiogram (ECG) by combining several signal quality indices (SQIs) was proposed by Behar et al.[59] to suppress false cardiac arrhythmia alarms. Different SQIs were combined with machine learning methods. Clifford et al.[60] evaluated first the quality of arterial blood pressure (ABP) to either suppress an alarm directly or to use additional information from the ABP signal for the decision if the alarm should be suppressed. Aboukhalil et al.[15] evaluated both timing and signal abnormality information from ABP to suppress false ECG arrhythmia alarms. Instead of ABP, Deshmane[61] used signal quality and onset information of photoplethysmograms (PPG).

The assessment of signal quality also has an important role in multi-parameter approaches. Li et al.[62] estimated heart rate (HR) by fusing ECG, ABP, and PPG, and using SQIs and a Kalman filter. HR features were then used for deciding whether an alarm should be suppressed. Optionally, authors used SQIs and several other features from the signals with machine learning to suppress the false alarms.

The PhysioNet/Computing in Cardiology Challenge 2015 (the Challenge)[63] provided an open data set with ECG, ABP, PPG, and respiratory data from intensive care inviting competitors to develop algorithms to reduce false arrhythmia alarms for five life-threatening arrhythmia types: asystole (ASY), extreme bradycardia (EBR), extreme tachycardia (ETC), ventricular tachycardia (VTA), and ventricular flutter or fibrillation (VFB). The data consists of 750 records for the training set and 500 records for the unrevealed test set. Both in training and test set, half of the records are 5 minutes long and the other half

contains an additional 30 seconds after the alarm. In every record, the alarm occurs at 5 minutes from the beginning of the record. Every record contains 3–4 signals of which two are always ECG leads. The additional signals are either or both ABP and PPG signals, and in some cases a respiratory signal.

Many of the well-performing algorithms the Challenge had signal quality or noise level assessment implemented in them.[64656667] Plesinger et al.[68] used testing of regular heart activity in multiple signals, and if no regular activity was detected, specific arrhythmia test was performed. Fallet et al.[69] had an approach based on robust HR estimation and Spectral Purity Index (SPI). In addition, machine learning approaches were presented by Hoog Antink and Leonhardt[70] and Kalidas and Tamil.[71]

The algorithm presented here compares the detected heart beats from ECG, ABP, and PPG signals, and selects which combination of two signals is the most reliable. The estimation of the reliability of this signal pair is based on the F_1-score[72] of the beats that are detected simultaneously in the two signals. The F_1-score is a measure combining sensitivity and precision and giving both an equal weight. After selecting the signal pair, arrhythmia specific features are computed from both signals. The features consist of HR features and SPI features. In addition, the algorithm uses the F_1-score as a quality feature for the classification. The classification between true and false alarms is performed with five separate Random Forest classifiers, one for each arrhythmia type. The performance is compared between two sets of features: having the the F_1-score as only quality feature and when adding local noise level around R-peaks as an additional feature.

2.2 METHODS

The overall flowchart of our alarm reduction algorithm is presented in Fig. 2.1. In this section we will first describe the beat detection from ECG and pulsatile signals (2.2.1) followed by the signal selection (2.2.2). Next, the feature computation from the selected signal pair is described in section (2.2.3). Finally, classification and performance evaluation are presented in sections (2.2.4) and (2.2.5), respectively.

2.2.1 Beat detection

The first step of our algorithm is to detect heart beats from ECG and pulsatile signals. Before the beat detection, the signals were downsampled from 250 Hz to 125 Hz.

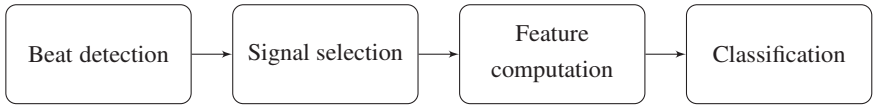

Figure 2.1: Flowchart of the algorithm.

2.2.1.1 ECG beat detection

The beat detection from ECG is performed with a low-complexity R-peak detector.[73] First, in a preprocessing stage, the ECG is convoluted with a Mexican hat wavelet. The absolute value of the convoluted signal produces an output S from which the R-peaks are detected. The absolute value enables the use of a single threshold and suits for our case where the lead of the ECG is unknown.

The detection of R-peaks is executed in an iterative fashion in four stages: segment selection, threshold determination, peak detection, and signal-to-noise ratio (SNR) estimation. The segment for R-peak detection is selected in such a fashion that only one QRS complex is expected in the segment. The limits for the segment are based on the position of the previously detected R-peak and on the assumption that the heart rate is in the range of 32-210 beats per minute (bpm).

The threshold for every R-peak detection is determined by the previous threshold, T_{prev}, and a new threshold estimate, $\widehat{T}$, following a first order autoregressive process

$$T = \alpha \cdot \widehat{T} + (1 - \alpha) \cdot T_{prev}, \tag{2.1}$$

where α is a coefficient describing the dynamic behaviour of the threshold. The value of α was set to the default value 1/3. The new threshold estimate, $\widehat{T}$, is the product of the maximum amplitude of the preprocessed signal, S_{max}, and the local noise level estimate, N_l, in the signal segment.

The N_l is indicative of the local SNR and is scaled to the range of [0, 1]. For the scaled noise level estimate, first an estimate of SNR is needed.

SNR is classically determined by

$$\mathrm{SNR} = 10 \cdot \log_{10}\left(\frac{P_s}{P_n}\right), \tag{2.2}$$

where P_s is the signal power and P_n is the noise power. For the purposes of R-peak detection algorithm, P_s is defined as the power around the detected R-peak $\hat{p}$, $S[\hat{p}]^2$, and P_n as the maximal power N_{max}^2 in the segments between

consecutive R-peaks. N_{max} is the maximum of S in the segments before and after the detected peak. P_s and P_n can now be replaced in Eq. (2.2) and the power of 2 can be taken from the logarithm as a multiple in front of the logarithm. In order to reduce the computational complexity of the function, the decadic logarithm is replaced by a binary logarithm and a low-complexity estimate of SNR is rewritten as

$$\widehat{\mathrm{SNR}} = \log_2(S[\hat{p}]) - \log_2(N_{max}). \tag{2.3}$$

Except for a scaling factor, the low-complexity estimate of the function corresponds to its higher complexity equivalent, and $\widehat{\mathrm{SNR}}$ corresponds to signal quality. The $\widehat{\mathrm{SNR}}$ is further scaled as the local noise level estimate

$$N_l = \frac{6 - \widehat{\mathrm{SNR}}}{8}, \tag{2.4}$$

where a low value of N_l indicates the minimal noise level and $N_l \geq 6/8$ indicates an $\widehat{\mathrm{SNR}}$ below 0 dB.

After determining the threshold, the first preprocessed sample crossing the threshold is the peak position candidate. The peak position candidate is updated if in the vicinity a sample with a higher amplitude is found. If no candidate is found in the segment, two other iterations are performed with an extended segment and lowered threshold. If after three iterations no R-peak is found, the segment moves forward with 1 s.

2.2.1.2 Pulsatile signals beat detection

The pulse detection from the ABP and PPG was performed with an open-source ABP pulse onset detection algorithm, *wabp*.[74] The algorithm is available from PhysioNet.[63]

The key concept of the algorithm is to transform the ABP waveform into a slope sum function (SSF) signal. The purpose of the SSF is to enhance the upslope of the waveform and suppress the remainder of the waveform. Before the SSF transformation, the signal is low-pass filtered to suppress high frequency noise that might affect the onset detection. The windowed and weighted SSF, z, at time i is defined as

$$z_i = \sum_{k=i-w}^{i} \Delta u_k, \quad \Delta u_k = \begin{cases} y_k - y_{k-1} & : \quad y_k - y_{k-1} > 0 \\ 0 & : \quad y_k - y_{k-1} \leq 0 \end{cases} \tag{2.5}$$

where w is the length of the window and y_k the low-pass filtered ABP signal.[74]

The final step of the algorithm is to establish a decision rule for the detection of each SSF pulse onset. This is done in two steps. First, adaptive thresholding is applied to detect the SSF pulses that have appropriate amplitude. Second, a search is employed locally around the detection point to confirm the detection and to identify the likely onset of the pulse.

This algorithm is developed for ABP signals, but was used here for PPG pulse onset detection as well. The scale of the PPG signal was adjusted to correspond to the scale of an ABP signal before the pulse detection.

2.2.2 Signal selection

False alarms are usually caused by an artifact or a disturbance in the signal that leads the physiological event the alarm is intended for, e.g. a flat signal can be misinterpreted as an asystole event because no heart beats are detected. Important in the false alarm reduction is to evaluate what is the quality of the signals or how accurate the features obtained from the signals are. If we can be certain that a feature, e.g. heart rate, is accurately measured, we can rely on the decision based on this feature.

Accurate signal quality measures during arrhythmia are difficult to develop and may not always work as desired. Behar et al.[59] combined several ECG SQIs for classifying ECG signal quality during arrhythmia. Their conclusion was that SQIs should be developed separately for every arrhythmia and sufficient data would be needed to develop classifiers for a quality classification.

In our previous work,[75] the signals selected for the feature computation were always one ECG lead and one pulsatile signal. However, there are cases where both ECG signals or the available pulsatile signal or signals may be corrupted. Figure 2.2 shows an example where both ECG leads are of bad quality, but both pulsatile signals have good quality instead.

In the current approach, the aim is to select for the feature computation the signal pair that has the most accurate beat detection on average. If a beat is detected in several signals, the beat is likely to be a real beat. Since we do not know beforehand in which of the signals the beats are detected most accurately, we cannot determine which of the signals and the detected beats in that signal are the reference. Therefore, measures such as sensitivity, also called true positive rate (TPR), and precision are not adequate for our purpose. They would produce two values per a signal pair because both signals would be needed to be considered as a reference. Sensitivity and precision are defined as

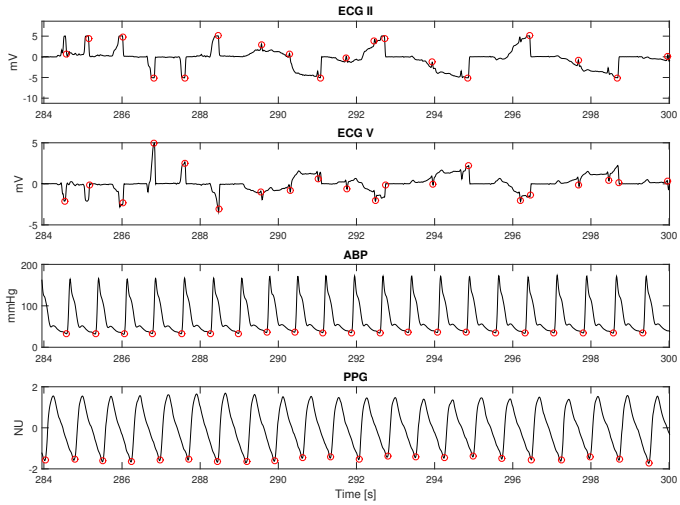

Figure 2.2: An example of bad quality ECG signals when pulsatile signal are good. The red circles indicate detected beats.

$$sensitivity = \frac{TP}{TP + FN}, \tag{2.6}$$

and

$$precision = \frac{TP}{TP + FP} \tag{2.7}$$

where TP are true positives, FP false positives, and FN false negatives.

We select to use a method based on F_1-score to select the signal pair that has the most beats detected in both signals simultaneously. The F_1-score is a harmonized mean between sensitivity and precision

$$\mathrm{F}_1 = \frac{2 \cdot sensitivity \cdot precision}{sensitivity + precision}. \tag{2.8}$$

Inserting the equations (2.6) and (2.7) into (2.8) the F_1-score can be computed as

$$\mathrm{F}_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}. \tag{2.9}$$

In the beat comparison, TPs are the beats detected both in the reference signal

and in the signal that is compared. FPs are the beats detected in the compared signal, but not in the reference signal, and FNs the beats detected in the reference without detection in the compared signal. In a signal pair, when the roles of the signals are changed having now the compared signal as a reference, the number of FPs becomes the number of FNs and vice versa. Therefore, F_1-score gives one score for a signal pair that measures the agreement between the detected beats in the signals and there is no need to select either of the signals as a reference. When F_1-score is 1, all the beats in both signals match with each other, and when F1-score is 0, none of the beats match. In Figure 2.2 the F_1-score between ABP and PPG is 1. Previously, F_1-score has been used as a signal quality measure for ECG by comparing beats detected with different beat detectors from the same signal.[76] In our case, the compared beats are from different signal sources, i.e., ECG, ABP, and PPG.

Between the signals, two beats are considered as a match when they occur within 100 ms from each other. ECG, ABP, and PPG are measured from different locations of the body and are based on different measurement principles. The information about a heart beat is not visible simultaneously. The beat will be visible first in the ECG which is measured from the chest and the hemodynamic response caused by the beat is visible later in ABP and then in PPG. The delay between different signal types is compensated when finding matching beats between ECG and ABP, ECG and PPG, and ABP and PPG. The delay is computed as the mean delay from a period of 10 consecutive beats when the standard deviation of the 10 consecutive delays is less than 5 % of the mean delay. If such a period is not present in the signal then the delay is not compensated.

F_1-score is computed for every signal pair, excluding respiratory signals, in a window before the alarm. The same window will be later used for the selected pair for feature computation which will be explained in Section 2.2.3. Since the length of the arrhythmia event before triggering an alarm varies depending on type of arrhythmia, the window length was optimized for every arrhythmia type separately. The optimized window lengths for different arrhythmia types vary from 14 to 16 seconds.

Based on the F_1-score, the most appropriate signal pair was selected for the feature computation. The signal combination of ECG lead and a pulsatile signal or both pulsatile signals with the greatest F_1-score was selected for ASY, if the F_1-score was greater than zero, and for EBR and ETC if the F_1-score was greater than 0.5. Otherwise, the signal pair with the maximum F_1-score was selected. For VFB and VTA both ECG leads were always selected, and F_1-score was computed for the ECG pair.

2.2.3 Feature computation

During arrhythmia the heart rate is not in the normal range or there are irregularities in the heart beats. For every arrhythmia, one or more features were designed characterizing the arrhythmia based on the definition given in the Challenge description.[63]

EBR and ETC are arrhythmias in which HR is either lower or higher than normal. The features for these two arrhythmias were based on their definitions: the minimum HR of five consecutive beats for EBR and maximum HR of 17 consecutive beats for ETC. During ASY there are no beats for at least four seconds, which can be characterized by the maximum interval between two consecutive beats.

In VFB the heart exhibits a rapid fibrillatory, flutter, or oscillatory waveform for at least four seconds.[63] VTA, on the other hand, is characterized by a number of consecutive beats originating from the ventricles with an HR greater than 100 bpm. Therefore, a measure based solely on heart rate or inter-beat intervals is not sufficient for identifying the two arrhythmia.

Previously, good results for VFB and VTA classification have been reported with Spectral Purity Index.[69] The SPI was initially presented for electroencephalogram (EEG) analysis as a dimensionless parameter between 0 and 1 reflecting the signal bandwidth.[77] The parameter has the maximum value 1 for a pure sine wave and diminishes as the bandwidth of the signal increases. The SPI of a signal is defined as the ratio between the squared, running second-order moment $\bar{\omega}_2$, and the running total power $\bar{\omega}_0$ and fourth-order moment $\bar{\omega}_4$,[78]

$$\Gamma_{SPI} = \frac{\bar{\omega}_2^2(n)}{\bar{\omega}_0(n)\bar{\omega}_4(n)}. \tag{2.10}$$

The spectral moments were implemented in the time domain according to Sörnmo and Laguna.[78] As done in the approach of Fallet et al.,[69] before computing SPI, ECG signals were first downsampled to 35 Hz and smoothed using a 5–sample moving average filter. The window length for estimation of the spectral moments in the time domain was selected to be 4 s for VFB, since the length of the fibrillatory waveform should be at least 4 s. For VTA, the window length was 2 s. The SPI was then averaged in a 1-second window, and the maximum and minimum of the averaged SPI in the window before the alarm were calculated as features.

The arrhythmia specific features used in the classification are listed in Table 2.1. The window length for computing the features varied between the arrhythmias. For ASY and VFB the window was 14 s, for EBR 15 s, and

for ETC and VTA 16 s before the alarm. Moreover, the F_1-score used in the signal selection was added to the feature sets which are given to the classifier as an input. To evaluate whether the F_1-score alone is a sufficient quality feature for the classification, the feature sets for classification were created with and without adding also the median local noise level N_l of the ECG (see Eq. (2.4)).

Table 2.1: Features computed from the selected signal pair

Arrhythmia	Features
ASY	Maximum inter-beat interval
EBR	Minimum heart rate of 5 consecutive beats
ETC	Maximum heart rate of 17 consecutive beats
VFB	Maximum SPI
VTA	Maximum and minimum SPI, maximum heart rate

Two feature sets were created for every arrhythmia type with the above features: one with adding F_1-score and other adding both the F_1-score and median N_l.

2.2.4 Classification

Our algorithm uses five different Random Forest classifiers, each trained for one type of arrhythmia. The choice for the classifier was made after first comparing a large number of classification algorithms. The Random Forests performed overall the best and were therefore selected as the classifier for our algorithm.

A Random Forest is a collection of a large number of tree-structured classifiers in which every tree in the classifier casts a unit vote for the most popular class. The trees in the Random Forest are grown by selecting randomly the inputs or combinations of inputs at each node of the tree to determine the split.[79] Single-tree approaches have been presented previously with good results for integrating multiple signals for artifact detection in neonatal ICU[80] and patient specific alarming models.[81] A binary classification tree was used also in the Challenge entry of Hoog Antink and Leonhardt.[70]

The classification accuracy improves when instead of having a single tree more trees form an ensemble.[79] Random Forests have been previously presented in an alarm classification setting as an analogy to a statistical hypothesis test for "situation is alarm relevant" vs. "situation is not alarm relevant".[82] Several physiological measures were given as an input to the Random Forest and the rate of false alarms was reduced by 45 to 30 % on average. In our approach, we use fewer and more event targeted features as inputs for the Random Forest.

In the final algorithm, the records with F_1-score zero are assigned directly as

false alarms and are not classified with the Random Forest. Therefore, before training the models, feature vectors for records having F_1-score zero were removed from the training set. In total 24 records of false alarms and one true alarm were removed. In addition, the ETC record 't409l', labeled as a false alarm was removed from the training set. In the record, both pulsatile signals had clear recognizable beats that matched completely with each other and the HR was 157-158 bpm for at least 17 consecutive beats, therefore this alarm was assumed to be true.

The classifiers were tested both with 100 and 500 trees. Asystole was the only arrhythmia type for which the performance in the training set improved when increasing the number of trees to 500, and therefore the asystole classifier was selected to consist of 500 trees. The remaining four classifiers consist of 100 trees, since the performance did not improve by adding more trees.

For nearly all the arrhythmia types, the distribution between the two classes, i.e. true and false alarms, was skewed. To balance the class distribution, the classifiers were trained using a cost matrix C,

$$\mathrm{C} = \begin{pmatrix} 0 & \frac{A_{true}}{A_{false}} \\ 5 & 0 \end{pmatrix}, \qquad \text{if } A_{true} \geq A_{false}$$

and

$$\mathrm{C} = \begin{pmatrix} 0 & 1 \\ 5 \cdot \frac{A_{false}}{A_{true}} & 0 \end{pmatrix}, \qquad \text{if } A_{true} < A_{false}$$

where A_{true} is the number of true alarms and A_{false} the number of false alarms in the training set. C(1,2) is the penalty given for a misclassified false alarm and C(2,1) the penalty given for a misclassified true alarm. Misclassification of true alarms is a more severe error than misclassification of false alarms. The multiple five was adopted from the automated score computed by the Challenge test system, which is defined in Section (2.2.5). In the data set for VFB, there were only 6 true alarms compared to 52 false alarms. Hence, for the classifier for VFB, including the multiple of five would have increased the weight too much and was therefore omitted.

2.2.5 Performance evaluation

The performance of the algorithm was evaluated with three different measures: true positive rate (TPR) or sensitivity defined in Eq. (2.6), true negative rate (TNR), and a Challenge score which is a weighted accuracy. The TNR and the score are computed as

$$TNR = \frac{TN}{TN + FP} \tag{2.11}$$

and

$$score = \frac{100 \cdot (TP + TN)}{TP + TN + FP + 5 \cdot FN}. \tag{2.12}$$

In the training set, the estimates for the performance measures were produced with k-fold cross-validation. The number of sets k was set to 10 when there were more than 10 samples in the smaller class. Otherwise, k was set to the size of the smaller class to ensure that there was at least one sample from both of the classes. The k sets were generated in a way that the class distribution in every set represents the class distribution of the training set. For the unrevealed test set, the performance measures were computed by the scoring system.

The algorithm was implemented and evaluated in Matlab 2014b (The MathWorks Inc., Natick, MA).

2.3 RESULTS

The results for the algorithm are listed in Table 2.2. On the left are the results without using information about the median N_l around the R-peaks in ECG and on the right when information about the N_l is added.

Table 2.2: Results without / with local noise level

Arrhythmia	Training set			Test set		
	TPR	TNR	Score	TPR	TNR	Score
ASY	90 / 90	89 / 90	85.72 / **86.63**	83 / 89	97 / 98	88.62 / **92.02**
EBR	95 / 93	85 / 83	**85.00** / 80.46	92 / 92	72 / 72	71.56 / 71.56
ETC	98 / 98	88 / 88	**93.35** / 91.33	100 / 100	0 / 80	95.50 / **99.10**
VFB	83 / 67	94 / 94	**88.58** / 83.11	89 / 78	73 / 96	70.79 / **81.82**
VTA	86 / 89	66 / 63	62.57 / **63.14**	84 / 88	75 / 71	67.56 / **68.14**
Real-time	-	-	-	92 / 93	78 / 80	75.00 / **77.39**
Retrospective	-	-	-	93 / 95	84 / 83	79.20 / **81.58**

The higher score between the two feature combinations is written in bold.

In the training set, adding the local noise level improved the results for two arrhythmia types: ASY and VTA. With the best performing feature combinations, scores of 85 or higher were achieved for all the arrhythmias except VTA. The TPRs were 83–98 % and the TNRs 63–94 %.

The best results in the test set were achieved when the local noise level was

added to the features, except for EBR where no change occurred. The overall TPRs were 93 and 95 % depending on whether the test set was real-time or when additional retrospective data after the alarm was included. The overall TNRs were 80 and 83 % and the scores 77.39 and 81.58 for real-time and retrospective data, respectively.

The best TPR was for ETC and was 100 %, i.e., no true alarms were missed. The TNR for ETC was 80 %. In the set there were 5 false alarms,[63] which means that all except one false alarm were suppressed. The best TNR (98 %) was for ASY. Based on the overall score, the performance in alarm classification was the best for ETC and then for ASY. The worst performance was for VTA both in the training and test set.

2.4 Discussion and conclusion

In this chapter, an alarm classification algorithm was presented based on a signal comparison and selection with F_1-score, computation of arrhythmia relevant features, and a classification with Random Forest classifiers. The signal selection based on F_1-score does not use any signal specific quality information. The F_1-score gives a value that represents how well beats detected in the signals are in agreement with each other. Therefore, F_1-score provides a means to quantify the reliability of features based on detected beats in two signals. This is in contrast to other methods that rely on signal quality estimation of single signals. Different beat detectors may perform differently in the presence of different types of noise and an SQI does not necessarily include the information on how reliable the beat detection is.

The alarm classification when having only F_1-score as a quality measure gave relatively good results. The overall TPR was 92 % and overall score 75.00. When both F_1-score and the median local noise level were used as quality indicators, the overall score increased to 77.39 and overall TPR of the algorithm was 93% in real-time events. This TPR is nearly as good as the best TPR achieved (94 %) with the same test set in the Challenge (September 2015).[63] Misclassifying a true alarm as a false alarm is more severe than not suppressing a false alarm. From a clinical point of view the TPRs should be further increased for alarm suppression algorithms. The score computed for overall comparison of the algorithms weighted misclassified true alarms five times more severe than misclassified false alarms. A higher weight for false negative classifications in the evaluation score could be also considered.

Interestingly, for bradycardia the results remained the same independent of the addition of the local noise level. Moreover, the results are poorer in the test set. Looking at the results in the Challenge, Krasteva et al.[66] had the best

reported score (93.81) for bradycardia alarm classification in the test set. This score was achieved by including features from both ECG and from pulsatile signals in the classification. Their results were better in the test set than in the training set. They also report a relatively large decrease in performance in the test set compared to the training set when only ECG based features are used. This could indicate that ECG based features are not sufficient for accurate classification of bradycardia alarms and that there might be information in the pulsatile signals that is better represented in the test set than in the training set. Adding a feature from pulsatile signals could improve the classification of bradycardia alarms.

For three arrhythmia types, the results improved when the performance was evaluated on the hidden test set. Usually, the results in the training set are better because classifiers are optimized for the training data. This suggests that the training set might not represent the test set completely and the data of the test set might be less noisy. Moreover, the amount of data particularly for VFB and ETC was very small in one of the classes, containing only 8 ETC false alarms and 6 VFB true alarms.[63] Increasing the amount of data could help in producing more robust solutions.

A system with a hidden test data for evaluating the algorithm performances enables a consistent and an objective way to compare different solutions. Using a closed system, however, has also its limitations. Further analysis of the results remains limited because the cases where the algorithm fails in the test set cannot be seen and the differences in data between training and test set cannot be analyzed. In addition, the complete information how the results are produced is not available. Some of the scores given by the system do not seem to comply with the number of false and true alarms reported in[63] if the scores are simply calculated according to the Eq. (2.8). The cause for this discrepancy remains unknown.

The classification model was selected as the Random Forest for all the arrhythmias. In the solution of Hoog Antink and Leonhardt[70] different machine learning techniques were used depending on the arrhythmia. In their approach, very different strategies depending on arrhythmia provided an optimal solution. The performance of our algorithm could be improved by optimizing the selection of the classifier for every type of arrhythmia separately.

The number of false alarms was 61 % and 69 % in the training and test set, respectively. The distribution of false and true alarms varies between arrhythmia types, but for four out of five types the number of false alarms was greater than the number of true alarms. The data is collected from monitors from three different manufactures and from four different hospitals in USA and Europe.[63] Hence, the data does not represent one particular manufacturer or hospital, but a more general situation. There is a need for alarm reduction

algorithms, and our algorithm based on signal selection from multiple signals and alarm classification by machine learning provides promising results. Further improvements can be made by possibly adding features and selecting the classifiers for every arrhythmia separately.

3

COMPARISON BETWEEN HOSPITAL SETTING AND DAILY LIFE

Abstract - *Atrial fibrillation (AF) is the most commonly experienced sustained arrhythmia, and it increases risks of stroke and congestive heart failure. Unobtrusive wearable solutions with photoplethysmography (PPG) have been proposed for AF detection and the performance has been mainly evaluated for short-term measurements in controlled measurement settings. In this study, we evaluate the predictive value of features from PPG for AF detection under both hospital and free-living conditions. PPG from the wrist was measured from 18 patients before and after cardioversion and from 16 patients (4 with 100% AF) for 24 hours. Single-lead ECG and 24-hour Holter were used respectively as gold standards. Six PPG-based inter-beat interval (IBI) variability and irregularity features were computed in three different sliding time windows. Thresholds for AF classification for every individual feature were determined with the data from the hospital conditions and tested with the measurements from free-living conditions. Overall, the best classification results were obtained by using a 120-s window, pNN40 resulting as the best feature. On average, the sensitivity was higher in the hospital conditions (92.3% vs. 71.6%) and the specificity higher in the free-living conditions (60.7% vs. 84.9%). In conclusion, testing the classification perfomance in free-living conditions is essential to properly evaluate AF detection models.*

Based on: L.M. Eerikäinen, L. Dekker, A.G. Bonomi, R. Vullings, F. Schipper, J. Margarito, H.M. de Morree, R.M. Aarts, "Validating Features for Atrial Fibrillation Detection from Photoplethysmogram under Hospital and Free-living Conditions", *Computing in Cardiology*, 2017,

3.1 INTRODUCTION

ATRIAL fibrillation (AF) is the most commonly experienced sustained arrhythmia and its prevalence increases with age. The arrhythmia increases the risk of stroke to five-fold and the risk of congestive heart failure to three-fold.[83] Early diagnosis of AF has a great importance, especially for the prevention of stroke. However, AF can be asymptomatic and therefore can remain undiagnosed. For detecting paroxysmal events, long-term or frequent monitoring is needed.

A measurement technique suitable for unobtrusive long-term monitoring is photoplethysmography (PPG). PPG is an optical measurement, which records blood volume changes in the vascular bed of the tissue, enabling extraction of cardiovascular parameters, such as heart rate.

PPG-based solutions intended eventually for long-term monitoring purposes have been proposed for AF detection based on features determining the irregularity or variability of the inter-beat intervals (IBIs).[49,84,85] However, the results in these studies have been reported only on short-term recordings up to 10 minutes. We previously showed that a Markov model could predict AF in free-living conditions using PPG data.[86] During continuous long-term monitoring the measurements are more prone to noise and the feature values might be less accurate compared to the short-term setting. In addition, different solutions are using different time windows for the feature computation. Here, we evaluate the most common irregularity and variability features for IBIs with three different window lengths, in both a hospital condition before and after an electrical cardioversion procedure, and in a free-living condition during 24-hour measurements.

3.2 DATA

The data for the study were recorded in two different settings in Eindhoven, the Netherlands: before and after an electrical cardioversion (CV) procedure in the hospital and in 24-hour measurements in free-living conditions. The study was approved by the local medical ethical committee and every patient provided written informed consent before participating. An overview of the datasets is presented in Table 3.1. During 24-hour measurements, patients either had 100% AF (4 patients) or no AF.

Table 3.1: Datasets

	Number of patients	Males (%)	Age (y) (m ± sd)	Total rec. length, non-AF (hh:mm)	Total rec. length, AF (hh:mm)
CV	18	56	75 ± 11	13:41	16:26
24h	16	63	65 ± 14	298:32	89:57

3.2.1 Measurements in hospital conditions

The measurements in hospital conditions were performed in the department where patients are treated with electrical cardioversion. 20 patients assigned for AF treatment were included in this part of the study. The patients were measured approximately one hour before and one hour after the procedure with PPG and accelerometer sensors at the wrist with a data logging device equipped with the Philips Cardio and Motion Monitoring Module (CM3 Generation-3, Wearable Sensing Technologies, Philips, Eindhoven). As a reference, a single-lead electrocardiogram (ECG) was measured from the chest with Actiwave Cardio (CamNtech Ltd., Cambridge, United Kingdom). At the beginning and at the end of the recording a synchronization protocol was performed by shaking both devices simultaneously.

The rhythm before and after the procedure was evaluated by a clinical expert by looking at the ECG. Figure 3.1 shows an example of 30 s of PPG signal and corresponding IBIs before and after the cardioversion. Two patients with unsuccessful cardioversion were excluded from further analysis to include only patients with both AF and regular rhythm. Baseline characteristics and medication information of the patients were collected afterwards from the patient record.

3.2.2 Measurements in free-living conditions

The measurements in free-living conditions were performed in 16 patients assigned for a 24-hour Holter examination. PPG and accelerometer data were measured at the non-dominant wrist with the same wrist-wearable device as in the hospital conditions. The ECG was recorded with a 12-lead Holter monitor (H12+, Mortara, Milwaukee, WI, USA). The Holter monitor was attached first to the patient by following normal hospital procedures and a synchronization protocol was performed by tapping the wrist-wearable device and pressing the event button on the Holter simultaneously. The same procedure was performed when the patient arrived at the hospital the following day to return the devices. During the measurement period, patients were keeping a diary of their activities, complaints, and medication. The diary was handed in at the time when the measurement ended.

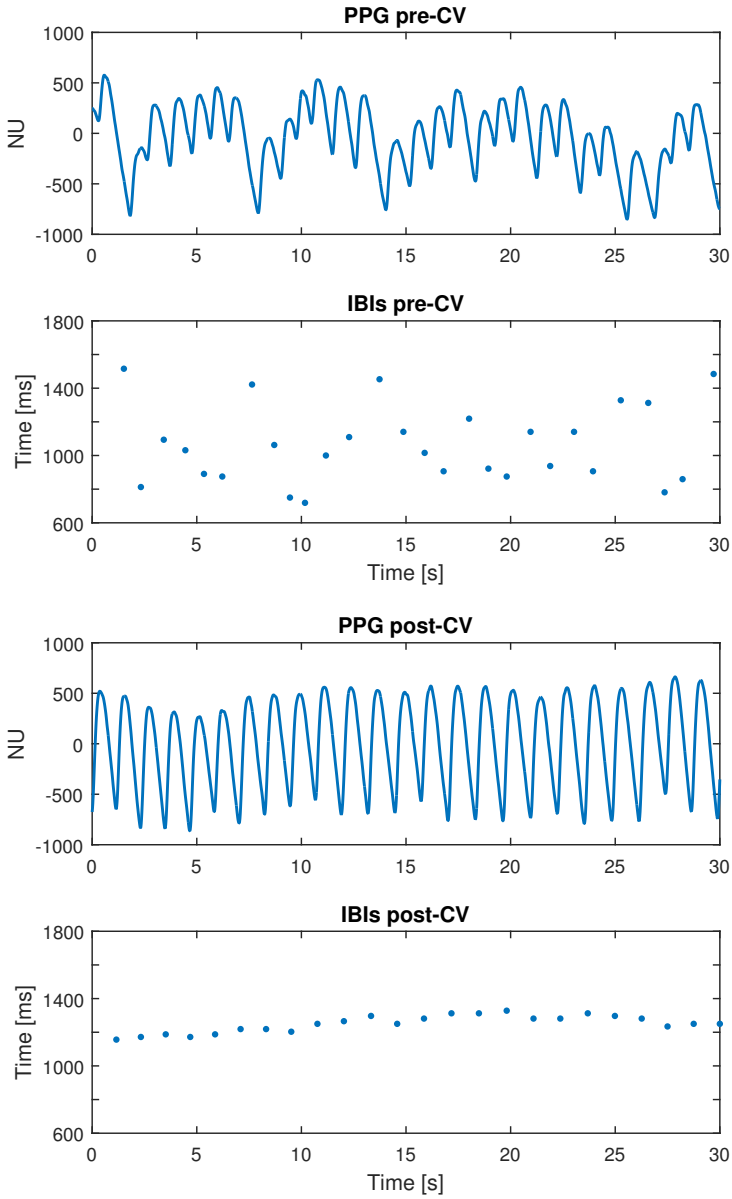

Figure 3.1: 30-s segments of PPG and corresponding IBIs before the cardioversion (above) and after the cardioversion (below) for a representative patient.

The ECG recordings were analyzed by trained analysts, supported by software (Veritas, Mortara, Milwaukee, WI, USA) that automatically detects the time and type of the beat. Every heart beat in the ECG was labelled either as sinus rhythm, AF, premature supraventricular or ventricular contraction,

artifact, or unknown. The output of the software was verified or corrected by the analysts. In addition, baseline characteristics and medication intake information were collected.

3.3 METHODS

The goal of the analysis was to compare features describing irregularity or variability of IBIs and the ability of the features to classify the rhythm as AF and non-AF in the measurement settings described in Section 3.2. Before starting the feature computation, the PPG was filtered with a 0.3 Hz high-pass filter and a 5 Hz low-pass filter. Heart beats were detected from the PPG pulses and after the pulse extraction, the beat information in PPG and ECG were aligned. The IBIs from PPG were computed as time differences between two consecutive pulses.

3.3.1 Features

For feature computation, outlier removal was made based on IBI length and IBIs $<$ 200 ms and $>$ 2200 ms were discarded. The features expressing variability or irregularity of the IBI sequence studied were Root Mean Square of Successive Differences (RMSSD), Shannon Entropy (ShE), the percentage of interval differences of successive intervals greater than 40 ms (pNN40) and greater than 70 ms (pNN70), and Sample Entropy (SampEn). 40 ms and 70 ms were selected based on Corino et al.[85] where the combination of the two features was found to be the most discriminative feature combination for AF.

ShE is a measure that has been successfully used to quantify the irregularity of IBI sequences during AF.[39,49] For calculating the entropy, first the probability distribution of the IBIs is computed assigning the intervals to fixed number of bins with equal size. The probablity of the IBI to fall in the bin i is

$$p(i) = \frac{n_{(i)}}{l - n_{outliers}}, \tag{3.1}$$

where $n_{(i)}$ is the number of IBIs that fall in the bin i, l the total number of IBIs in the window, and $n_{outliers}$ the number of IBIs considered as outliers. When the probabilities for every bin are known, ShE is

$$\text{ShE} = -\sum_{i=1}^{N} p(i) \frac{log(p(i))}{log(N)}. \tag{3.2}$$

N is the number of bins and was selected to be $N = 16$, which is the minimum

number of bins to obtain a reasonable accuracy.[39]

SampEn evaluates similar patterns in the time series and a lower value indicates more self-similarity in the time series. In detail, SampEn is the negative natural logarithm of the conditional probability that two sequences similar to each other at m points are similar also at $m + 1$ points. SampEn was computed according to[87]

$$\text{SampEn} = -\ln(A/B) = -\ln(A) + \ln(B), \tag{3.3}$$

where A is the number of similar sequences of length $m + 1$ and B the number of similar sequences of length m within tolerance r. Two SampEn features were generated by setting m equal to 1 and 2 (SampEn1 and SampEn2), and r equal to 0.25 times the standard deviation of the series as in.[85]

The features were computed in three different window lengths: 30 s, 60 s, and 120 s, by sliding with 30 s.

3.3.2 Performance metrics

The statistical measures used to assess the predictive value of the features were sensitivity (Sens = TP/(TP+FN)), specificity (Spec = TN/(TN+FP)), and accuracy (Acc = (TP+TN)/(TP+FP+TN+FN)), where TP is the number of true positives, TN true negatives, FP false positives, and FN false negatives.

3.3.3 Cross-validation

The measurements in the hospital conditions were considered to be more controlled because of their shorter duration and the patients were in supine position during the entire measurement period. Therefore, the hospital dataset was used as a training set for defining the thresholds for every individual feature for every window length. This was done with a stratified leave-one-out cross-validation. One patient was held as a test data and the set of remaining patients was used to define the threshold which would give an optimum cut-off point on the receiver operating characteristic curve according to Youden index. The procedure was repeated 18 times leaving each patient for testing one time.

The classification to AF and non-AF in the free-living conditions was based on the thresholds defined with the dataset in the hospital conditions. The mean of the thresholds of the cross-validation were selected as the final ones for every feature and window length.

3.4 RESULTS

Table 3.2: Sensitivity and specificity in the hospital

Feature	30 s		60 s		120 s	
	Sens	Spec	Sens	Spec	Sens	Spec
ShE	89.1	54.3	89.1	53.9	88.9	54.5
RMSSD	88.2	44.7	87.9	44.9	88.5	44.2
pNN40	93.5	64.6	94.7	63.9	97.1	66.0
pNN70	91.1	61.1	95.7	59.4	93.9	60.2
SampEn1	86.1	63.6	85.4	68.3	93.2	67.4
SampEn2	88.0	57.5	89.2	66.4	92.3	71.7
Mean	89.3	57.7	90.3	59.5	92.3	60.7

The sensitivity and specificity of the AF classification in the hospital conditions for every window length are presented as mean values over all patients in Table 3.2 and accuracy is presented in Table 3.3. The standard deviation for sensitivity varied between 5.5–25.7%, for specificity between 17.5–41.5%, and for accuracy between 11.4–19.7%. The highest sensitivity (97.1%) and the highest accuracy (83.8%) were obtained with pNN40 and the highest specificity was with SampEn2 (71.7%) with a 120-s window.

Table 3.3: Accuracy in the hospital conditions

Feature	30 s	60 s	120 s
ShE	73.4	73.2	73.4
RMSSD	68.4	68.4	68.4
pNN40	83.8	81.5	83.8
pNN70	80.7	79.9	79.1
SampEn1	75.3	76.4	80.9
SampEn2	74.7	77.8	82.3
Mean	76.1	76.2	78.0

The results for the free-living conditions were calculated only in terms of sensitivity and specificity, and are listed in Table 3.4. The standard deviations for sensitivity ranged between 9.5–13.8% and for specificity 2.8–13.3%. The highest sensitivities were obtained with a 120-s window and were similar for all the features ranging from 69.0% to 72.7%. pNN40 was the feature with the highest specificity (94.3%). Accuracy was not computed due to having only four patients with AF and 12 without AF in the dataset. The accuracies would not be comparable to the accuracies in the hospital dataset which is more balanced.

Table 3.4: Sensitivity and specificity in free-living

Feature	30 s		60 s		120 s	
	Sens	Spec	Sens	Spec	Sens	Spec
ShE	40.0	96.1	60.5	92.5	72.3	89.3
RMSSD	40.0	87.8	60.3	82.8	72.4	78.7
pNN40	40.1	96.5	60.3	95.1	72.3	94.3
pNN70	40.3	96.3	60.7	93.8	72.7	93.1
SampEn1	37.7	78.0	57.1	78.9	70.7	76.5
SampEn2	35.9	73.7	55.7	76.1	69.0	77.6
Mean	39.0	88.1	59.1	86.5	71.6	84.9

3.5 DISCUSSION

This is the first study evaluating features for variability and irregularity of IBIs both in hospital and free-living measurement conditions. When comparing the sensitivity and specificity, the results showed a difference between the two conditions. On average, e.g. with 120-s window length, the sensitivity was higher in the hospital conditions (92.3%) compared to the free-living conditions (71.6%). The specificity, on the contrary, was higher in free-living (84.9%) than in the hospital (60.7%).

In addition to different measurement conditions, the different patient profiles might cause differences in the performance. In hospital conditions, after the electrical cardioversion there might still be irregularities, such as premature contractions, present in the rhythm. Five patients experienced a large number of irregulaties after the procedure which explains the low mean specificity and the high standard deviation of specificity (up to 41.5%). In free-living conditions, the density of premature contractions was lower on average which might explain the higher specificity.

The thresholds for the features were trained with the data recorded in the hospital. A large amount of irregularities in the non-AF group in the data set might have caused the thresholds to be higher than optimal thresholds for the free-living conditions causing the sensitivity to drop.

In free-living conditions, sensitivity improved significantly when increasing the window length. Specificity did decrease, but to a smaller extent. This indicates that a longer window length gives better classification results. This was expected, because the type of features used in this study become more reliable with more data. Interestingly, in the hospital conditions the window length did not seem to influence significantly the classification performance.

3.6 CONCLUSION

The classification performance of the PPG-derived features changed between the hospital and free-living conditions. Thus, testing the classification performance in free-living conditions is essential to properly evaluate AF detection models.

Part II

MONITORING IN DAILY LIFE

4

COMPARISON BETWEEN ECG AND PPG

Abstract - *Background: Atrial fibrillation (AF) is the most commonly experienced arrhythmia and it increases the risk of stroke and heart failure. The challenge in detecting the presence of AF is the occasional and asymptomatic manifestation of the condition. Long-term monitoring can increase the sensitivity of detecting intermittent AF episodes, however that being either cumbersome or invasive and costly with electrocardiography (ECG). Photoplethysmography (PPG) is an unobtrusive measuring modality enabling heart rate monitoring, and promising results have been presented in detecting AF. However, there is still limited knowledge about the applicability of the PPG solutions in free-living conditions. Methods: In this study, we compared inter-beat interval derived features for AF detection between ECG and wrist-worn PPG. The data consisted of 24-hour ECG, PPG, and accelerometer measurements from 27 patients (8 AF, 19 non-AF). In total, seven features (Shannon entropy, Root Mean Square of Successive Differences (RMSSD), normalized RMSSD, pNN40, pNN70, sample entropy, and coefficient of sample entropy (CoSEn)) were compared. Body movement was measured with the accelerometer and used with three different thresholds to exclude PPG segments affected by movement. Results: CoSEn resulted as the best performing feature from ECG with Cohen's kappa 0.95. When the strictest movement threshold was applied, the same performance was obtained with PPG (kappa = 0.96). In addition, pNN40 and pNN70 reached similar results with the same threshold (kappa = 0.95 and 0.94), but were more robust with respect to movement artifacts. The coverage of PPG was 24.0–57.6% depending on the movement threshold compared to 92.1% of ECG. Conclusion: The inter-beat interval features derived from PPG are equivalent to the ones from ECG for AF detection. Movement artifacts substantially worsen PPG-based AF monitoring in free-living conditions, therefore monitoring coverage needs to be carefully selected. Wrist-worn PPG still provides a promising technology for long-term AF monitoring.*

Based on: L.M Eerikäinen, A.G. Bonomi, F. Schipper, L.R.C. Dekker, R. Vullings, H.M. de Morree, R.M. Aarts,"Comparison between electrocardiogram- and photoplethysmogram-derived features for atrial fibrillation detection in free-living conditions", *Physiol. Meas.*, vol. 39, No. 8, 2018, ©Institute of Physics and Engineering in Medicine, 2018.

4.1 INTRODUCTION

ATRIAL fibrillation (AF) is the most commonly experienced arrhythmia. It affects 1-2% of the general population and the prevalence is expected to increase in the coming years. AF increases the risk of stroke, heart failure, hospitalization, and death.[88]

AF is a progressive disease which starts with occasional events, called paroxysmal AF, and slowly progresses to persistent and permanent AF.[88] The challenge in early detection of AF is the occasional nature of the events, but also that AF can be asymptomatic. In a group of patients with paroxysmal AF, the episodes were more often asymptomatic than accompanied with symptoms.[89]

The standard practice for diagnosing AF is with electrocardiography (ECG). However, ECG has its limitations. The sensitivities of 12-lead ECGs and transtelephonic ECGs are between 30–40% whereas for 24/48-hour Holters 44–60%.[27] The added value of prolonged continuous monitoring for diagnosing AF has been shown when monitoring survivors from cryptogenic stroke either continuously with an insertable cardiac monitor (ICM) compared to Holter screening.[2325] In a 3-year period, 8 times more patients were diagnosed with AF with an ICM compared to the Holter control group. Implantable devices are costly and require surgical procedures whereas Holter monitors can cause irritation from the electrodes, are cumbersome to wear, thus not suitable for long-term monitoring. Therefore, there is a demand for more convenient long-term monitoring solutions for diagnosing AF.

Photoplethysmography (PPG) is an unobtrusive measurement modality which enables measuring of different physiological parameters, such as heart rate.[289091] Use of PPG for AF detection has been studied with different technologies, such as with smartphones,[9250499394] with finger probes in a clinical environment[9596] and with wrist-worn devices.[848697988599]

Wrist-worn PPG devices are easy to use and comfortable to wear, therefore providing a promising solution for long-term monitoring. Although the wrist-worn PPG based AF detection algorithms showed promising classification performance, so far there is limited knowledge about their applicability to free-living conditions where the measurements are affected by various type of movement artifacts. The majority of the studies have been conducted in fairly controlled conditions and with short measurements up to 10 minutes[84979885] with only two exceptions. Pantelopoulos et al.[99] have presented results with overnight measurements and in our previous study we presented a Markov-model approach when using 24-hour data.[86]

The aim of this study is to compare state-of-the-art inter-beat interval (IBI) derived features commonly used for AF detection from ECG and PPG in free-living conditions. Information about body movement is used to investigate the effect of movement artifacts to their discriminative power during daily living.

4.2 METHODS

4.2.1 Data

The dataset for the analysis was collected in patients scheduled for a 24-hour Holter measurement. Patients were contacted by a cardiologist and given at least one week to consider the participation to the study. The participants gave a written informed consent before the start of the measurements. The dataset was collected in the Catharina Hospital, Eindhoven, The Netherlands.

The data consisted of 24-hour ECG measurement with a 12-lead Holter monitor (H12+, Mortara, Milwaukee, WI, USA), PPG, and 3-axis accelerometer measurements from the non-dominant wrist with a data logging device equipped with the Philips Cardio and Motion Monitoring Module (CM3 Generation-3, Wearable Sensing Technologies, Philips, Eindhoven). The PPG sensor was based on reflective mode using two green light LEDs. The sampling frequency of both PPG and accelerometry was 128 Hz and the dynamic range of the accelerometer was $\pm$ 8 g.

For syncronization purposes, at the start of the measurement, the event button of the Holter monitor was pressed and data logger tapped at the same time instant. During the recording period patients marked in a diary the daily activities, possible symptoms, and medication intake. At the end of the measurement, patients returned to the hospital and the same syncronization procedure was repeated. Recording devices were detached and the diary was handed in. Information about the daily activities in the diary was not used in this study. In addition, baseline characteristics, medical characteristics, and information about medication were collected.

The ECG data were visually analyzed by a clinical expert using an automated rhythm detection software (Veritas, Mortara, Milwaukee, WI, USA). The software extracted beat times from the ECG and identified every beat either to normal, supraventricular premature beat (SVPB), ventricular premature beat (VPB), AF, paced, artifact, or unknown. The rhythm was then confirmed or corrected by the expert. The raw ECG data was not available for further research purposes, and therefore the beat times and beat labels were used in the data analysis.

In total 30 patients were recruited. 8 patients had continuous AF, 19 patients normal rhythm with premature beats, 2 patients atrial flutter, and one patient had a very noisy ECG reference. The patients with atrial flutter and very noisy ECG reference were excluded from the analysis. The patient characteristics of the remaining 27 patients are presented in Table 4.1.

Table 4.1: Patient characteristics

Baseline characteristics	**AF (N = 8)**	**non-AF (N = 19)**
Sex, male, n (%)	5 (62.5)	10 (52.6)
Age, years, M ± SD (range)	69 ± 11 (43 – 79)	67 ± 13 (34 – 87)
Height, cm, M ± SD (range)	166.5 ± 8.6 (152 – 179)	171.8 ± 8.7 (151 – 185)
Weight, kg, M ± SD (range)	86.5 ± 26.6 (71 – 149)	83.2 ± 20.3 (52 – 113)
BMI, $\mathrm{kg/m^2}$, M ± SD (range)	30.9 ± 7.5 (24.6 – 48.1)	27.9 ± 5.5 (20.2 – 39.3)
Medical characteristics		
Structural heart disease		
Coronary artery disease, n (%)	1 (12.5)	3 (15.8)
Heart failure, n (%)	1 (12.5)	1 (5.3)
Heart valve disease, n (%)	0 (0)	1 (5.3)
Risk factor		
Hypertension, n (%)	4 (50.0)	4 (21.0)
Hyperlipidemia, n (%)	0 (0)	0 (0)
Diabetes Mellitus, n (%)	2 (25.0)	0 (0)
Obstructive sleep apnea, n (%)	1 (12.5)	0 (0)
Medication		
Beta-blocker, n (%)	6 (75.0)	9 (47.4)
Calcium channel blocker, n (%)	4 (50.0)	3 (15.8)
Statin, n (%)	3 (37.5)	8 (42.1)
Anti-arrhythmic drug class I, n (%)	1 (12.5)	5 (26.3)
Anti-arrhythmic drug class III, n (%)	1 (12.5)	2 (10.5)
Digoxin, n (%)	2 (25.0)	1 (5.3)
Anticoagulation, n (%)	8 (100)	15 (79.0)

4.2.2 Preprocessing and data synchronization

The raw PPG data was downsampled from 128 Hz to 64 Hz and bandpass filtered to range from 0.3 to 5 Hz. The pulses were detected by identifying fiducial points in the PPG waveform, i.e. the troughs, by detecting local minima. For finding the local minima, the points where the first derivative goes from negative to positive were selected. To prevent detecting too many local minima, an adaptive threshold was used to exclude locally insignificant ones. The threshold was obtained by filtering the bandpass-filtered PPG signal with a first order lowpass filter (time constant 125 ms). The search for the minima

was enabled only when the PPG signal was below the threshold. Additionally, maximum magnitude of acceleration, after removal of gravity, was assessed every second from the accelerometer. If a threshold of 0.1 g was exceeded, the

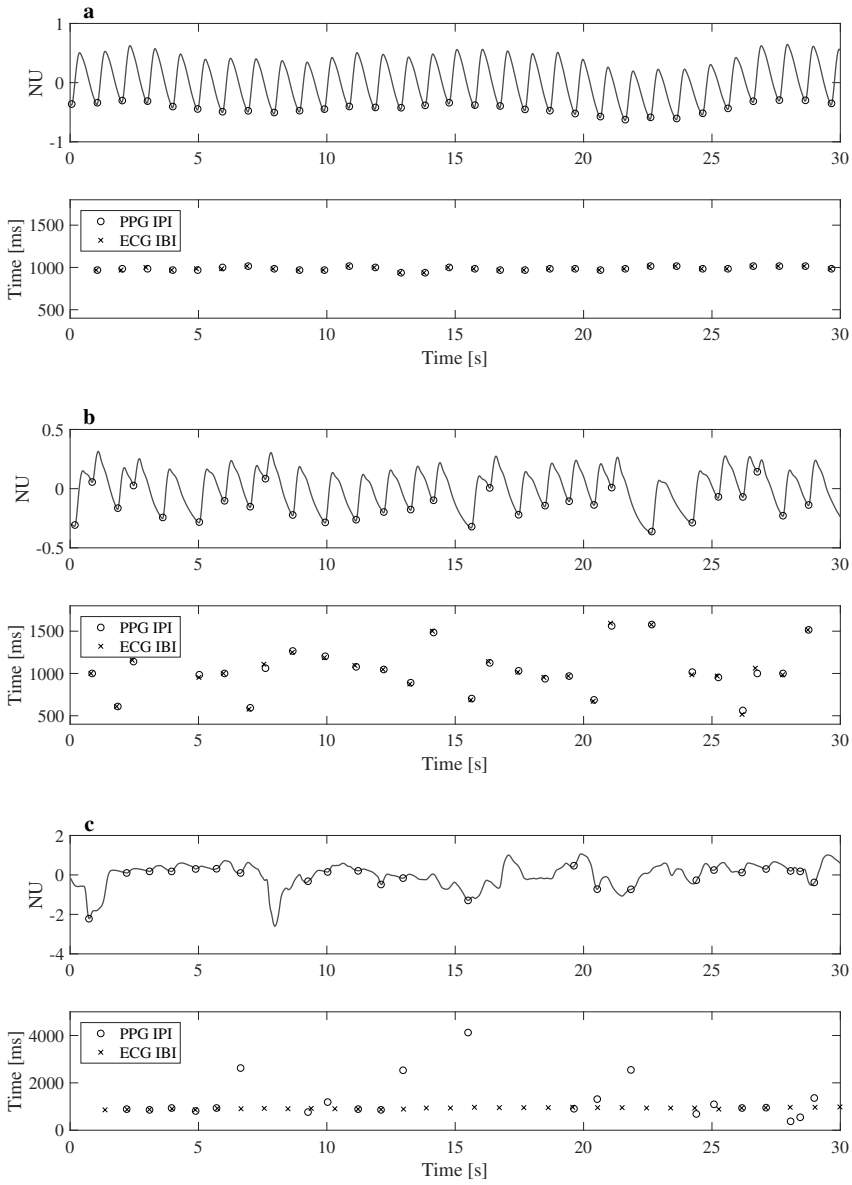

Figure 4.1: Examples of 30s segments of PPG signals (above) and corresponding IBIs and IPIs (below) during sinus rhythm (a), atrial fibrillation (b), and movement (c).

local minima in that period were not considered.

The time between the PPG fiducial points was used to calculate the inter-pulse intervals (IPI). Similarly, the inter-beat intervals (IBI) were calculated from the ECG as the time difference between the beat times given by the Holter software. The raw ECG signal was not available for further data analysis purposes.

The IBI and IPI series were recorded with different devices each of which having their own clock. The clocks may exhibit a time-offset and may run at different speeds, i.e. there might be a drift. For synchronising the IBI and IPI series, first a set of short IPI subsequences was taken from the full IPI sequence. For each IPI subsequence from this set, the best matching position in the IBI sequence was determined. Each match gave a time-offset between the start time of the IPI subsequence and the start time of the best matching position in the IBI sequence. From these time-offsets the clock offset and drift were determined. In addition, the accelerometer signal was aligned based on the offset defined by the fit. In Fig. 4.1 example sequences of 30s of PPG signal and corresponding IBIs and IPIs during sinus rhythm, AF, and movement are presented.

4.2.3 Features

In total seven IBI derived features for AF detection from the literature were compared in this study. The features are pNN40, pNN70, Shannon entropy (ShE), Root Mean Square of Successive Differences (RMSSD), normalized RMSSD (nRMSSD), sample entropy (SampEn), and coefficient of sample entropy (CoSEn). Prior to computing the features, outlier IBIs/IPIs were discarded by removing the ones < 200 ms and > 2200 ms.

In the study of Corino et al.[85] a wide range of features from PPG were analyzed for AF detection. The percentage of differences of successive IBIs that exceed 40 ms or 70 ms (pNN40 and pNN70) were found to be the best discriminative feature combination.

Shannon entropy is an entropy estimate often used to determine regularity of an IBI sequence to distinguish AF.[92504994] The values in the sequence, in this case IBIs or IPIs, are divided in bins and the probability $p(i)$ of a value being in the bin i is

$$p(i) = \frac{n_{(i)}}{l - n_{outliers}}. \tag{4.1}$$

$n(i)$ is the number of values in the bin i, l length of the sequence, and $n_{outliers}$ the number of outliers in the sequence. The bins were equally spaced in the

range from 200 ms to 2200 ms. After having the probabilities for every bin, ShE can be calculated as follows

$$\mathrm{ShE} = -\sum_{i=1}^{N} p(i)\frac{log(p(i))}{log(N)}, \tag{4.2}$$

where N is the number of bins. We used 16 bins as that has been shown to be the minimum number of bins to obtain a reasonable accuracy.[39]

RMSSD and nRMSSD are features used to assess the beat-to-beat variability and have been studied for AF detection from PPG.[9250498594] RMSSD of an IBI sequence of a length l is

$$\mathrm{RMSSD} = \sqrt{\frac{1}{l-1}\sum_{j=1}^{l-1}(IBI(j+1) - IBI(j))^2} \tag{4.3}$$

The nRMSSD is the RMSSD divided by the mean IBI (or IPI) of the sequence.

SampEn assesses the similar patterns in a sequence, lower value indicating more self-similarity. SampEn is the negative natural logarithm of the conditional probability that two sequences that match with each other at m points, i.e. the difference between the two sequences of length m is smaller than tolerance r, they also match when m + 1 points are compared. SampEn was calculated according to[87]

$$\mathrm{SampEn} = -\ln(A/B) = -\ln(A) + \ln(B), \tag{4.4}$$

where A is the number of matches with template length $m + 1$ and B is the number of matches with length m, where m was set to 1, and r 0.25 times the standard deviation of the sequence in line with.[85]

CoSEn is an entropy estimate proposed by Lake and Moorman[40] that is optimized for AF detection and calculated as

$$\mathrm{CoSEn} = \mathrm{SampEn} + \ln(2r) - \ln(\mathrm{mean}(IBI)), \tag{4.5}$$

where r is the tolerance used for computing SampEn.

The features were computed in sliding time windows of 30 s, 60 s, and 120 s with a 30 s shift. Windows that had less than 20, 40, and 80 intervals for 30 s, 60 s, and 120 s window, respectively, were excluded from the analysis.

For the computation of SampEn and CoSEn, the window needed to contain at least 9 consecutive IBIs after removing outliers, otherwise the window was excluded.

4.2.4 Movement intensity

The information from the accelerometer was used to evaluate the movement of the wrist. Movement intensity was defined as

$$Movement\ intensity = \sum_{ax=1}^{3} \left[\frac{1}{l_{acc}} \sum_{i=1}^{l_{acc}} (acc(i)_{ax} - m_{ax})^2 \right] \tag{4.6}$$

where ax is the axis of the accelerometer, l_{acc} the length of acceleration sequence, and m_{ax} the mean acceleration over the sequence on that axis.

Movement intensity was used in the feature computation to discard windows exceeding a predefined movement threshold. Three different thresholds were set for comparison: 75%ile, 50%ile, and 25%ile of the movement distribution of all patients.

4.2.5 Performance metrics

The discriminative power of the features was determined with the following performance metrics: sensitivity, specificity, accuracy, positive predictive value (PPV), F_1-score and Cohen's kappa. When TP being true positives, TN true negatives, FP false positives, and FN false negatives, sensitivity, specificity, accuracy, and PPV are calculated as follows:

$$Sensitivy = \frac{TP}{TP + FN}, \tag{4.7}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{4.8}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{4.9}$$

and

$$PPV = \frac{TP}{TP + FP}. \tag{4.10}$$

F_1-score is a harmonic mean of precision (PPV) and recall (sensitivity) and is based on the efficiency score of Van Rijsbergen[72]

$$\text{F}_1\text{-score} = 2 \cdot \frac{PPV \cdot sensitivity}{PPV + sensitivity}. \tag{4.11}$$

Cohen's kappa[100] is a measure describing the inter-rater agreement of two categorical variables. Kappa is calculated with the following formula

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}, \tag{4.12}$$

where p_o is the observed agreement and p_e the chance agreement. Observed agreement p_o is calculated the same way as accuracy and is the percentage of true observations from all the observations. The chance agreement p_e is

$$p_e = \frac{1}{N_{all}^2} \sum_k n_{k1} n_{k2}, \tag{4.13}$$

where k is the class, N number of observations, and n_{ki} the number of times rater i predicted class k. In our case, the two raters are the reference and the output of the automatic classification based on the feature.

4.2.6 Cross-validation

The cut-off values for AF detection for every feature were determined by leave-one-subject-out cross-validation. The data from one subject were held for testing whereas the data from the remaining 26 subjects were used for training. Due to the imbalance in the number of subjects between the AF and non-AF groups, the data from AF group were upsampled in the training set to balance the class distribution.

The cut-off value for every feature with every window length and movement intensity threshold was determined in the training set by using Receiver Operating Characteristics (ROC) curve and Youden index.[101] Figure 4.2 is an example of the ROC curves and operative points defined by the Youden index during the training phase with features computed from ECG with 30 s window and when data of one patient have been left for testing. For every defined cut-off value, the procedure of holding data for testing from one patient was repeated 27 times.

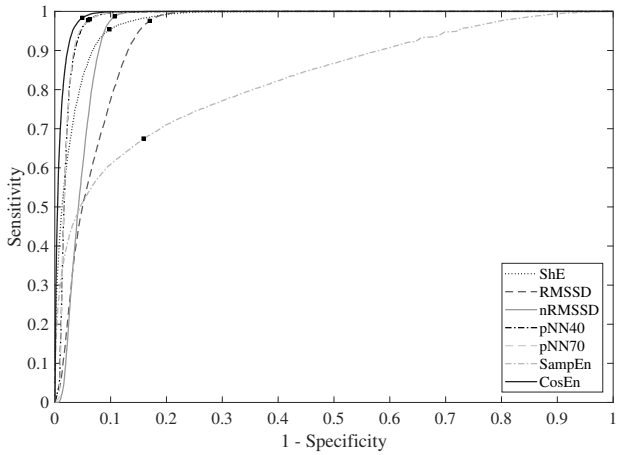

Figure 4.2: An example of ROC curves during the training phase with features computed from ECG with a 30 s window. The black squares on the curve are the operative points defined by the Youden index.

4.3 RESULTS

The AF classification performance of every feature was calculated by aggregating the results obtained with the test data from every round of cross-validation. The results were computed with ECG by varying the window length and with PPG by varying both the window length and movement threshold. Cohen's kappa was selected as the metric to compare the performance of individual features, since it is a metric not affected by the imbalance between the two classes, i.e. AF and non-AF. For the comparison between features computed from ECG, Fig. 4.3 shows a histogram of kappa for every feature when varying the window length. Based on this comparison, CoSEn is the strongest feature from ECG with kappa 0.901–0.950. For every feature the longest window gave the best performance.

The effect of the movement threshold to PPG derived features was compared using a 120 s window. Furthermore, for comparing the effect of the window length, the strictest movement threshold (25%ile) was used. Figure 4.4 shows on the left a histogram of kappa of the features when window length is kept constant but movement threshold varies. On the right, there is a histogram when movement threshold is kept constant and window length varies. When the movement threshold is set to reject more movement, the performance increases for all the features. The results of varying the window length are in line with the results from ECG and with a longer window length better kappa is obtained. CoSEn is again one of the strongest features with $\mathrm{kappa}_{120s(25\%ile)}$ 0.956, but

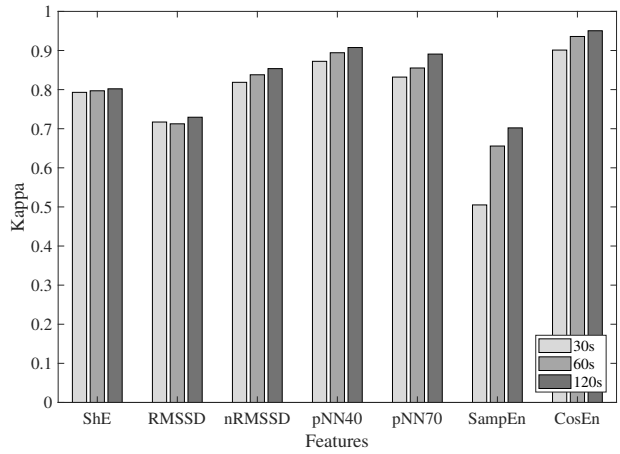

Figure 4.3: Cohen's kappa of features computed from ECG with different window lengths.

additionally pNN40 and pNN70 appear as strong features for AF classification from PPG with $\mathrm{kappa}_{120s(25\%ile)}$ 0.953 and 0.945, respectively.

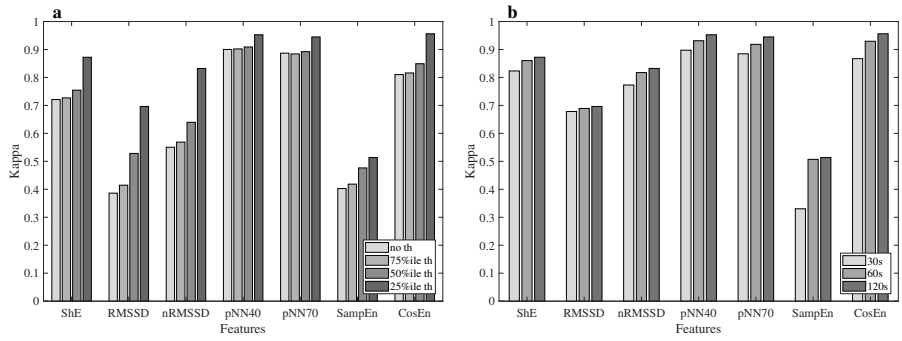

Figure 4.4: Cohen's kappa of features computed from PPG with different movement thresholds using 120 s window (a) and window lengths using 25%ile threshold (b).

Restricting sufficiently the accepted amount of movement resulted in an increase in the performance. However, when windows for feature computation are discarded from the analysis due to the movement, the coverage, which is defined as the percentage of 30 s instants with a feature value, also decreases. In Fig. 4.5, the mean coverage of all patients with ECG and PPG when varying the movement threshold is presented with all the different window lengths. On average, the coverage with ECG with different window lengths is 92.1%. With PPG the average coverage with no movement threshold, 75%ile, 50%ile, and 25%ile thresholds is 57.6%, 54.9%, 45.4%, and 24.0%, respectively. The

movement thresholds are determined separately for different window lengths and therefore varying the window length does not significantly influence the coverage.

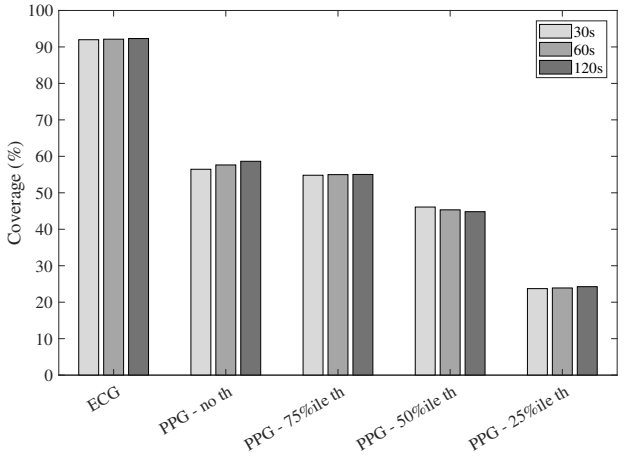

Figure 4.5: Mean overage of ECG and PPG with different movement thresholds and window lengths.

CoSEn resulted as the best feature from both from ECG and PPG with 25%ile threshold when kappa was compared. Table 4.2 shows sensitivity, specificity, PPV, accuracy, kappa, and F_1-score of CoSEn with both ECG and PPG when varying window length and movement threshold. When the 25%ile threshold is used, the classification performance with PPG approaches the results with ECG with all the metrics. With 60 s and 120 s window, the results with PPG are at the level of ECG. Figure 4.6 shows kappa with CoSEn and pNN40 from PPG as a function of movement threshold compared to how the coverage changes when the movement threshold is changed. Kappa with CoSEn from ECG is marked as a reference, since it was the highest kappa obtained with ECG. It is visible how kappa increases when the movement threshold is stricter and both CoSEn and pNN40 from PPG eventually reach the same kappa as with ECG. With higher movement thresholds pNN40 performs better, thus being more robust against movement artifacts. On the contrary to kappa, coverage decreases when excluding more movement from the analysis.

Table 4.2: Performance with CoSEn from PPG and ECG with different window lengths and movement thresholds

Window		Sensitivity	Specificity	PPV	Accuracy	Kappa	F_1-score
30s	ECG	*0.981*	*0.950*	*0.886*	*0.958*	*0.901*	*0.931*
	PPG - no th	0.952	0.819	0.684	0.857	0.691	0.796
	PPG - 75%ile th	0.950	0.829	0.695	0.864	0.703	0.803
	PPG - 50%ile th	0.960	0.866	0.746	0.893	0.761	0.839
	PPG - 25%ile th	0.968	0.934	0.855	0.944	0.867	0.908
60s	ECG	*0.980*	*0.971*	*0.931*	*0.973*	*0.936*	*0.955*
	PPG - no th	0.942	0.884	0.773	0.901	0.777	0.849
	PPG - 75%ile th	0.948	0.890	0.784	0.907	0.791	0.858
	PPG - 50%ile th	0.959	0.916	0.829	0.929	0.837	0.889
	PPG - 25%ile th	**0.975**	**0.968**	**0.928**	**0.970**	**0.929**	**0.951**
120s	ECG	*0.983*	*0.978*	*0.948*	*0.980*	*0.951*	*0.965*
	PPG - no th	0.959	0.898	0.798	0.916	0.810	0.871
	PPG - 75%ile th	0.960	0.901	0.805	0.919	0.816	0.876
	PPG - 50%ile th	0.966	0.919	0.839	0.934	0.849	0.898
	PPG - 25%ile th	**0.984**	**0.980**	**0.955**	**0.981**	**0.956**	**0.970**

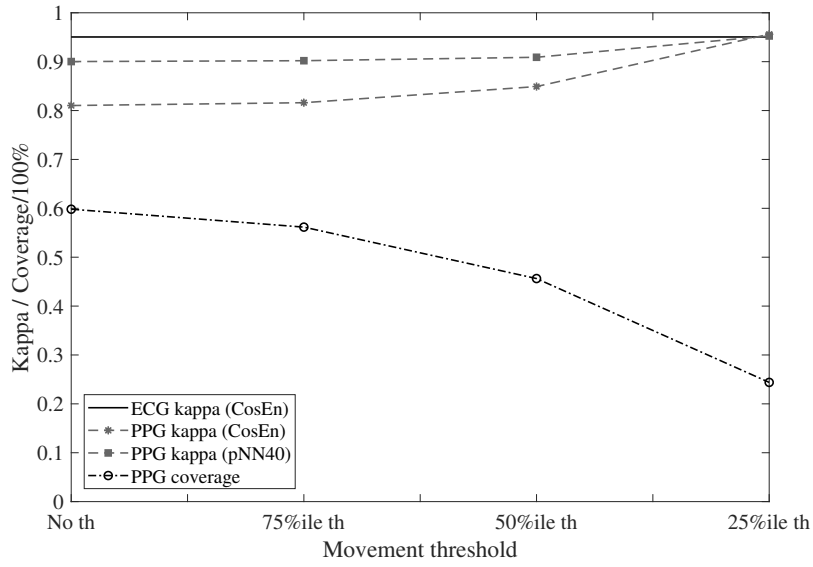

Figure 4.6: Increase of kappa with CoSEn and pNN40 from PPG and decrease of coverage as a function of movement threshold. No movement threshold was applied when computing kappa with CoSEn from ECG, but for easier comparison it is marked as a line of reference.

4.4 DISCUSSION

In this study, we compared for the first time commonly used IBI-features derived from ECG and PPG for AF detection in free-living conditions with 24-hour measurements. CoSEn resulted as the most powerful individual feature from ECG and with strict movement threshold from PPG reaching high sensitivity, specificity, and kappa with both measurement modalities. With the 25%ile threshold for movement, pNN40 calculated from PPG gave similar kappa (0.952) compared to CoSEn (0.956). When accepting more movement, pNN40 performed better, therefore being more robust. Even without using the movement information the coverage was substantially reduced by the movement artifacts, being on average 58% compared to the 92% with ECG. During high intensity movement the pulses were not detected from the PPG and these segments were excluded from the analysis even without using any movement threshold. This also explains why the coverage and performance remain at the same level when 75%ile threshold was applied.

The results indicate that when periods of PPG data affected by movement are discarded from the analysis, i.e. when we expect stable measurement conditions and better signal quality, the PPG works equally well as ECG Holter measurement. The impact of presence of simulated muscle artifact noise on the performance of IBI-based AF detection algorithms for ECG has been previously studied by Oster and Clifford.[102] They showed a linear increase in the performance when SNR increased. In large part our result are in line with their findings. Reducing the movement artifacts had the highest impact on specificity and PPV, thus reducing the false positives. However, on the contrary to the results of Oster and Clifford[102] also sensitivity improved. In our current study, the movement thresholds were not optimized in terms of trade-off between classification performance and coverage. That is left for further research. One option could also be to incorporate movement or signal quality information to the classification model as a feature[9798] to further improve accuracy and increase coverage.

We compared only individual features to make a more objective comparison between measurement modalities, i.e. ECG and PPG. Alternatively, we could have compared the AF detection of a classification model combining either ECG or PPG derived features. With the current way, the comparison is independent from the choice of the feature combination and classification model. These choices might be different when optimized for ECG and PPG, depending also whether PPG is affected by movement artifacts or not. Better classification performance could be possibly obtained when more features are combined. Therefore the classification performance obtained with an individual feature is not intended to reflect the maximum performance that is possible to obtain

with PPG in free-living conditions. Especially, adding information beyond IBI-derived features, such as morphology features[94,99] and spectral features,[98] could possibly further boost the performance.

There are some limitations in the study. The dataset was not large enough to divide the data in a separate training and an unseen test set, and therefore results with cross-validation are presented. There is an imbalance between the two classes which affects some of the performance metrics, such as accuracy, PPV, and F_1-score. Therefore, these metrics are not comparable to the results obtained in other studies with balanced class distributions. In addition, the division into two groups was made solely based on the rhythm, i.e. whether AF was present or not, and the patient characteristics between these groups resulted to be slightly different. However, in such a small dataset, the possible influence of these differences to the results is difficult to assess.

Another limitation of the study is that all the patients with AF had continuous AF. Ideally, the aim was to measure events of paroxysmal AF, but no paroxysmal AF was detected in our study population. It was not possible to determine before the measurement if a patient would have a paroxysmal event during the measurement period. This also reflects the current problem with 24-hour Holter monitoring that rare events are missed, if they occur outside the monitoring period.[27] Therefore, with the current dataset it is not possible to assess how accurately paroxysmal events are detected and whether the window length influences that. Nevertheless, this is the first study comparing ECG and PPG for AF detection during daily life and the results, even with only continuous AF, are promising. Further studies with prolonged PPG measurements to multiple days or weeks, which is difficult and uncomfortable to measure with a Holter, can most likely better capture subjects with paroxysmal events to the study population and give information about their detection with PPG.

As mentioned earlier, even when the movement intensity was not considered in the analysis, the coverage of the rhythm classification with PPG was on average 58% due to the inability to detect pulses. When adding the assessment of movement intensity, the coverage reduced even more. This causes a limitation for the use of IBI-based methods for continuous monitoring to detect paroxysmal events which might occur during the periods when coverage is lost. The detection of these AF events would be therefore partly dependent on the frequency, duration, and daily distribution of the events. However, the lost coverage could be compensated with prolonged monitoring period up to weeks or even months. The impact of lost coverage on the sensitivity of detecting paroxysmal events and the added value of prolonged monitoring with PPG should be assessed with further studies. Moreover, development of methods to improve the coverage could help to overcome the issue with data loss.

In general, the comparison between performance of different algorithms developed for AF detection from PPG is difficult. Algorithms are developed in different settings, and with datasets having different characteristics, e.g. AF vs. subjects with sinus rhythm and AF vs. subjects with other rhythms, such as presence of premature contractions. We have previously shown that results from one measurement setting and patient group are not directly applicable to another setting and patient group with different characteristics.[103] In addition, the amount of data points to compute a feature, i.e. window length, which is not equal between different solutions, influences the results. This was also shown in the work of Tang et al.[96] when comparing models using 1-minute, 2-minute, and 10-minute data.

4.5 Conclusion

Comparable results in AF detection are possible to obtain with PPG and ECG when using a single feature and when discarding PPG signals during movement identified with the accelerometer. On the one hand this leads to a limited coverage, but on the other hand PPG devices can be worn for much longer periods than ECG recorders compensating for the lost coverage. The prolonged monitoring period might have an added value in detecting paroxysmal AF. Therefore, wrist-worn PPG devices provide a promising solution for long-term monitoring of AF. Future studies should be performed to assess the impact of the coverage loss on the sensitivity of detecting paroxysmal AF events.

5

ARRHYTHMIA DETECTION

Abstract - *Objective: Photoplethysmography (PPG) enables unobtrusive heart rate monitoring, which can be used in wrist-worn applications. Its potential for detecting atrial fibrillation (AF) has been recently presented. Besides AF, another cardiac arrhythmia increasing stroke risk and requiring treatment is atrial flutter (AFL). Currently, the knowledge about AFL detection with PPG is limited. The objective of our study was to develop a model that classifies AF, AFL, and sinus rhythm with or without premature beats from PPG and acceleration data measured at the wrist in daily life. Methods: A dataset of 40 patients was collected by measuring PPG and accelerometer data, as well as electrocardiogram as a reference, during 24-hour monitoring. The dataset was split into 75%-25% for training and testing a Random Forest (RF) model, which combines features from PPG, inter-pulse intervals (IPI), and accelerometer data, to classify AF, AFL, and other rhythms. The performance was compared to an AF detection algorithm combining traditional IPI features for determining the robustness of the accuracy in presence of AFL. Results: The RF model classified AF/AFL/other with sensitivity and specificity of 97.6/84.5/98.1% and 98.2/99.7/92.8%, respectively. The results with the IPI-based AF classifier showed that the majority of false detections were caused by AFL. Conclusion: The PPG signal contains information to classify AFL in the presence of AF, sinus rhythm, or sinus rhythm with premature contractions. Significance: PPG could indicate presence of AFL, not only AF.*

Based on: L.M Eerikäinen, A.G. Bonomi, F. Schipper, L.R.C. Dekker, H.M. de Morree, R. Vullings, R.M. Aarts,"Detecting Atrial Fibrillation and Atrial Flutter in Daily Life Using Photoplethysmography Data", *IEEE Journal of Biomedical and Health Informatics*, vol. 24, No. 6, 2020,

5.1 Introduction

Atrial fibrillation (AF) is a cardiac arrhythmia that has been estimated to affect approximately 3% of the adult population, the prevalence increasing at older age.[104,105] The arrhythmia is associated with increased morbidity, such as stroke and heart failure.[5,106] Therefore, a timely diagnosis and start of the treatment of AF is essential, and new solutions for unobtrusive, low-cost, and possibly prolonged monitoring are increasingly studied. Especially solutions for long-term monitoring that work in daily life are needed for detecting intermittent episodes of AF that may be missed if monitoring period is short.

Another cardiac arrhythmia causing a similar stroke risk as AF, but is less common, is atrial flutter (AFL).[10710] In AFL the atrial rhythm is regular and the ventricular rate is dependent on atrioventricular conduction and on whether the flutter is typical or atypical. The guidelines for anticoagulation and aims for AFL management are similar as for AF.[5] In addition, many patients with AFL develop later AF[108] or both arrhythmias may coexist.[10] Although the aims in management are similar, the treatment strategies for the two arrhythmias differ. AF is more often treated with medication whereas cardiac ablation is more common in treating AFL,[5,107] the success rate of ablations for a specific type of AFL being 90–95%.[5] Rate control in AFL is often more difficult to achieve than in AF[5] and antiarrhythmic therapy of AF may also cause AFL.[107] Knowing the type of arrhythmia causing the stroke risk is therefore important as antiarrhythmic therapies differ.

Photoplethysmography (PPG) is an optical measurement modality that can be used in physiological measurements, such as heart rate monitoring.[28,91] Reflective PPG is often used for wearable solutions, e.g. wristband devices. The potential of using PPG measured at the wrist to detect AF has been investigated in several studies with promising results.[84,85,97,98,109–118] Most of the approaches have focused on discriminating AF from normal sinus rhythm (NSR)[109–111] or, more in general, non-AF rhythms without further dividing the rhythms into different classes.[84,97,98,112–117] For classification of multiple rhythms, Corino et al.[85] proposed a method to classify the rhythms into AF, NSR, and other arrhythmias, whereas the approach of Fallet et al.[118] focused on classifying AF, NSR, and ventricular arrhythmias. Furthermore, the potential of using PPG signals to classify multiple cardiac rhythms has been studied with PPG measured with smartphones.[49,51,119]

While AF detection with PPG has been widely studied, the literature about using PPG signals to classify AFL is limited. In the study of Corino et al.,[85] the class for other arrhythmias included 9 subjects having either AFL, ventricular premature beats (VPB), atrial tachycardia, and variable conduction. The

sensitivity and specificity for this class were 75.8% and 76.8%, respectively. In other studies in which AFL is mentioned to appear in the dataset, the data has been often either excluded[112,113,117,120] or it has been included in the group of non-AF rhythms.[116] Kashiwa et al.[117] excluded continuous AFL, but did not distinguish short AFL episodes among AF episodes. The approaches that included AFL in a different class than AF, were based on the information obtained from the inter-pulse intervals (IPIs).[85,116] From electrocardiography (ECG) studies we know that rhythm irregularity is not specific for AF, but occurs also during AFL.[121] Therefore, analysis on IPI irregularity patterns may not be sufficient for an accurate classification of these two arrhythmias. AFL can also manifest as a very regular rhythm, which can be a challenge when distinguishing from sinus rhythm. Adding additional information, e.g. from the PPG waveform, may be helpful in detecting different rhythms.

The objective of our study was to develop a classifier for classification of AF, AFL, and other rhythms, using PPG and acceleration data measured at the wrist in daily life. The category other rhythms included NSR and sinus rhythm accompanied with premature beats originating either from the atria or the ventricles. First, we developed an AF classifier based on rhythm irregularity, derived from the IPIs with commonly used features for AF detection, to benchmark the classification performance. This was done in three different ways: by considering AFL as a non-AF rhythm together with sinus rhythm and premature beats, considering AFL with AF because of the similar stroke risk, and by excluding AFL completely. Second, we added new features, such as features from the PPG waveform, to improve the classification performance and provide sufficient information for classifying multiple rhythms, i.e. AF, AFL, and other.

5.2 METHODS

5.2.1 Data

The dataset for this study consisted of simultaneous ECG, PPG, and accelerometry measurements in 40 patients undergoing a 24-hour Holter measurement as part of routine clinical care. The patients were contacted by a cardiologist and given at least one week to consider their participation in the study. Before the start of the measurements, the participants gave written informed consent. The study (NL53827.100.15) was approved by the medical ethical committee MEC-U (Medical Research Ethics Committees United) in the Netherlands, and the data was collected in the Catharina Hospital, Eindhoven, the Netherlands.

The ECG was measured with a 12-lead Holter monitor (H12+, Mortara,

Table 5.1: Rhythm type distributions in the training and test sets

Rhythm type	Training set (N = 29)	Test set (N = 10)
NSR	48.5% (15)	45.7% (5)
SVPB	0.5% (15)	0.4% (5)
VPB	0.6% (12)	0.3% (5)
AFL	10.8% (3)	22.4% (2)
AF	39.6% (11)	31.0% (3)

NSR = normal sinus rhythm, SVPB = supraventricular premature beat, VPB = ventricular premature beat, AFL = atrial flutter, AF = atrial fibrillation

Milwaukee, WI, USA). The PPG and 3-axis accelerometer measurements at the non-dominant wrist were made with a data logging device equipped with the Philips Cardio and Motion Monitoring Module (CM3 Generation-3, Wearable Sensing Technologies, Philips, Eindhoven, the Netherlands). The PPG sensor was based on reflective mode using two green LEDs. The sampling frequency of both PPG and accelerometry was 128 Hz and the dynamic range of the accelerometer was $\pm$ 8 g. The accuracy of heart rate measurement when using the same PPG-sensor has been previously reported in.[122]

The ECG data were visually analyzed by a clinical expert using an automated rhythm detection software (Veritas, Mortara, Milwaukee, WI, USA). The software extracted beat times from the ECG and identified every beat either as normal, supraventricular premature beat (SVPB), VPB, AF, paced, artifact, or unknown. The rhythm was then confirmed or corrected by the expert. The software labeled also atrial flutters as AF, which were corrected after the visual inspection by the expert.

Out of the 40 patients, 14 had continuous AF during the recording period, 20 had normal sinus rhythm with premature contractions, 4 had continuous atrial flutter, 1 had continuous atrial flutter with atrial tachycardia, and one patient had a very noisy ECG reference. The patient with the very noisy ECG reference was excluded from the analysis, because no classification could be made based on the reference data.

For developing the classification models, the dataset was divided into two parts: a training set and a test set. The training set was used for training of the model and the test set was kept as unseen data to test the classification performance. The split was made by assigning 75% of the patients to the training set and 25% to the test set based on the rhythm, with the aim to have similar rhythm distributions in both datasets. These are presented in Table 5.1 based on the percentage of beats in every class and the number of patients

Table 5.2: Patient characteristics

Training set	AF	AFL	Other
Age, years	66 ± 11	63 ± 7	69 ± 12
BMI, $\mathrm{kg/m^2}$	31.2 ± 7.0	30.3 ± 5.6	28.6 ± 6.1
Min HR, bpm	43 ± 11	42 ± 27	49 ± 8
Mean HR, bpm	79 ± 16	78 ± 28	71 ± 12
Max HR, bpm	136 ± 19	131 ± 26	108 ± 17
Record length per patient, h	23.1 ± 1.1	23.4 ± 0.2	23.5 ± 1.0
Total record length, h	208	70	353
Test set			
Age, years	76 ± 6	70 ± 4	62 ± 16
BMI, $\mathrm{kg/m^2}$	24.3 ± 3.3	28.3 ± 3.6	26.9 ± 3.3
Min HR, bpm	38 ± 17	47 ± 4	49 ± 9
Mean HR, bpm	80 ± 20	84 ± 3	72 ± 14
Max HR, bpm	144 ± 43	159 ± 15	119 ± 14
Record length per patient, h	22.6 ± 0.5	23.2 ± 0.1	23.8 ± 0.4
Total record length, h	68	46	119

The values are presented as mean $\pm$ standard deviation over subjects.
AF = atrial fibrillation, AFL = atrial flutter, HR = heart rate

having that rhythm. The atrial tachycardia was included in the class of AFL. The patient characteristics in both datasets are presented in Table 5.2.

5.2.2 Preprocessing and pulse detection

The raw PPG data was uploaded from the data logging device and processed offline. As a preprocessing step prior to pulse detection, the PPG signal was downsampled from 128 Hz to 64 Hz and bandpass filtered to range from 0.3 to 5 Hz. The pulses were detected by identifying the fiducial points, i.e. the troughs, in the waveform by detecting the local minima. The time difference between two consecutive fiducial points was calculated to obtain the IPIs. The IPI time series were used to match the pulses to the labeled ECG beat times. The method for detecting the pulses and synchronizing the PPG and ECG beat information is described more in detail in.[113] Examples of 30-second segments of the PPG waveforms and the corresponding IPIs for different rhythm types are presented in Fig. 5.1 and Fig. 5.2. Figure 5.2 shows differences in the rhythm characteristics between different AFL subjects.

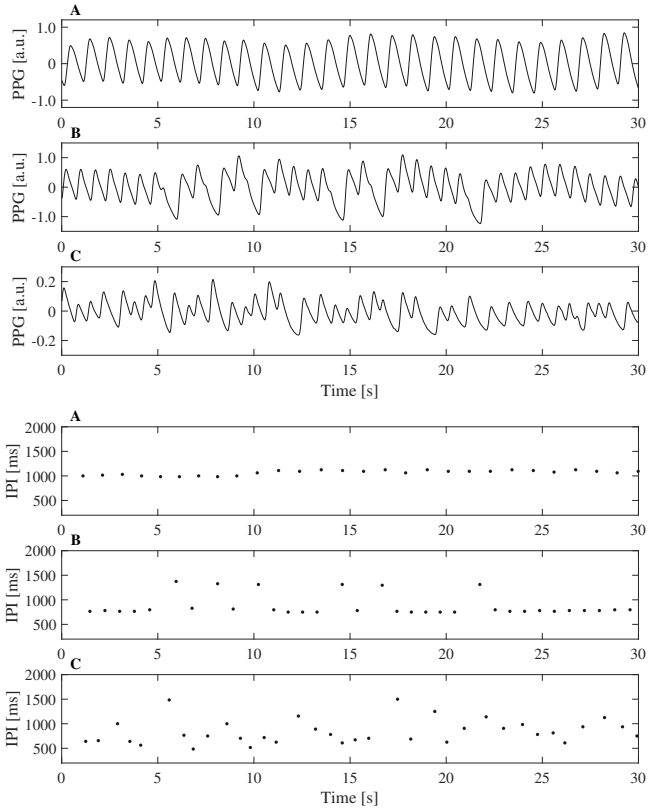

Figure 5.1: PPG waveforms and the corresponding IPIs of sinus rhythm (A), sinus rhythm with SVPBs (B), and atrial fibrillation (C).

5.2.3 Modeling architecture

In this study, two models were developed: an AF classification model based on the traditionally used IPI variability patterns (benchmark model), and a multi-rhythm model to classify AF, AFL, and other rhythms. Figure 5.3 shows the block diagram of the two models. The benchmark model takes as input only features computed from IPI series, whereas the multi-rhythm model uses the IPI series, PPG waveform, and accelerometer data as input for the feature computation.

The PPG, accelerometer, and IPI time series data were segmented in 30-second non-overlapping windows for computing the features and every window was labeled based on the rhythm. The beat labels from ECG were used as the ground truth. When the majority of beats were AF, the window was labeled as AF. This was the same for AFL. When there were any SVPBs in the window, and no AF or AFL, the window was labeled as SVPB, and the same for VPB.

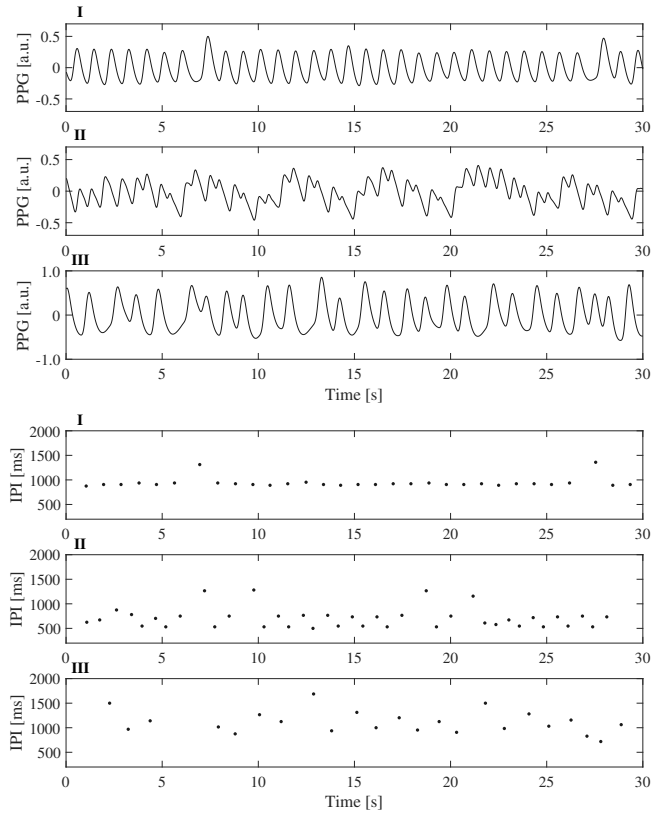

Figure 5.2: PPG waveforms and the corresponding IPIs from three different atrial flutter subjects (I-III).

The windows were labeled as sinus rhythm when they did not contain any of the previously mentioned rhythms. If more than half of the beats were labeled as artifact, the window was discarded from the analysis.

For binary classification, which was done with the benchmark model ((1) in Fig. 5.3), the classification was performed in three different ways. First, all the windows labeled as AF were considered as one class, and AFL, NSR, SVPB, and VPB together as another class, i.e. non-AF rhythms. Second, AFL was considered as the same class with AF, and finally, AFL was completely removed from the analysis. For the multi-rhythm classification ((2) in Fig. 5.3), AFL was separated as its own class, and NSR, SVPB, and VPB, treated as one class to which we refer as 'other'. Premature beats were not separated as their own class because there where too few examples compared to AF, AFL, and NSR. In addition, they usually are not considered to require medical treatment.

5.2.4 Features

The features to be used for the classification were calculated for every 30-second window in two stages: first for the benchmark AF classifier using only IPI-features, and second for the multi-rhythm classifier with the aim of improving overall accuracy and classify both AF and AFL separately. All the features used in the analysis are summarized in Table 5.3.

Prior to the feature computation, IPIs that were shorter than 200 ms and longer than 2200 ms were excluded from the analysis. This was done because pulses could be sometimes falsely detected or missed, leading to incorrect IPIs. For more robust feature calculation, the features were calculated only if the window contained 20 or more IPIs.

For the benchmark AF classifier, six features characterizing the variability or entropy of the IPI series, which have been used for AF classification in the literature,[39,40,49,85] were computed:

- Shannon Entropy (ShEn)
- Normalized Root Mean Square of Successive Differences (nRMSSD)
- pNN40 and pNN70
- Sample Entropy (SampEn)
- Coefficient of Sample Entropy (CoSEn)

We have previously studied the discriminative power of these features individually for AF detection with PPG and compared that to ECG.[113] Here, these features were used to build the first model to classify AF either with or without AFL.

For the multi-rhythm classifier, additional features were calculated in order to improve the detection accuracy and enable separate AFL classification. The features were derived either directly from the PPG waveform, from the IPI series, or from the accelerometer data. The features of the PPG waveform were computed from the signal segments after subtracting the mean value and dividing by its standard deviation. The new feature set consisted in total of 16 features, which will be described here.

One category consists of features that have been considered to measure the signal quality of PPG. A feature that has been used for motion artifact detection is kurtosis.[123] It is a statistical measure describing the tails of the distribution defined as

$$K = \frac{E(x-\mu)^4}{\sigma^4}, \tag{5.1}$$

where μ and σ are the mean and standard deviation of x, respectively, and $E(n)$ is the expected value of the quantity n.

In the PPG analysis domain, Hjorth descriptors called mobility and complexity,[124] $\mathcal{H}_1$ and $\mathcal{H}_2$, have previously been used for analyzing the quality of the signal.[125] The descriptors have initially been developed for electroencephalogram (EEG) analysis[124] and represent the mean frequency and half the bandwidth, respectively. The descriptors are calculated from spectral moments of the signal, the nth order spectral moment being defined as

$$\bar{\omega}_n = \int_{-\pi}^{\pi} \omega^n S(e^{j\omega}) d\omega, \tag{5.2}$$

where $S(e^{j\omega})$ is the power spectrum. From the moments with different orders, $\mathcal{H}_1(n)$ is defined as

$$\mathcal{H}_1(n) = \sqrt{\frac{\bar{\omega}_2(n)}{\bar{\omega}_0(n)}}, \tag{5.3}$$

and $\mathcal{H}_2(n)$ as

$$\mathcal{H}_2(n) = \sqrt{\frac{\bar{\omega}_4(n)}{\bar{\omega}_2(n)} - \frac{\bar{\omega}_2(n)}{\bar{\omega}_0(n)}}. \tag{5.4}$$

The spectral moments were implemented in this work in the time domain according to Sörnmo and Laguna.[78]

A similar feature to $\mathcal{H}_2(n)$, which also has its origin in EEG analysis, is Spectral Purity Index (SPI):[77,78]

$$\mathrm{SPI}(n) = \frac{\bar{\omega}_2^2(n)}{\bar{\omega}_4(n)\bar{\omega}_0(n)}. \tag{5.5}$$

Later SPI has been used also for ECG analysis in detecting false alarms of ventricular tachycardia and fibrillation/flutter.[126,127] Recently, the method was also applied to PPG signals in order to distinguish ventricular arrhythmias from sinus rhythm and AF.[118] In addition, it has been considered as a signal quality metric for PPG.[97]

Shannon Entropy in the time domain was included as a feature in the first stage to study the variability of the IPI series, but it can be also extended to the frequency domain. Spectral Entropy (SE)[128] measures the spectral complexity of the time series and can be calculated as

$$\mathrm{SE} = \sum_f p_f \log\left(\frac{1}{p_f}\right), \tag{5.6}$$

where p_f is the power spectral density normalized with the total spectral power. Previously, Fallet et al.[118] have studied SE for discrimination of AF and ventricular arrhythmias from PPG.

In addition to the features from the frequency domain, additional IPI features were introduced. Modeling RR intervals as a Markov process has been used for AF detection initially by Moody and Mark.[31] In the model, each RR interval was considered to be either short, regular, or long. When every RR interval was assigned to one of these states, transitions between states and transition probability matrices for different rhythms could be calculated. From the transition probabilities, a score that represents the likelihood of the rhythm can be derived. With a similar approach having more states, a good performance in detecting AF from the IPI series has been recently shown.[112] In the current work, the Markov model was used in order to distinguish between AF and AFL.

The transition probability matrices to model the Markov process were calculated for AF and AFL from the IPI series. These gave the score S that indicates whether the rhythm is more likely to be AF than AFL.

$$S_{ij} = \log\left(\frac{p_{ij}^{AFL}}{p_{ij}^{AF}}\right), \tag{5.7}$$

where p_{ij}^{AFL} is the probability that after an interval belonging to state i, an interval belonging to state j occurs during AFL, and p_{ij}^{AF} is the same for AF.

In addition, as a new feature, the same procedure was used to calculate a score when using pulse-to-pulse-to-pulse intervals (PPPI) instead of IPIs. PPPIs were calculated as the time difference between two fiducial points by skipping one pulse in between, the length of the interval corresponding to the sum of two consecutive IPIs. The patterns formed by PPPIs are different from IPIs and this can be especially helpful for distinguishing when irregularity is due to an alternating IPI-pattern, such as during AFL, instead of due to AF. Both IPI and PPPI series were normalized and the scores filtered according to,[31] having as the coefficient $k = 0.25$. The number of states for the used model was 12. Because a score is produced for every interval, a mean of the scores in the window was taken.

To include information reflecting the heart rate, the maximum, minimum, median, and standard deviation of the IPIs were included as features. In addition, from the PPG waveform, maximum, minimum, mean, and standard

deviation of the pulse amplitude, i.e. the difference between the peak and the onset of the pulse, were calculated.

Finally, from the accelerometer data the norm of the accelerations on the three axis was calculated. From the norm, the standard deviation and the maximum absolute value in each window were included to the feature set.

5.2.5 Feature selection

Feature selection was employed to select the optimal set of IPI variability features for the benchmark AF classification models. Based on our previous work,[113] all the six features reflecting variability or entropy of the IPI sequence have individually a strong discriminative power in AF classification. Moreover, all these features try to capture relatively similar information and could be redundant. Therefore, the Minimal-Redundancy-Maximal-Relevance (mRMR) criterion[129] was selected as the method to rank the features.

The mRMR method tries to find features that maximize the mean value of mutual information between all individual features and the target class, i.e. the maximal relevance. However, it is likely that these features have a large dependency on each other. Therefore, the minimal redundancy criterion is added, and it is based on the mutual information between the individual features. The balance between the two criteria is optimized by finding the set that maximizes the difference between the maximal relevance and minimal redundancy.

The calculations were made with the implementation provided by Peng et al.[129] made available on.[130] The method requires discretization of the features and that was made by having three states when using thresholds at mean $\pm$ standard deviation.[130]

5.2.6 Classifiers

The benchmark AF classifier (Fig. 5.3 (1)) for the binary classification using the selected IPI-features as input was a generalized logistic regression model. The probability for the 30-second window to contain AF was given by the function:

$$y_{AF}(t) = \frac{e^{X(t) \cdot b}}{1 + e^{X(t) \cdot b}}, \tag{5.8}$$

where t is the index of the window, $X(t)$ a vector containing the feature values for the window at time t, and b a vector of the model coefficients.

The threshold for the probability was selected as the one that maximizes

Youden index[101] J defined as

$$J = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} + \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} - 1, \tag{5.9}$$

where TP are true positives, FN false negatives, TN true negatives, and FP false positives.

The multi-rhythm classifier (Fig. 5.3 (2)) to perform the classification into AF, AFL, and other, was a Random Forest (RF) model[79] taking as input the selected IPI-features and the additional 16 PPG, IPI, and accelerometry features. RFs are ensembles of decision trees that are grown in parallel by selecting a random subset of features to grow each tree. The final classification is based on a majority vote given by the classifications of each individual tree. RFs have a few beneficial characteristics, such as that they are relatively robust to noise and outliers, and can give useful internal estimates about the error and importance of the variables. The latter is an advantage with the small dataset because performing a feature selection for the multi-class problem is more challenging than for the binary case.

5.2.7 Cross-validation

For selecting the number of features for the benchmark AF classifier, leave-one-subject-out cross-validation with the training set was used. The performance was calculated by using the data of one subject for testing and the data of the remaining subjects for training. The number of features went from one to six by adding the features in the order given by the mRMR. The number of features reaching the highest accuracy was selected.

The multi-rhythm classifier was also evaluated with the training set by using 10-fold cross-validation. In order to maintain an equal distribution of the three classes in every fold, the dataset was divided in 10 sets of equal size that were stratified by the classes but not by patients because of the number of AFL patients. The training and testing was performed 10 times with each of the 10 sets serving as a test set once.

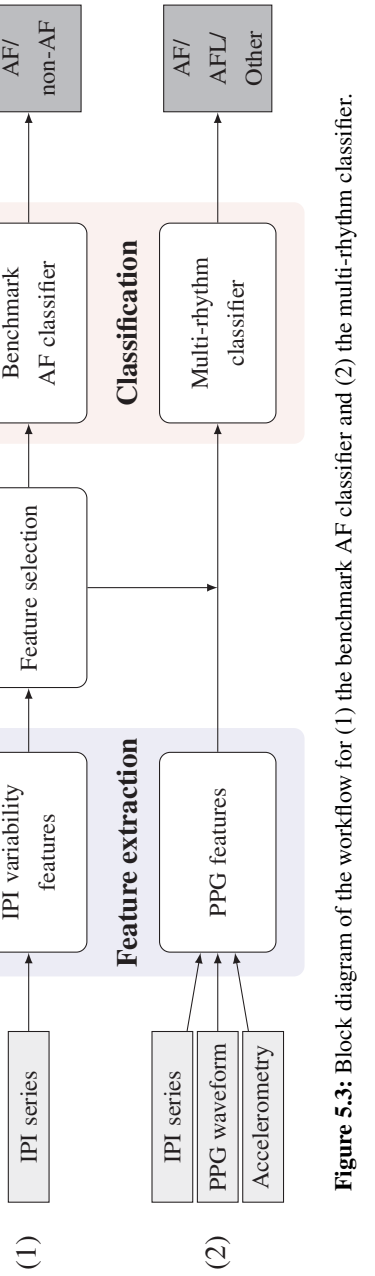

Figure 5.3: Block diagram of the workflow for (1) the benchmark AF classifier and (2) the multi-rhythm classifier.

Table 5.3: Features for rhythm classification

IPI-features	PPG waveform features	Acceleration features
Shannon Entropy (ShEn)	Kurtosis (K)	Standard deviation*
Normalized Root Mean Square of Successive Differences (nRMSSD)	Hjorth mobility ($\mathcal{H}_1$)	Maximum absolute value*
pNN40	Hjorth complexity ($\mathcal{H}_2$)	
pNN70	Spectral Purity Index (SPI)	
Sample Entropy (SampEn)	Spectral Entropy (SE)	
Coefficient of Sample Entropy (CoSEn)	Maximum pulse amplitude	
Mean Markov score AF/AFL - IPIs	Minimum pulse amplitude	
Mean Markov score AF/AFL - PPPIs	Mean pulse amplitude	
Maximum IPI	Standard deviation of pulse amplitudes	
Minimum IPI		
Median IPI		
Standard deviation of IPIs		

*From the norm of the accelerations on the three axis.
IPI = inter-pulse intervals, PPPI = pulse-to-pulse-to-pulse interval

5.3 RESULTS

5.3.1 Training set

The mRMR selection method on IPI features was used for each of the three class divisions. For AF vs. non-AF classification gave the following ranking independent whether AFL was included in non-AF rhythms or completely excluded from the analysis: pNN70, SampEn, ShEn, CoSEn, pNN40, nRMSSD, starting from the most relevant one. When AF and AFL vs. other classification was considered, only pNN70 and pNN40 switched order with each other in aforementioned ranking. Based on the leave-one-subject-out cross-validation, the best accuracy was obtained by combining the first three features. The Receiver Operating Characteristics (ROC) curve of the models combining these three features are presented in Fig. 5.4 when AFL is either included in or excluded from non-AF rhythms, or completely excluded from the analysis. The results are calculated with 43.6% of the data after the windows not containing a sufficient number of IPIs or the reference was labeled as artifact were excluded. The median (lower - upper quartiles) coverage per patient was 47.0% (29.6 - 56.2)%.

Figure 5.4 shows the operating points for the three binary AF classifiers selected by maximizing J. The model AF vs. AFL and other had sensitivity of 93.6% and specificity of 88.2%. Misclassification of AF occurred primarily in presence of SVPB, and AFL, i.e. other arrhythmia generated from the atria, as shown in Fig. 5.5 on the left. The presence of VPB caused relatively little false detections of AF. The model that classified AF and AFL vs. other had better sensitivity and specificity (95.2% and 90.4%). With this model the false negative classifications were mainly due to AFL, as shown on the right of Fig. 5.5. The best performance for binary classification with sensitivity of 98.2% and specificity of 90.9%, was obtained when AFL was not present in the data.

The multi-rhythm classifier was a RF model consisting of 100 trees and classified the 30-second windows into AF, AFL, or other rhythm. The results calculated with cross-validation are listed in Tables 5.4 and 5.5. Table 5.4 is the confusion matrix of the classifications and in Table 5.5 the results are presented as one class vs. all the rest in terms of sensitivity, specificity, positive predictive value (PPV), and accuracy.

5.3.2 Test set

The results of the test set were first calculated with the benchmark AF models. The test set had a coverage of 41.6%, with the remaining windows excluded

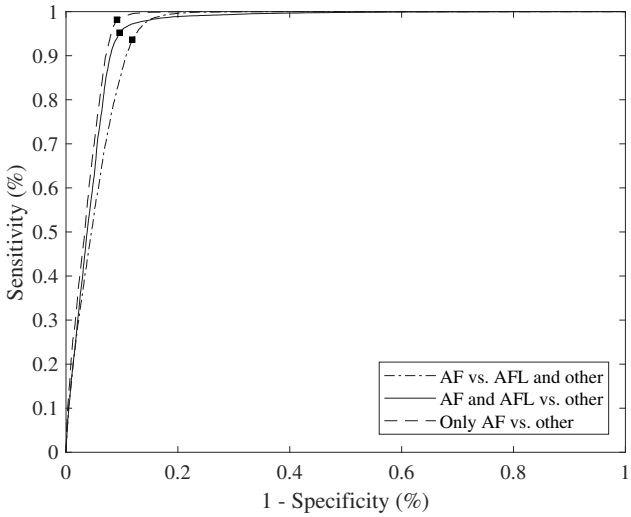

Figure 5.4: Receiver Operating Characteristics of the binary classification models with three IPI-features. The square depicts the operating point defined by the Youden index.

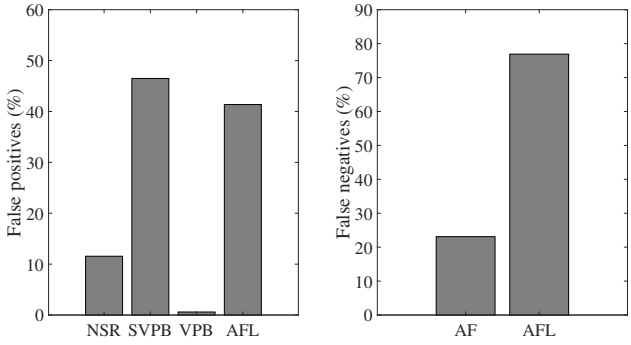

Figure 5.5: The distributions of the rhythm classes of the false positive AF detections (left) and false negative AF and AFL detections (right) with IPI-models in the training set.

due to insufficient IPIs or the ECG reference being labeled as artifact. The median (lower - upper quartiles) proportion of windows included for each subject was 42.2% (31.1 - 50.1)%.

For AF vs. AFL and other classification, the sensitivity and specificity were 96.1% and 96.2%, respectively. When AF and AFL were included in the same class against the rest, the sensitivity decreased to 58.9% and specificity to 92.5%. The best classification performance was again obtained when AFL was completely excluded from the analysis, the sensitivity being 99.1% and specificity 95.4%.

Table 5.4: Confusion matrix of the classifications with the multi-rhythm classifier of the training set

		Predicted label			
		AF	AFL	Other	N. windows
True label	AF	**97.2%**	0.7%	2.2%	11468
	AFL	9.9%	**86.3%**	3.8%	2261
	Other	3.0%	0.0%	**97.0%**	21281

Table 5.5: Classification performance of the multi-rhythm classifier with the training set

Rhythm type	Sensitivity	Specificity	PPV	Accuracy
AF	97.2%	96.4%	92.8%	96.7%
AFL	86.3%	99.7%	95.5%	98.9%
Other	97.0%	97.6%	98.5%	97.2%

The results of the multi-rhythm model are presented in Tables 5.6 and 5.7. The sensitivity and specificity for detecting AF were 97.2% and 98.2%, respectively. AF and other rhythms were rarely detected as AFL, and the specificity for AFL was 99.7%. The most false detections are with AFL detected as other rhythm.

5.4 Discussion

This is the first study showing that both AF and AFL detection is possible from PPG data in daily life. We demonstrated that many of the false positive classifications of a benchmark AF classification model were due to instances of AFL. When considering AFL to be classified with AF as the same class, it was often missed by the benchmark model. The presented multi-rhythm classification algorithm showed much improved performance, particularly making less false positive AF detections, when trained to classify both AF as well as AFL. As the results with the benchmark AF models show, the rhythm characteristics when using the IPI-features did not belong to either of the binary classes, i.e. AF and other, and decreased the classification performance. Therefore, considering AFL as a separate class can be beneficial also in terms of improving AF detection and not only for giving classification for the rhythm type itself.

Table 5.6: Confusion matrix of the classifications with the multi-rhythm classifier of the test set

		Predicted label			
		AF	AFL	Other	N. windows
True label	AF	**97.6%**	0.7%	1.6%	3264
	AFL	1.7%	**84.5%**	13.8%	2793
	Other	1.8%	0.1%	**98.1%**	6747

Table 5.7: Classification performance of the multi-rhythm classifier with the test set

Rhythm type	Sensitivity	Specificity	PPV	Accuracy
AF	97.6%	98.2%	95.0%	98.1%
AFL	84.5%	99.7%	98.7%	96.4%
Other	98.1%	92.8%	93.8%	95.6%

Combining information from the PPG signal, IPIs, and accelerometer improved the classification accuracy and enabled discrimination of AFL from AF and the other rhythm types. Adding features from the PPG waveform helped in detecting AFL compared to using only IPI-information. In previous work, a comparison of PPG pulses gave different results in a patient suffering from a regular form of typical AFL than in patients with AF or other rhythms.[131] Moreover, some of the features derived from the PPG waveform have been used to discriminate ventricular arrhythmias.[118] The PPG waveform characteristics have been also studied in the context of force-interval relationship during AF[132] and mechanical alternans.[133] This could indicate that the PPG waveform itself contains information about different rhythms and cardiac function.

Discriminating AFL from AF has been also possible based on RR interval series when a multilevel model of the atrioventricular node was used.[134] Here, we included the Markov model approach to process the IPI and PPPI series in order to classify these two rhythms. Thus, the RR intervals or IPIs also contain valuable information when processed in an adequate manner.

The validity of the windows to be analyzed was judged based on the number of pulses detected in that window, but no further signal quality analysis was made. The method detecting the pulses already considers the body acceleration and therefore the number of pulses indirectly already reflects this. In addition, some of the features included in the RF model have been developed or used previously for PPG signal quality assessment, and information of the body acceleration was also given as an input. This may have improved the

classification of the more noisy segments.

The study suffers from some limitations. In the dataset, all the patients that suffered from AF, had it continuously. For datasets with intrapatient rhythm variability, data measured before and after electrical cardioversion have been used in different studies to investigate AF detection with PPG. This reflects a hospital setting but not ambulatory monitoring, which is where the wrist-worn applications can really add value. Studies that show AF detection with PPG in ambulatory setting are very few. Shen et al.[135] had measurements from 3 to 8 hours and mention subjects with rhythms that change over time. However, these include AF and eight other rhythms and the proportion of paroxysmal AF remains unclear. Sološenko et al.[116] presented in their study a dataset measured for approximately 22 hours per subject in cardiac rehabilitation. Similarly to our study, all the AF subjects had continuous AF.

The number of AFL subjects in the dataset was less compared to the other groups. AFL can have different rhythm characteristics depending on the type and therefore can vary between subjects. The results between the training set and test set in AFL classification differ slightly due to the different types of AFL cases, e.g. very regular AFL being misclassified as other rhythm in the test set. Yet, the results remain relatively similar when compared to the benchmark AF models.

Premature beats were not separately classified in this work because of the small number compared to other beat types nor were they suppressed in order to reduce false positive detections. Furthermore, the classification was done in windows which makes identifying individual beats difficult. As Fig. 5.5 shows, premature beats had an impact on false positives when only IPI-based features were used in the classification. Adding PPG-waveform and accelerometry features helped in improving specificity by also reducing false positives caused by premature beats and not only by AFL. Understanding the effect on the classification accuracy caused by higher burden of premature beats and variability in their beat patterns between patients remains for future research.

The dataset was divided into training and test set in order to leave some of the recordings untouched while developing the models. The split was done by patients and no data from a patient assigned to the training set ended up in the test set. Therefore, it was not possible to match the rhythm class distributions completely between the two sets. The characteristics of the sets may, therefore, not be entirely comparable, which is reflected especially in the results of the benchmark AF model.

The choice to use RF was made because of the class distribution and the ability of RF to give information about the feature importance. The model has many advantages, but one drawback is poor interpretability. For clinical

applications, more transparent options of models may be more suitable when larger datasets are available.

The selected approach was based on calculating the features in 30-second non-overlapping windows. The features based on the IPIs require a sufficient amount of data for the calculations, and missed pulses caused the data to be discarded, thereby reducing coverage to approximately 45%. The coverage of our approach could be improved by reducing the number of IPIs required per window or by using overlapping windows. However, the effects on the classification performance should be studied. Sološenko et al.[116] used an approach to classify every beat separately, which resulted in a higher coverage (89.2%) during 24-hour measurements. However, the sensitivity for AF with this coverage was only 72.0%. When 50% of the data was judged as analyzable, the sensitivity increased to 97.2% and specificity was 99.6%.

5.5 Conclusion

In this study, we demonstrated that PPG and acceleration measurements at the wrist can be used to discriminate between AF, AFL, and other rhythms in daily life. We showed that with an AF vs. non-AF model and AF and AFL vs. other rhythms model that used only information derived from inter-pulse intervals, the false detections were for a large part caused by AFL. The multi-rhythm model included more information from the wrist measurement, such as features from the PPG waveform and accelerometer data. This model was not only able to improve the overall performance of AF detection, but could also classify AFL with high accuracy. The results of this study indicate that the PPG signal contains sufficient information, derived both from the waveform and IPIs, to accurately classify between AF, AFL, and other rhythms. Thus, PPG could provide promising means to detect AFL along with AF.

6

BEYOND ARRHYTHMIA DETECTION

Abstract - *Force-interval relationships (FIRs) of the heart represent the relationships between inter-beat intervals (IBIs) and strength of the ventricular contractions. These relationships are typically measured invasively and are altered from normal in heart failure (HF). An unobtrusive and continuous measurement of FIRs could be beneficial when HF and atrial fibrillation (AF) coexist in order to understand if AF causes progression of HF. We hypothesize that FIRs could be assessed during AF with IBIs and hemodynamic changes captured unobtrusively by photoplethysmography (PPG) at the wrist. FIRs were assessed by using Spearman's rank correlation between the pulse onset change in the PPG waveform and either the preceding or pre-preceding IBIs (r_{pre} and $r_{pre-pre}$) in 5-minute segments. 32 patients (14 continuous AF, 18 no AF) were measured during the night with PPG and electrocardio-graphy as a reference. The mean and standard deviation of r_{pre} were -0.25 $\pm$ 0.08 and 0.05 $\pm$ 0.12 ($p < 0.0001$), and of $r_{pre-pre}$ 0.60 $\pm$ 0.09 and 0.16 $\pm$ 0.14 ($p < 0.0001$), during AF and sinus rhythm, respectively. Areas under the Receiver Operating Characteristics curve were 0.987 and 0.998, respectively. Thus, during AF the IBIs correlate with the beat-to-beat changes of blood volume measured with PPG, likely to indicate that FIRs can be measured unobtrusively with the PPG signal.*

Based on: L.M. Eerikäinen, A.G. Bonomi, L. Dekker, F. Schipper, R. Vullings, R.M. Aarts, "Force-Interval Relationships of the Heart Measured With Photoplethysmography During Atrial Fibrillation", *Computing in Cardiology*, 2018,

6.1 INTRODUCTION

FORCE-INTERVAL relationships (FIRs) of the heart, i.e. post-extrasystolic potentiation (PESP) and mechanical restitution (MR), represent the relation between the length of a preceding or a pre-preceding inter-beat interval (IBI) and the strength of contraction that follows.[136] In PESP, the beat following an extrasystole is potentiated, i.e. the contraction is stronger, which is a phenomenon known already for 120 years.[137] The MR, in turn, represents the recovery of the contractile strength after an extrasystolic beat.[136]

In heart failure (HF) the contractile capacity of the heart is reduced and a number of studies have shown that FIRs of a failing heart are different to those of non-failing hearts.[138–140] The cause of the difference is not entirely understood, but is presumably due to altered Ca^{2+} handling of the myocardial cells.[141] Studies of FIRs have been commonly conducted with invasive measures in laboratory settings, although recently Sinnecker et al.[142] stated that FIR assessed by an unobtrusive blood pressure measurement predicts mortality in survivors of myocardial infarction with atrial fibrillation (AF).

AF is the most commonly experienced sustained cardiac arrhythmia. It is characterized by an abnormal electrical activity of the atria which causes the heart to beat irregularly. Studies have shown that, related to this irregularity, the contractility of the heart varies beat-to-beat as a consequence of FIRs.[143–145] AF often coexists with HF and the clinical outcome of patients with this coexistence is particularly poor compared to those having only one of the two conditions.[146] For some HF patients with reduced ejection fraction, AF causes symptomatic deterioration whereas for others AF does not affect the patient's condition, the difficulty being the identification of the two groups of patients.[147]

Photoplethysmography (PPG) is an unobtrusive measurement modality that has recently gained much interest due to its ease of use and applicability for long-term monitoring, e.g. in wrist-based wearable devices. PPG measures blood volume changes in the vascular bed of the tissue and enables heart rate detection from the pulses in the signal. In case of AF, promising results have been presented for detection of AF episodes with continuously measured wrist-worn PPG by assessing rhythm irregularities.[86] However, there is limited knowledge whether continuous PPG monitoring during AF could have additional value beyond rhythm irregularity assessment, considering that PPG also captures hemodynamic effects.

We hypothesize that the beat-to-beat contractility variations during AF may be reflected as pulse-to-pulse variations in the PPG waveform. The assumption is that during irregular rhythm FIRs can be observed and there is a stronger

correlation between IBIs and variations in pulse morphology. Here we present a study about the relationships between IBIs and PPG signal morphology variations during AF that could possibly serve as an unobtrusive method to assess FIRs.

6.2 METHODS

6.2.1 Data

The dataset consisted of 24-hour measurements of electrocardiography (ECG), PPG and accelerometer from 40 subjects. PPG and accelerometer were measured with a wrist-worn data logging device equipped with the Philips Cardio and Motion Monitoring Module (CM3 Generation-3, Wearable Sensing Technologies, Philips, Eindhoven) and an ECG reference with a 12-lead Holter monitor (H12+, Mortara, Milwaukee, WI, USA). The measurement protocol and devices are described in more detail in.[86]

The ECG data were visually analyzed by a clinical expert using an automated rhythm detection software (Veritas, Mortara, Milwaukee, WI, USA). The ECG beat times were extracted by the software and every beat was classified either as normal, supraventricular or ventricular premature contraction, AF, paced, artifact, or unknown. The rhythm was then confirmed or corrected by the expert. In addition to rhythm information, the collected data included patient diaries of the daily activities, and baseline characteristics, medical characteristics, and information about medication which were retrieved from the medical records.

For the analysis, patients with atrial flutter or very noisy ECG reference were excluded. This resulted in 32 patients of which 14 had continuous AF (age (years, m $\pm$ sd): 68 $\pm$ 11, males: 71%, BMI (kg/m^2, m $\pm$ sd): 29.8 $\pm$ 6.9) and 18 showed normal sinus rhythm with premature contractions (age, (years, m $\pm$ sd): 67 $\pm$ 14, males: 50%, BMI (kg/m^2, m $\pm$ sd): 28.1 $\pm$ 5.6).

6.2.2 Features

The PPG data was downsampled from 128 Hz to 64 Hz and preprocessed by filtering with a bandpass filter to the range from 0.3 to 5 Hz. The pulses were detected by identifying fiducial points in the PPG waveform such as the onset of the pulse. The major fluctuations in the PPG signal were removed by subtracting a cubic spline fitted to the point of the maximum gradient of the rising pulse slope.[148] The IBIs were calculated as the time differences between two consecutive pulse onset times. Based on the IBI sequence from PPG and

the IBI sequence from the ECG reference beat-times, the reference beat labels were aligned with the PPG pulses.

The analysis was limited to the period at night between the times that patient had reported going to bed and getting up in the morning with the purpose of having less movement artifacts in the signal. The time stamps were selected manually around the reported sleep and awake times by looking at the accelerometer data.

Figure 6.1 shows an example of recorded PPG waveform during AF with six PPG pulses and the variables that from here on will be referred as pre-preceding inter-beat interval ($\mathrm{IBI}_{pre-pre}$), preceding inter-beat interval (IBI_{pre}), and difference in the PPG signal between two consecutive pulse onsets (i) and (i-1) (onset_{diff}). IBI_{pre} is the time difference between the pulses (i) and (i-1) and $\mathrm{IBI}_{pre-pre}$ between the pulses (i-1) and (i-2). The onset_{diff} could indirectly reflect the change in end-diastolic volume. This is, in turn, related to the change in stroke volume and end-systolic volume, which are known to be influenced by FIRs.[145]

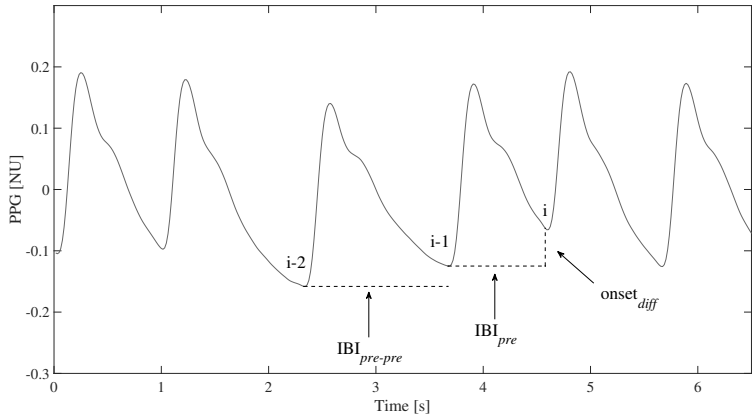

Figure 6.1: Example of PPG pulses, onset_{diff}, IBI_{pre}, and $\mathrm{IBI}_{pre-pre}$ during AF.

The relationship between IBIs and onset_{diff} were analyzed with Spearman's rank correlation. Spearman's rank correlation was selected because of the non-linearity of the relationship. IBIs $<$ 200 ms and $>$ 2000 ms were discarded from the analysis before assessing the correlation. The correlation coefficients between $\mathrm{IBI}_{pre-pre}$ and onset_{diff} ($r_{pre-pre}$), and IBI_{pre} and onset_{diff} (r_{pre}) were computed in 5-minute segments shifting by 60s and including segments containing at least 180 IBIs. In every segment top and bottom 2% of onset_{diff} values and corresponding IBIs were discarded to exclude outliers. Moreover, segments including premature contractions were excluded from the analysis

since these also induce FIRs and the focus of the analysis was in FIRs during AF.

6.2.3 Statistical analysis

The difference between the distributions of correlation coefficients r_{pre} and $r_{pre-pre}$ during AF compared to sinus rhythm was assessed with Mann-Whitney U-test. In addition, Receiver Operating Characteristics (ROC) were analyzed in order to understand to which extent the two correlations are different in the AF and sinus rhythm groups.

6.3 RESULTS

The relationships between the inter-beat intervals ($\mathrm{IBI}_{pre-pre}$ and IBI_{pre}) and onset_{diff} when computed in a 5-minute segment during AF are shown in Fig. 6.2. It illustrates a similar positive trend of onset_{diff} in relation to pre-preceding IBI and a negative trend in relation to preceding IBI which is similar to the relationships of the IBIs and end-systolic volume as in[145] during AF.

Histograms of correlation coefficients r_{pre} and $r_{pre-pre}$ computed for IBI_{pre} and onset_{diff}, and $\mathrm{IBI}_{pre-pre}$ and $\mathrm{onset}_{\mathrm{diff}}$, respectively, are presented in Fig. 6.3. The mean and standard deviation of r_{pre} were -0.25 $\pm$ 0.08 and 0.05 $\pm$ 0.12, for AF and sinus rhythm, respectively, and of $r_{pre-pre}$ 0.60 $\pm$ 0.09 and 0.16 $\pm$ 0.14. In both cases the Mann-Whitney U-test confirmed that the difference in the distributions of r_{pre} and $r_{pre-pre}$ between the AF and sinus rhythm group was statistically significant ($p < 0.0001$). Of all computed r_{pre} 94.84% were significantly different from zero ($p < 0.05$) during AF with respect to 23.66% when AF was not present. For $r_{pre-pre}$ the percentages of nonzero correlation were 99.98% and 58.37% for AF and sinus rhythm, respectively.

Figure 6.4 shows the ROC curves for the correlation coefficients r_{pre} and $r_{pre-pre}$, with the areas under the curve 0.987 and 0.990, respectively.

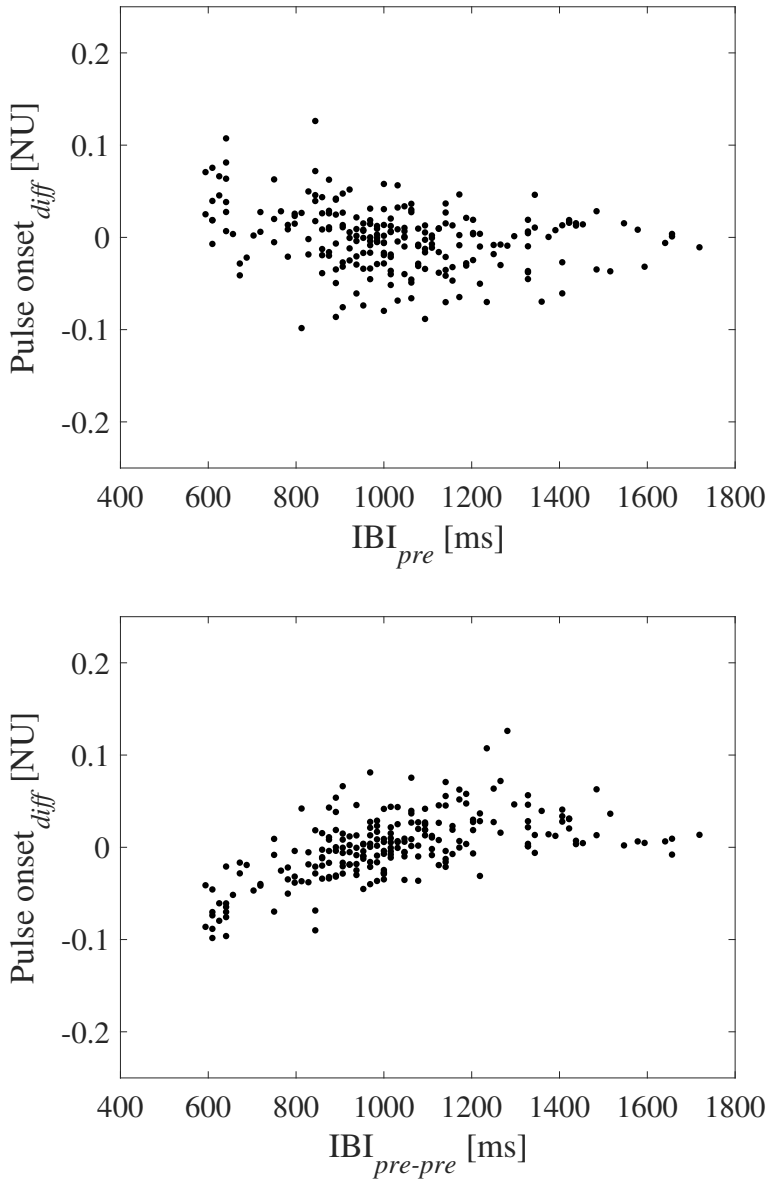

Figure 6.2: The relation between preceding IBI (IBI_{pre}) and pulse onset difference (onset_{diff}) (top) and between pre-preceding IBI ($\mathrm{IBI}_{pre-pre}$) and pulse onset difference (below) in a 5-minute segment when every point represents a beat.

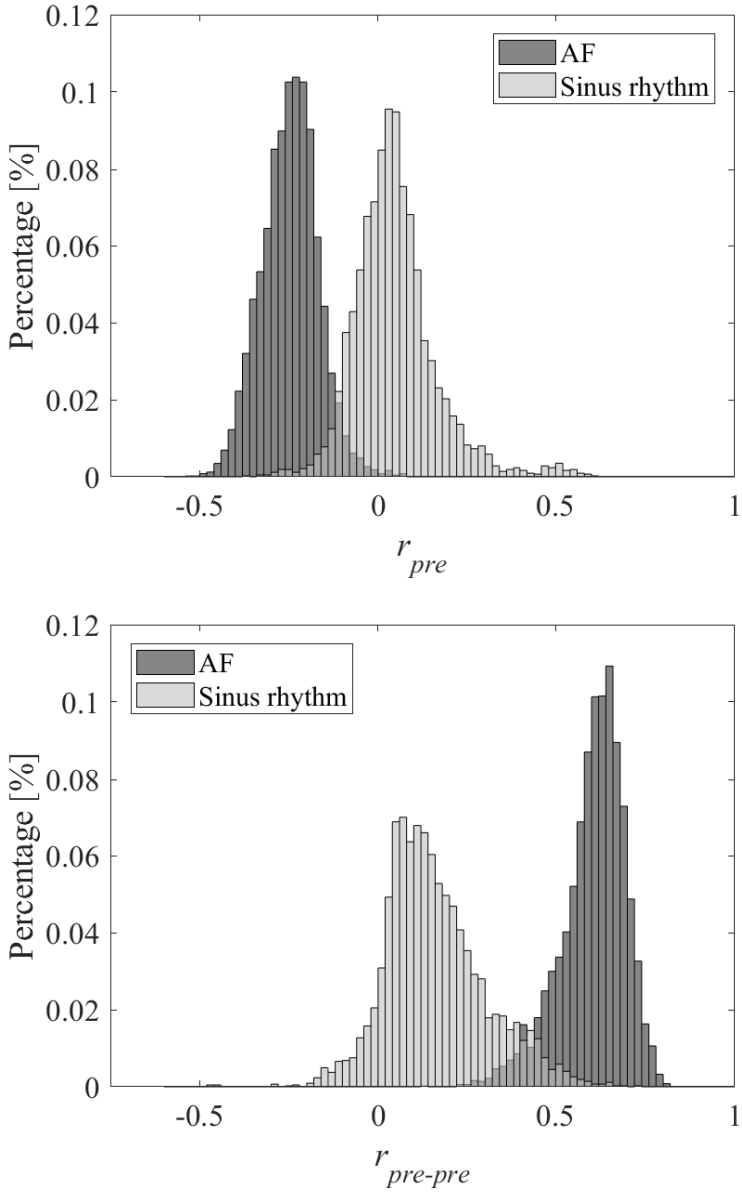

Figure 6.3: Histograms of r_{pre} (left) and $r_{pre-pre}$ (right) during AF and sinus rhythm.

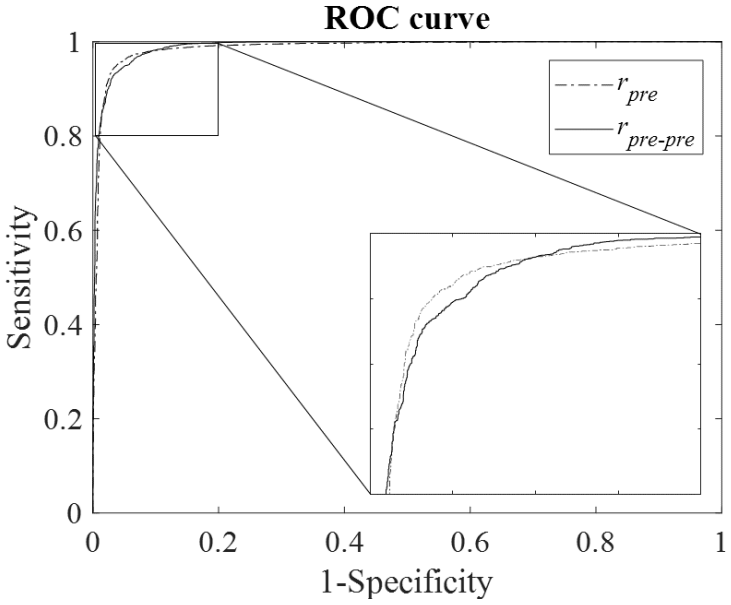

Figure 6.4: ROC curves for r_{pre}, and $r_{pre-pre}$.

6.4 DISCUSSION

We presented two features that could represent FIRs in the PPG signal during AF. The beat-to-beat variability of IBIs during AF shows a relation with blood volume changes which cannot be observed when there is little variability, i.e. during sinus rhythm. The mean correlation coefficients between IBIs and pulse onset changes during AF obtained in this study are similar to the ones presented by Brookes et al.[145] with IBIs and end-systolic volume. The unobtrusive continuous assessment of FIRs, and whether they change with time, can be particularly interesting when AF and HF coexist, since AF can increase symptoms and cause further progression of HF.[146] Moreover, AF can induce HF. Continuous monitoring could possibly help in noticing early on deterioration in the patient's condition, enabling interventions or changes to treatment and preventing further deterioration.

As a limitation, an invasive reference measure of contractility was missing in this study. In addition, new studies are required for investigating whether the unobtrusive measures from PPG presented here show a difference between HF patients and patients with non-failing hearts.

6.5 Conclusion

The relationship between IBIs and pulse-to-pulse morphology variations in the PPG signal during AF could reflect the FIRs of the heart. This will provide an interesting unobtrusive parameter to study further in the group of patients with AF and HF, first by confirming the findings with experiments in a controlled setting using an invasive measurement as a reference, and later on with the aim of improving disease management in this patient group.

Epilogue

7

DISCUSSION AND FUTURE PROSPECTS

This chapter discusses the research conducted in the thesis and proposes directions for further research.

Partly based on: L.M. Eerikäinen, A.G. Bonomi, L.R.C. Dekker, R. Vullings, R.M. Aarts, "Atrial fibrillation monitoring with wrist-worn photoplethysmography-based wearables: State-of-the-art review", *Cardiovascular Digital Health Journal*, vol. 1, No. 1, 2020, Creative Commons Attribution 4.0 License.

7.1 General conclusions and discussion

The most important tool currently for monitoring cardiac arrhythmias is electrocardiography (ECG). However, it has its limitations and alternative monitoring solutions have been investigated to serve as complementary modalities for ECG. In the context of intensive care, other measurement modalities, such as arterial blood pressure (ABP) and photoplethysmography (PPG), are present and can be used for deriving information about heart rate and hemodynamics of the patients. For continuous long-term monitoring, surface ECG is not suitable because of the skin irritation and poor signal quality over time of wet gel electrodes, whereas invasive alternatives are costly.

This thesis studied the additional value of pulsatile waveforms, ABP and PPG, to arrhythmia monitoring in clinical setting and daily life. In *Part I*, first, the problem of high number of false alarms for critical arrhythmias was approached using available waveforms to assess signal quality and selecting the best quality data for classifying alarms. Using a multi-modal approach resulted in a significant reduction of false alarms. Second, it was shown that atrial fibrillation (AF) detection with PPG measured at the wrist when developed with data coming from a clinical setting, did not translate in terms of detection accuracy directly to daily life. This suggests that measurement context and patient population are important when developing AF detection methods with PPG for a specific purpose. *Part II* builds on this conclusion and focuses on development and evaluation of methods for AF detection with wrist-worn PPG in daily life with 24-hour measurements. The influence of movement to the detection accuracy was first assessed with the conclusion that the acceleration data can be used for selecting data of good quality and to improve accuracy, but at the cost of measurement coverage. An improved coverage with comparable accuracy was obtained by extracting more information from the PPG waveform than in the previous approach and including movement information in the model instead of a preprocessing step. In addition, this information enabled distinction of AF and atrial flutter (AFL). Finally, the relationship between waveform characteristics and inter-beat intervals when extracted during irregular and regular rhythms was studied, and this relationship could relate to the force-interval relationship of the heart. From the findings in this thesis, it can be concluded that pulsatile waveforms add valuable complementary information to ECG in cardiac arrhythmia monitoring and that characteristics of PPG waveform should be studied beyond extracting information that is surrogate for known ECG features, such as heart rate and inter-beat intervals. Accurate cardiac arrhythmia detection is possible using PPG signals and potentially useful for screening AF and preventing strokes.

The following sections will discuss more in detail the reduction of false cardiac arrhythmia alarms in clinical setting (Section 7.1.1.1) and in daily life (Section 7.1.1.2), and Section 7.1.2 will focus more on atrial fibrillation detection with wrist-worn PPG in general. Key findings, contributions, and limitations related to the work conducted in this thesis are highlighted as bullet points. Section 7.1.2 also summarizes the current knowledge on AF detection based on the development during the years 2016-2019, discussing the potential and the knowledge gaps related to the technology when aimed for long-term monitoring. More details about the studies about AF detection with wrist-worn PPG-based wearables are listed in the Appendix.

7.1.1 Reduction of false positive cardiac arrhythmia detections

Accurate arrhythmia detection when using automated detection methods is extremely important. For different use cases, the implications of false detections vary, but nonetheless these should be avoided. In intensive care, excess of false alarms causes desensitization to the alarms, which can lead to missing real events requiring intervention. In addition, audible alarms increase noise level that causes stress both to patients and caregivers. In daily life monitoring, false alarms will cause worry to the patients and burden the healthcare system by initiating investigations that are not necessary. The cause of false positive cardiac arrhythmia detections is often poor signal quality, which can be due to movement or poor sensor contact.

7.1.1.1 Clinical setting

In intensive care, the types of warnings given by the monitoring system vary depending on the type of the detected event. Audible alarms are given to the life-threatening cardiac arrhythmias. The triggered alarm might be based on limited or poor quality information available and deemed false after visual inspection of more information. Available information for classification was increased in Chapter 2 by comparing two ECG leads and pulsatile signals, ABP and PPG, to determine the two best quality signals from the available ones instead of relying only on ECG. Information about heart rate was derived from these signals and used for classifying arrhythmia alarms either true or false. The results presented in Chapter 2 show a significant decrease in the number of false alarms and are comparable to the ones obtained with other approaches developed with the same public dataset in the context of the PhysioNet/CinC Challenge 2015 the dataset was initially published for.[149]

- Adding more information to the automated cardiac alarm classification by comparing signals from different sources for determining signal quality, lead to reducing the number of false alarms in intensive care.
- Alarms for ventricular tachycardia were the most difficult to classify. From pulsatile signals, assessing whether a heart beat originates from the atria or the ventricles is much more challenging than from ECG.
- More features from the pulsatile signals in addition to heart rate could be valuable for classification. The dataset used in the current analysis posed some challenges, e.g. the number, composition, and quality of the available signals varied between recordings, leading to missing pulsatile signals for some of the subjects.

7.1.1.2 Daily life

For daily life use, PPG is a suitable non-invasive and low-cost technology to extract heart rate and rhythm information. First, smartphone camera-based PPG measurements have been proposed for discriminating AF,[50,51,93] which was followed by several studies investigating AF detection with wrist-worn devices, eventually the aim being non-invasive long-term monitoring of AF in daily life.

PPG is prone to artifacts due to movement or poor sensor contact and therefore robustness of the AF detection algorithm is an important aspect to consider when using PPG measured at the wrist. Poor quality data can be either discarded from the analysis or the algorithm developed must be robust to the presence of noise. The accuracy of AF detection related to measurement setting and data quality was studied in Chapters 3 and 4.

- The classification performance of AF detection based on wrist-worn PPG differed between data measured in clinical setting and in daily life due to differences in activity level and patient characteristics.
- In the clinical setting, sensitivity was higher and specificity lower compared to the daily life setting. This can be due to that in the clinical setting, non-AF rhythms contained proportionally more premature contractions than in daily life because the data was measured after the cardioversion procedure.
- Selecting PPG data segments for rhythm classification based on accelerometer data measured simultaneously at the wrist, improved the classification performance, especially specificity and positive predictive value, thus reducing false positives.

In the literature, part of the studies made with using wrist-worn PPG to detect AF do not report assessing and excluding outliers or bad quality data.[110,117,150–154] Nemati et al.[97] and Shashikumar et al.[98] did not exclude data, but selected the best channel from 8 measured PPG channels for analysis based on signal quality metrics, which were derived from the raw PPG signal and accelerometer data. Deep learning approaches have been used in order to avoid discarding bad quality data[109,135] or to develop a separate signal quality assessment algorithm to select the data to be analyzed.[114] Other methods to discard data as noisy include assessing motion level from the accelerometer data,[112,113] using standard deviation of the acceleration signals[118] or deviation from the acceleration of gravity,[85] combining information on signal-to-noise ratio and accelerometer data,[120] analyzing pulse morphology,[116] and excluding outlier beats or inter-beat intervals.[84,155] Harju et al.[155] selected the included beats based on whether they were correctly detected according to the ECG reference measurement. Combining signal quality metrics within the classification algorithm is discussed in Section 7.1.2.

The number of false positive AF detections can be also heavily influenced by other irregular rhythms. One study assessed the feasibility of AF screening with a smartphone camera,[156] which was based on a spot measurement of PPG initiated by the user, with 12 328 participants in the general population. Among the AF detections given by the algorithm, 23.7% were AF, but 34.2% were triggered by other irregularities, mainly frequent and irregular ectopic beats. Therefore, for development of AF detection methods, presence of other rhythm irregularities is important to include in the development data. In Chapter 5, multi-rhythm classification from wrist-worn PPG was studied.

- In AF detection based on traditionally used inter-beat interval features, when atrial flutter was also present, the number of false positives were mainly caused by atrial flutter and premature atrial contractions.
- Combining features derived from inter-beat intervals, PPG waveform, and accelerometer data with machine learning and classifying the rhythms into three classes instead of AF against the rest, better accuracy and smaller number of false positives was obtained.

7.1.2 PPG-based atrial fibrillation detection in daily life

Atrial fibrillation can have serious health consequences, such as stroke or heart failure.[5] The prevalence of AF is expected to increase in the future[157] and therefore solutions for timely diagnosis and treatment are needed in order to prevent stroke and other adverse events caused by this arrhythmia. From a population health perspective there is a need for effective screening strategies.[158]

Several studies have shown that continuous and prolonged monitoring with implantable devices increases detection of AF in populations that have high stroke risk or survived a stroke.[23,25,106,159] Continuous long-term monitoring can be performed currently with implantable loop recorders (ILRs) that are subcutaneous devices recording ECG. They have been designed to overcome the limitations of intermittent monitoring and lack of information on AF burden and density. The indication for monitoring can be suspected AF, e.g. in case of stroke, or need for long-term monitoring for AF management.[92] However, these devices are costly. ILRs are considered cost-effective in cryptogenic stroke patients,[160] but relevant cost-savings could be made if non-invasive technologies were available for long-term monitoring.

The challenge with the new proposed solutions today is that, as stated in the recent EHRA position paper, in relation to AF screening "the role of consumer cardiac monitoring using wearables in combination with apps is completely undefined".[161] Two large scale studies monitoring general population in an ambulatory setting, the Apple Heart Study[162] and the Huawei Heart Study,[163] have been conducted by monitoring volunteers with wrist-worn PPG for several days and investigating whether this leads to diagnosing AF. Yet, from a technical perspective, the capabilities and limitations of the PPG-based wrist-worn applications need to be well understood.

In this thesis work, the use of wrist-worn PPG as a possible solution to detect AF and the topic was approached from different perspectives. Apart from this thesis work, other studies on AF detection with wrist-worn PPG have been published since 2016, which this section aims to summarize while

discussing different relevant aspects of the research related to AF detection with wrist-worn PPG aimed for long-term monitoring.

7.1.2.1 Study populations

In the development and validation of new methods, such as detection of AF episodes using PPG, the size and diversity of the study population is important in order to understand how generalizable the proposed solutions are. In the context of AF, diversity can be related to age, presence of comorbidities, medication, and different cardiac rhythms. When considering use of PPG-technology for long-term monitoring, the health status of a person may have an impact on the activity level. This in turn may influence the accuracy of the technology because PPG measurement is sensitive to movement artifacts.

The study populations in terms of number of subjects vary from 11 to 1617 and most of the studies include less than 100 volunteers.[84,85,97,98,109,110,113,114,116,118,135,150,151,154,155,164] Six studies had a study population larger than 100 subjects,[111,112,117,120,152,153] the largest being 1617 volunteers monitored in ambulatory setting.[152]

The rhythm characteristics in the study populations vary between studies. Some focus only on the distinction between AF and sinus rhythm (SR) whereas others aim at classifying AF in the presence of other rhythms. Besides SR, other rhythms include presence of premature atrial contractions (PAC)[112,113] and premature ventricular contractions (PVC),[85,112,113] ventricular arrhythmia,[118] and atrial flutter, atrial tachycardia, and variable conduction.[85] Corino et al.[85] and Fallet et al.[118] classified other rhythms as a separate class from AF and SR. A number of studies did not specifically report what rhythms the non-AF or other rhythms included.[97,98,116,135,151,152,154] Shen et al.[135] mention that there were 8 other arrhythmias present besides AF. In addition, characteristics of AF vary between permanent, persistent, and paroxysmal, because AF characteristics evolve with time. Therefore, heart rate and complexity of the rhythm can be different depending on the type of AF.

- In the work presented in this thesis, the number of subjects for analysis were from 16 to 40 in daily life and 20 undergoing cardioversion in the hospital.
- The data from daily life used contained only continuous AF and paroxysmal AF episodes were not present. From other type of rhythms, the data set included sinus rhythm, PACs, PVCs, and atrial flutter, adding variability of other irregular rhythms.
- The limited number of subjects and a lack of paroxysmal AF cases are limitations of the dataset used in this work.

In general, the studies have been conducted with older populations in which AF is more prevalent. Most of the studies report ages for both AF and non-AF groups, but some report only range or one average age. The mean age in the group suffering from AF is in majority of studies above 70 years, the lowest being 55.7 years. In the non-AF groups, the ages were on average slightly lower.

7.1.2.2 Measurement setting and duration

The advantage of wrist-worn PPG-devices is that they are non-invasive and comfortable to wear for long time periods in daily life. The extended monitoring time can be helpful when aiming to detect intermittent AF episodes, but the ambulatory monitoring poses more challenges because of movement artifacts and compliance with the use of the device.

The measurement setting in most of the studies has been either hospital or laboratory conditions where patients are either in supine or sedentary position.[84,85,97,98,109–112,117,118,120,150–153,155,164] The measurements have been made before and after cardioversion,[109,110,112,150,152] before or during catheter ablation or electrophysiological study,[84,117,118] in emergency care or cardiac ward,[153] with post-operative patients,[155,164] admitted patients,[85,97,98,151] and in laboratory setting.[111,120,150] The recordings made in these settings lasted from few minutes up to a few hours.

Fewer studies have been performed in ambulatory setting either with patients scheduled for a Holter,[112,113] patients in cardiac rehabilitation,[116] patients with implantable cardiac monitors (ICM),[154] and an unspecified group of patients and healthy subjects.[135,152] Figure 7.1 illustrates the percentage of studies that monitored subjects on average the record length presented on the x-axis or longer. The figure shows that 45% of the studies contain recordings shorter than 30 minutes. For the studies that had both hospital and ambulatory

cohorts, the record length based on the ambulatory cohort is included in the graph. The work in this thesis is mainly based on 24-hour data from patients undergoing Holter monitoring (Chapters 3, 4, 5, and 6), but also data from patients undergoing cardioversion was used (Chapter 3).

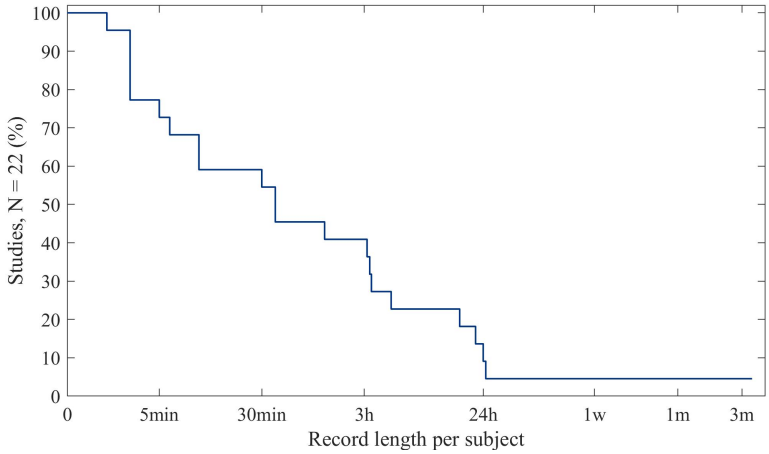

Figure 7.1: The percentage of studies in which the average record length per subject was the length on x-axis or longer.

- Comparison of AF classification performance between hospital setting and daily life setting shows that evaluating the classification accuracy in daily life is essential when developing AF detection models with wrist-worn PPG for daily life.
- 24-hour data from patients undergoing Holter monitoring can already represent a variety of daily activities and a realistic daily life setting.
- 24-hour data may still be limited when day-to-day variations are of interest or when AF is paroxysmal and very infrequent.

7.1.2.3 Methods for AF detection and PPG data quality assessment

Various methods have been used for detecting AF episodes. The methods vary in how easy they are to interpret and how easily their detection accuracy is affected when the signal is noisy. Excluding noisy data can improve detection accuracy. Therefore, automatic data quality assessment also plays a role in automated AF detection.

A common method to detect the presence of AF from a PPG signal is to assess irregularity of the rhythm from the inter-beat interval (IBI) series. Different features representing variability or entropy can be used to detect irregularity in the beat sequence. These features have been used by setting thresholds to determine whether AF is present,[113,117,153] or have been combined with machine learning to give a classification of the rhythm.[84,85,164] The IBI series have been also evaluated using a Markov model, which is a model that assigns a probability of being AF to each interval.[112,155] In Chapters 3, 4, and 5 AF detection with IBI features was studied, and in Chapter 6 a new feature based on the relationship between IBIs and PPG waveform changes was presented.

- AF detection with PPG-derived IBI features obtained comparable classification accuracy in daily life compared to ECG when data PPG data quality was assessed based on accelerometer signal.
- Heavy influence of movement in daily life was causing missing data and reduced measurement coverage because the PPG pulses were not detected or when data was chosen to be discarded as noisy.
- Information from accelerometer is not directly correlated to artifacts in PPG signal and artifacts can be also caused by poor sensor contact or micromotions that are not visible in accelerometer signal.
- The relationship between inter-beat intervals and pulse-to-pulse morphology variations in the PPG signal during AF was distinguishable compared to during sinus rhythm, and could reflect the force-interval relationship of the heart.

In addition to time domain analysis of IBIs, other features that could characterize the rhythm have been extracted from the signal. These include mainly features from the frequency domain.[98,118] Signal quality metrics have also been combined with the rhythm information, and the quality metrics have been derived either from the PPG signal itself or from the accelerometer data characterizing movement that is often measured with PPG at the wrist.[97] IBI features have been combined with this additional information with machine learning.[97,98,118] Chapter 5 studied combining multiple features for arrhythmia classification in daily life.

- Accurate multi-rhythm classification was obtained with PPG-derived features from time domain and frequency domain along with accelerometer derived features using machine learning.

In addition to traditional machine learning methods, deep learning approaches have been proposed for AF detection from wrist-worn PPG.[98,109,114,135,151,152] Shashikumar et al.[98,151] extracted first time series information and frequency domain information as the input for their deep learning network. Two approaches have been based on extracted heart rate and activity data.[152,154] Gotlibovych et al.[109] combined the PPG waveform and accelerometer data in their model whereas Aliamiri et al.[114] used this information first with one network to assess the quality of the PPG data and classified the rhythm with a second network from the good quality PPG data. Only one of the deep learning approaches was based on using only the PPG waveform without additional quality assessments or adding other sensor information.[135] Because of the small size of the dataset used in this thesis work, the use of deep learning methods was not explored.

7.1.2.4 Detection performance and measurement coverage

For screening applications, it is important to be able to detect the presence of AF, but at the same time not to cause false positive detections. Especially, the falsely detected AF episodes can cause worry and harm to the patient as well as burden the healthcare system. The movement artifacts, which are more usually present in daily life, can become an issue. Not making decisions about presence of AF when the measured data is noisy improves detection accuracy, but this has a negative impact on the measurement coverage. This in turn can lead to missing AF episodes. Accuracies for AF detection were presented in Chapters 3, 4, and 5.

- Comparison of different window lengths for feature computation showed using that longer windows (120 s) gave the best classification results.
- An accurate detection of AF using wrist-worn PPG and acceleration signals in daily life was obtained. Based on a single inter-beat interval feature, the highest sensitivity and specificity were 98.4% and 98.0%, respectively, with a measurement coverage of 25%. When combining more information from PPG signal and accelerometry with machine learning, coverage increased to 42.2%, and sensitivity and specificity were 97.6% and 98.2%, even in the presence of atrial flutter.
- Atrial flutter was classified with 84.5% sensitivity and 99.7% specificity.
- The more selective the data quality assessment is, the lower the coverage but better the detection performance.

In the literature, most often the detection performance of the algorithms is reported as sensitivity, specificity, and accuracy. All the three metrics are not always reported and some studies report additional metrics. The studies reporting accuracy, present values above 90% except for one study.[84,85,97,98,110,112–114,116,118,120,151] Solosenko et al.[116] obtained accuracy of 87% when 89.2% of the 21-hour measurements were analyzed. However, when the quality requirement for the data to be analyzed was made stricter and 50% of the data were analyzed, the performance improved in terms of sensitivity from 72.0% to 97.2%, but the improved accuracy was not reported.

Within the studies that present sensitivity and specificity,[84,85,97,98,109,111–113,116–118,120,152,153,155,164] the median sensitivity is 96.2% and specificity 97.7%. The highest sensitivity reported was 100% when specificity was 93.1%[150] whereas in the study where 100% specificity was reported, sensitivity was 96% and 93% depending on the dataset.[112] Gotlibovych et al.[109] have reported both sensitivity and specificity close to 100% (99.8% and 99.9%, respectively).

In addition to sensitivity and specificity, some studies have reported positive predictive value (PPV).[111–113,117,118,120,152–154] In hospital settings, the median PPV of the studies was 97.5% varying from 82.0% to 100%.[111,112,117,118,120,152,153] In ambulatory settings, false positives are more common. Tison et al.[152] reported PPV of 90.9% when patients undergoing cardioversion were monitored, but PPV of 7.9% in ambulatory setting. In the long-term study of Wasserlauf et al.[154] PPV was 39.9%. In 24-hour Holter monitoring with very strict restriction on not analyzing data affected by motion, reported PPV was 95.5%.[113] Bonomi et al.[112] maintained PPV of 100% both in cardioversion and Holter cohorts with a slight drop in sensitivity from 96% to 93%. In their additional study population of 120 subjects free from arrhythmia, 21% of the subjects were detected with irregularities attributable to AF and these episodes corresponded to <0.2% of the total monitoring time.

In the studies that do not exclude any data, the measurement coverage can be assumed to be 100%. With short measurements and often in a hospital setting, the coverage has been 50-95%[111,118,120,135] after noise removal. In ambulatory setting, more disturbances to the measurement are present, which can influence the coverage. Shen et al.[135] reported final results with 100% coverage with measurements from 3 to 8 hours in ambulatory setting. In other ambulatory studies, with the measurement durations of approximately 24 hours, the coverage has been 47-48%,[112] and 50-89.2%.[116]

7.1.2.5 Clinical evidence and relevance

Currently, ECG is the gold standard for diagnosis of AF. PPG technologies, as a tool for supporting clinical decisions in patients, could be valuable for evaluating the need and requirements for ECG monitoring when AF is suspected. In addition, spot measurements of ECG could be made with wrist-worn devices that combine both PPG and ECG[97,165] or with separate ECG devices connected to smartphones or tablets[166] when continuous PPG-based measurement indicates rhythm irregularity.

The overview of studies investigating AF detection with wrist-worn PPG measurements presented here shows that an excellent accuracy can in general be obtained. However, when focusing on long-term monitoring, the results of the studies should always be reflected on the circumstances, such as the diversity of the study population and the setting in which the monitoring was performed. The majority of studies has been conducted with hospitalized patients and in a setting in which the study subjects are less active than during ambulatory monitoring. Less studies have been performed in ambulatory setting where movement artifacts are more likely to be present. Within these studies the results vary more, especially the number of false positives and PPV. Unreliable and bad quality data can be discarded to improve detection accuracy, but this induces a data loss, leading to a trade-off between the performance and measurement coverage. Development of new techniques for robust detection methods are needed in order to obtain high accuracy and coverage in long-term ambulatory monitoring. In addition, studies assessing the impact of measurement coverage on the diagnostic yield are needed.

In many studies, the subjects with AF are often known to have AF or otherwise have a high probability of arrhythmia. For example, hospitalized patients and patients undergoing cardiac procedures such as cardioversion or cardiac ablation. These populations may have different general health conditions, comorbidities, and medication as compared to the more general population targeted for screening.

The Apple Heart Study[162] and the Huawei Heart Study,[163] aimed to study feasibility of AF screening in large populations by monitoring participants for several days with wrist-worn PPG. Of the study populations, 0.5%[162] and 0.2%[163] were notified about irregular rhythms as an indication to contact a clinical expert. This seems reasonable when considering the prevalence of AF. As a part of their inclusion criteria, both studies required the participants to own both an adequate smartphone and smartwatch. The prevalence of AF increases with older age,[167] but especially smartwatches are usually owned by younger people.[168] In the Huawei Heart Study, the mean age of the participants was 34.7 ($\pm$11.5) years with only less than 6% being above 54 years.[163] In the Apple Heart Study, this proportion of participants was 15.9%.[162] Mass screening

of the population that owns a consumer device, these being younger than the groups at risk, is likely not to target the people suffering from undiagnosed AF and increases the risk of causing false positives in a presumably healthy population.

7.1.2.6 Comparison of studies

The heterogeneity of study populations, measurement settings, methodological decisions, such as the segment length to be classified, and using different metrics to report the results, makes it difficult to compare the methods and results objectively. For the development of more accurate methods for detection of AF, and possibly other arrhythmias, this is important. Currently, there is a lack of publicly available labelled datasets of wrist-worn PPG measurements in ambulatory setting, which would make this comparison possible. Availability of a large dataset, which is representative for the population and measurement conditions, would enable a fair comparison of different solutions and open the possibility for more researchers to work on solutions. This would accelerate the technical development and lead to more accurate solutions faster than today. For ECG analysis, such progress has been made possible because of the availability of various open source datasets and software, such as the ones on PhysioNet.[169]

Different studies report their results with different metrics and often measurement coverage has not been explicitly reported. Considering the importance of false positives, reporting PPV is an important addition to sensitivity and specificity. Coverage should be explicitly reported in order to understand if there is a data loss which can possibly have an impact on the diagnostic yield. In addition, subject characteristics, such as age and heart rhythms present during the monitoring, may influence the performance of the detection method and therefore should be included.

7.2 FUTURE PROSPECTS

The current limited number of studies about detecting AF with measuring wrist-worn PPG in ambulatory setting calls for more studies showing how well the technology works in daily life, especially on detecting individual episodes and the duration of episodes. The optimal balance between measurement coverage and detection accuracy is not yet fully understood. If full measurement coverage with good accuracy can be obtained, continuous monitoring could enable better assessment of AF burden and not only detecting presence of AF. Moreover, more insight could be obtained on the daily patterns of the AF episodes, whether they are related to triggers and have symptoms connected

to them, and if medication is adequate for the management of AF. For the latter, accuracy of heart rate measurement during AF episodes is of great importance in order to assess whether heart rate stays within the desired limits. Recently in the TACTIC-AF study,[170] intermittent anticoagulation guided by continuous monitoring of AF burden with pacemakers and implantable cardiodefibrillators obtained promising results in patients with rare AF episodes and low to moderate stroke risk. More research on the topic is warranted, but these results on guided medication intake are an additional indication that unobtrusive continuous AF monitoring could have value beyond screening if the technical challenges related to the measurement coverage can be overcome.

The current knowledge on the ability to distinguish different types of arrhythmia from PPG is still limited. More detailed distinction to different type of arrhythmias could add value both in clinical setting and in daily life. In daily life monitoring, the number of premature contractions and whether they are atrial or ventricular could provide further understanding about symptoms, possible false detections of AF, and risk of developing other arrhythmia. In clinical setting, recognition of ventricular beats and arrhythmias from pulsatile waveforms could further help in improving accuracy for alarming. In addition, further development by using artificial intelligence to analyze data from multiple sources during hospital monitoring could help in providing right alarms on the right time.

The results in this thesis and in the literature show that accurate AF detection from PPG is possible with the current methods. There are still challenges for the methods to work accurately with poor signal quality and in the presence of other arrhythmias. The interesting directions for methodological development are in improving AF detection under circumstances that can induce many artifacts, and distinguishing different types of arrhythmias from each other and whether they originate from the atria or ventricles. In ECG analysis, both visual and automated, P-wave analysis has an important role in AF detection and diagnosis. Considering the underlying measurement principle of PPG, the measured signal reflects ventricular activity. Whether with very advanced data analysis techniques a surrogate for P-wave analysis from PPG signals can be found, remains to be investigated. In addition, analyzing measured PPG signals during different stages of AF, i.e. paroxysmal, permanent, etc., or even prediction of possible AF can be interesting for future research. For prediction of AF, underlying comorbidities are also important. Alternatively, AF can lead to heart failure and studying long-term PPG-based monitoring of AF could be of interest for potentially detecting deterioration of cardiac function earlier[171] and leading to adequate treatment.

Similar studies as the Apple Heart Study and the Huawei Heart Study targeting populations at risk of having AF are currently missing. The results

and the accuracy of the solutions may change when most of the monitored people have higher risk of having AF, but possibly also other arrhythmias and comorbidities. In addition, there is a need for randomized trials to assess whether detecting AF with wrist-worn devices in daily life leads to safe and effective treatment for reducing strokes. The relation between stroke risk and duration of device detected AF has been studied with implantable devices with varying results.[172,173] Therefore, the insight on stroke risk related to AF detected with wrist-worn wearables is important in order to understand the value of the technology, on the one hand in terms of maintaining people healthier and improving their quality of life, and on the other hand saving costs and improving the healthcare system.

BIBLIOGRAPHY

[1] Gregory Y. H. Lip and D. Gareth Beevers. ABC of Atrial Fibrillation: History, Epidemiology, and importance of atrial fibrillation. *BMJ*, 311(7016):1361, 1995.

[2] Regis A. Desilva. George Ralph Mines, ventricular fibrillation and the discovery of the vulnerable period. *Journal of the American College of Cardiology*, 29(6):1397–1402, 1997.

[3] Mark E. Silverman. From rebellious palpitations to the discovery of auricular fibrillation: Contributions of Mackenzie, Lewis and Einthoven. *The American Journal of Cardiology*, 73(5):384–389, 1994.

[4] G. Y. H. Lip et al. Atrial fibrillation. *Nat Rev Dis Primers*, 2(16016):1–26, 2016.

[5] Paulus Kirchhof et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Europace*, 18(11):1609–1678, 2016.

[6] Jocasta Ball, Melinda J Carrington, John J V McMurray, and Simon Stewart. Atrial fibrillation : Profile and burden of an evolving epidemic in the 21st century. *International Journal of Cardiology*, 167:1807–1824, 2013.

[7] N. Saoudi et al. A classification of atrial flutter and regular atrial tachycardia according to electrophysiological mechanisms and anatomical bases: A statement from a joint expert group from the working group of arrhythmias of the European society of cardiology and the North American society of pacing and electrophysiology. *European Heart Journal*, 22(14):1162–1182, 2001.

[8] Kenneth A. Ellenbogen and Jayanthi N. Koneru. Atrial Tachycardia. In *Cardiac Electrophysiology: From Cell to Bedside: Seventh Edition*, pages 681–699. Elsevier Inc., seventh ed edition, 2018.

[9] Chrishan Joseph Nalliah, Saurabh Kumar, Prashanthan Sanders, and Jonathan M. Kalman. Typical and Atypical Atrial Flutter: Mapping and Ablation. In *Cardiac Electrophysiology: From Cell to Bedside: Seventh Edition*, pages 713–723. Elsevier Inc., seventh ed edition, 2018.

[10] Sok Sithikun Bun, Decebal Gabriel Latcu, Francis Marchlinski, and Nadir Saoudi. Atrial flutter: More than just one of a kind. *European Heart Journal*, 36(35):2356–2363, 2015.

[11] Hooman Kamel et al. Paroxysmal supraventricular tachycardia and the risk of ischemic stroke. *Stroke*, 44(6):1550–1554, 2013.

[12] Sana M. Al-Khatib et al. 2017 AHA/ACC/HRS Guideline for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death. *Circulation*, 138(13):e272–e391, 2018.

[13] Kristin E. Sandau et al. Update to Practice Standards for Electrocardiographic Monitoring in Hospital Settings: A Scientific Statement From the American Heart Association. *Circulation*, 136(19):e273–e344, 2017.

[14] Barbara J. Drew, Patricia Harris, Jessica K. Zègre-Hemsey, Tina Mammone, Daniel Schindler, Rebeca Salas-Boni, Yong Bai, Adelita Tinoco, Quan Ding, and Xiao Hu. Insights into the problem of alarm fatigue with physiologic monitor devices: A comprehensive observational study of consecutive intensive care unit patients. *PLoS ONE*, 9(10), 2014.

[15] Anton Aboukhalil, Larry Nielsen, Mohammed Saeed, Roger G. Mark, and Gari D. Clifford. Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. *Journal of Biomedical Informatics*, 41:442–451, 2008.

[16] M. C. Chambrin. Alarms in the intensive care unit: How can the number of false alarms be reduced? *Critical Care*, 5(4):184–188, 2001.

[17] D. Balogh, E. Kittinger, A. Benzer, and J. M. Hackl. Noise in the ICU. *Intensive Care Medicine*, 19:343–6, 1993.

[18] Avinash Konkani and Barbara Oakley. Noise in hospital intensive care units-a critical review of a critical topic. *Journal of Critical Care*, 27(5):522.e1–522.e9, 2012.

[19] Peter Zimetbaum and Alena Goldman. Ambulatory arrhythmia monitoring: Choosing the right device. *Circulation*, 122(16):1629–1636, 2010.

[20] Jonathan S. Steinberg et al. 2017 ISHNE-HRS expert consensus statement on ambulatory ECG and external cardiac monitoring / telemetry. *Heart Rhythm*, 14(7), 2017.

[21] Tommaso Sanna. Long-term monitoring to detect atrial fibrillation with the indwelling implantable cardiac monitors. *International Journal of Stroke*, 13(9):893–904, 2018.

[22] Suneet Mittal, Colin Movsowitz, and Jonathan S. Steinberg. Ambulatory external electrocardiographic monitoring: Focus on atrial fibrillation. *Journal of the American College of Cardiology*, 58(17):1741–1749, 2011.

[23] Tommaso Sanna et al. Cryptogenic stroke and underlying atrial fibrillation. *The New England Journal of Medicine*, 370(26):2478–86, 2014.

[24] James A. Reiffel et al. Incidence of previously undiagnosed atrial fibrillation using insertable cardiac monitors in a high-risk population: The REVEAL AF study. *JAMA Cardiology*, 2(10):1120–1127, 2017.

[25] Johannes Brachmann et al. Uncovering atrial fibrillation beyond short-term monitoring in cryptogenic stroke patients: Three-year results from the cryptogenic stroke and underlying atrial fibrillation trial. *Circulation: Arrhythmia and Electrophysiology*, 9(1):1–10, 2016.

[26] H. T. Tu, S. Spence, J. M. Kalman, and S. M. Davis. Twenty-eight day Holter monitoring is poorly tolerated and insensitive for paroxysmal atrial fibrillation detection in cryptogenic stroke. *Internal Medicine Journal*, 44(5):505–508, 2014.

[27] Spencer Z. Rosero, Valentina Kutyifa, Brian Olshansky, and Wojciech Zareba. Ambulatory ECG monitoring in atrial fibrillation management. *Progress in Cardiovascular Diseases*, 56(2):143–152, 2013.

[28] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28:R1–R39, 2007.

[29] Toshiyo Tamura, Yuka Maeda, Masaki Sekine, and Masaki Yoshida. Wearable Photoplethysmographic Sensors—Past and Present. *Electronics*, 3(2):282–302, 2014.

[30] Ralph Wijshoff. *On photoplethysmography artifact reduction and applications*. PhD thesis, Eindhoven University of Technology, 2016.

[31] G. B. Moody and R. G. Mark. A new method for detecting atrial fibrillation using R-R intervals. In *Computers in Cardiology*, volume 10, pages 227–230, 1983.

[32] Saeed Babaeizadeh, Richard E. Gregg, Eric D. Helfenbein, James M. Lindauer, and Sophia H. Zhou. Improvements in atrial fibrillation detection for real-time monitoring. *Journal of Electrocardiology*, 42(6):522–526, 2009.

[33] K. Tateno and L. Glass. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ARR intervals. *Med. Biol. Eng. Comput*, 39:664–671, 2001.

[34] M. G. Tsipouras, D. I. Fotiadis, and D. Sideris. An arrhythmia classification system based on the RR-interval signal. *Artificial Intelligence in Medicine*, 33(3):237–250, 2005.

[35] Hans D. Esperer, Chris Esperer, and Richard J. Cohen. Cardiac arrhythmias imprint specific signatures on Lorenz plots. *Annals of Noninvasive Electrocardiology*, 13(1):44–60, 2008.

[36] Jinho Park, Sangwook Lee, and Moongu Jeon. Atrial fibrillation detection by heart rate variability in Poincare plot. *Biomedical engineering online*, 8:38, 2009.

[37] Shantanu Sarkar, David Ritscher, and Rahul Mehra. Communications A Detector for a Chronic Implantable Atrial. *IEEE Transactions on Biomedical Engineering*, 55(3):1219–1224, 2008.

[38] Lijuan Zhang, Tianci Guo, Bin Xi, Yang Fan, Kun Wang, Jiacheng Bi, and Ying Wang. Automatic recognition of cardiac arrhythmias based on the geometric patterns of Poincaré plots. *Physiological Measurement*, 36(2):283–301, 2015.

[39] S. Dash, K. H. Chon, S. Lu, and E. A. Raeder. Automatic real time detection of atrial fibrillation. *Annals of Biomedical Engineering*, 37(9):1701–1709, 2009.

[40] D. E. Lake and J. R. Moorman. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *AJP: Heart and Circulatory Physiology*, 300(1):H319–H325, 2011.

[41] Alireza Ghodrati, Bill Murray, and Stephen Marinello. RR Interval Analysis for Detection of Atrial Fibrillation in ECG Monitors. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008(1):601–604, 2008.

[42] Marta Carrara et al. Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy. *Physiological Measurement*, 36(9):1873–1888, 2015.

[43] Helmut Pürerfellner et al. P-wave evidence as a method for improving algorithm to detect atrial fibrillation in insertable cardiac monitors. *Heart Rhythm*, 11(9):1575–1583, 2014.

[44] Prashanthan Sanders et al. Performance of a new atrial fibrillation detection algorithm in a miniaturized insertable cardiac monitor: Results from the Reveal LINQ Usability Study. *Heart Rhythm*, 13(7):1425–1430, 2016.

[45] Oliver Faust, Alex Shenfield, Murtadha Kareem, Tan Ru San, Hamido Fujita, and U. Rajendra Acharya. Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. *Computers in Biology and Medicine*, 102(May):327–335, 2018.

[46] Yong Xia, Naren Wulan, Kuanquan Wang, and Henggui Zhang. Detecting atrial fibrillation by deep convolutional neural networks. *Computers in Biology and Medicine*, 93(July 2017):84–92, 2018.

[47] Zachi I. Attia et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet*, 394(10201):861–867, 2019.

[48] Awni Y. Hannun et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1):65–69, 2019.

[49] Jo Woon Chong, David D. McManus, and Ki H. Chon. Arrhythmia discrimination using a smart phone. *IEEE Journal of Biomedical and Health Informatics*, 19(3):1–4, 2015.

[50] David D. McManus et al. A novel application for the detection of an irregular pulse using an iPhone 4S in patients with atrial fibrillation. *Heart Rhythm*, 10(3):315–319, 2013.

[51] David D. McManus et al. PULSE-SMART: Pulse-based arrhythmia discrimination using a novel smartphone application. *Journal of Cardiovascular Electrophysiology*, 27(1):51–57, 2016.

[52] Sylvia Siebig, Silvia Kuhls, Michael Imhoff, Ursula Gather, Jürgen Schölmerich, and Christian E. Wrede. Intensive care unit alarms–how many do we need? *Critical Care Medicine*, 38(2):451–456, 2010.

[53] S. T. Lawless. Crying wolf: false alarms in a pediatric intensive care unit. *Critical Care Medicine*, 22:981–5, 1994.

[54] C. Tsien and J. Fackler. Poor prognosis for existing monitors in the intensive care unit. *Critical Care Medicine*, 25:614–9, 1997.

[55] Kelly Creighton Graham and Maria Cvach. Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms. *American Journal of Critical Care*, 19(1):28–34, 2010.

[56] A. Mäkivirta, E. Koski, A. Kari, and T. Sukuvaara. The median filter as a preprocessor for a patient monitor limit alarm system in intensive care. *Computer Methods and Programs in Biomedicine*, 34(2-3):139–144, 1991.

[57] M. Borowski, S. Siebig, C. Wrede, and M. Imhoff. Reducing False Alarms of Intensive Care Online-Monitoring Systems: An Evaluation of Two Signal Extraction Algorithms. *Computational and Mathematical Methods in Medicine*, 2011:1–11, 2011.

[58] Omid Sayadi and Mohammad B. Shamsollahi. Life-threatening arrhythmia verification in ICU patients using the joint cardiovascular dynamical model and a bayesian filter. *IEEE Transactions on Biomedical Engineering*, 58(10):2748–2757, 2011.

[59] Joachim Behar, Julien Oster, Qiao Li, and Gari D. Clifford. ECG signal quality during arrhythmia and its application to false alarm reduction. *IEEE Transactions on Biomedical Engineering*, 60(6):1660–1666, 2013.

[60] G. D. Clifford, A. Aboukhalil, J. X. Sun, W. Zong, B. A. Janz, G. B. Moody, and R. G. Mark. Using the blood pressure waveform to reduce critical false ECG alarms. *Computers in Cardiology*, pages 829–832, 2006.

[61] Anagha Vishwas Deshmane. False arrhythmia alarm suppression using ECG, ABP, and photoplethysmogram. Master's thesis, Massachusetts Institute of Technology, 9 2009.

[62] Q. Li, R. G. Mark, and G. D. Clifford. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiological Measurement*, 29:15–32, 2008.

[63] Gari D. Clifford, Ikaro Silva, Benjamin Moody, Qiao Li, Danesh Kella, Abdullah Shahin, Tristan Kooistra, Diane Perry, and Roger G. Mark. The PhysioNet / Computing in Cardiology Challenge 2015 : Reducing false arrhythmia alarms in the ICU. *Computing in Cardiology*, pages 273–276, 2015.

[64] Paula Couto, Ruben Ramalho, and Rui Rodrigues. Suppression of false arrhythmia alarms using ECG and pulsatile waveforms. *Computing in Cardiology*, pages 749–752, 2015.

[65] Chathuri Daluwatte, Lars Johannesen, Jose Vicente, Christopher G. Scully, Loriano Galeotti, and David G. Strauss. Heartbeat fusion algorithm to reduce false alarms for arrhythmias. *Computing in Cardiology*, pages 745–748, 2015.

[66] V. Krasteva, I. Jekova, R. Leber, R. Schmid, and R. Abächerli. Validation of arrhythmia detection library on bedside monitor data for triggering alarms in intensive care. *Computing in Cardiology*, pages 737–740, 2015.

[67] Wei Zong. Reduction of false critical ECG alarms using waveform features of arterial blood pressure and / or photoplethysmogram signals. *Computing in Cardiology*, pages 289–292, 2015.

[68] Filip Plesinger, Petr Klimes, Josef Halamek, and Pavel Jurak. False alarms in intensive care unit monitors : detection of life-threatening arrhythmias using elementary algebra, descriptive statistics and fuzzy logic. *Computing in Cardiology*, pages 281–284, 2015.

[69] Sibylle Fallet, Sasan Yazdani, and Jean-Marc Vesin. A multimodal approach to reduce false arrhythmia alarms in the intensive care unit.

Computing in Cardiology, pages 277–280, 2015.

[70] Christoph Hoog Antink and Steffen Leonhardt. Reducing false arrhythmia alarms using robust interval estimation and machine learning. *Computing in Cardiology*, pages 285–288, 2015.

[71] Vignesh Kalidas and Lakshman Tamil. Enhancing accuracy of arrhythmia classification by combining logical and machine learning techniques. *Computing in Cardiology*, pages 733–736, 2015.

[72] C. J. van Rijsbergen. *Information Retrieval*. Butterworths, London, 2nd edition, 1979.

[73] Michael J. Rooijakkers, Chiara Rabotti, S. Guid Oei, and Massimo Mischi. Low-complexity R-peak detection for ambulatory fetal monitoring. *Physiological Measurement*, 33:1135–1150, 2012.

[74] W. Zong, T. Heldt, G.B. Moody, and R.G. Mark. An open-source algorithm to detect onset of arterial blood pressure pulses. *Computers in Cardiology*, pages 259–262, 2003.

[75] Linda M. Eerikäinen, Joaquin Vanschoren, Michael J. Rooijakkers, Rik Vullings, and Ronald M. Aarts. Decreasing the false alarm rate of arrhythmias in intensive care using a machine learning approach. *Computing in Cardiology*, pages 293–296, 2015.

[76] Marco A. F. Pimentel, Mauro D. Santos, David B. Springer, and Gari D. Clifford. Heart beat detection in multimodal physiological data using a hidden semi-Markov model and signal quality indices. *Physiological Measurement*, 36(8):1717–1727, 2015.

[77] I. I. Goncharova and J. S. Barlow. Changes in EEG mean frequency and spectral purity during spontaneous alpha blocking. *Electroencephalography and clinical neurophysiology*, 76(3):197–204, 1990.

[78] Leif Sörnmo and Pablo Laguna. *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Academic Press, 2005.

[79] L. Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.

[80] C. L. Tsien, I. S. Kohane, and N. McIntosh. Multiple signal integration by decision tree induction to detect artifacts in the neonatal intensive care unit. *Artificial Intelligence in Medicine*, 19(3):189–202, 2000.

[81] Ying Zhang and Peter Szolovits. Patient-specific learning in real time for adaptive monitoring in critical care. *Journal of Biomedical Informatics*, 41(3):452–60, 2008.

[82] W. Sieben and U. Gather. Classifying Alarms in Intensive Care - Analogy to Hypothesis Testing. *Artificial Intelligence in Medicine*, pages

130–138, 2007.

[83] A. John Camm et al. 2012 focused update of the ESC Guidelines for the management of atrial fibrillation. *Europace*, 14:1385–413, 2012.

[84] Mathieu Lemay, Sibylle Fallet, Philippe Renevey, Josep Sol, Etienne Pruvot, and Jean-Marc Vesin. Wrist-Located Optical Device for Atrial Fibrillation Screening : A Clinical Study on Twenty Patients. *Computing in Cardiology*, pages 681–684, 2016.

[85] Valentina D. A. Corino, Rita Laureanti, Lorenzo Ferranti, Giorgio Scarpini, Federico Lombardi, and Luca T. Mainardi. Detection of atrial fibrillation episodes using a wristband device. *Physiological Measurement*, 38:787–799, 2017.

[86] Alberto G Bonomi, Fons Schipper, Linda M Eerikäinen, Jenny Margarito, Ronald M Aarts, Saeed Babaeizadeh, Helma M De Morree, and Lukas Dekker. Atrial Fibrillation Detection using Photoplethysmography and Acceleration Data at the Wrist. *Computing in Cardiology*, pages 277–280, 2016.

[87] J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol*, 278:H2039–H2049, 2000.

[88] A. J. Camm and et al. Guidelines for the management of atrial fibrillation: The Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *European Heart Journal*, 31(19):2369–2429, 2010.

[89] R. L. Page, W. E. Wilkinson, W. K. Clair, E. A. McCarthy, and E. L. Pritchett. Asymptomatic arrhythmias in patients with symptomatic paroxysmal atrial fibrillation and paroxysmal supraventricular tachycardia. *Circulation*, 89(1):224–227, 1994.

[90] Giulio Valenti and Klaas R. Westerterp. Optical heart rate monitoring module validation study. *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, pages 195–196, 2013.

[91] Judith van Andel, Constantin Ungureanu, Ronald Aarts, Frans Leijten, and Johan Arends. Using photoplethysmography in heart rate monitoring of patients with epilepsy. *Epilepsy & Behavior*, 45:142–145, 2015.

[92] Jinseok Lee, Bersain A. Reyes, David D. McManus, Oscar Mathias, and Ki H. Chon. Atrial fibrillation detection using a smart phone. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1177–80, 2012.

[93] Pak Hei Chan et al. Diagnostic Performance of a Smartphone-Based

Photoplethysmographic Application for Atrial Fibrillation Screening in a Primary Care Setting. *Journal of the American Heart Association*, 5(7):1–8, 2016.

[94] Tim Schäck, Yosef Safi Harb, Michael Muma, and Abdelhak M Zoubir. Computationally efficient algorithm for photoplethysmography-based atrial fibrillation detection using smartphones. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 104–108, 2017.

[95] S. M. Shan, S. C. Tang, P. W. Huang, Y. M. Lin, W. H. Huang, D. M. Lai, and A. Y. A. Wu. Reliable PPG-based algorithm in atrial fibrillation detection. In *2016 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 340–343, Oct 2016.

[96] Sung-Chun Tang, Pei-Wen Huang, Chi-Sheng Hung, Shih-Ming Shan, Yen-Hung Lin, Jiann-Shing Shieh, Dar-Ming Lai, An-Yeu Wu, and Jiann-Shing Jeng. Identification of atrial fibrillation by quantitative analyses of fingertip photoplethysmogram. *Scientific Reports*, 7(October 2016):45644, 2017.

[97] Shamim Nemati, Mohammad M. Ghassemi, Vaidehi Ambai, Nino Isakadze, Oleksiy Levantsevych, Amit Shah, and Gari D. Clifford. Monitoring and detecting atrial fibrillation using wearable technology. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3394–3397, 2016.

[98] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S Nemati. A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology. *IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 141–144, 2017.

[99] Alexandros Pantelopoulos, Anthony Faranesh, Andreea Milescu, Paige Hosking, Subramaniam Venkatraman, and Conor Heneghan. Screening of atrial fibrillation using wrist photoplethysmography from a Fitbit tracker. *Iproceedings*, 3(1):e17, 2017.

[100] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37, 1960.

[101] W. J. Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950.

[102] Julien Oster and Gari D. Clifford. Impact of the presence of noise on RR interval-based atrial fibrillation detection. *Journal of Electrocardiology*, 48(6):947–51, 2015.

[103] Linda M. Eerikäinen et al. Validating features for atrial fibrillation detec-

tion from photoplethysmogram under hospital and free-living conditions. *Computing in Cardiology*, pages 3–6, 2017.

[104] Staffan Björck, Bo Palaszewski, Leif Friberg, and Lennart Bergfeldt. Atrial fibrillation, stroke risk, and warfarin therapy revisited. *Stroke*, 44(11):3103–3108, 2013.

[105] Moti Haim et al. Prospective national study of the prevalence, incidence, management and outcome of a large contemporary cohort of patients with incident non-valvular atrial fibrillation. *Journal of the American Heart Association*, 4(1):1–12, 2015.

[106] Jeff S. Healey et al. Subclinical atrial fibrillation and the risk of stroke. *New England Journal of Medicine*, 366(2):120–129, 2012.

[107] Richard L. Page et al. 2015 ACC/AHA/HRS Guideline for the management of adult patients with supraventricular tachycardia: Executive summary. *Journal of the American College of Cardiology*, 67(13):1575–1623, 2015.

[108] S. C. Halligan, B. J. Gers, and R. D. Brown. The natural history of lone atrial flutter. *Annals of Internal Medicine*, 40(4):265–268, 2004.

[109] Igor Gotlibovych et al. End-to-end deep learning from raw sensor data: atrial fibrillation detection using wearables. *arXiv preprint*, 2018.

[110] Chengming Yang et al. Using PPG signals and wearable devices for atrial fibrillation screening. *IEEE Transactions on Industrial Electronics*, 2019.

[111] Yong-Yan Fan, Yan-Guang Li, Jian Li, Wen-Kun Cheng, and Zhao-Liang Shan. Diagnostic performance of a smart device with photoplethysmography technology for atrial fibrillation detection: Pilot study (Pre-mAFA II Registry). *JMIR mHealth and uHealth*, 7(3):1–11, 2019.

[112] Alberto G. Bonomi et al. Atrial fibrillation detection using a novel cardiac ambulatory monitor based on photo-plethysmography at the wrist. *Journal of the American Heart Association*, 7(15), 2018.

[113] Linda M. Eerikäinen et al. Comparison between electrocardiogram- and photoplethysmogram-derived features for atrial fibrillation detection in free-living conditions. *Physiological Measurement*, 39(8), 2018.

[114] Alireza Aliamiri and Yichen Shen. Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor. In *2018 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2018*, volume 2018, pages 442–445, 2018.

[115] Maxime Voisin et al. Ambulatory Atrial Fibrillation Monitoring Using

Wearable Photoplethysmography with Deep Learning. *arXiv preprint*, 2018.

[116] Andrius Sološenko, Andrius Petrėnas, Leif Sörnmo, Birutė Paliakaitė, and Vaidotas Marozas. Detection of atrial fibrillation using a wrist-worn device. *Physiological Measurement*, 40:025003, 2019.

[117] Asami Kashiwa et al. Performance of an atrial fibrillation detection algorithm using continuous pulse wave monitoring. *Annals of Noninvasive Electrocardiology*, 24(2):1–8, 2019.

[118] Sibylle Fallet et al. Can one detect atrial fibrillation using a wrist-type photoplethysmographic device? *Medical & Biological Engineering & Computing*, 57(2):477–487, 2018.

[119] Ming Zher Poh et al. Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms. *Heart*, pages 1921–1928, 2018.

[120] Marcus Dörr et al. The WATCH AF trial: SmartWATCHes for detection of atrial fibrillation. *JACC. Clinical electrophysiology*, 5(2):199–208, 2019.

[121] David E. Krummen, Gregory K. Feld, and Sanjiv M. Narayan. Diagnostic accuracy of irregularly irregular RR intervals in separating atrial fibrillation from atrial flutter. *American Journal of Cardiology*, 98(2):209–214, 2006.

[122] Francesco Sartor, Jos Gelissen, Ralph van Dinther, David Roovers, Gabriele B. Papini, and Giuseppe Coppola. Wrist-worn optical and chest strap heart rate comparison in a heterogeneous sample of healthy individuals and in coronary artery disease patients. *BMC Sports Science, Medicine and Rehabilitation*, 10(1):1–10, 2018.

[123] Nandakumar Selvaraj, Yitzhak Mendelson, Kirk H. Shelley, David G. Silverman, and Ki H. Chon. Statistical approach for the detection of motion/noise artifacts in photoplethysmogram. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 4972–4975, 2011.

[124] Bo Hjorth. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3):306–310, 1970.

[125] Eduardo Gil, José María Vergara, and Pablo Laguna. Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children. *Biomedical Signal Processing and Control*, 3:267–277, 2008.

[126] Sibylle Fallet, Sasan Yazdani, and Jean-Marc Vesin. False arrhythmia

alarms reduction in the intensive care unit: a multimodal approach. *Physiological Measurement*, (37):1217–1232, 2016.

[127] Linda M. Eerikäinen, Joaquin Vanschoren, Michael J. Rooijakkers, Rik Vullings, and Ronald M. Aarts. Reduction of false arrhythmia alarms using signal selection and machine learning. *Physiological Measurement*, 37:1204–1216, 2016.

[128] I. A. Rezek and S. J. Roberts. Stochastic complexity measures for physiological signal analysis. *IEEE Transactions on Biomedical Engineering*, 45(9):1186–1191, 1998.

[129] H Peng, F Long, and C Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

[130] Hanchuan Peng. mRMR (minimum Redundancy Maximum Relevance Feature Selection). `http://home.penglab.com/proj/mRMR/`. [Online; accessed 11-February-2019].

[131] G. B. Papini et al. Sinus or not: a new beat detection algorithm based on a pulse morphology quality index to extract normal sinus rhythm beats from wrist-worn photoplethysmography recordings. *Physiological Measurement*, 39(11), 2018.

[132] Linda M. Eerikäinen et al. Force-interval relationships of the heart measured with photoplethysmography during atrial fibrillation. In *Computing in Cardiology Conference*, volume 45, pages 1–4, 2018.

[133] Tudor Besleaga et al. Detecting mechanical alternans utilizing photoplethysmography. In *Computing in Cardiology Conference*, volume 45, pages 1–4, 2018.

[134] Eberhard P Scholz et al. Discriminating atrial flutter from atrial fibrillation using a multilevel model of atrioventricular conduction. *Heart Rhythm*, 11(5):877–884, 2014.

[135] Yichen Shen, Maxime Voisin, Alireza Aliamiri, Anand Avati, Awni Hannun, and Andrew Ng. Ambulatory atrial fibrillation monitoring using wearable photoplethysmography with deep learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1909–1916, 2019.

[136] M. W. Cooper. Postextrasystolic potentiation. Do we really know what it means and how to use it? *Circulation*, 88:2962–2971, 1993.

[137] O. Lagendorff. Untersuchungen am überlebenden Säugetierherzen III. Vorübergehende Unregelmässigkeiten der Herzschlages und ihre Aus-

gleichung. *Pflügers Archiv*, 70:473–486, 1898.

[138] J. P. Merillon and et al. Post-extrasystolic left ventricular peak pressure with and without left ventricular failure. *Cardiovascular Research*, 13:338–344, 1979.

[139] B. F. Hoffman, H. J. Bartelstone, B. J. Scherlag, and P. F. Cranefield. Effects of postextrasystolic potentiation on normal and failing hearts. *Bulletin of the New York Academy of Medicine*, 41(5):498–534, 1965.

[140] W. A. Seed and et al. Relationships between beat-to-beat interval and the strength of contraction in the healthy and diseased human heart. *Circulation*, 70(5):799–805, 1984.

[141] David J. Sprenkeler and Marc A. Vos. Post-extrasystolic potentiation: link between Ca2+ homeostasis and heart failure? *Arrhythmia & Electrophysiology Review*, 5(1):20–26, 2016.

[142] Daniel Sinnecker and et al. Force-interval relationship predicts mortality in survivors of myocardial infarction with atrial fibrillation. *International Journal of Cardiology*, 182:315–320, 2015.

[143] S. M. Hardman, M. I. Noble, and W. A. Seed. Postextrasystolic potentiation and its contribution to the beat-to-beat variation of the pulse during atrial fibrillation. *Circulation*, 86(4):1223–1232, 1992.

[144] Suzanna M. C. Hardman, Mark I. M. Noble, Tony Biggs, and W. Anthony Seed. Evidence for an influence of mechanical restitution on beat-to-beat variations in haemodynamics during chronic atrial fibrillation in patients. *Cardiovascular Research*, 38(1):82–90, 1998.

[145] C. I. Brookes, P. A. White, M. Staples, P. J. Oldershaw, A. N. Redington, P. D. Collins, and M. I. Noble. Myocardial contractility is not constant during spontaneous atrial fibrillation in patients. *Circulation*, 98(17):1762–1768, 1998.

[146] Liang-Han Ling, Peter M. Kistler, Jonathan M. Kalman, Richard J. Schilling, and Ross J. Hunter. Comorbidity of atrial fibrillation and heart failure. *Nature Reviews Cardiology*, 13(3):131–147, 2015.

[147] Atul Verma, Jonathan M. Kalman, and David J. Callans. Treatment of patients with atrial fibrillation and heart failure with reduced ejection fraction. *Circulation*, 135(16):1547–1563, 2017.

[148] Lin Yang, Song Zhang, Xiaoyang Li, and Yimin Yang. Removal of pulse waveform baseline drift using cubic spline interpolation. *2010 4th International Conference on Bioinformatics and Biomedical Engineering, iCBBE 2010*, pages 1–3, 2010.

[149] Gari D. Clifford et al. False alarm reduction in critical care. *Physiological Measurement*, 37(8):E5–E23, 2016.

[150] Aviram Hochstadt, Ehud Chorin, Sami Viskin, Arie Lorin Schwartz, Natan Lubman, and Raphael Rosso. Continuous heart rate monitoring for automatic detection of atrial fibrillation with novel bio-sensing technology. *Journal of Electrocardiology*, 52:23–27, 2019.

[151] Supreeth P. Shashikumar, Amit J. Shah, Gari D. Clifford, and Shamim Nemati. Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks. *arXiv preprint*, 2018.

[152] Geoffrey H. Tison et al. Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiology*, 3(5):409–416, 2018.

[153] E. S. Väliaho, P. Kuoppa, J. A. Lipponen, T. J. Martikainen, H. Jäntti, T. T. Rissanen, I. Kolk, M. Castrén, J. Halonen, M. P. Tarvainen, and J. E. K. Hartikainen. Wrist band photoplethysmography in detection of individual pulses in atrial fibrillation and algorithm-based detection of atrial fibrillation. *EP Europace*, pages 1031–1038, 2019.

[154] Jeremiah Wasserlauf, Cindy You, Ruchi Patel, Alexander Valys, David Albert, and Rod Passman. Smartwatch performance for the detection and quantification of atrial fibrillation. *Circulation: Arrhythmia and Electrophysiology*, 12(6):1–9, 2019.

[155] Jarkko Harju, Adrian Tarniceriu, Jakub Parak, Antti Vehkaoja, Arvi Ylihankala, and Ilkka Korhonen. Monitoring of heart rate and inter-beat intervals with wrist plethysmography in patients with atrial fibrillation. *Physiological Measurement*, 39, 2018.

[156] Frederik H. Verbrugge, Tine Proesmans, Johan Vijgen, Wilfried Mullens, Maximo Rivero-Ayerza, Hugo Van Herendael, Pieter Vandervoort, and Dieter Nuyens. Atrial fibrillation screening with photo-plethysmography through a smartphone camera. *EP Europace*, pages 1–9, 2019.

[157] Deirdre A. Lane, Flemming Skjøth, Gregory Y.H. Lip, Torben B. Larsen, and Dipak Kotecha. Temporal trends in incidence, prevalence, and mortality of atrial fibrillation in primary care. *Journal of the American Heart Association*, 6(5):1–9, 2017.

[158] Ben Freedman et al. Screening for atrial fibrillation. *Circulation*, 135(19):1851–1867, 2017.

[159] Paul Ziegler et al. Detection of previously undiagnosed atrial fibrillation in patients with stroke risk factors and usefulness of continuous monitoring in primary stroke prevention. *American Journal of Cardiology*,

110(9):1309–1314, 2012.

[160] Alex Diamantopoulos et al. Cost-effectiveness of an insertable cardiac monitor to detect atrial fibrillation in patients with cryptogenic stroke. *International Journal of Stroke*, 11(3):302–312, 2016.

[161] Andreas Goette et al. EHRA White Paper: knowledge gaps in arrhythmia management — status 2019. pages 1–28, 2019.

[162] Marco V. Perez et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. *The New England Journal of Medicine*, 381(20):1909–1917, 2019.

[163] Yutao Guo et al. Mobile photoplethysmographic technology to detect atrial fibrillation. *Journal of the American College of Cardiology*, 74(19), 2019.

[164] Zienab Rezaei Yousefi, Jakub Parak, Adrian Tarniceriu, Jarkko Harju, Arvi Yli-Hankala, Ilkka Korhonen, and Antti Vehkaoja. Atrial fibrillation detection from wrist photoplethysmography data using artificial neural networks. In *World Congress on Medical Physics and Biomedical Engineering 2018*, pages 399–404. Springer Singapore, 2018.

[165] Michael V. McConnell, Mintu P. Turakhia, Robert A. Harrington, Abby C. King, and Euan A. Ashley. Mobile health advances in physical activity, fitness, and atrial fibrillation: moving hearts. *Journal of the American College of Cardiology*, 71(23):2691–2701, 2018.

[166] Julian P. J. Halcox et al. Assessment of remote heart rhythm sampling using the AliveCor heart monitor to screen for atrial fibrillation the REHEARSE-AF study. *Circulation*, 136(19):1784–1794, 2017.

[167] Gerald V. Naccarelli, Helen Varker, Jay Lin, and Kathy L. Schulman. Increasing prevalence of atrial fibrillation and flutter in the United States. *The American journal of cardiology*, 104(11):1534–9, 2009.

[168] Christopher C. Cheung, Kenneth G. Gin, and Jason G. Andrade. Watch Out. *JACC: Clinical Electrophysiology*, 5(4):525 LP – 526, apr 2019.

[169] Ary L. Goldberger et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):E215–20, 2000.

[170] Jonathan W. Waks et al. Intermittent anticoagulation guided by continuous atrial fibrillation burden monitoring using dual-chamber pacemakers and implantable cardioverter-defibrillators: Results from the Tailored Anticoagulation for Non-Continuous Atrial Fibrillation (TACTIC-AF). *Heart Rhythm*, 15(11):1601–1607, 2018.

[171] Ronald M. Aarts. Determination system for determining a heart failure risk. US Patent, 2019. US20190083043.

[172] Taya V. Glotzer et al. The relationship between daily atrial tachyarrhythmia burden from implantable device diagnostics and stroke risk the trends study. *Circulation: Arrhythmia and Electrophysiology*, 2(5):474–480, 2009.

[173] Isabelle C. van Gelder et al. Duration of device-detected subclinical atrial fibrillation and occurrence of stroke in ASSERT. *European Heart Journal*, 38(17):1339–1344, 2017.

LIST OF AUTHOR'S PUBLICATIONS

JOURNAL ARTICLES

[J-1] **Eerikäinen, L.M.**, Bonomi, A.G., Dekker, L.R.C., Vullings, R., Aarts, R.M., "Atrial fibrillation monitoring with wrist-worn photoplethysmography-based wearables: State-of-the-art review", *Cardiovascular Digital Health Journal*, 2020.

[J-2] Regis, M., **Eerikäinen, L.M.**, Haakma, R., van den Heuvel, E.R., Serra, P., "Robust probabilistic modelling of photoplethysmography signals with application to detection and classification of premature contractions", Submitted.

[J-3] **Eerikäinen, L.M.**, Bonomi, A.G., Schipper, F., Dekker, L.R.C., de Morree, H.M., Vullings, R., Aarts, R.M., "Detecting atrial fibrillation and atrial flutter in daily life using photoplethysmography data", *IEEE Journal of Biomedical and Health Informatics*, 2020.

[J-4] Papini, G.B., Fonseca, P., **Eerikäinen, L.M.**, Overeem, S., Bergmans, J.W.M., Vullings, R., "Sinus or not: a new beat detection algorithm based on a pulse morphology quality index to extract normal sinus rhythm beats from wrist-worn photoplethysmography recordings", *Physiological Measurement*, 2018.

[J-5] **Eerikäinen, L.M.**, Bonomi, A.G., Schipper, F., Dekker, L.R.C., Vullings, R., de Morree, H.M., Aarts, R.M., "Comparison between electrocardiogram and photoplethysmogram derived features for atrial fibrillation detection in free-living conditions", *Physiological Measurement*, 2018.

[J-6] Bonomi, A.G., Schipper, F., **Eerikäinen, L.M.**, Margarito, M., van Dinther, R., Muesch, G. de Morree, H.M., Aarts, R.M., Babaeizadeh, S., McManus, D.D., Dekker, L.R.C., "Atrial fibrillation detection using a novel cardiac ambulatory monitor based on photo-plethysmography at the wrist", *Journal of American Heart Association*, 2018.

[J-7] **Eerikäinen, L.M.**, Vanschoren, J., Rooijakkers, M., Vullings, R., Aarts, R.M., "Reduction of false arrhythmia alarms using signal selection and machine learning", *Physiological Measurement*, 2016.

INTERNATIONAL CONFERENCE PROCEEDINGS

[C-1] **Eerikäinen, L.M.**, Bonomi, A.G., Schipper, F., Dekker, L.R.C., Vullings, R., de Morree, H.M., Aarts, R.M., "How accurately can we detect atrial fibrillation using photoplethysmography data measured in daily life?", *Computing in Cardiology*, Singapore, September 2019.

[C-2] **Eerikäinen, L.M.**, Bonomi, A.G., Dekker, L.R.C., Schipper, F., Vullings, R., Aarts, R.M., "Force-interval relationships of the heart measured with photoplethysmography during atrial fibrillation", *Computing in Cardiology*, Maastricht, The Netherlands, September 2018.

[C-3] **Eerikäinen, L.M.**, Dekker, L.R.C., Bonomi, A.G., Vullings, R., Schipper, F., Margarito, J., de Morree, H.M., Aarts, R.M., "Validating features for atrial fibrillation detection from photoplethysmogram under hospital and free-living conditions", *Computing in Cardiology*, Rennes, France, September 2017.

[C-4] Bonomi, A.G., **Eerikäinen, L.M.**, Schipper, F., Aarts, R.M., de Morree, H.M., Dekker, L.R.C., "Detecting episodes of brady- and tachycardia using photo-plethysmography at the wrist in free-living conditions", *Computing in Cardiology*, Rennes, France, September 2017.

[C-5] Bonomi, A.G., Schipper, F., **Eerikäinen, L.M.**, Margarito, J., Aarts, R.M., Babaeizadeh, S., de Morree, H.M., Dekker, L.R.C., "Atrial fibrillation detection using photo-plethysmography and acceleration data at the wrist", *Computing in Cardiology*, Vancouver, Canada, September 2016.

[C-6] **Eerikäinen, L.M.**, Vanschoren, J., Rooijakkers, M., Vullings, R., Aarts, R.M., "Decreasing the false alarm rate of arrhythmias in intensive care using a machine learning approach", *Computing in Cardiology*, Nice, France, September 2015.

REGIONAL CONFERENCES

[R-1] **Eerikäinen, L.M.**, Bonomi, A.G., Schipper, F., Dekker, L.R.C., de Morree, H.M., Vullings, R., Aarts, R.M., "Atrial fibrillation and atrial flutter detection from wrist-worn photoplethysmography data", *18th National Day on Biomedical Engineering*, Brussels, November 2019.

[R-2] **Eerikäinen, L.M.**, Bonomi, A.G., Dekker, L.R.C., Schipper, F., Vullings, R., Aarts, R.M., "Photoplethysmography – Unobtrusive means to measure force-interval relationships of the heart during atrial fibrillation?", *7th Dutch Bio-Medical Engineering Conference*, Egmond aan Zee, January 2019.

[R-3] **Eerikäinen, L.M.**, Dekker, L.R.C., Vullings, R., Bonomi, A.G., Aarts, R.M., "Post-extrasystolic potentiation detection with photoplethysmography from the wrist – A preliminary study", *6th Dutch Bio-Medical Engineering Conference*, Egmond aan Zee, January 2017.

OTHER PUBLICATIONS

[O-1] **Eerikäinen, L.M.**, Bonomi, A.G., Dekker, L.R.C., Aarts, R.M., "Atrial fibrillation episodes detected using photoplethysmography – do we know which are true?" [Letter to the Editor], *Journal of American College of Cardiology*, 2020.

ACKNOWLEDGEMENTS

When looking back the five years that brought me to this point, I realize how many people have been part of this PhD experience. I'd like to dedicate these last pages to thank the people involved.

First and foremost, I want to express my gratitude to my first promotor prof.dr. Ronald Aarts for the guidance, encouragement, and inspiration. Thank you for believing in my abilities and giving me a push to aim higher when I needed that. Without that I would have not taken up some challenges that lead to important opportunities. Next to work related discussions, I enjoyed our talks about many shared interests, such as arts and culture, from which I also could learn a lot.

Thank you my second promotor prof.dr. Lukas Dekker for your supervision, support, and clinical expertise that has been extremely valuable to me. I very much enjoyed our discussions about clinical aspects and value and what I could learn from them. Despite your busy schedule with all the clinical work and academic responsibilities, you've always responded very fast if there was something I needed.

Next I'd like to thank my co-promotor dr.ir. Rik Vullings. Thank you for all the guidance and valuable advice during these years. In addition to everything what I've learned under your supervision about technical work and writing articles, I got to gain experience in international collaboration in the projects in which we were together. Thank you for trusting me responsibilities and giving me an opportunity to grow (and to have some additional travels).

Thank you, dr. Alberto Bonomi, for supervising me at Philips Research. I'm grateful for all the time you've invested in my guidance and discussing ideas. You've helped me to look at things from new perspectives and keep continuing when there were challenges. I've learned a lot from you and I'm happy that I can continue working with you as colleagues.

I'd like to thank the committee members prof.dr.ing. Steffen Leonhardt, prof.dr. Isabelle van Gelder, and prof.dr.ir. Gerard de Haan for their time and effort to assess this thesis, and for the understanding and flexibility regarding changes in planning during these new, uncertain, and challenging times.

I'm grateful to prof.dr.ir. Jan Bergmans that I had the opportunity to pursue PhD in the SPS group. It has been a great environment to work and grow.

Thank you Jan for all the support and trust during these years. I've appreciated how easy it has been to approach you when I needed advice, and the interest you've had in my development and in which direction I want to go to.

Prof.dr. Massimo Mischi, dr. Chiara Rabotti, and Rik, thank you for creating such a nice working environment and community in the BM/d group. Over these years, many people in the group have become much more than just colleagues. Thank you Marina for being a friend whom I can always trust. Rogier, thank you for always finding the right words of encouragement to say. Simona, I'm very lucky to have you not only as a colleague and a friend, but also as a neighbor (with Edu) just few floors up. Ruud, thank you for the great company and somehow always finding time in your busy agenda to catch up. Paul, the last period of the PhD was a bit easier when having someone to share the same steps with, thank you for that. Thank you Nienke for your contagious enthusiasm and for introducing me to, e.g. Efteling and popular Dutch TV programs. Elisabetta, I'm happy that after 6 years we ended up in the same university again, your positivity always improves my days. Eleni, thank you for sharing the amazing trip to Vietnam from which I have great memories. Federica, my neighbor in the office for a long time, thank you for all the nice chats. Thank you Gabriele for sharing the teaching responsibilities and for the fun company. And there are many more with whom I've had the pleasure to spend daily lunches, coffees, and fun BM/d events. Thank you Aline, Anastasiia, Anouk, Bas, Ben, Beatrijs (I really liked and appreciated our 'Nederlandse kwart'), Celine, Cristina, Eleonora, Guy, Iris, Laura, Lieke, Lin, Liz, Luuk (thank you for being so helpful and a great teaching assistant), Maretha, Michiel (thank you for sharing your algorithm with a starting PhD student), Myrthe, Nish, Peiran, Salvatore, Stefan, Tom, Xufei, and Yizhou.

In addition to the people in the BM/d group, there are more people in the SPS group that I'd like to acknowledge. Carla, Marieke, Sabine, and Judith, thank you for all the help with different forms, budget numbers, and administrative tasks during the PhD, and for encouraging me to speak Dutch with you even if it was sometimes slow and required some patience. Marco, Thijs, Amir, Ivan, José, thank you for the company in the office and at lunches during the first year of my project. Special thanks to Amir also for the collaboration when I was working with my first paper. Moreover, thank you dr. Pierre Woerlee for everything I've learned when assisting with the course.

I'm happy I could spend part of my time at Philips Research. Thank you Jörg Habetha and Johannes Nürnberg for hosting me in your departments. I'd like to thank dr.ir. Reinder Haakma for the guidance and the opportunity for being part of the team. Thank you also to all the wonderful colleagues that I had the pleasure to work and/or share the office with: Helma, Jenny, Fons, Leonie, Jos, Koen, Pedro, Lieke, and Marija. Last but not least, Giulio, thank you for

helping me when needed and sharing your wise thoughts and experience about doing a PhD.

This PhD experience would not have been the same without the Data Science Flagship and the fellow PhD students there. Thank you Alberto, Alok, Bart, Chao, Felipe, Heleen, Humberto, Marie, Mark, Marta, Mustafa, Niek, and Saskia for sharing the challenges and joys of being part of the Flagship. With warmth I think of our dinners, (milestone) drinks, barbeques, city trips, and also the workshops and courses we followed together which were much more fun with you friends. I'd also like to thank some of the PhD supervisors in the Flagship that I had the opportunity to interact with. Thank you prof.dr. Edwin van den Heuvel for the many interesting discussions we had and for challenging my way of thinking. Thank you dr. Joaquin Vanschoren for sharing your machine learning expertise and helping with the CinC challenge. Also thank you prof.dr. Uzay Kaymak, prof.dr. Volkher Scharnhorst, dr. Arjen-Kars Boer, and dr. Anna Wilbik for sharing your ideas and questions in the joint meetings.

Large part of this work would have not been possible without the measurements at Catharina Hospital. I'm extremely grateful to dr. Loes Verborg, dr. Nicole Sturkenboom, dr. Rob Eerdekens, the personnel at the Holter and cardioversion departments, and all the patients who made the trials possible. Thank you Maarten, Esther, Ingrid, Wilma, and Harald for all the help and support in setting up the next steps.

In the eWatch project, I had the opportunity to be part of an European project and get to know project partners not only in the Netherlands, but also in Belgium and in Turkey. I'd like to thank Ernst, Francesco, Reinder, Ruben, and Zhang for the collaboration and travel company, and all the partners for the nice project meetings and dinners.

I'd also like to thank my friends outside the university. You have been essential in taking time off from the PhD. Irena, I admire how devoted you are to your family and friends, and I'm lucky to be one of them. Thank you for being there for me. Thank you Elina and Peter for barbeques, movie nights, relaxing times in your garden, and always being easy to reach. When missing Finland, Merimieskirkko in Rotterdam has been a great place to ease this with its sauna, library, Finnish shop, and lovely people. I'd like to especially thank Riitta and other members of the book club for the nice gatherings and book recommendations. Without Merimieskirkko, I also would have not met Ilona who became a dear friend. Thank you Ilona for all the engaging conversations that shaped my perspective, and for the visits to new places, especially the ones surrounded by nature, when you still lived in the Netherlands.

I've always enjoyed learning new languages and while living in the Netherlands of course I wanted to study also Dutch. Petri, ik ben blij dat ik Nederlands met jou heb geleerd maar vooral dat ik meer dan een leraar heb gevonden. Ik wil je en Jan bedanken voor de gezellige avonden bij jullie thuis en voor de hulp en steun die ik van jullie heb gekregen.

Ai miei amici, Alice, Alice, Alessandro, Anthony, Eugenia, Francesco, Luca, Marta, Risilda, Silvia e Valentina, grazie di cuore di esserci. Sono stata molto fortunata di avervi conosciuto a Milano e avervi nella mia vita anche se sono anni che viviamo nei paesi diversi. Ogni volta che possiamo vederci mi sento di essere con la famiglia e mi dispiace che per la causa della distanza le volte sono troppo poche.

Kiitos Annalle kannen sunnittelusta ja kärsivällisyydestä kuunnella ja toteuttaa kaikki muutosehdotukset. Anneli, Gitta, Heini, Iiris, Jasmin, Kata, Maiju, Mia, Mikko, Outi, Pamir, Tommi, ja Valtteri, vaikka joskus kuluu pitkäkin aika näkemättä teitä, ei se tapaamishetkellä koskaan tunnu siltä. On ihanaa, että olen aina voinut palata Suomessa vanhojen ystävien seuraan, purkaa huolet ja jakaa ilot, ja tuntea olevani kotona. Se on ollut tärkeä voimavara näinä vuosina ja auttanut jaksamisessa saattaa tämä työ loppuun. Kiitos ystävät, että olette aina tukena.

Lopuksi kiitos perheelleni kaikesta tuesta ja rohkaisusta, jota ilman en olisi päässyt näin pitkälle. Laura, ilman esimerkkiäsi minulla ei olisi ollut rohkeutta lähteä vaihto-oppilaaksi 17-vuotiaana. Uskon, että ilman sitä kokemusta en olisi uskaltanut myöhemmin tarttua moneen muuhun jännittävään tilaisuuteen, enkä olisi nyt kirjoittamassa näitä väitöskirjan kiitossanoja. Äiti ja isi, kiitos, että olette valaneet uskon siihen, että pystyn saavuttamaan sen mihin ryhdyn. Kiitos myös, että olette rohkaisseet tekemään sellaisia valintoja, jotka tuntuvat oikeilta minulle, vaikka se joskus tarkoittaisi lähtemistä maailmalle ja kauemmas teistä.

– Linda

About the author

Linda Eerikäinen was born in 1987 in Helsinki, Finland. In 2010, she received her B.Sc. degree in Biomedical Engineering from Aalto University, Finland. After completing her bachelor studies, she pursued a double degree program in Biomedical Engineering obtaining M.Sc. degrees both from Politecnico di Milano, Italy, in 2014, and from Aalto University with distinction in 2015. She conducted her graduation project at GE Healthcare Finland, working on the topic of automatic seizure detection from electroencephalography in intensive care. In February 2015, she started a PhD project on cardiac arrhythmia monitoring in the Signal Processing Systems group at Eindhoven University of Technology (TU/e), of which the results are presented in this thesis. The project was a collaboration project between TU/e, Philips Research, and Catharina Hospital Eindhoven. Since January 2020, she works as a scientist in the department of Patient Care and Measurements at Philips Research in Eindhoven.

APPENDIX

STUDIES ABOUT ATRIAL FIBRILLATION DETECTION WITH WRIST-WORN DEVICES MEASURING PHOTOPLETHYSMOGRAPHY

Author	**Number of subjects**	**Rhythm type**	**Age (years)**	**Device**	**Duration of recording**	**Detected episode duration**	**Measure-ment setting**	**Reference data**	**Detection performance**	**Measure-ment coverage**
Lemay et al.[84]	20	AF and SR	Not reported	Proprietary device	2213 10s-epochs, one epoch per detected beat	Single beat	Hospital before catheter ablation	12-lead ECG annotated by epoch	Sens: 99.38% Spec: 56.64% Acc: 93.76%	Not explicitly reported
Nemati et al.[97]	46 (36+10)	AF and other rhythms	18-89 Simband	Samsung	3.5-8.5 min per subject	Each record	Hospital	Single-lead ECG	Sens: 97% Spec: 94% Acc: 95%	Not explicitly reported
Corino et al.[85]	70	AF, other arrhythmia, and SR	AF: 76 ± 9 (58-89) SR: 40 ± 17 (27-75) Other: 65 ± 15 (48-92)	Empatica E4	10 min per subject	2 min	Hospital	Not reported	Sens: 75.4% Spec: 96.3%	20% (the best 2 min segment selected

Shashikumar et al.[98]	98	AF and other rhythms	18-89	Samsung Simband	Approx. 5 min per subject	30 s	Hospital	Single-lead ECG	Sens: 89% Spec: 96% Acc: 91.8%	Not explicitly reported
Aliamiri et al.[114]	19	AF and SR	Not reported	Samsung gear	1443 30 s epochs device	30 s	Not reported	ECG	Acc: 98.18%	50%
Bonomi et al.[112]	20 + 40 (+120)	AF and other rhythms	73.1 $\pm$ 11.6 (45-87) and 67.4 $\pm$ 12.1 (34-87)	Data logger with Philips C3M module	1.9–2.8 h and 21.9–39.2 h per subject	1 min	Hospital ECV and daily life	Single-lead ECG and 12-lead ECG	Sens: 96% and 93% Spec: 100% and 100% Acc: 98% and 97%	47% and 48%
Eerikäinen et al.[113]	30	AF and other rhythms	AF: 69 $\pm$ 11 (43-79) Other: 67 $\pm$ 13 (34-87)	Data logger with Philips C3M module	24 h per subject	2 min	Daily life	12-lead ECG	Sens: 98.4% Spec: 98.0% Acc: 98.1%	25%
Fallet et al.[118]	17	AF, ventricular arrhythmias, and SR	57 $\pm$ 13	Proprietary device	3056 10s-epochs	10 s	Hospital during ablation procedure	12-lead ECG	Sens: 96.2% Spec: 92.8% Acc: 95.0%	71%

Gotlibovych et al.[109]	53	AF and SR	37-85	Prototype fitness tracker device	180 h data (36 h AF)	0.8 s	Hospital ECV and during sleep outside hospital	ECG	Sens: 99.8% Spec: 99.9%	100%
Harju et al.[155]	30	AF and SR	AF: 74.8 $\pm$ 8.3 SR: 67.5 $\pm$ 10.7	PulseOn optical heart rate monitor	1-2 h per subject	1 min	Hospital during post-operative care	ECG	Sens: 99.0% Spec: 92.96%	Not explicitly reported
Hochstadt et al.[150]	20	AF and SR	74.1 $\pm$ 8.7	CardiacSense	30 min per subject	30-250 beats	Hospital or laboratory	ECG	Sens: 100% Spec: 93.1%	Not explicitly reported
Shashikumar et al.[151]	97	AF and other rhythms	18-89	Samsung Simband	5-10 min per subject	30 s	Hospital	ECG	Spec: 100% Acc: 95%	Not explicitly reported
Tison et al.[152]	51 + 1617	AF and other rhythms	66.1 $\pm$ 10.7 and AF: 55.7 $\pm$ 14.2 Other: 41.4 $\pm$ 11.9	Apple Watch	40 min and 18.5 million 5 s measure-ments	5 s	Hospital CV and daily life	12-lead and self-reported persistent AF	Sens: 98% and 67.7% Spec: 90.2% and 67.6%	Not reported

Yousefi et al.[164]	30	AF and SR	AF: 74.8 ± 8.3 SR: 67.5 ± 10.7	PulseOn optical heart rate monitor	1.5 h per subject	30 IBIs	Hospital during post-operative care	ECG	Sens: 99.2% ± 1.3% Spec: 99.54 ± 0.64%	73.3%
Dörr et al.[120]	672	AF and SR	AF: 77.4 ± 9.1 SR: 75.6 ± 9.8	Samsung Gear Fit 2	5 min per subject	1 min	University center	iECG	Sens: 93.7% Spec: 98.2% Acc: 96.1%	78%
Fan et al.[111]	108	AF and SR	AF: 66.56 ± 13.17 SR: 58 ± 14.78	Huawei Band 2	3 min per subject	1 min	Hospital	12-lead ECG	Sens: 95.36% Spec: 99.70%	95%
Kashiwa et al.[117]	20 + 116	AF and other rhythms	62.7 ± 10.9	Pulse wave monitor, Seiko Epson	368 h	6 min	Hospital during electro-physiological study	2-channel ECG	Sens: 84.1% Spec: 97.7%	100%
Shen et al.[135]	82	AF and other rhythms	Not reported	Not reported	8 h and 3 h per subject	30 s	Daily life	ECG patch	Only AUC reported (94.8%)	100%

Solosenko et al.[116]	34	AF and other rhythms	AF: 72.9 ± 8.9 Other: 67.5 ± 10	Wrist-worn device developed at Kaunas University of Technology	21 h per subject	One pulse	Cardiac rehabilitation	Single-lead ECG	Sens: 72.2% Spec: 99.7% Acc: 87.0%	89.2%
Väliaho et al.[153]	213	AF and SR	AF: 72.0 ± 14.3 SR: 54.5 ± 18.6	Empatica E4	5 min per subject	2 min	Hospital, emergency care or cardiac ward	3-lead ECG	Sens: 96.2% Spec: 98.1%	Not reported
Wasserlauf et al.[154]	24	AF and other rhythms	72.1 ± 7.2	Apple Watch	110 ± 35.7 days	1 h	Daily life	ICM	Sens: 97.5%	Not reported
Yang et al.[110]	11	AF and SR	63 ± 12	Prototype device	0.75-12 h	10 s	Hospital ECV	ECG	Acc: 92.71%	Not reported

Acc = accuracy, AF = atrial fibrillation, ECV = elective cardioversion, ICM = insertable cardiac monitor, IBI = inter-beat interval, Sens = sensitivity, SR = sinus rhythm, Spec = specificity