ZEINAB REZAEI YOUSEFI
ATRIAL FIBRILLATION DETECTION FROM PHOTOPLETHYSMOGRAPHY DATA USING ARTIFICIAL NEURAL NETWORKS
Master's thesis

Examiner: Assist. Prof. Antti Vehkaoja

The examiner and topic of the thesis were approved on 31 January 2018

ABSTRACT

Atrial fibrillation (AF) is one of the most common types of cardiac arrhythmia- especially in elderly and hypertensive patients, leading to increased risk of heart failure and stroke. Therefore, early screening and diagnosis can reduce the AF impact. The development of photoplethysmography (PPG) technology has enabled comfortable and unobtrusive physiological monitoring of heart rate with a wrist-worn device. It is important to examine the possibility of using PPG signal to diagnose AF in real-world situations.

There are several recent studies classifying cardiac arrhythmias with artificial neural networks (ANN) based on RR intervals derived from ECG, but no one has evaluated ANN approach for wrist PPG data. The aim of this MSc thesis is to present an ANN-based classifier to detect AF episodes from PPG data. The used classifier is multilayer perceptron (MLP) that utilizes backpropagation for learning. This classifier is able to distinguish between AF and non-AF rhythms. The input feature of the ANN is based on the information obtained from an interbeat interval (IBI) sequence of 30 consecutive PPG pulses.

The PPG dataset was acquired with PulseOn (PO) wearable optical heart rate monitoring device and the recordings were performed in the post-anesthesia care unit of Tampere University Hospital. The study was approved by the local ethical committee. The guidelines of the Declaration of Helsinki were followed. In total 30 patients with multiple comorbidities were monitored during routine postoperative treatment. 15 subjects had sinus rhythm (SR) and 15 had AF during the recording. The average duration of each recording was 1.5 hours. The monitoring included standard ECG as a reference and a wrist-worn PPG monitor with green and infrared light sources.

As IBIs extracted from the PPG signals are highly sensitive to motion artefacts, IBI reliability was automatically evaluated using PPG waveform and acceleration signals before AF detection. Based on the achieved results, the ANN algorithm demonstrated excellent performance at recognizing AF from SR, using wrist PPG data.

PREFACE

This thesis work was carried out at the Faculty of Biomedical Science and Engineering at Tampere University of technology during the year 2017-2018 and was conducted in the PulseOn Oy as a part of a development project which was cooperated with Tampere University Hospital.

I wish to express my sincere gratitude to my supervisor at TUT Assistant Professor Antti Vehkaoja and at PulseOn Oy Doctoral student Jakub Parak for their guidance, encouragement and patience during the time of finalizing this Master of Science thesis. I would also like to express my deep appreciation to Professor Ilkka Korhonen for giving me this opportunity to cooperate in the algorithms team of PulseOn and be a part of this very interesting project and for all his insightful advice. I would like also to thank all the colleagues of PulseOn Oy specially Dr. Adrian Tarniceriu for providing an inspiring and communicative environment and the collaborators in Tampere University Hospital for providing measurement data and expert advice.

Finally, my special thanks are reserved to my family who have been there for me whenever their support was needed throughout all my life.

In Tampere, Finland, on 22 May 2018
Zeinab Rezaei Yousefi

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ACC	Accuracy
AF	Atrial Fibrillation
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under the Curve ROC
AV	Atrioventricular
AVNN	Average of NN intervals
bpm	Beats per minute
ECG	Electrocardiogram
FN	False Negative
FP	False Positive
HMM	Hidden Markov Model
IBI	Interbeat interval
LDA	Linear Discriminant Analysis
LED	Light Emitting Diode
MLP	Multilayer Perceptron
OHR	Optical Heart Rate
pNNx	Proportion of pairs of successive NNs that differ by more than x ms
PAC	Premature Atrial Contraction
PO	PulseOn
PVC	Premature Ventricular Contraction
QDA	Quadratic Discriminant Analysis
RMSSD	Root mean square standard deviation
ROC	Receiver Operating Characteristic
RRI/RR interval	R-peak to R-peak interval (in ECG waveform)
SA	Sinoatrial
SamEn	Sample Entropy
SEN	Sensitivity
SDNN	Standard deviation of the NN intervals
SGD	Stochastic Gradient Descent
ShEn	Shannon entropy
SPE	Specificity
SQE	Signal Quality Estimation
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

1. INTRODUCTION

1.1 Motivation

Design and development of wearable sensing technologies have enabled large-scale acquisition of health data in real life and long-term monitoring of physiological status of the users. Moreover, recent advances in machine learning have improved the ability of machines and software to classify, quantify and identify patterns in biomedical signals. By increasing the costs of healthcare and the aging of the world population, there is a need to monitor the health status of patients out of the hospital and during their daily activities. These wearable sensing technologies are becoming more popular and affordable and facilitate inexpensive unobtrusive solutions for continuous health and activity status monitoring. Nowadays, these technologies have drawn lots of attention from the industry and research communities to produce devices and algorithms and to evaluate their performance for ambulatory health monitoring.

One application area of such wearable sensing devices is cardiac monitoring and specifically automatic arrhythmia detection. Atrial fibrillation is the most frequently occurring type of cardiac arrhythmia that is affecting the life of over 10% of the population above 75 years old. The prevalence of AF in adult population is increasing with the aging of the world-wide population and according to recent studies almost one in four middle-aged adults in the US and Europe will develop AF [7–9]. It is therefore crucial to develop new algorithms that will aid in the analysis of heart rhythm and detection of atrial fibrillation.

The electrical excitation of the healthy heart, starts at the sinus node and following that spreads through the atrium and ventricles of the heart [10]. The regular rhythm of a normal heart is approximately 60-100 bpm that is called sinus rhythm. In contrast, during atrial fibrillation abnormal electrical signals travel through the atria. These irregular impulses originate from the randomly contraction of the muscle fibers of the heart's upper chambers rather than the sinus node and lead to a semi constantly circulating stimulation. Therefore, the time intervals between the two heartbeats become disorganized and irregular.

In spite of the fact that the AF mechanism is caused by electrical disturbances in the heart, this pathology can also affect the photoplethysmography signal, since it produces an irregular rhythm and therefore irregular flow in the blood vessels [11]. The photoplethysmogram offers an alternative method to ECG for heart rate - often called pulse rate in case of PPG –acquisition, which is convenient regarding to its recording and allows for self-monitoring, thus it has the potential to early AF diagnosis and to reduce the need of clinical staff. The advantage of PPG signal is that it can be obtained non-invasively and the sensing device is easy-to-set up and economically efficient.

Automatic detection of atrial fibrillation has been studied by many researchers and among various methods conducted in this field the analysis of the heart rate dynamics is shown to be a reliable method to enhance rhythm monitoring and distinguish between AF and SR. Several machine learning approaches have been proposed in the literature to assess the different PPG features extracted from pulse rate dynamics for accurate detection of AF [12–15]. A large number of studies have evaluated the use of smart-phones and smart-watches in medical practice as they have become prevalent [11, 16–20]. A major challenge is the early screening and detection of AF, because in the early stages or for some subjects, AF can be asymptomatic. Furthermore, AF in the early stages occurs irregularly with unpredictable times of appearance and durations. This asymptomatic and irregular or paroxysmal nature of AF motivates the development of solutions that may help in automatic detection of AF at the early stages. Although PPG waveform contains valuable information about cardiovascular health, these pulses can be easily corrupted by different sources of artefact, especially in ambulatory monitoring. Since the reliability of information is highly important when using the data in clinical decision making, a fast and accurate quality estimation of the PPG signal is needed to recognize corrupted data from valid data before any diagnosis and prediction.

According to a study in 2013 [7], the estimated number of people suffering from AF globally was 33.5 millions, and the findings prove that there is a progressive increase in overall incidence and prevalence of AF-associated morbidity and mortality. Therefore, AF has a significant effect on the quality of life of a large number of people and there is a high demand for developing methods for an accurate systematic AF detection. Automatic detection of AF using PPG recording devices is an ongoing research and there is still room for improvement.

1.2   Problem statement and objectives

This thesis work was done for the PulseOn Oy as a part of the development of an AF detection algorithm using the PPG signals acquired from the PulseOn heart rate monitoring device and as the continuation to the project of the PPG signal quality estimation.

In spite of the fact that PPG signal acquisition is more convenient than the current ECG methods, as stated earlier it is more susceptible to be corrupted by different kinds of artefact especially in the long-term monitoring. These artefacts may cause variations in the interbeat interval (IBI) time series derived from the PPG signal. As a result they may be confused by irregularities due to the AF episodes. Therefore, a main challenge of AF detection problem using PPG signals is how to address the difference between corrupted IBIs (due to artefact) and irregular IBIs (due to AF).

The limitation of the existing AF detection methods is that either they do not consider the reliability of the IBIs or they only observe the motion artefacts by measuring the amount of movement and discard segments corresponding to the high motion level. But the proposed

approach in this thesis is that prior to perform the AF detection, the PPG signals are subject to a processing procedure for quality estimation and besides motion the quality of the whole waveform of each PPG pulse is assessed.

Data of this thesis was acquired from patients with either continuous SR or continuous AF during the whole time of the recording and the goal is an automatic classification of pulses into two target classes: AF and SR. A multilayer perceptron with backpropagation learning method is implemented to carry out the classification task. The input features extracted from the PPG signals are time domain parameters, representing the pulse rate characteristics. The focus of this thesis report is on the applied machine learning method, extracted features for AF detection and the impact of signal quality estimation on pulse classification and due to the confidentiality policy of the PulseOn Oy the details about signal quality estimation algorithm are kept sealed.

In summary, the key objectives of this thesis are itemized as follows:

- To develop and to evaluate a machine learning method based on ANN to detect AF pulses using features derived from PPG analysis with considering the quality of the PPG pulses,

- To investigate the impact of the artefact detection by comparing the performances of the ANN method using PPG derived features before and after the PPG signal quality estimation with the performance of the ANN method using the ECG derived features as a gold standard,

- To examine the ability of the proposed method for detection of those AF events that occur episodically and last for a short period of time,

- To compare the performance of different classifiers for AF detection using PPG derived features.

1.3   List of publications

The following publications resulted from the work conducted during this thesis:

- Adrian Tarniceriu, Jarkko Harju, **Zeinab R. Yousefi**, Antti Vehkaoja, Jakub Parak, Arvi Yli-Hankala, Ilkka Korhonen, "The Accuracy of Atrial Fibrillation Detection from Wrist Photoplethysmography. A Study on Post-Operative Patients", 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, July 17-21, 2018, Honolulu, HI, USA.

- **Zeinab R. Yousefi**, Jakub Parak, Adrian Tarniceriu, Jarkko Harju, Arvi Yli-Hankala, Ilkka Korhonen, Antti Vehkaoja, "Atrial fibrillation detection from wrist photoplethys-mography data using artificial neural networks", World Congress on Medical Physics and Biomedical Engineering, June 2018, Prague, Czech Republic.

1.4   Thesis structure

This thesis is divided into five chapters, organized as follows. Chapter 2 provides theoretical background and is divided into three sections. First, a brief description of clinical background about the physiology of the heart, cardiac arrhythmia specifically AF and an overview of the photoplethysmography is presented. Afterwards, the chapter shifts into an overview of related works that have been done on the atrial fibrillation detection using PPG signal. Then technical background is presented by describing the multilayer perceptron and backpropagation and popular features used for AF detection. In Chapter 3, the data acquisition process and details of the implemented classifier are explained. Chapter 4 represents the obtained results and discussion about them. Finally, Chapter 5 contains the main conclusions of this work and subjects of future work.

2. THEORETICAL BACKGROUND

2.1 Clinical background

This section presents a brief explanation of the physiology of the heart and different types of rhythms and applied methods to diagnosis arrhythmias.

2.1.1 Physiology of the heart

Heart is the essential component of the circulatory system which acts as a muscular pump and keeps the blood pumping through the arteries and veins to other organs and tissues of the body [1]. Heart has four hollow chambers or cavities –two upper chambers called atria and two lower ones called ventricles. The frontal diagram showing interior chambers is shown in Figure 2.1.
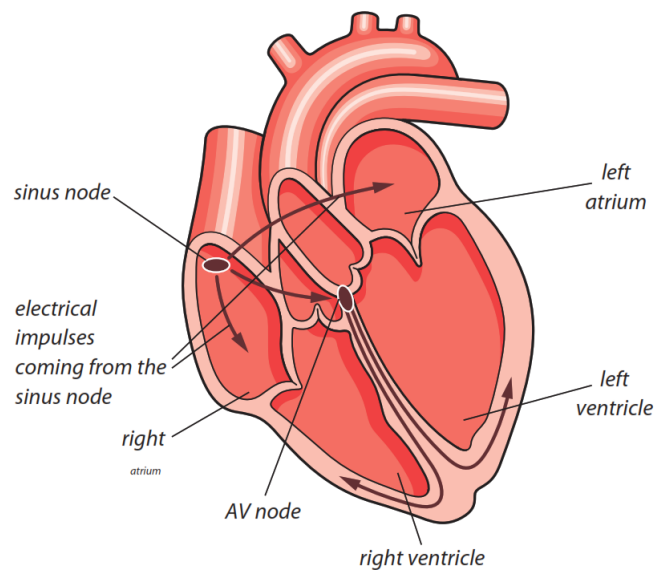
Figure 2.1. *Frontal section showing interior chambers of the heart [image from www.bhf. org.uk]*

Pumping action of the heart is triggered by electrical impulses that are produced by the sinoatrial (SA) node (also called heart's natural pacemaker) that is located on the right atrium of the heart [1]. These electrical impulses spreads over the atria to the Atrioventricular (AV) node causing the contraction of atria and squeezing the blood into two ventricle chambers. The impulses then spread from AV node to the ventricles through an electrical pathway causing the ventricles contraction and squeezing the blood out of the heart to the rest part of the body.

Heart is beating based on the needs of the body. During the rest, organs require less amount of blood and oxygen, therefore, heart rate and blood pressure decrease. While during the activity and training, organs need an increasing amount of blood and oxygen, resulting in a rise in heart rate and blood pressure. The most straightforward and common way of obtaining information about the function and electrical activity of the heart is electrocardiogram (ECG). The electrodes that are placed on the skin surface can detect electrical potentials of the body surface produced by the heart and therefore, visualize in the ECG the electrical changes associated with each heartbeat. Electrical activity of the heart arises from depolarization and repolarization of the heart muscle during a single heartbeat.

One normal cardiac cycle shown in Figure 2.2 includes P-wave, QRS-complex and T-wave that each represent special electrical event [1]. P-wave represents atrial depolarization, QRS-complex indicates the rapid depolarization or contraction of the right and left ventricles. T wave illustrates the re-polarization of the ventricles.
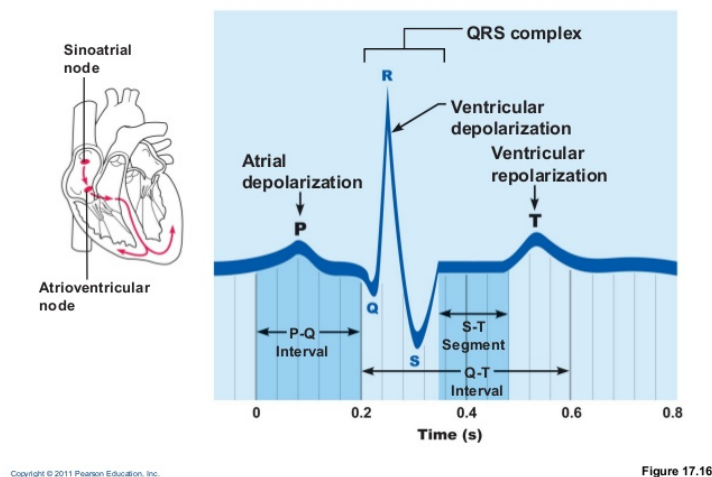
Figure 2.2. *A normal electrocardiogram wave [1][image from www.apsubiology.org]*

In addition to those previously named waves, intervals such as the PR interval, PR segment, QT interval and ST segment are useful to interpret the condition of the heart. Heart rate (HR) is defined as the number of times the heart beats (contracts) over a certain amount of time [21]. The unit is beats per minute (bpm). The heart rate can be deducted from RR interval which is the time interval between the two R-peaks of the QRS-complex. The heart rhythm is the pattern of the heartbeating. It can be described as normal or abnormal, regular or irregular and fast or slow. Cardiac rhythm has the ability to be used as a medical diagnostic parameter to indicate normal and abnormal conditions of the heart. Abnormal rhythms are called arrhythmia (or dysrhythmia) that include slow, fast or irregular cardiac rhythm. Generally, arrhythmia is a disturbance in the regular rhythm of the heartbeat and is caused by abnormalities in impulse formation or conduction. Aging, illness and physiological status can affect the beat-to-beat rhythm.

2.1.2 Sinus rhythm vs. atrial fibrillation

Heart normal rhythm is called sinus rhythm. As already mentioned normal cardiac impulses are started by the natural pacemaker of the heart (SA node) and travel down to the rest of the body. A normal heart is beating in an organized, sequential manner. The rate of sinus rhythm is 60–100 bpm. When the sinoatrial node fails to excite the atrium and initiate an impulse, other parts of atrium may generate an impulse to stimulate the atrium that leads to irregular and abnormal heartbeats. Such abnormal heartbeats may increase or decrease the blood pressure which can lead to paralysis or stroke or even sudden death. There are four main types of cardiac arrhythmia including Premature (extra) beats, Supraventricular or atrial arrhythmias, ventricular arrhythmias and bradyarrhythmias [22].

Premature beats are usually harmless and happen naturally and depending on the occurring point are called premature atrial contraction (PACs) or premature ventricular contraction (PVCs) [22]. The former takes place in the atria (upper chambers of the heart) and the latter occurs in the ventricles (the heart's lower chambers). Premature beats occur earlier than expected time and interrupting the normal heart rhythm.

Supraventricular or atria arrhythmias are irregular heart rates that start in the atria or atrioventricular (AV) node. Three most common atrial arrhythmias consists of [22]:

- Atrial Flutter (changing rate, usually regular with sawtooth pattern)
- Atrial Fibrillation (changing rate, always irregular)
- Supraventricular Tachycardia (rate >150 bpm)

Ventricular arrhythmias begin in the ventricles. Ventricular tachycardia and ventricular fibrillation are two common arrhythmias in this category. Bradyarrhythmias occur when the heart rate is less than 60 bpm.

The arrhythmia that is the case of this study is the Atrial Fibrillation or AFib/AF for short. Electrical impulses that circle uncoordinatedly across the muscles of the atria and cause them to shiver or twitch are known as fibrillation. Subsequently, the lower chambers of the heart do not receive regular impulses and therefore contract without regular pattern. This can lead to uncontrolled and irregular heartbeat. Figure 2.3 shows the electrical conduction and ECG wave during sinus rhythm and atrial fibrillation. During atrial fibrillation, electrical impulses initiate rapidly from multiple sites in both atria, triggering 400 or more atrial contractions per minute. This is felt as an always irregular, sometimes rapid heartbeat. AF is classified based on its temporal pattern of occurrence as [23]:

- **Paroxysmal AF** comes and goes between periods of completely normal heartbeats and typically lasts from at least 1 minute to hours up to 7 days, but it is not there all the time and terminates within 7 days.
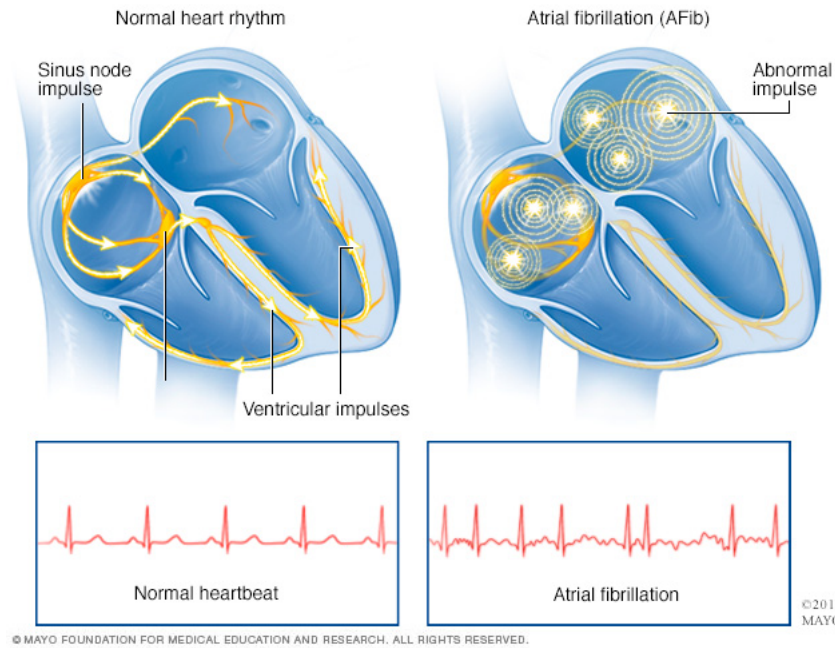
Figure 2.3. *Electrical conduction and ECG wave during normal sinus rhythm and atrial fibrillation [image from www.nhlbi.nih.gov]*

- **Persistent AF** lasts longer than 7 days at a time and does not resolves spontaneously and usually needs treatment with medicines or a procedure called electrical cardioversion to restore sinus rhythm.

- **Long-standing persistent AF** lasts more than 1 year duration. In some references it is included in the category of persistent AF.

- **Permanent AF** is there all the time, and physician or patient decide to abandon attempts to restore sinus rhythm and the heart never returns to a normal sinus rhythm.

Usually, the symptoms of AF can include the following:

- Heart palpitations

- Shortness of breath or being breathless

- Feeling faint

- Fatigue, dizziness and syncope.

However, in some people AF may be asymptomatic and several studies have shown that AF episodes occur without symptoms [24–26]. Therefore, the essential part of both treatment and pre-emptive therapy is to find a reliable way to detect AF and the most optimal ways are those that provide continuous long-term monitoring.

2.1.3 Diagnosis of arrhythmias

There are several tests that doctors suggest when they are suspect that the patient may have symptoms of a kind of arrhythmia. Some diagnostic tests are performed in an electrophysiology lab and need an expert to do the tests and analyse the results, while others consist of devices that patient wears during his/her daily routine to detect less-frequent irregularities. The following experiments can help doctors in checking the cardiac rhythm and can be applied to diagnose an arrhythmia:

- **Electrocardiogram** — ECG/EKG can record the electrical activity of the heart including the timing and duration of each electrical phase associated with each heartbeat [27]. In the conventional standard 12-lead ECG ten electrodes are worn on the chest surface and the limbs while the patient is resting. The test is taken often in a laboratory. ECG has the potential to show features that could suggest someone may have a kind of arrhythmia only if it is happening at the time of the test. ECG experiment is painless and non-invasive.

- **Holter monitor** — It is also called ambulatory ECG monitoring and was introduced by Dr. Norman Holter in 1957 [28]. Holter monitor is a small portable ECG machine that is used typically for 24-48 hours. It is worn around the patient waist and 3-5 electrodes need to be taped on the patient chest skin. The device can record ECG over a 24-hour period – through day and overnight. Routine daily life is allowed while recording the ECG using Holter monitor. Diagram of a 24-hour Holter monitor is shown in Figure 2.4.
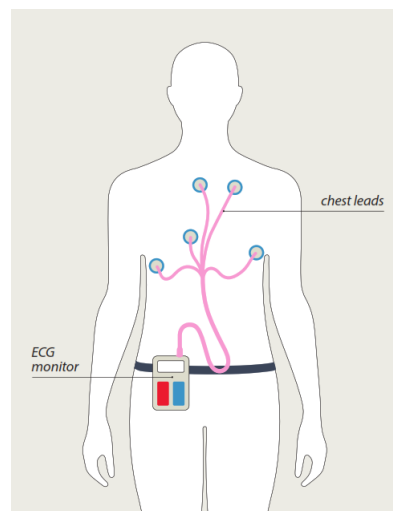
Figure 2.4. *Diagram of a Holter monitor [image from www.bhf.org.uk]*

- **Event monitor** — If the symptoms of an arrhythmia are not happening frequently, doctors may suggest wearing a small recording device that is called a cardiac event recorder. Whenever a patient notices typical symptoms of an arrhythmia, he/she can manually activate the device to record the heart electrical activity for a few minutes by placing the device on the the fingers or on the chest wall using a chest belt [28].

- **Implantable loop recorder (ILR)** — It is a kind of implantable cardiac monitoring device that is used specifically for AF and other arrhythmias monitoring. This implantable device can be inserted under the chest skin and worn for several years. It is a small and slim device that is able to automatically or manually record the heart electrical activity. ILRs can provide the cardiac rhythm that occurred just prior to patient symptoms using loop recording. It means that not only the ongoing current events are recorded, but also the temporary memory remains the recordings for minutes before device activation [28, 29].

As previously stated, the initial diagnosis of cardiovascular diseases including AF has been extensively made utilizing a 12-lead ECG or Holter monitoring. The ECG wave can determine the origin of the heart's electrical activities and therefore the cardiac rhythm, heart rate variability and abnormalities in the cardiac conduction can be monitored using ECG. Electrocardiography is the "gold standard" using either hard wire or telemetry transmission. However, applying the above mentioned ECG monitors still has some drawbacks including [30]:

- 12-lead ECG is done in nurse/doctor office and requires a lot of wires and can make the movement too difficult or sometimes impossible.

- It is possible, for example in paroxysmal AF, the arrhythmia episodes may not occur during the measurement period in the nurse/doctor office.

- Electrodes of 12-lead ECG must have a firm contact to skin and materials of the electrode patches are irritating in continued use or can be allergenic. Allergenic materials can not be used. (For sensitive people, may possibly cause allergic reaction).

- Chest belts are not traditionally used for clinical examinations, but there are some recent attempts to validate their clinical use [31, 32]. However, chest belts need to be moistened and can easily become uncomfortable. If it is not tightened, then belt movement can cause error in the measurement and also irritation.

- Although event recorder device has fewer lifestyle restrictions than 12-lead ECG and Holter monitoring, if it is manually activating device, then it can not be a good solution because asymptomatic events cannot be recorded automatically.

The key features of ambulatory monitoring are capability of long-term monitoring, easy to use and non-invasive nature. Recent technological advancements are capable to tackle many of the above obstacles and facilitated ambulatory heart monitoring during daily activities providing continuous information of heart rate and rhythm from days to weeks. Optical measurement is one of the applied technologies in this area. Measurement can be done easily by sending light to the skin on the wrist with small LEDs and there is no need for wires or gears. The next section describes an optical technique that can be applied effectively to arrhythmia detection.

2.1.4 Photoplethysmography

By developing sensor technology, small wrist heart rate monitoring devices have been popular to help sport physiologists to analyze the body response to different exercises and training. By increasing the interests in easy ways of health monitoring, photoplethysmography has turned to be an alternative method for ECG when estimating heart rate (HR) and heart rate variability (HRV) [2].

Photoplethysmography is an easy-to-set up and economically efficient sensing device that works based on optical principle to evaluate the variations of light propagation inside the tissues during cardiac cycle. The device consists of a light source (red, infrared or green) and a photo-detector (PD) at the skin surface to detect the small changes in light intensity related to blood volume changes in the microvascular bed of tissue [33]. Heart pumps the blood and produces the pulse, that can be felt, for example at the artery in the wrist. This peripheral pulse wave is synchronous with each heartbeat. Consequently, the pulse rate and rhythm can be measured.

If the light is illuminated into the skin, different biological tissue including skin pigments, bone, arterial and venous blood can absorb the light. A biological tissue consists of several different media, each one has its own length and light absorption coefficient. By making this assumption that the illuminated media indicates only a vein or an artery, the blood pressure pulse changes inside the vessels during cardiac cycle and this change results in varying the light absorption and reflection by blood.

One can measure the amount of light that have reflected back out and can find out how much light has been absorbed by blood. The light attenuation is modeled by the Beer-Lambert law (Eq. 2.1). By this law, the light intensity shows an exponentially decay as a function of length of medium (l) the light passes through and light absorption coefficient (α) that is a properties of the medium at a determined wavelength. I_0 is the intensity of the illuminated light beam. By plotting the amount of absorption over time, the resulted waveform represents pulsatile changes of arterial blood volume on that tissue that correspond to the heart rate [2].

$$I = I_0 e^{-\alpha l} \tag{2.1}$$

An example of a photoplethysmographic waveform is shown in Figure 2.5. PPG waveform consists of two components: constant component and changing component, which are called as DC and AC as an analogy to constant and alternating current [2]. DC component corresponds to the amount of light absorbed by the tissue which is a constant value due to the invariant structure of the tissue and the average blood volume of venous blood and diastolic volume of the atrial blood. The DC component can also change slowly because of respiration, vasomotor activities and thermoregulation. On the other hand, the AC component represents the pulsatile arterial blood and changes in the amount of blood

volume between the systolic and diastolic blood pressure phases. The AC component is placed over the DC part and its frequency corresponds to the heart rate .
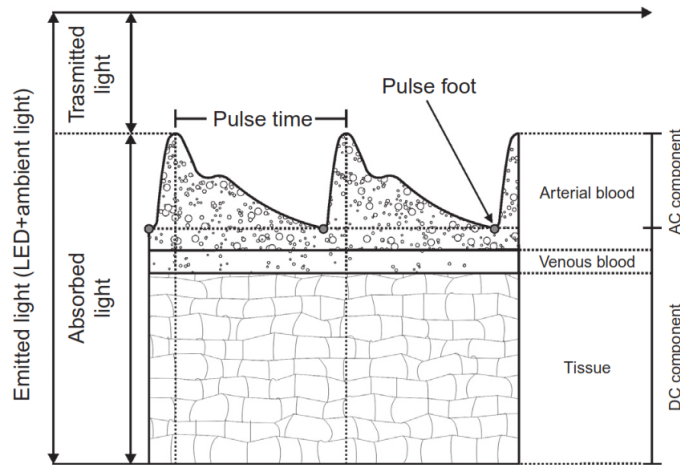
Figure 2.5. *Depiction of the constant and changing components of a typical PPG signal. The AC component related to blood volume changes with each heartbeat is placed over a DC component that is associated with the constant light absorption due to the tissue. [2]*

Wearable PPG can work on two different modes based on the place of the photo-detector that is capturing the light. Figure 2.6 represents these two modes. Traditional way of PPG measurement works in transmittance mode in which the transmitted light is detected by a photo-detector at the opposite side of the LED source. Although, this mode is better in obtaining higher quality signal, it cannot be applied for all body locations and it can be placed on fingertip, nasal septum or earlobe to be more effective. But these measurement sites are more susceptible to environmental temperature [3].

On the other hand, there is reflectance mode that has recently gained interest and that is more suitable for long term monitoring due to its convenient location and unobtrusiveness. The different measurement sites include the forearm, wrist, ankle and forehead. In this mode, the reflected or back-scattered light from tissue or blood vessels is detected by a photo-detector located next to the LED [3].

PPG technology was introduced in clinical routine in 1972 for monitoring of oxygen saturation (pulse oximetry) [34]. Recently, in addition to oxygen saturation (SpO_2) PPG has been applied for evaluation of HR, but its sensitivity to movement artefacts were restricted
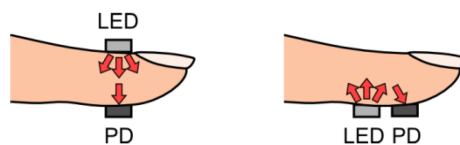
Figure 2.6. *Placement of the LED and photo-detector in transmittance and reflectance modes of photoplethysmography [3]*

the applicability of PPG specially in ambulatory monitoring. However, recent studies show that this promising technique can be helpful for other clinical practices including continuous cardiac and respiratory event monitoring, early screening and diagnostic of various cardiac diseases [15, 33, 35, 36].

Different factors can affect the quality of PPG signal including implemented sensing setup, probe attachment site and contact pressure, subject movement and posture, poor blood perfusion, ambient light and environmental temperature [2]. These factors increase errors in diagnostics based on PPG signals and are needed to be considered when using PPG signals as resource data for diagnosis.

2.2 Review of related work

As explained earlier, the classical way for AF screening is using portable ECG devices and screening the ECG waves. There are many proposed algorithms for AF detection using information derived from the ECG. However, development of wearable technologies is leading to an interesting alternative solution for the ECG-based devices that is more affordable and simpler to use by general public. Although, there is an increasing interest in this domain, it has not been extensively studied. In the section that follows, a brief description of the works done on the AF detection based on the PPG signals is presented and a summary of these works is listed in Table 2.1. The abbreviations and acronyms present in this table are introduced in later section.

Table 2.1. Review of recent researches on the AF detection using PPG data

Author-Year	Recording Device	Extracted Features	Classifier	Performance
Lee et al. (2013) [16]	Smart Phone (2-minutes recording)	RMSSD, ShEn, SampEn	Comparison with threshold derived from MIT-BIH databases	ACC RMSSD 98.44%, ACC ShEn 84.94%, SEN SampEn 95.22%
Ferranti et al. (2015) [11]	Empatica,wristband	SampEn, ShEn, SD, RMSSD, nRMSSD, coefficient of variation, pNN50, Lorenz plot distribution of subsequent interval deltas, shape analysis	PCA and wrapper types feature selection, support vector machine classification	ACC 90% SEN 96.67%
Chong et al. (2015) [17]	Smartphone iPhone 4S (2-minute recordings)	RMSSD, Sh En, Poincare plot, pulse rise and fall times	Threshold-based classification rules	SPE 98.86%, discriminates PVCs and PACs from AF. SEN 96.84% and 97.83%
Chan et al. (2016) [18]	Smart Phone (17.1seconds recording)	Lack of repeating patterns in the PPG waveform due to the irregular rhythm of AF	support vector machine based on the self-similarity of the waveform	SEN 92.9%, SPE 97.7 %, PPV 53.1%, NPV 99.8%
McMANUS et al. (2016) [19]	Smartphone iPhone 4S (2-minute recordings)	RMSSD, RR Differences, ShEn	Threshold-based classification rules	SEN 97%, SPE 93.5%, Acc 95.1%
Krivoshei et al. (2016) [12]	iPhone 4S (5 min video file)	nRMSSD, ShEn, SD1/SD2 from Poincare plot	Threshold-based classification rules	SEN 85%, SPE 95%
Fallet et al. (2016) [13]	wrist-type device (CSEM)	Ratio of the power of the fundamental frequency and the first harmonic to the total power of the pre-processed PPG signal	Thresholding on the Adaptive organization index (AOI)	AOI values: 0.45 ± 0.11 for AF, 0.73 ± 0.19 for SR
Nemati et al. (2016) [14]	watch-based wearable device (Samsung Simband)	Sample Entropy, STD	Elastic Net logistic model	ACC 95%, SEN 97%, SPE 94%, AUROC 0.99
Bonomi et al. (2016) [15]	wrist-wearable device	Probability of AF using First-order 11-state Markov Model	Threshold on the calculated probability	SEN $97 \pm 2\%$, SPE $99 \pm 3\%$
Schäck et al. (2017) [20]	Smart Phones (20 Seconds recording)	RMSSD, Sh En, mean, median, SD, mean absolute deviation, crest time, peak rise height, VLF, LF, HF, kurtosis of the spectrum	wrapper type feature selection, support vector machine classification	ACC, SEN, SPE for ShEn-mPRH 100%

Most of the highly developed methods for AF detection rely on the extracting interbeat-intervals from PPG signals acquired from fingertip, wrist or earlobe, and investigate the possibility of decision making based on IBI series statistics. Some authors have preferred simple methods such as a set of decision rules while others selected more developed machine learning classifiers such as support vector machine.

Recent studies have applied PPG measured from smart-phone cameras [16, 17, 20] and calculated a group of statistical features to distinguish between AF and SR. For instance, in one study [16] three statistical methods including RMSSD, shannon entropy (ShEn), and sample entropy (SampEn) acquired from a collection of pulsatile time series were used in iPhone-based AF detection. The authors applied MIT-BIH AF and MIT-BIH NSR databases to derive threshold values for these three features. Using these threshold values, they could reach the beat-to-beat sensitivity of 0.9763, 0.7461 and 0.9258 and specificity of 0.9961, 1.0 and 0.9980 for RMSSD, ShE and SampE, respectively. The achieved accuracies were 0.9844, 0.8494, and 0.9522. Authors of another study proposed a smartphone-based arrhythmia discrimination algorithm that is able to distinguish between normal SR, AF, premature ventricular contractions (PVCs) and premature atrial contraction (PACs) [17]. Features they applied included: RMSSD, Shannon Entropy, trajectory patterns of Poincare plot and pulse rise and fall times. The specificity of normal SR detection was 0.9886, and sensitivities of discrimination between PVCs and PACs from AF were 0.9684 and 0.9783, respectively.

In [20] both time-domain and frequency-domain features were computed. A time window of 20 seconds were used for the time-domain. The frequency-domain features were calculated for every 5 seconds segments with 80% overlap and then were averaged over the same 20 seconds time windows. The authors also utilized some features to detect motion artefacts during the recording and automatically exclude them. Finally, they applied feature selection and support vector machines for classification and achieved 100% detection accuracy of AF on the clinically recorded data. The distribution of their subjects was as 20 measurements of AF, 294 of SR and 12 of vibration (strong hand movement).

Although smart-phones are easy to use and affordable monitoring devices that patients can utilize during their daily life, they are intermittent-type measurement solutions and cannot be applied continuously in ambulatory applications. Therefore, the problem of happening arrhythmia at the time of measurement is still unsolved.

Ferranti and Laureanti [11] developed a decision-making system trained on information derived from blood volume pressure (BVP) signal acquired from Empatica E4 wristband [37]. The recording duration was 10 minutes. They extracted 16 diagnostic indexes including time-domain, frequency-domain, shape analysis and nonlinear indexes to classify patient's health status. Selection of the most relevant indexes were done through PCA and wrapper method. By applying the SVM classification on the selected features they could reach to the accuracy of 90% and sensitivity of 96.67% in AF detection.

In another research [14], Nemati et al. proposed and validated an AF detection algorithm using PPG and accelerometry data. The recordings were obtained from a multi-sensor wrist-worn device (the Samsung Simband). The features that they extracted include: SampEn, standard deviation of IBI and a robust version of standard deviation by excluding the IBIs outside of the $0.05 - 0.95$ percentile range. Additionally, they calculated two signal quality indexes (SQI). One index was using the Hjorth's purity quality metric [15]. This signal quality index purity is zero for random noise and it is one if the signal is sinusoidal. Another signal quality index was the average of the accelerometer amplitude. Finally, a channel with the highest signal quality was selected and only the features of that chosen channel were used for AF detection. 46 subjects were participating in this study, 15 with AF and 31 non symptomatic. The duration of recordings were 3.5 to 8.5 minutes. The accuracy of 95%, sensitivity of 97% and specificity of 94% were achieved in this work.

Fallet et al. [13] measured the level of disorganization of the various PPG signals during AF using an adaptive organization index (AOI). This index was defined as the ratio of the power of the fundamental frequency to the total power of the PPG signal. Adaptive band-pass filters were used to compute the fundamental harmonic. Their study population included 18 patients undergoing catheter ablation of cardiac arrhythmias. They had four categories in their dataset: SR, AF, regularly paced rhythm and irregularly paced rhythm. The mean of AOI values were measured as 0.45 ± 0.11 for AF, 0.73 ± 0.19 for SR, 0.78 ± 0.20 for regularly paced rhythm, and 0.61 ± 0.19 for irregularly paced rhythm. The area under the ROC curve was 0.864 between AF and SR classes.

Finally, Bonomi et al. [15] in 2016, proposed a method based on a first-order Markov model to detect AF from PPG signal acquired from a wrist-wearable device that was equipped with a PPG sensor along with an accelerometer. In this model the probability of AF given the irregular pattern in the interbeat time series was calculated, then using a predefined threshold the output of this Markov model was associated to either AF or SR class. The recorded accelerometer signal was used to determine the amount of motion for each interbeat time interval. When the motion level exceeded a previously selected threshold the pulses determining such IBIs were discarded. Their proposed approach achieved the sensitivity of $97 \pm 2\%$ and the specificity of $99 \pm 3\%$ for AF detection. Due to the motion artefacts the average of $36 \pm 9\%$ of monitoring period were not classified.

2.3   Technical background

This chapter begins with introduction to the concept of machine learning and specifically Artificial Neural Network, its fundamental principles and use scenarios. Afterwards, a brief overview of features related to arrhythmia detection is presented accompanied with a discussion about evaluation and performance metrics.

2.3.1 Machine learning

Machine learning is a fast-growing subfield of computer science. The principal idea of machine learning is to program machines (computers) so that they can learn and specifically are able to extract meaningful patterns or knowledge from a collection of data sets by themselves [38]. In a sense, machine learning can be viewed as a branch of Artificial Intelligence (AI), since it can refer to the changes in systems performing AI tasks. It is able to turn experience into expertise or "to detect meaningful patterns in complex sensory data" [38]. A widely quoted definition of machine learning is: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E" [39].

Exponentially growth of data production as well as the massive increase in computing power had special effect on machine learning so that it can now be applied in many fields such as search engines, computer vision, transportation, health monitoring, medical diagnosis, marketing, financial services, natural language processing among others. Depending on the available data and the purpose of the task, there are three major categories in machine learning: supervised, unsupervised and semi-supervised learning.

Supervised learning

The task in supervised learning is to teach the algorithm with interrelated input (independent) and output (dependent) data [40, 41]. Algorithms are trained based on a set of input variables (features) x_i and output variables (targets) y_i. According to the presented m training instances $D = \{(x_i, y_i); i = 1, ..., m\}$, which is called training set, the system learns the properties of the data and selects a hypothesis which minimizes the differences between predictions $\hat{y}_i$ and desired output y_i. Particularly, the task in supervised learning is that, given a training set, infer a function $h : \mathcal{X} \longmapsto \mathcal{Y}$ that maps the input vector to the output or target values such that the $\hat{y} = h(x)$ is a good prediction for the corresponding value of y. Here, $\mathcal{X}$ denotes the space of input values, and $\mathcal{Y}$ the space of output values.

Depending on the nature of target variables the supervised learning problem is called regression or classification. In the former the target values are continuous. While in the latter y can only take discrete values. In both of these problems, the main task is to find related patterns inside the training set and learn a model using these patterns such that it can predict labels of unseen samples. The block diagram of a typical supervised classification system is shown in Figure 2.7.

Unsupervised learning

Unlike supervised learning, in unsupervised learning the target labels are not known or available to the system. The training data only include the observations x without the output or classes. The system tries to learn the data properties and find out the hidden structure within data. The most popular application of unsupervised learning is clustering.
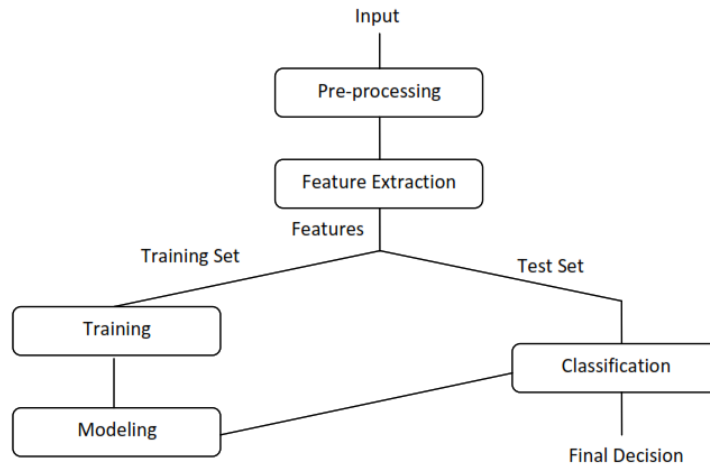
Figure 2.7. *Block diagram of a typical supervised classification system*

In this problem, the goal is to find clusters of data that are sharing similar properties in such a way that data points are placed in a specific cluster based on their scores in clusters' criteria. [38]

Semi-supervised learning

Semi-supervised learning is the combination of supervised learning and unsupervised learning [42] . The goal of this type of learning is to train the model using both labeled and unlabeled data. The reason to combine these methods is that in many applications labeling of the training data set is usually difficult, time-consuming and expensive. While unlabeled data can be collected easily and inexpensively. A mixture of two previously mentioned techniques can be applied to improve the learning performance. Self-training is the most common semi-supervised learning algorithm. A classifier is built using a small amount of annotated data and then iteratively classifies the unlabeled data and augments the most confident examples into the labeled training set.

2.3.2 Artificial Neural Networks Classifier

As previously stated, the goal in classification is to take input vectors and according to their properties assign them to different categories. Various types of algorithms can be used for classification problems. The following section describes briefly Artificial Neural Network and specifically multilayer perceptron method which is the applied method in this thesis. Afterwards, the learning process in MLP that can happens during the backpropagation is explained. Then it is followed by a brief review about the other classifiers that are used for comparison with the ANN at the results and discussion chapter.

Artificial Neural Network is a machine learning method which is inspired by the function

of human brain that is composed of a set of neurons and synapses [43]. Every neuron is connected by multiple synapses to other neurons in many layers. Figure 2.8 depicts a simple biological neural network. At a very simple level, neurons are basically computational units of the brain and synapses are signal transferring units. Dendrites (inputs) are channeled to axon terminals (outputs) and take electrical input (called "spikes") to the output through the axon channel.
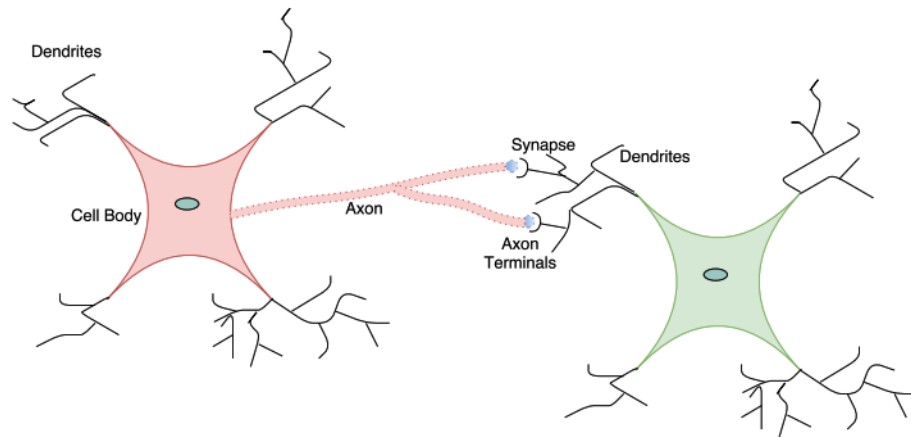
Figure 2.8. *A simple biological neural network*

Neural Networks have proven to be noteworthy for several machine learning tasks such as market prediction [44], image classification [45], speech recognition [46], audio feature extraction [47] etc. Even though the ANNs are well-known for their ability of learning, these networks have some limitations including the long training process, need of large datasets and requirements for determining the architecture and parameters of the network. However, thanks to the advancement in computer hardware technologies and developments in training algorithms ANNs have had a big recent resurgence. Still the biggest drawback of ANN for clinical utilization is the difficulty of interpreting the model specifically in the clinical environments. ANN is a black box once it is trained and knowing the process that is going underneath of this black box with regards to the decisions that the ANN is making is not trivial (it can be done but it is not as straightforward as with other machine learning algorithms).

2.3.3 Multilayer perceptron

A multilayer perceptron (MLP), is a computational model that processes information through a collection of interconnected computational units or nodes. This nonlinear mathematical model can be applied for regression and classification in numerous applications [4]. The operating principle of a simple neuron (shown in Figure 2.9) is that all input values are multiplied by weights. Then the resulting values are summed together and passed through a nonlinear activation function to provide the probability of the occurrence of an event. Activation function plays an important role to make the NN a nonlinear model and serves

as a decision function. ANN incorporates a bias term as well in order to prevent zero-sum skewing the learning of the neural network. These steps can be summed in the following equation for neuron k as:
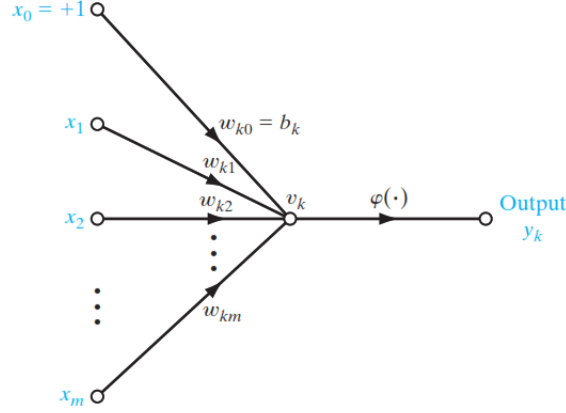
Figure 2.9. *A simple neuron model [4]*

$$v_k = \sum_{i=1}^{m} w_{ki} x_i + b_k, \tag{2.2}$$

where x_i's are input features, w_{ki}'s are multiplied weights and b is the bias term. The above expression can be simplified by letting $b_k = x_0$, so that v_k can be rewritten as:

$$v_k = \sum_{i=0}^{m} w_{ki} x_i = \mathbf{w_k^T x}, \tag{2.3}$$
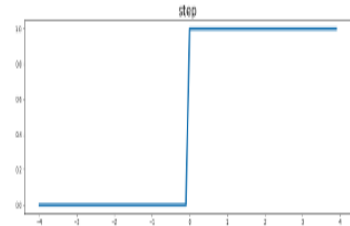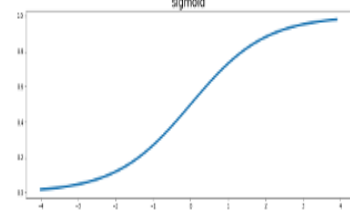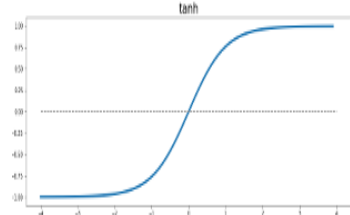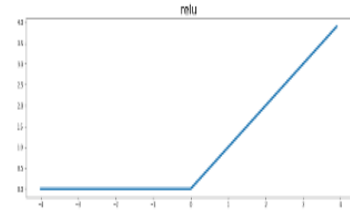
Afterwards, the output will be generated by passing v_k through the activation function φ. This function transform the input to the ranges of [0,1] (or sometime [-1,1]). It can be chosen depending on different implementation purposes.

$$y_k = \varphi(\mathbf{w_k^T x}), \tag{2.4}$$

Therefore, it can be seen from the Eq. 2.4 that the neural network is a nonlinear mapping function that maps a set of input variables to a set of output variables by controlling the weights vector W. Activation function is an abstract representation of action potential rate in biological neurons. Commonly used activation functions are listed below and their expression and figures are shown in Table 2.2:

- **Binary Step function** is a simple discrete-valued function whose value is zero for negative argument and 1 otherwise. It is not commonly used in MLP due to its discontinuity.

Table 2.2. *Most common used activation functions*

Name	Plot	Equation
Binary step		$\varphi(x) = \begin{cases} 0 & for\ x < 0 \\ 1 & for\ x > 0 \end{cases}$
Logistic sigmoid		$\varphi(x) = \sigma(x) = \frac{1}{1+e^{-x}}$
Hyperbolic tangent (tanh)		$\varphi(x) = tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
Rectified linear unit (ReLU)		$\varphi(x) = \begin{cases} 0 & for\ x < 0 \\ x & for\ x \geq 0 \end{cases}$

- **Logistic sigmoid** is a smooth approximation to the binary step. It is a continuous function with the output between 0 and 1.

- **Hyperbolic tangent (tanh)** is a bipolar version of the logistic sigmoid function with the outputs between -1 and 1.

- **Rectified linear unit (ReLU)** also known as ramp function that is the positive part of its input. ReLU has efficient gradient propagation specially in so called deep neural networks.

As illustrated in Figure 2.10 a multilayer perceptron or ANN model is organized in layers each containing several neural units. The basic topology in a feed-forward network includes three layers: input layer to receive the input signal, output layer that makes the prediction or decision about the input, and in between of those two, hidden layers which are the computation engine of ANN. The number of hidden layers can vary in different applications. Typically, neurons or units of each layer in the network are connected to all units of the previous layer. This configuration is called fully connected network.

The number of input neurons corresponds to the number of input variables and the number of neuron in output layers equals to the number of target labels. The number of hidden
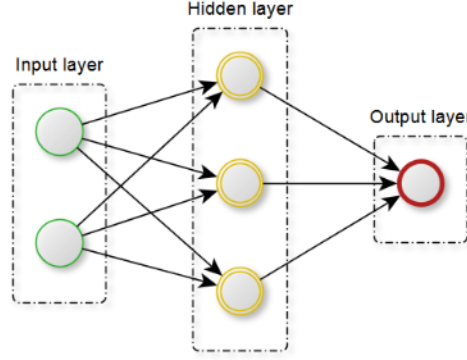
Figure 2.10. *Typical representation of ANN topology. The input, hidden, and output nodes are illustrated by nodes, and the weights are shown by links between the nodes. Arrows determine the direction of information flow during forward propagation.*

layers depends on the design decisions. Too few neurons may lead to underfitting since they are not able to satisfactorily discover the patterns in a data set. On the other hand, using too many neurons in the hidden layers may result in overfitting in which the network learns the details of the training data too well while it lacks the ability to generalize to the new and unseen data. Furthermore, the training time increases by an extraordinary large number of neurons in the hidden layers.

2.3.4 Backpropagation

The simplest MLP architecture is feed-forward network and the idea behind of it is a directed acyclic graph whose nodes are computational units and information flows only in one forward direction from the input nodes, through the hidden layer to the output nodes. There are not any cycles between these units in the network. The learning problem in feed-forward network is to find the optimal combination of weights so that the outputs of the network approximate the targets as closely as possible.

While, the most popular learning technique used in MLP is backpropagation in which there is a complex joining of multiple fully connected perceptrons resulting to a better understanding about the dataset. By applying a continuous activation function, the neural network is differentiable with respect to the parameters of the network (weights) which plays a crucial role in network training [40]. Backpropagation algorithm benefits from weight adjustment by error correction method. Training in a neural network is an iterative procedure in which parameters (weights of the network) are updated after each epoch until some stopping criteria is met. Epoch is a completed two-pass (one forward pass and one backward pass) of all the training samples. Learning process in backpropagation occurs through two phases: propagation and weights update. Consider an MLP with input features $\mathbf{x}(n)$ and desired response $\mathbf{d}(n)$ illustrate together a training sample as follows:

$$\mathscr{F} = \{\mathbf{x}(n), \mathbf{d}(n)\} \quad \text{for n = 1, ..., N} \tag{2.5}$$

At the beginning, the network is initialized with randomly chosen weights. The Figure 2.11 represents neuron j with the input signals that are produced by neurons at the previous layer [4].
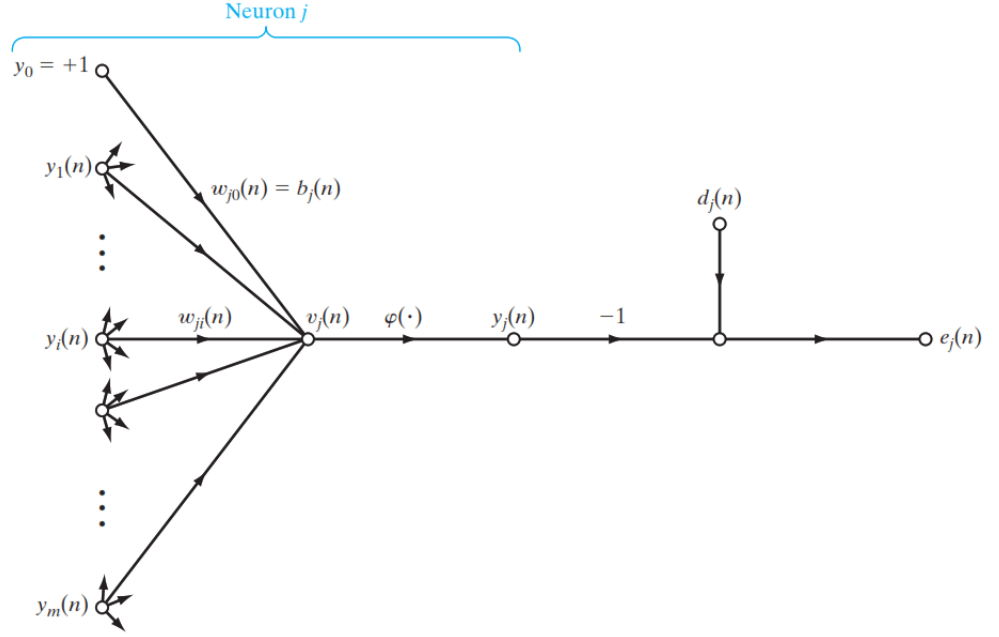
15

Figure 2.11. *Neuron j feeding by the signals produced by the neurons of the previous layer [4]*

Therefore the input of the activation function related to the neuron j is

$$v_j(n) = \sum_{i=0}^{m} w_{ji}(n) y_i(n) \tag{2.6}$$

where m is the total number of inputs of the neuron j. Afterwards, the output of neuron j is then

$$y_j(n) = \varphi_j(v_j(n)), \tag{2.7}$$

Propagation phase is similar to feed-forward network until $y_j(n)$ the output of neuron j in the output layer is generated. Then the error signal between desired response $d_j(n)$ and predictions (actual response) $y_j(n)$ is defined as:

$$e_j(n) = d_j(n) - y_j(n) \tag{2.8}$$

where $d_j(n)$ is the jth element of the desired response vector $\mathbf{d}(n)$. Using the Least Mean Square algorithm The loss function of neuron j is defined by:

$$E_j(n) = \frac{1}{2}e_j^2(n) \tag{2.9}$$

The total loss function of the whole network is calculated by summing the loss of all the neurons in the output layer:

$$E(n) = \sum_{j \in C} E_j(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \tag{2.10}$$

where the set C contains all the neurons present in the output layer. The goal is to make $y_j(n)$ and $d_j(n)$ identical for all $j \in C$, by minimizing the loss function of the network. This error signal propagates backwards through the network until the weights have their corresponding error values that represents their impact to final output error. After feeding the network with data sample n , the weights must be updated for the next epoch $n+1$:

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n) \tag{2.11}$$

where each weight is adjusted by

$$\Delta w_{ji}(n) = -\eta \frac{\partial E(n)}{\partial w_{ji}(n)} \tag{2.12}$$

where parameter η is called learning rate of the backpropagation algorithm that is the step size of each iteration in the negative gradient direction. The negative gradient in Eq. 2.12 stands for gradient descent in weight space, which means the weights get updated in a direction that decreases the value of total loss function. Therefore, the whole learning problem is reduced to the calculation of the gradient of the loss function of network with respect to its weights.

Based on the chain rule, this gradient can be determined as:

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = \frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \tag{2.13}$$

Based on the Eq. 2.10
$$\frac{\partial E(n)}{\partial e_j(n)} = e_j(n) \tag{2.14}$$

By using the Eq. 2.8, we get
$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \tag{2.15}$$

Next, differentiating Eq. 2.7 with respect to $v_j(n)$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi'(v_j(n)) \tag{2.16}$$

Finally, the use of the Eq. 2.6 yields

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_j(n) \tag{2.17}$$

Therefore, substituting the Eq. 2.14-2.17 into the Eq. 2.13 yields

$$\Delta w_{ji}(n) = -\eta \delta_j(n) y_j(n) \tag{2.18}$$

where the local gradient $\delta_j(n)$ is determined by

$$\delta_j(n) = e_j(n) \varphi'(v_j(n)) \tag{2.19}$$

Depending on the place of the neuron j in the network, two distinct cases can be identified. In one case, neuron j is present in the output layer and calculation of associated error signal is straightforward.
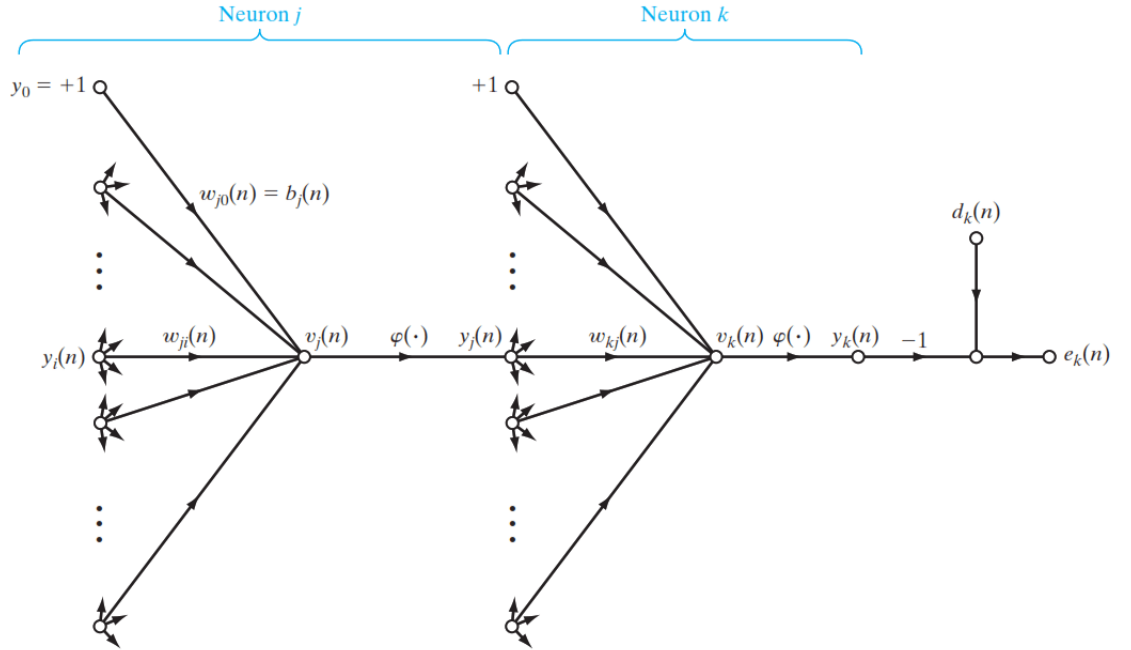
Figure 2.12. *Hidden neuron j connecting to the output neuron k [4]*

In another case, the neuron j is a hidden node instead of output node. In this situation (depicted in Figure 2.12), there is not any specified desired response to calculate the error signal using Eq. 2.8. Therefore error signal of a neuron in a hidden layer is determined recursively and working backwards using the error signals of all the neurons that are directly connected to this hidden neuron.

The local gradient for the hidden neuron j is redefined as

$$\delta_j(n) = -\frac{\partial E(n)}{\partial y_j(n)}\frac{\partial y_j(n)}{\partial v_j(n)} \tag{2.20}$$

To calculate the first term in the above equation, one can proceed as follows:

Total loss can be rewritten as

$$E(n) = \frac{1}{2}\sum_{k \in C} e_k^2(n) \tag{2.21}$$

where k is related to the neuron in the output layer. Therefore, differentiating Eq. 2.21 yields

$$\frac{\partial E(n)}{\partial y_j(n)} = \sum_k e_k \frac{\partial e_k(n)}{\partial y_j(n)} \tag{2.22}$$

By applying the chain rule

$$\frac{\partial E(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial v_k(n)}\frac{\partial v_k(n)}{\partial y_j(n)} \tag{2.23}$$

It can be seen in the Figure 2.12 that

$$e_k(n) = d_k(n) - y_k(n) = d_k(n) - \varphi_k(v_k(n)) \tag{2.24}$$

$$\frac{\partial e_k(n)}{\partial v_k(n)} = -\varphi_k'(v_k(n)) \tag{2.25}$$

From Figure it can be noted that

$$v_k(n) = \sum_{j=0}^{m} w_{kj}(n)y_j(n) \tag{2.26}$$

By differentiating Eq. 2.26, we get

$$\frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n) \tag{2.27}$$

Finally, substituting to the Eq. 2.23 yields

$$\frac{\partial E(n)}{\partial y_j(n)} = \sum_k e_k(n)\varphi_k'(v_k(n))w_{kj}(n) = -\sum_k \delta_k(n)w_{kj}(n) \tag{2.28}$$

Therefore, the local gradient $\delta_j(n)$ for the case of j as a hidden layer is described by

$$\delta_j(n) = \varphi'(v_j(n))\sum_k \delta_k(n)w_{kj}(n) \tag{2.29}$$

The backpropagation algorithm can be summed up as follows:

1. Propagate the input through the network in forward direction.

2. Propagate the errors through the network in backward direction:

   - Start from output layer $l = L$ and adjust the weights of all layers $l = 1,...,L$ using the following equation

$$w_{jk}^l(n+1) = w_{jk}^l(n) + \Delta w_{jk}^l(n) \tag{2.30}$$

   where

$$\Delta w_{jk}^l(n) = \eta \delta_j^l(n) y_j^{l-1}(n) \tag{2.31}$$

   where

$$\begin{cases} e_j^l(n)\varphi_j'(v_j^l(n)) & \text{if neuron j is in the output layer L} \\ \varphi_j'(v_j^l(n))\sum_k \delta_k^{l+1}(n)w_{kj}^{l+1}(n) & \text{if neuron k is in the hidden layer l} \end{cases} \tag{2.32}$$

3. Repeat the above steps until the stopping criterion is true, either the predefined maximum number of training epochs goes beyond or the error goes below a predefined threshold.

2.3.5 Other classifiers

There are various possible choices for classification tasks and the selection depends totally on the application and the type and size of the data that needs to be analyzed. Some well-known examples of classification algorithms are Naive Bayes classifier, decision trees, boosted trees, random forests, nearest neighbor, support vector machines (SVM) and discriminant analysis. Among these mentioned algorithms linear classifiers, such as linear discriminant analysis (LDA) and SVM with linear kernel, and nonlinear classifiers, such as ANN, SVM with nonlinear kernels and quadratic discriminant analysis represent the highly developed supervised classifiers. The main difference between linear and nonlinear classifiers is related to the shape of the decision boundary which is a straight line or plane in the linear classifiers and curved line or surface in the nonlinear classifiers.

Since in this research, the performance of the proposed ANN is compared to other classifiers including SVM (with linear and nonlinear kernels), LDA and QDA, hence a brief review of these classifiers is provided without describing an in-depth knowledge about their key parameters.

Liner and quadratic discriminant analyses

The LDA is basically a classical dimensionality reduction technique used in various application domains. It is a popular statistical procedure that "maximizes the between-class variance and minimizes the within-class variance" to obtain the maximal separability [40]. Given a dataset including two classes, the LDA method finds an optimal projection vector to map the n-dimensional input into a line in such a way that the samples from the same

class are grouped as much as possible, and different classes are separated from each other as far as possible. The LDA assumes that the data within each class are normally distributed, and covariance matrix of both classes are equal. The QDA is closely related to the LDA and similarly sets assumption that the samples are normally distributed within each class. However, dissimilar to LDA, QDA assumes that each class has different covariance matrix.

Support vector machine

The SVM is basically a linear binary classifier arising from statistical learning theory that defines a hyperplane as the decision boundary in such a way that the margin (the distance from the nearest training points) between two classes is maximized [48]. The SVM can be extended to deal with more difficult cases of nonlinearly separable patterns by mapping the data into a higher dimensional feature space where it becomes linearly separable. This mapping is done using optimal kernel method.

2.3.6   Feature extraction

Diagnosis of atrial fibrillation using ECG signals can be done by analysis of the morphological properties of the electrocardiogram - meaning the absence of the p-wave. Unfortunately, in the PPG signal there is not any morphological property like the absence of p-wave in the ECG signal, to enable us to detect abnormal atrial activity and distinguish between AF and SR. Morphological features of the PPG such as pulse amplitude, width and area were not used for AF detection. Therefore, other diagnostic metrics such as high disparity of the length of IBIs can be helpful for automatic detection of AF using only PPG signals.

Several statistical features can be extracted from PPG signals to distinguish between AF and SR. Heart Rate Variability (HRV) is the most useful signal for understanding the functioning of the autonomic nervous system on the HR. Many physiological factors that are adjusting the normal heart rhythm can be reliably reflected in the HRV [49]. HRV is related to the analysis of variation in the beat to beat intervals of the heart and can provide a measure for cardiac condition that manual visual inspection of PPG or even ECG morphology cannot provide. As the halmark of atrial fibrillation is its irregularity in the heartbeat therefore typical HRV parameters can be applied as features for AF detection. These features are divided mostly into three categories that are associated with time domain, frequency domain and nonlinear method of analysis.

Time domain parameters: Time domain representation of a signal describes the signal variations with respect to the time. The time duration between one beat and another beat is called NN (Normal-to-Normal) interval. Normal means that abnormal beats such as ectopic beats have been removed and compensated [50]. Most of the time domain features are extracted from statistical properties of a window of N beats. Some of the most widely used time domain features in the literature are the following:

- **Mean (AVNN):** Average value for N interbeat intervals within a segment.

$$Mean = \bar{I} = \frac{1}{N} \sum_{i=1}^{N} I_i \tag{2.33}$$

- **RMSSD:** Root mean square of successive difference between adjacent NN intervals in each segment. RMSSD captures the variation of the N adjacent peak differences and provides an estimate of short-term components of HRV [50].

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (I_{i+1} - I_i)^2} \tag{2.34}$$

- **SDNN:** Standard deviation of N interbeat intervals within each segment.

$$SDNN = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (I_i - \bar{I})^2} \tag{2.35}$$

- **pNNx:** Number of successive difference of intervals which differ by more than x ms, divided by the total number of NN intervals within each segment. The common value for x is 50 ms.

$$pNN50 = P(|I_{i+1} - I_i| > 50ms) \tag{2.36}$$

Frequency domain parameters: The interbeat intervals can be decomposed into their frequency constituents using fast Fourier transformation (FFT). Various parameters can be extracted by frequency analysis of PPG signal. Frequency-based HRV analysis or spectral analysis determines that how large the power of the input signal is inside a certain frequency band. It has the ability to discriminate between the sympathetic and parasympathetic contents of the NN intervals [51]. These features are not ideal solutions in practice as they need at least 2 to 5 minutes of recordings [50]. For example very low frequency area cannot be represented correctly in a short-term measurement.

- **ULF:** Ultra Low Frequency (between 0 and 0.0033 Hz)
- **VLF:** Very Low Frequency (between 0.0033 and 0.04 Hz)
- **LF:** Low frequency (between 0.04 and 0.15 Hz)
- **HF:** High frequency (between 0.15 and 0.40 Hz)
- **LF/HF:** The ratio of low to high frequency power.

Other than these mentioned frequency domain features, there are some other features such as power spectral density (PSD), short term Fourier transform (STFT) and continuous wavelet transform (CWT) that can be extracted as features for HRV analysis.

Nonlinear analysis: Nonlinear methods are designed mostly based on the chaos theory and fractal geometry:

- **Entropy:** It is a measure of uncertainty and unpredictability of fluctuations in a time series. For instance, a random white noise signal has the highest entropy value because the pattern of the signal is not certainly predictable. If there are a large number of similar values in a time series, entropy will be small. Approximate (ApEn), sample (SamEn), Shanon (ShEn) and spectral entropy (SpEn) are all measures of randomness and complexity of a time series [52]. Spectral entropy reflects the complexity of the input signal in the frequency domain and is given by:

$$H = \sum_f p_f log(\frac{1}{p_f}) \tag{2.37}$$

where p_f is the probability density function (PDF) of the process at frequency f. Entropy analysis is useful methodology but it is hard to implement in short segment of time series [53].

- **SD1/SD2:** This feature is used to analyze the Poincare plot that is a graph of each interbeat-interval against the prior interval that creates a scatter plot [54]. One example of Poincare plot is shown in Figure 2.13. SD1 is the standard deviation of Poincare plot perpendicular the line of identity ($IBI_{i+1} = -IBI_i$) and SD2 is the standard deviation of Poincare plot along the line identity ($IBI_{i+1} = IBI_i$). SD1/SD2 is the ratio of SD1 to SD2 that measures the unpredictability of the IBI time series.
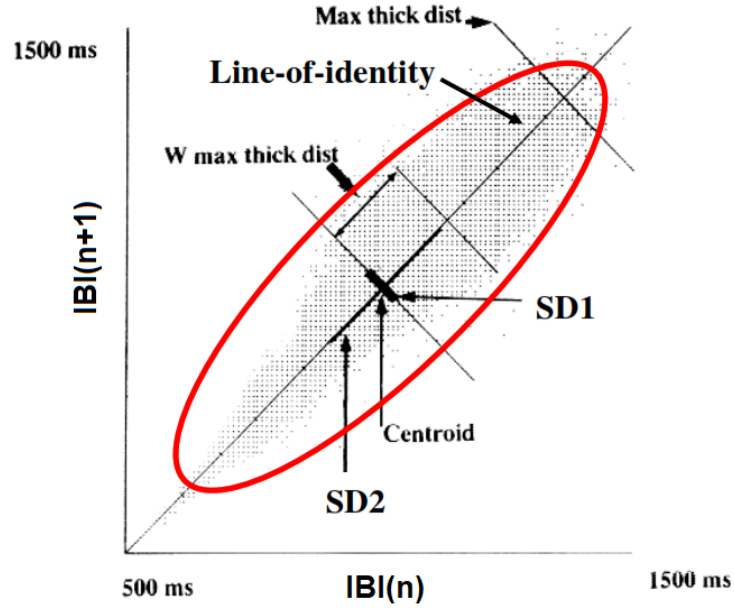
Figure 2.13. *One example of Poincare plot [image from www.physionet.org with a slight modification]*

2.3.7 Cross validation

A successful machine learning algorithm should be able to progress from individual examples to broader generalization. One of the most important concern in learning process

is generalization that means how well a model can perform on new and previously unseen samples [55]. The lack of generalization means that when a new example is being tested, a big error will be generated. It is possible to build a model which mimics very well the training set with high accuracy. It can happen when the model is designed by iterating many times on a very limited size data set. In this situation the model is suffering from the overfitting and lack of generalization.

In order to prevent this problem and reliably assess the performance of the model in unseen data set, a different test set is required. If there is a rich dataset, the simplest way is to divide the dataset into training, validation and test sets [56]. Training and validation sets are used during the training phase. For instance in ANN classifiers training set is used to find the optimal weights and validation set is applied for tuning the parameters of the network architecture such as number of hidden layers. Test set is used at the final phase for performance evaluation of the specified model. However, if the available data is limited, we can benefit from the idea of cross validation.

The principle of cross validation is to split the training data into k distinct subsets and for each model, evaluate it as follows: train the model on $k-1$ groups to find the model parameters and then evaluate the model on the remaining group. This procedure is then repeated as many times as there are segments of training data in a way that every part is used once as a validation set. Finally, the performance scores from the k runs are the mean value of all test set performances. This method is called k-fold cross validation and it is illustrated in Figure 2.14 for the case of $k = 5$. In the simplest case the partitions are in equal size.
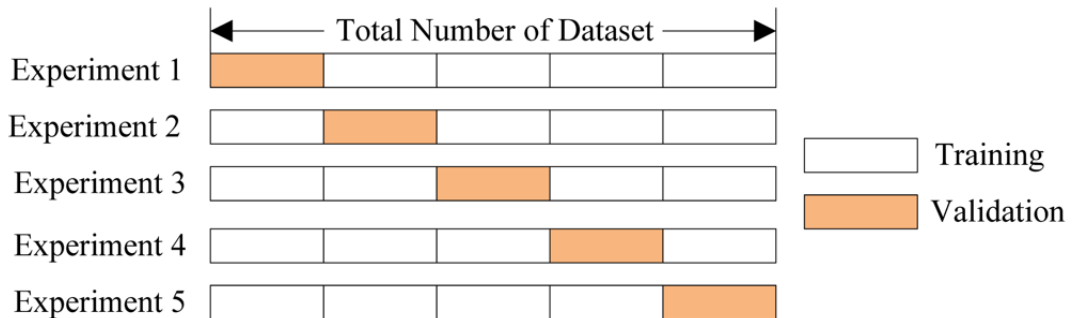
Figure 2.14. *k-fold cross validation representation [image from https://www.kaggle.com/ dansbecker/cross-validation]*

2.3.8 Evaluations and performance measures

The idea of performing a detection and diagnosis test is to assure that a patient has a particular disease, to the extent that we can take decision for the next steps. Therefore, it is important to have some metrics to assess the performance of a proposed classifier or diagnostic test. The most common illustrative way to measure the performance of a classifier is using a table called confusion matrix or contingency table. It visualizes the

variety of performance measures in a classification task. Most of the important performance metrics then can be drawn from the confusion matrix entries. For a binary classification problem with positive (diseased) and negative (non-diseased) classes the confusion matrix is a two-by-two table, shown in Figure 2.15, constructed by ground truth and predicted labels including four important concepts [57]:

Figure 2.15. *A confusion matrix template for the binary classification*

- True Positive (TP): the number of correctly classified diseased samples,
- True Negative (TN): the number of correctly classified non-diseased samples,
- False Positive (FP): the number of misclassified diseased samples,
- False Negative (FN): the number of misclassified non-diseased samples.

The appropriate performance measure depends on the task needed to be done and the type of data used for modeling. There are several measures that cann be applied at the same time to estimate the performance of a machine learning algorithm. For instance, the simplest measure is accuracy (ACC) that shows the percent of correctly classified samples. On the other hand, error rate (ERR) is the percent of incorrect classifications that means:

$$ACC = 1 - ERR = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.38}$$

$$ERR = \frac{FP + FN}{TP + TN + FP + FN} \tag{2.39}$$

However, this measure assumes equal costs for misclassification and the accuracy alone is not able to specify the efficiency of the classifier and does not provide any information about different types of error. In contrast, other measures such as sensitivity (SEN) and specificity (SPE) that are used in this study provide a measure of true positive rate (TPR) and true negative rate (TNR), respectively.

Sensitivity or recall shows the ability of a test to correctly classify 'diseased' or positive samples and is calculated as:.

$$SEN = \frac{TP}{TP+FN} \tag{2.40}$$

and specificity is the ability of a classifier to correctly classify non-disease samples or is the probability of the test being negative when disease absent.

$$SPE = \frac{TN}{TN+FP} \tag{2.41}$$

Most of the classification algorithms provide a continuous probability value as an output, therefore, a threshold is required to assign the output to a specific class (positive or negative in binary classification). All of the above mentioned performance measures depend on the discrimination threshold. For instance, in order to obtain a reliable sensitivity and specificity, it is needed to choose a proper threshold. In order to overcome the subjectivity effect on the threshold selection process, a figure of the sensitivity (TPR) over false positive rate (FPR) is plotted as the threshold (partition value) changes. FPR is determined as:

$$FPR = 1 - SPE = \frac{FP}{TN+FP} \tag{2.42}$$

This curve is called receiver operator characteristic (ROC) and shows the relationship between sensitivity and specificity of a classifier with different partition values. ROC analysis can quantify how accurate is a medical diagnostic test or system to distinguish between positive and negative cases [58]. Another common performance measure used in this study is area under the ROC curve (AUC) that is an index derived from ROC curve and it represents the overall accuracy of a test. A perfect test gives an AUC of 1.0, while random chance has an AUC of 0.5. The larger the AUC the better overall diagnostic performance. This metric can be applied as a measure for comparison of accuracy between different model results.

3. MATERIALS AND METHODS

This chapter describes the data collection procedure and the implemented system including pre-processing, signal quality estimation, feature extraction and classification.

3.1 Subjects

In total 30 subjects participated to the study. The patient population consisted of adult patients who had undergone an elective surgery immediately before the recording and were recovering from anesthesia. The cardiac rhythm of patients was not controlled by any drugs. Patient population information is shown in table 3.1. The subjects were classified into two groups: SR and AF. They were classified by a medical doctor at the beginning of the measurement, based on their current rhythm. It was visually verified that the cardiac rhythm of the patients assigned to the two groups was either SR or AF during the whole time of the recording. However, one of the AF classified patient was later excluded from the final analysis because it was noticed that the cardiac rhythm of that patient was switching between sinus rhythm and atrial fibrillation several times during the recording.

Table 3.1. *Patients population information*

	Female	male	Age($\mu \pm \sigma$)
Continuous AF	8	6	74.8 ± 8.3
Normal SR	7	8	67.5 ± 10.7

Figure 3.1. *PulseOn Optical Heart Rate monitor [5]*

An ethical approval was obtained from the local ethical review board of Pirkanmaa Hospital District (ETL R17024), and the study was registered at clinicaltrials.gov (NCT03081793) before study initiation. The procedures followed in this study were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000 and 2008. The test subjects gave their written consent to participate in this study and had the right to leave at any time.

3.2 Data acquisition

The wrist PPG signal together with a tri-axis acceleration data was recorded using a PulseOn Optical Heart Rate (PO OHR) monitor (PulseOn Technologies Ltd., Espoo, Finland [5]). PulseOn device (shown in Figure 3.1) is a wrist worn device designed for continuous, real-time data acquisition in daily life. PulseOn OHR uses reflection mode PPG with utilizing two light wavelengths (green and infrared) to record the PPG signal. It is also equipped with a tri-axis accelerometer. The PPG signal was acquired at a sampling frequency of 25 Hz. The device was worn as instructed by the manufacturer, with the main part of the device facing upward, similar to a regular wrist watch shown in Figure 3.2, about one finger width from the wrist bone and tightened by the responsible person of data acquisition, to ensure optimal skin contact for the whole recording.

Figure 3.2. *PulseOn Optical Heart Rate monitor on the patient's wrist during acquisition*

The ECG signal was recorded with the GE Carescape^{TM} B850 monitor (GE Healthcare Ltd. [59]) to be used as the ground truth. The sampling frequency of ECG signal were 300 Hz, respectively. RR intervals (RRI) were acquired from the reference ECG signal using Kubios HRV software, version 2.2 (Kubios Ltd. [60]). The ECG waveforms were subject of the visually inspection by an expert to ensure that no R-waves were missed, correct any R-wave detection errors, and confirm that the whole measurement was either SR or AF.
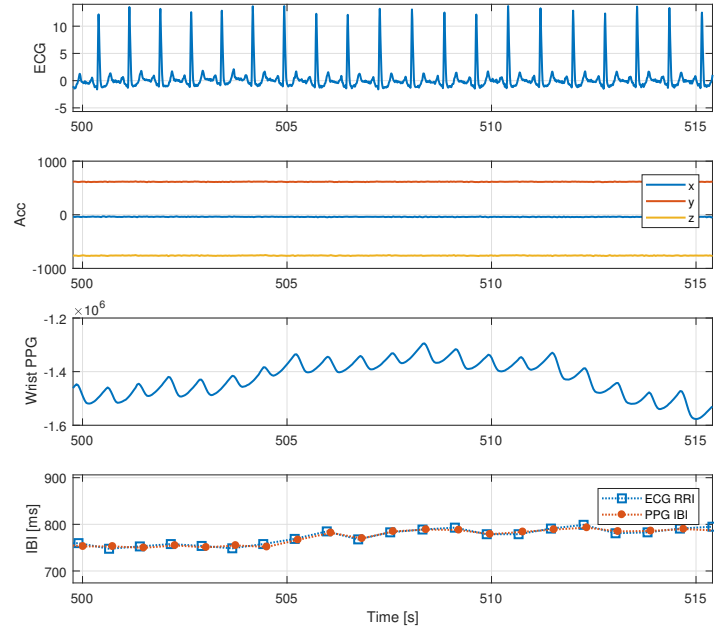
Figure 3.3. *Example of measured waveforms (ECG, Acceleration, PPG) and corresponding IBI and RRI for an SR patient*

The data recordings were performed in the post-anesthesia care unit of Tampere University Hospital. The patients were laying down in bed and remained in the steady state for the whole data acquisition process. The duration of each recording was 1.5 hours in average. In total, there were 23.52 hours of data recorded for the SR data sets and 21.96 hours for the AF data sets.

3.3 Implemented AF detector

3.3.1 Signal preprocessing

Signal processing and analysis of the recorded dataset was performed in Matlab R2017a. The raw PPG signal needs to be filtered to remove the baseline wandering related to the respiration of the subjects and the movement of the device with respect to the skin. A proprietary signal processing algorithm of PulseOn was applied to filter and then segment the PPG signal beat by beat and identify the individual PPG pulses. Accuracy of estimated HR and IBI from PPG signals obtained from PulseOn device have already been evaluated in [61,62]. Sample ECG, Acceleration, PPG waveforms and the corresponding RR intervals and IBI values for SR and AF subjects are depicted in Figure 3.3 and 3.4, respectively.

To assess the clinical utility of PPG pulses, it is needed to define a method to estimate the signal quality of each pulse. The applied method for signal quality estimation in this study is based on the PulseOn proprietary method that is briefly explained in the next section.
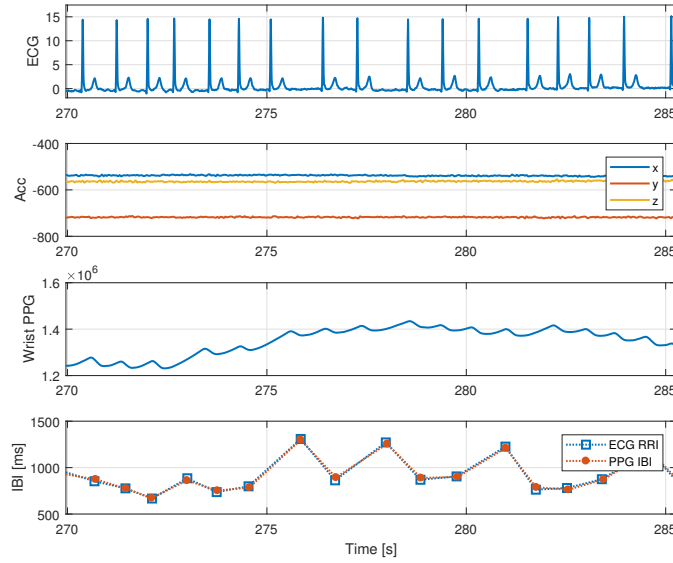
Figure 3.4. *Example of measured waveforms (ECG, Acceleration, PPG) and correspond-ing IBI and RRI for an AF patient*

3.3.2 Signal quality estimation

Different factors can affect the quality of the PPG signal including probe attachment site, contact pressure, subject movement and posture and environmental temperature [2]. These factors will increase errors in diagnostics based on PPG signals and finally can lead to a large number of missed or incorrect detection and false alarms in clinical decision making setups which apply PPG signal as input.

The raw PPG signals can be heavily corrupted by motion artefacts. Examples of motion corrupted IBI for an SR data set and an AF data set are shown in Figures 3.5 and 3.6. Therefore, the pulses which are detected during the high motion areas are unreliable and must be classified as artefacts. However, considering only the motion is not efficient for signal quality estimation due to the presence of other sources of disturbances than the motion, that are corrupting the PPG signal. Several quality indices such as amplitude of PPG pulses, amplitude to duration ratios, variance of PPG pulses, direct matching and linear resampling of pulses can be fed to a decision tree [63–65].

In this study, the IBI quality is estimated using the PulseOn proprietary algorithms that is based on the waveform morphology of the PPG pulses and the amount of motion. An example of the functioning of signal quality estimation (SQE) classifier is shown in Figure 3.7. In 160-165 s interval with high motion signal (large variations in the tri-axis acceleration signal), PPG signal is corrupted and the IBI is not accurate in comparison with the RRI from reference ECG. As it can be seen in low motion area, the PPG signal quality increases and the IBIs are accurate enough, therefore, pulses are considered reliable (or correct) in this area.
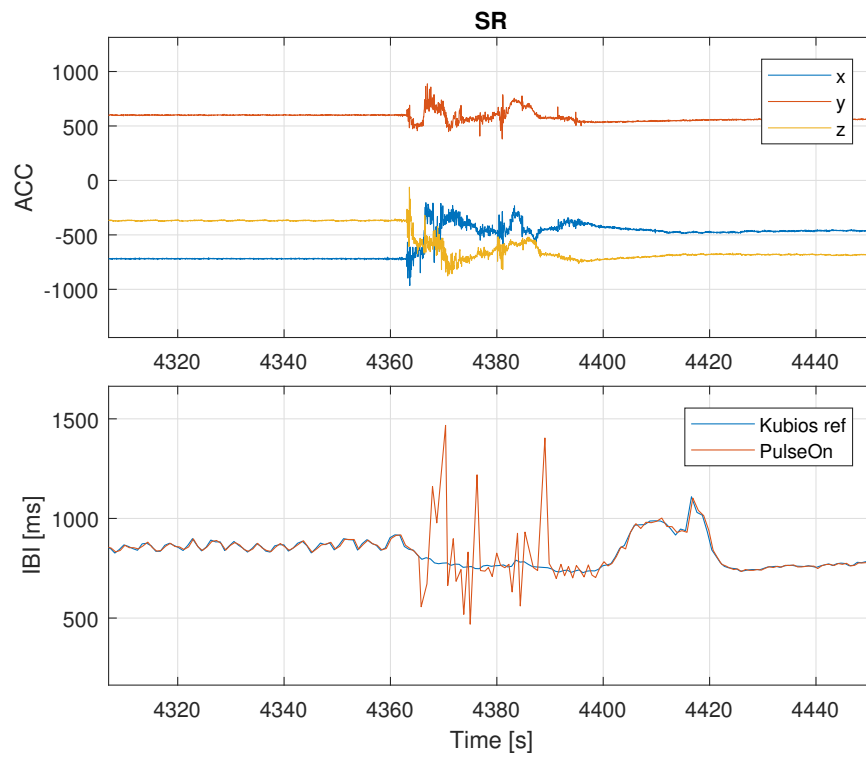
Figure 3.5. *Example of corrupted IBI due to high motion in SR patient*
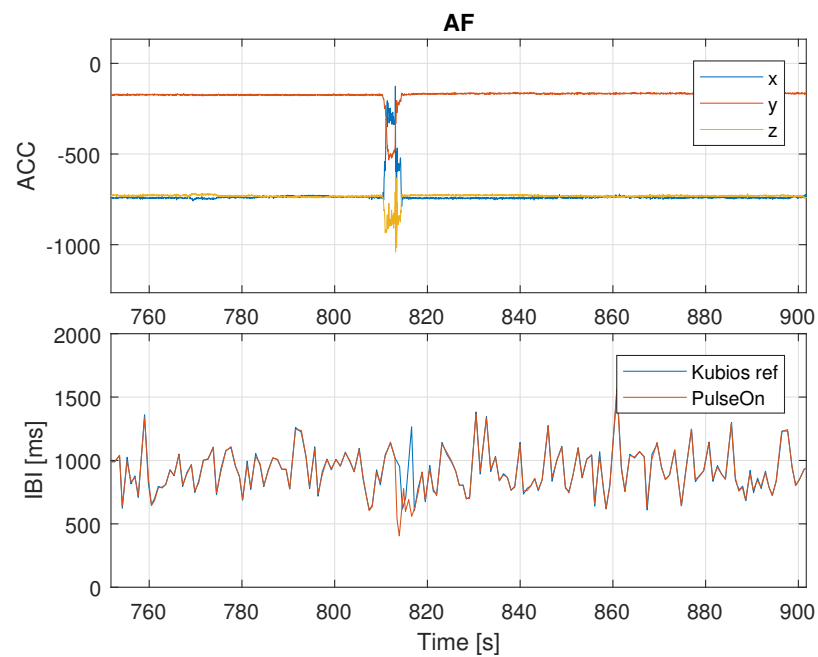
Figure 3.6. *Example of corrupted IBI due to high motion in AF patient*
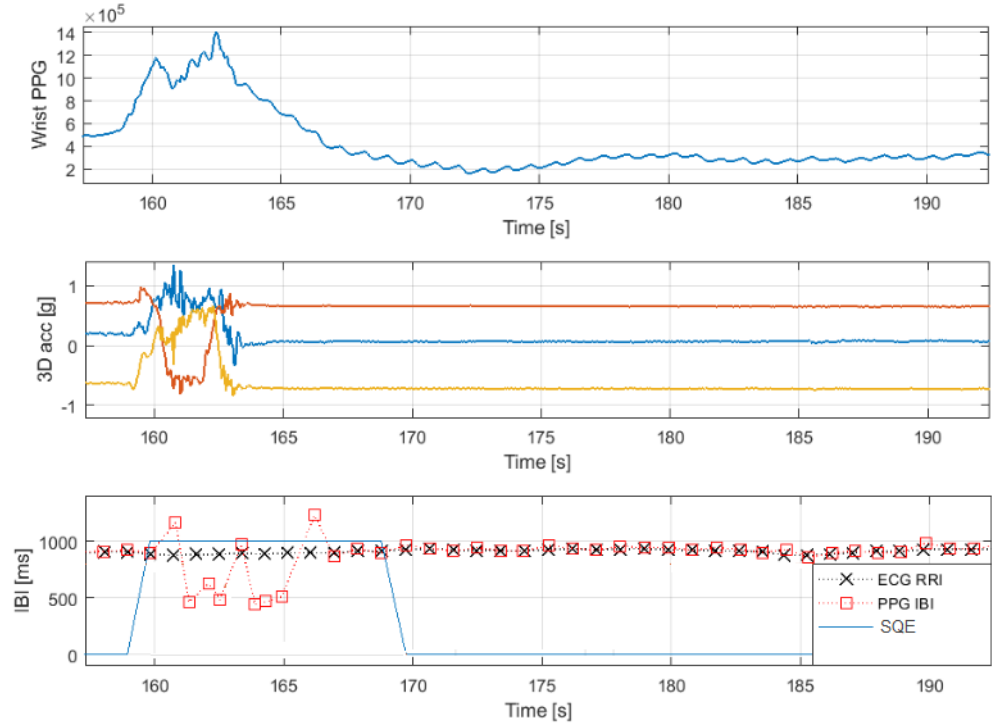
Figure 3.7. *Example of the functioning of signal quality estimation classifier (blue line shows the output of PulseON signal quality estimation algorithm. It is zero for correct pulses and non-zero for artefact pulses) [6]*

3.3.3 Feature extraction

Features applied in this study are identical to those features which are commonly used in the arrhythmia detection using ECG signal. This is due to the fact that the PPG signal is produced originally from the mechanical contraction and pumping action of the heart. This mechanical activity itself is the result of electrical activity of the heart which can be recorded by ECG device. Therefore the IBIs of the PPG signal are equivalent to the RRI of ECG signal. These features are based on the analysis of the pulse rhythms and more precisely based on the analysis of the interbeat-intervals. As it can be seen in Figure 3.3, the SR IBIs in the area with low motion are almost constant. However, AF arrhythmia variations will lead to IBI with large amplitude varieties (Figure 3.4).

In order to extract the features used in ANN-based classification, the IBI sequence is divided into sliding windows of $N = 30$ consecutive IBIs. For instance, the two consecutive IBI sequences are as follows:

$$T_1 = \{t_1, t_2, \cdots, t_N\},$$

$$T_2 = \{t_2, t_3, \cdots, t_{N+1}\}$$
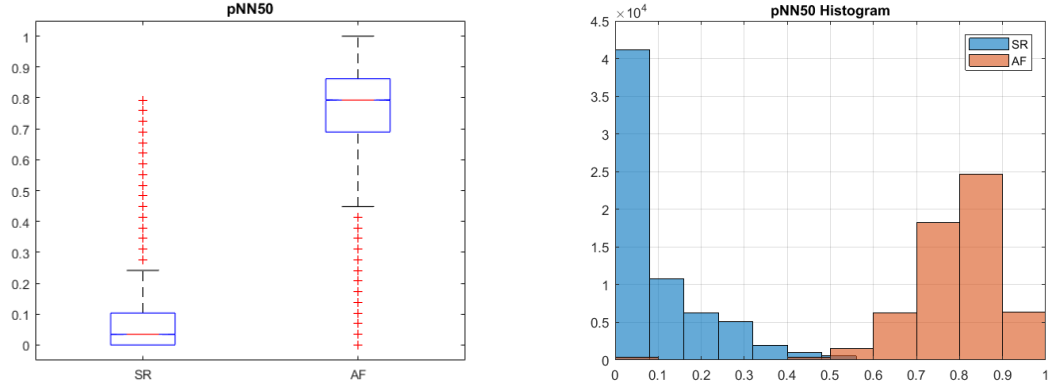
Figure 3.8. *Comparison of statistical values of pNN50 between normal SR and AF subjects*

The length of the analyzed window needs to be taken into consideration. A longer window length provides a more robust estimation of the interbeat-interval segment content, since more data is available, although, the computational cost is higher. In contrast, a short window length allows for faster calculations and is more effective to address the challenging problem of detection of those paroxysmal AF which usually has unpredictable onset and may have very short duration, e.g. less than 1 minute. For each segment the features were computed from the statistics of the successive interbeat-intervals: the transition matrix cells introduced in [66], the fraction of successive IBIs that differ by more than 50 ms (pNN50), the standard deviation of IBIs and the root mean square of successive difference of IBIs as the input features for ANN.

Figures 3.8-3.10 compare the statistical value distributions of pNN50, SDNN and RMSSD of normal SR and AF subjects, respectively. These features are used to quantify beat-to-beat variability, and since higher variability is present in AF than in SR, these features are expected to be higher than those of SR IBI sequence. It is apparent that there is no overlap in spreads or 75% SR is below 75% AF sets, resulting in a significant difference between groups SR and AF. However, it can be seen that there are outliers in all of these features and as a result the AF detection should not solely be based on one feature and none of these features alone cannot be used to reliably discriminate between AF and SR. Figure 3.11 shows the feature space of pNN50, RMSSD and SDNN features for different arrhythmia classes (SR and AF).

Another input feature that has been recognized as a promising feature is the transition matrix cells achieved from the IBIs. The idea is to model the interbeat-interval sequence as a 3-states Markov process [66]. In this model, each beat is assigned to one of the three states {Short, Regular, Long} based on its IBI length. A beat is classified as Short if the associated IBI does not exceed 85% of the mean IBI and it is assigned to Long state if the associated IBI exceeds 115% of the mean IBI, otherwise it is classified as Regular. [66] The mean IBI is determined recursively by the following equation:

$$meanIBI(n) = 0.75 \times meanIBI(n-1) + 0.25 \times IBI(n) \qquad (3.1)$$
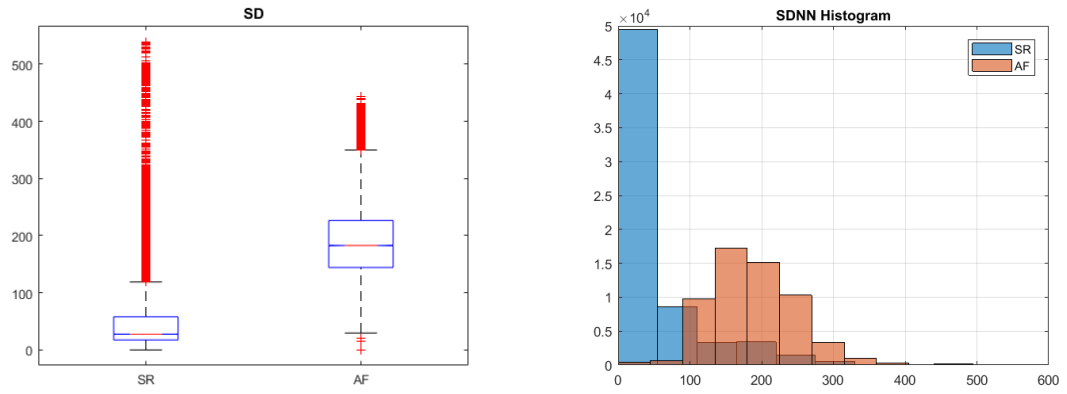
Figure 3.9. *Comparison of statistical values of SDNN between normal SR and AF subjects*
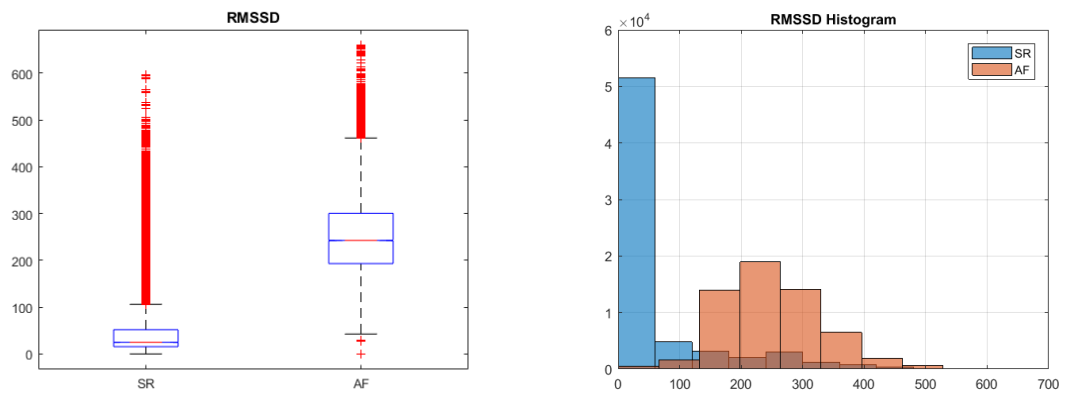
Figure 3.10. *Comparison of statistical values of RMSSD between normal SR and AF subjects*
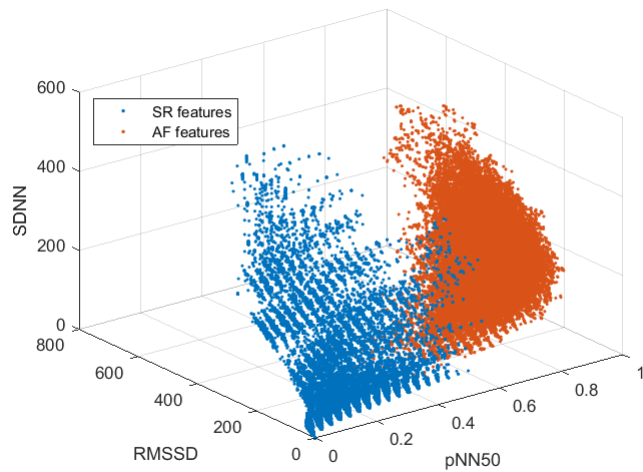
Figure 3.11. *The feature space of pNN50, RMSSD and SDNN features for different arrhythmia classes*
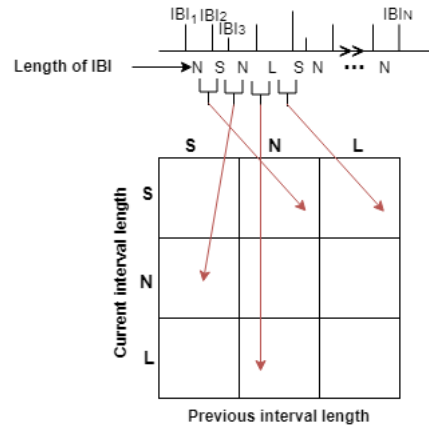
Figure 3.12. *Deriving transition probability matrix from an interbeat interval sequence*

		from		
		Short	Regular	Long
	Short	0	0	0
to	Regular	0	29	0
	Long	0	0	0

Table 3.2. *An example of SR transition statistics*

		from		
		Short	Regular	Long
	Short	1	5	2
to	Regular	7	6	3
	Long	0	4	1

Table 3.3. *An example of AF transition statistics*

The $P \times P$ transition matrix represents the transition probabilities between different states (short, regular and long). Element p_{ij} in this matrix is the probability of moving from state i to state j such that $\sum_{j=1}^{n} p_{ij} = 1, \quad \forall i$ which means that each row of transition probability matrix must sum to unitary since the elements of a row corresponds to the transition probabilities of a certain state. [66] Figure 3.12 illustrates the process of obtaining the transition probability matrix. Since a sliding window of N interbeat-intervals is used to provide this matrix, therefore, there is a one-to-one correspondence between beats and matrices that allows us to label each beat individually rather than in a segment of N beats.

Transition statistics for two selected segments from SR and AF sets are illustrated at Tables 3.2 and 3.3, respectively. Consequently, the corresponding transition probability matrices are calculated as in Tables 3.4 and 3.5.

		from		
		Short	Regular	Long
	Short	0	0	0
to	Regular	0	1	0
	Long	0	0	0

Table 3.4. *An example of SR transition probability matrix*

		from		
		Short	Regular	Long
	Short	0.125	0.625	0.25
to	Regular	0.4375	0.375	0.1875
	Long	0	0.8	0.2

Table 3.5. *An example of AF transition probability matrix*
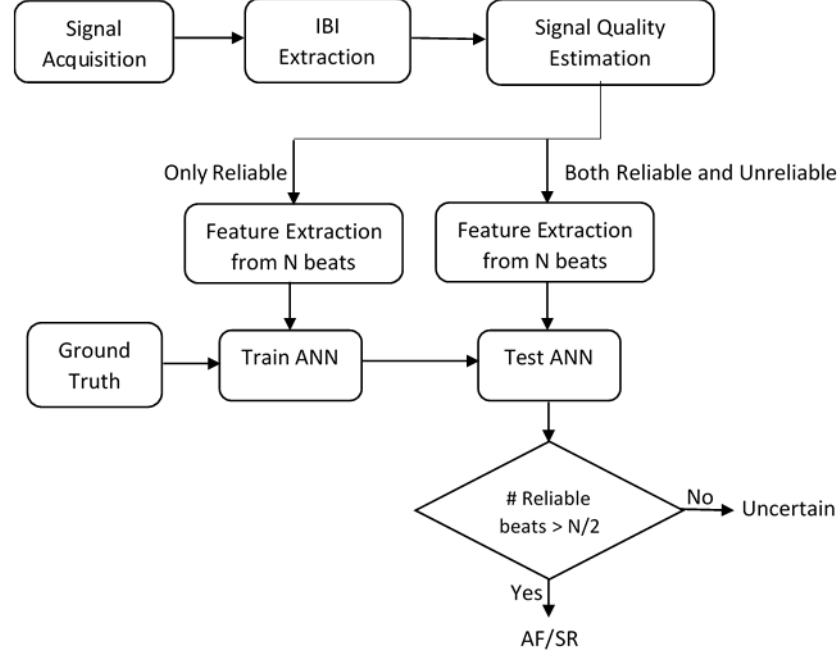
Figure 3.13. *Block diagram of AF detection algorithm*

3.3.4 AF detection using Artificial Neural Networks

The ANN design and evaluation were carried out by using the Matlab 2017a patternnet Toolbox. Given the above features, a multilayer perceptron is designed as a classifier to discriminate between AF and SR. The developed MLP is referred to as ANN in this thesis. An ANN architecture including 12 input units (9 inputs for transition matrix and 3 for the other features), 10 hidden layers and one output neuron is proposed in this work.

In order to train the network, only reliable or correct beats obtained from signal quality estimation algorithm were used. However, to test the network IBIs of all beats (both correct and artefact beats) were imported to the network. The mean IBI is calculated such that the Eq. 3.1 is updated if the beat is reliable (or classified as the correct beat), otherwise the mean IBI will remain unchanged. The final decision of the classifier is based on the output of the signal quality estimation algorithm. If more than the half of beats within each window were artefact (unreliable), the output decision will be assigned as uncertain.

If the number of reliable beats is more than the half of the beats within each window, instead of using a predefined probability threshold to process the output of the ANN, first an N-point moving average is utilized and then the obtained value is compared to two thresholds T_1 and T_2. If this value surpasses the higher threshold, then the beat corresponding to the given interval is labeled as AF. If the value is less than the lower threshold then the beat is labeled as SR. Otherwise, if the value is between the two thresholds, the label remains unchanged similar to the previous assigned label. The reason is due to this fact that the heart tends to hold its current rhythm, instead of changing it frequently [66].

4. RESULTS AND DISCUSSION

In this thesis an algorithm was developed and validated to detect atrial fibrillation with a high performance by only using PPG and accelerometry signals obtained from the PulseOn wrist-band optical heart rate monitor. In order to assess the generalization performance of the AF detection algorithm, k-fold cross-validation was applied, such that each time one of the 29 subsets were used as the test set and the remaining 28 subsets were put together to form a training set. In this study, the number of pulses that is required to trigger an AF detection was set to be as minimum as possible. The applied length of window is 30 beats to extract features.

This chapter summarizes the results of this study and their relevance and includes the most important results, the possible causes of the errors and discuss about the reliability of the results.

4.1 Comparison of RRI-based, IBI-based and IBI-SQE-based

Table 4.1 summarizes the information about the number of detected beats and pulses in ECG recording and PPG measurement, respectively. Total amount of heartbeats acquired from ECG recording after excluding ectopic beats were 202227, in which 98777 were from AF sets and 103450 from SR sets. Total amount of detected pulses from PulseOn wrist-band were 183188 with 90870 pulses in AF sets and 92318 in SR sets, respectively. According to the signal quality estimator 32933 of AF pulses and 25522 of SR pulses (totally 58455 pulses, 32% of the whole data) were assigned unreliable due to the low quality of their PPG pulses that can be the result of hand movement or other kinds of artefacts.

Table 4.1. *The number of AF/SR beats in the ECG measurement and AF/SR pulses in acquired PPG*

	AF(#)	SR(#)	Total(#)	Artefacts(#)
ECG	98777	103450	202227	-
PPG	90870	92318	183188	58455

Each input feature was fed separately to the ANN classifier, to investigate the accuracy of features for AF detection, independently. By pooling the outcome from all test sets across all 29 folds, the overall sensitivity, specificity, accuracy and area under curve for each feature separately are summarized in Table 4.2. The best AF sensitivity is achieved from RMSSD and the best achieved specificity is related to the feature of transition matrix. As it was discussed earlier, and based on the represented values in this table the overall

performance for a single feature is not accurate enough and using combined features allows the classifier to take advantage of their discrimination abilities to obtain potentially better results.

Table 4.2. *Patine-by-patient sensitivity, specificity, accuracy and area under curve for each feature separately*

	SEN(%)	**SPE(%)**	**ACC(%)**	**AUC**
pNN50	97.18 ± 5.91	94.11 ± 7.45	95.63	0.887
SDNN	95.82 ± 6.35	59.19 ± 27.54	78.62	0.687
RMSSD	97.34 ± 4.23	59.44 ± 28.74	67.45	0.688
Transition Matrix	91.67 ± 10.08	96.21 ± 4.83	93.72	0.945
All features	97.87 ± 4.84	95.83 ± 6.31	96.93	0.977

Since this study utilizes simultaneously recorded ECG signal, it is possible to compare the results of the proposed method with features acquired from RR intervals that are less prone to be corrupted by the artefacts compared with the IBIs. From now on, all of the four input features are fed to the ANN classifier. Beat-by-beat analysis of the RRI-based classification is shown in Table 4.3. By applying the ANN classifier on the features acquired from RR intervals 99.9% SR beats were classified correctly, while 92.8% AF beats were classified as AF and 7.2% were classified wrongly as SR beats.

The ANN classification results for the IBI-based scenario in which all beats are being considered, be it reliable or not are also shown in Table 4.3 under the section "IBI-based". 97.6% of AF pulses were classified correctly as AF and 96.26% SR pulses were classified as SR. Compared to the RRI-based, the percentages of beats classified as AF (both in SR and AF sets) have increased, and this was being expected because the artefact beats produce irregularity in IBI sequences, therefore the algorithm recognizes these irregularities as AF events.

Table 4.3. *Beat-by-beat analysis of classifier outputs, for two scenarios: RRI-based in which features are derived from ECG and IBI-based in which features are derived from IBI sequence but without considering the reliability of the IBIs*

	RRI-based		**IBI-based**	
	AF(#)	**SR(#)**	**AF(#)**	**SR(#)**
AF sets	91721	7056	88698	2172
SR sets	134	103316	3451	88867

Table 4.4 shows the beat-by-beat analysis for the proposed IBI-SQE-based method in this thesis which considers the reliability of the beats and if more than the half of the beats inside the window are artefacts, then the output is labeled as uncertain.

The patient-by-patient sensitivity, specificity and area under curve across all 29 folds, for these three scenarios are shown in Table 4.5. The last row shows the percentages of the data that is classified in each scenario. The ECG signals are less affected by artefacts normally

Table 4.4. *Beat-by-beat analysis of classifier outputs, for IBI-SQE-based: features derived from IBI with considering the quality estimation of the pulses*

	AF(#)	Uncertain(#)	SR(#)
AF	62303	28085	482
SR	325	20436	71557

Table 4.5. *Comparison of patient-by-patient performance of ANN classifier for RRI-based, IBI-based (not considering the pulses reliability) and IBI-SQE-based (with applying signal quality estimation)*

	RRI-based	IBI-based	IBI-SQE-based
SEN (%)	93.83 ± 18.65	97.87 ± 4.84	95.12 ± 15.32
SPE (%)	99.84 ± 0.59	95.83 ± 6.31	99.4 ± 2.2
AUC	0.9896	0.9767	0.9983
Classified data (%)	100	100	73.5

present in ambulatory conditions. Most of the false positives in IBI-based scenario are originated from those pulse waveforms that were corrupted by hand movement artefacts. Increasing the false positives may reduce the specificity when the device is used in real-life situation due to the presence of more motion artefacts. In order to overcome this limitation, the IBI quality needed to be checked before running the AF detection algorithm to exclude those segments that are corrupted or too noisy.

The reason for higher value of sensitivity (the ability to detect true AF events) of "IBI-based" approach compared to "IBI-SQE-based" is that the artefact beats (32% of the data) produce irregularity in IBI sequences and the algorithm recognizes these irregularities as AF events. In addition to that, further investigation revealed that the reason of relatively low mean and very high standard deviation values of sensitivity for the "RRI-based" and "IBI-SQE-based" is related to one dataset acquired from a patient (set #23) which were classified as AF since there were not any visible p-wave in the ECG signal. However, compare to the other AF patients there were much more regularity in the IBI sequence of this patient that made it similar to the SR cases. One part of the measured waves from this patient is shown in Figure 4.1.

The patient-by-patient performance metrics of these three scenarios after excluding the data set #23 from the experiment are shown in Table 4.6. The sensitivity, specificity and AUC of both methods RRI-based and IBI-based were improved and closely match together.

According to the signal quality estimator applied in this study 30% of the detected pulses were classified as unreliable or artefact beats (after excluding the problematic data set). However, based on the proposed algorithm, only 23.3% data were labeled as uncertain due to the low quality. The reason of this achievement is that even in the presence of limited number of artefact pulses within each window if there is still a sufficient number of reliable IBIs, the classifier is able to accurately classify the pulse as AF or SR. However,
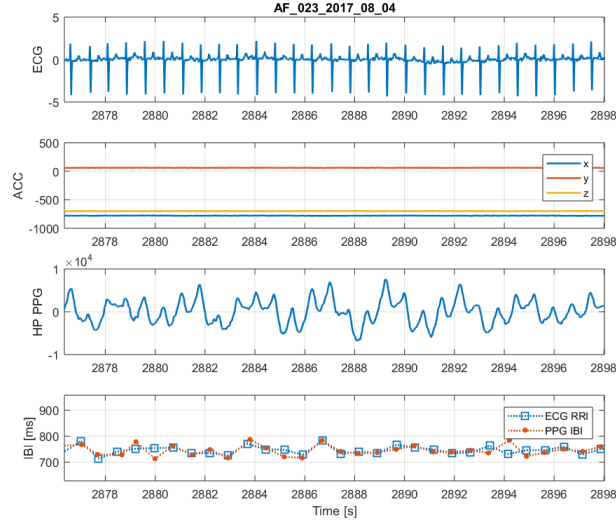
Figure 4.1. *Measured waveforms of the AF patient #23 with regular IBI which seems to belong to the SR group*

Table 4.6. *Comparison of patient-by-patient performance of ANN classifier for RRI-based, IBI-based and IBI-SQE-based method after excluding the problematic data set #23*

	RRI-based	**IBI-based**	**IBI-SQE-based**
SEN (%)	98.8 ± 2.25	98 ± 1.95	98.94 ± 3.61
SPE (%)	99.7 ± 0.9	95.58 ± 6.54	99.54 ± 1.2
AUC	0.999	0.984	0.998
Classified data (%)	100	100	76.7

if more than half of the pulses from the window are artefacts, it is better to not provide wrong decision and the output of the classifier is set to uncertain. $78 \pm 22.6\%$ of the SR group were labeled as AF/SR and $75.2 \pm 18\%$ of the AF group were labeled as AF/SR. These numbers are achieved on a patient-by-patient basis. Therefore, in order to have the improved performance in the proposed IBI-SQE-based method compared to the scenario of IBI-based, it is inevitable that some of the data is discarded and finally labeled as uncertain. However, discarding a part of the data and not assigning an output is naturally preferred over to having a wrong decision when the data is not accurate enough.

4.2 Examining the IBI-SQE-based method on manually mixed dataset

The patients recruited in this study were either with continuous SR or continuous AF. But the proposed method is not limited to detect only permanent AF and can be applied in the same way for paroxysmal AF patients. The only restriction is that the AF episodes preferably need to last for at least 30 beats. The proposed algorithm has been tested on artificially concatenated data from both sets. An example is shown in Figure 4.2. Two AF
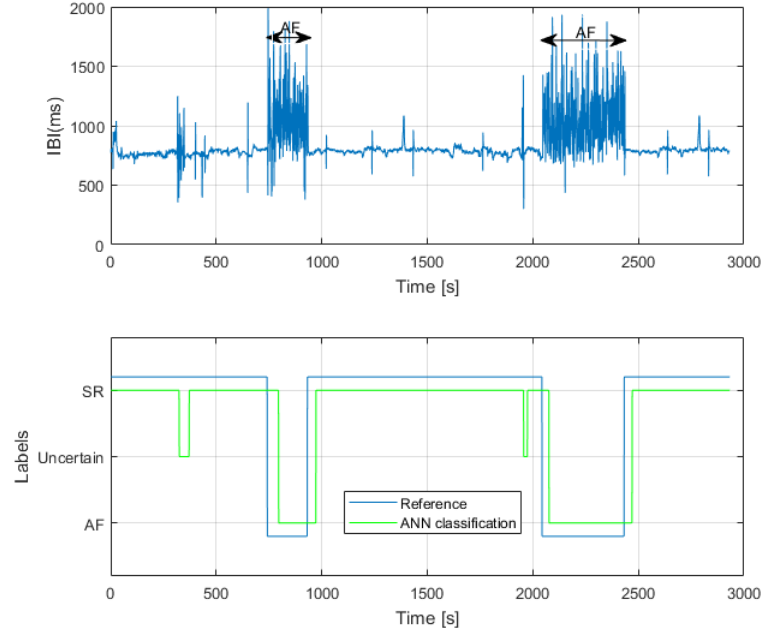
Figure 4.2. *ANN classification results using IBI-SQE-based algorithm for an artificially mixed data set (artificial paroxysmal AF data*

episodes have been inserted in an SR data set at times 700 - 900 s and 2000 - 2500 s. It is evident that the proposed algorithm is able to detect the AF episodes correctly. There are some irregular IBIs due to the artefacts present in the areas which the pulses are from an SR set (e.g. at time 350 - 400 s). These irregularities produce a pattern similar to the AF sets. However, as it can be seen from Figure 4.2 the output of the proposed algorithm is uncertain in this period and it is not wrongly classified as AF.

4.3 Comparison of different classifiers

In AF detection problem, if the feature space of AF and SR is not perfectly separable for a large number of measurements, non-linear classifications such as ANN, QDA and SVM with nonlinear kernel can be performed. Typically, in high-dimensional spaces, data can more easily be separated linearly and lead to better generalization than is achieved in low dimensional spaces.

The performance of the ANN classifier using IBI-SQE-based algorithm was compared to two linear (LDA and SVM with linear kernel) and two nonlinear (QDA and SVM with Gaussian kernel) classifiers. The input features of these classifiers are similar to the applied features for ANN including: pNN50, RMSSD, SDNN and transition matrix derived from IBI time series. Cross-validation was also performed to assess the reproducibility of these classifiers.

There are some performance measures that are widely used in the literature to compare

different classifiers, including the accuracy rate, precision and recall, ROC area under curve and the time spent for classification [67]. In this study, since it is a more practical analysis of the classifiers, the sensitivity, specificity and AUC are compared together.

Table 4.7 represents these metrics for four classifiers and Figures 4.3 and 4.4 compare the boxplots of the performance of these classifiers, in terms of sensitivity, specificity and AUC. It should be noted that these results are achieved after excluding data set #23.

Table 4.7. *Classification performances of different classifiers on a patient-by-patient basis. The results are obtained after excluding data set #23.*

	ANN	QDA	SVM(Gaussian)	LDA	SVM(Linear)
SEN (%)	98.94 ± 3.61	96.5 ± 7.21	98.67 ± 3.5	99.95 ± 0.18	99.94 ± 0.18
SPE (%)	99.54 ± 1.2	96.9 ± 5.02	99.7 ± 1.11	99.4 ± 1.15	99.97 ± 0.12
AUC	0.998	0.987	0.997	0.999	0.999

The comparative analysis among ANN, SVM (linear and gaussian kernels), LDA and QDA pointed out that, for classification performance, linear SVM achieved the highest values closely followed by LDA. Then nonlinear classifiers of ANN, SVM with Gaussian kernel achieved almost similar results. The QDA algorithm produces the worst performance compared to other classifiers.

The results show that the linear classifiers provide slightly more accurate classification compared to ANN classifier. One reason can be the limited number of recorded pulses. ANN usually needs a large amount of data to be trained accurately without being trapped in overfitting. Furthermore, linear classifiers have less complexity compared to the nonlinear classifiers. Therefore, they are more suitable for practical implementation such that the final decision making can be as a part of the application embedded in the device. In addition to that, it can simplify the real-time detection of AF events, especially for the occurrences of the paroxysmal AF.

Totally, these results demonstrate that wrist-worn PPG device is accurate enough to detect AF and has the ability to become an alternative for the ECG-based AF detection that is cumbersome and more expensive.
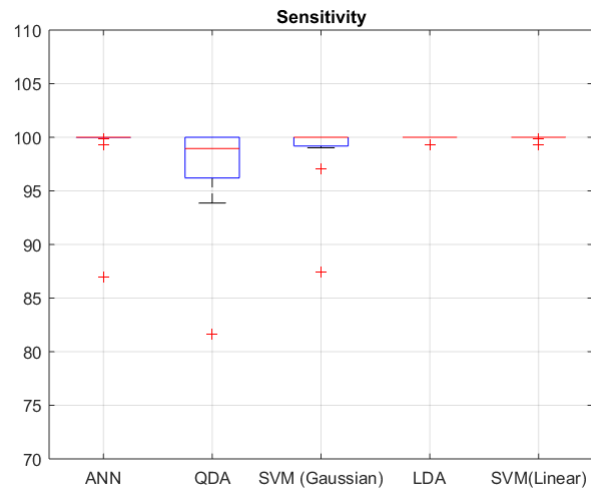
Figure 4.3. *Comparison of boxplots of sensitivity for ANN, SVM, LDA and QDA*
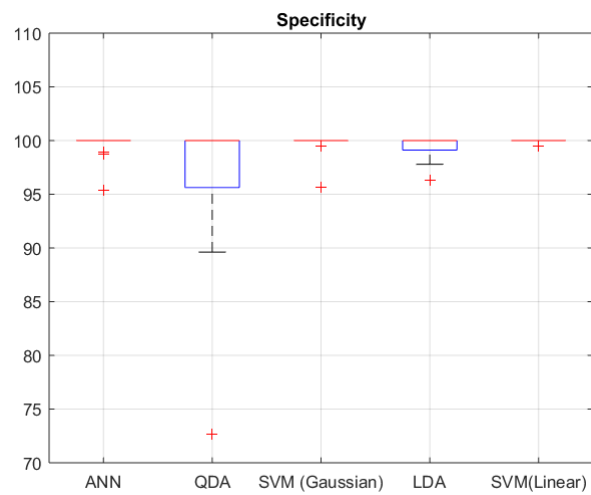

Figure 4.4. *Comparison of boxplots of specificity for ANN, SVM, LDA and QDA*

5. CONCLUSIONS AND FUTURE WORK

Currently, AF monitoring is an expensive and burdensome procedure. Rapidly increasing of population of at-risk individuals for this potentially life-threatening arrhythmia due to the aging and obesity epidemic, demand a low-cost, low-burden and more reliably available AF detection technology. The development of new long-term ambulatory monitoring technologies enables unobtrusive patient monitoring for days and weeks instead of hours. These technologies along with a successful algorithm provide patients with access to a convenient wrist-band device and may enable superior sensitivity in AF detection out of hospital. However, it is important to recognize that these tools are validated in extensive number of clinical trials. Therefore more tests are needed.

The main goal of this thesis was the development and validation of a machine learning based approach for AF detection using only data acquired from PPG signal. The objective of this research was accomplished very well by the development of a multilayer perceptron that is an ANN based algorithm. It is shown that AF can be accurately detected using only pulsatile PPG data acquired from PulseOn optical heart rate monitor. One key challenge the study included was the reliability of the PPG pulses that was addressed by application of signal quality estimation. The achieved performance was comparable to that of an ECG-based algorithm.

The developed algorithm was tested also on a manually mixed dataset to check the capability of the proposed method to correctly recognize the isolated AF episodes. Although, the results were satisfying, the performance analysis remains to be proven for real paroxysmal AF subjects.

The proposed ANN classifier was also compared to other nonlinear and linear classifiers including SVM (with linear and Gaussian kernels), LDA and QDA. With the current amount of data sets the linear classifiers showed slightly better performance. Although the cross-validation experiments have proved very promising, collecting more data from patients and checking the reproducibility of these results on a larger cohort study is needed.

This study has important limitations that have to be pointed out. The patients were stationary during the recording. Also, the subjects were hospitalized patients therefore the SR IBIs were very stable. It would be different if the subjects were younger and in better shape, even athletes and had much higher HRV.

An interesting direction for future work is the application of Recurrent Neural Networks (RNN) particularly those using Long Short-Term Memory (LSTM) which are popular models for learning from sequence data. Deep neural network architecture have achieved

great success in effectively capture long term temporal dependencies in time series [68]. Since the cardiac interbeat-intervals provide a time series, therefore the AF detection problem can be a battlefield for LSTM to join.

Another possible improvement is adding the diagnosis of other type of arrhythmias such as atrial flutter, ventricular and supraventricular arrhythmias. The main challenge is the confusion of supraventricular tachycardia, atrial flutter and atrial fibrillation which makes sense given that all of them are atrial arrhythmias.

At the end, further research work is still needed to explore the feasibility and acceptability of the proposed technique under realistic device utilization scenarios for 24/7 monitoring during daily activities and sleep.

REFERENCES

[1]   E. Marieb, *Essentials of Human Anatomy and Physiology (10th)*, vol. 10. San Francisco, CA: Benjamin Cummings/Pearson Education, 2011.

[2]   M. Lemay, M. Bertschi, J. Sola, P. Renevey, J. Parak, and I. Korhonen, "Application of optical heart rate monitoring," in *Wearable Sensors*, pp. 105–129, Elsevier, 2015. Available: https://pdfs.semanticscholar.org/c999/ f808598037ca5b3ee064b38310d9b59f2bf9.pdf "[Online; accessed 15-Dec-2017]".

[3]   T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable photoplethysmographic sensors—past and present," *Electronics*, vol. 3, no. 2, pp. 282–302, 2014. Available: http://www.mdpi.com/2079-9292/3/2/282/htm "[Online; accessed 15-Dec-2017]".

[4]   S. S. Haykin, S. S. Haykin, S. S. Haykin, and S. S. Haykin, *Neural networks and learning machines*, vol. 3. Pearson Upper Saddle River, NJ, USA:, 2009. Available: http://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009. pdf "[Online; accessed 15-Dec-2017]".

[5]   PulseOn Oy. Available: http://www.pulseon.fi "[Online; accessed 10-Jan-2018]".

[6]   A. Tarniceriu, J. Harju, Z. Rezaei Yousefi, A. Vehkaoja, J. Parak, A. Yli-Hankala, and I. Korhonen, "The accuracy of atrial fibrillation detection from wrist photoplethysmography. a study on post-operative patients," in *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'18) at the Honolulu, HI, USA*, IEEE, 2018.

[7]   S. S. Chugh, R. Havmoeller, K. Narayanan, D. Singh, M. Rienstra, E. J. Benjamin, R. F. Gillum, Y.-H. Kim, J. H. McAnulty, Z.-J. Zheng, *et al.*, "Worldwide epidemiology of atrial fibrillation: a global burden of disease 2010 study," *Circulation*, pp. CIRCULATIONAHA–113, 2013.

[8]   L. Friberg and L. Bergfeldt, "Atrial fibrillation prevalence revisited," *Journal of internal medicine*, vol. 274, no. 5, pp. 461–468, 2013.

[9]   J. Heeringa, D. A. van der Kuip, A. Hofman, J. A. Kors, G. van Herpen, B. H. C. Stricker, T. Stijnen, G. Y. Lip, and J. C. Witteman, "Prevalence, incidence and lifetime risk of atrial fibrillation: the rotterdam study," *European heart journal*, vol. 27, no. 8, pp. 949–953, 2006.

[10] B. Olshansky, M. K. Chung, S. M. Pogwizd, and N. Goldschlager, *Arrhythmia Essentials E-Book*. Elsevier Health Sciences, 2016.

[11]   L. Ferranti and R. Laureanti, "Atrial fibrillation detection in ppg signal recorded through a wristband device," 2015. Available: https://www.politesi.polimi.it/bitstream/10589/115204/3/ AtrialFibrillationdetectioninPPGsignalrecordedthroughawristbanddevice.pdf "[Online; accessed 20-May-2018]".

[12]   L. Krivoshei, S. Weber, T. Burkard, A. Maseli, N. Brasier, M. Kühne, D. Conen, T. Huebner, A. Seeck, and J. Eckstein, "Smart detection of atrial fibrillation," *Europace*, vol. 19, no. 5, pp. 753–757, 2016.

[13]   S. Fallet, M. Lemay, P. Renevey, C. Leupi, E. Pruvot, and J.-M. Vesin, "An adaptive organization index to characterize atrial fibrillation using wrist-type photoplethysmographic signals," in *Computing in Cardiology Conference (CinC), 2016*, pp. 337–340, IEEE, 2016.

[14]   S. Nemati, M. M. Ghassemi, V. Ambai, N. Isakadze, O. Levantsevych, A. Shah, and G. D. Clifford, "Monitoring and detecting atrial fibrillation using wearable technology," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pp. 3394–3397, IEEE, 2016.

[15]   A. G. Bonomi, F. Schipper, L. M. Eerikäinen, J. Margarito, R. M. Aarts, S. Babaeizadeh, H. M. de Morree, and L. Dekker, "Atrial fibrillation detection using photo-plethysmography and acceleration data at the wrist," in *Computing in Cardiology Conference (CinC), 2016*, pp. 277–280, IEEE, 2016.

[16]   J. Lee, B. A. Reyes, D. D. McManus, O. Maitas, and K. H. Chon, "Atrial fibrillation detection using an iphone 4s," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 203–206, 2013.

[17]   J. W. Chong, N. Esa, D. D. McManus, and K. H. Chon, "Arrhythmia discrimination using a smart phone," *IEEE journal of biomedical and health informatics*, vol. 19, no. 3, pp. 815–824, 2015.

[18]   P.-H. Chan, C.-K. Wong, Y. C. Poh, L. Pun, W. W.-C. Leung, Y.-F. Wong, M. M.-Y. Wong, M.-Z. Poh, D. W.-S. Chu, and C.-W. Siu, "Diagnostic performance of a smartphone-based photoplethysmographic application for atrial fibrillation screening in a primary care setting," *Journal of the American Heart Association*, vol. 5, no. 7, p. e003428, 2016.

[19]   D. D. McManus, J. Lee, O. Maitas, N. Esa, R. Pidikiti, A. Carlucci, J. Harrington, E. Mick, and K. H. Chon, "A novel application for the detection of an irregular pulse using an iphone 4s in patients with atrial fibrillation," *Heart Rhythm*, vol. 10, no. 3, pp. 315–319, 2013.

[20]   T. Schäck, Y. S. Harb, M. Muma, and A. M. Zoubir, "Computationally efficient algorithm for photoplethysmography-based atrial fibrillation detection using smart-

phones," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, pp. 104–108, IEEE, 2017.

[21] M. Landén, B. Wennerblom, H. Tygesen, K. Modigh, K. Sörvik, C. Ysander, A. Ekman, H. Nissbrandt, M. Olsson, and E. Eriksson, "Heart rate variability in premenstrual dysphoric disorder," *Psychoneuroendocrinology*, vol. 29, no. 6, pp. 733–740, 2004.

[22] "Arrhythmia." National Heart, Lung, and Blood Institute, 2017. Available: https://www.mayoclinic.org/diseases-conditions/atrial-fibrillation/ symptoms-causes/syc-20350624 "[Online; accessed 19-Jan-2018]".

[23] E. N. Prystowsky, B. J. Padanilam, and R. I. Fogel, "Treatment of atrial fibrillation," *Jama*, vol. 314, no. 3, pp. 278–288, 2015.

[24] H. Oral, S. Veerareddy, E. Good, B. Hall, P. Cheung, K. Tamirisa, J. Han, J. Fortino, A. Chugh, F. Bogun, *et al.*, "Prevalence of asymptomatic recurrences of atrial fibrillation after successful radiofrequency catheter ablation," *Journal of cardiovascular electrophysiology*, vol. 15, no. 8, pp. 920–924, 2004.

[25] C. W. Israel, G. Grönefeld, J. R. Ehrlich, Y.-G. Li, and S. H. Hohnloser, "Long-term risk of recurrent atrial fibrillation as documented by an implantable monitoring device: implications for optimal patient care," *Journal of the American College of Cardiology*, vol. 43, no. 1, pp. 47–52, 2004.

[26] G. Quirino, M. Giammaria, G. Corbucci, P. Pistelli, E. Turri, A. Mazza, A. Perucca, C. Checchinato, M. Dalmasso, and S. S. Barold, "Diagnosis of paroxysmal atrial fibrillation in patients with implanted pacemakers: relationship to symptoms and other variables," *Pacing and clinical electrophysiology*, vol. 32, no. 1, pp. 91–98, 2009.

[27] E. J. d. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "Ecg-based heartbeat classification for arrhythmia detection: A survey," *Computer methods and programs in biomedicine*, vol. 127, pp. 144–164, 2016.

[28] R. Kohno, H. Abe, and D. G. Benditt, "Ambulatory electrocardiogram monitoring devices for evaluating transient loss of consciousness or other related symptoms," *Journal of arrhythmia*, vol. 33, no. 6, pp. 583–589, 2017.

[29] F. Giada, E. Bertaglia, B. Reimers, D. Noventa, and A. Raviele, "Current and emerging indications for implantable cardiac monitors," *Pacing and Clinical Electrophysiology*, vol. 35, no. 9, pp. 1169–1178, 2012.

[30] A. Haavikko, "Evaluation of performance of an optical heart rate sensor," 2014. Available: https://dspace.cc.tut.fi/dpub/handle/123456789/22628 "[Online; accessed 20-May-2018]".

[31] "Beat2phone ecg meter." VITAL SIGNUM OY. Available: www.beat2phone.com "[Online; accessed 20-May-2018]".

[32] "Qardiocore ecg monitor." Qardio. Inc. Available: www.getqardio.com "[Online; accessed 20-May-2018]".

[33] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.

[34] J. G. Webster, *Design of pulse oximeters*. CRC Press, 1997.

[35] T. Conroy, J. H. Guzman, B. Hall, G. Tsouri, and J.-P. Couderc, "Detection of atrial fibrillation using an earlobe photoplethysmographic sensor," *Physiological measurement*, vol. 38, no. 10, p. 1906, 2017.

[36] E. Gil, J. M. Vergara, and P. 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*, vol. 3, no. 3, pp. 267–277, 2008.

[37] "Empatica e4." Empatica Inc. Available: http://www.empatica.com "[Online; accessed 10-Jan-2018]".

[38] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[39] T. M. Mitchell *et al.*, "Machine learning. 1997," *Burr Ridge, IL: McGraw Hill*, vol. 45, no. 37, pp. 870–877, 1997.

[40] M. B. Christopher, *Pattern recognition and machine learning*. Springer-Verlag New York, 2016.

[41] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd international conference on Machine learning*, pp. 161–168, ACM, 2006.

[42] M. F. A. Hady and F. Schwenker, "Semi-supervised learning," pp. 215–239, Springer, 2013.

[43] F. Rosenblatt, "Principles of neurodynamics. perceptrons and the theory of brain mechanisms," tech. rep., CORNELL AERONAUTICAL LAB INC BUFFALO NY, 1961.

[44] J. S. Zirilli, *Financial prediction using neural networks*. International Thomson Computer Press, 1996.

[45]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[46]   A. Waibel, "Modular construction of time-delay neural networks for speech recognition," *Neural computation*, vol. 1, no. 1, pp. 39–46, 1989.

[47]   N. Kanda, R. Takeda, and Y. Obuchi, "Elastic spectral distortion for low resource speech recognition with deep neural networks," in *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*, pp. 309–314, IEEE, 2013.

[48]   B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond.* MIT press, 2001.

[49]   U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and biological engineering and computing*, vol. 44, no. 12, pp. 1031–1051, 2006.

[50]   A. J. Camm, M. Malik, J. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. Kleiger, *et al.*, "Heart rate variability. standards of measurement, physiological interpretation, and clinical use," *European heart journal*, vol. 17, no. 3, pp. 354–381, 1996.

[51]   M. Mohebbi and H. Ghassemian, "Detection of atrial fibrillation episodes using svm," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 177–180, IEEE, 2008.

[52]   J. Lee, D. McManus, and K. Chon, "Atrial fibrillation detection using time-varying coherence function and shannon entropy," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pp. 4685–4688, IEEE, 2011.

[53]   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," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 300, no. 1, pp. H319–H325, 2010.

[54]   F. Shaffer and J. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, p. 258, 2017.

[55]   S. Arlot, A. Celisse, *et al.*, "A survey of cross-validation procedures for model selection," *Statistics surveys*, vol. 4, pp. 40–79, 2010.

[56]   P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of database systems*, pp. 532–538, Springer, 2009.

[57]   Y. A. E.-W. Skaik, "Understanding and using sensitivity, specificity and predictive values.," 2008.

[58]   T. A. Lasko, J. G. Bhagwat, K. H. Zou, and L. Ohno-Machado, "The use of receiver operating characteristic curves in biomedical informatics," *Journal of biomedical informatics*, vol. 38, no. 5, pp. 404–415, 2005.

[59]   GE Healthcate Ltd. Available: http://www.gehealthcare.com "[Online; accessed 10-Jan-2018]".

[60]   Kubios Ltd. Available: http://www.kubios.com "[Online; accessed 10-Jan-2018]".

[61]   J. Parak, A. Tarniceriu, P. Renevey, M. Bertschi, R. Delgado-Gonzalo, and I. Korhonen, "Evaluation of the beat-to-beat detection accuracy of pulseon wearable optical heart rate monitor," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 8099–8102, IEEE, 2015.

[62]   J. Pietilä, S. Mehrang, J. Tolonen, E. Helander, H. Jimison, M. Pavel, and I. Korhonen, "Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities," in *EMBEC & NBC 2017*, pp. 145–148, Springer, 2017.

[63]   M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current cardiology reviews*, vol. 8, no. 1, pp. 14–25, 2012.

[64]   J. A. Sukor, S. Redmond, and N. Lovell, "Signal quality measures for pulse oximetry through waveform morphology analysis," *Physiological measurement*, vol. 32, no. 3, p. 369, 2011.

[65]   C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, and L. Tarassenko, "Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring," *IEEE journal of biomedical and health informatics*, vol. 19, no. 3, pp. 832–838, 2015.

[66]   S. G. Artis, R. Mark, and G. Moody, "Detection of atrial fibrillation using artificial neural networks," in *Computers in Cardiology 1991, Proceedings.*, pp. 173–176, IEEE, 1991.

[67]   D. R. Amancio, C. H. Comin, D. Casanova, G. Travieso, O. M. Bruno, F. A. Rodrigues, and L. da Fontoura Costa, "A systematic comparison of supervised classifiers," *PloS one*, vol. 9, no. 4, p. e94137, 2014.

[68]   M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," *arXiv preprint arXiv:1710.06122*, 2017.