

# Detection of Atrial Fibrillation Episodes Using a Wristband Device

Valentina D.A. Corino<sup>1</sup> & Rita Laureanti<sup>1</sup> & Lorenzo Ferranti<sup>1</sup>  
& Giorgio Scarpini<sup>2</sup> & Federico Lombardi<sup>2</sup> & Luca T. Mainardi<sup>1</sup>

<sup>1</sup>Department of Electronic, Information and Bioengineering, Politecnico di Milano, Milan, Italy

<sup>2</sup> Fondazione IRCCS Ca Granda Ospedale Maggiore Policlinico, U.O.C. di Malattie Cardiovascolari, Università degli Studi di Milano, Dipartimento di Scienze Cliniche e di Comunità, Milan, Italy

E-mail: [valentina.corino@polimi.it](mailto:valentina.corino@polimi.it)

**Abstract.** Undiagnosed atrial fibrillation (AF) patients are at high risk of cardioembolic stroke or other complications. The aim of this study was to analyze the blood volume pulse (BVP) signals obtained from a wristband device and develop an algorithm for discriminating AF from normal sinus rhythm (NSR) or from other arrhythmias (ARR). Thirty patients with AF, 9 with ARR and 31 in NSR were included in the study. The recordings were obtained at rest from Empatica E4 wristband device and lasted 10 minutes. The analysis, on a two-minute segment, included spectral, variability and irregularity analysis performed on the inter-diastolic interval series, and similarity analysis performed on the BVP signal. Variability parameters were the highest in AF, the lowest in NSR and intermediate for ARR, as an example pNN50 values were, respectively,  $81\pm 8$ ,  $20\pm 5$ ,  $55\pm 27$  ( $p<0.05$ ). The similarity parameters were the highest in NSR, the lowest in AF and intermediate for ARR, as an example using a threshold for assessing similarity of  $\pi/4$ :  $0.90\pm 0.09$ ,  $0.40\pm 0.20$ ,  $0.58\pm 0.23$ , all  $p<0.05$ . The rhythm classification was preceded by over-sampling (using Synthetic Minority Over-sampling Technique) the class of ARR, being it the smallest class. Then, the features selection was performed (using the Sequential Forward Floating Search algorithm) which identified two variability parameters (pNN70 and pNN40) as the best selection. The classification by the k-nearest neighbor classifier reached an accuracy of about 0.9 for NSR and AF, and 0.8 for ARR. Using pNN70 and pNN40, the specificity for the three rhythms was  $Sp_{NSR}=0.928$ ,  $Sp_{AF}=0.963$ ,  $Sp_{ARR}=0.768$ , while the sensitivity was  $Sp_{NSR}=0.773$ ,  $Sp_{AF}=0.754$ ,  $Sp_{ARR}=0.758$ .

## 1. Introduction

Atrial fibrillation (AF) may be paroxysmal, i.e., it can occur episodically and terminate spontaneously, and is often asymptomatic [24], making the identification of subjects with this rhythm disorder even more difficult. Undiagnosed AF patients are at high risk of cardioembolic stroke or other complications and even an increased risk for death is associated to AF [28, 2, 12]. Thanks to long-term oral anticoagulation therapy about

two-thirds of AF related ischemic strokes can be prevented [11]. However, in about one fourth of patients with stroke or transient ischemic attack, AF is diagnosed only after the event [29], in absence of any preventive therapy. Thus, detection of AF before the occurrence of ischemic stroke is one of the most important and future strategy in preventing stroke, as clinical benefit has been shown by using oral anticoagulation therapy in patients with moderate to high risk of stroke [15]. Opportunistic or systematic screening of the general population has been proposed as a tool for an appropriate identification of patients with asymptomatic AF [4]. Recently, new technologies have been developed to monitor heart rate by means of everyday sensors, as mobile phone cameras [19], thumb ECG [13], and videocameras [14, 8]. These everyday sensors offer the possibility to record and study biosignals providing information which could help in discriminating patients with AF and other arrhythmias (ARR) from subjects in normal sinus rhythm (NSR), without symptoms. One of these devices is the Empatica wristband that can record the blood volume pulse (BVP) by using a photoplethysmographic (PPG) sensor.

The aim of this study was to analyze the BVP signals obtained from the wristband device and develop an algorithm to discriminate AF from NSR, and, more interestingly, from ARR.

## 2. Methods

### 2.1. Patients

We analyzed BVP signals recorded from 70 patients admitted to the Maggiore Policlinico Hospital in Milan, Italy. Thirty patients had persistent AF (AF group), 9 suffered from other arrhythmias (ARR group), and 31 were healthy subjects (NSR group). The group of patients with other arrhythmias included patients with atrial flutter, with many ectopic ventricular beats, with atrial tachycardia and variable conduction. Clinical characteristics are shown in Table 1. All data were recorded between March and November 2015.

The study conforms with the Declaration of Helsinki, and was approved by the Ethics Committee of Maggiore Policlinico Hospital in Milan (Italy). All patients gave their written informed consent for the procedures related to the study.

### 2.2. Protocol

All recordings were performed with the subject in a supine position, at rest. The subject was asked to stay as still as possible to reduce motion artifacts. While the patient layed in a relaxed position, the Empatica E4 wristband was applied on the wrist of the non-dominant arm, with the main part of the device facing upward, in a similar way to a regular wrist watch. Ten-minute recording was acquired for each subject. The Empatica E4 wristband is a wearable wireless device designed for continuous, real-time data acquisition in daily life. The device is equipped with sensors for the registration

**Table 1.** Clinical characteristics of the study population.

|                     | NSR           | AF           | ARR           |
|---------------------|---------------|--------------|---------------|
| N                   | 31            | 30           | 9             |
| Gender(male/female) | 16/15         | 12/18        | 8/1           |
| Age (years)         | 40±17 (27-75) | 76±9 (58-89) | 65±15 (48-92) |
| Diabetes            | 0             | 6            | 2             |
| Hypertension        | 3             | 20           | 4             |
| Beta-blockers       | 2             | 22           | 3             |
| Flecainide          | 4             | 0            | 0             |
| Amiodarone          | 0             | 1            | 1             |
| ACE-inhibitor       | 4             | 12           | 4             |

NSR: normal sinus rhythm, AF: atrial fibrillation, ARR: other arrhythmias.

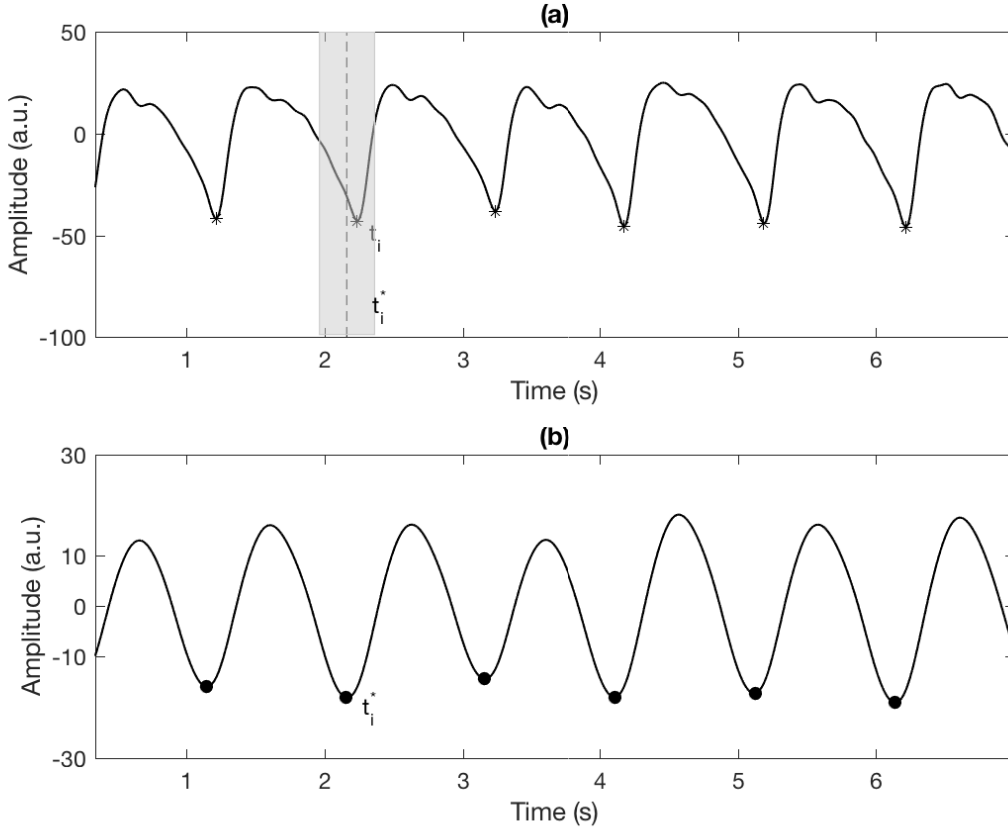
of different biosignals: an electrodermal activity sensor, an infrared thermopile, a 3-axis accelerometer and a PPG sensor which measures the BVP signal. The BVP is sampled at 64 Hz.

### 2.3. Signal preprocessing

The first preprocessing step is the detection and removal of noise: the raw BVP signal may be noisy, due to artifacts, mainly caused by the patient moving his/her arm with the device on, during the measurement. To identify segments corrupted by noise, data from the 3-axis accelerometer were used: the norm of the accelerations on the three axes was computed and the deviation from the acceleration of gravity  $g$  calculated. The absolute value of the deviation was then compared to a threshold. When the threshold was exceeded, the algorithm classified that portion of the signal as noisy and discarded from further analysis. The threshold was empirically set to  $0.07g$ .

The second preprocessing step is the detection of systolic peaks and diastolic minima in the BVP signal. First, the diastolic minima are found by low-pass filtering the BVP signal using a moving-average of 23 samples ( $\sim 0.36$  seconds), acting on the signal as a low-pass filter, similarly as in [22]. Therefore, a smoother signal is obtained where the local minima are easily identifiable. These positions are taken as coarse temporal reference for each diastolic peak, whose timing is then refined by searching the minimum on the original signal in a 0.2-s window around it (see Figure 1). To automatically locate the systolic peaks, first all local maxima in the BVP signal are detected. Then, using the information on the location of the diastolic minima, the first local maximum after each diastolic minimum is defined as a systolic peak (with the restriction that there can be only one peak for every cycle).

From the systolic peaks and diastolic minima, the inter-systolic and the inter-diastolic intervals series are computed, respectively. These series can be used as a surrogate of the RR series, thus providing information on the heart rate of each subject.



**Figure 1.** Diastolic minima detection: the local minima ( $t_i^*$ ) are found on the low-pass filtered signal (b) and reported on the original signal (a), where the minimum ( $t_i$ ) are found in a 0.2-s window (the grey rectangle) centered in  $t_i^*$ , represented by the dashed line.

#### 2.4. Signal characterization

Twenty-four indexes were computed belonging to the following three classes: i) Spectral analysis, ii) Variability and Irregularity analysis, iii) Shape analysis. Parameters from the first two classes were computed on the inter-systolic and inter-diastolic interval series, whereas shape analysis was performed on the BVP signal. All parameters were computed on a two-minute segment for each subject. All subjects had at least one two-minute segment without movement artifacts.

*2.4.1. Spectral Analysis* Power spectral analysis of the inter-systolic and inter-diastolic interval series was performed by means of an AR model:

$$y(n) = \sum_{k=1}^p a_k y(n-k) + e(n) \quad (1)$$

where  $e(n)$  is a gaussian white noise process,  $n$  is the discrete time index,  $p$  is the model order and the  $a_k$ 's are the AR model coefficients. In the  $z$ -domain, the model transfer

function becomes

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} = \frac{z^p}{\prod_{i=1}^p (z - z_i)} \quad (2)$$

where the  $z_i$  are the model poles. In this study, the model coefficients were estimated using the Levinson Durbin algorithm, the Andersons test [16] was used to check the validity of the model, and the model order was selected by Akaike information criterion [1]. Using Cauchy's residue theorem, the AR spectrum,  $P(\omega)$ , can be divided into a sum of  $p$  components [30]. Consequently, the spectrum can be decomposed into bell-shaped curves, named the spectral components. The central frequency  $f_i$  and the power  $P_i$  of the  $i$ -th spectral component can be computed as [21]

$$f_i \approx f_s \frac{\angle(z_i)}{2\pi} \quad P_i = \mu \frac{\sigma_e^2}{f_s} \text{Re} \{ \gamma(z_i) \} \quad (3)$$

where  $\angle(\cdot)$  is the phase expressed in radians, and  $\mu = 2$  for complex pole pairs and  $\mu = 1$  for real ones;  $\sigma_e^2$  is the prediction error variance, and  $\gamma$  is the pole residue.

The spectral decomposition algorithm [21, 3] was used to measure the central frequency and the power of the spectral components falling in the low frequency (LF, 0.03 - 0.15) and high frequency (HF, 0.15 - 0.40 Hz) bands.

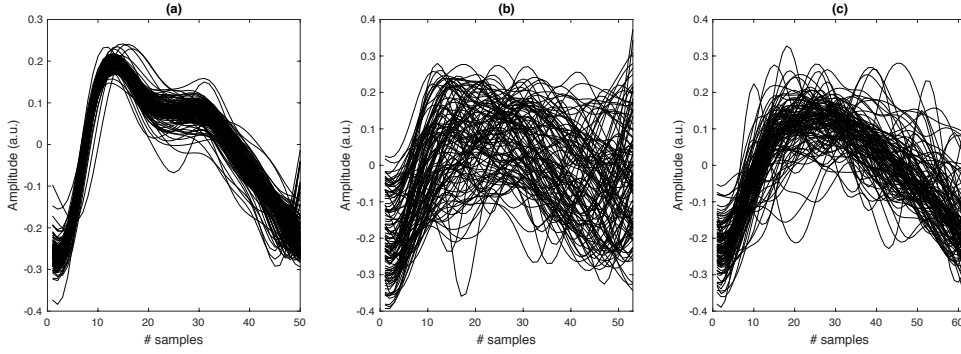
*2.4.2. Variability and Irregularity Analysis* Variability and irregularity quantify different properties: variability is related to the dispersion of data, whereas irregularity is related to the degree of unpredictability of the data fluctuations.

Variability analysis of the inter-diastolic intervals series includes the mean (M), the standard deviation (SD), the root of the mean squared differences of successive intervals, (rMSSD) and the percentage of interval differences of successive intervals greater than  $x$  ms (pNN $x$ , with  $x = [10, 20, \dots, 100]$ ).

Irregularity of the inter-diastolic intervals series was assessed by sample entropy (SampEn), that quantifies the unpredictability of fluctuations. The presence of repetitive patterns of fluctuation in a time series makes it more predictable than a time series in which such patterns are absent. SampEn reflects the likelihood that similar patterns of observations will not be followed by additional similar observations. A time series containing many repetitive patterns, i.e., a regular and predictable series, has a relatively small SampEn; a less predictable, i.e., more irregular process has a higher SampEn. In particular, SampEn is the negative natural logarithm of the conditional probability that two sequences of length  $m$  that match within tolerance  $r$  will also match at the  $m + 1$  length. Defining as A the total number of matches of length  $m + 1$  and B the total number of matches of length  $m$ , SampEn is computed as [26]

$$\text{SampEn} = -\ln(A/B) = -\ln(A) + \ln(B) \quad (4)$$

In this study,  $m$  was used equal to 1 and 2, while  $r$  equal to 0.25 times the standard deviation of the series, as commonly used [?].



**Figure 2.** Example of normalized pulses for three patients in (a) NSR, (b) AF and (c) ARR.

*2.4.3. Shape Analysis* To assess wave similarity, each wave is represented as a point of the  $p$ -dimensional real space, the normalized waves are points belonging to the  $p$ -dimensional unitary sphere. Hence, the morphological dissimilarity between two waves is evaluated by using the standard metric of the sphere to compute their distance [10]

$$D_{i,j} = \arccos(\mathbf{w}_i^N \cdot \mathbf{w}_j^N) \quad (5)$$

where  $\mathbf{w}_i^N$  and  $\mathbf{w}_j^N$  represent the  $i$ -th and  $j$ -th normalized waves, i.e.,  $\mathbf{w}_i^N = \mathbf{w}_i / \|\mathbf{w}_i\|$  and  $(\cdot)$  denotes the scalar product. A measure of similarity between waves is obtained by calculating the relative number of similar pairs of waves in the recording. The similarity depends on the threshold  $\epsilon$  used in evaluating the similarity, that is, two waves are considered to be similar when their distance is lower than  $\epsilon$ . In this study, one pulse in the BVP signal is considered as wave and five different thresholds are tested in this study:  $\epsilon = [\pi/2, \pi/4, \pi/8, \pi/16, \pi/32]$ , defining  $\text{Sim}_1, \text{Sim}_2, \text{Sim}_3, \text{Sim}_4, \text{Sim}_5$ . As an example, in Figure2 the similarity between pulses in the three rhythms is shown. It can be observed that in NSR the waves are very similar, whereas during AF the waves are more different, and they are in between for the patient with arrhythmias.

### 2.5. Statistical Analysis

One-way ANOVA test was performed to compare the computed parameters during NSR, AF and ARR. If the p value of the ANOVA test was significant, an unpaired t-test or Wilcoxon test with Holm's correction was applied. A  $p < 0.05$  was considered statistically significant. All analyses and statistical tests were performed using MATLAB R2012b (The MathWorks, USA).

### 2.6. Classification

To distinguish NSR from AF and ARR, feature selection was first performed, then followed by classification.

A two-step features selection procedure was used: first, features with a significant p-value of the ANOVA test were selected; second, a sequential forward floating search

(SFFS) algorithm was used to identify a small subset of optimal features. The SFFS algorithm [25] is briefly described in the following. Starting from the empty set of features, the feature  $x_i$  that maximizes the objective function  $J$  when combined with the  $k$  features that have already been selected ( $Y_k$ ) is added (forward step). The objective function values with different number of features  $J(k)$  is memorized, where  $k$  indicates the number of features. After the forward step, a backward step is performed. The backward step consists in removing from  $Y_k$  the feature that makes the objective function  $J^*(k)$  larger than  $J(k)$ , where  $J^*(k)$  is the objective function after removing one feature. The backward step is repeated as long as  $J^*(k)$  is larger than  $J(k)$ , with  $k$  decreasing, with the constrain that the last added feature cannot be removed. In this study, the objective function was the average of the accuracy for AF and for ARR, in order to decrease the false negative rate.

A dataset (like the one in our study) is imbalanced if the classes are not approximately equally represented. A way to overcome this problem is to re-sample the original dataset, by oversampling the minority class. The algorithm used in this study is Synthetic Minority Over-sampling Technique (SMOTE): the minority class is over-sampled by creating synthetic examples. Briefly, the minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining the  $q$  minority class nearest neighbors ( $q$  was chosen equal to 3) [5]. In this study, we over-sampled the class of ARR, so that all the three classes were equally represented.

After the feature selection step, the classification was performed using the k-nearest neighbor classifier. In the training phase, 2/3 of the data are used to build the model, whereas in the test phase the remaining 1/3 of the data are classified according to the model generated in the training phase [17]. The k-nearest neighbor classifier with k equal to 3 was used in this study [18]. Leave- $p$ -out cross-validation ( $p = 30$ ) was performed with 100 bootstrap repetitions, from which performance metrics were averaged [17].

### 3. Results

#### 3.1. Signal characterization

Tables 2 and 3 show results obtained analyzing the inter-diastolic intervals series in the three groups along with the p-values of the ANOVA test (results for the inter-systolic intervals series are not shown as they were very similar). In particular, Table 2 reports the spectral parameters and it can be observed that some of them are significantly different when comparing AF or ARR to NSR, but they are not able to distinguish AF from ARR. In particular, the power in the HF band was found to be significantly lower in NSR than in patients with arrhythmias.

Many of the variability and irregularity parameters are significantly different among all the three groups, as shown in Table 3. Among the variability parameters, all the  $pNNx$  are significantly different when comparing NSR to arrhythmias as well as when

**Table 2.** Frequency domain parameters in the three groups (mean  $\pm$  one standard deviation).

|                | NSR   | AF  | ARR   | p-value<br>K-W | Pairwise<br>significance  |
|----------------|---|---|---|----------------|---------------------------|
| LF (frequency) | 0.081 (0.061-0.110)   | 0.079 (0.055-0.105)                                 | 0.098 (0.053-0.104)   | ns             | -                         |
| HF (frequency) | 0.244 (0.192-0.320)   | 0.217 (0.194-0.256)                                 | 0.211 (0.0198-0.272)  | ns             | -                         |
| LF (power)     | $5 \cdot 10^{-7}$ ( $2 \cdot 10^{-7}$ - $1 \cdot 10^{-5}$ ) | $2 \cdot 10^{-5}$ ( $10^{-5}$ - $6 \cdot 10^{-5}$ ) | $9 \cdot 10^{-6}$ ( $2 \cdot 10^{-6}$ - $2 \cdot 10^{-4}$ ) | ns             | -                         |
| HF (power)     | $2 \cdot 10^{-7}$ ( $7 \cdot 10^{-8}$ - $4 \cdot 10^{-6}$ ) | $10^{-5}$ ( $9 \cdot 10^{-6}$ - $3 \cdot 10^{-5}$ ) | $6 \cdot 10^{-6}$ ( $3 \cdot 10^{-6}$ - $2 \cdot 10^{-5}$ ) | 0.008          | NSR vs. AF<br>NSR vs. ARR |

NSR: normal sinus rhythm, AF: atrial fibrillation, ARR: other arrhythmias.

**Table 3.** Variability and irregularity parameters in the three groups (mean  $\pm$  one standard deviation).

|                       | NSR                 | AF                  | ARR                 | p-value<br>ANOVA | Pairwise<br>significance  |
|-----------------------|---------------------|---------------------|---------------------|------------------|---------------------------|
| M                     | 834 (772-918)       | 877 (808-1035)      | 918 (763-997)       | ns               | -                         |
| SD                    | 89 (70-174)         | 330 (266-509)       | 223 (176-580)       | 0.001            | NSR vs. AF<br>NSR vs. ARR |
| pNN10                 | 84 (78-88)          | 97 (96-98)          | 92 (86-96)          | 0.0006           | All                       |
| pNN20                 | 51 (40-65)          | 93 (89-94)          | 82 (62-90)          | $10^{-11}$       | All                       |
| pNN30                 | 51 (40-65)          | 93 (89-94)          | 82 (62-90)          | $10^{-11}$       | All                       |
| pNN40                 | 29 (20-45)          | 87 (82-92)          | 69 (39-79)          | $10^{-19}$       | All                       |
| pNN50                 | 15 (10-28)          | 84 (76-88)          | 57 (32-74)          | $10^{-22}$       | All                       |
| pNN60                 | 15 (10-28)          | 84 (76-88)          | 57 (32-74)          | $10^{-22}$       | All                       |
| pNN70                 | 8 (5-20)            | 78 (66-83)          | 48 (25-69)          | $10^{-24}$       | All                       |
| pNN80                 | 7 (3-12)            | 75 (60-81)          | 41 (25-66)          | $10^{-24}$       | All                       |
| pNN90                 | 7 (3-12)            | 75 (60-81)          | 41 (25-66)          | $10^{-24}$       | All                       |
| pNN100                | 5 (1-11)            | 70 (54-77)          | 35 (23-65)          | $10^{-22}$       | All                       |
| rMSSD                 | 108 (66-273)        | 470 (390-726)       | 356 (267-785)       | 0.0006           | NSR vs. AF<br>NSR vs. ARR |
| SampEn <sub>m=1</sub> | 10.31 (0.395-1.396) | 1.349 (1.156-1-458) | 0.365 (0.190-0.821) | 0.0004           | All                       |
| SampEn <sub>m=2</sub> | 0.924 (0.391-1.370) | 1.297 (1.089-1.496) | 0.340 (0.147-0.758) | 0.0002           | All                       |

NSR: normal sinus rhythm, AF: atrial fibrillation, ARR: other arrhythmias.

comparing AF to ARR. In particular, pNN $x$  tend to have high values for patients with AF, low values for subjects in NSR and values in-between in presence of ARR. The irregularity parameter, SampEn, is significantly different across the groups: SampEn is higher during AF than during NSR or in patients with ARR. Moreover, SampEn is higher in subject in NSR than during ARR.

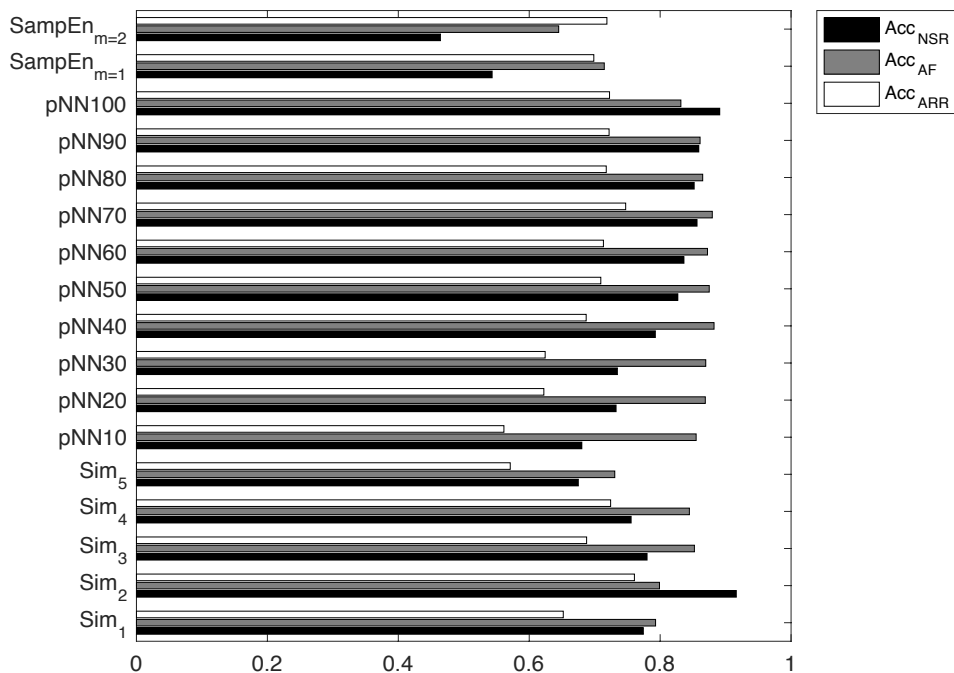
Tables 4 shows the results on the similarity of the waves morphology for different thresholds (different rows). It can be observed that in each rhythm the smaller the threshold, the lower the value of the similarity index. When comparing the different rhythms, for all the tested thresholds, NSR always has the highest similarity, AF has the lowest values and the ARR group has intermediate values.



**Table 4.** Similarity indexes in the three groups (mean  $\pm$  one standard deviation).

|                  | Threshold | NSR              | AF                        | ARR              | p-value<br>K-W | Pairwise<br>significance |
|------------------|-----------|------------------|---------------------------|------------------|----------------|--------------------------|
| Sim <sub>1</sub> | $\pi/2$   | 0.98 (0.98-0.99) | 0.92 (0.82-0.94)          | 0.96 (0.92-0.98) | $10^{-6}$      | All                      |
| Sim <sub>2</sub> | $\pi/4$   | 0.92 (0.85-0.98) | 0.44 (0.20-0.52)          | 0.66(0.41-0.78)  | $10^{-16}$     | All                      |
| Sim <sub>3</sub> | $\pi/8$   | 0.64 (0.44-0.87) | 0.10 (0.02-0.13)          | 0.36 (0.08-0.56) | $10^{-15}$     | All                      |
| Sim <sub>4</sub> | $\pi/16$  | 0.19 (0.09-0.51) | 0.01 ( $10^{-4}$ -0.02)   | 0.08 (0.01-0.28) | $10^{-7}$      | All                      |
| Sim <sub>5</sub> | $\pi/32$  | 0.02 (0.01-0.12) | $10^{-5}$ (0- $10^{-4}$ ) | 0.002 (0-0.04)   | $10^{-4}$      | All                      |

NSR: normal sinus rhythm, AF: atrial fibrillation, ARR: other arrhythmias.



**Figure 3.** Mean accuracy for NSR, AF and ARR, obtained using only the parameter as indicated on the y-axis.

### 3.2. Rhythm classification

Globally, seventeen parameters were found significantly different in the three rhythms (as shown in the previous section): Figure 3 shows the mean accuracy for the three rhythm using only one of these 17 parameters at a time. It can be noted that the classification with one variability (pNN $x$ ) parameter makes the accuracy for AF and NSR the highest. Good accuracy for AF and NSR is obtained also with the similarity parameters, whereas irregularity parameters are worse in classifying the three rhythms, being the worst for NSR. The accuracy for ARR is about 0.7 with almost all the parameters.

Figure 4 shows the accuracy for NSR, AF and ARR, obtained using an increasing number of parameters, as selected by the SFFS algorithm. In Figure 4, the  $n$ -th dot represents the mean accuracy (over the 100 repetitions) obtained using  $n$  parameters. It can be observed that passing from one to two parameters, the mean accuracy increases,

**Table 5.** Selected parameters (using at maximum six) and the corresponding accuracy (mean $\pm$  one standard deviation) for the three rhythms.

| Selected Features  | Accuracy NSR      | Accuracy AF       | Accuracy ARR       |
|--|-------------------|-------------------|--------------------|
| pNN70  | 0.86 $\pm$ 0.06   | 0.88 $\pm$ 0.05   | 0.75 $\pm$ 0.07    |
| pNN70, pNN40   | 0.88 $\pm$ 0.05 * | 0.91 $\pm$ 0.05 * | 0.80 $\pm$ 0.06 ** |
| pNN70, pNN40, Sim <sub>5</sub>   | 0.87 $\pm$ 0.06   | 0.92 $\pm$ 0.05   | 0.80 $\pm$ 0.07    |
| pNN70, pNN40, Sim <sub>5</sub> , Sim <sub>4</sub>  | 0.88 $\pm$ 0.05   | 0.91 $\pm$ 0.05   | 0.80 $\pm$ 0.06    |
| pNN70, pNN40, Sim <sub>5</sub> , Sim <sub>4</sub> , SampEn <sub>m=1</sub>                    | 0.88 $\pm$ 0.06   | 0.92 $\pm$ 0.04   | 0.80 $\pm$ 0.06    |
| pNN70, pNN40, Sim <sub>5</sub> , Sim <sub>4</sub> , SampEn <sub>m=1</sub> , Sim <sub>3</sub> | 0.87 $\pm$ 0.06   | 0.92 $\pm$ 0.05   | 0.81 $\pm$ 0.06    |

\* p < 0.01, \*\* p < 0.0001 Accuracy using two features vs. accuracy using one feature.

**Table 6.** Sensitivity and specificity using tow parameters (pNN70 and pNN40) for the three rhythms.

| Rhythm | Specificity | Sensitivity |
|--------|-------------|-------------|
| NSR    | 0.928       | 0.773       |
| AF     | 0.963       | 0.754       |
| ARR    | 0.768       | 0.758       |

NSR: normal sinus rhythm, AF: atrial fibrillation, ARR: other arrhythmias.

**Table 7.** Sensitivity and specificity using two parameters (pNN70 and pNN40) for the three rhythms.

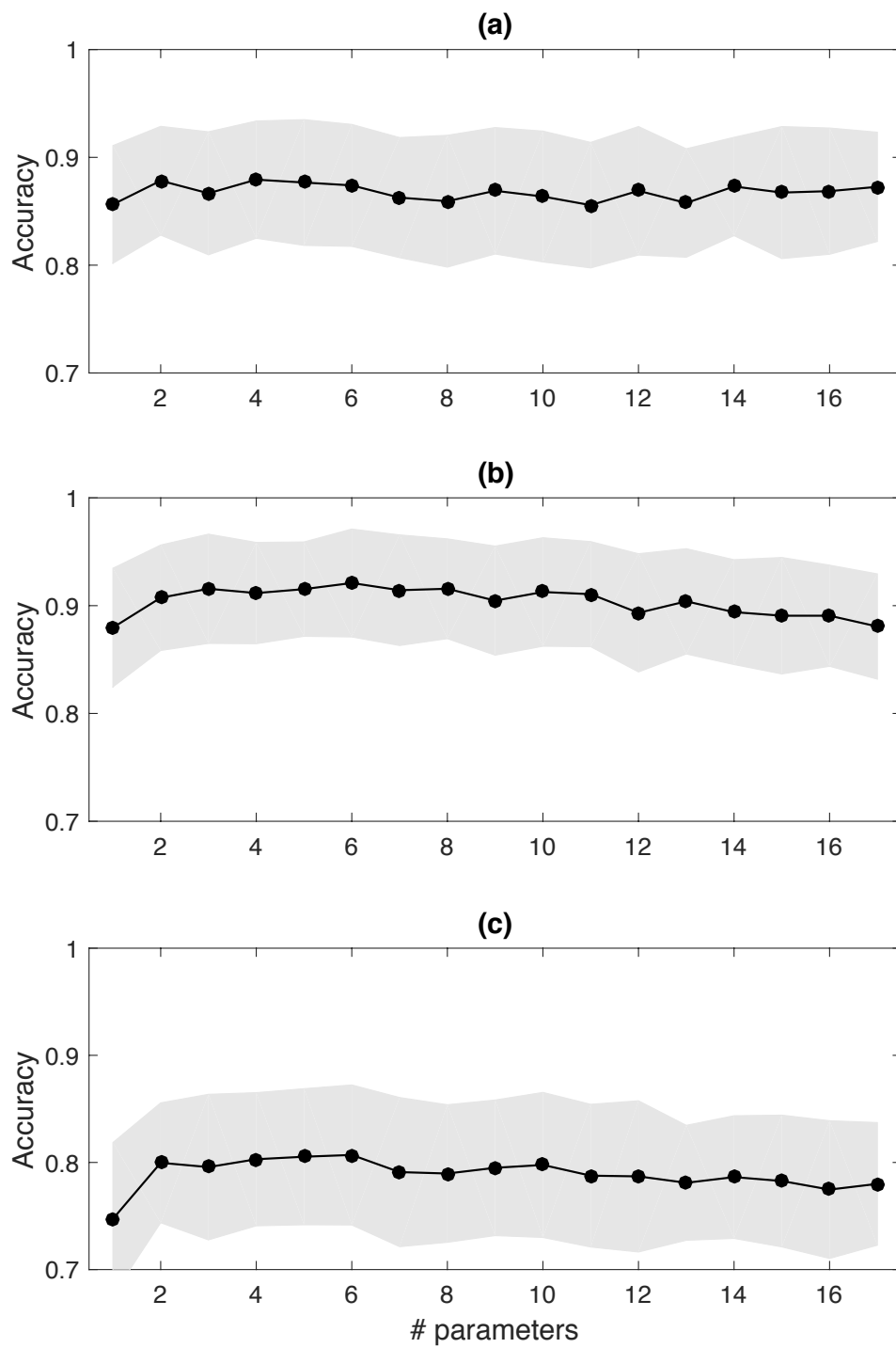
| Rhythm | Specificity | Sensitivity |
|--------|-------------|-------------|
| NSR    | 0.990       | 0.874       |
| AF     | 0.960       | 0.880       |
| ARR    | 0.885       | 0.083       |

NSR: normal sinus rhythm, AF: atrial fibrillation, ARR: other arrhythmias.

then it is almost constant, and then when new parameters are added, slightly decreases. Table 5 shows the best selected parameters using up to six parameters. It can be noted that the mean accuracy significantly increased using two parameters compared to only one parameter, for all the three rhythms. Using two parameters, namely pNN70 and pNN40, seems to be the best compromise which balances the number of features and the accuracy values. In particular, using these two parameters, the specificity and sensitivity are shown in Table 7.

#### 4. Discussion



In this study, we assessed for the first time, the possibility to discriminate AF from NSR and ARR by using biosignals recorded by a wristband device. The main finding is that AF can be well detected, with high sensitivity and high specificity, using only few parameters computed on the BVP signal. The differences between AF and NSR signals



**Figure 4.** Mean  $\pm$  the standard deviation (black line and the grey area, respectively) of the accuracy for (a) NSR, (b) AF (c) and ARR.

are usually very pronounced, as the heart rate during AF, without atrioventricular node block, is much more irregular [7]. This causes the RR series and, similarly, the inter-diastolic intervals series to be more variable and more irregular during AF [6]. In addition, during AF as R waves may not be coupled with an adequate left ventricular output to generate discrete pulses, thus the arterial blood pressure and consequently the BVP signal may look different from those in NSR [23]. However, AF irregularity may cause problems in discriminating AF from ARR, that can include atrial tachycardia, atrial flutter, and premature ventricular contractions. In respect to this, it is worth noting that the classifier was able to correctly classify AF from ARR with an average false negative rate of  $\sim 25\%$ .

Previous studies have shown that recording the BVP signal by a pulse oximeter, similar information to that from HRV analysis can be obtained [20]. Similar-BVP signals have been recently extracted from everyday sensors as smartphone [19] or video cameras [8], with the aim of detecting AF episodes. In [19], AF was detected from pulsatile signals in the human fingertip using the camera of an iPhone 4s. They computed parameters similar to those in our study assessing variability and irregularity (rMSSD, SampEn and Shannon entropy). The accuracy using rMSSD was 0.98 (without cross-validation), but the protocol included 25 patients with AF before and after cardioversion, thus in NSR, without the presence of the ARR group. In [8], they used a video camera to record an individual's face and extract the subtle beat-to-beat variations of skin color reflecting the cardiac pulsatile signal. They analyzed recordings from 11 patients with AF undergoing electrical cardioversion, before and after the procedure, thus comparing AF and NSR, introducing a novel quantifier of pulse variability called the pulse harmonic strength. Among the 407 epochs of 15 seconds of synchronized ECG and videoplethysmographic signals, the pulse harmonic strength was associated with a 20% detection error rate, while the error rate of the automatic ECG-based measurements ranged between 17% and 29%. It is worth noting that some differences exist between these studies and ours. Our protocol included patients with ARR, such as atrial tachycardia, that makes the correct classification of AF more difficult: in our classification, the AF group was very rarely misclassified as NSR, being the false negative rate (AF classified as NSR) only 0.004. Moreover, the wrist-band device allows the recording without patient interaction, i.e., once the device is on, the patient does not need to be compliant.

With the aim of detecting paroxysmal AF, there is the need of monitoring the patients for longer periods. A recent study assessed the possibility of using a handheld device (thumb ECG) to record short ECG [13]. The main finding was that intermittent ECG recording was superior to routine 24-hour Holter ECG in manually detecting relevant paroxysmal arrhythmias in a patient population reporting symptoms of palpitations, dizziness/presyncope. The intermittent recordings were repeated over four weeks, both regularly twice daily and when having symptoms. During this period the compliance was high, as the 95 included patients had a median of 59 registrations. Another possibility for long monitoring is using an  insertable cardiac monitor [27]. In this study [27] the  insertable cardiac monitor correctly identified 37 of the 38 patients

with Holter-detected AF (diagnostic sensitivity of 97.4%) and 97 of the 100 patients without AF according to Holter analysis (diagnostic specificity of 97.0%). The main disadvantage of these two devices are, respectively, the need of high compliance of the patients and the invasive nature of the insertable cardiac monitor. On the contrary, the wrist-band device can be used for longer periods, without causing discomfort to the patients and without needing their compliance.

Given the increasing number of people at risk for AF, and the high prevalence of paroxysmal and asymptomatic AF, a wider screening of people at risk, using non-invasive comfortable devices may be helpful. Empatica E4 wristband device, used in this study, is a wearable wireless device designed for continuous, real-time data acquisition in daily life, being non-invasive and worn like a regular wrist watch, without causing discomfort to the patient, not even after prolonged acquisitions. The results highlighted the possibility to discriminate AF from NSR and more interestingly from ARR based on two-minute recording, thus laying the groundwork for longer recordings for patients at risk of AF. The results were obtained with the patients laying still, in controlled condition. Therefore additional studies will be needed to assess the feasibility of AF detection by using wristband device during daily activities, where artifacts caused by daily activity may prevent the detection of events. The first step toward this final goal was to analyze the BVP wristband signals and develop an algorithm for discriminating AF from sinus rhythm and other arrhythmias. To test this possibility, we analyzed the most stable situation to avoid external disturbance. Finally, it is worth noting that the most relevant parameters were those assessing waveform similarity as well as variability of the inter-diastolic intervals series. These variability parameters, despite their simplicity, have been previously found to be predictive of long-term clinical outcome in a population of patients with mild-to-moderate heart failure and AF [9]. Thanks to their simplicity, they might be even implemented on the device itself.

## 5. Acknowledgments

The Authors want to thank Empatica for providing the Empatica E4 device for recording the signals.

- [1] Akaike, H.: Statistical predictor identification. *Ann Inst Statist Math* **22**, 203–217 (1970)
- [2] Anter, E., Jessup, M., Callans, D.: Atrial fibrillation and heart failure: Treatment considerations for a dual epidemic. *Circulation* **119**(18), 2516–2525 (2009)
- [3] Baselli, G., Cerutti, S., Civardi, S., Lombardi, F., Malliani, A., Merri, M., Pagani, M., Rizzo, G.: Heart rate variability signal processing: A quantitative approach as an aid to diagnosis in cardiovascular pathologies. *Int. J. Biomed. Comput.* **20**, 51–70 (1987)
- [4] Benjamin, E., et al.: Prevention of atrial fibrillation: Report from a national heart, lung, and blood institute workshop. *Circulation* **119**, 606–618 (2009)
- [5] Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W.: Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* pp. 321–357 (2002)
- [6] Corino, V., Holmqvist, F., Mainardi, L., Platonov, P.: Beta-blockade and  $\alpha$ 1-adenosine receptor agonist effects on atrial fibrillatory rate and atrioventricular conduction in patients with atrial fibrillation. *Europace* **16**, 587–594 (2014)
- [7] Corino, V., Sassi, R., Mainardi, L., Cerutti, S.: Signal processing methods for information

- enhancement in atrial fibrillation: Spectral analysis and non-linear parameters. *Biomedical Signal Processing and Control* **1**, 271–281 (2002)
- [8] Couderc, J., Kyal, S., Mestha, L., Xu, B., Peterson, D., Xia, X., Hall, B.: Detection of atrial fibrillation using contactless facial video monitoring. *Heart Rhythm* **12**, 195–201 (2015)
- [9] Cygankiewicz, I., Corino, V., Vazquez, R., Bayes-Genis, A., Mainardi, L., Zareba, W., A. Bayes de Luna, P.P.f.: Reduced irregularity of ventricular response during atrial fibrillation is associated with poor outcome in patients with heart failure. *American Journal of Cardiology* **116**, 1071–1075 (2015)
- [10] Faes, L., Nollo, G., Antolini, R., Gaita, F., Ravelli, F.: A method for quantifying atrial fibrillation organization based on wave-morphology similarity. *IEEE Transactions on Biomedical Engineering* **49**(12), 1504–1513 (2002)
- [11] Hart, R., Pearce, L., Aguilar, M.: Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med* **146**, 857–867 (2007)
- [12] Healey, J., et al.: Subclinical atrial fibrillation and the risk of stroke. *N Engl J Med* **366**, 120–129 (2012)
- [13] Hendriks, T., Rosenqvist, M., Wester, P., Sandström, H., Hörnsten, R.: Intermittent short ecg recording is more effective than 24-hour holter ecg in detection of arrhythmias. *BMC Cardiovascular Disorders* **14**, 41 (2014)
- [14] Iozzia, L., Cerina, L., Mainardi, L.: Relationships between heart-rate variability and pulse-rate variability obtained from video-PPG signal using ZCA. *Physiological Measurement* **37**, 1934–1944 (2016)
- [15] J.C. Hsu, a.T.M., Kennedy, K., Katz, D., Marzec, L., Lubitz, S., Gehi, A., Turakhia, M., Marcus, G.: Oral anticoagulant therapy prescription in patients with atrial fibrillation across the spectrum of stroke risk insights from the NCDR PINNACLE registry. *JAMA Cardiol* **1**, 55–62 (2016)
- [16] Kay, S.M., Marple, S.L.: Spectrum analysis: a modern perspective. *IEEE Trans. Biomed. Eng.* **69**, 1380–1419 (1981)
- [17] Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *IJCAI’95 Proceedings of the 14th international joint conference on Artificial intelligence*, vol. 2, pp. 1137–1143 (1995)
- [18] Kononenko, I.: Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med* **23**, 89–109 (2001)
- [19] Lee, J., Reyes, B.A., McManus, D.D., Mathias, O., Chon, K.H.: Atrial fibrillation detection using an iPhone 4S. *IEEE Transactions on Biomedical Engineering* **60**(1) (2013)
- [20] Lu, S., Zhao, H., Ju, K., Shin, K., Lee, M., Shelley, K., K.: Can photoplethysmography variability serve as an alternative approach to obtain heart rate variability information? *Journal of Clinical Monitoring and Computing* **22**, 23–29 (2008)
- [21] Mainardi, L.: On the quantification of heart rate variability spectral parameters using time-frequency and time-varying methods. *Phil Trans Royal Soc A* **367**, 255–275 (2009)
- [22] Mainardi, L., Corino, V., Belletti, S., Terranova, P., Lombardi, F.: Low frequency component in systolic arterial pressure variability in patients with persistent atrial fibrillation. *Autonomic Neuroscience* **151**(2), 147–153 (2009)
- [23] Mainardi, L., Corino, V., Belletti, S., Terranova, P., Lombardi, F.: Low frequency component in systolic arterial pressure variability in patients with persistent atrial fibrillation. *Auton Neurosci* **151**, 147–53 (2009)
- [24] Page, R., Wilkinson, W., Clair, W., McCarthy, E., Pritchett, E.: Asymptomatic arrhythmias in patients with symptomatic paroxysmal atrial fibrillation and paroxysmal supraventricular tachycardia. *Circulation* **89**, 224–227 (1994)
- [25] Pudil, P., Novovicova, J., Kittler, J.: Floating search methods in feature selection. *Pattern Recognition Letters* **15**(1119–1125) (1994)
- [26] Richman, J.S., Moorman, J.R.: Physiological time-series analysis using approximate entropy and

- sample entropy. *American Journal of Physiology - Heart and Circulatory Physiology* **278**(6), H2039–H2049 (2000)
- [27] Sanders, P., H. Pürerfellner, a.E.P., Sarkar, S., Bacco, M.D., B. Maus, a.L.D.: Performance of a new atrial fibrillation detection algorithm in a miniaturized insertable cardiacmonitor: Results from the revealLINQ usability study. *Heart Rhythm* **13**, 1425–1430 (2016)
- [28] Soliman, E.Z., et al.: Atrial fibrillation and the risk of myocardial infarction. *JAMA Internal Medicine* **174**(1), 107–114 (2014)
- [29] Sposato, L., Cipriano, L., Saposnik, G., Vargas, E., Riccio, P., Hachinski, V.: Diagnosis of atrial fibrillation after stroke and transient ischaemic attack: asystematic review and meta-analysis. *Lancet Neurol* **14**, 377–387 (2015)
- [30] Zetterberg, L.H.: Estimation of parameters for a linear difference equation with application to EEG analysis. *ab-MB* **5**, 227–275 (1969)