Detection of Atrial Fibrillation using 12-lead ECG for Mobile Applications

R. Couceiro, P. Carvalho, J. Henriques, R. P. Paiva, M. Antunes†
Centre for Informatics and Systems, University of Coimbra, Coimbra, Portugal
†Centre of Cardio-thoracic Surgery of the University Hospital of Coimbra, Coimbra, Portugal

Abstract— **Atrial Fibrillation (AF) is the most common arrhythmia and is associated with an increased risk of heart-related deaths and the development of conditions such as heart failure, dementia, and stroke. Affecting mostly elderly people, AF is associated with high comorbidity, increased mortality and is a major socio-economic impact in our society. Therefore, the detection of AF episodes in personalized health (p-Health) environments can be decisive in the prevention of major cardiac threats and in the reduction of health care costs.**

In this paper we present a new algorithm for detection of AF based on the assessment of the three main physiological characteristics of AF: 1) the irregularity of the heart rate; 2) the absence of the P-wave and 3) the presence of fibrillatory waves. Several features were extracted from the analysis of 12-lead electrocardiogram (ECG) signals, the best features were selected and a support vector machine classification model was adopted to discriminate AF and non-AF episodes. Our results show that the inclusion of features from the analysis of the recovered atrial activity was able to increase the performance of the algorithm: sensitivity of 88.5 % and specificity of 92.9%.

In the WELCOME project it is being designed a novel light vest with an integrated sensor system that collects several signals, including 12-lead ECG signals. The proposed algorithm is currently integrated in the WELCOME feature extraction module, which is responsible for receiving raw signals, extraction higher level features (e.g. occurrence of AF episodes) and provide them to the clinical decision process.

I. Introduction

Atrial Fibrillation (AF) is the most common sustained cardiac arrhythmia and is associated with significant morbidity, mortality and decreased life quality, specially in elderly, where its prevalence increases to 8% [1]. It is predicted that by 2060 AF will affect 17.9 million people in Europe and 6-12 million in United States by 2050 [2]. Despite not being lethal, AF is associated with an increased risk of heart failure, dementia, and stroke.

AF results from multiple re-entrant wavelets in the atria, which leads to its partial disorganization. In the electrocardiogram, AF can be recognized by the absence of the P-wave before the QRS-complex, which is replaced by a "sawtooth" shaped wave, and by the appearance of an irregular cardiac frequency, or both.

Based on these characteristics, several methods have been proposed in literature to detect AF using single-lead or multi-lead ECG signals. To assess the irregularity of the heart rhythm, methods based on the analysis of a Hidden Markov Model transition probabilities [3], linear and non-linear analysis of auto-regressive (AR) models [4] and histogram-based statistical analysis [5] have been proposed. To extract the atrial activity (AA), two main directions have been followed: i) single-lead ECG analysis and multi-lead ECG analysis. Techniques such as blind source separation, spatio-temporal cancellation and artificial neural networks are the most promising in these two research fields. In single-lead ECG analysis, the main approaches for QRS-T cancelation are based on wavelet transforms [6, 7] and template-based approaches (e.g. [8]), while in the multi-lead ECG analysis, the main methods are based on the blind source separation techniques, such as independent component analysis (ICA) [9]. However, the majority of the studies proposed lack of proper analysis regarding the extraction of relevant features from the extracted AA, capable of being used in the discrimination between AF and non-AF episodes.

In our previous work [10] we proposed an algorithm for detection of AF based on the single-lead ECG analysis and combining features assessed from heart rate (HR) and atrial activity.

In this paper, we propose a novel algorithm for the detection of AF, which is based on the assessment of the HR irregularity and on the atrial activity, using a 12-lead ECG approach. In this algorithm, the atrial activity is retrieved using ICA and several frequency domain features were extracted. The relevance of the extracted features was evaluated using the F-score metric and the best features (three from HR analysis, one from P-wave detection and four from AA analysis) were selected for classification purposes.

The remainder of the paper is organized as follows. In section II the algorithm structure and feature extraction are presented. The data collection and main results are presented and discussed in section III. In section IV the main conclusions are outlined.

Manuscript received September 10, 2015. This work was supported by CISUC (Center for Informatics and Systems of University of Coimbra) and by EU projects Welcome (PTDC-EEI-PRO-2857-2012) and iCIS (CENTRO-07-ST24-FEDER-002003).

R. Couceiro, P. Carvalho, J. Henriques and R. P. Paiva are with the University of Coimbra, Department of Informatics Engineering, Science and Technology Faculty of the University of Coimbra, Pólo II, Coimbra, Portugal (e-mail: {rcouceir, carvalho, jh, ruipedro}@dei.uc.pt).

M. Antunes is with the Centro de Cirurgia Cardio-Torácica of the Hospital da Universidade de Coimbra, Coimbra, Portugal (e-mail: antunes.cct.huc@sapo.pt).

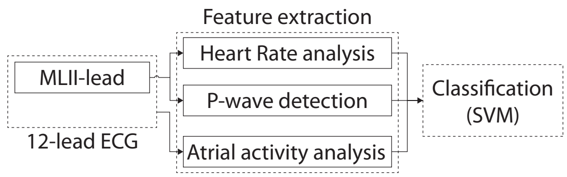

Figure 2. Structure of the proposed algorithm.

II. METHODS

The proposed method consists of two main phases, which are: 1) the feature extraction and; 2) the classification. In the feature extraction phase, the ECG signals are processed and analyzed in order to extract relevant features for the discrimination between AF and non-AF episodes, which are used downstream during the classification phase.

A. Feature extraction

The first step of the proposed algorithm is the segmentation of the MLII-lead ECG signal, i.e. the detection of its characteristics waves (P-wave, QRS-complex and T-wave). Here, an algorithm similar to the one proposed by Sun *et al.* [11] has been adopted.

1) Heart rate analysis

In heart rate (HR) analysis the main objective is the extraction of features that are able to quantify the regularity of the RR intervals in the ECG. To this matter, the RR sequence was modelled using a Markov process (see Figure 1) with three states [3]: small (S_1), regular (S_2) and long (S_3) RR intervals.

From the transition probabilities between each state, one constructed a transition probability (TrP) matrix, which characterizes the regularity (or irregularity) of the heart cycles. The probability of the state S_2 and the probability of transition from S_2 to S_2 state quantifies the regularity of the heart rate, i.e. a high S_2-to-S_2 probability shows that is very likely to find two consecutive RR intervals with the same (regular) length. In fact, these are the first features (F_1 and F_2) that characterize the RR regularity:

$$F_1 = P(S_2) \quad (1)$$

$$F_2 = P(S_i, S_j) = P(S_i|S_j) \times P(S_i) \quad (2)$$

where *i=2* and *j=2* are the labels corresponding to the second state (regular RR interval).

From the analysis of the TrP matrix we found that AF and non-AF episodes present very characteristic distributions.

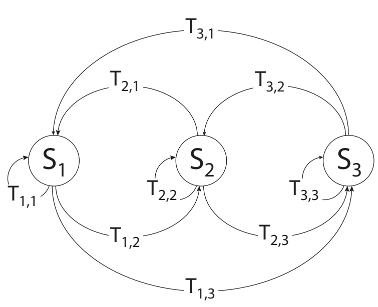

Fig 1. Structure of the Hidden Markov model used to analyze the HR.

While the TrP matrices corresponding to non-AF episodes present a dirak-impulse-like distribution concentrated around the S_2-to-S_2 transition, the TrP matrices from AF episodes present a much flatter probabilistic distribution, i.e. it is more likely to find transitions between RR intervals with different lengths during AF episodes. Based on this finding we proposed the assessment TrP matrix dispersion by measuring its entropy (*H*), as defined in (3).

$$F_3 = \sum_{i=1}^{3} P(S_i) \times \sum_{j=1}^{3} P(S_j|S_i) \times log_2 P(S_j|S_i) \quad (3)$$

Additionally, the similarity between a probabilistic distribution under analysis and a model representative of AF episodes (collected from MIT-BIH Atrial Fibrillation database) was also assessed using the Kullback–Leibler divergence, as defined in (4).

$$F_4 = \sum_{x=1}^{3} \sum_{y=1}^{3} P(x,y) log\left(\frac{P(x,y)}{\overline{P_{AF}(x,y)}}\right) \quad (4)$$

where $\overline{P_{AF}(x,y)}$ is the defined AF model and $P(x,y)$ is the distribution under analysis. More details about these features can be found in [10].

2) P-wave detection

The first step in the analysis of the atrial activity is to search for the presence of P-waves before the QRS complex. While during non-AF episodes the P-waves are commonly distinguishable, during AF episodes the P-waves are replaced by a "sawtooth" like waveform resultant from the fibrillatory process. To correctly evaluate the presence or absence of P-waves, the Pearson correlation (ρ) coefficient is calculated between the segmented P-waves and a P-wave model and the rate of P-waves per window (F_6) is assessed by:

$$F_5 = R_{Pwaves} = \frac{N_{SP}}{N_{RR}} \quad (5)$$

where N_{SP} is the number of selected P waves (with ρ greater than 0.2) and N_{RR} is the number of cardiac cycles detected in the analysed window.

3) Atrial activity analysis

The third main characteristic of AF is the discoordination of the atrial activation, which is a result of the disorganization in the path of the electrical impulses in the atria. In the ECG, the result is the replacement of the commonly seen P-waves by fibrillatory waves, with typical frequencies ranging from 5 to 8 Hz (herein defined as AF_{int}). Moreover, the spectrum of AF episodes presents no harmonics and the amplitudes above 15 Hz are minimal [9].

In order to analyse this process, it is essential to retrieve the signal components related with the AA, i.e. to cancel or extract the QRS complex and the T wave (QRST) from the analysed signals. To recover the atrial components of the ECGs, we applied independent component analysis (ICA) as proposed in [9].

First, all the ECG signals were upsampled to 1kHz using a shape-preserving piecewise cubic interpolator, aiming the improvement of the frequency resolution in the subsequent analysis. Next, the signals were normalized regarding their amplitude and preprocessed. In the preprocessing, the power line interference is canceled using a 50Hz notch filter, while the baseline wandering and thermal noise are reduced using a 0.5-60 Hz band-pass filter. The separation process was performed using the FastICA algorithm in consecutive 10s windows, shifted by 10s increments. The identification of the components related with the AA was performed using a kurtosis-based source reordering, where the components with sub-Gaussian statistical properties ($k<0$) were assigned to AA, and the components with Gaussian ($k=0$) and super-Gaussian properties ($k>0$) were assigned to noise (and/or artifacts) and VA, respectively. After the separation process is concluded, the components corresponding to the AA are summed into a single AA source and the power spectral density (PSD) was estimated using the Welch's (WOSA) method. In the PSD estimation one used a Hamming window with 4096 samples, a section overlap of 2048 samples and a discrete Fourier transform (DFT) with 8192 points.

From the analysis of the estimated PSDs, five features were extracted. Although AF episodes are characterized by a spectrum peak within the AF frequency region, occasionally, due to difficulties in the separation process or in the peak detection, no peak is found within this region. Therefore, the first AA feature (F_6), was defined as the distance from the spectrum maximum peak to the frequency interval characteristic of AF episodes, i.e. 5 to 8 Hz.

In contrast with AF spectrums, which present a very characteristic frequency spectrum with a major peak in the AF_{int}, non-AF episodes present a spectrum dispersed along a wider frequency range. This observation leads to the definition of more two AA features, which are the entropy of the spectrum (F_7) and the Kullback–Leibler divergence between the spectrum and a generalized bell-shaped membership function (F_8):

$$f(x,a,b,c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \quad (6)$$

where the parameters $a=2$, $b=6$ and $c=6$ control the shape and position of the function in the AF_{int}. Let $P(w)$ be the spectrum under analysis and $Q(w)$ be the aforementioned bell-shaped function, the features F_7 and F_8 are defined as follows:

$$F_7 = -\sum_{w\in W} P(w)\times log_2 P(w) \quad (7)$$

$$F_8 = -\sum_{w\in W} P(w) log \frac{P(w)}{Q(w)} \quad (8)$$

where w is the frequency bin and W is the range of spectrum frequencies.

Additionally, the dispersion of the spectrum was also assessed by the number of spectrum peaks above half height the maximum peak (F_9) and by the weight of the main peak spectrum frequencies (F_{10}), as defined in (9) and (10)

$$F_{10} = \frac{\sum_{w\in W} P(w)\times Q(w)}{\sum_{w\in W} P(w)} \quad (9)$$

where W_P is the range of frequencies corresponding to the main spectrum peak.

To assess the weight of the spectrum frequencies above 15 Hz the last feature was defined as:

$$F_{11} = \frac{\sum_{w>15} P(w)}{\sum_{w\in W} P(w)} \quad (10)$$

In Figure 3 we illustrate the main characteristics of the AF and non-AF spectra and the rationale behind the extracted features.

B. Classification

The classification between AF and non-AF episodes was performed on a 10 second window basis using a support vector machine classification model (C-SVC algorithm) with a radial basis function.

III. RESULTS AND DISCUSSION

A. Study protocol

In this study AF and non-AF episodes from 12 patients were considered. From those, 1 episode (2 records of 30 mins.) was selected from the "St.-Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database" and 11 episodes (11 records of 60 mins.) were selected from the 12-lead ECG database collected under the project

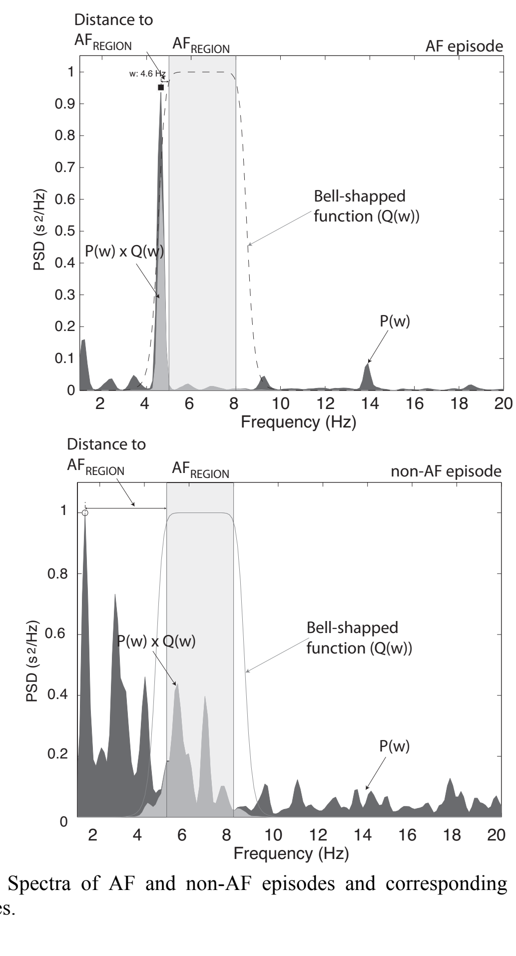

Fig 3. Spectra of AF and non-AF episodes and corresponding extracted features.

TABLE II
RESULTS ACHIEVED BY THE PROPOSED MULTI-LEAD AND SINGLE-LEAD AF DETECTION ALGORITHMS.

Algorithm	SE(%) avg ± std	SP(%) avg ± std	PPV(%) avg ± std
Single-lead algorithm [10]	79.0 ± 3.0	91.4 ± 0.5	86.6 ± 2.2
Multi-lead algorithm	88.5 ± 1.4	92.9 ± 0.3	90.6 ± 1.4

"Cardiorisk - Personalized Cardiovascular Risk Assessment through Fusion of Current Risk Tools".

The selected records were partitioned into records of 5 mins leading to the construction of a dataset consisting of 144 records of 5 mins length, in which 72 records present AF and 72 records present other rhythms other than AF.

B. Feature selection

The selection of the features most suitable for detection of AF episodes was performed based on the F-score metric. A ROC analysis was performed for each feature using a 6-fold cross validation approach, leading to the selection of eight features. The best features were extracted from the HR analysis (F_4 and F_3), followed by three features from the AA analysis (F_6, F_8 and F_{10}). Three features from both HR and AA analysis (F_2, F_9, F_{11}) presented a F-score below the 50% and therefore were not selected.

C. Validation

The validation of our algorithm was performed using a 6-fold cross validation approach, where the dataset was randomly partitioned into 6 equal size subsets. From the 6 subsets, 5 subsets were used for training (with episodes from 10 patients) and 1 subset (with episodes from the remaining 2 patients) was used for testing. The cross-validation process was repeated 6 times for each of the 6 subsets. This process was repeated 20 times and the average and standard deviation (avg ± std) of the sensitivity (SE), specificity (SP) and positive predictive value (PPV) was evaluated.

In Table II we present the results achieved by the single-lead [10] and multi-lead algorithms in the testing subsets. It is possible to observe that the multi-lead algorithm performed better than the single-lead algorithm. The analysis of AA recovered from 12-lead source separation provided relevant features that enabled the increase of approximately 9% the algorithms SE, 1% in the algorithms SP and 4% in the algorithms PPV. These results show that source separation techniques such as ICA can provide a valuable insight about AA and enable the extraction of reliable features for AF detection.

IV. CONCLUSIONS

In this paper we presented a novel algorithm for detection of AF episodes based on the analysis of 12-lead ECG signals. The proposed algorithm is based on the analysis of the three main characteristics of AF: the irregularity of the RR interval, the absence of the P-wave and the presence of the fibrillatory wave. The extraction of features from the separated atrial activity is the main innovative aspect of the proposed algorithm. Experimental results showed that the extracted features are relevant to this topic and the algorithm was able to achieve better discrimination performance when compared to the previously proposed single-lead solution. Based on these evaluations, it is possible to conclude that the extraction and analysis of atrial activity from multi-lead ECG signals is an important contribution to the enhancement of AF detection problem.

The proposed algorithm is currently integrated in the WELCOME feature extraction module, which is responsible for the off-line extraction of higher level features in a cloud server and for providing them to the clinical decision process.

ACKNOWLEDGMENT

This work was supported by CISUC (Center for Informatics and Systems of University of Coimbra) and by EU projects Welcome (PTDC-EEI-PRO-2857-2012) and iCIS (CENTRO-07-ST24-FEDER-002003).

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] R. C. Davis, F. D. R. Hobbs, J. E. Kenkre, A. K. Roalfe, R. Iles, G. Y. H. Lip, *et al.*, "Prevalence of atrial fibrillation in the general population and in high-risk groups: the ECHOES study," *EP Europace,* vol. 14, pp. 1553-1559, 2012-11-01 00:00:00 2012.

[2] R. B. Schnabel, X. Yin, P. Gona, M. G. Larson, A. S. Beiser, D. D. McManus, *et al.*, "50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study," *The Lancet,* vol. 386, pp. 154-162.

[3] G. B. Moody and R. G. Mark, "A new method for detecting atrial fibrillation using R-R intervals.," in *Computers in Cardiology*, 1983, pp. 227-230.

[4] S. Cerutti, L. T. Mainardi, A. Porta, and A. M. Bianchi, "Analysis of the dynamics of RR interval series for the detection of atrial fibrillation episodes," in *Computers in Cardiology 1997*, 1997, pp. 77-80.

[5] K. Tateno and L. Glass, "A method for detection of atrial fibrillation using RR intervals," in *Computers in Cardiology 2000*, 2000, pp. 391-394.

[6] L. Senhadji, F. Wang, A. Hernandez, and G. Carrault, "Wavelets extrema representation for QRS-T cancellation and P wave detection," in *Computers in Cardiology, 2002*, 2002, pp. 37-40.

[7] C. Sanchez, J. Millet, J. J. Rieta, F. Castells, J. Rodenas, R. Ruiz-Granell, *et al.*, "Packet wavelet decomposition: An approach for atrial activity extraction," in *Computers in Cardiology, 2002*, 2002, pp. 33-36.

[8] S. Shkurovich, A. V. Sahakian, and S. Swiryn, "Detection of atrial activity from high-voltage leads of implantable ventricular defibrillators using a cancellation technique," *Biomedical Engineering, IEEE Transactions on,* vol. 45, pp. 229-234, 1998.

[9] J. J. Rieta, F. Castells, C. Sanchez, V. Zarzoso, and J. Millet, "Atrial activity extraction for atrial fibrillation analysis using blind source separation," *Biomedical Engineering, IEEE Transactions on,* vol. 51, pp. 1176-1186, 2004.

[10] R. Couceiro, P. Carvalho, J. Henriques, M. Antunes, M. Harris, and J. Habetha, "Detection of atrial fibrillation using model-based ECG analysis," in *ICPR 2008. 19th International Conference on Pattern Recognition, 2008.*, 2008, pp. 1-5.

[11] Y. Sun, K. Chan, and S. Krishnan, "Characteristic wave detection in ECG signal using morphological transform," *BMC Cardiovascular Disorders,* vol. 5, pp. 1-7, 2005.