Physiologically Motivated Classification of Atrial Fibrillation

R. Couceiro, J. Henriques, R. P. Paiva, M. Antunes, P. Carvalho

Abstract— **Atrial Fibrillation (AF) is the most common arrhythmia and it is estimated to affect 33.5 million people worldwide. AF is associated with an increased risk of mortality and morbidity, such as heart failure and stroke and affects mostly older persons and persons with other conditions (e.g. heart failure and coronary artery disease). In order to prevent such life threatening and life quality reducing conditions it is essential to provide better algorithms, capable of being integrated in low-cost personalized health systems.**

This paper presents a new algorithm for AF detection, which is based on the analysis of the three physiological characteristics of AF: 1) Irregularity of heart rate and; 2) Absence of P-waves and 3) Presence of fibrillatory waves. Based on these characteristics several features were extracted from 12-lead electrocardiograms (ECG) and selected according to their discrimination ability. The classification between AF and non-AF episodes was performed using an SVM classification model.

Our results show that the identification of the fibrillatory patterns, using the proposed features, extracted from the analysis of 12-lead ECG improves the performance of the algorithm to a sensitivity of 88.5% and specificity 92.9%, when compared to our previous single-channel approach.

The proposed algorithm is currently integrated in the feature extraction module that is being developed under the WELCOME project.

I. Introduction

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia and is still the major cause of stroke, heart failure, sudden death and cardiovascular morbidity [1]. It is estimated that this arrhythmia affects about 3% of adults and its prevalence increases to 8% in older persons [1]. By 2060 it is anticipated that 17.9 million people will suffer from AF in Europe and 6-12 million in United States by 2050 [2].

In UK, the costs associated with AF-related complications and treatment are estimated in 1% of total healthcare spending, while in the US these reach 6.0–26.0 billion dollars (in 2008), and it is estimated that the burden represented by AF to healthcare systems will continue to increase unless AF is timely prevented and treated [1].

AF is the result of the continuous conduction of multiple re-entrant wavelets propagating through the atria in a chaotic fashion leading to its partial disorganization. From the inspection of the electrocardiogram, it can be identified by the absence of a normal atrial activity (P-wave before the QRS-complex), being replaced by characteristic fibrillatory waves, by an irregular heart rate, or both.

Several approaches have been proposed in literature to detect the AF based on these characteristics, using single and multi-lead ECG signals. In [3], Moody et al. proposed the analysis of the Hidden Markov Model transition probabilities to assess the irregularity of the heart rate. Additionally, linear and non-linear analysis of auto-regressive (AR) models [4] and histogram-based statistical analysis [5] have also been proposed.

In the analysis of the atrial activity (AA), two main research lines have been followed using techniques such as blind source separation, spatio-temporal cancellation and artificial neural networks. In single-lead ECG analysis, researchers focus mainly on the QRS-T cancelation based on wavelet transforms (e.g. [6, 7]) and template-based approaches (e.g. [8, 9]). In the multi-lead ECG analysis, the main methods focus on the identification and characterization of the atrial activity based on blind source separation techniques, such as independent component analysis (ICA) [10]. In the latest, a proper analysis of the features extracted from the AA and capable of being used in the discrimination between AF and non-AF episodes is still lacking.

In our previous work [11] we proposed an algorithm for detection of AF combining features related to the heart rate (HR) irregularity and to the changes in the atrial activity, extracted from the analysis of a single-lead ECG signal.

In this work, we propose a novel algorithm for the detection of AF, which complements the analysis of the HR irregularity with features extracted from the temporal and spectral analysis of the AA 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 and five from AA analysis) were selected for classification purposes.

II. Methods

The proposed algorithm 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. The

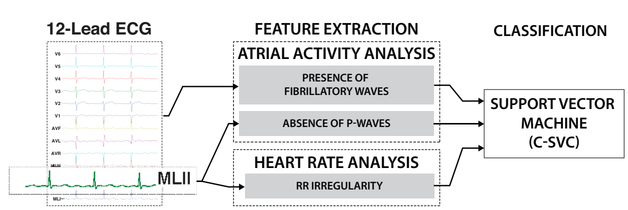

Figure 1. Structure of the proposed algorithm.

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).

Preprint submitted to 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Received February 17, 2017.

algorithm scheme is presented in Figure 1.

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.* [12] 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 2) with three states [3]: small (S_S), regular (S_R) and long (S_L) 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_R and the probability of transition from S_R to S_R state quantifies the regularity of the heart rate, i.e. a high S_R-to-S_R 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:

$$F_1 = P(S_R) \tag{1}$$

$$F_2 = P\left(S_i, S_j\right) = P\left(S_i | S_j\right) \times P(S_i) \tag{2}$$

where $i=R$ and $j=R$ 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. While the TrP matrices corresponding to non-AF episodes present a dirak-impulse-like distribution concentrated around the S_R-to-S_R 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).

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

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).

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

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 [11].

2) 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.

To discriminate this process two approaches have been followed from the single and multi-lead perspective.

a) Absence of P-waves

To access this characteristic we adopted a single-lead analysis approach, where the main objective is to seek 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:

$$F_5 = R_{Pwaves} = \frac{N_{SP}}{N_{RR}} \tag{5}$$

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.

b) Presence of fibrillatory waves

To identify the presence of fibrillatory waves, a multi-lead approach has been used. Here, the main objective is to seek for the presence of these waves, which are characterized by spectrum frequencies ranging from 5 to 8 Hz (herein defined as AF_{int}). Another characteristic of AF episodes is the absence of harmonics and the almost nonexistence of amplitudes above 15 Hz in the spectrum [10].

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 [10].

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

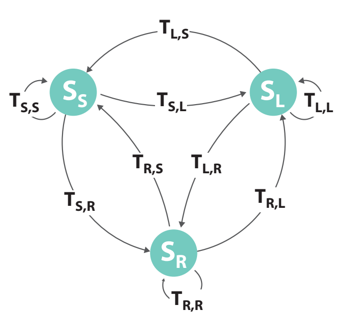

Figure 2. Structure of the Hidden Markov model used to analyze the HR.

Preprint submitted to 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Received February 17, 2017.

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 spectrum entropy (F_7) and the Kullback–Leibler divergence between the spectrum and a generalized bell-shaped membership function (F_8):

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

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:

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

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

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)

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

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 we defined F_{11} as:

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

The last extracted feature (F_{12}) is the kurtosis value of the component corresponding to the atrial activity, i.e. the lowest kurtosis value of the separated components.

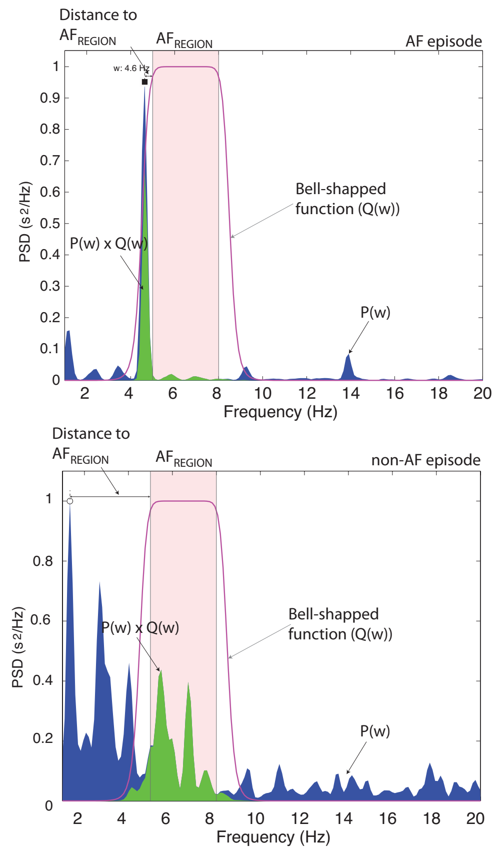

Figure 3. Spectra of the atrial activity of AF and non-AF episodes and corresponding extracted features.

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. Datasets

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 "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 (FS) metric. A ROC analysis was performed for each feature using a 6-fold cross validation approach, leading to the selection of

Preprint submitted to 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Received February 17, 2017.

TABLE II
PERFORMANCE OF THE EXCTRACTED FEATURES

Feature domain / number	FS(%) avg ± std	SE(%) avg ± std	SP(%) avg ± std	PPV(%) avg ± std
HR / 4	83,9 ± 2,9	81,1 ± 3,4	83,9 ± 2,9	83,9 ± 2,9
HR / 3	75,3 ± 4,3	68,1 ± 4,5	75,3 ± 4,3	75,3 ± 4,3
AA / 6	68,6 ± 2,7	64,9 ± 2,4	68,6 ± 2,7	68,6 ± 2,7
AA / 8	67 ± 2,7	68 ± 0,5	67 ± 2,7	67 ± 2,7
AA / 10	64,1 ± 4,4	65,1 ± 3,4	64,1 ± 4,4	64,1 ± 4,4
AA / 5	61,8 ± 2,9	63,2 ± 2,1	61,8 ± 2,9	61,8 ± 2,9
HR / 1	53,5 ± 3,9	45,3 ± 4,4	53,5 ± 3,9	53,5 ± 3,9
AA / 7	50,2 ± 2,3	60,8 ± 1,4	50,2 ± 2,3	50,2 ± 2,3
HR / 2	47,1 ± 2,9	36,5 ± 2	47,1 ± 2,9	47,1 ± 2,9
AA / 12	41,2 ± 3,6	43,9 ± 3,7	41,2 ± 3,6	41,2 ± 3,6
AA / 11	32,4 ± 5	39,9 ± 7,3	32,4 ± 5	32,4 ± 5
AA / 9	31,7 ± 2,8	25,5 ± 2,5	31,7 ± 2,8	31,7 ± 2,8

eight features (presented in Table II). The best features were extracted from the HR analysis (F_4 and F_3), followed by four features from the AA analysis (F_6, F_8, F_{10} and F_5). Four features from both HR and AA analysis (F_2, F_{12}, F_9 and F_{11}) presented a F-score below the 50% and therefore were not selected.

C. *Algorithm 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 III we present the results achieved by the single-lead [11] 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.

TABLE III
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 [11]	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

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 three main characteristics of AF: the irregularity of the RR interval and the remodelling of the atrial activity, described by the absence of the P-waves, which are replaced by a fibrillatory waves. 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.

Future work will focus on the testing of other source separation techniques and combinations in order to achieve a better separation between atrial and ventricular activity, as well as to evaluate this algorithm in a larger database.

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).

REFERENCES

[1] P. Kirchhof, S. Benussi, D. Kotecha, A. Ahlsson, D. Atar, B. Casadei, *et al.*, "2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS," *European Heart Journal,* vol. 37, pp. 2893-2962, 2016.

[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] K. Ivanko, N. Ivanushkina, and Y. Karplyuk, "Atrial electrical activity extraction for atrial fibrillation assessment," in *2016 IEEE 36th International Conference on Electronics and Nanotechnology (ELNANO)*, 2016, pp. 192-197.

[10] 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.

[11] 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.

[12] 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.