Atrial fibrillation detection through heart rate variability using a machine learning approach and Poincare plot features

J.P. Sepulveda-Suescun[1], J. Murillo-Escobar[1], R.D. Urda-Benitez[1], D.A. Orrego-Metaute[1] and A. Orozco-Duque[1]

[1] Instituto Tecnológico Metropolitano, Grupo de Investigación e Innovación Biomédica Gi2B, Medellin, Colombia

Abstract— **Atrial Fibrillation (AF) is a common cardiac arrhythmia, and it has a high rate of morbidity and mortality. In this paper, an algorithm for automatic AF episodes detection based on novel low computational cost features is proposed. The features are based on Poincare plots calculated from heart rate variability signal. A supervised classification technique, Support Vector Machines, optimized with Particle Swarm Optimization, was implemented. The data was obtained from MIT-BIH Atrial Fibrillation and Normal Sinus Rhythm Databases. This method shows an accuracy of 92.9% to detect AF spontaneous episodes in signals from AF patients, and 97.8% to classify between AF episodes from AF patients and episodes from subjects with normal sinus rhythm. The proposed method can be employed in real time applications due to its performance as well for its low computation time around 8.8 ms per episode.**

Keywords— **Atrial fibrillation, SVM, Poincare plot, PSO, heart rate variability.**

I. INTRODUCTION

Atrial fibrillation (AF) is one of the most common cardiac arrhythmia worldwide, having a significant mortality rate, because AF may increases the blood coagulation, which cause 15% of strokes [1, 2]. The AF in the electrocardiogram (ECG) is characterized by irregularity present in the RR intervals and the replacement of the P wave by rapid oscillations and fibrillatory waves that vary in time, amplitude and shape [3], therefore, the diagnosis of paroxysmal AF presents a major challenge for clinicians and researchers because of its asymptomatic nature and spontaneous, especially in its early stages[4]. Do to this, there is an urgent need to develop methods for accurate and rapid detection of paroxysmal AF.

Different methods have been proposed to detect AF from the ECG signal using wavelet transform, Shannon entropy sequence analysis of the ECG signal and machine learning methods as genetic programming optimized simulated annealing and artificial neural networks[5, 6, 7, 8, 9], however, this strategy represents a high computational cost since it is sometimes necessary to segment the P wave or analyzing raw signal where there is information that is not relevant. Therefore, it has been proposed AF detection by RR intervals, because this offers a more efficient strategy from the computational point of view, in this context have been developed methodologies based on statistical characteristics of the Heart Rate Variability (HRV), which is a more robust method because the time intervals R-R are less affected by noise [10, 11].

In this paper, we present a method based on the estimation of novel features from RR interval and HRV to detect the presence of AF and Normal sinus rhythm using signals from MIT-BIH Atrial fibrillation database and MIT-BIH Normal Sinus Rhythm Database. We proposed a method based on Poincare plot with low computational cost in a support Vector machine optimize by particle swarm optimization.

II. MATERIALS AND METHODS

The proposed methodology to the AF detection is show in Fig 1. As it can be seen, the detection system comprises three steps: preprocessing, characterization and classification, each step are detailed in the following subsections. With this methodology we intend to differentiate short events of AF to Normal Sinus Rhythm (NSR), in this sense we conduce two experiments, the first one using only signals that had at least one episode of AF or NSR, given that in literature is the most common approach, the second experiment we employ only signals from AF patients during NSR with spontaneous episodes of AF, having a greater proximity to real application due to the spontaneous nature of this affection.

A. *Data Base Description*

MIT-BIH Atrial Fibrillation Database (AFDB) is the most popular and often used publicly available database for AF detection, this data base has 23 ECG recording of subjects such had suffered significant arrhythmia each arrhythmia. Each record has a duration of about 10 hours. A total of 291 AF episodes, 14 of atrial flutter, 12 of junctional rhythm and 288 of NSR were found. [7]. MIT-BIH Normal Sinus Rhythm Database (NSRDB) includes 18 ECG recordings of subjects

I. Torres et al. (eds.), *VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th -28th, 2016*,
IFMBE Proceedings 60,
DOI: 10.1007/978-981-10-4086-3_142

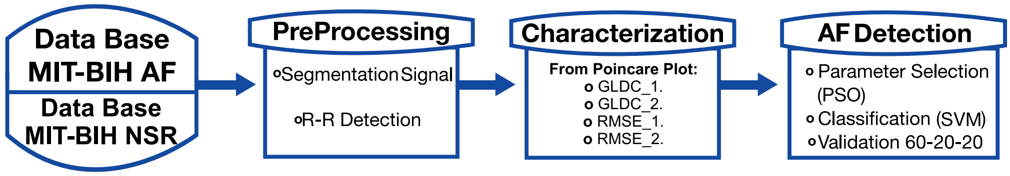

Fig. 1. Proposed Methodology

such had not suffered significant arrhythmia [12].

The analysis of developed algorithm for AF detection was made in two experiments. The first one using signals with at least one event of AF or NSR, in this regard we employ AFDB to extract signals with AF and the NSR signals were get from NSRDB. For the second experiment we use only signals that has both rhythms, which were found only in AFDB.

B. *Preprocessing*

The preprocessing begin, with the extraction the RR intervals from the ECG signals within the database. For this task, the entries in the database referenced with (.ATR) were used, then was assessed the duration of the events of atrial fibrillation and normal sinus rhythm, for the AF and NSR events greater to 30 seconds were selected from the signals. A windows with length of 30 seconds were construct for each event, centered around the middle point. For calculate the Heart Rate Variability (HRV) we start detecting R peaks employing the Pan-Tompkins method, due to is capability of been implemented in real time [13].

C. *Characterization*

The *Poincare Plot* is a scatter-plot of the pairs ($HRV[n]$, $HRV[n+m]$), showing the relationship among m consecutive beats [14]. The Poincare plot in NSR episode looks like a cluster of points grouped linearly unlike, the Poincare plot in AF episode evidenced a complete dispersal of the points (Fig 2), this occurs because of the higher erratic variability of the HRV in AF episode with respect to the NSR episode.

From the Poincare plot, linear dependence between $HRV[n]$ and $HRV[n+1]$ is analyzed, signals with AF present greater degree of dispersal, while NSR segments shows greater linear dependence, for this was used the generalized linear dependence coefficient (**GLDC_1**) proposed in [15] which is calculated as:

$$D = 1 - |R|^{1/(p-1)} \tag{1}$$

Where p is the number of involved variables and $|R|$ represents the determinant of correlation matrix, R is obtained from the covariance matrix S between $HRV[n]$ and $HRV[n+1]$ as $R = D^{-1/2} S D^{-1/2}$, D is a square matrix that retains only the main diagonal of S.

Moreover we extend this analysis to measure the generalized linear dependence coefficient between five consecutive heart beats (**GLDC_2**) using (1), in this case the correlation matrix S is calculated between $HRV[n]$, $HRV[n+1]$ $\cdots$ $HRV[n+5]$. Now, assuming that in NSR is possible find

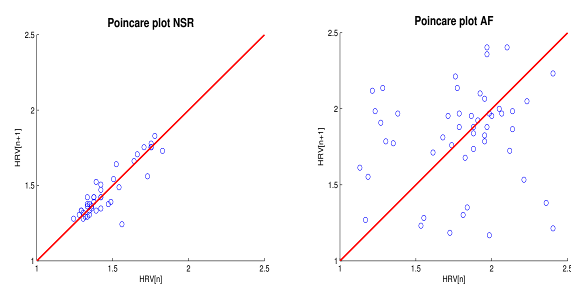

Fig. 2. Poincare plots of NSR and AF

$HRV[n+1]$ as linear function of $HRV[n]$ using regression, we propose measures such capacity through the difference between the actual $HRV[n+1]$ and the predicted one, following the next procedure. Let L be a vector that contain $HRV[n]$ for $n = 1, \cdots l-1$, and l is the number of heart beats of the signal, P is a vector with $HRV[n+1]$, if we estimate P as $\widehat{P} = \omega L$ is possible measure the Root Mean Square Error (RMSE) between $\widehat{P}$ and P (**RMSE_1**). The weight coefficient ω can be found using the normal equation as:

$$\omega = \left(L^T L\right)^{-1} L^T P \tag{2}$$

Is possible extend this analysis for five consecutive heart beats, for this case P will be a vector with $HRV[n+5]$, for $n = 1, \cdots l-5$ and L be a matrix in which every column represents $HRV[n]$, $HRV[n+1]$ $\cdots$ $HRV[n+4]$, repeating the above formulation is possible find the RMSE between the actual P and the predicted (**RMSE_2**).

Aditionally to Poincare plot featues, standard deviation of HRV (**STD_HRV**) was implemented due to it has been widely study in different papers [9, 16, 11].

D. AF Detection

The goal of this paper is to classify each signal segment into AF or NSR, using the proposed extracted features. To achieve this, we use Support Vector Machines (SVM) as our classification algorithm. SVM is supervised classifier proposed for Vapnik in 1995, the idea behind this methods is find an hyperplane that separate the data in their corresponding classes. Given a training data $\{x_i, y_i\}$, where $x_i \in \mathbf{R}^d$, $i = 1, 2 \cdots l$ and $y_i \in \{-1, 1\}$, this way the hyperplane can be found solving the dual problem formulation (3), this formulation allow us for non-linear solutions introducing a kernel function $\phi(x_i, x_j)$, note that in its dual form the optimization problem appears only in terms of the Lagrange multipliers α_i [17]. In this paper we employ the Radial Basis Kernel (RBK) due to it is the dot product in the infinite dimensional space which minimizing both the estimation and approximation errors of classifier in this regard the RBK function is $\phi(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$.

$$\begin{gathered} \max \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j \phi(x_i, x_j) \\ \text{Subject to:} \begin{cases} 0 \leq \alpha_i \leq C \\ \sum_{i=1}^{l} \alpha_i y_i = 0 \end{cases} \end{gathered} \tag{3}$$

For the well performance of the SVM is necessary a good selection of the parameter, we selected the C and γ of RBK through Particle Swarm Optimization (PSO) which algorithm is described in [18].

The data obtained in the characterization stage were subjected to a linear normalization between 0 and 1, to avoid redundancy in the data and increase the performance. Subsequently the data were randomly split into three subsets, one with 60% of the data to train the SVM, and two sets more, each with 20% of the data, we use the first subset (validation) for PSO to get the parameter combination that maximize the accuracy, the reminder data were employ to assessed the overall performance of the proposed system in terms of the accuracy, sensitivity and specificity.

III. RESULTS AND DISCUSSION

A. Preprocessing

For the first experiment in which is necessary had 30 seconds of signal that has at least one event of AF or NSR, we obtain from AFDB 226 segments of signal with AF and 234 of NSR from NSRDB. In the second setup we extract from AFDB signals that has both of the study rhythms getting 226 segments of signal with AF and 264 of NSR.

B. Characterization

We extracted five features of the HRV, one of these commonly used in literature and the other four calculated from the Poincare plot were proposed in this paper for detecting AF. Our studies show that these features can be used as good markers for detection of AF as can be seen in Fig 3, from the boxplot is clear that features based in the analysis of five consecutive heart beats represent a most robust approach due can be less sensible to anomaly beats like premature ventricular contractions in contrast to statistical methods as the used in [11, 7].

C. AF Detection

The performance of the methodology is listed in Table 1, the proposed methodology on the experiment two presents performance values about 4 points below the results reported by other approaches [7], but these have the disadvantage of working with complete ECG signal and use characteristics with a very high computational cost, allowing these methodologies presented better performances but away from the possibility to be implemented in real time, contrasting with this experiment of our methodology, which were estimated characteristics of the same database and with a low computational cost, meanwhile in the experiment one the obtained performance is equivalent to another approaches [6, 9]. The PSO gave for both experiments a parameter $C = 131$ and $\gamma = 9.5$.

Table 1. Classification Performance

	Sen	Spe	Acc	F1
Exp 1	97.9%	97.8%	97.8%	97.9%
Exp 2	91.3%	94.3%	92.9%	92.3%

Sen: Sensitivity, Spe: Specificity, Acc: Accuracy, F1: F1 score, Exp 1: Experiment 1, Exp 2: Experiment 2

The mean time for the characterization of each of the signal segments for both experiments of our methodology, with Matlab 2014a on an Intel Core CPU to 2.0GHz, 8 RAM and Windows of 64 bits, was of 8.8 *mS*.

IV. CONCLUSION

In this paper, we presented novels features from Poincare plot, based on the HRV signal extracted from ECG recordings, focused to detect spontaneous atrial fibrillation events. These features show good difference leading to better classification using supervised techniques. The developed experiments show the feasible implementation of a low computational cost methodology for discrimination of paroxysmal AF episodes from NSR in the same HRV signals. The proposed

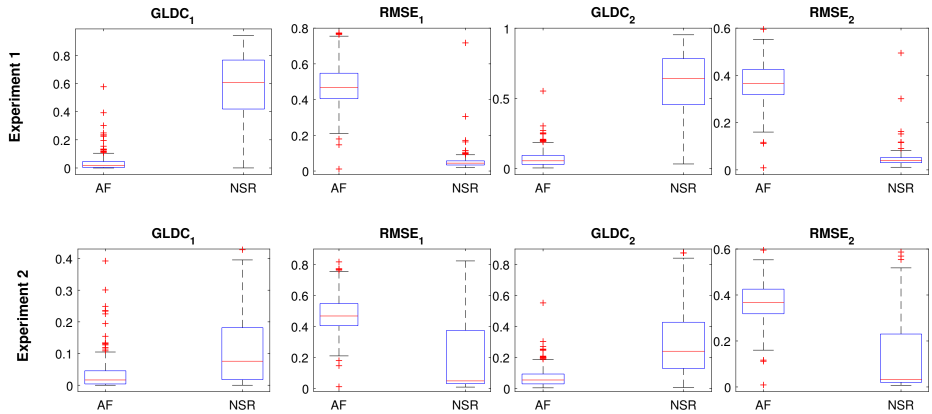

Fig. 3. Boxplots proposed features

method will be implemented in embedded systems in futures developments due its factible implementation in real time.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Chen J, Yang Z G, Xu H Y, Shi K, Long Q H, Guo Y K. Assessments of pulmonary vein and left atrial anatomical variants in atrial fibrillation patients for catheter ablation with cardiac CT *Eur Radiol.* 2016.
2. Mohebbi Maryam, Ghassemian Hassan. Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal *Computer Methods and Programs in Biomedicine.* 2012;105:40–49.
3. Fuster Valentin, Rydén Lars E, Cannom David S, et al. ACC/AHA/ESC 2006 Guidelines for the Management of Patients with Atrial Fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice *Circulation.* 2006;114:257–354.
4. Yoon Kwon Ha, Thap Tharoeun, Jeong Chang Won, et al. Analysis of Statistical Methods for Automatic Detection of Congestive Heart Failure and Atrial Fibrillation with Short RR Interval Time Series in *2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*:452–457 2015.
5. Ghodrati Alireza, Murray Bill, Marinello Stephen. RR Interval Analysis for Detection of Atrial Fibrillation in ECG Monitors in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*;2008:601–604 2008.
6. Lian Jie, Wang Lian, Muessig Dirk. A simple method to detect atrial fibrillation using RR intervals. *American journal of cardiology.* 2011;107:1494–1497.
7. Asgari Shadnaz, Mehrnia Alireza, Moussavi Maryam. Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine *Computers in Biology and Medicine.* 2015;60:132–142.
8. Petrenas Andrius, Marozas Vaidotas, Lukosevicius Arünas, Sörnmo Leif. A Noise-Adaptive Method for Detection of Brief Episodes of Paroxysmal Atrial Fibrillation in *Computers in Cardiology*:739–742 2013.
9. Yaghouby Farid, Ayatollahi Ahmad, Bahramali Reihaneh, Yaghouby Maryam, Alavi Amir Hossein. Towards automatic detection of atrial fibrillation: A hybrid computational approach *Computers in Biology and Medicine.* 2010;40:919–930.
10. Tateno K., Glass L.. A method for detection of atrial fibrillation using RR intervals in *Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163)*:391–394 2000.
11. Ruan Xiuhua, Liu Changchun, Liu Chengyu, Wang Xinpei, Li Peng. Automatic detection of atrial fibrillation using R-R interval signal in *The 4th International Conference on Biomedical Engineering and Informatics (BMEI 2011)*:644–7 2011.
12. Goldberger A. L., Amaral L.A.N., Glass L., et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals *Circulation.* 2000;101:215–220.
13. Pan J, Tompkins W J. A real-time QRS detection algorithm. *IEEE transactions on bio-medical engineering.* 1985;32:230–236.
14. Brennan M., Palaniswami M., Kamen P.. Do existing measures of Poincar?? plot geometry reflect nonlinear features of heart rate variability? *IEEE Transactions on Biomedical Engineering.* 2001;48:1342–1347.
15. Peña Daniel, Rodríguez Julio. Descriptive measures of multivariate scatter and linear dependence *Journal of Multivariate Analysis.* 2003;85:361–374.
16. Maji U., Mitra M., Pal S.. Differentiating normal sinus rhythm and atrial fibrillation in ECG signal: A phase rectified signal averaging based approach in *International Conference on Control, Instrumentation, Energy and Communication, CIEC 2014*:176–180 2014.
17. Burges Christopher J.C.. A Tutorial on Support Vector Machines for Pattern Recognition *Data Mining and Knowledge Discovery.* 1998;2:121–167.
18. Kennedy J, Eberhart R. Particle swarm optimization *Neural Networks, 1995. Proceedings., IEEE International Conference on.* 1995;4:1942–1948.

Author: Juan Pablo Sepúlveda Suescún
Institute: Instituto Tecnológico Metropolitano
Street: Calle 73 No 76A - 35
City: Medellín
Country: Colombia
Email: jsepulveda1331@hotmail.com