An Efficient Pattern Recognition Kernel-Based Method for Atrial Fibrillation Diagnosis

Zouhair Haddi[1], Bouchra Ananou[1], Youssef Trardi[1], Jean-François Pons[2], Stéphane Delliaux[3], Mustapha Ouladsine[1], Jean-Claude Deharo[4]

[1]Aix Marseille University, University of Toulon, CNRS, LIS, Marseille, France
[2]Aix Marseille University, University of Toulon, CNRS, IM2NP, Marseille, France
[3]Aix Marseille University., INSERM, INRA, C2VN, Marseille, France
[4]Centre Hospitalier Universitaire de la Timone, Marseille, France

Abstract

The aim of this work is to develop an efficient diagnosis method for atrial fibrillation (AF) arrhythmia based on inter-beat interval time series analysis and relevance vector machine (RVM) classifier. Automatic and fast AF diagnosis is still a major concern for the healthcare professional. Several algorithms based on univariate and multivariate analysis have been developed to detect AF. The published results do not show satisfactory detection accuracy especially for brief duration as short as one minute. Although RVM has been applied on tasks such as computer vision, natural language processing, speech recognition etc., this is the first attempt to adopt RVM for AF diagnosis. Four publicly-accessible sets of clinical data (AF Termination Challenge Database, MIT-BIH AF, Normal Sinus Rhythm RR Interval Database, and MIT-BIH Normal Sinus Rhythm Databases) were used for assessment. All time series were segmented in 1 min RR interval window and then three specific features were calculated. The RVM classifier was trained on 2000 randomly selected samples from the merged database. The results showed that the RVM model performed better than do existing algorithms, with 99.20% for both sensitivity and specificity.

1. Introduction

Among cardiac arrhythmias, atrial fibrillation is of particular concern because of its recognized increased risk of stroke, heart failure, morbidity and all-cause of mortality [1]. Referred to as a "silent killer" because it does not always produce symptoms that alert patients/Doctors to the problem, its automatic detection is strongly advisable. Moreover, due to its unpredictable occurrence and the short-time periods the patient may cope with it, long-time electrocardiography (ECG) recordings are regularly needed. The analysis of ECG/Holter recordings is a crucial point to diagnose patient health status. Indeed, the absence of P-waves (replaced by rapid oscillations or fibrillatory waves) or irregularities of R–R intervals, or combination of both characteristics are the main morphological features to detect AF episodes. In recent years, several algorithms have been developed to detect AF specifically when short duration occurs [2]. Most of them are based on inter-beat interval time series analysis. Although progress in the published results is seen, there is still scope for improvement which needs to be addressed especially for brief duration as short as one minute or less. In this context, J. Lian et al. [3] have proposed map based algorithm of R-R intervals versus change of R-R intervals (RdR) to detect AF. The map is divided by a grid with 25-ms resolution in 2 axes and nonempty cells are counted to classify AF and non-AF episodes. The best sensitivity and specificity have been achieved for window size of 128 R-R time series (95.9% and 95.4% respectively). Ghodrati et al., have investigated two R-R irregularity measures for AF detection in ECG monitors: Normalized absolute deviation of R-R intervals and normalized absolute difference between successive R-R intervals. The receiver operating characteristic analysis results showed 89.33% as overall of sensitivities of three AF datasets (MIT AFDB, MIT–BIH Arrhythmia DB and Drager AFDB). S. Hargittai has investigated the performance of several features, extracted from segments of 80 beats to detect AF [5]. He confirmed that the use of the scatter plot of successive RR differences (dRR Lorenz Plot) and Sample Entropy yielded an overall error rate of about 5% for Physionet datasets (MIT-BIH arrhythmia, atrial fibrillation and long-term atrial fibrillation databases).

X. Ruan et al. have attempted to distinguish AF from normal sinus rhythm using scatter plot of one-minute R-R interval signal [6]. Four geometrical features were used, namely, Vector Angular Index (VAI), Vector Length Index (VLI), Dispersion of points along the diagonal line

(SD2) and perpendicular to the diagonal line (SD1). Each index is tested using Mann-Whitney test to see if there is significant difference between the two groups. The best parameters that give very high performance were SD1 and SD2 which both of them reached 98.3% and 100% of sensitivity and specificity respectively. These results have been obtained based on datasets containing only 120 recordings (60 from AF Termination Challenge Database and 60 normal ECG recordings coming from healthy people) of one-minute duration which is not significant for a universally applicable tool.

In the current study, three of the four aforementioned geometrical features were used, namely, VAI, VLI, and SD1 have been exploited as input to PCA and RVM to diagnosis AF of 1-min episode.

2. Methods

2.1. Databases

PCA and RVM were evaluated on four publicly-accessible sets of clinical data: AF Termination Challenge Database, MIT-BIH AF, Normal Sinus Rhythm RR Interval Database, and MIT-BIH Normal Sinus Rhythm Databases. All time series were segmented in 1-min RR interval windows (total of 47156 and 4902 time series for NSR and AF respectively). Usually, authors trained their algorithms on one of these datasets and tested them on the remaining. We believe that if the four datasets are merged (and categorized into AF and NSR groups), this could give rise to huge variability and therefore to better characterize AF rhythm [7].

2.2. Feature extraction

Three features were extracted from the scatter plot (defined as a diagram in which each R-R interval is plotted as a function of the previous R-R interval) to classify AF from NSR. The scatter plot of NSR seems as sticky, in that nearly all the scatter points are centralized along the diagonal line. However, the scatter plot of AF seems as unfolded fan-shaped, in which all the points are dispersed around the whole plot. These features are calculated as follows:

VAI is calculated as the mean of all the absolute value of angular differences between the lines plotted from every scatter point to the original point and the diagonal line, measuring the angular dispersion of all the points. VLI is calculated as the standard deviation of all distances of scatter points from the original point, measuring the distance dispersion of all the points. They can be defined as:

$$VAI = \frac{\sum_{i=1}^{N} |\theta_i - 45|}{N} \qquad (1)$$

$$VLI = \frac{\sqrt{\sum_{i=1}^{N}(l_i - L)^2}}{N} \qquad (2)$$

where θ_i is the angle between the line plotted from every scatter point to the original point and the x-axis, l_i is length between every scatter point and the original point, L is the mean of all the l_i, N is the number of scatter points.

SD1 is calculated as the standard deviation of the distances of points from y = x axis, measuring the width of the ellipse and indicating the short-term variability and is defined as:

$$SD1 = STD\left(\frac{|RR_{n+1} - RR_n|}{\sqrt{2}}\right) \qquad (3)$$

where RR_n is an R-R interval series with $n=1,2,…,N$-1, RR_{n+1} is the same as RR_n index-shifted by 1 and $STD(x)$ represents the standard deviation of x.

2.3. Pattern recognition methods

PCA is a very well-known unsupervised method often employed in ECG signal processing [7,8]. The main objective of PCA consists in expressing the information contained in a dataset by a smaller number of variables called principal components. These principal components are linear combinations of the original response vectors. The principal components are chosen to contain the maximum data variance and to be orthogonal. Hence, PCA allows the reduction of multidimensional data to a lower dimensional approximation, while simplifying the interpretation of the data by the first two or three principal components (PC1, PC2, and PC3) in two or three dimensions and preserving most of the variance in the data [9,10].

RVM—a method first introduced by Tipping [11]—is a data-driven method with the Bayesian treatment of the support vector machine (SVM) pipeline [12,13] as defined by Equation 4. RVM is a Bayesian sparse kernel model that introduces a prior distribution over the model weights that are governed by a set of hyper-parameters [14]. In comparison with the equivalent SVM, the most compelling feature of the RVM is its superior generalization performance and a shorter time for prediction because relatively few "relevance vectors" (RVs) are used in effecting the prediction [15]. In other words, these RVs, calculated by using a type-II maximum likelihood procedure, constitute the key observations of a phenomenon, provide insight on its generic nature and retain low computational complexity. The RVM also provides posterior probabilistic outputs.

$$f(x_t, w) = \sum_{i=1}^{N} w_i K(x_t, x_i) + w_0 \qquad (4)$$

Where N is the total number of observations $\{x_i\}$, x_t is a test vector, $w = [w_0, ..., w_N]$ are the weights determined by maximizing the margin of separation between the two classes, w_0 is a bias term, and $K(x_t, x_i)$ is a kernel function.

When RVM is applied to classification, the target conditional distribution $p(y|x)$ is assumed to follow a Bernoulli distribution. Assume that we have a set of training i-vectors $X = \{X_s, X_b\}$ and $y_i = 1$ when $x_i \in X_s$ and $y_i = 0$ when $x_i \in X_b$, the likelihood of the training data set can be written as [10]:

$$p(y|w) = \prod_{i=1}^{N} \sigma(f(x_i; w))^{y_i} \{1 - \sigma(f(x_i; w))\}^{1-y_i} \quad (5)$$

where $y_i \in \{0,1\}$; $N = X_s + X_b$; $y = [y_1, ..., y_N]^T$;
$$w = [w_0, ..., w_N]^T \quad (6)$$

and $\sigma\{.\}$ is the logistic sigmoid link function defined as

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (7)$$

To avoid over-fitting, RVM defines a zero-mean Gaussian prior distribution over w:

$$p(w|\alpha) = \prod_{i=0}^{N} \mathcal{N}(w_i|0, \alpha_i^{-1}) = \mathcal{N}(w|0, A^{-1}) \quad (8)$$

where $\alpha = [\alpha_0, \alpha_1, ..., \alpha_N]^T$, α_i is the hyper-parameter associated with weight w_i and $A = diag(\alpha_0, \alpha_1, ..., \alpha_N)$.

Using Eq. 5 and Eq. 8, we can obtain the posterior distribution of w:

$$p(w|y, \alpha) = \frac{p(y|w)p(w|\alpha)}{\int p(y|w)p(w|\alpha)dw} = \frac{g(w)}{p(y|\alpha)}, \quad (9)$$

where $g(w)$ is defined as $g(w) \equiv p(y|w)p(w|\alpha)$

Taking logarithm of $g(w)$:

$$\log g(w) = \log p(y|w)p(w|\alpha)$$

$$= \sum_{i=1}^{N} \{y_i \log[\sigma(f(x_i; w))] + (1 - y_i)\log[1 - \sigma(f(x_i; w))] - \frac{1}{2}w^T A w + C$$

$$= \sum_{i=1}^{N} \{y_i \log[\sigma(\phi(x_i)^T w)] + (1 - y_i)\log[1 - \sigma(\phi(x_i)^T w)] - \frac{1}{2}w^T A w + C \quad (10)$$

where $\phi(x_i)^T w = f(x_i; w)$

$$\log g(w) \approx \log w_0 - \frac{1}{2}(w - w_0)^T H(w - w_0)$$

where H is a Hessian matrix

Thus

$$q(w) = \frac{|H|^{\frac{1}{2}}}{(2\pi)^{\frac{N+1}{2}}} exp(-\frac{1}{2}(w - w_0)^T H(w - w_0))$$
$$= \mathcal{N}(w|w_0, H^{-1})$$

which is a Gaussian distribution with mean w_0 and covariance matrix H^{-1}.

Finally, we use $q(w)$ to approximate the posterior $p(w|y, \alpha)$ around the mode w_0.

3. Results and discussion

3.1. PCA analysis results

PCA analysis was performed to elucidate whether it was possible to discriminate AF and NSR groups. Before applying PCA, a mean centring pre-processing technique was applied to the AF and NSR dataset. Fig. 1 shows the obtained PCA plot (PC 1 vs. PC 2). The variances explained by the first and the second principal components are 98.59%. Although two principal components contain large amount of original information, some samples of AF and NSR classes in two-dimensional space overlap with each other. This means that the problem under investigation requires very powerful classification tool in order to separate correctly the two clusters.
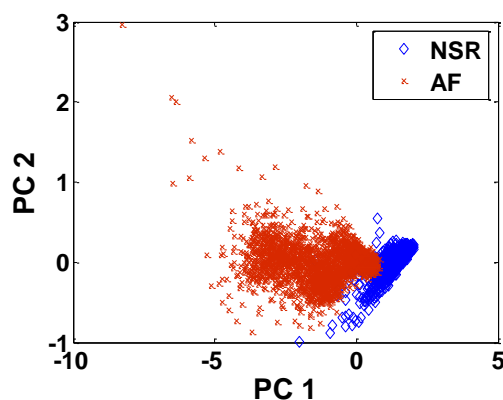

Figure 1. Scores plot of PCA performed on AF and NSR of 1-min episodes.

3.2. RVM results

Since the aim of this work was to identify AF episodes of short duration, RVM was applied to develop the

classifier model. In the training process of a generalized linear model (4) under the RVM scheme described in the previous section, only the weights and hyper-parameters are optimized. It is assumed that the kernel functions are given. Yet the performance of the model depends intensely on the choice of the kernel functions and the value of their parameters. In the work presented in this paper the kernel function is a radial basis function with σ equal 4. The selected RVM classifier has reached a 99.20% success rate in the identification of the AF and NSR groups. Table 1 shows the confusion matrix of the RVM classifier. Rows indicate true categories and columns predicted categories. As it can be noticed in this table, of 2000 1-min segments of AF and NSR, only 16 mistakes occurred: 8 1-min episodes belonging to AF were misclassified as belonging to NSR, and 8 1-min episodes belonging to NSR were misclassified as belonging to AF. In other words, 99.20% have been reached as sensitivity and specificity.

Table 1. RVM classification results.

Actual	Predicted	
	AF	NSR
AF	992	8
NSR	8	992

4. Conclusion

The main purpose of this work is to develop an efficient diagnostic method able to exploit short heartbeat segments for an accurate automatic atrial fibrillation screening. The AF diagnosis method was based on Relevance vector machine. This machine learning technique has been optimized to recognize correctly AF and NSR of one minute episode. For upcoming works, we attempt to detect very short AF episodes (~ 30s) from RR interval signals.

Acknowledgements

References

[1] K. Miyazawa, G.YH. Lip, Atrial Fibrillation, *Medicine*, vol. 46, no. 10, pp. 627–631, Oct. 2018.

[2] J-F. Pons, Z. Haddi, J-C. Deharo, A. Charaï, R. Bouchakour M. Ouladsine, S. Delliaux, Heart Rhythm Characterization Through Induced Physiological Variables, *Scientific Reports*, vol. 7, Jul. 2017, Art. no. 5059.

[3] J. Lian, L. Wang, D. Muessig, A Simple Method to Detect Atrial Fibrillation Using RR Intervals, *Am. J. Cardiol.*, vol. 107, no. 10, pp. 1494–1497, May 2011.

[4] A. Ghodrati, B. Murray, S. Marinello, RR Interval Analysis for Detection of Atrial Fibrillation in ECG Monitors, *Conf. Proc of 30th Annual International IEEE EMBS Conference*, pp. 601-604. Aug. 2008.

[5] S. Hargittai, Is it Possible to Detect Atrial Fibrillation by Simply using RR Intervals?, *Computing in Cardiology,* vol. 41, pp. 897–900, Sept. 2014.

[6] X. Ruan, C. Liu X. Wang, P. Li, Automatic Detection of Atrial Fibrillation Using R-R Interval Signal, *Conf. Proc of 4th International Conference on Biomedical Engineering and Informatics (BMEI),* vol. 4, pp. 644–647, Oct. 2011.

[7] Z. Haddi, J-F Pons, S. Delliaux, B. Ananou, J-C Deharo, A. Charaï, R. Bouchakour, M. Ouladsine, A Robust Detection Method of Short Atrial Fibrillation Episodes, *Computing in Cardiology,* vol. 44, pp. 1–4, Sept. 2017.

[8] D. Raine, Ph. Langley, E. Shepherd, S. Lord, S. Murray, A. Murray, J.P. Bourke, Principal Component Analysis of Atrial Fibrillation: Inclusion of Posterior ECG Leads Does Not Improve Correlation with Left Atrial Activity, *Med Eng Phys.,* vol. 37 no. 2, pp. 251–255, Feb. 2015.

[9] Z. Haddi, A. Amari, F. E. Annanouch, A, Ould Ali, N. El Bari, B. Bouchikhi, Potential of a Portable Electronic Nose for Control Quality of Moroccan Traditional Fresh Cheeses, *Sensor Letters*, vol. 9, no. 6, pp. 2229–2231, Dec. 2011.

[10] Z. Haddi, F. Annanouch, A. Amari, A. Hadoune, B. Bouchikhi, N. El Bari, Application of a Portable Electronic Nose Device to Discriminate and Identify Cheeses with Known Percentages of Cow's and Goat's Milk, IEEE Sensors Conference, pp. 771–774, Nov. 2010.

[11] M. Tipping, Sparse Bayesian Learning and The Relevance Vector Machine, *J. Mach. Learn. Res.,* vol. 1, pp. 211–244. Jun. 2001.

[12] Y. Trardi, B. Ananou, Z. Haddi, M. Ouladsine, Multi-Dynamics Analysis of QRS Complex for Atrial Fibrillation Diagnosis, *5th International Conference on Control, Decision and Information Technologies (CoDIT'18),* pp. 1067–1072, Apr. 2018.

[13] Y. Trardi, B. Ananou, Z. Haddi, M. Ouladsine, A Novel Method to Identify Relevant Features for Automatic Detection of Atrial Fibrillation, *26th Mediterranean Conference on Control and Automation (MED)*, pp. 613–618, Jun. 2018.

[14] J. Hu, P.W. Tse, A Relevance Vector Machine-Based Approach with Application to Oil Sand Pump Prognostics, *Sensors*, vol. 13, no. 9, pp. 12663–12686, Sep. 2013.

[15] C. Rasmussen, C. Williams, Gaussian Processes for Machine Learning, *MIT Press: Cambridge, MA, USA, London, UK,* pp. 149–150, 2006.

Address for correspondence.

Zouhair Haddi
Aix Marseille University
Laboratoire d'Informatique et Systèmes (LIS)
Domaine Universitaire de Saint-Jérôme
Avenue Escadrille Normandie-Niemen
13397 Marseille, France
zouhair.haddi@lis-lab.fr