Classification of High Resolution ECG from Chagasic Patients with Wavelet Based Bayesian Models

R Prado[1], I García[2], P Gomis[3]

[1]Dpto. de C.C. y Estadística & CESMa, Universidad Simón Bolívar, Venezuela
[2]Dpto. Formación General y Cs. Básicas & CESMa, Universidad Simón Bolívar, Venezuela
[3]Dpto. de Tecnología Industrial & GBBA, Universidad Simón Bolívar, Venezuela

Abstract

The problem of classifying multiple signals has been studied by several authors from different perspectives. Techniques such as late potential (LP) measurements in the high-resolution electrocardiogram (HRECG), abnormal intra-QRS potentials (AIQP), and time-frequency measurements, have been used for evaluating signals of patients with chagasic myocarditis. The goal of these indexes is to identify different stages of the disease, however their predictive performance is not fully satisfactory. In this work, we extend the previous developments using a wavelet based Bayesian modelling approach for analyzing and classifying unfiltered HRECG signals. We evaluate the predictive capabilities of the proposed model through a Bayesian classification based on maximum posterior probability. The technique provides considerably higher specificity and accuracy rates and may improve the predictive values of HRECG in assessing chagasic myocarditis.

1. Introduction

In this work, we classified HRECG signals of patients with Chagasic myocarditis, affection caused by the deterioration that produces Chagas' disease to the cardiac muscle. This deterioration consists on damaged or inflamed tissue, which may produce bundle branch block (BBB), cardiac muscle expansion, pains chest, tachycardia, etc. A common end point is sudden death due to electrical instability in myocardial contraction.

The clinical diagnosis of Chagas' heart disease is based on positive serology and a combination of clinical findings. However, previous studies have suggested that HRECG - like non-invasive method - can be used for studying changes in the cardiac signals [1] and allow to characterize the disease in different levels with major precision.

The analysis of HRECG that have been developed until now is based on the use of indexes to detect abnormal signal, like late potentials (LP), frequency and time-frequency domain indexes and abnormal intra-QRS potentials (AIQP).

The LP's consists of capturing the abnormal signals that happen at the end of the QRS complex. The main disadvantage of their use for characterizing chagasic signals is that the duration of QRS is extended due to blockade problems.

The AIQP are based on the localization of abnormal potentials in any place within complex QRS [2]. These indexes have been efficient to detect chagasic patients with ventricular risk of tachycardia, but the results of the model can be slanted and in practice we needed very high order models.

In order to improve the predictive value of HRECG for chagasic myocarditis detection, we have developed a Bayesian model that allows obtaining the pattern beat of the patients in each stage of the disease, which is used for calculating the probability that a new patient (without classifying) belongs to any chagasic group. This probability is calculated using predictive Bayes factor.

Among the advantages that have the use of wavelet Bayesian Models are: firstly, that it allows to describe different stages of the disease using a reduced number of wavelet coefficients. Secondly, it allows to incorporate a prior distribution on the parameters of the model (which improves a posteriori predictive value of the model) and also, it allows to express uncertainty in terminus of probabilities, i.e., in this case, we can calculate the probability that a patient belongs to certain stage of the disease and to compute the probability that a certain coefficient of the wavelet transformation is important to characterize the interest group.

2. Methods

2.1. Data acquisition

The dataset has been acquired at Institute of Tropical Medicine, Caracas, Venezuela. Ten minutes of

0276-6547/01 $17.00 © 2001 IEEE

Computers in Cardiology 2001;28:501-504.

continuous ECG was recorded in each subject, using orthogonal XYZ leads, sampled at 1000 Hz and digitalized at 16 bits. The HRECG was obtained with averaging terminated at a noise-endpoint of 0.3 μV RMS. Signal averaging was performed with a Predictor system (Corazonix Corp., Oklahoma City).

The HRECG signals were recorded in 80 subjects grouped as: 11 healthy subjects and 33 seropositive patients without evidence of cardiac damage (group I), 25 chagasic patients with evidence of cardiac involvement including BBB (group II) and 11 patients in similar conditions of the previous group but with documented episodes of ventricular tachycardia (group III).

2.2. The model and best basis selection for DWT

We aim to extend our previous work using a wavelet based Bayesian modeling [3], presenting briefly the model employed.

Firstly, we have M signals corresponding to high resolution signal averaged electrocardiograms of patients, originally classified into p different groups; each group formed by m_i patients with $i = 1,\Lambda\ ,n$. Let $y_{il}(t)$ be the averaged signal at time t for patient l in group i. Then, the following model is assumed

$$y_{il}(t) = f(t) + \alpha_i(t) + \varepsilon_{il}(t) \quad , \quad \varepsilon_{il}(t) \sim N(0,\sigma^2)$$

Next, we applied a discrete wavelet transformation (DWT), to obtain $d_{il} = Wy_{il}$, but previously, we follow [4] to select the best bases in relation to a cost functional C that maps bases. This is, given the vector d_{il}, the cost of the basis B is $C(Bd_{il})$, where $d_{il} = By_{il}$ is the representation of y_{il} in the basis B. Then, we chose the entropy cost function like a cost measure. The entropy cost function of a vector d_{il} is

$$\xi(d_{il}) = -\sum_{i=1}^{n} p_i \log(p_i) \qquad (2.2.1)$$

where $p_i = \left|d_{il}(j,k)\right|^2 \Big/ \left\|d_{il}\right\|^2$. Since, the functional $\xi(d_{il})$ is not additive, we used the related functional

$$C(d_{il}) = -\sum_{i=1}^{n} \left|d_{il}\right|^2 \log\left|d_{li}\right|^2 \qquad (2.2.2)$$

Minimizing (2.2.1) is equivalent to minimizing (2.2.2).

After, applying a DWT with the basis selection,

$$d_{il} = Wy_{il} = \theta + \tau_i + \varepsilon_{il}^* \,, \quad \varepsilon_{il}^* \sim N(0,\sigma^2)$$

where

$$\theta(j,k) \sim (1-\pi_j^{\theta})I_0(\theta(j,k)) + \pi_j^{\theta} N\left(\theta(j,k)\middle| 0, u/2^j\right)$$

$$\tau_i(j,k) \sim (1-\pi_j^{\tau_i})I_0(\tau_i(j,k)) + \pi_j^{\tau_i} N\left(\tau_i(j,k)\middle| 0, v_i/2^j\right)$$

with $\pi_j^{\theta} = \alpha^j$ the prior probability of $\theta(j,k) \neq 0$, $\pi_j^{\tau_i} = \beta_i^j$ and $I_0(.)$ denotes the indicator function.

Finally, a Markov Chain Monte Carlo (MCMC) simulation scheme is used to implement posterior inference in the proposed model.

2.3. Classification techniques

To make Bayesian inference about unknown observable, we work in the follow form: after the data have been observed (the wavelet discrete transform of HRECG signals classified in p groups), we calculated the posterior probability that a "new" signal belongs to one of p groups and then we assigned the new signal to the group for which the posterior probability is greater. It is, Let $D = \{d_{il}, i = 1,\Lambda\ ,p; l = 1,\Lambda\ ,n_i\}$ be a set of M HRECG signals represented in the wavelet domain and $\tilde{D} = \{\tilde{d}_a, a = 1,\Lambda\ ,q\}$ a set of q new signal also represented in the wavelet domain. The probability that $\tilde{d}_a$ belongs to group j is given by

$$P(M_j \middle| \tilde{d}_a, data) = \frac{q_j\, p(\tilde{d}_a \middle| data, M_j)}{\sum_{i=1}^{p} q_i\, p(\tilde{d}_a \middle| data, M_i)} \qquad (2.3.1)$$

where:

- M_i is defined by the parameters $\mathrm{A}_i = \{\theta, \tau_i, \sigma^2\}$ for $i = 1,\mathrm{K}\ ,p$.
- q_i is the prior probability that $\tilde{d}_a$ belongs to group i.
- $p(\tilde{d}_a \middle| datos, M_i)$ denote the predictive density of $\tilde{d}_a$ given model M_i and the available data.

$$p(\tilde{d}_a \middle| data, M_i) = \int p(\tilde{d}_a \middle| data, \mathrm{A}_i, M_i)\, p(\mathrm{A}_i \middle| data)\, d\mathrm{A}_i$$

Note that if $q_i = 1/p$ for all i, equation (2.3.1) can be write as

$$p(M_j|\tilde{d}_a, data) = \left(\sum_{i=1}^{p} \frac{p(\tilde{d}_a|data, M_i)}{p(\tilde{d}_a|data, M_j)} \right)^{-1} = \left(\sum_{i=1}^{p} B_{ij}^{*}(\tilde{d}_a) \right)^{-1}$$

with $B_{ij}^{*}(\tilde{d}_a)$ the predictive Bayes factor for M_i versus M_j. A log Bayes factor of 1 (-1) indicates evidence in favor of model i (j), while the value 0 indicates no evidence for any model.

3. Results

A window of 2^9 ms beginning 50 ms before QRS onset was chosen to analyze each HRECG beat. This window includes the whole QRS complex and part of the ST segment. The DWT was applied using the MATLAB library for wavelet analysis *Wavelab v.701* [5].

Then, we used the procedure described in section 2.2 corresponding to the selection of the best basis with the entropy cost function, resulting that among the Daubechies family, the best basis is that one with 5 vanishing moments. Among the Coiflet family, that one with smaller entropy has 8 vanishing moments. In addition, others wavelet basis were considered, leading to similar results in terms of posterior inferences and classification The results presented correspond to transformation using a Daubechies 5 basis and they are part of a preliminary analysis described in [6].

Table 1. Distribution of HRECG signals of groups I and II for cross-validation.

Number of formed groups	Size of each group	Signals of group I	Signals of group II
9	7 signals	4	3
1	6 signals	3	3

We evaluate the predictive performance of our model to discriminate (i) between subjects in groups I or II and (ii) between subjects in groups I or III. For this, we split the 69 signals of groups I and II into 10 groups for cross-validation purposes, constructed as shown in table 1.

Of the 10 formed groups, 9 are used to train the model and the rest is used to classify. This process is repeated 10 times, leaving one HRECG signal from each group for classification.

Similarly, for discrimination of signals from groups I and III, we split the 55 signals of the following form:

Table 2. Distribution of HRECG signals of groups I and III for cross-validation.

Number of formed groups	Size of each group	Signals of group I	Signals of group III
8	5 signals	4	1
1	7 signals	6	1
1	8 signals	6	2

The classification criterion used was the assignment of each particular signal to the group with maximum posterior probability, as defined in equation (2.3.1). Then, the sensitive and specificity (percentage of abnormal and normal signals, respectively, identified by the classification procedure) was calculated. Also, we computed the accuracy (percentage of subjects correctly identified). These results were compared with those obtained in [7] and [8], where the classification was performed via linear discriminant functions. Table 3 shows a comparison between the specificity, sensitivity and accuracy rates obtained using the wavelet based Bayesian models and the rates obtained using late potential indexes based on short-time Fourier representations (LPST) [7][8] .

Table 3. Specificity, sensitivity and accuracy rates.

	Specificity (%)	Sensitivity (%)	Accuracy (%)
Wavelet based Ba yesian m odels (I, II)	81.82	60.00	73.91
LPST indexes (I, II)	24.24	65.38	42.37
Wavelet based Ba yesian m odels (I, III)	90.91	90.91	90.91
LPST indexes (I, II)	45.45	15.15	22.72

4. Discussion and conclusions

Wavelet based Bayesian methods provide considerably higher specificity, sensibility and accuracy rates when discriminating HRECG signals of groups I and III of the

studied chagasic database. LPST indexes show a higher sensitivity rate in the discrimination of groups I and II than wavelet based Bayesian methods (65.30% vs 60%), however, the specificity and accuracy rates for the wavelet based Bayesian methods are considerably higher.

Our Bayesian model based classification approach shows, in general, a much better predictive performance than classification schemes based on short-time frequency late potential indexes. In addition, the proposed model based method provides a probabilistic assessment of the uncertainties involved in classifying the signals.

In order to improve the predictive performance, further developments should include extensions to multivariate models that take in consideration the information provided by the XYZ leads simultaneously. Also, different prior structures should be considered.

Acknowledgements

This work was supported in part by Dirección de Investigación, Universidad Simón Bolívar, Sede del Litoral,Venezuela (S1-NCBA-11-99 and S1-NCBA-3-99).

References

[1] Eliazari MV. Chagasic myocardiopathy historical perspective. Medicina 1999; 59-2:25-40.

[2] Gomis P, Suárez K , Passariello G, Mendoza I, Caminal P, Lander P. Abnormal intra-QRS signals and late potentials in the high resolution ECG associated with Chagasic myocarditis. In Computers in Cardiology 1996, volume 23. Piscataway: IEEE Computer Society Press, 1996; 633-6.

[3] García I. Modelaje Bayesiano de Ondículas aplicadas a señales cardíacas en pacientes con Mal de Chagas. Master's thesis, Universidad Simón Bolívar, 1999.

[4] Vidakovic B. Statistical Modelling by Wavelets. New York, USA: Wiley Series in Probability and Statistics, 1999.

[5] WaveLab Reference Manual, 1999. URL http: // www-stat.stanford.edu/ wavelab.

[6] García I, Prado R and Gomis P. Wavelet based Bayesian methods for characterizing high-resolution ECG signals. In Computes in Cardiology 2000, volume 27: 403-406.

[7] Suárez K. Evaluation of Chagas disease through HRECG time-frequency representations. Master's thesis, Universidad Simón Bolívar, 1997.

[8] Suárez K, Gomis P, Silva J. Analysis of late potentials associated with chagasic myocarditis by means of time-frequency representations. In Computes in Cardiology 1998, volume 25: 397-400.

Address for correspondence.

Irene García.
CESMa
Universidad Simón Bolívar
APARTADO 89000, Caracas 1080-A
Venezuela.
irene@cesma.usb.ve
fax:+58212-9063233/3234