Use of Approximation Entropy for Stratification of Risk in Patients With Chagas Disease

M Vizcardo†, A Ravelo‡
†Departamento de Física, Universidad Nacional San Agustín, Arequipa, Perú, mvizcardoc@uns.edu.pe
‡Instituto for Technological Develoment and Innovation in Communications, Universidad Las Palmas de Gran Canarias, Spain

Abstract

According to the World Health Organization, the number of people infected with Trypanosoma Cruzi is estimated between 6 and 7 million, the causative agent of Chagas disease, and in 560000 people exposed to the risk of affectation.

The approximate entropy was used to quantify the regularity of the tachograms of patients with Chagas disease. The study population consisted of three groups of volunteers: 92 controls (C), 102 patients with positive serology without cardiac involvement diagnosed by conventional non-invasive methods (CH1) and 107 patients with positive serology and mild to moderate incipient heart failure (CH2). Clinical evaluation, Machado-Gerreiro serology examination, chest x-rays, echocardiogram, electrocardiogram and ambulatory record Holter (24 hours), were made to the three groups. We analyzed RR segments of 5 minutes, 288 segments, corresponding to 24 hours per patient. We found significant differences between the Control and CH2 groups, which is used to stratify risk in the CH1 group.

1. Introduction

Chagas disease is caused by a flagellated parasite: Trypanosoma cruzi, transmitted by an insect of the genus Triatoma and also by blood transfusions. In Latin America the number of people infected is approximately 6 million, with a population exposed to the risk of infection of 568,000[1]. In 40 % of the population infected with Trypanosoma cruzi (T. cruzi) there is cardiac involvement [2, 3, 11]. These estimates explain why this disease is a serious public health problem in the countries where it is endemic. In the evolution of Chagas disease we can distinguish an initial acute phase of infection and a prolonged intermediate chronic phase, in which the disease is often clinically silent, and the usual diagnostic techniques do not provide a robust criterion to predict whether a seropositive asymptomatic patient will suffer cardiac involvement. It is our interest to develop a non-invasive methodology low cost, that allows to see the dysautononia or dysfunction in the course of the 24 hours and with this it could be used to detect early any cardiac alterations produced by the T. cruzi.

The approximate entropy (ApEn) was developed by Steve Pincus[4][5] is based on the entropy of Kolmogorov-Sinai K-S a formula for entropy K-S, proposed by Grassberger and Procaccia[6] and modified by Takens[7], a modification of ApEn was made by Richman[8] sample entropy (Sampen), both which measures the irregularity and complexity of a time series of data.

2. Database

The results of this work were obtained by processing the electrocardiogram (ECG) and obtaining the RR interval. The following test is carried out: clinical evaluation, serological Machado-Gerreiro test, chest x-rays, echocardiogram, electrocardiogram and ambulatory Holter registration (24 hours). The volunteers are classified into three groups: 92 healthy people called group control C; 102 patients infected with only positive serology (clinical evaluation, chest x-rays, echocardiogram, electrocardiogram and Holter were normal) called CH1 group; and 107 seropositive patients with incipient cardiac involvement first-degree atrioventricular block (BAV), sinus bradycardia (BS) or right bundle branch block of the bundle of His (BRDHH), that were not being treated with medications, called the CH2 group. All were outpatients, and informed consent was obtained from all of them. The ECG signals were recorded at 500 Hz with 12 bits of resolution, a set of 288 frames of 5 minutes was obtained. We have used the database of the Institute of Tropical Meditation (IMT) of the Central University of Venezuela.
3. Methods

3.1. Preprocessed

For the detection of the QRS complex the program based on the Pan-Tompik algorithm[9] was used, then the 288 tachograms of the RRs of 5 minutes were generated, they were made a post processed the "ada" filter[10], to eliminate the eptopic beat, and then to find the HRV indices.

3.2. Feature extraction

ApEn and SampEn are based on the comparisons of the component-to-component embedding vectors \(m = 2\) and with a threshold of 20% of the standard deviation of the intervals \(RR\) \((r = 0.20)\). The difference between the entropies is that ApEn does not take into account the comparisons of the embedment vectors with itself and the way to calculate the logarithm. To calculate ApEn:

\[
ApEn(m, r, N) = - \frac{1}{N - m} \sum_{i=1}^{N-m} \log \left( \frac{A_i}{B_i} \right)
\]

(1)

where \(m\) is the embedment dimension, \(r\) is the threshold that is typically 20% of the value of the standard deviation and \(N\) is the number of data.

SampEn is calculated as:

\[
SampEn(m, r, N) = - \log \left( \frac{\sum_{i=1}^{N-m} A_i}{\sum_{i=1}^{N-m} B_i} \right) = - \log \left( \frac{A}{B} \right)
\]

(2)

similar a ApEn \(m, r\) and \(N\).

3.3. Classifier

We will use the kruskalwallis test that matches the 288 frames corresponding to the Control-CH1, Control-CH2 and CH1-CH2 groups to find the data for 5 minutes if \(p \leq 0.05\) are no similar. Also use the logistic classifier to the circadian profiles of the average values of the \(RR\) intervals and finally the analysis of boxplot to analyze a frame in special

4. Results

4.1. Approximate Entropy (ApEn)

ApEn was calculated at 288 tachograms of 5 minutes, 24-hour circadian profiles were obtained (Fig 1a).

- Between 00:00 hours and 06:00 the control group C has slightly higher values than CH1, CH1 is slightly larger than CH2.
- Between 06:30 and 11:20 there is no greater difference between the three groups.

- We noticed that between 11:20 and 21:30 there is a sustained difference of group C
- At 14:30 it is the biggest difference between groups C and CH2.

![Approximate Entropy ApEn](image)

Figure 1. a) ApEn values of the 24 hours (288 frames) b) Probability value of the groups: Control-CH1, Control-CH2, CH1-CH2

Now we will use the logistic classifier to groups C and CH2, it is possible to classify up to 90% of specificity of group C and 80% of sensitivity in group CH2 between 14:15 h. to 15:40 h.

With these results we apply the Kruskal-Wallis test where

![Kruskal-Wallis test](image)

Figure 2.

we can see in figure 1b that around frame 175 (14:30 hours) the values of the probabilities are less than \(p < 0.05\), for the groups: Control-CH1 and Control-CH2 but not for CH1-CH2 as shown in the figure, can also be seen in table

![Logistic classifier average ApEn in groups of 5 min.](image)
4.2. Sample Entropy (SampEn)

We apply the SampEn algorithms to the 288 segments of 5 minutes to the three groups and we obtain the average values of the \( RR \) intervals, this can be seen in Fig 4a.

- Between 00:00 hours and 08:30 there is no greater difference between the three groups. The control group C has slightly higher values than CH1, CH1 is slightly larger than CH2.
- Between 08:30 and 10:00 the groups CH1 and CH2 have a higher value than group C.
- However, from 12:30 to 16:15 we noticed that there is more activity in group C than groups CH1 and CH2.
- At 14:10 it is the biggest difference between groups C and CH2.

Now we will use the logistic classifier to groups C and CH2, it is possible to classify up to 80% of specificity of group C and 70% of sensitivity in group CH2 between 14:30 h. to 15:50 h.

We applied the Kruskal-Wallis test to the values of SampEn to the groups: Control-CH1, Control-CH2, where we noticed that around frame 175 the values of probabilities \( p < 0.05 \) where the groups are different from each other, this result is the same as ApEn.

We analyze the segment 160 of the result of the SampEn values that corresponds to the 13:30 hours, using the box-plot we noticed that it is possible to show the risk to the CH1 group, just like ApEn.

5. Discussion and conclusions

The average ApEn values between 11:20 and 21:30 show (Fig. 1a) that group Control has higher values than the group CH1 and CH2, this could be due to a decrease or enervation in its barometric response, this is also observed in SampEn (Fig 4a) at shorter time intervals from 12:30 to 16:15.

As the ApEn and SampEn measure the irregularity and
complexity, when applying to the groups Control, CH1 and CH2 we note that it is measuring variability of the HRV, that is, the CH1 and CH2 groups are enervated or diminished due to having less variability, in the afternoon hours. The logistic classifier and Kruskal-Wallis test allows us to differentiate both groups (control and CH2) both in ApEn and SampEn in the afternoon hours, therefore it is possible to use the classifier to can be used in the stratify group CH1.

Acknowledgments
Vicentadoro de Investigación de la Universidad Nacional San Agustín de Arequipa
Instituto for Technological Develoment and Innovation in Communications, Universidad Las Palmas de Gran Canarias Conflicts of Interest
The authors declare no conflict of interest

References

[27] Catherine Masiel Merejo Pea and et.al., Dysautonomy in different death risk groups (Rassi score) in patients with Chagas heart disease, Pacing Clin Electrophysiol. (2018);41:238245.