

Application of Permutation Entropy in the Stratification of Patients With Chagas Disease

D Cornejo[†], M Rodriguez[†], L Diaz[†], E Alvarez[‡], M Vizcardo[†]

[†]Escuela Profesional de Física, Universidad Nacional de San Agustín de Arequipa

[‡]Escuela de Física, Universidad Central de Venezuela, Venezuela

Abstract

Chagas disease is an endemic disease that in recent decades has ceased to be a rural disease to become mainly an urban disease, becoming a public health problem because it is a life-threatening disease and because 70 million people are at risk of infection. This disease, which has cardiac involvement in its chronic phase, can often be silent and asymptomatic, because of this the establishment of early markers in this type of patients is of great interest. To achieve this goal this study proposes the use of the permutation entropy (PE) which has been shown satisfactory results in the analysis of hearth rhythm and the differentiation between healthy people and people who suffered of congestive heart failure, also known as heart failure.

This study analyzes three groups: 83 volunteers (Control), 102 patients with positive serology and no cardiac involvement diagnosed by conventional non-invasive methods (CH1) and 107 patients with positive serology and mild to moderate incipient heart failure (CH2). The used data comes from the 24-hour ECG record, the RR intervals from each of the three groups are shown in 288 frames of 5 minutes.

The analysis performed shows significant differences between the three groups, with 90% to 100% specificity and sensitivity, making this study a promising precedent for the development of a low-cost, non-invasive method for detecting possible heart failure in asymptomatic patients.

1. Introduction

Chagas disease is an endemic disease potentially life-threatening, which is caused by a flagellated parasite: *Trypanosoma cruzi* (T. cruzi). According to the WHO this disease is found mainly in Latin America countries with an estimated 6 to 7 million of people infected[1]; however in the last decades this disease which was mainly rural has changed to be mostly an urban disease having as a consequence an increased number of cases in North America as well as in European and African countries. Chagas disease has two phases: an initial acute phase of infection and a prolonged intermediate chronic phase, in the last one the

disease is often clinically silent and asymptomatic, therefore a large number of people with the disease are not diagnosed. Because of this fact, an estimate of 70 million people are at risk of infection[1] due to the reception of blood or organs from sick patients who were undiagnosed or untreated. Studies shows that 40% of the population with the disease also has a cardiac compromise[2–5] which most common manifestation is congestive heart failure (CHF). Previous studies have analyzed the variability in the ECG signals of patients with CHF to get a way to obtain a premature diagnosis, most of these studies had great results[6, 8–11]. One way to analyze the variability of a system is by using different entropies; however, permutation entropy proposed by Brandt and Pompe[12] is particularly appropriate for the study of the complexity of the systems, this entropy was applied in different fields of study including the biomedical area[13–16] where it had good results distinguishing different heart rate variability[17]. So taking into account the heart failure like a characteristic symptom of the last and most dangerous phase of the disease and the successfully application of the PE to analyze complex systems like ECG's, is of interest to use these results to develop a low-cost non-invasive methodology that allows to distinguish dysautonomia or dysfunction in the course of 24 hours, getting an early marker of the cardiac alterations produced by T. cruzi

2. Database

2.1. Preprocessed

This study used the 24-hour ECG holter register, which was divided into 288 segments (frames) corresponding to 5 minutes of ECG (approximately 300 beats) for each volunteer or patient. The data was post processed with the ada filter[7], to eliminate the eptopic beat, and then to find the HRV indices.

2.2. Feature extraction

The permutation entropy (PE) defined by Brand and Pompe[12] was the used tool to process the data, and it

is defined as:

$$H(n) = - \sum p(\pi) \log p(\pi) \quad (1)$$

It is a way to measure the complexity in data based in the comparison between one value of a time series with its neighbors; in other words, it encodes patterns (each one denoted by π) within the data and count the probability $p(\pi)$ of each pattern to appear to calculate the entropy. For the actual calculation, the base 2 is used for the logarithm.

2.3. Kruskal Wallis test

The Kruskal Wallis test is a non-parametric test, so it doesn't assume the data came from a normal distribution that can be completely described with the mean and the standard deviation. It can be used to compare two or more samples with different sample sizes.

This test was used to find significant differences in the samples using the p-value which is a probability that measure the evidence against the null hypothesis (both samples came from the same population). Lower probabilities have stronger evidence against this null hypothesis.

In this case the critical p-value is 0.05, therefore if $p\text{-value} < 0.05$ it indicates greater differences between the groups: Control-CH1, Control-CH2 and CH1-CH2; it means that it would be proven that the samples come from different populations.

2.4. Logistic regression

The logistic regression is a statistical model that uses the logistic function to predict a probabilistic relationship of the variables, admitting the case in which the dependent variable is polytomous, comprising a purely quantitative procedure; it is especially useful when the dependent variable is dichotomous, being bounded in the interval (0,1).

It was used as a classifier of the values obtained by the PE in each frame (with the purpose of finding intervals of time where the groups can be differentiated) to certain groups, because its level of prediction and interpretation of the relationship of the variables according to probabilities is very high.

2.5. ROC curve

The receiver operating characteristic curve (ROC curve) is a plot that shows the efficacy of a binary classifier to make the right classification in terms of the true positive rate and the false positive rate. The graphic shows the sensitivity TPR in the y axis and the false positive rate FPR (also referred as 1-TNR where TNR is the true negative rate or specificity) in the x axis. In that way the points that end up close to the (0,1) coordinate are the values in which the binary classifier had the most success.

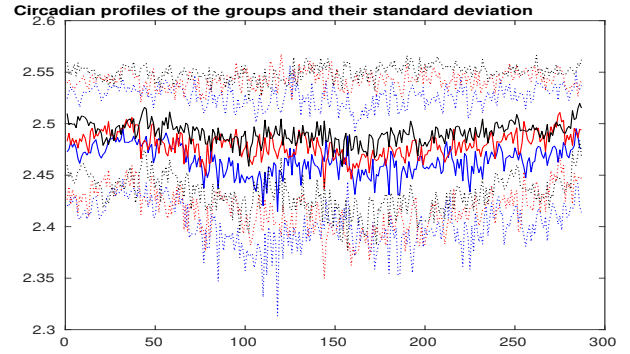


Figure 1. Circadian profiles of 24-hour PE mean values and standard deviation (288 frames), Control group (blue), CH1 group (red) and CH2 group (black)

3. Results

The figure 1 shows:

- The average values obtained for PE of the Control group in the circadian profiles (blue) are always below the average values of the CH2 group (red) and has the greatest differences in all the frames with the exceptions in the frames 0 till 50, and near the frames 150 and 200, which corresponds to the midnight till 4:00 hours, and 12:00 to 16:00 hours, respectively
- All day exists a similar behavior between the average values of the CH2 group PE (red) and the average values of the CH1 group (black) with two exceptions: In the frames 0 to 70 which corresponds from midnight to 5:00 hours approximately and from frames 80 to around 110 corresponding to 6:00 hours until approaching to 9:00 hours
- In the frames 45 to 75 which corresponds from near the 4:00 hours until 6:00 hours, it is possible to notice that the circadian profile shows a similar behavior between the average values of the CH1 group (black) and the Control group (blue)
- The CH1 group (black) has higher entropy levels than the control group but lower than the CH2 group (red), being in the middle of these in the frames 210 to 270, that is since 17:00 hours to 21:00 hours approximately

The Kruskal Wallis test is represented in figure 2, it shows that:

- It is possible to find significant differences between the Control and CH1 groups in different frames, but the most relevant differences are in the frames 80 to 140 since around 6:00 hours to 11:00 hours approximately and 160 to 190, which corresponds from nearly 13:00 hours until 15 hours.
- In the case of the Control and CH2 groups, many differences are evident throughout the 24-hour trial, being the frames 70 to 140, and 220 to 288 regions of great interest. These regions corresponds since about the 5:00 hours un-

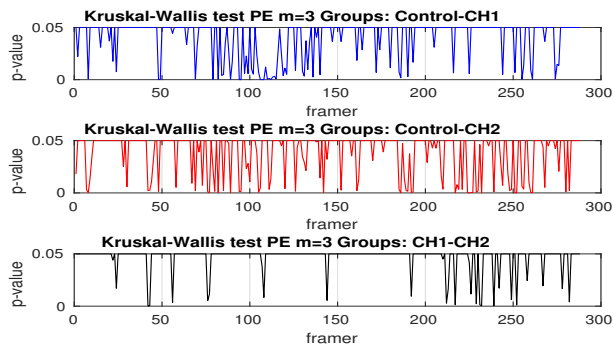


Figure 2. Kruskal-Wallis test of PE of the 288 framer

til 11:00 hours approximately and around the 18:00 hours until the end of the day, respectively.

• Finally, between the CH1 and CH2 groups the Kruskal-Wallis test evidences a similar behavior since the beginning of the day until the frame 200 corresponding to 16:00 hours; however, its possible to find significant differences starting in the frame 220, near the 18:00 hours, being the frames 220 to 250 where the greatest differences are found, corresponding since about the 18:00 hours to 20:00 hours.

To discriminate between the Control and CH2 groups of PE is convenient to use the logistic regression, for this objective 50% of the frames (144) were chosen randomly to train the program and the other 50% for validation. This process is represented in Figure 3. Because the data is cho-

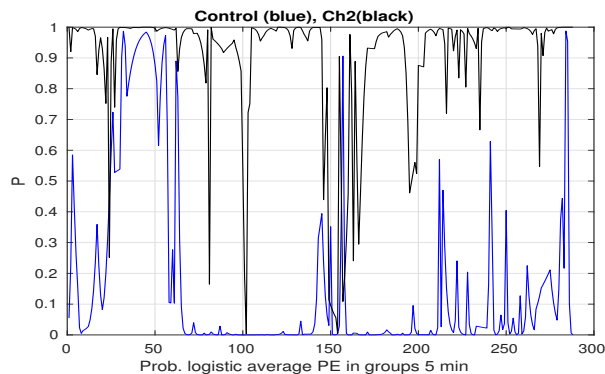


Figure 3. Logistic classifier of the average values of PE in the 288 framer

sen randomly, the graphs obtained may vary; however, it is generally possible to observe a region of interest between the frames 70 to 140 corresponding to about 5:00 hours until 11:00 hours approximately. It's possible to verify the region of interest shown by the logistic classifier agree with the analysis carried out by the Kruskal Wallis test and the observation of the circadian profile from the means of the permutation entropy values of each group.

Using the data generated by the logistic classifier, a ROC curve is performed to determine the sensitivity and specificity of the obtained classifier. The ROC space of the region of interest between the frames 105-142 is represented in Figure 4, it shows that: Approximately 20% of the total frames are classified in the region that comprises from 90% to 100% of specificity and sensitivity. This result shows the great precision of the method applied.

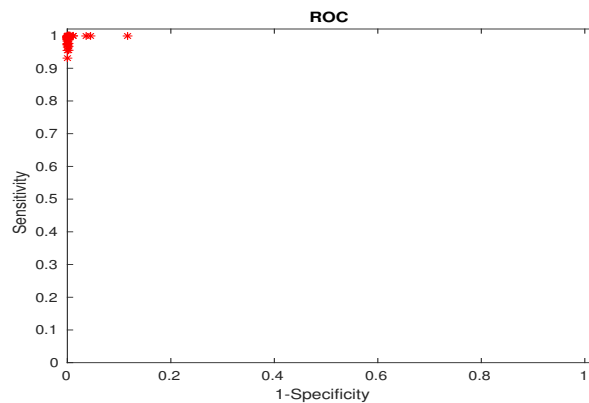


Figure 4. ROC curve for the frames 105 - 142

4. Discussion and conclusion

- The PE is a good way to measure the heart rate variability, when applied to the patients of each group, it could be determined that there is a notorious difference between them, being the greatest difference between the control and CH2 groups, generating areas of interest in the frames 70 to 140 and 220 to 288, corresponding to around the 6:00 hours until 11:00 hours and since about 18:00 hours until the end of the day. It was evident that the patients belonging to the control group show a lower value of PE than the patients belonging to the CH2 group, taking into account that a characteristic symptom of the chronic phase of Chagas disease is the CHF, these results are in agreement with previous studies made about patients that suffer CHF.
- Likewise, it was observed that the patients in the CH1 group show a wide divergence with the control patients between midnight and 4 pm approximately (frames 0 to 250), and in the first five hours of the day (frames 0 to 60) there is no significant difference between the three groups.
- Thanks to the marked deviation between the control and CH2 groups and the different regions of interest from the logistic regression, which sensibility and sensitivity were validated with the use of the ROC curve, the present study can be taken like a precedent for the development of a method for early detection of this disease made for asymptomatic patients (CH1).

Acknowledgments

Universidad Nacional de San Agustín de Arequipa

References

- [1] Panamerican Health Organization. Control, interruption of transmission and elimination of Chagas disease as a public health problem. Guide of evaluation, verification and validation. Washington, D.C.: OPS / WHO; 2019.
- [2] Moleiro, F., Rodriguez, A., Misticchio, F., Ruesta, V., Octavio, J., lvarez, E., Fernndez, A., Jimnez, J., Marcano, A., Vizcardo, M. (2001) Utilidad de la aplicacin de tcnicas de modelado no lineal en el anlisis de electrocardiogramas de pacientes con infeccin chagsica. *Rev Esp Cardiol.* 54: 1081-1090
- [3] Moleiro F. & Mendoza I, Miocardiopatía Chagsica Crnica. Estudio Epidemiológico, *Acta Cient. Venez.* 1980, 31:66
- [4] Hagar, J & Rahimtoola, S., *N. Eng. J. Med.*, 325, 763, 1991
- [5] Di Lorenzo Oliveira et al. (2020) Risk Score for Predicting 2-Year Mortality in Patients with Chagas Cardiomyopathy from Endemic Areas: SaMi-Trop Cohort Study. *Journal of the American Heart Association* (Vol.9, Issue 6).
- [6] Li, Yang et al. (2018). Short-term QT interval variability in patients with coronary artery disease and congestive heart failure: a comparison with healthy control subjects. *Medical & Biological Engineering & Computing.* 57. 1-12. 10.1007/s11517-018-1870-8.
- [7] N Wessel, A Voss, *Renormalised Entropy: A New Method of Non-Linear Dynamics for the Analysis of Heart Rate Variability*, *Computers in Cardiology* 1994
- [8] Chandrakar Kamath. (2014). Entropy Measures of Irregularity and Complexity dor Surface Electrocardiogram Time Series in Patients with Congestive Heart Failure. *Journal of Advances in Computer Research* (Vol.6, Issue 4).
- [9] Vicnesh, Jahmunah & Oh, Shu Lih & Koh, Joel En Wei & Ciaccio, Edward & Chua, Kuang & Tan, Ru San & Acharya, U Rajendra. (2019). Computer-aided diagnosis of congestive heart failure using ECG signals A review. *Physica Medica.* 62. 95-104. 10.1016/j.ejmp.2019.05.004.
- [10] Borowska, Marta. (2015). Entropy-Based Algorithms in the Analysis of Biomedical Signals. *Studies in Logic, Grammar and Rhetoric.* 43. 10.1515/slgr-2015-0039.
- [11] T. Liu, W. Yao, M. Wu, Z. Shi, J. Wang, X. Ning, Multiscale permutation entropy analysis of electrocardiogram, *Physica A* (2016), <http://dx.doi.org/10.1016/j.physa.2016.11.102>
- [12] Bandt, Christoph & Pompe, Bernd. (2002). Permutation Entropy: A Natural Complexity Measure for Time Series. *Physical review letters.* 88. 174102. 10.1103/PhysRevLett.88.174102.
- [13] Zanin, Massimiliano & Zunino, Luciano & Rosso, Osvaldo & Papo, David. (2012). Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review. *Entropy.* 14. 1553. 10.3390/e14081553.
- [14] Vizcardo, M. & Ravelo, A. , Use of Approximation Entropy for Stratification of Risk in Patients With Chagas Disease, *Computing in Cardiology*, 2018, doi: 10.22489/CinC.2018.234
- [15] Vizcardo, M. & Manrique, M. & Garcia, A. & Gomis, P., Application of the Entropy of Approximation for the Nonlinear Characterization in Patients with Chagas Disease, *Computing in Cardiology*, 2019,doi:10.23919/CinC49843.2019.9005876
- [16] Vizcardo M, et.al., Nonlinear characterization of ECGs in patients with Chagas' disease, 2019, *Biomedical Physics & Engineering Express*, doi = 10.1088/2057-1976/ab03f7, IOP Publishing, vol 5, number 2, pag 025042.
- [17] Bian, Chunhua & Qin, Chang & Ma, Qianli & Shen, Qinghong. (2012). Modified permutation-entropy analysis of heartbeat dynamics. *Physical review. E, Statistical, nonlinear, and soft matter physics.* 85. 021906. 10.1103/PhysRevE.85.021906.

Address for correspondence:

Miguel Vizcardo Cornejo
Escuela Profesional de Física, Av. Independencia sn, Arequipa,
Perú
mvizcardocunsa.edu.pe