XLI CILAMCE
XLI Ibero-Latin American Congress on Computational Methods in Engineering
16-19 November, 2020 FOZ DO IGUAÇU PARANÁ - BRAZIL

Classification of Intra Cardiac Atrial Fibrillation using High Order Statistics and Machine Learning.

Luis Fillype da Silva[1], Jonathan Araujo Queiroz[1], Allan Kardec Barros[1]

[1]*Dept. of Electricity Engineering Course, University Federal of Maranhão - UFMA*
Av. dos Portugueses, 1966 - Vila Bacanga, 65080-805, Maranhão, Brasil
silvaluis_@outlook.com, queirozjth@gmail.com, allan@ufma.br

Abstract. The electrocardiogram is an exam that quantifies the electrical activity of the heart, allowing the detection of the heart rhythm, the number of beats per minute and the diagnosis of various cardiac pathologies. This article aims to obtain a classification model based on the beats of two groups of individuals: healthy and unhealthy. The feature extraction methodology was used and adapted for the classification of atrial fibrillation. Classifications were performed in two-dimensional and three-dimensional space in the database, obtaining an accuracy of 97% to 100%.

Keywords: Feature extraction, Electrocardiogram, Classification of ECG signals.

1 Introduction

The importance of effective diagnosis of heart disease has driven the development of computational methods for the more accurate detection of these heart diseases. According to the World Health Organization [4], the major causes of death in the world today are strongly related to problems arising from the heart.

Atrial fibrillation (AF), the most common sustained cardiac arrhythmia in clinical practice, is an important and growing public health problem. It is estimated that about 2.2 million individuals are affected in the USA and almost 4.5 million in Europe. Its prevalence varies with age, affecting about 10% of patients over 80 years of age. Considering the age factor, studies project an increase of up to five times in the prevalence of AF in the year 2050 [4].

Thus, the analysis of the electrocardiogram (ECG) becomes of paramount importance for the previous detection of abnormalities, as it is one of the most used non-invasive techniques to aid the diagnosis of Atrial Fibrillation [1], as well as to evidence the propensity to development or the existence of other diseases of this kind [4].

Works such as that by Queiroz et. al. [1] analyze the variation in tension that occurs in a heartbeat *t* interval using kurtosis. In Lovelock's article, Wirtz, Hemzo (2018) [5] a method is proposed based on deep convolutional neural networks for the classification of heartbeat capable of accurately classifying five different arrhythmias. This article proposes to extract the entire heartbeat of an ECG and to cluster two groups using high-order statistics, subsequently performing the classification in two Machine Learning algorithms.

2 Materials and method

In Fig.1, the methodology followed for the preparation of this article is illustrated. First, the databases to be used were defined, separating them into two sets of data: individuals who have normal sinus rhythm and individuals who have atrial fibrillation. Then, the pre-processing of the database signals and data organization for the feature extraction step was performed. In this step, the values of variance, asymmetry and kurtosis of the data sets of each base are calculated. After that, combinations of characteristics were selected, which are represented by the aforementioned statistics, and placed as input for the classifiers used in this article. At the end of the process, the classification metric values are returned to evaluate the algorithm.

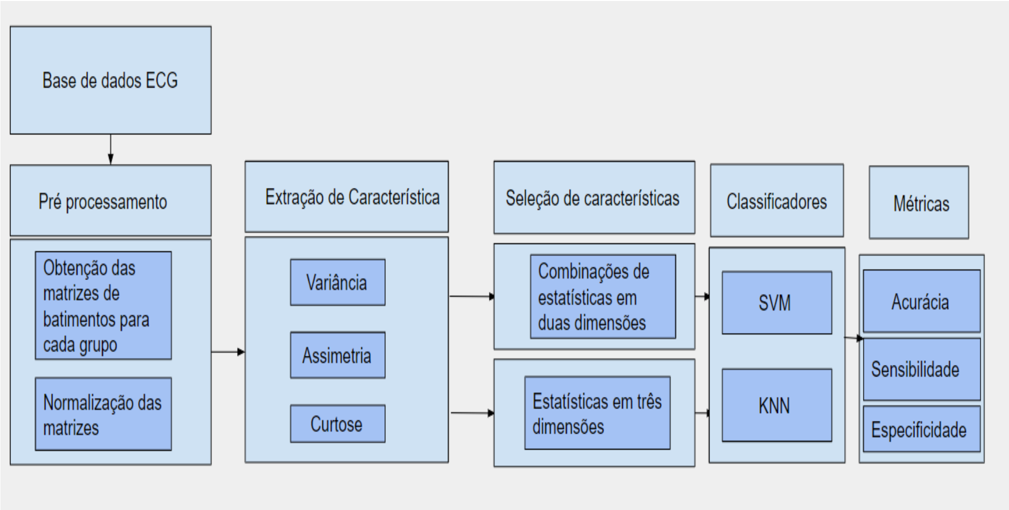

Figure 1. Proposed methodology.

1.2 Data base

The databases used were the MIT-BIH Normal Sinus Rhythm and the Intracardiac Atrial Fibrillation Database, both available from Goldberger et al [2] [3]. The database of patients with normal sinus rhythm contains 18 records, of which 13 were used for this analysis. The database of patients with intra-atrial fibrillation contains 8 patients, all of which are used.

From both bases, ECG signals characteristic of the DII lead, the most used in the world, were selected. The entire duration of the signal was selected, sampled at a frequency of 256 Hz to extract the beats.

1.1 Pre-processing

Each selected signal was segmented in order to obtain the respective beat, according to what is proposed by Queiroz et. al. [1]. The beats of each group were grouped, with a matrix A being generated by the concatenation of the beats of the healthy group, and a B matrix of beats of the group with intracardiac atrial fibrillation, as described in eq. (1):

$$Bn,m = [Bn,a \quad Bn,b \; ... \; Bn,z \quad]. \tag{1}$$

where *n* represents the number of beats, and m represents the total of all columns of all beats.

After that, the mean of its set was subtracted from the sign, dividing the result by Shannon's entropy, constituting Eq. (2) below:

$$Bn,m = \frac{Bn,m - \frac{1}{N}\sum_1^N \; Bn,m}{-\sum_1^n \; Pn,m(log_2\frac{1}{N})}. \tag{2}$$

where *P* represents the probability associated with each beat obtained.

1.1 Feature extraction

In this stage, he used the extraction methodology by means of high-order statistics, until the fourth, proposed in Queiroz et. al. [1]. Through the equations to be described below, a vector was obtained for each of the associated statistics: variance, kurtosis and skewness. eq. (3), (4) and (5) represent them, respectively.

$$\sigma_X^2 = E(X^2) - ((E(X))^2. \tag{3}$$

$$\lambda_X = E\big[\big((X - E(X)\big)\sigma^{-1}\big]^3. \tag{4}$$

$$\kappa_X = E\big[\big((X - E(X)\big)\sigma^{-1}\big]^4. \tag{5}$$

1.1 Evaluation metrics

Values of accuracy, sensitivity and specificity, described by eq. (6), (7) and (8) below, were used as metrics for the classifiers.

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN}\ x\ 100. \tag{6}$$

$$Sensitivity = \frac{VP}{VP + FN}\ x\ 100. \tag{7}$$

$$Specificity = \frac{VN}{VN + FP}\ x\ 100. \tag{8}$$

In the equations, *VP* corresponds to the number of true positives, *VN* the true negatives, *FP* for false positive records and *FN* for false negative classifications.

3 Simulation and Results

This work analyzed the beats extracted from ECGS of several healthy patients with atrial fibrillation, in order to classify them. For the classification stage, matrices were generated, where each column is represented by variance, asymmetry and kurtosis, respectively. Such matrices were the inputs of the classifiers K-nearest Neighbors (KNN) and Support Vector Machine (SVM) in order to verify which statistical moments have greater accuracy, sensitivity and specificity. In addition, a set of high-order statistics was presented for each signal in order to obtain which combination provides more efficiency in the classification process, where σ_X^2 represents the variance of the individuals beats in the database, κ_X the kurtosis of the base and λ_X the skewness of data.

Below, in Fig. 2 and Fig. 3, the beats of a signal from each database used are represented. In Fig.1, the beats of an individual with healthy beats are shown, and in Fig. 2, the beats of an individual belonging to the group of intracardiac atrial fibrillation.

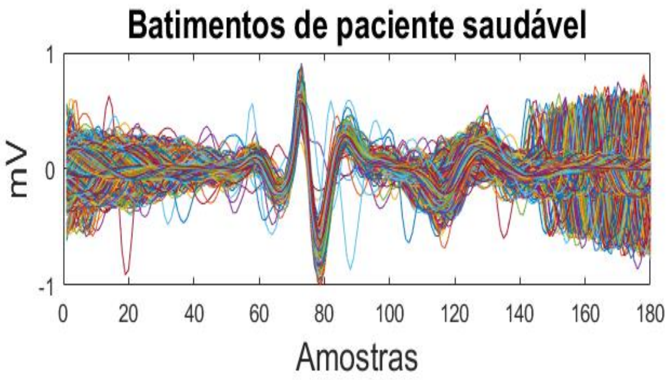

Figure 1. Beats extracted from a healthy patient.

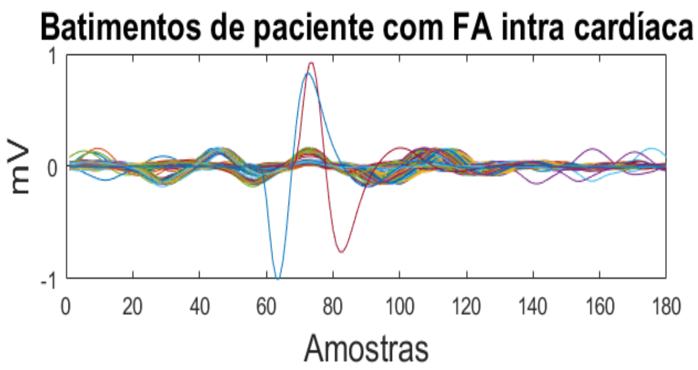

Figure 2. Beats extracted from a patient with Intra cardiac Atrial Fibrillation.

In Tab. 1 shows the results obtained from the accuracy of the classifications of each group, with each two-dimensional and three-dimensional combination. Tab. 2 shows results related to the sensitivity of each classifier. Finally, in Tab. 3, the values obtained for specificity are contained.

Table 1. Accuracy of classifiers

Statistics	KNN	SVM	Metric
σ_X^2+κ_X	100%	100%	Accuracy
σ_X^2+λ_X	87%	57%	Accuracy
$\kappa_X + \lambda_X$	37%	29%	Accuracy
σ_X^2+$\lambda_{X\,+}\kappa_X$	100%	97%	Accuracy

Table 2. Sensitivity of classifiers

Statistics	KNN	SVM	Metric
σ_X^2+κ_X	100%	100%	Sensitivity
σ_X^2+λ_X	92%	90%	Sensitivity
$\kappa_X + \lambda_X$	50%	47%	Sensitivity

$\sigma_X^2 + \lambda_X + \kappa_X$	100%	96.5%	Sensitivity

Table 3. Specificity of classifiers

Statistics	KNN	SVM	Metric
$\sigma_X^2 + \kappa_X$	87%	90%	Specificity
$\sigma_X^2 + \lambda_X$	83%	81%	Specificity
$\kappa_X + \lambda_X$	42%	37%	Specificity
$\sigma_X^2 + \lambda_X + \kappa_X$	85%	87%	Specificity

4 Discussion

This work presented analysis of the characteristics in two and three dimensions. In Fig. 4, the analysis in two dimensions shows the separation of the data using variance and kurtosis, as shown in Tab. 1, 2 and 3, and also reaffirming the results discussed in Queiroz et. al. [1] and in Lucena et. al. [7], who point out that kurtosis may be an appropriate approach to measure sparse signals, such as ECG. In Fig. 5, the combination of kurtosis and skewness did not achieve a very clear separation between groups. In Fig. 6, using the variance and asymmetry of the data, a better separation can be obtained.

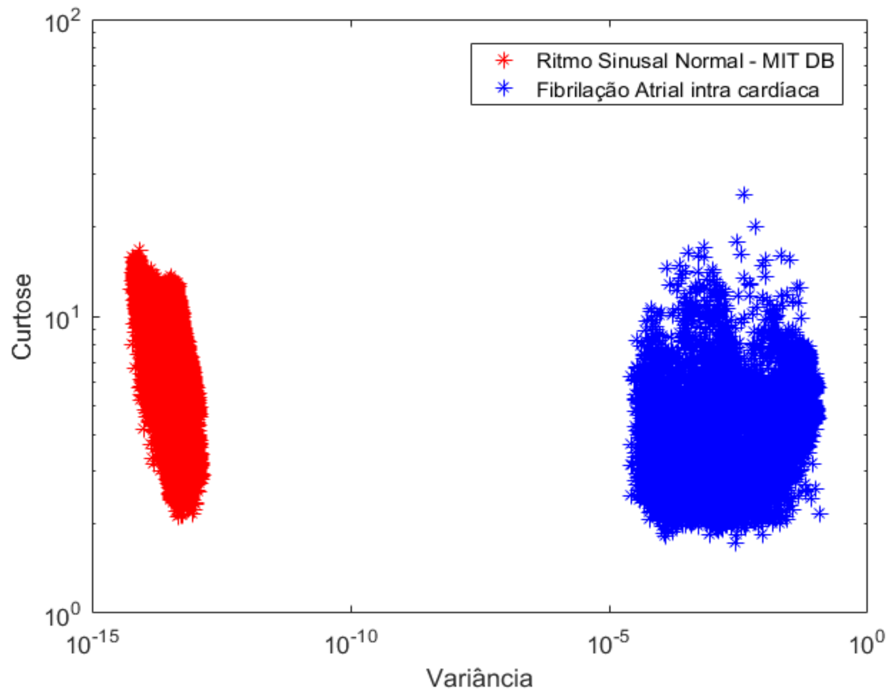

Figure 4. Features expressed by Variance and Kurtosis.

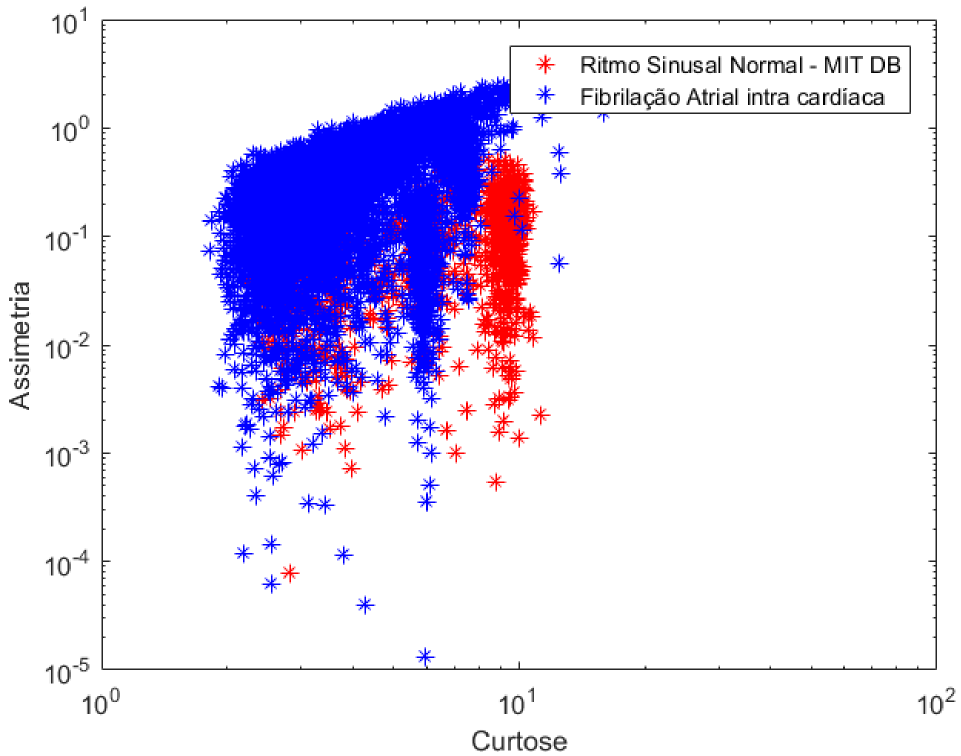

Figure 5. Features expressed by Kurtosis and Skewness.

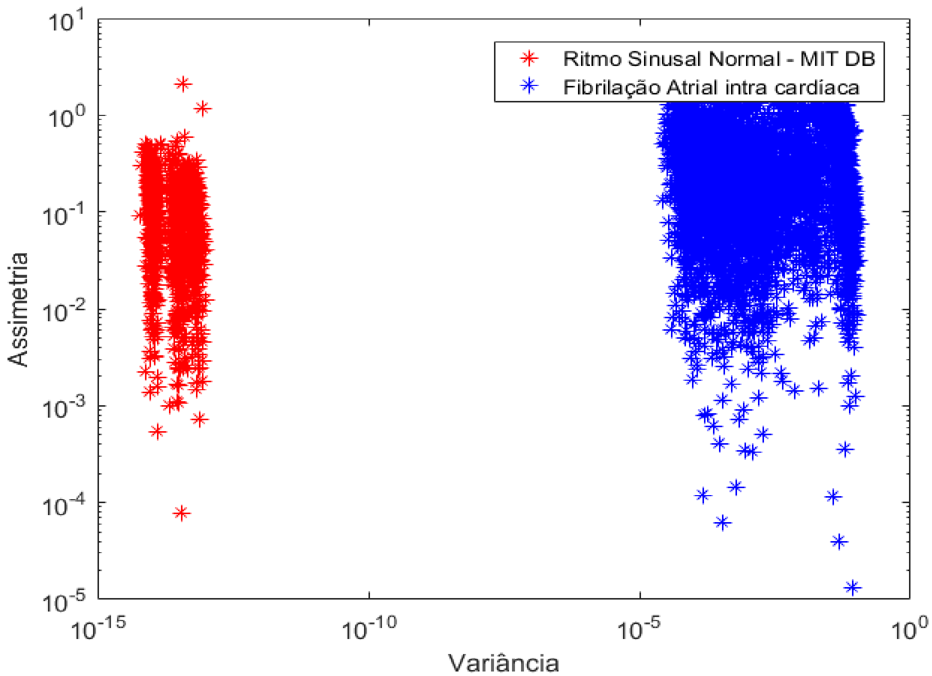

Figure 6. Features expressed by Variance and Skewness.

It was observed that combinations that use variance and kurtosis in their representations of characteristics obtained greater separation between the two groups. This is because ECG of individuals with atrial fibrillation has the highest variance, while individuals with normal sinus rhythm have a smaller variance in the data. In Fig.

7, a three-dimensional representation is shown, taking into account the variance, asymmetry and kurtosis of the set of beats of each group.

It can be seen that although a 3D representation provided a better visualization of the data, the accuracy of using only two characteristics had a greater result in the classifiers used.

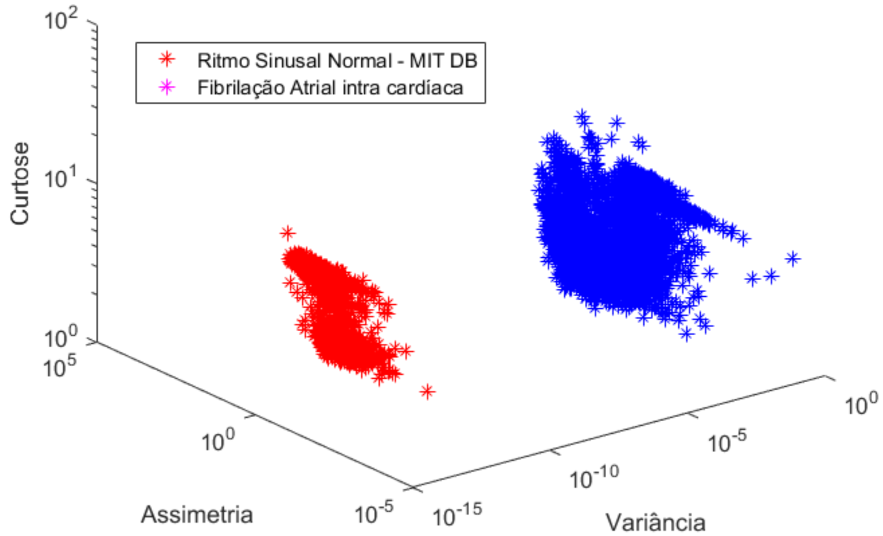

Figure 7. Features expressed by Variance, Skewness and Kurtosis.

5 Conclusions

The proposed method reinforces other studies that show the effectiveness of using high-order statistics to extract characteristics and classify heart diseases such as atrial fibrillation. In addition, the results obtained by the method presented in this article can be used to assist decision making in clinical processes, in addition to detecting arrhythmias autonomously. In future work, different cardiovascular diseases and techniques can be used to improve pre-processing, as well as to apply other classifiers to evaluate metrics.

Acknowledgements. We are greatufull to the CNPQ, and the Biológicas Information Processing Laboratory of UFMA.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] Queiroz JA, et al, Diagnostic decision support systems for atrial fibrillation based on a novel electrocardiogram approach, Journal of Electrocardiology (2017). Avaidable on https://www.researchgate.net/publication/320682762_Diagnostic_decision_support_systems_for_atrial_fibrillation_based_on_a_novel_electrocardiogram_approach. Acesso em 17 de Junho 2020.

[2] v. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation. 101 (23), pp. e215–e220." (2000). Avaidable on: https://archive.physionet.org/physiobank/database/afdb/

[3] Goldberger, A., et al. "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation. 101 (23), pp. e215–e220." (2000). Avaidable on: https://archive.physionet.org/physiobank/database/nsrdb/

[4] World Health Organization (WHO) Cardiovascular diseases. Disponível em https://www.who.int/health-topics/cardiovascular-diseases Acesso em 17 de Junho 2020.

[5] KACHUEE, M.; FAZELI, S; SARRAFZADEH , M. ECG Heartbeat Classification: A Deep Transferable Representation. 2018. Avaidable on <https://arxiv.org/pdf/1805.00794.pdf>. Acesso em 19 de Junho 2020.

[6] RECALL- Registro Brasileiro Cardiovascular de Fibrilação Atrial: Sociedade Brasileira de Cardiologia, 2012. Avaidable on <http://cientifico.cardiol.br/pesquisa/2014/registros/pdf/Protocolo_RECALL_20121212.pdf>

[7] Lucena F, Barros AK, Príncipe JC, Ohnishi N. Statistical coding and decoding of heartbeat intervals. PLoS One 2011;6(6):e20227. Avaidable on https://www.researchgate.net/publication/51239927_Statistical_Coding_and_Decoding_of_Heartbeat_Intervals. Acesso em 19 de Junho 2020.