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

Support method for the diagnosis of Atrial Fibrillation using 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 examination that provides a graphical representation of the electrical activity of the heart. Through it, it is possible to observe the rhythm of heart beats, the number of beats per minute, in addition to enabling the diagnosis of various arrhythmias. This article aims to develop a classification model based on the beats of two groups of individuals, healthy and Atrial Fibrillation. The methodology for the extraction of characteristics based on the morphology of the cardiac signal was adapted to classify Atrial Fibrillation. Classifications were performed in two-dimensional and three-dimensional space, obtaining accuracy from 95% to 100%.

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

1 Introduction

According to the World Health Organization [4], approximately 17.3 million people worldwide are victims of heart disease each year. In Brazil, an estimated 950 people are killed each day by cardiovascular disease, according to the Brazilian Society of Cardiology. Thus, the effective diagnosis of heart disease has driven the development of computational methods that assist in the detection of diseases of this kind.

An examination that quantifies the electrical activity of the heart, making it possible to detect the heart rate and the number of beats per minute, the analysis of the electrocardiogram (ECG). ECG is essential to predict, detect and diagnose various cardiac problems, such as atrial fibrillation, as it is one of the most used non-invasive techniques to assist in this diagnosis [1].

Queiroz et. al. [1] investigate the variation in tension that occurs in a heartbeat t interval using kurtosis. Kachuee, Fazeli, Sarrafzadeh[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 used in this article is illustrated. The databases to be used were defined, separating them into two groups: healthy individuals and individuals who have Atrial Fibrillation. Preprocessing of the database signals was carried out, organizing them for the extraction of features. In this step, the values of variance, skewness and kurtosis of the data set of each base are calculated.

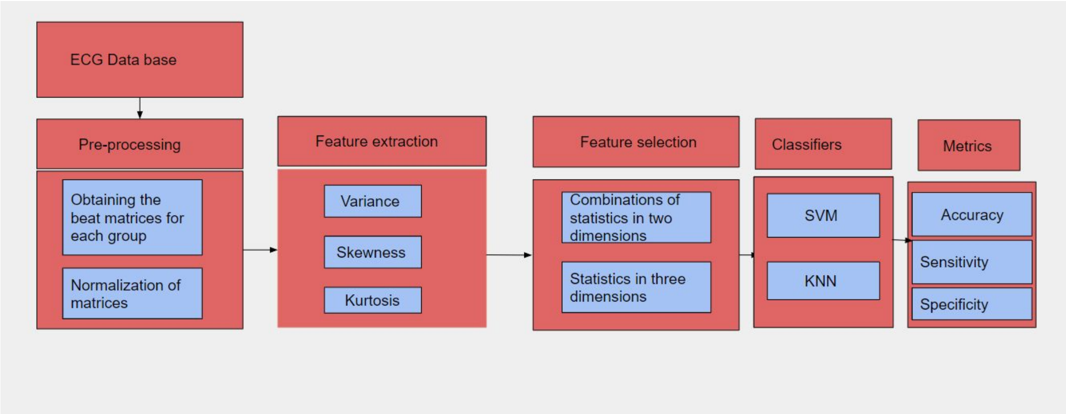

Figure 1. Proposed methodology.

After that, combinations of features were selected, which are represented by high-order 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.

2.1 Data base

The data sets MIT-BIH Rhythm Sinus Normal and the MIT-BIH Atrial Fibrillation Database were used, 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 atrial fibrillation contains 13 patients, all of which are used.

2.2 Pre-processing

ECG signals characteristic of the DII lead, the most used in the world, were selected from both bases. The entire duration of the signal, sampled at a frequency of 256 Hz, was used to extract the beats of each patient for analysis and subsequent extraction of characteristics. After that, each selected signal was segmented in order to obtain the respective beat, according to what is proposed by Queiroz et. al. [1].

Thus, the beats of each group were grouped, generating a matrix A by concatenating the beats of the healthy group, and a B matrix of beats of the group with Atrial Fibrillation, as described in eq. (1):

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

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, given by eq. (2), was subtracted from the sign, and dividing the result by Shannon's entropy:

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

where P represents the probability associated with each beat obtained.

2.3 Feature extraction

The extraction methodology was adapted using high-order statistics, proposed by Queiroz et. al. [1]. A vector was obtained for each of the associated statistics: variance, kurtosis, and asymmetry, which will be the inputs for the classifiers. eq. (3), (4) and (5) represent them, respectively.

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

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

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

2.4 Evaluation metrics

Values of accuracy, sensitivity and specificity, described by eq. (6), eq. (7) and eq. (8) below, were used as evaluation 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 paper analyzed the beats extracted from ECGS for healthy patients with Atrial Fibrillation, in order to classify them. For the classification stage, matrices were generated, where each column is represented by variance, skewness 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 the most efficiency in the classification process.

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 atrial fibrillation group.

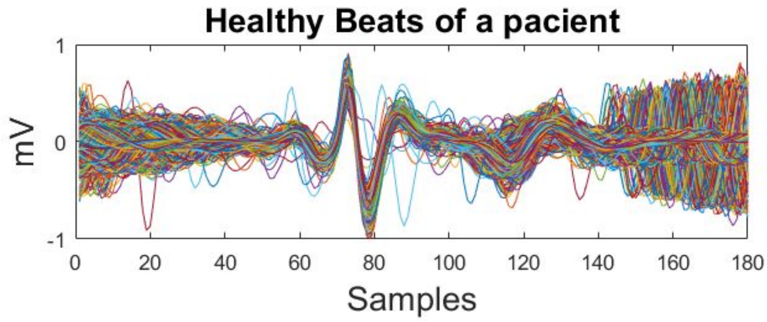

Figure 2. Beats extracted from a healthy patient.

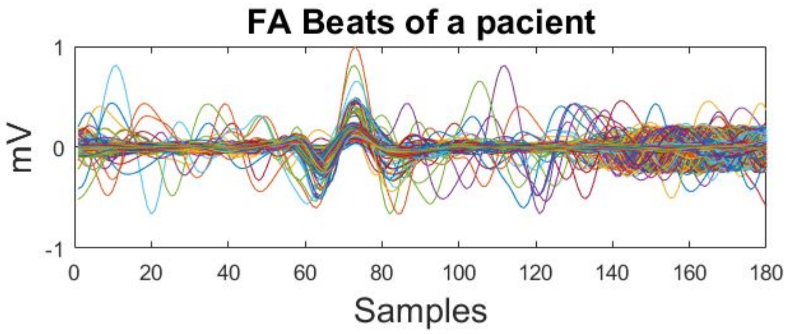

Figure 3. Beats extracted from a patient with Atrial Fibrillation.

Fig. 4 shows the results obtained from the accuracy of the classifications for each group, with each two-dimensional and three-dimensional combination. Fig. 5 shows results related to the sensitivity of each classifier. Finally, in Fig. 6, the values obtained for specificity are contained.

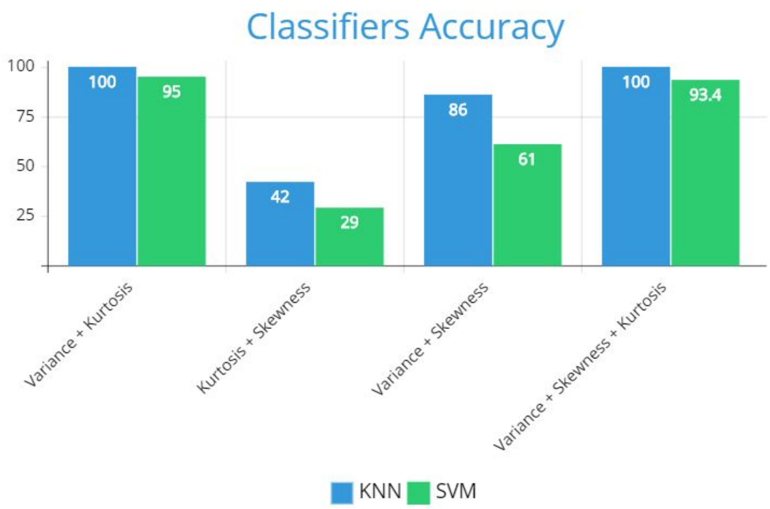

Figure 4. Accuracy of classifiers for each statistical combination.

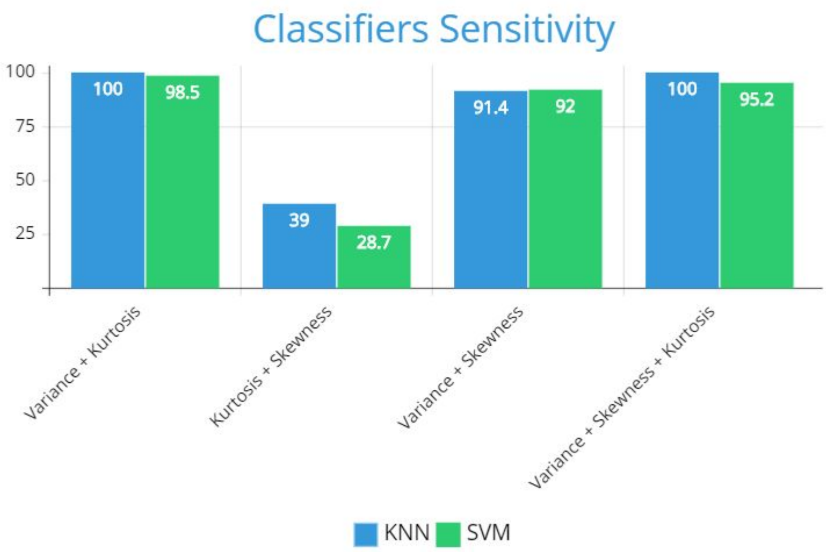

Figura 5. Sensitivity of classifiers for each statistical combination.

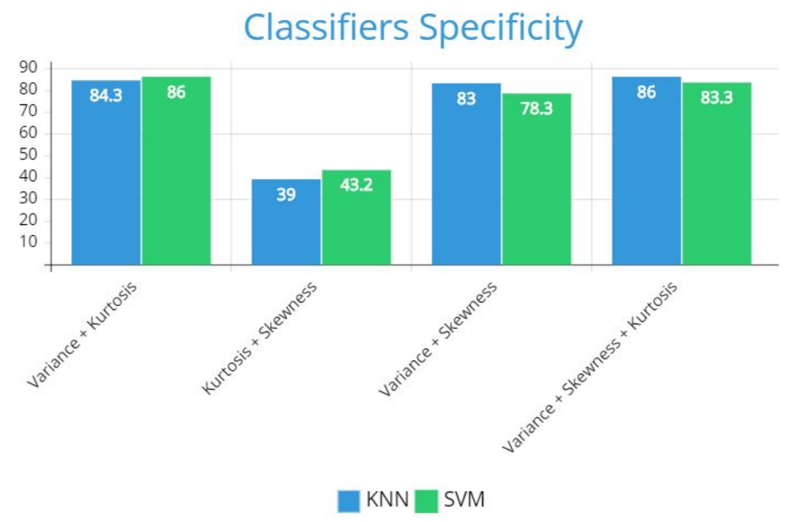

Figure 6. Specificity of classifiers for each statistical combination.

4 Discussion

This work presented analysis of the characteristics in two and three dimensions for the classification of beats with atrial fibrillation. In Figure 7, the analysis in two dimensions shows a separation of the data using variance and kurtosis, as shown in Fig. 7, Fig.8 and Fig.9, 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. 8, the combination of kurtosis and skewness did not achieve a clear separation between groups. In Fig. 6, using the variance and skewness of the data, a better separation can be obtained.

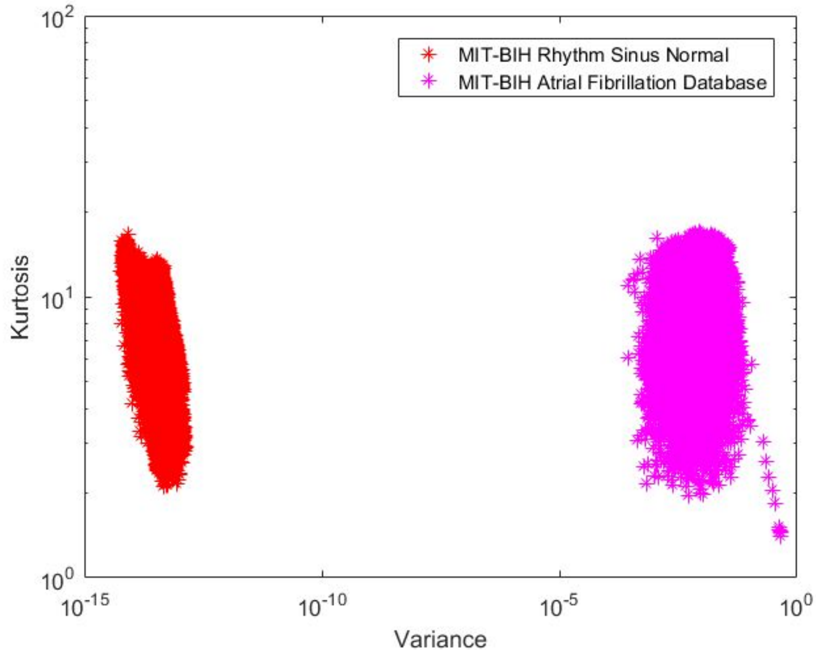

Figure 7. Features expressed by Variance and Kurtosis.

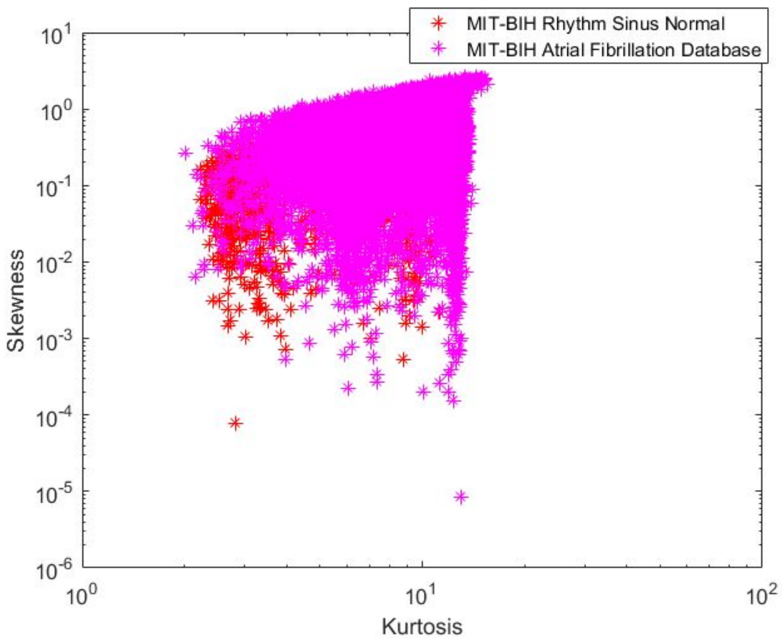

Figure 8. Features expressed by Kurtosis and Skewness.

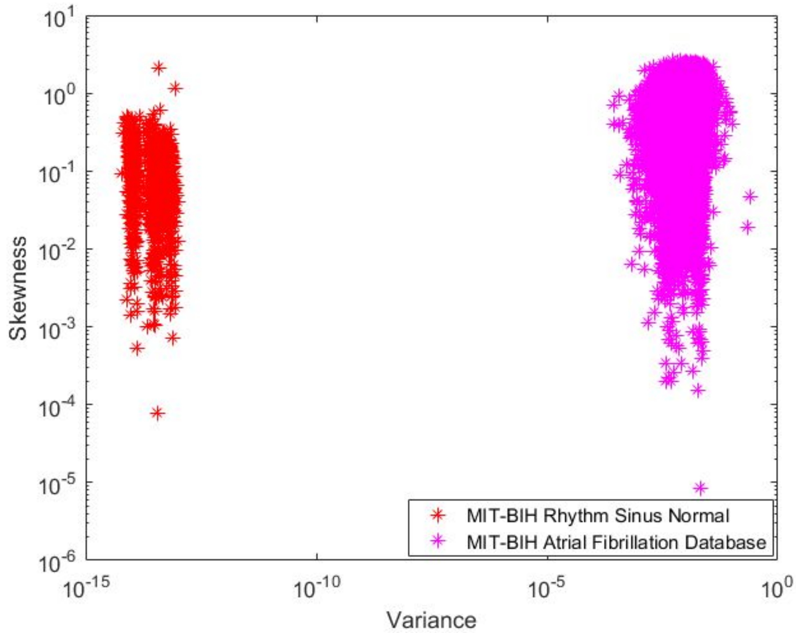

Figure 9. Features expressed by Variance and Skewness.

It can be seen that the combinations that used variance and kurtosis in their representations of features obtained greater separation between the two groups. This was due to the fact that individuals with atrial fibrillation have a greater variance, while individuals with normal sinus rhythm have less variance in the data.

In Fig. 10, a three-dimensional representation is shown, taking into account the variance, skewness and kurtosis of the set of beats of each group.

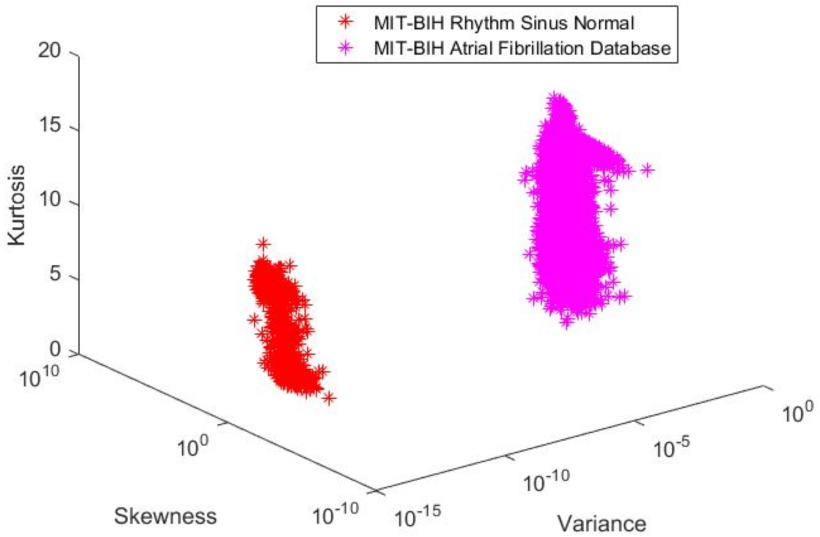

Figure 10. Features expressed by Variance, Skewness and Kurtosis.

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

5 Conclusions

In this article, the effectiveness of using high-order statistics to extract characteristics and classify heart diseases such as atrial fibrillation was reinforced. The results obtained can be used as a basis for decision making of a clinical nature, detecting arrhythmias autonomously. In future works, different cardiovascular diseases can be used in the methodology and techniques can be used to improve the pre-processing, as well as apply other classifiers to evaluate the metrics.

Acknowledgements. We are grateful 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. Avaidable on 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. Disponível em <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.