Prediction of Left Ventricular Ejection Fraction Using an ECG-based LSTM Model in Chagas Disease Patients

João Gabriel S. Ferreira[1], Luís Otávio Rigo Jr.[2], Roberto Coury Pedrosa[3], João Paulo do Vale Madeiro[1]

[1] Federal University of Ceará, Fortaleza, Brazil
[2] University of International Integration of Afro-Brazilian Lusofonia, Redenção, Brazil
[3] Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

Objective: *Classify the left ventricular ejection fraction of chagasic patients into preserved and non-preserved by using electrocardiogram signals.*

Context: *Left ventricular ejection fraction is an important indicator of heart failure and predictor of sudden death. To estimate this indicator, echocardiography is necessary, which is usually more expensive and restrictive than electrocardiography.*

Methods: *Initially, we separated the signals into two classes: ejection fraction less than 0.5 (class 1) and ejection fraction greater than or equal to 0.5 (class 2). We used a Tukey's boxplot to separate noisy beats from non-noisy ones based on their duration. Next, we applied an LSTM (Long-Short Term Memory) network to classify sets of 200 beats of each signal. Finally, we applied an artificial neural network to obtain a class for the entire signal, using the LSTM outputs of each set of 200 beats.*

Results: *We obtained, as the best result, an accuracy of 0.79 and a F_1-score of 0.78.*

Conclusion: *We obtained satisfactory results. However, we believe that they can be improved by a more sophisticated beat selection method and a more robust LSTM model.*

Keywords—**LSTM. Ejection fraction. ECG.**

1. Introduction

In this work, we deal specifically with Chagas disease (CD), an infection caused by the protozoan *Trypanosoma cruzi* which, in its chronic phase, can cause serious cardiovascular problems, such as heart failure [1]. According to the Brazilian Ministry of Health, 1.9 to 4.6 million people have chronic cases of the disease in Brazil [1].

The Rassi [2, 3] score represents an important advance in the prognosis of CD. However, this score is only applicable to patients who already have the cardiac form of CD and requires a complete set of tests, such as 12-lead ECG, echocardiography, thorax radiography, and 24-hour Holter [3].

Among the most relevant metrics for this score, the left ventricular ejection fraction (LVEF), estimated by echocardiography, is notable. LVEF corresponds to the percentage of blood ejected from the left ventricle into the aorta during systole, constituting an important indicator of heart failure [4].

Access to public health is limited in Brazil, especially in remote areas, where many patients remain on long waiting lists for echocardiography. On the other hand, ECGs are more common, cheaper, and more suitable than echocardiograms. Several studies [3, 5, 6] prove the capability of machine learning algorithms to identify diseases from electrocardiographic signals.

Often, these algorithms do not perform well in real-world scenarios due to the unique characteristics of the signals of each person, leading to unreliable performance when trying to classify a new patient's ECG signal [7].

A common approach to this problem is to add to the training set a short signal (with a few minutes) from the patient who wants to be diagnosed [7, 8].

The present study, on the other hand, seeks a model capable of accurately predicting LVEF from the analysis of generic one-lead ECG signals, which means that no specific components are required for the diagnosis of a defined patient.

2. Methodology

2.1. Data description

We used Holter signals from 219 patients with Chagas heart disease, obtained at the Clementino Fraga Filho University Hospital – Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. Each signal is 4 hours long, with

a sampling rate of 128Hz. The local ethics committee approved the survey (number 45360915.1.1001.5262).

In addition, we received the LVEF value for each patient/exam and files containing, for each beat of each signal, the positions of the beginning, peak and end of the P wave, the beginning and end of the QRS complex, the R peak and the beginning, peak and end of the T wave.

2.2. Preprocessing

There is no consensus when defining a threshold between compromised and preserved LVEF, so this can be 35%, 40%, 45%, or 50% [9]. Therefore, we chose to use the guidelines of the American Heart Association [10], which define LVEF $\leq$ 40% as reduced, 41% $\leq$ LVEF $\leq$ 49% as slightly reduced and LVEF $\geq$ 50% as preserved.

In this work, we opt for a binary classification, where class 1 represents a not preserved LVEF ($\leq$ 49%) and class 2 represents a preserved LVEF ($\geq$ 50%).

Then, we divided our dataset into three subsets: training set (with 72.25% of data), validation set (with 12.75% of data) and test set (with 15% of data).

Next, we define a beat window (BW) as starting at the start of each P wave and ending at the end of the subsequent T wave. For each signal, we removed the beats whose length exceeded the lower or upper limits of Tukey's boxplot and, after that, we applied a zero padding to leave the remaining BWs with the same length.

The process of removing outliers using Tukey's boxplot was necessary due to anomalies (artifacts) that appeared in some signals (see Figure 1). Usually, excessively long or short BWs are artifacts.

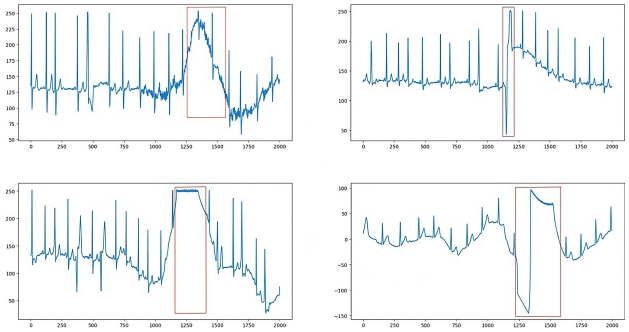


Figure 1: Examples of artifacts present on signals.

However, the removal process breaks (to some extent) with the sequential character of the signals, making the classification task difficult. An issue is discussed later.

2.3. Processing

LSTM networks, due to their ability to store information for long periods of time, are ideal for dealing with time-series patterns, such as ECG signals. Therefore, we propose an LSTM network architecture that receives almost sequential sets of 200 BWs (approximately 5 minutes) as input. We say almost sequential, as the chaining of these BWs may have been affected by removals made by Tukey's boxplot.

We also chose to train the LSTM network with sets of BWs for two main reasons. First, training with sets of BWs is faster than training with several single BWs. Second, with a sequence of BWs as input, the network can extract information that could not be obtained with an isolated single BW (such as the distance between the R peaks, for example).

The proposed network is shown in Figure 2. The analyzed model is initially composed of N consecutive LSTM layers, each containing U units.

Next, we add M dense layers, where the first has k neurons, the second $\lfloor k/2^1 \rfloor$, the third $\lfloor k/2^2 \rfloor$, and so on. We chose to use a ReLU activation function in all dense layers, except for the output layer. Each fully connected layer is followed by a dropout layer of rate D.

Finally, we have the output layer with 2 neurons. We choose the softmax activation function, which, in this case, is equal to the sigmoid function.

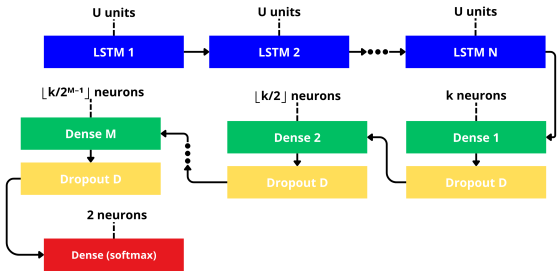


Figure 2: Proposed architecture.

Furthermore, for the calculation of the error, we opted for the categorical cross-entropy function.

All the mentioned variables (N, U, M, k, D) were defined through Bayesian optimization (BO). Below, we can observe the intervals where we pre-set the algorithm to seek optimal values.

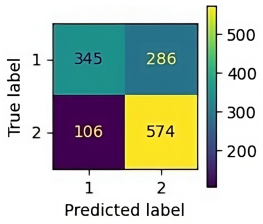
- Number of LSTM layers (N): $1 \leq N \leq 4$, $N \in \mathbb{Z}$
- Number of units of each LSTM layer (U): $U = 20 \cdot u$, where $3 \leq u \leq 8$, $u \in \mathbb{Z}$
- Number of dense layers (M): $1 \leq M \leq 4$, $M \in \mathbb{Z}$
- Number of neurons in the first dense layer (k): $k = 2^K$, where $5 \leq K \leq 7$, $K \in \mathbb{Z}$
- Dropout rate for each dense layer (D): $0.3 \leq D \leq 0.6$, $D \in \mathbb{R}$
- Batch size (L): $L = 2^l$, where $6 \leq l \leq 9$, $l \in \mathbb{Z}$
- Optimizer (ot): $ot = \{$Adam, RMSprop, Adadelta, Adagrad, Adamax, SGD$\}$

- Learning rate (η): $\eta = \{0.1, 0.01\}$

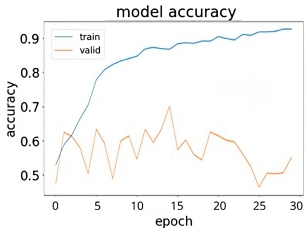
We adopted, as covariance function, the Matérn function with $\nu = 1.5$ and, as acquisition function, the LSC function with $\kappa = 2.576$, standard parameters for the library *Bayesian Optimization* [11]. The objective function is the accuracy in the validation set.

We configured the BO to perform 35 random iterations and 85 iterations of the Bayesian algorithm itself. In Figures 3a and 3b, it is possible to see, respectively, the confusion matrix of the best model and its learning curve.

Observing Figure 3b, it is possible to realize an overfitting, probably caused by the gaps left in the ECG signal by the Tukey's boxplot removal process.



(a) Validation set confusion matrix.



(b) Learning curve.

Figure 3: Best performing model results.

As already mentioned, the classification process shown in Figure 3a refers to sets of 200 BWs. Therefore, we need a strategy to classify a full signal based on the classification results of its sets of 200 BWs. Remember that, for each set of classified BWs, the network returns (due to the *softmax* function) a tuple of two elements corresponding to the probability that the input belongs to class 1 and class 2, respectively.

Therefore, we adopted the strategy explained in the following subsection.

2.4. Classification Method

Imagine, for example, a signal with 1000 BWs, that is, 5 sets of 200 BWs. If, the output generated by LSTM, for each of these sets, is: [0.6, 0.4], [0.1, 0.9], [0.2, 0.8], [0.55, 0.45] and [0.65, 0.35], we can, then, add all corresponding probabilities. Therefore, we have a total of 0.6+0.1+0.2+0.55+0.65 = 2.1 for class 1 and 0.4+0.9+0.8+0.45+0.35 = 2.9 for class 2.

Instead of just verify which value is bigger, it is possible to pass the tuple [2.1, 2.9] as input of an Artificial Neural Network (ANN). This procedure seems more interesting, as the ANN can establish weights and thresholds for the probabilities of each class, improving the comparison process.

As, in general, we have few sets of 200 BWs per signal, we proposed a very simple architecture for the ANN, which can be seen in Table 1.

Layer	Nº of neurons
Dense	32
Dense	8
Dense	4
Softmax	2

Table 1: Proposed ANN model.

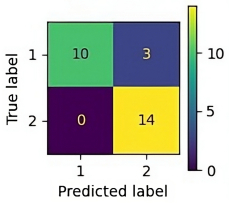


Figure 4: ANN confusion matrix.

For training the ANN, we used the Adam optimizer, a batch size equal to 8 and a learning rate equal to 0.001. The model was trained for 800 epochs. The resulting confusion matrix (for the validation set) can be seen in Figure 4

3. Results

In this section, we apply the LSTM model, followed by the ANN, to the test set.

Initially, applying the LSTM network to classify the sets of BWs, we obtain the results shown in Figure 5 and in Table 2

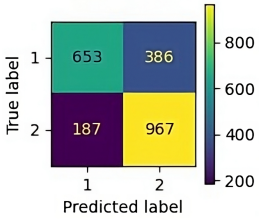


Figure 5: Confusion matrix related to the classification of sets of 200 BWs in the test set.

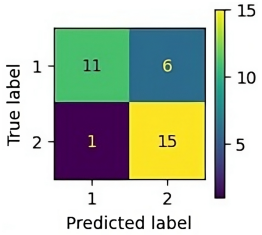


Figure 6: Confusion matrix related to the ANN classification.

	Precision	Recall	F_1-score
Class 1	0.78	0.63	0.70
Class 2	0.71	0.84	0.77
Accuracy			0.74
Macro	0.75	0.73	0.73
Weighted	0.74	0.74	0.74

Table 2: Metrics related to the classification of sets of 200 BWs in the test set.

In Table 2, we call a metric 'macro' when it is equivalent to the arithmetic mean of the individual metrics of each class. Furthermore, we call a metric 'weighted' when it is equivalent to the weighted mean of the individual metrics of each class, where the weights are proportional to the number of samples of the respective classes.

Finally, applying the ANN provides to obtain the results seen in Figure 6 and Table 3.

	Precision	Recall	F_1-score
Class 1	0.92	0.65	0.76
Class 2	0.71	0.94	0.81
Accuracy			0.79
Macro	0.82	0.79	0.78
Weighted	0.82	0.79	0.78

Table 3: Metrics related to the ANN classification.

4. Conclusion

The results evidence the capability of the proposed model to differentiate ECG signals from patients with preserved and non-preserved LVEF. However, the metrics obtained (accuracy of 0.79 and F_1-score of 0.78) reveal that this model isn't ready to be applied in real-world scenarios, where metrics above 0.9 would be desirable.

In part, these limitations are caused by Tukey's boxplot within the BWs selection. Since this method only considers the duration of the BWs, it is not capable to really differentiate noisy and non-noisy beats. Furthermore, the removal of beats breaks with the sequential character of the signals, making the learning process even harder.

Therefore, in future works, it would be extremely important to implement a more robust noise detection algorithm, such as noise automatic classification algorithm (NACA) [12] or quality measurement algorithm (QMA) [13].

Acknowledgments

We are thankful for the financial support of the Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico - Funcap, Process PS1-0186-00439.01.00/21.

References

[1] Doença de Chagas. Ministério da Saúde Apr 2022;.

[2] Rassi Jr A, Rassi A, Little WC, Xavier SS, Rassi SG, Rassi AG, Rassi GG, Hasslocher-Moreno A, Sousa AS, Scanavacca MI. Development and validation of a risk score for predicting death in chagas' heart disease. New England Journal of Medicine 2006;355(8):799–808.

[3] Silva LEV, Moreira HT, Bernardo MMM, Schmidt A, Romano MMD, Salgado HC, Fazan Jr R, Tinos R, Marin-Neto JA. Prediction of echocardiographic parameters in chagas disease using heart rate variability and machine learning. Biomedical Signal Processing and Control 2021; 67:102513.

[4] Albuquerque N, Oliveira A, Mota da Silva J, Peres A, Araujo T. Relationship between cardiac ejection fraction and blood pressure in coronary heart patients. Rev Rene 11 2017;18:584–590.

[5] Ulloa-Cerna AE, Jing L, Pfeifer JM, Raghunath S, Ruhl JA, Rocha DB, Leader JB, Zimmerman N, Lee G, Steinhubl SR, et al. rechommend: An ecg-based machine learning approach for identifying patients at increased risk of undiagnosed structural heart disease detectable by echocardiography. Circulation 2022;146(1):36–47.

[6] Aziz S, Ahmed S, Alouini MS. Ecg-based machine-learning algorithms for heartbeat classification. Scientific Reports 2021;11(1):1–14.

[7] Kiranyaz S, Ince T, Hamila R, Gabbouj M. Convolutional neural networks for patient-specific ecg classification. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015; 2608–2611.

[8] Li Y, Pang Y, Wang J, Li X. Patient-specific ecg classification by deeper cnn from generic to dedicated. Neurocomputing 2018;314:336–346.

[9] Solal AC, Leurs I, Assyag P, Beauvais F, Clerson P, Contre C, Thebaut JF, Genoun M, of Cardiologists FNC, et al. Optimization of heart failure medical treatment after hospital discharge according to left ventricular ejection fraction: the future survey. Archives of cardiovascular diseases 2012; 105(6-7):355–365.

[10] Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, Deswal A, Drazner MH, Dunlay SM, Evers LR, et al. 2022 aha/acc/hfsa guideline for the management of heart failure: a report of the american college of cardiology/american heart association joint committee on clinical practice guidelines. Journal of the American College of Cardiology 2022;79(17):e263–e421.

[11] Nogueira F. Bayesian Optimization: Open source constrained global optimization tool for Python, 2014–. URL `github.com/fmfn/BayesianOptimization`.

[12] de Moraes JL, Paixão GM, Gomes PR, Mendes EM, Ribeiro ALP, Beda A. A novel algorithm to assess the quality of 12-lead ecg recordings: validation in a real telecardiology application. Physiological Measurement 2023; 44(3):035006.

[13] Hayn D, Jammerbund B, Schreier G. Qrs detection based ecg quality assessment. Physiological measurement 2012; 33(9):1449.

Address for correspondence:

Dr. João Paulo do Vale Madeiro.

Department of Computing Science, Federal University of Ceará, Campus do Pici, 60440-900, Fortaleza, Ceará, Brazil.

Email: jpaulo.vale@dc.ufc.br