Machine Learning-Based Cardiac Arrhythmia Detection in Electrocardiogram Signals

João Vitor Mendes Pinto dos Santos[1]*, Thamiles Rodrigues de Melo[1]
[1]SENAI CIMATEC University Center; Salvador, Bahia, Brazil

The cardiovascular system is vital for human physiology, regulating blood circulation. Cardiovascular Diseases (CVDs), including cardiac arrhythmias, can disrupt the heartbeat rhythm, impacting blood circulation. Black-box computational modeling of this system can facilitate the development of novel methods and devices to assist in diagnosing and treating CVDs. Artificial Neural Networks (ANNs) represent an effective black-box approach. Implementation involves selecting a database, separating training and test sets, and defining the model structure. The MIT-BIH database is commonly utilized to train computational models to detect cardiac arrhythmias. However, preliminary results with the ANN model trained using MIT-BIH data failed to meet the expected objectives, presenting numerous challenges. Nonetheless, given its nascent stage, there remains potential for optimizations, rendering it a prospective tool for diagnosing cardiac arrhythmias.
Keywords: Electrocardiogram. Cardiac Arrhythmias. Artificial Intelligence. Machine Learning.

The cardiovascular system plays a pivotal role in human physiology, orchestrating blood circulation throughout the body. Its proper functioning is paramount for maintaining optimal blood pressure and flow and the effective distribution of oxygen and nutrients essential for sustaining physiological well-being. However, deviations in cardiovascular function, often precipitated by cardiovascular diseases (CVDs), can significantly compromise these vital processes. CVDs stand as a leading cause of mortality globally, contributing to approximately 17.9 million deaths annually, accounting for 31% of all global deaths [1].

The heart's electrical activity, crucial for regulating cardiac function, can be comprehensively represented through an electrocardiogram (ECG) (Figure 1). The ECG waveform encapsulates essential information about the various electrical waves governing the cardiac cycle. Within the spectrum of CVDs, cardiac arrhythmias emerge as electrical aberrations that disrupt the temporal intervals between QRS complexes in the ECG signal.

Received on 10 February 2024; revised 15 May 2024.
Address for correspondence: João Vitor Mendes Pinto dos Santos. SENAI CIMATEC University Center. Av. Orlando Gomes, 1845 - Piatã, Salvador, Bahia, Brazil. E-mail: joao.v.m.p.santos@ieee.org.br.

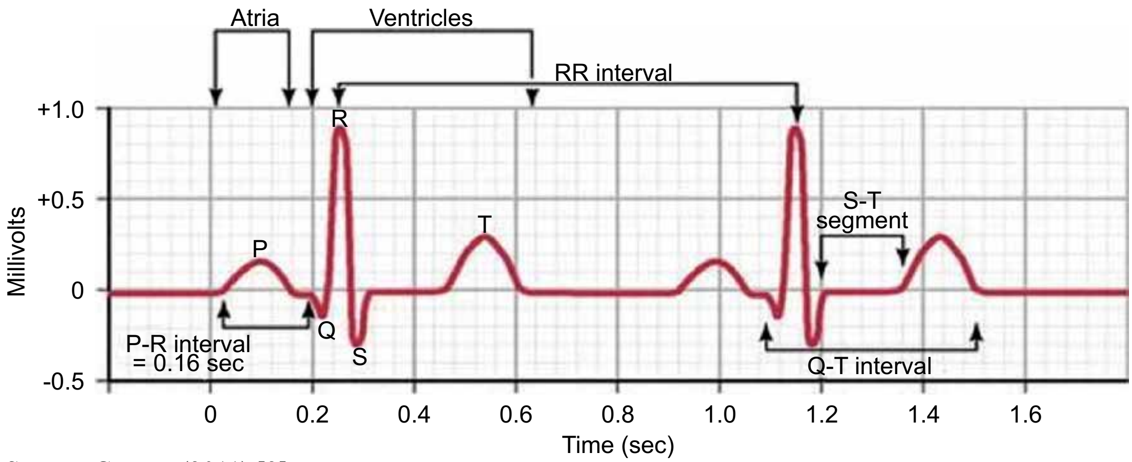
J Bioeng. Tech. Health 2024;7(2):114-116

These disturbances manifest as irregular heartbeat rhythms, thereby perturbing the harmonious functioning of the cardiovascular system [2].

Computational modeling of the human cardiovascular system offers insights into heart physiology and facilitates the development of diagnostic and therapeutic methods for cardiovascular diseases. However, like many physical phenomena, the cardiovascular system is characterized by Non-Linear and Time-Varying Systems (NLVT), posing challenges for classical analytical modeling and control methods. Black box modeling, a system identification method, circumvents this challenge by heuristically deriving an approximate mathematical model based on input-output relationships without explicitly modeling the system's physical behavior. Artificial Intelligence (AI), particularly Machine Learning (ML), has emerged as a powerful tool for implementing black box modeling solutions. ML, a subarea of AI, enables the creation of artificial neural networks (ANNs), computational models trained using datasets to perform desired functions

In this context, this research aims to develop an ANN model capable of accurately identifying the location of R peaks in electrocardiogram (ECG) signals. This model aims to evaluate the patient's heart rhythm and aid in diagnosing cardiac arrhythmias [3].

Figure 1. Capture of a typical ECG signal.



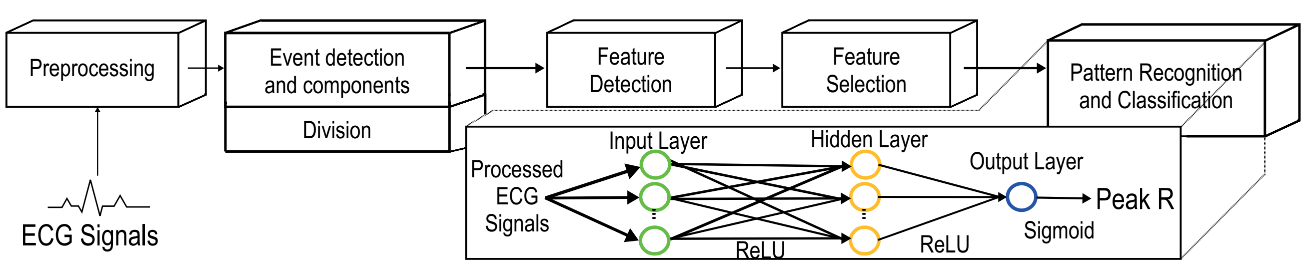
Source: Guyton (2011) [2].

Materials and Methods

Figure 2 illustrates the typical process of training mathematical models using machine learning (ML) neural network architectures. Following this process, the electrocardiogram (ECG) database selected for this study was the MIT-BIH database [4], renowned as one of the most frequently utilized databases in the scientific literature [5]. The chosen artificial neural network (ANN) architecture was the multilayer perceptron (MLP) classifier. The configuration of the MLP, as depicted in Figure 2, was established with three layers: the input layer comprising 20 neurons, employing the rectified linear unit (ReLU) activation function, is connected to the hidden layer, which consists of 10 neurons, also activated by the ReLU function. Subsequently, the hidden layer is linked to the output layer utilizing the sigmoid activation function, generating the desired outcome. The MLP classifier will be implemented within the TensorFlow and sci-kit-learn libraries within the Python programming language development environment.

The training and validation steps of the computational model involve utilizing a subset of records from the database [4] for training purposes and a separate subset for testing and validating the model. Initially, a unitary set containing record 100 was utilized for training, selected due to its high signal quality and sinus rhythm falling within the expected range for a healthy heart [4]. For the test set, six records were randomly chosen. This approach assesses the model's efficacy in recognizing patterns across various records, regardless of their previously known condition. Averages of Precision (PRC), Recall (RC), F1 Score

Figure 2. Block diagram of computer-aided arrhythmia classification systems.



Adapted from Hammad and colleagues (2021) [3].

(F1), and Accuracy (ACC) metrics were computed to evaluate the performance of the model. These metrics were derived from True Positive (Tp), True Negative (Vn), False Positive (Fp), and False Negative (Tn) values, as outlined in the literature [6].

Results and Discussion

Figure 3 depicted the results of the predictions generated by the trained artificial neural network (ANN) model. The outcomes reveal three distinct behaviors: Accurate identification of a single peak per R wave, as illustrated in Figure 3(a) Multiple markings on the same peak within a single R wave, exemplified in Figure 3(b) The absence of any markings denoting R peaks, as illustrated in Figure 3(c).

The results presented in Table 1 indicate that the model predictions exhibit a low rate of False Positive (Fp) values but a high rate of False Negative (Fn) values. Consequently, the model demonstrates high precision but low recall, as precision measures the proportion of true positive predictions among all positive predictions. In contrast, recall measures the proportion of true positive predictions among all actual positive instances. On a scale from 0 to 1, the average F1 score suggests that the model's performance is moderate, considering the significant disparity in recall values. Notably, the trained model proves ineffective for record 114, as no detected R peak corresponds to a True Positive (Tp), resulting in accuracy being the only relevant metric.

Overall, the model exhibits high accuracy, primarily because this metric accounts for all

Figure 3. Temporal template for predicting R peaks: (a) correct marking; (b) multiple markings on the same peak; (c) absence of marking.

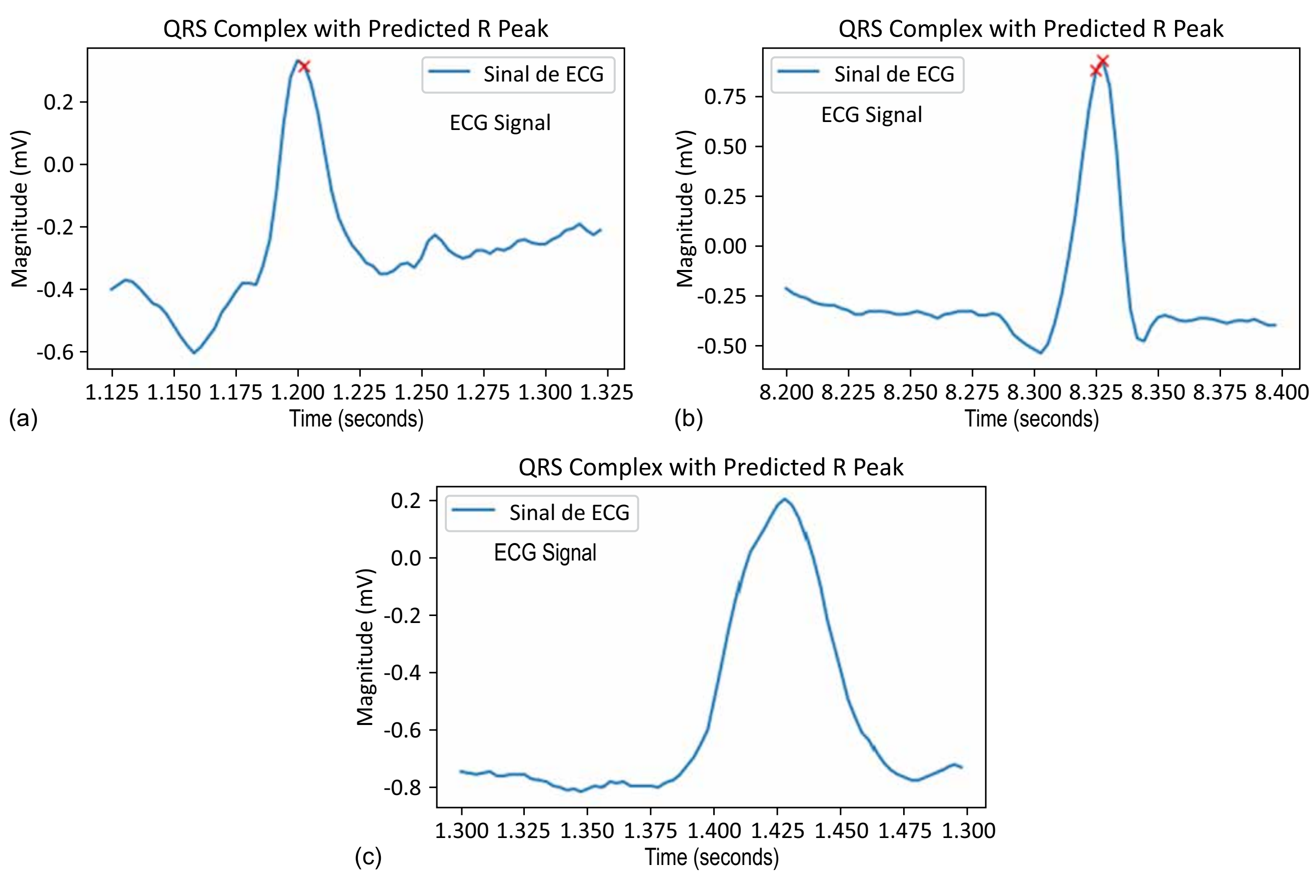


Table 1. Metrics adopted for evaluating the performance of the trained ANN model.

Test Recorder	$PRC = \frac{Tp}{Tp+Fp}$	$RC = \frac{Tp}{Tp+Fn}$	$F1 = \frac{2*PRC*RC}{PRC+RC}$	$ACC = \frac{Tp+Tn}{Tp+Tn+Fp+Fn}$
105	0.83	0.66	0.73	0.99
114	0	0	0	0.99
121	0.89	$4.26*10^{-3}$	$8.48*10^{-3}$	0.99
200	0.90	0.38	0.53	0.99
215	0.93	0.67	0.77	0.99
228	0.78	0.36	0.49	0.99
Median	0.72	0.35	0.42	0.99

outcomes and the inherent imbalance between R-peaks classes and non-R-peaks. Notably, R-peak signals are inherently more challenging to detect, contributing to the model's observed performance.

A contributing factor to these results may be utilizing only one record for training the model, leading to an inadequately sized database. Other factors potentially influencing performance include noise in the ECG signal and the accuracy of R peak annotations used during training. These challenges pertain to the ECG signal's pre-processing and event and wave detection stages. To address these issues, Hammad and colleagues and Xiao and colleagues, in their systematic literature reviews, suggest employing traditional and wavelet filters for noise reduction and algorithms such as the Pan-Tompkins algorithm and the Savitzky filter-Golay.

Conclusion

This study proposed a system for identifying cardiac arrhythmias using artificial neural networks (ANNs) to aid medical diagnoses. The ANN model developed within the Python development environment demonstrated proficiency in identifying R peaks from the ECG signal. While the achieved results have yet to meet the desired research objectives fully, it is essential to note that the model's development is in its nascent stage, with numerous optimization opportunities awaiting exploration. Furthermore, several more advanced ANN architectures could be investigated in future studies, including Convolutional Neural Networks, Long Short-Term Memory networks, and hybrid architectures.

Acknowledgments

The authors thank the National Council for Scientific and Technological Development (CNPq) for their financial support in conducting this research.

References

1. OPAS. Organização Pan-Americana da Saúde. Doenças Cardiovasculares. 2017. Available at: <https://www.paho.org/pt/topicos/doencas-cardiovasculares>. Acessed on Sept. 29, 2023.
2. Guyton AC, Hall JE. Tratado de Fisiologia Médica. 12. ed. São Paulo, SP: Elsevier, 2011.
3. Hammad M, Kandala RNVPS, Abdelatey A et al. Automated detection of shockable ECG signals: A review. Information Sciences 2021;571:580–604.
4. Moody GB, Mark RG. MIT-BIH Arrhythmia Database. Available at:<https://physionet.org/content/mitdb/>. Acessed on March 15, 2024.
5. Xiao Q, Lee K, Mokhtar SA et al. Deep learning-based ECG arrhythmia classification: A systematic review. Applied Sciences 2023;13(8):4964.
6. Nahiduzzaman, Nayeem J, Ahmed T et al. Prediction of heart disease using multi-layer perceptron neural network and support vector machine. In: 2019 4th International Conference on Electrical Information and Communication Technology (EICT). 2019:1–6.