OpenHeart: Detection of Chagas Disease from Single-Lead Electrocardiogram

Edo Ikurumi[1]

[1]Department of Advanced Cardiology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Abstract

Chagas disease can cause severe cardiac complications, yet early detection is often limited by a lack of access to diagnostic tools like 12-lead electrocardiograms (ECG). This study investigates using features from short, single-lead ECGs—increasingly available via consumer devices—for detection. We developed machine learning models using ECGs from the Sami-Trop (N=362) and PT-XL (N=385) datasets.

A Convolutional Neural Network (CNN) model achieved promising performance on cross validation, with an Area Under the Curve (AUC) of 0.93. A support vector machine (SVM) trained on RR intervals achieved an AUC of 0.63 on the same. The models achieved, mean challenge scores on the test set of 0.056 and 0.078 respectively. In comparison, a support vector machine (SVM) trained on engineered RR interval features achieved a cross-validation AUC of 0.63. On the hidden test set, these models achieved mean challenge scores of 0.056 and 0.078 respectively.

These results highlight the varying performance of different modeling approaches and suggest that while deep learning models can be effective, feature-based methods provide a simpler baseline for this detection task.

1. Introduction

This research was undertaken as part of the PhysioNet challenge for CinC 2025 [1,2,3]. Chagas disease is fundamentally a disease of poverty inextricably linked to the social and environmental conditions that facilitate its transmission. Consequently, a low-cost approach to diagnosis and monitoring is highly desirable. In Brazil, the burden of Chagas disease is not uniformly distributed but presents with distinct hotspots. Furthermore, acute outbreaks are also a concern, necessitating a portable diagnostic solution.

Chagas cardiomyopathy, a significant consequence of the disease, frequently leads to cardiomegaly and is known to be arrhythmogenic. This suggests that detection methods for these cardiac issues could also be applicable to Chagas disease. In recent years, advancements in single-lead ECG devices have significantly improved their capability to detect arrhythmias. Alongside this, there have been notable developments in remote photoplethysmography (PPG), which involves capturing PPG signals from video data. This presents a highly portable solution with the potential for algorithms that utilise R-R intervals, offering a non-invasive and accessible approach.

The primary objective of this paper was to consider algorithms that can ultimately be deployed in the most cost-effective and portable manner possible, aligning with the urgent need for accessible healthcare solutions in regions affected by Chagas disease. By exploring these technological advancements, we aim to contribute to the development of practical tools that can address the diagnostic and monitoring challenges posed by this debilitating condition, particularly in resource-limited settings.

2. Methods

2.1. Overview

My research aimed to develop a model that primarily relies on lead I ECG data although an alternative approach utilising focusing on rhythm features was also explored. Due to the inherent limitations of information available in lead I ECG, particularly concerning the rhythm data that can be definitively extracted, my investigation was structured in two distinct phases. Initially I conducted an in-depth analysis of rhythm features and their correlation with Chagas disease, followed by the subsequent creation and evaluation of the predictive model.

As early detection of Chagas disease is crucial, an additional analysis involving only ECGs with no detected abnormalities was also carried out.

2.2. Datasets

This study utilised several public ECG databases. The primary Chagas-positive cohort was sourced from the

Sami-Trop dataset [4]. The primary control (Chagas-negative) cohort was taken from the PTB-XL dataset [5].

For the CNN model development, 803 records from Sami-Trop and 803 from PTB-XL were used, with all recordings chosen being longer or equal to 7 seconds in length. Longer records were truncated to a uniform 7-second length.

For the feature-based model development, 362 records from Sami-Trop and 362 from PTB-XL were used, with all recordings chosen being longer or equal to 10 seconds in length. Longer records were truncated to a uniform 10-second length. Additionally, any records where less than 8 RR intervals were detected were excluded from training.

For local evaluation, records from the CODE-15 dataset [6] were used using the respective length parameters for each model approach.

A critical sub-analysis on "normal" ECGs was performed to test for early-stage detection capabilities. For this purpose, three distinct datasets were created. The first, SET A, was a strictly controlled comparison between 62 "normal" ECGs from Sami-Trop and 66 ECGs from PTB-XL explicitly annotated as "sinus rhythm normal ekg" to isolate subtle Chagas-related changes. The second, SET B, was a less strict comparison, matching the 62 "normal" Sami-Trop records against 63 records from PTB-XL with the general code "NORM," which could include minor variations like sinus tachycardia or bradycardia. The third, SET C, served as an external validation set, using 78 "normal" ECGs with a 1:1 ratio of Chagas to control from the CODE-15 dataset.

For preprocessing, the raw single-lead ECG signal was processed to identify R-peaks using the well-established Pan-Tompkins algorithm. The resulting sequence of RR intervals was then filtered to remove outliers and retain only normal-to-normal sinus beat intervals for robust feature calculation.

In the case of inference on the CNN model, records shorter and 7 seconds were padded, while longer records were truncated. For the feature models, any record with less than 6 detected RR intervals was deemed negative by default.

2.3. Rhythm Analysis

A comprehensive set of 31 features was engineered from the preprocessed 10-second ECG signals to capture a wide range of cardiac dynamics. These were divided into two main categories. The first category consisted of 24 RR interval features focused on quantifying heart rate variability (HRV). These included standard time-domain statistical measures such as the mean and median of RR intervals, heart rate (HR), SDNN, RMSSD, pNN20, and pNN50. Geometric features were also derived from the Poincaré plot, including the standard deviations along the short (SD1) and long (SD2) axes and their ratio, alongside histogram-based features like the Triangular Index (TI) and the Baevsky Stress Index (BSI). Finally, non-linear measures of signal complexity were computed to capture subtle irregularities, including the correlation dimension, RR entropy (RREN), SD entropy, RR wavelet entropy (RRWEN), and SD wavelet entropy. The second category comprised 7 QRS amplitude features to capture morphological changes in the QRS complex. These included the mean, median, standard deviation, and entropy of the QRS peak amplitudes, as well as the minimum, maximum, and range (gap) of amplitudes within the 10-second window.

2.4. Models

Three distinct modeling strategies were pursued to find an optimal balance between performance and generalizability.

The first was a lightweight 1D Convolutional Neural Network (CNN) trained on the raw ECG waveform as an end-to-end model. This model was trained using the Adam optimizer with a cosine learning rate scheduler with soft restarts, an initial learning rate of 1e-2, and a final learning rate of 1e-5, for a maximum of 140 epochs.

The second and third strategies involved traditional machine learning classifiers trained on the engineered features. For both feature-based models, the input features were first standardised to have a zero mean and a standard deviation of one.

The RR Model was a K-Nearest Neighbors (KNN) classifier with 100 neighbors considered, using only the R-R interval features.

The third and most comprehensive approach, the RR+AMP Model, was a random subspace ensemble model composed of discriminant analysis learners; this model used both the R-R interval and the ECG QRS complex amplitude features. This ensemble used a subspace size of 16 predictors and was trained for 30 learning cycles.

2.4. Evaluation

The evaluation protocol was multi-faceted. Initial model development and hyperparameter tuning were performed using a 5-fold cross-validation scheme on the primary training data. To assess generalization to unseen data, the trained models were then evaluated on the external CODE-15 dataset. Finally, performance was reported on a hidden test set as part of the official PhysioNet Challenge, providing an unbiased assessment of the final models. The statistical significance of each engineered feature was independently assessed using the Kruskal-Wallis test using the local training and evaluation sets. Models trained for the additional analysis on “normal” ECGs were only evaluated using cross-

validation.

3. Results

3.1. Model Performance

The performance of the three models is summarised in Table 1. The CNN model achieved a high cross-validation AUC of 0.9359 but performed poorly on local validation set of CODE-15 with an AUC of 0.3557, indicating significant overfitting. In contrast, the feature-based models generalised better. The RR+AMP model yielded the highest cross-validation AUC of 0.7220, demonstrating a modest improvement over the RR-only model which had a cross-validation AUC of 0.6827. When validated locally using the CODE-15 set, the RR+AMP model had an AUC of 0.6867 and the RR model an AUC of 0.6900, showing a consistent performance across different datasets.

Table 1. Results of each model on local testing.

Model	Training data cross-validation AUC	CODE-15 validation
CNN	0.9359	0.3557
RR	0.6827	0.6900
RR+AMP	0.7220	0.6867

For the testing on the hidden tsets, the CNN model was submitted during the main phase of the challenge, while the feature models were both submitted for the Hackathon during the conference. As a result, which was feature was eventually evaluated is unknown. The CNN model achieved a mean challenge score of 0.056 and the feature model a mean challenge score of 0.078.

Table 2. Results of each model on the hidden sets.

Model	Hidden validation challenge score	Hidden test mean challenge score
CNN	0.101	0.056
Feature	0.062	0.078

3.2. Analysis on Normal ECGs

The analysis restricted to clinically "normal" ECGs also identified highly significant features, suggesting that subtle autonomic dysregulation is detectable even without overt abnormalities. For SET A (strictly normal controls), SDNN($p=3.46\times10-8$) and Poincaré long diagonal st.d. ($p=3.85\times10-8$) were the most powerful discriminators.

A simple model trained on SET A achieved an AUC of 0.8119, with 80.6% specificity and 74.2% sensitivity, highlighting a strong potential for early detection. Performance decreased on SET B (AUC 0.6897) and SET C (AUC 0.6127), where control groups were less stringently defined or sourced from a different dataset, respectively.

Table 3. Normal ECG model evaluation

Dataset	Training data cross-validation AUC
SET A	0.8119
SET B	0.6897
SET C	0.6127

4 Discussion

This study demonstrates the potential of using machine learning models on features from short, single-lead ECGs to detect Chagas disease. Our findings indicate that feature-engineered models, particularly the RR+AMP model, provide a more robust and generalizable performance compared to our end-to-end CNN deep learning approach. Biases in the data and the lack of data cleansing, could be the cause for the poor performance of the CNN model on the hidden sets. For the feature models, the fact that they cannot process records with too few RR intervals, might have contributed to the model's low performance.

The most significant finding is the ability to discriminate between Chagas and healthy individuals even when their ECGs are classified as "normal". The high performance of a model on SET A (AUC 0.8119) suggests that subtle disruptions in cardiac autonomic control, captured by HRV metrics like SDNN and Poincaré plot analysis, precede the development of clinically apparent ECG abnormalities. This aligns with the known pathophysiology of Chagas disease, where autonomic nervous system damage is an early manifestation. This capability is crucial for early-stage screening, allowing for timely intervention before irreversible cardiac damage occurs. A limitation of this approach is that such a model might confuse other diseases that affect HRV for Chagas disease.

The primary limitation of this work is the use of short ECG recordings, which is suboptimal for stable HRV feature calculation. Standard HRV analysis typically requires recordings of at least 30 seconds. The instability of features derived from short segments may have constrained the overall model performance.

In conclusion, this research provides strong evidence

that accessible, single-lead ECG analysis can be a powerful tool for Chagas disease screening. By leveraging computational analysis of RR intervals and QRS amplitude, it is possible to detect subtle pathological signs that may be missed by routine interpretation. As technology for heart rhythm monitoring becomes ubiquitous through smartphones and wearables, this approach has the potential to significantly improve access to care and enable large-scale screening efforts in at-risk populations.

Future work should focus on validating these methods on longer recordings (e.g., 20-30 seconds) to enhance feature stability and further improve model accuracy. Additionally, an important investigation will be to compare the ability of these features to discriminate Chagas disease with no ECG findings from other diseases that may also affect the autonomic nervous system and thus HRV.

References

[1] Reyna MA, Koscova Z, Pavlus J, Weigle J, Saghafi S, Gomes P, Elola A, Hassannia, MS, Campbell, K, Bahrami Rad A, Riberio AH, Ribeiro ALP, Sameni R, Clifford GD. Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025; 52: 1-4.

[2] Reyna MA, Koscova Z, Pavlus J, Weigle J, Saghafi S, Gomes P, Elola A, Hassannia, MS, Campbell, K, Bahrami Rad A, Ribeiro AH, Ribeiro ALP, Sameni R, Clifford GD. Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025; 2025: URL: https://arxiv.org/abs/2510.02202, DOI: 10.48550/arXiv.2510.02202.

[3] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation; 101(23): e215-e220.

[4] Ribeiro ALP, Ribeiro AH, Paixao GMM, Lima EM, Horta Ribeiro M, Pinto Filho MM, et al. Sami-Trop: 12-lead ECG traces with age and mortality annotations. Zenodo; 2021.

[5] Wagner P, Strodthoff N, Bousseljot R, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3). PhysioNet. 2022. RRID:SCR_007345.

[6] Ribeiro AH, Paixao GMM, Lima EM, Horta Ribeiro M, Pinto Filho MM, Gomes PR, et al. CODE-15%: a large scale annotated dataset of 12-lead ECGs. Zenodo; 2021.

Address for correspondence:

Edo Ikurumi
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
ikrmi.edo@gmail.com