Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025

Matthew A Reyna[1], Zuzana Koscova[1], Jan Pavlus[1], James Weigle[1], Soheil Saghafi[1], Paulo Gomes[2], Andoni Elola[3], Mohammad Sina Hassannia[1], Kiersten Campbell[1], Ali Bahrami Rad[1], Antônio H Ribeiro[4], Antonio LP Ribeiro[2], Reza Sameni[1,5,*], Gari D Clifford[1,5,*]

[1]Department of Biomedical Informatics, Emory University, USA
[2]Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; Telehealth Center from Hospital das Clínicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[3] Department of Electronic Technology, University of the Basque Country UPV/EHU, Eibar, Gipuzkoa, Spain
[4] Department of Information Technology, Uppsala University, Uppsala, Sweden
[5]Department of Biomedical Engineering, Georgia Institute of Technology, USA
[*]Joint senior authors

Abstract

The George B. Moody PhysioNet Challenge 2025 invited teams to develop algorithmic approaches for identifying Chagas disease from electrocardiograms (ECGs).

Chagas disease is a parasitic infection that is endemic to South America and Central America and primarily transmitted by insects. Chronic Chagas disease can cause cardiovascular disease and digestive disorders. Serological testing capacities for diagnosing Chagas disease are limited, but the manifestation of Chagas cardiomyopathy in ECGs can support the prioritization of patients for confirmatory testing and treatment.

This Challenge provided multiple innovations. First, we leveraged several datasets with labels from patient reports and serological testing, providing a large dataset with weak labels and smaller datasets with strong labels. Second, we enriched the data to help teams generalize to unseen data sources. Third, we designed an evaluation metric that captured the local serological testing capacity to frame the machine learning problem as a triage task.

Over 630 participants from 111 teams submitted over 1300 entries during the Challenge, representing diverse approaches from academia and industry worldwide. Results on our hidden test data indicate strong performance, but poor generalization to a novel database with largely asymptomatic patients.

1. Introduction

The George B. Moody PhysioNet Challenges are annual competitions that support the development of open-source approaches to complex physiological and clinical problems [1]. For the PhysioNet Challenge 2025, we invited teams to develop algorithms for using electrocardiograms (ECGs) to identify cases of Chagas disease and to help prioritize potential Chagas disease patients for confirmatory diagnosis and treatment.

Chagas disease is a tropical parasitic disease that is caused by protozoan *Trypanosoma cruzi* and primarily transmitted by triatomine bugs. It is endemic to South America and Central America, affecting more than 8 million people worldwide with 30,000 to 40,000 annual infections and 10,000 to 14,000 annual deaths [2, 3]. Currently, there is no human vaccine for Chagas disease.

After an acute phase, which usually occurs in childhood, Chagas disease enters a life-long chronic phase [4, 5]. In the early stages of infection, Chagas disease has no or mild symptoms, and can be treated by specific drugs to prevent the progression of the disease. In the later stages of infection, Chagas disease can cause cardiomyopathy, leading to heart failure, cardiac arrhythmias, and thromboembolism, and is associated with a higher risk of death. Serological testing has shown a high prevalence of Chagas disease in some areas, and such tests can be used for diagnosis in individual patients, but serological testing capacities are limited. However, Chagas cardiomyopathy often manifests in electrocardiograms (ECG), providing a signal for Chagas disease that can support diagnosis and treatment.

2. Methods

We curated several datasets, devised and implemented a clinically relevant evaluation metric, and evaluated entries

 ISSN: 2325-887X DOI: 10.22489/CinC.2025.068

from 111 teams for the PhysioNet Challenge 2025 [6].

2.1. Challenge Data

The Challenge data included 12-lead ECG data and Chagas disease labels from multiple public and private sources:

- The **CODE-15%** dataset [7] contains 345779 ECG records from 233770 Brazilian patients with self-reported Chagas labels. Approximately 2% of patients reported positive cases of Chagas disease. The signals are approximately 10s in length, and the sampling frequency is 400Hz. This dataset is public and part of the Challenge training set.
- The **SaMi-Trop** dataset [8] contains 1631 ECG records from Brazilian patients with Chagas cardiomyopathy. All patients were serologically validated for Chagas disease. The signals are approximately 10s in length, and the sampling frequency is 400Hz. This dataset is public and part of the Challenge training set.
- The **PTB-XL** database from the Physikalisch-Technische Bundesanstalt (PTB), Brunswick, Germany [9] contains 21799 ECG records from 18869 European patients. All patients were assumed to be Chagas negative because Chagas disease is not endemic to Europe, but this assumption was not confirmed with serological testing and may not hold in every case. The signals are 10s in length, and the sampling frequency is 500Hz. This dataset is public and part of the Challenge training set.
- The **REDS-II** database [10] contains 1979 ECG records from 631 Brazilian patients with both positive and negative Chagas labels from serological testing. The original dataset was constructed so that the Chagas positive and negative cases were balanced, so we oversampled the Chagas-negative cases to approximate the population positivity rate. The signals are approximately 10s in length, and the sampling frequency is 300Hz. This dataset is private and part of the Challenge validation and test sets.
- The **SaMi-Trop 3** database contains 3855 ECG records from Brazilian patients with both positive and negative Chagas labels from serological testing. The original dataset was constructed so that the Chagas positive and negative cases were balanced, so we oversampled the Chagas-negative cases to approximate the population positivity rate. The signals are approximately 10s in length, and the sampling frequency varies. This dataset is private and part of the Challenge test set.
- The **ELSA-Brasil** database [11] contains 13739 ECG records from 13739 Brazilian patients with both positive and negative Chagas labels from serological testing. Approximately 2% of patients tested positive for Chagas disease. The signals are approximately 10s in length, and the sampling frequency is 300Hz. This dataset is private and part of the Challenge test set.

We reformatted the data in a WFDB-compatible format so that data from different sources shared a consistent format. We truncated zero-padded ECG signals to remove added zeros, and we removed empty signals. We replaced ages above 89 with a single age of 90 as needed to deidentify the data. The REDS-II dataset and the SaMi-Trop 3 were constructed to approximately balance the data with comparable numbers of positive and negative Chagas disease cases, so we oversampled the Chagas-negative records in these datasets to approximately match the prevalence rate of the ELSA-Brasil data, which has a 2.04% positivity rate. For both Chagas-positive and Chagas-negative records in these two datasets, we also added small amounts of various forms of noise, applied filters that are representative of different ECG devices, and resampled the data to different sampling frequencies to create new but highly similar records that support algorithmic generalizability. Please see [6] for details about how we augmented the data for the Challenge.

2.2. Challenge Objective

The PhysioNet Challenge 2025 invited teams to develop algorithms for using ECGs to identify cases of Chagas disease and to help prioritize potential Chagas disease patients for confirmatory diagnosis and treatment.

2.2.1. Challenge Timeline

This year's Challenge was the 26th George B. Moody PhysioNet Challenge [1]. As in previous years, the Challenge had an unofficial phase and an official phase.

The unofficial phase (9 January 2025 to 9 April 2025) introduced the teams to the Challenge. We publicly shared the Challenge objective, training data, example algorithms, and evaluation metric and invited the teams to submit their code for training and evaluation, scoring at most five entries from each team on the validation set. Between the unofficial and official phases, we took a hiatus (10 April 2025 to 28 May 2025) to improve the Challenge. The official phase (29 May 2025 to 20 August 2025) continued the Challenge. We updated the Challenge data, example algorithms, and evaluation metric and again invited teams to submit their code for training and evaluation, scoring at most ten entries from each team on the validation set.

After the end of the official phase, we attempted to evaluate at most one entry from each team on the test set to determine the winners of the Challenge.

We announced the results at the end of the Computing in Cardiology (CinC) 2025 conference, where the teams presented, defended, and published their work. Participation in CinC was a requirement for rankings and prize eligibility. We will publicly release the algorithms after the end of the Challenge and the publication of these papers.

2.2.2. Challenge Evaluation

The serological testing capacities for Chagas disease in Brazil and many other parts of South America and Central America are limited, so we evaluated the ability of algorithms to use ECGs to prioritize potential Chagas disease patients for serological testing. We defined the Challenge score as the fraction of patients with Chagas disease in a patient cohort that a team's algorithm prioritizes in the top 5% of the cohort, which is roughly the serological testing capacity in the region. The team with the highest Challenge score on the hidden test set won the Challenge.

Mathematically, let X be a dataset of n ECG records, and let p_x be the algorithm's estimated probability of a positive case of Chagas disease for an ECG record $x \in X$. Let $x_1 = \text{argmax}_{x \in X} p_x$ be the record with the highest probability, x_2 the record with the second-highest probability, etc. We define $X_\alpha = \{x_k \in X : k \leq \alpha n\} \subseteq X$ as the collection of records with the $k = \lfloor \alpha n \rfloor$ highest probabilities. The Challenge score is the true positive rate for X_α, i.e., the fraction of Chagas-positive cases in X_α out of the total number of Chagas-positive cases in X, for $\alpha = 0.05$.

If there are records $x, y \in X$ with equal probabilities $p_x = p_y$ such that $x \in X_\alpha$ and $y \notin X_\alpha$, then the Challenge score randomly breaks ties between the tied records so that $\alpha n - 1 \leq |X_\alpha| \leq \alpha n$ and returns the expected (mean) score over all permutations of the tied records.

While we also reported traditional evaluation metrics on the validation and test sets to provide additional context for the algorithms, the Challenge score specifically considers the use of algorithms to prioritize potential Chagas disease patients for confirmatory testing and treatment.

3. Challenge Results

During the Challenge, a total of 111 teams submitted 1317 algorithms, including 185 successful entries and 266 unsuccessful entries during the unofficial phase as well as 372 successful entries and 494 unsuccessful entries during the official phase. After the Challenge, we were able to score 65 teams on both validation and test sets; a total of 41 teams met all of the requirements to be ranked.

Figure 1 shows the scores on the validation and test sets. The median Challenge score decreased from 0.279 on the REDS-II data in the validation set to 0.275 (1.4% lower) on the REDS-II data in the test set, to 0.236 (15% lower) on the SaMi-Trop 3 data in the test set, and to 0.100 (64% lower) to the ELSA-Brasil data in the test set. Table 1 summarizes the performance of the three highest-ranked teams. The full team summaries, additional scores, and the full Challenge criteria for rankings are available in [6, 12].

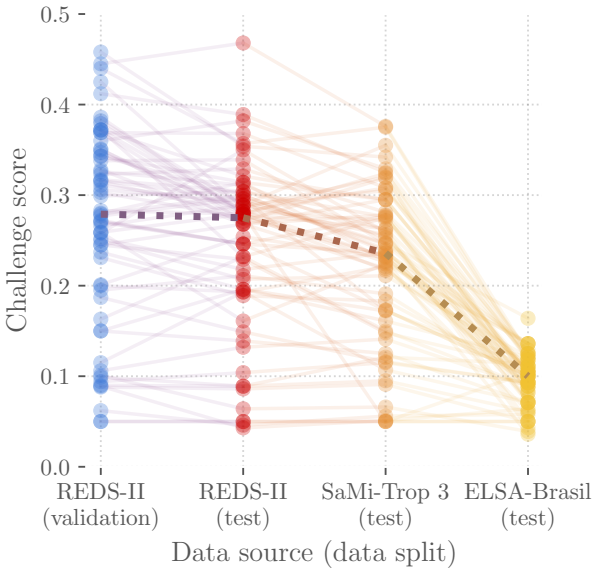

Figure 1: Challenge scores (y-axis) on the data sources (x-axis) in the validation and test sets. Each point is the score of a different team on a different dataset, and each thin solid line connects the scores for a team across datasets. The thick dotted lines show the median score of the teams across the datasets.

4. Discussion

The changes in the performance of the algorithms across the different data sources in the hidden validation and test sets reflect the difficulty of generalizing to unseen data.

The REDS-II and SaMi-Trop 3 datasets contained similar patient populations, including many symptomatic patients with Chagas cardiomyopathy; the small differences in performance between the REDS-II data in the validation and test sets and the SaMi-Trop 3 data in the test set are caused by sequential overfitting through repeated scoring on the validation set and the use of different ECG devices and clinical practices, respectively.

The ELSA-Brasil dataset contained largely asymptomatic patients without Chagas cardiomyopathy or known cardiovascular disease; the larger differences in performance between the other data sources and the ELSA-Brasil data are likely due to the different patient populations and the difficulty of identifying Chagas disease in ECGs without cardiovascular disease symptoms.

5. Conclusions

This article describes a large compendium of public and private 12-lead ECGs from several sources with both self-reported and serologically validated Chagas disease labels. The combination of standard 12-lead ECGs with a large

Rank	Team name	REDS-II (validation)	REDS-II (test)	SaMi-Trop 3 (test)	ELSA-Brasil (test)	Mean (test)
1	Biomed-Cardio	0.445	0.468	0.376	0.125	0.323
2	DlaskaLabMUI	0.440	0.357	0.375	0.118	0.283
3	AIChagas	0.360	0.382	0.329	0.129	0.280

Table 1: Challenge scores on the validation set, which contains data from the REDS-II dataset, and the test set, which contains data from the REDS-II dataset, the SaMi-Trop 3 dataset, and the ELSA-Brasil dataset, for the three highest-ranked Challenge teams according to the mean score across the three datasets in the test set.

database with weak labels and several smaller databases with strong labels poses a classical machine learning problem in a real-world clinical setting, and the use of three completely hidden data sources with varying patient populations, ECG devices, and other factors helps to support and assess model generalizability to unseen data.

The use of an evaluation metric that incorporates the confirmatory testing capacity for Chagas disease helps to support the development of clinically relevant machine learning models for Chagas disease detection as part of an effort to prioritize patients for limited resources [6].

Acknowledgements

This research was supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB: R01EB030362); the National Center for Advancing Translational Sciences of the National Institutes of Health (NCATS: UL1TR002378); as well as AliveCor and MathWorks under unrestricted gifts. AE received support by the MCIN/ AEI/10.13039/501100011033/, by FEDER Una manera de hacer Europa through grant PID2021-122727OB-I00, and by the Basque Government through grant IT1717-22. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the above entities.

References

[1] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[2] Cucunubá ZM, Gutiérrez-Romero SA, Ramírez J, Velásquez-Ortiz N, Ceccarelli S, Parra-Henao G, et al. The epidemiology of Chagas disease in the Americas. The Lancet Regional Health Americas 2024;37:100881.

[3] Nepomuceno de Andrade G, Bosch-Nicolau P, Nascimento BR, Martins-Melo FR, Perel P, Geissbühler Y, et al. Prevalence of Chagas disease among Latin American immigrants in non-endemic countries: an updated systematic review and meta-analysis. Lancet Regional Health Europe 2024;46:101040.

[4] Sabino EC, Nunes MCP, Blum J, Molina I, Ribeiro ALP. Cardiac involvement in Chagas disease and african trypanosomiasis. Nature Reviews Cardiology 2024;Epub ahead of print.

[5] Nunes MCP, Beaton A, Acquatella H, Bern C, Bolger AF, Echeverría LE, et al. Chagas cardiomyopathy: An update of current clinical knowledge and management: A scientific statement from the American Heart Association. Circulation 2018;138(12):e169–e209.

[6] Reyna MA, Koscova Z, Pavlus J, Saghafi S, Weigle J, Elola A, et al. Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025, 2025. URL https://arxiv.org/abs/2510.02202.

[7] Ribeiro A, Ribeiro M, Paixão G, Oliveira D, Gomes P, Canazart J, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications 2020;11(1):1760.

[8] Cardoso C, Sabino E, Oliveira C, de Oliveira L, Ferreira A, Cunha-Neto E, et al. Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile. BMJ Open 2016;6(5):e0011181.

[9] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, et al. PTB-XL, a large publicly available electrocardiography dataset. Scientific Data 2020;7:154.

[10] Nunes MCP, Buss LF, Silva JLP, Martins LNA, Oliveira CDL, Cardoso CS, et al. Incidence and predictors of progression to Chagas cardiomyopathy: Long-term follow-up of Trypanosoma cruzi–seropositive individuals. Circulation November 2021;144(19):1553–1566. ISSN 1524-4539.

[11] Pinto-Filho M, Brant L, Dos Reis R, Giatti L, Duncan B, Lotufo P, et al. Prognostic value of electrocardiographic abnormalities in adults from the Brazilian longitudinal study of adults' health. Heart 2021;107(19):1560–1566.

[12] George B. Moody PhysioNet Challenge 2025. https://physionetchallenges.org/2025/. Accessed: 2025-09-30.

Address for correspondence:
Matthew A Reyna
DBMI, 101 Woodruff Circle, 4th Floor East, Atlanta, GA 30322
matthew.a.reyna@emory.edu