Detection of Chagas Disease Using Tensor Decomposition and Wavelet Scattering Transform of ECG Signals

Shivnarayan Patidar

National Institute of Technology Goa, India

Abstract

Early detection of Chagas disease is critical for enabling timely medical intervention and preventing cardiovascular complications ahead. Electrocardiogram (ECG) signals offer crucial insights into the disease's progression, underscoring the need for advanced signal processing and machine learning-based diagnostic tools. This study proposes a novel framework for accurate and efficient identification of Chagas disease using standard 12-lead ECG signals. The methodology begins with signal preprocessing to ensure uniformity and reduce noise. Then, four categories of features are extracted: (a) statistical features from heart rate variability, (b) statistical features from all 12 ECG leads, and (c) wavelet scattering transform (WST) features from leads II, V1, and V3 and (d) patient metadata. These features are input to a RUSBoost classifier, designed to address class imbalance while performing binary classification. The proposed framework was evaluated on the PhysioNet/CinC Challenge 2025 dataset and achieved a mean Challenge score of 0.205 on the hidden test set, securing the 21st rank under the team name Medics.

1. Introduction

Chagas disease, caused by the protozoan parasite Trypanosoma cruzi, remains a significant public health concern in Central and South America. It is estimated that Chagas disease affects between 6 and 7 million people globally, causing nearly 10,000 deaths annually [1]. Often beginning asymptomatically, Chagas disease poses a significant challenge to clinicians, as early pathological changes may go undetected until life-threatening events occur [2, 3]. One of the most severe clinical manifestations of Chagas disease is chronic Chagas cardiomyopathy, which affects nearly 30% of infected individuals and is characterized by heart failure, life-threatening arrhythmias, thromboembolism, and sudden cardiac death [2–5]. Therefore, early and accurate detection is critical for timely intervention and improved prognosis.

Serological testing is essential for diagnosing Chagas disease but is limited by testing capacity. ECG, a widely accessible and cost-effective tool, can help screen patients and reduce the need for unnecessary serological tests. Combinations of common ECG abnormalities in Chagas disease—such as bundle branch blocks, atrioventricular (AV) blocks, and ventricular arrhythmias—can indicate early cardiac involvement and signal potential progression to cardiomyopathy [6]. Manual ECG interpretation is time-consuming, requires expertise, and is prone to variability—challenges that are critical in resource-limited, Chagas-endemic regions. Automated ECG analysis, powered by recent advances in signal processing, machine learning, offers a scalable and efficient alternative for accurate diagnosis and risk assessment[4].

Several studies have explored signal processing techniques along with AI and machine learning for early and accurate detection of Chagas disease using ECG or heart rate variability (HRV) data. Carl Jidling et al. developed a convolutional neural network (CNN) model trained on millions of ECGs, showing strong performance in detecting chronic Chagas cardiomyopathy, though less effective in early stages [7]. Vizcardo and Ravelo used Approximate Entropy (ApEn) from 24-hour HRV to differentiate disease stages [8]. María et al. achieved 98% accuracy using deep learning on ApEn features, enhanced by data augmentation to overcome small dataset limitations [5].

To address the above mentioned challenges, this work proposes a novel method aligned with the George B. Moody PhysioNet Challenge [9–11] on Chagas disease with large and diverse dataset—an initiative that promotes transparency, reproducibility, and cross-study comparison—ultimately advancing the clinical adoption of AI-driven ECG diagnostics for Chagas disease.

Rest of paper is organized as follows. Our methodology is in section 2. Through numerous experiments, results and evaluations are provided in section 3. Finally, section 4 presents conclusions and future directions.

2. Methodology

The proposed methodology involves four main steps: signal pre-processing, application of scattering transform on selected ECG leads, feature extraction, and binary classification. The overall system design is illustrated in Figure 1, with details of each step described below.

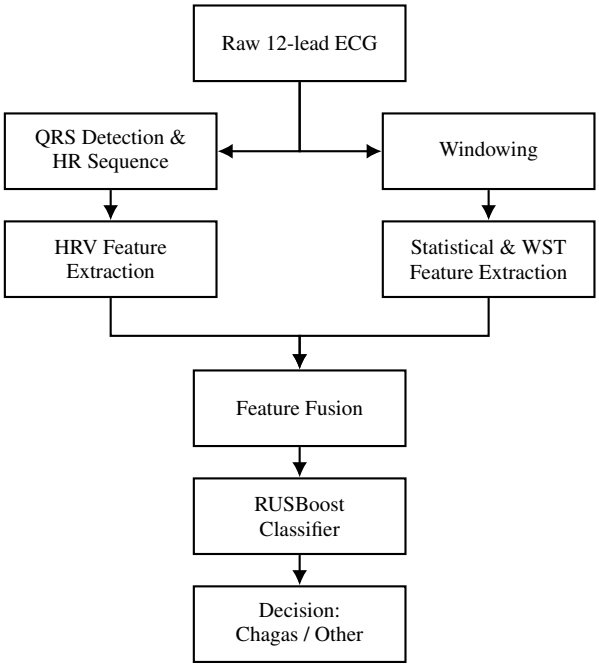

Figure 1: Proposed framework

2.1. Dataset Overview

The PhysioNet Challenge 2025 training dataset for Chagas disease combines three publicly available 12-lead ECG sources. The CODE-15 % dataset [12] provides over 300,000 weakly labeled ECGs from Brazil (2010–2016, 400 Hz, 7.3–10.2 s). The SaMi-Trop dataset [13] contributes 1,631 serologically confirmed Chagas-positive cases from Brazil (2011–2012, 400 Hz, 7.3–10.2 s). The PTB-XL dataset [14] adds 21,799 European ECGs (1989–1996, 500 Hz, $>$10 s), serving as presumed negatives. Training data are publicly available, while validation and test sets are withheld to ensure unbiased evaluation with region-specific Chagas prevalence.

2.2. Pre-processing

The proposed methodology utilizes standard 12-lead ECG signals, along with heart rate signals derived from either Lead II or V2. Heart rate signals are extracted from the raw ECG by first detecting R-peaks using the Pan and Tompkins algorithm applied to Lead II. If Lead II is affected by low voltage or noise, Lead V2 is used as an alternative. The RR intervals, representing the time between successive QRS complexes, are then calculated to generate RR-interval signals in seconds. These are converted into heart rate signals in beats per minute.

2.3. Wavelet Scattering Transform

The wavelet scattering transform (WST) combines wavelet transforms with nonlinear modulus and low-pass filtering operations to extract features that are robust to translation and small deformations [15].

The process begins with the original signal $s(t)$. By convolving it with a scaling function $\phi(u)$, one obtains the *zeroth-order* scattering coefficient:

$$ST_0 s = s(t) * \phi(u).$$

Next, the signal is filtered by a family of wavelets $\psi_\lambda(t)$, followed by a modulus operation to yield the *first-order* coefficients:

$$ST_1 s(u, \lambda_1) = |s(t) * \psi_{\lambda_1}(t)| * \phi(u).$$

Where, wavelet parameter λ encodes both scale and frequency. This procedure is iteratively repeated: at each stage, the modulus output is convolved with additional wavelets, and finally averaged by the scaling function. For example, the *second-order* coefficients are:

$$ST_2 s(u, \lambda_1, \lambda_2) = \big|\, |s(t) * \psi_{\lambda_1}(t)| * \psi_{\lambda_2}(t) \,\big| * \phi(u).$$

In general, the process produces a hierarchical sequence of scattering coefficients:

$$\{ST_0, ST_1, ST_2, \ldots, ST_n\},$$

which together form the scattering representation of the signal.

2.4. Feature Extraction

Four feature sets were computed: (a) statistical features from all 12 ECG leads, (b) WST features from leads II, V1, and V3 (4 s window), (c) HRV-based statistics, and (d) patient metadata (age, gender). A total of 30 statistical features captured temporal, amplitude, and morphological aspects of ECG and HR sequences, including measures of central tendency, dispersion, distribution shape, amplitude, and energy. Also, Shannon wavelet entropy, log energy entropy and Hjorth parameters (activity, mobility, complexity) were computed to quantify signal complexity.

2.5. Chagas Detection Model: RUSBoost

In this study, Random Under-Sampling Boosting (RUSBoost) is used as a classification method. Random undersampling (RUS) addresses the issue of underlying imbalanced datasets of chagas by randomly removing samples from the majority class, thereby creating a balanced dataset for training. Boosting, on the other hand, is an ensemble technique that builds several weak learners and aggregates their predictions through majority voting to produce a final decision. RUSBoost integrates RUS with boosting in AdaBoost framework [16], as an ensemble method, generally achieve better performance than individual models in such scenarios. Additionally, RUSBoost eliminates the need for manual threshold adjustment, as it relies on the collective majority vote of the weak learners.

3. Results and Discussion

The proposed framework was carried out in MATLAB R2024b version. Feature-level empirical analysis revealed that the WST features, particularly from leads II, V1 and V3, are clinically relevant with 4-second windowing. All possible combinations of filter bank quality factors for a wavelet scattering network comprising up to three filter banks, employing Gabor wavelet as the mother wavelet are systematically evaluated. Among these configurations, the setting [32,8,4] yielded the highest macro F1-score. Notably, instead of the conventional mean aggregation, a max operation was applied to the scattering wavelet transform coefficients of the selected ECG leads, which helped emphasize peak values and preserve salient morphological details that may otherwise be attenuated through averaging. In all 2682 features are obtained comprising 2288 from WST, 390 statistical features from ECGs and HRV sequence, gender and age, sign and threshold information from QRS detection. Figure 2 illustrates the WST coefficients for both normal and Chagas cases. The use of an optimal RUSBoost with maximum split of 100, a minimum leaf size of 10, and an ensemble of 297 trees effectively mitigated class imbalance. Table 1 and 2 summarize the results obtained under various experimental configurations on the training and testing datasets, respectively. The effect of tensor decomposition is also investigated. The feature vector was reshaped into a 53×10×5 tensor, and Tucker decomposition with optimal ranks 53×8×4 was applied. After including the previously excluded features, the total feature count was 1698, and performance remained nearly unchanged with tensor representation.

Table 1: Test results on validation and hidden datasets.

Dataset	Challenge Score	AUROC	AUPRC	F1 Chagas	Accuracy
(a) Validation data	26.6	67.2	12.0	14.0	92.1
(b) REDS-II test data	28.9	73.7	24.9	16.1	92.9
(c) SaMi-Trop test data	25.6	69.8	9.0	16.1	80.8
(d) ELSA Brazil test data	7.1	47.5	2.0	0	97.9
Avg. test results	20.5	63.67	11.96	10.73	90.53

Table 2: Challenge scores for our selected entry.

Training	Validation	Test	Ranking
44.12	26.6	20.5	21

4. Conclusion

This work proposes a novel Chagas detection framework combining WST and RUSBoost. WST applied on three selected ECG leads, recursively uses wavelet transforms with nonlinear modulus and averaging to extract stable features, while RUSBoost addresses class imbalance for reliable binary classification. Future work may explore more data, WST parameter tuning, tuning tensor decomposition, and additional feature sets or classifiers.

Acknowledgments

The author would like to thank MathWorks for providing a free academic license of MATLAB R2024b.

References

[1] Cucunubá ZM, Gutiérrez-Romero SA, Ramírez JD, Velásquez-Ortiz N, Ceccarelli S, Parra-Henao G, Henao-Martínez AF, Rabinovich J, Basáñez MG, Nouvellet P, Abad-Franch F. The epidemiology of Chagas disease in the Americas. The Lancet Regional Health Americas 2024; 37:100881.

[2] Rassi A, Rassi A, Marin-Neto JA. Chagas disease. Lancet 2010;375(9723):1388–1402.

[3] Nunes MCP, Beaton A, Acquatella H, Bern C, Bolger AF, Echeverría LE, Dutra WO, Gascon J, Morillo CA, Oliveira-Filho J, Ribeiro ALP, Marin-Neto JA, On behalf of the American Heart Association Rheumatic Fever E, of the Council on Cardiovascular Disease in the Young; Council on Cardiovascular KDC, Nursing; S, Council S. Chagas Cardiomyopathy: An Update of Current Clinical Knowledge and Management: A Scientific Statement From the American Heart Association. Circulation 2018; 138(12):e169–e209.

[4] Silva LEV, Moreira HT, Bernardo MMM, Schmidt A, Romano MMD, Salgado HC, Fazan R, Tinós R, Marin-Neto JA. Prediction of echocardiographic parameters in Chagas disease using heart rate variability and machine learn-

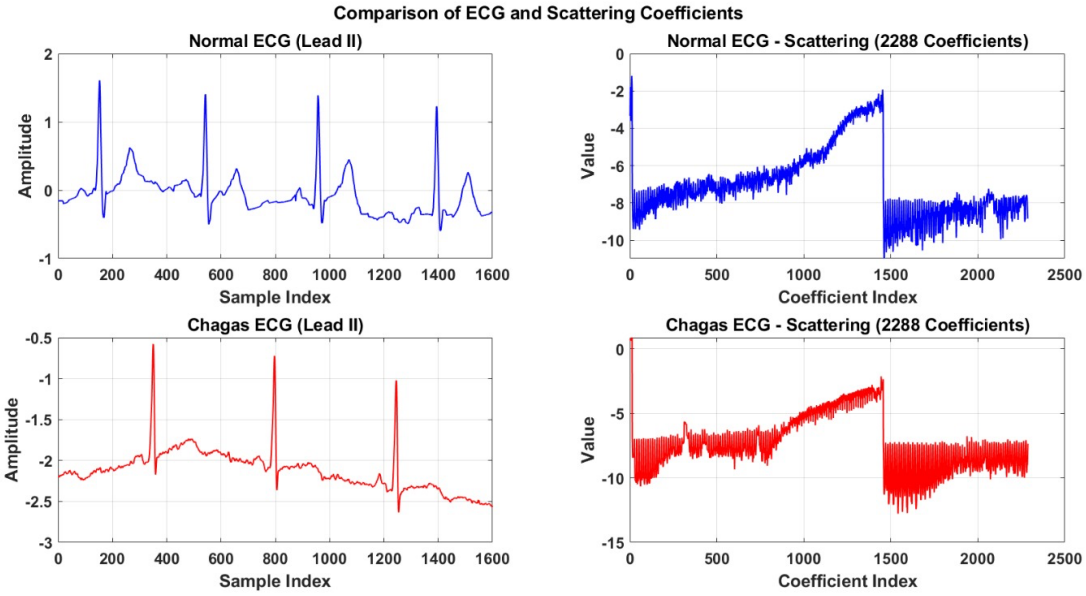

Figure 2: Comparison of normal and Chagas ECG signals (Lead II) and their corresponding scattering coefficients (2288).

ing. Biomedical Signal Processing and Control 2021; 67:102513.

[5] Rodriguez MF, Cornejo DR, Díaz LA, Ravelo-García A, Alvarez E, Cabrera-Caso V, Condori-Merma D, Cornejo MV. Optimal artificial neural network for the diagnosis of Chagas disease using approximate entropy and data augmentation. In Computing in Cardiology, volume 50. 2023; 1–4.

[6] Ribeiro ALP, Nunes MdCP, Teixeira MM, Rocha MdOC. Diagnosis and management of chagas disease and cardiomyopathy. Nature Reviews Cardiology 2012;9(10):576–589.

[7] Jidling C, Gedon D, Schön TB, Oliveira CDL, Cardoso CS, Ferreira AM, Giatti L, Barreto SM, Sabino EC, Ribeiro ALP, Ribeiro AH. Screening for Chagas disease from the electrocardiogram using a deep neural network. PLoS Neglected Tropical Diseases 2023;17(7):e0011118.

[8] Vizcardo M, Ravelo A. Use of approximation entropy for stratification of risk in patients with Chagas disease. In Computing in Cardiology, volume 45. 2018; 1–4.

[9] 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 2000;101(23):e215–e220.

[10] Reyna MA, Koscova Z, Pavlus J, Weigle J, Saghafi S, Gomes P, Elola A, Hassannia MS, Campbell K, Bahrami Rad A, et al. Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025. In Computing in Cardiology, volume 52. 2025; 1–4.

[11] Reyna MA, Koscova Z, Pavlus J, Saghafi S, Weigle J, Elola A, Seyedi S, Campbell K, Li Q, Bahrami Rad A, Ribeiro A, 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.

[12] Ribeiro AH, Ribeiro MH, Paixão GM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MP, Andersson CR, Macfarlane PW, Meira Jr W, Schön TB, Ribeiro AL. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications 2020;11(1):1760.

[13] Cardoso CS, Sabino EC, Oliveira CD, de Oliveira LC, Ferreira AM, Cunha-Neto E, Bierrenbach AL, Ferreira JE, Haikal DS, Reingold AL, Ribeiro AL. Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile. BMJ Open 2016; 6(5):e011181.

[14] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. Scientific Data 2020; 7(1):154.

[15] Mallat S. Group invariant scattering. Communications on Pure and Applied Mathematics 2011;65(10):1331–1398.

[16] Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. RUSBoost: A hybrid approach to alleviating class imbalance. IEEE Transactions on Systems Man and Cybernetics Part A Systems and Humans 2010;40(1):185–197.

Address for correspondence:

Shivnarayan Patidar
National Institute of Technology Goa, Goa, India
shivnarayan.patidar@nitgoa.ac.in