Using Sparse Gaussian Process Regression to Detect Chagas Disease from a 12-Lead Electrocardiogram

Simoni Mishra[1], Sanjeev Bhavnani[2], Vinay M Pai[3], Jun Beom Park[4], John Metzcar[5]

[1]Boston University, Boston, USA
[2]Scripps Clinic, San Diego, USA
[3]Health Tequity, LLC, Potomac, USA
[4]Kaiser Permanente, Sacramento, USA
[5]University of Minnesota, Bloomington, USA

Abstract

Chagas disease affects an estimated 6-7 million people and can lead to electrical conduction defects, arrhythmias, and cardiomyopathy. We proposed the use of a Gaussian Process Regression (GPR)-based algorithm for detecting Chagas disease from real-world clinical electrocardiograms (ECGs), as an entry to the George B. Moody PhysioNet Challenge 2025 (Team: ThreeRareMen). Raw ECG data were obtained from PTB-XL and SaMi-Trop datasets; the composite kernel based GPR models were developed and trained using R-wave detection and normalization. Features included morphological characteristics, heart rate variability, and frequency-domain band powers. Model was trained on features extracted from single-beat ECG segments and evaluated on a held-out subset of the training set. The model performance was poor, yielding an accuracy of 7%, area under the receiver operating characteristic of 0.143, and area under precision-recall curve of 0.055. (Note that these performance metrics were obtained on our internal test set; they should not be used for comparison against other Challenge teams methods). No single feature provided a strong separation between classes, suggesting the need for more sophisticated features with this model. Inspection of misclassified cases showed that many false negatives occurred in records with subtle ECG abnormalities, while false positives were often associated with noisy or atypical signals.

1. Introduction

Chagas disease, caused by *Trypanosoma cruzi*, affects an estimated 6-7 million people – mostly in Central and South America- and can lead to conduction defects, arrhythmias, and cardiomyopathy. Confirmatory diagnosis is limited in many settings, so fast electrocardiogram (ECG)-based screening could help prioritize patients for testing and treatment [1].

The objective of this effort is to design, train, and evaluate Gaussian process regression (GPR)-based algorithms for detecting Chagas disease from standard clinical ECGs. We aim for high discrimination and clinically useful sensitivity at fixed specificity, with robustness to class imbalance, lead configurations, and site/domain shifts. We leverage GPR's calibrated uncertainty to support clinician trust and provide feature- and lead-level explanations. All models and code will be fully reproducible to enable deployment as a triage tool for prioritizing confirmatory serological testing and timely care. We develop and evaluate GPR methods to detect Chagas disease from standard clinical ECGs in the George B. Moody PhysioNet / Computing-in-Cardiology Challenge 2025 [2–4]. The work covers ECG preprocessing (WFDB parsing, denoising, R-peak detection, beat/window segmentation), feature extraction at beat and record levels (morphology and heart rate variability (HRV) descriptors, with optional dimensionality reduction), and GPR models with radial basis function (RBF) kernels for calibrated uncertainty and interpretability. We benchmark our approach against simple baselines and report the area under the receiver operating characteristic (AUROC) curve, area under the precision-recall curve (AUPRC), and sensitivity at fixed specificity. We also assess calibration and robustness to class imbalance, reduced-lead configurations, and cross-site shifts. All experiments follow subject-wise data splits.

We chose GPR because it fits the clinical and data constraints of ECG-based Chagas screening: it performs strongly in small–to–moderate datasets, provides well-calibrated uncertainty per prediction for safe triage, and offers interpretability that highlights which ECG/HRV features and leads matter [5]. With flexible kernels (RBF + noise), GPR naturally models smooth ECG morphology and variability while remaining robust to site and lead differences; its probabilistic output can be thresholded at fixed specificity and abstain when uncertainty is high [6].

2. Methods

We built an analytical pipeline to detect Chagas disease from standard ECGs. The pipeline steps included: reading and cleaning the signals, finding R-peaks, cutting fixed-length beats, extracting features, training a GPR with an automatic relevance determination kernel, and combining beat-level outputs into a record-level score. We report accuracy metrics and check that results hold across different leads and sites.

2.1. Assumptions

We assumed that the labels reflect true status and are time-aligned with ECG acquisition and that the ECG morphology/HRV contain sufficient signal for Chagas discrimination. Additionally, it is assumed that the metadata (sampling rate, lead order) are accurate, signal quality is adequate after standard denoising, and R-peak detection yields consistent segments. Train/validation/test splits were assumed to be subject-wise and independent; evaluation cohorts were broadly similar to development data, or remaining shift was addressed in robustness analyses. Missing/variant leads were handled via predefined strategies (drop / impute / reduced-lead models) without systematic bias. Class priors at evaluation were not radically different from development, or any differences were mitigated via threshold tuning.

2.2. Software

Signal processing and modeling used NeuroKit2 0.2.12 (ECG cleaning, R-peak detection), WFDB 4.3.0 (PhysioNet I/O); models were saved with joblib 1.5.1. Visualization/workflow used tqdm 4.67.1.

2.3. Data Processing

Raw ECG data were obtained from the PhysioNet Challenge datasets, which included the PTB-XL[7], Sami-Trop [8], and CODE-15% [9] datasets. For this manuscript, only the PTB-XL and SaMi-Trop datasets were used. The data processing pipeline consisted of several steps to prepare the signals for machine learning analysis:

2.3.1. Beat Extraction

Each ECG file was cleaned, and R-peaks were detected using either NeuroKit2 or a custom bandpass filter and peak detection algorithm. Around each detected R-peak, a fixed window (typically 0.25 seconds before and 0.45 seconds after was extracted and resampled to a standard length. The first clean, artifact-free beat following preprocessing was selected as single representative beat for each ECG. Use of a single beat simplified kernel-based modeling and ensured consistent feature extraction across samples

2.3.2. Feature Extraction

The resulting beat was z-score normalized and used to compute morphological, HRV, and frequency-domain features. Morphological features included R-peak amplitude, QRS duration, and waveform symmetry metrics. Frequency-domain features were derived from short-time Fourier transforms to capture band power across relevant frequency ranges. Feature extraction was performed using a combination of custom signal processing routines and the NeuroKit2 library, ensuring the inclusion of clinically relevant and statistically informative attributes.

2.3.3. Labeling and Dataset Preparation

Each beat was assigned a Chagas label based on the metadata accompanying its source record. The final dataset consisted of extracted feature vectors and corresponding labels for model training and evaluation.

2.4. Model Architecture

The primary objective of this study was to develop a robust machine learning model for the detection of Chagas disease from ECG signals. The model architecture was designed to leverage both domain-specific feature extraction and probabilistic classification.

2.4.1. GPR Model

The extracted features served as input to a GPR model. GPR is a non-parametric, Bayesian approach to regression and classification that models the underlying data distribution and provides uncertainty estimates for predictions. During development, several kernel combinations – including constant, radial basis function (RBF), Matern, and white noise kernels – were evaluated for model performance. The final model utilized a composite kernel, consisting of a constant kernel multiplied by a radial basis function (RBF) kernel, with an added white noise kernel. Hyperparameters for the kernel were optimized using multiple restarts to ensure robust convergence.

2.4.2. Calibration and Threshold Optimization

To improve the reliability of probabilistic outputs, the model incorporated a calibration step using Platt scaling or isotonic regression. The final binary classification was determined by optimizing the decision threshold to maximize the F1 score or the official Challenge score metric.

2.4.3. Training and Evaluation

The dataset was split into training and testing subsets to prevent overfitting and to provide an unbiased estimate of model performance. Model training was performed on the extracted features, and evaluation metrics—including AUROC, AUPRC, accuracy, F-measure, and the Challenge score — were computed using the official evaluation script provided by the Challenge organizers.

3. Results

The proposed pipeline was evaluated using the official PhysioNet Challenge scoring script. The model was trained on features extracted from single-beat ECG segments and evaluated on a held-out test set. Performance metrics area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), accuracy, and F-measure.

After training and calibration, the model achieved the following results on the test set (Table 1):

Table 1. Performance of GPR-based Chagas detector after training and probability calibration

Metric	Score
AUROC	0.143
AUPRC	0.055
Accuracy	0.070
F-measure	0.130

To assess classification performance, we examined the confusion matrix (Figure 1). The overall accuracy was 7%. The ROC curve (Figure 2) demonstrates the model's inability to distinguish between Chagas disease and control cases across various thresholds, with an AUROC of 0.143. The precision-recall curve (Figure 3) provides more evidence of the model's difficulty to identify positive cases, particularly in the context of class imbalance, with an AUPRC of 0.055. Note that these performance metrics were obtained on our internal test set and are not official Challenge scores; therefore, they cannot be used for comparison against other Challenge teams methods.

4. Discussion and Conclusions

The results indicate that, while the model was able to learn some discriminative patterns from the ECG features, overall performance was poor. The AUROC and AUPRC values suggest limited ability to distinguish between

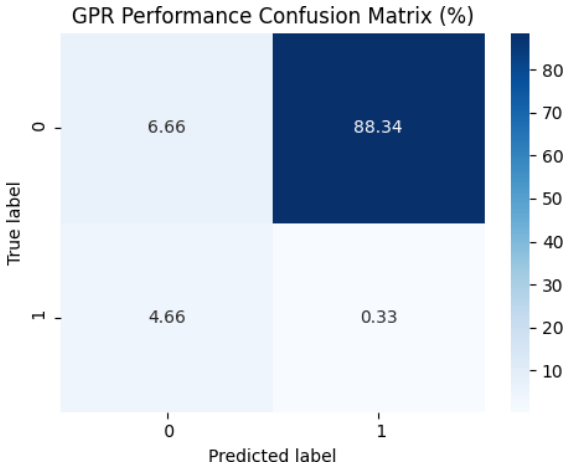

Figure 1. The confusion matrix shows the percentage of true positives, true negatives, false positives, and false negatives on the test set.

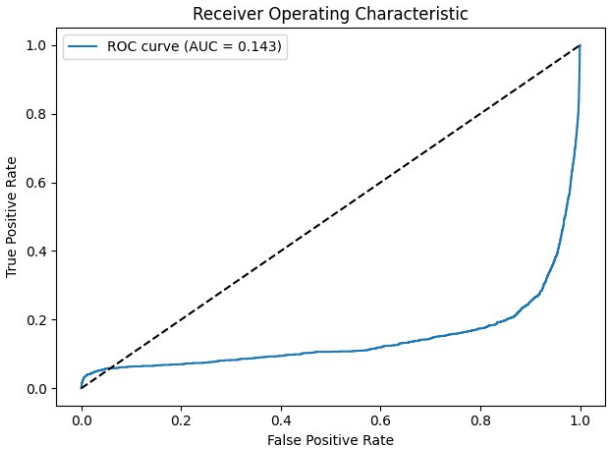

Figure 2. Receiver operating characteristic (ROC) curve for the GPR model on the test set.

Chagas disease and control cases, and the Challenge score reflects the difficulty of the task on this dataset.

To further understand model performance, confusion matrices and ROC curves were generated. The confusion matrix revealed a tendency of the model to favor the majority class, resulting in low sensitivity and Chagas disease detection. ROC and precision-recall curves confirmed the low discriminative power, with curves remaining close to the diagonal and baseline, respectively. Analysis of feature importance indicated that heart rate variability metrics and QRS morphology contributed most to the model's predictions. However, no single feature provided strong separation between classes, suggesting that additional or more sophisticated features may be more

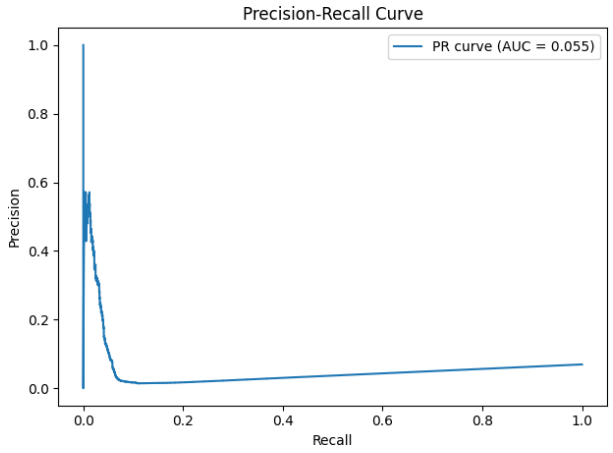

Figure 3. Precision-recall curve for the GPR model on the test set.

necessary. Inspection of misclassified cases showed that many false negatives occurred in records with subtle ECG abnormalities, while false positives were often associated with noisy or atypical signals. This highlights the challenge of detecting Chagas disease from ECG data alone and suggests that further improvements may require more advanced signal processing or integration of clinical metadata. Overall, the results demonstrate the feasibility of using GPR for automated Chagas disease detection from ECG signals but also underscore the need for further methodological enhancements.

Future work will focus on expanding the feature set, optimizing model parameters, and exploring ensemble approaches to improve classification performance. We plan an additional analysis that incorporates an AI-based assessment of a 12-lead ECG for the detection of cardiomyopathy with low ejection fraction (EF <40%) by using a regulated, commercially available CNN algorithm [10]**.** Preliminary data has shown the feasibility of this CNN algorithm in a cross-sectional study within the SaMi-Trop cohort to detect Chagas cardiomyopathy on a 12 lead ECG with an accuracy of 83% [11]**.** By enriching for probable Chagas cardiomyopathy, we can determine how a higher pretest probability affects GPR discrimination, calibration, and positive predictive value / negative predictive value at fixed specificity and explore a simple combined GPR + low EF model. This analysis will indicate whether low EF- based triage improves screening yield without inflating false positives.

We anticipate that additional data and data-curation may help improve predictive ability of such models. For example, longitudinal ECG data from each patient before and after seropositivity. If available, following changes in ECG readouts during disease progression of patients receiving delayed treatment versus those who are in active treatment may allow detection of subtle changes over time. This could provide insight into occult signs of disease progression that precedes symptom manifestation (from conduction abnormalities and myocardial fibrosis occurring at much later stage). We must also be aware of factors that influence the training set, including ECG dataset obtained from various location (different endemic areas with varying disease prevalence), varying degrees of disease manifestation at the time of initial medical evaluation due to difference in access to care, and inaccuracies introduced from self-reported Chagas disease positive population.

References

[1] Lidani KCF, Andrade FA, Bavia L, et al. Chagas Disease: From Discovery to a Worldwide Health Problem. *Front Public Health* 2019; 7: 166.

[2] Goldberger AL, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*; 101. Epub ahead of print 13 June 2000. DOI: 10.1161/01.CIR.101.23.e215.

[3] Matthew A. Reyna, Zuzana Koscova, Jan Pavlus, et al. Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025. DOI: 10.48550/arXiv.2510.02202.

[4] 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.

[5] Schulz E, Speekenbrink M, Krause A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *J Math Psychol* 2018; 85: 1–16.

[6] Cheng L-F, Dumitrascu B, Darnell G, et al. Sparse multi-output Gaussian processes for online medical time series prediction. *BMC Med Inform Decis Mak* 2020; 20: 152.

[7] Wagner P, Strodthoff N, Bousseljot R-D, et al. PTB-XL, a large publicly available electrocardiography dataset. *Sci Data* 2020; 7: 154.

[8] Cardoso CS, Sabino EC, Oliveira CDL, et al. Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile. *BMJ Open* 2016; 6: e011181.

[9] Ribeiro AH, Ribeiro MH, Paixão GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun* 2020; 11: 1760.

[10] Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nat Med* 2019; 25: 70–74.

[11] Brito BODF, Attia ZI, Martins LNA, et al. Left ventricular systolic dysfunction predicted by artificial intelligence using the electrocardiogram in Chagas disease patients–The SaMi-Trop cohort. *PLoS Negl Trop Dis* 2021; 15: e0009974.

Address for correspondence:
Vinay M. Pai.
2930 Domingo Ave #1524, Berkeley, CA 94705.
vinay@healthtequity.net