Predicting Chagas Disease from ECGs Using Simulator-Augmented DNN-Derived Abnormality Scores

Yuta Hashimoto[1,2], Naoki Nonaka[2,3], Jun Seita[2,3]

[1]Department of Emergency and Critical Care Medicine, Saitama Medical University, Saitama, Japan
[2]Predictive Medicine Special Project, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan
[3]RIKEN Center for Interdisciplinary Theoretical and Mathematical Sciences, Tokyo, Japan

Abstract

This study proposes a two-stage framework for improving the detection of Chagas disease under conditions of scarce and imbalanced data. In Stage1, a ResNet18-based binary classification model was trained to estimate abnormality scores from 12-lead electrocardiograms, with simulator-generated synthetic ECGs incorporated to supplement underrepresented classes. In Stage2, these abnormality scores were combined with age and sex to form tabular features, which were then classified using a random forest and XGBoost model. The framework demonstrated improved detection performance through lead optimization and synthetic data augmentation. These results highlight the potential of integrating deep learning with simulation-based approaches to enhance Chagas disease screening in resource-limited settings.

This article is part of Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025[1] [2] [3]. The team name was RISMU. The team could not be scored on the hidden test data within the allotted test time and therefore was not ranked.

(Abbreviations: ECG, electrocardiogram; DNN, deep neural network; AUROC, area under the receiver operating characteristic.)

1. Introduction

Serological testing remains the gold standard for diagnosing Chagas disease; however, access is often limited in resource-constrained regions [4]. Hence, supplementary methods that efficiently identify high-priority individuals are needed. Electrocardiography (ECG), being non-invasive and low-cost, represents a valuable alternative that reflects conduction abnormalities and arrhythmias frequently observed in Chagas disease, which can be leveraged for automated risk estimation .

With advances in deep learning, the accuracy of automated ECG analysis has markedly improved. Deep neural networks (DNNs) trained in a supervised manner can achieve high precision in classifying diverse ECG abnormalities when provided with large-scale labeled datasets. Nevertheless, their performance critically depends on the availability of high-quality labeled data, the creation of which requires expertise, cost, and effort. As a result, rare diseases and distinctive waveform patterns are often underrepresented, potentially limiting model generalizability, especially in the case of disease-specific patterns such as those found in Chagas disease.

To address this challenge, synthetic data generation using ECG simulators based on known abnormal waveform morphologies has attracted growing attention. Combining synthetic and real data has been shown to enhance performance in tasks such as arrhythmia detection and myocardial infarction and is particularly effective for rare or underrepresented waveform patterns [5].

In this study, we propose a two-stage framework to improve the detection performance of Chagas disease in environments characterized by scarce and imbalanced data. In the first stage (Stage 1), a binary classification model estimates abnormal ECG scores associated with Chagas disease. In the second stage (Stage 2), these scores are combined with demographic features such as age and sex to predict disease presence. Furthermore, for abnormalities insufficiently represented in the real datasets, synthetic ECGs generated by a simulator are incorporated as supplementary training data.

The aim of this study is to enhance the detection accuracy of Chagas disease under limited data conditions through a two-stage approach that integrates deep learning with simulator-based synthetic data generation.

2. Methods

The proposed framework consisted of two stages: Stage1 for estimating ECG abnormality scores and Stage2 for binary classification of Chagas disease.

In Stage1, we constructed a binary classification model to quantitatively evaluate ECG abnormalities associated with Chagas disease. A ResNet18[6] architecture was employed, with 12-lead ECGs as input, and preprocessing was applied for each targeted abnormality class. A Sigmoid activation

function was placed in the output layer to compute scores corresponding to the abnormality labels. No handcrafted features were designed; instead, raw ECG waveforms were directly used as input. The resulting output scores were stored as pseudo-labels and subsequently used as input features for Stage2.

To address the severe class imbalance in real-world data, synthetic ECGs were generated for abnormalities underrepresented in Chagas disease. These synthetic signals were produced using a simulator based on Gaussian distributions and physiological parameters, following the method reported by Nonaka et al [5]. The generated synthetic data were incorporated as supplementary training data in Stage1.

In Stage2, a binary classification task was performed to predict the presence or absence of Chagas disease. Raw ECG waveforms were not directly utilized; instead, the pseudo-labels derived from Stage1 were combined with demographic features. Specifically, the features consisted of the Stage1 output scores, age (scaled by 100), and sex (encoded as male = 1, female = 0), resulting in a feature vector of three dimensions plus the number of abnormal ECG classes considered. These features were used to train a Random Forest classifier, which estimated the probability of Chagas disease in a binary setting. An overview of the pipeline is shown in Figure 1.

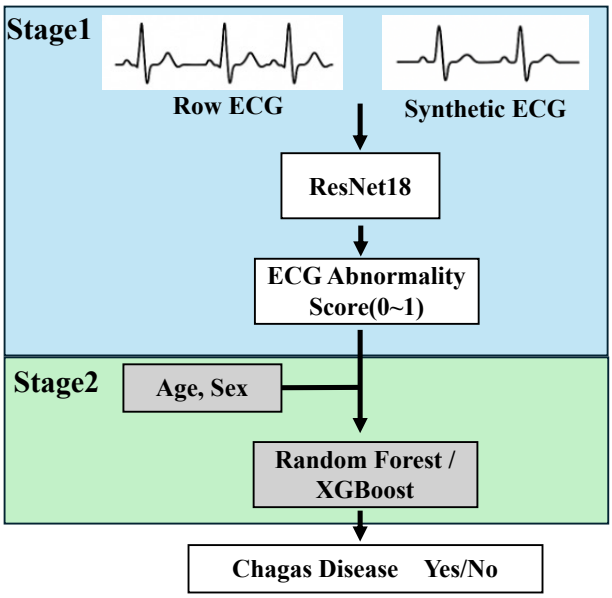

Figure 1. The framework of our proposed method

3. Tables and figures

3.1. Datasets

This study utilized publicly available real-world ECG datasets (PTB-XL [7] and SaMi-Trop [8]) in combination with synthetic ECG data generated by a simulator. PTB-XL was employed in both Stage1 and Stage2. In Stage1, it was used for training and evaluation of abnormal ECG classification, targeting seven labels: first-degree atrioventricular block (1AVB), incomplete right bundle branch block (IRBBB), complete left bundle branch block (CLBBB), long QT interval (LNGQT), acute myocardial infarction (AMI), atrial fibrillation (AFIB), and left posterior fascicular block (LPR). In Stage2, all PTB-XL records were assigned to the negative class based on demographic information, whereas SaMi-Trop served as the positive class to construct a binary classification task for Chagas disease.

For abnormalities that were present but insufficiently represented in the real datasets—namely second-degree atrioventricular block (2AVB; Wenckebach type/type II) and long QT interval (LNGQT)—synthetic ECGs were generated using a simulator capable of reproducing known waveform morphologies, thereby supplementing the limited real data.

3.2. Preprocessing

A unified preprocessing pipeline was applied to all datasets. First, the sampling frequency of each ECG signal was standardized by resampling through integer averaging or linear interpolation, ensuring a uniform target frequency of 500 Hz. When the waveform length was insufficient, zero-padding was applied to standardize the input dimensions. Subsequently, normalization was performed by applying z-score standardization (mean = 0, variance = 1) to each ECG signal, thereby correcting for differences in amplitude scaling across recordings.

3.3. Training setup and hyperparameters

The training was conducted on a Linux system equipped with four NVIDIA RTX A5000 GPUs (24GB each), CUDA 11.6, and Driver 510.47.03.

For Stage1, a ResNet18-based multi-label classification model was employed. The input consisted of 12-lead ECGs (500 Hz, fixed length of 5000 time step), and the output layer applied a sigmoid activation function to produce scores for seven distinct abnormality labels. The Adam optimizer was used, with the binary cross-entropy loss function. The initial learning rate was set to 0.0005, combined with a cosine annealing scheduler with warm-up. Data augmentation was performed by applying both temporal shift ratio and signal masking ratio. Training was performed for 500 epochs with a batch size of 256, and early stopping was implemented with a patience of 5, evaluated every five epochs. The learning rate and augmentation parameters were explored in preliminary experiments, and the optimal values were adopted for subsequent training.

For Stage2, the pseudo-labels obtained from Stage1 were

combined with demographic features to construct the classification task. Specifically, the input consisted of the pseudo-label scores, age, and sex (male = 1, female = 0), resulting in a three-dimensional tabular feature vector. A Random Forest classifier was employed, with the number of decision trees set to 10, 20, or 50, the maximum tree depth set to 5, 10, 20, or unrestricted, and class weights set either with balancing or without weighting. These hyperparameter combinations were explored, and model performance was evaluated using five-fold cross-validation with the area under the receiver operating characteristic curve (ROC-AUC) as the primary metric.

3.4. Experimental conditions

Data split

For the PTB-XL dataset, the division was based on the officially provided *strat_fold* information. Specifically, Folds 1–6 were used as training data for Stage1, Folds 7–8 as validation data for Stage01, and Folds 9–10 as evaluation data for Stage2. For the SaMi-Trop dataset, all cases were randomly divided into training, validation, and test sets for use in Stage2.

Label

In Stage1, binary classification was performed for each of the seven abnormalities in PTB-XL, namely 1AVB, IRBBB, CLBBB, LNGQT, AMI, AFIB, and LPR, based on the diagnostic labels provided in the dataset.

In Stage2, a binary classification task for Chagas disease was constructed by assigning all PTB-XL records to the negative class and all SaMi-Trop records to the positive class.

4. Results

All results reported in this section were obtained from the held-out portion of the training data and not from the hidden official test set.

4.1. Stage1: Optimization of the binary classification model

To examine the effect of data augmentation, we conducted parameter searches for *temporal shift ratio* and *signal masking ratio* using first-degree atrioventricular block (1AVB) as the target condition.

The temporal shift ratio was varied from 0.50 to 0.80 in increments of 0.05 during training. As a result, considering AUROC, F1 score, and loss in combination, a value of **0.55** was identified as the optimal setting.

With the temporal shift ratio fixed at 0.55, the signal masking ratio was varied between 0.10 and 0.30 in increments of 0.05. The best performance was achieved at **0.20**.

Based on these results, subsequent experiments adopted a temporal shift ratio of 0.55 and a signal masking ratio of 0.20 as the default augmentation settings.

The classification performance of the ResNet18 model showed AUROC values of **0.94** for 1AVB, **0.74** for IRBBB, **0.98** for CLBBB, **0.74** for LNGQT, **0.65** for AMI, **0.98** for AFIB, and **0.93** for LPR. Among these, 1AVB, CLBBB, AFIB, and LPR achieved AUROC values greater than 0.9. Furthermore, performance improved when input leads were optimized, with AUROC increasing to **0.96** for IRBBB using the V1 lead and to **0.85** for AMI using all 12 leads.

To compensate for abnormalities that were insufficiently represented in real data, synthetic ECGs generated by the simulator were incorporated. This led to performance values of **2AVB= 0.81** and **LNGQT = 0.77**.

Table1. Classification performance with synthetic ECGs.

Abnormality	without synthetic ECG	with synthetic ECG
2AVB	-	0.81
LNGQT	0.74	0.77

4.2. Stage2: Comparison of binary classification models

A total of 21,799 PTB-XL cases were used as negative examples, while 1,959 SaMi-Trop cases were used as positive examples. In Stage2, we compared three conditions: (i) models using demographic features only, (ii) models incorporating abnormality scores derived from the Stage1 DNN without synthetic data, and (iii) models additionally incorporating synthetic ECGs. The results are summarized in Table 2.

Table 2. Challenge score performance of Stage2 binary classification models (classification between PTB-XL and SaMi-Trop, not the official setting.)

Model	Demographics only	DNN + No synthetic data	DNN + Synthetic data
RF	0.0871	0.6147	0.6098
XGBoost	0.0000	0.6319	0.6626

For the Random Forest classifier, using demographic features alone resulted in a challenge score of **0.0871**. Incorporating DNN-derived abnormality scores without synthetic data achieved **0.6147**, while incorporating synthetic ECGs yielded **0.6098**.

For XGBoost, demographic features alone yielded a challenge score of **0.0000**. Incorporating DNN-derived scores without synthetic data improved performance to **0.6319**, and with synthetic ECGs further improved to **0.6626**.

4.3. Official challenge submission

Table 3. Official Challenge results for team RISMU.

Training	Validation	Test	Ranking
N/A	N/A	N/A	Not ranked

Challenge scores for our selected entry (team RISMU), including our team's ranking on the hidden test set. We used cross-validation on the public training data and did not receive official scores on the hidden validation/test sets. The team could not be scored on the hidden test data within the allotted test time and therefore was not ranked.

5. Discussion

In this study, we adopted a two-stage framework to capture ECG abnormalities characteristic of Chagas disease. In Stage1, a ResNet18-based binary classifier estimated abnormality scores. Several abnormalities such as 1AVB, CLBBB, AFIB, and LPR achieved AUROC values >0.9. Accuracy further improved with tailored lead selection: IRBBB detection improved with lead V1, and AMI with all 12 leads. For abnormalities with few real cases (2AVB, LNGQT), synthetic ECGs were generated. As a result, 2AVB achieved an AUROC of 0.81, whereas LNGQT improved only slightly (0.74→0.77), reflecting the difficulty of reproducing QT prolongation. In Stage2, Stage1 abnormality scores were combined with demographic features for binary classification using Random Forest and XGBoost. Without deep learning–derived features, performance was negligible. Incorporating abnormality scores raised AUROC to 0.5–0.6, and further adding synthetic data improved it to 0.6626 with XGBoost. These results indicate that combining demographic features with deep learning–based scores is beneficial, and that gradient boosting in particular better exploits synthetic data.

6. Conclusion

In this study, we proposed a two-stage framework consisting of ECG abnormality score estimation (Stage1) and binary classification (Stage2). The model achieved high AUROC values across several ECG abnormalities, confirming the effectiveness of lead selection and synthetic data augmentation. Combining deep learning–derived abnormality scores with demographic information improved classification performance, with XGBoost showing further gains when synthetic data were introduced. This work demonstrates that integrating real and synthetic data with deep learning can enhance Chagas disease screening accuracy under data scarcity. Future work will focus on external validation, improvement of synthetic ECG generation, and extension to other cardiac disorders toward developing a more generalizable screening system.

Acknowledgements

This work was supported by the SECOM Science and Technology Foundation. The authors declare that they have no conflicts of interest to disclose.

References

[1] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, e215–e220, Jun. 2000.

[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"; *In Computing in Cardiology 2025,* volume 52: 2025, 1-4.

[3] 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";URL: http://arxiv.org/abs/2510.02202,DOI:10.48550/arXiv.2510.02202.

[4] R. D. de Carvalho, A. H. Ribeiro, P. W. Macfarlane, T. B. Schön, and A. L. P. Ribeiro, "Screening for Chagas disease from the electrocardiogram using a deep neural network,"*PLoS Negl. Trop. Dis.*, vol. 17, no. 7, e0011118, Jul. 2023.

[5] N.Nonaka, J. Seita," Boosting ECG classification performance by pre-training with synthesized data,"*ECML-workshop(DEARING)2024*.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv*:1512.03385, Dec. 2015.

[7] 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.

[8] A. L. P. Ribeiro, A. P. Sabino, A. L. Lacerda, F. M. J. Santos, C. C. J. Silva, A. C. C. de Oliveira, R.C.R.Oliveira, A.R.A.Lima-Costa, and E.L.Chiari, "Design and implementation of the SaMi-Trop cohort: a cohort of Trypanosoma cruzi–seropositive patients in Brazil," *Rev. Soc. Bras. Med. Trop.*, vol. 46, no. 3, pp. 243–250, Jun. 2013.

Address for correspondence:

Jun Seita
Predictive Medicine Special Project, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan 230-0045, Japan
jun.seita@riken.jp