Detection of Chagas Disease Using Digital Electrocardiogram by Deep Transfer Learning of the InceptionTime Model

Arnaud Champetier[1], Christian Mueller[1]

[1]Cardiovascular Research Institute of Basel, Basel, Switzerland

Abstract

Chagas disease, caused by Trypanosoma cruzi, remains underdiagnosed in low-resource settings where serological testing is limited. Given the widespread availability of electrocardiography (ECG) and the conduction abnormalities characteristic of Chagas cardiomyopathy, ECG-based artificial intelligence offers a scalable alternative for early detection. A deep transfer learning model was developed based on a pretrained InceptionTime architecture, and fine-tuned on Brazilian (CODE-15%, Sami-Trop) and European (PTB-XL) ECG datasets. Recordings were preprocessed with filtering, downsampling, normalization, and 5-second segment extraction, with data augmentation applied during training. The training loss consisted of binary cross-entropy with a penalization term to emphasize the challenge metric. Inference combined predictions across multiple ECG segments and models. The approach achieved a cross-validation score of 0.42, a score of 0.382 on the validation set, and an average test score of 0.256 ranking 6th/40. These findings demonstrate the feasibility of deep transfer learning for ECG-based Chagas screening and its potential to expand diagnostic access in underserved regions.

1. Introduction

Chagas disease, caused by Trypanosoma cruzi, affects 6–8 million people in Latin America and causes an estimated 10,000 deaths annually [1]. Despite its significant burden, diagnosis relies on multiple serological assays, which are costly and less accessible in low-resource regions[1], [2]. Machine learning approaches have been applied to serological data for detecting the parasite in blood samples. However, these methods do not address the accessibility challenges in underserved areas [3], [4].

Chagas cardiomyopathy, developed in the chronic phase of the disease, is often marked by characteristic conduction disturbances—such as right bundle branch block, left anterior hemiblock, first-degree atrioventricular block, atrial fibrillation, and ventricular ectopy. These conduction disturbances not only define its clinical profile but have also been integrated into established prognostic scores [5], [6], [7], [8], [9]. Therefore, the electrocardiogram (ECG) offers a low-cost, widely available alternative to the serological testing. Yet, the direct use of ECG for Chagas disease screening remains underexplored, representing a critical opportunity to improve early detection and expand diagnostic reach [10].

The 2025 PhysioNet Challenge seeks to address this research gap by leveraging two Brazilian datasets—CODE-15% and Sami-Trop—and one European dataset, PTB-XL [11], [12], [13], [14], [15], [16], [17]. Our work aims to use a large pre-trained ECG model and fine tune it to better detect the Chagas disease.

2. Method

Our methodology was inspired by the preprocessing framework of the PhysioNet 2021 Challenge winner and by the binary outcome prediction model of Buscher et al., which relies on the InceptionTime backbone architecture [18], [19], [20].

The architecture comprises two residual blocks, each containing three inception modules. The first inception module processes the raw ECG signal by applying a convolution across all channels to generate a bottleneck layer with 32 filters. This representation is subsequently passed through three parallel convolutional layers, each with 32 filters but distinct kernel sizes. The outputs of these convolutions are concatenated with the result of a max-pooling operation applied directly to the raw ECG (Figure 1). Each subsequent inception module receives as input the output of the preceding module.

At the final stage, an adaptive average pooling layer reduces the temporal dimension to a single value, producing a feature vector of size 128 (32 × 4). A fully connected layer (128, 1) then transformed this into the ECG feature vector. To leverage demographic information, patient age and sex, encoded as numerical values, were concatenated with the ECG feature vector before classification. The combined vector was passed through two fully connected layers of sizes (3, 3) and (3, 1), respectively, yielding the final prediction. The InceptionTime backbone was initialized with pretrained weights from Buscher et al., derived from a large emergency department ECG cohort [18].

All ECG recordings were preprocessed prior to model input. A zero-phase third-order Butterworth band-pass

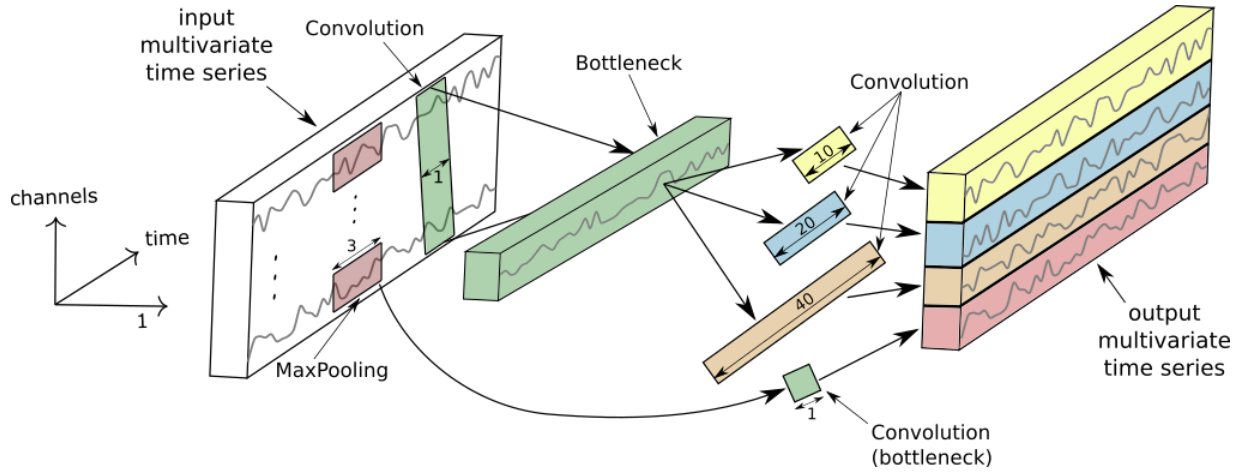

Figure 1: Inception module for time series classification (from the original paper). Bottleneck is visually described with a single dimension versus 32 in implementation.

filter (1–47 Hz) was applied, followed by downsampling to 200 Hz and z-score normalization. From each recording, a random 5-second segment was extracted. To improve model robustness, data augmentation was applied during training: random lead masking (10% of leads set to zero), two temporal masks covering approximately 6% of the signal window, and additive Gaussian noise sampled from $\mathcal{N}(0, 0.02)$.

The optimization objective was a penalized binary cross-entropy (pBCE) loss, designed to align with the challenge metric requirements. For a training batch with labels $y \in \{0,1\}$ and predictions $p \in [0,1]$, the loss is defined as:

$$L_{pBCE} = \alpha L_{BCE} + (1-\alpha) L_{FN} \quad (1)$$

Here, α is a hyperparameter set to 0.66 to balance the loss terms. The first term is the standard binary cross-entropy, and the second term penalizes false negatives above the 95th percentile of predicted probability:

$$L_{FN} = \sum_{i \in \mathfrak{B}} \mathbb{1}_{FN_i} \quad (2)$$

With:

$$\mathbb{1}_{FN_i} = \begin{cases} 1 & if\ y_i = 1\ and\ \hat{y}_i < \tau_{0.95} \\ 0 & otherwise \end{cases} \quad (3)$$

Where $\mathfrak{B}$ is the set of all indices in the batch, y_i the outcome, $\hat{y}_i$ the predicted outcomes, $\tau_{0.95}$ the 95th percentile of predicted outcomes in the batch.

A 5-fold cross-validation (CV) strategy was used, training five models independently with different data splits. Model training was conducted using the fastai framework (built on the PyTorch library) with the Adam optimizer. Training proceeded in two phases. In the first phase, all layers except the final linear classifier were frozen, and the model was pretrained for 5 epochs using the One-Cycle learning rate policy, with the learning rate varying from 1×10^{-2} to 1×10^{-4}. In the second phase, all layers, including the Inception modules, were unfrozen and trained for 25 epochs with the One-Cycle policy, with learning rates varying from 5×10^{-3} to 5×10^{-6}. A learning rate scheduler (ReduceLROnPlateau, patience = 2, reduction factor = 0.1) was applied, and early stopping was employed based on validation loss with a patience of 10 epochs.

For inference, the three models with the highest validation performance were selected. As the input length was restricted to 5 seconds, five random 5-second segments were extracted from each ECG recording. Each model generated probability estimates for all segments, which were averaged to obtain a segment-level prediction per model. The final output was the mean of the three model-level predictions.

Computations were performed using a NVIDIA GeForce RTX 3090 (24 GB), computational resources were allocated to hyperparameter optimization using the Optuna framework, followed by iterative refinement in subsequent training runs.

3. Results

Table 1: Losses, scores and rank across each dataset

Dataset	Custom Loss	Score	Official
CV training	1.10±0.06		
CV validation	1.14±0.03	0.42±0.01	
REDS-II validation		0.382	6/40
REDS-II (test)		0.289	17/40
SaMi-Trop 3 (test)		0.355	3/40
ELSA-Brasil (test)		0.125	7/40

Cross-validation training scores were not directly

computed, as this would have substantially increased training time. Instead, for each fold, the training loss and validation score corresponding to the epoch with the best validation loss were extracted. Table 1 reports the mean ± standard deviation of these values across the 5 folds. Alongside the challenge scores across each of the competition's datasets and the corresponding scores. Achieving a mean score of 0.256 across the 3 test datasets, we ranked 6th out of 40 official entries and 7th out of 65 total submissions

4. Discussion

The primary finding is a significant decrease in the challenge score from 0.382 on the REDS-II validation dataset to 0.289 on the REDS-II test dataset, indicating reduced model performance on unseen data from a comparable clinical setting. This variance in performance may be attributed to two key factors: first, the lack of more diverse data augmentation techniques to enhance model's ability to generalize across varied scenarios; second, the reliance on probabilities from individual ECG segments, rather than analyzing the entire ECG, likely contributed significantly to the performance decline by missing broader contextual patterns. Due to time constraints in the competition, the variance of the model performance due to intra-ECG segment predictions was not explored.

Future work could investigate the variability of deep learning models when applied to ECG segments or, similarly, examine prediction consistency across consecutive ECGs (intra-setting ECGs) to address these limitations and improve model robustness.

Several alternative methodologies were explored during the official phase, including the fine tuning of newly derived foundations models: ECGFounder and ECGFM [21], [22]. Training the classification head or the full ECGFounder model on ECG signals upsampled to 500 Hz resulted in unstable training dynamics. Also, we evaluated an approach in which the 27 cardiac outcome predictions produced by ECGFounder were used as input features for an XGBoost classifier. However, when assessed using 5-nested-fold cross-validation, this strategy achieved lower performance than our deep learning model. A student–teacher framework based on the ECGFM model, which incorporated 150 predictions into a distillation loss, was also investigated; however, this approach suffered from either unstable training dynamics or severe overfitting.

These limitations highlight the challenges of adapting large, pretrained ECG models within the competition's time constraints. Future work could examine generalization performance of ECGFounder and ECGFM under specific training strategies and explore advanced regularization schemes for distillation.

5. Conclusion

We developed and evaluated a deep learning model for Chagas disease detection from digital ECG signals, addressing the need for accessible diagnostic tools in low-resource regions. Leveraging a pretrained InceptionTime backbone and fine-tuning it on Brazilian and European datasets, we achieved competitive (6^{th}/40) performance in the 2025 PhysioNet Challenge.

Acknowledgments

This work was supported by the university hospital Basel, the university of Basel, the Swiss Heart Foundation and the Swiss National Science Foundation.

References

[1] A. S. de Sousa, D. Vermeij, A. N. Ramos, and A. O. Luquetti, "Chagas disease," *Lancet*, vol. 403, no. 10422, pp. 203–218, Jan. 2024, doi: 10.1016/S0140-6736(23)01787-7.

[2] J. A. Pérez-Molina and I. Molina, "Chagas disease," *Lancet*, vol. 391, no. 10115, pp. 82–94, Jan. 2018, doi: 10.1016/S0140-6736(17)31612-4.

[3] V. Uc-Cetina, C. Brito-Loeza, and H. Ruiz-Piña, "Chagas parasite detection in blood images using AdaBoost," *Comput. Math. Methods Med.*, 2015, doi: 10.1155/2015/139681.

[4] A. Ojeda-Pat, A. Martin-Gonzalez, C. Brito-Loeza, H. Ruiz-Piña, and D. Ruz-Suarez, "Effective residual convolutional neural network for Chagas disease parasite segmentation," *Med. Biol. Eng. Comput.*, vol. 1, p. 3, 2022, doi: 10.1007/s11517-022-02537-9.

[5] M. C. P. Nunes *et al.*, "Chagas cardiomyopathy: An update of current clinical knowledge and management: A scientific statement from the American Heart Association," *Circ.*, vol. 138, no. 12, pp. e169–e209, Sep. 2018, doi: 10.1161/CIR.0000000000000599.

[6] C. Di Lorenzo Oliveira *et al.*, "Risk score for predicting 2-year mortality in patients with Chagas cardiomyopathy from endemic areas: SaMi-Trop cohort study," *J. Am. Heart Assoc.*, vol. 13, doi: 10.1161/JAHA.119.014176.

[7] L. Z. Rojas *et al.*, "Electrocardiographic abnormalities in Chagas disease in the general population: A systematic review and meta-analysis," *PLoS Negl. Trop. Dis.*, vol. 12, no. 6, Jun. 2018, doi: 10.1371/journal.pntd.0006567.

[8] M. S. Marcolino, D. M. Palhares, L. R. Ferreira, and A. L. Ribeiro, "Electrocardiogram and Chagas disease: A large population database of primary care patients," *Glob. Heart*, vol. 10, no. 3, p. 167, Sep. 2015, doi: 10.1016/j.gheart.2015.07.001.

[9] A. Rassi *et al.*, "Development and validation of a risk score for predicting death in Chagas' heart disease," *N. Engl. J. Med.*, vol. 355, no. 8, pp. 799–808, Aug. 2006, doi: 10.1056/NEJMoa053241.

[10] C. Jidling *et al.*, "Screening for Chagas disease from the electrocardiogram using a deep neural network,"

PLoS Negl. Trop. Dis., vol. 17, no. 7, p. e0011118, Jul. 2023, doi: 10.1371/journal.pntd.0011118.

[11] M. A. Reyna *et al.*, "Detection of Chagas disease from the ECG: The George B. Moody PhysioNet Challenge 2025," *Comput. Cardiol.*, vol. 52, pp. 1–4, 2025.

[12] A. H. Ribeiro *et al.*, "CODE-test: An annotated 12-lead ECG dataset," 2020, doi: 10.5281/zenodo.3765780.

[13] A. L. P. Ribeiro *et al.*, "Tele-electrocardiography and big data: The CODE (Clinical Outcomes in Digital Electrocardiography) study," *J. Electrocardiol.*, vol. 57, pp. S75–S78, Nov. 2019, doi: 10.1016/j.jelectrocard.2019.09.008.

[14] P. Wagner *et al.*, "PTB-XL: A large publicly available electrocardiography dataset," *Sci. Data*, vol. 7, no. 1, 2020, doi: 10.1038/s41597-020-0495-6.

[15] A. L. P. Ribeiro *et al.*, "SaMi-Trop: 12-lead ECG traces with age and mortality annotations," 2021, doi: 10.5281/zenodo.4905618.

[16] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circ.*, vol. 101, no. 23, 2000, doi: 10.1161/01.CIR.101.23.E215.

[17] M. A. Reyna *et al.*, "Detection of Chagas disease from the ECG: The George B. Moody PhysioNet Challenge 2025," *Comput. Cardiol.*, Oct. 2025, doi: 10.48550/arXiv.2510.02202.

[18] A. Büscher *et al.*, "Deep learning electrocardiogram model for risk stratification of coronary revascularization need in the emergency department," *Eur. Heart J.*, Mar. 2025, doi: 10.1093/eurheartj/ehaf254.

[19] P. Nejedly *et al.*, "Classification of ECG using ensemble of residual CNNs with attention mechanism," *Comput. Cardiol.*, vol. 48, pp. 1–4, 2021, doi: 10.23919/CINC53138.2021.9662723.

[20] H. Ismail Fawaz *et al.*, "InceptionTime: Finding AlexNet for time series classification," *Data Min. Knowl. Discov.*, vol. 34, no. 6, pp. 1936–1962, Nov. 2020, doi: 10.1007/s10618-020-00710-y.

[21] J. Li *et al.*, "An electrocardiogram foundation model built on over 10 million recordings," *N. Engl. J. Med. AI*, vol. 2, no. 7, Jun. 2025, doi: 10.1056/AIOA2401033.

[22] S. Zhang *et al.*, "ECGFM: A foundation model for ECG analysis trained on a multi-center million-ECG dataset," *Inf. Fusion*, vol. 124, p. 103363, Dec. 2025, doi: 10.1016/j.inffus.2025.103363.

Address for correspondence:
Arnaud Champetier
Spitalstrasse 2, 4056 Basel, Switzerland
arnaudtanguy.champetier@usb.ch