ChagaL: Low-Rank Adaptation to Detect Chagas Disease from Electrocardiograms

Pierre Aublin[1,2], Stephanie Kühne[2], Philip Raake[2], Sebastian Zaunseder[1] and Dario Bongiovanni[2]

[1] Chair for Diagnostic Sensing, University of Augsburg, Germany
[2] Department of Internal Medicine I, University Hospital of Augsburg, University Augsburg, Germany

Abstract

This work aims to develop an algorithm for the automatic detection of Chagas disease using 12-lead ECG within the George B. Moody PhysioNet Challenge 2025. To that end, we adapted a residual network, initially trained in a South American population to estimate the patient biological age using the ECG. To perform this adjustment, we used Low-Rank Adaptation to update the weights of the convolutional layers of the networks. To mitigate labeling uncertainties, we trained several models using different sub-datasets , which are part of the challenge public training set. These models were then assembled to provide a prediction on the presence of Chagas disease from the ECG. Locally, the dataset was split into a 90/10 % training/internal held out sets stratified by demographics, chagas labels and origin of the database to optimize the hyperparameters. Hyperparameters that provided the best internal validation challenge score were used to train the final model in the full training set. Our model achieved a challenge metric score of 0.202 on the challenge test set leading our team, ChagAI, to be ranked 23rd among 40 participants. The results highlight the difficulties of detecting Chagas disease, even with a large training dataset. Despite a better than random level, the model requires further optimization to improve its performance if it is intended to be used in clinical practice.

1. Introduction

Chagas disease, caused by the protozoan parasite Trypanosoma cruzi, is a neglected tropical disease endemic to Latin America but increasingly recognized as a global health concern [1]. While the acute phase of the disease may present with mild or non-specific symptoms, its chronic form can persist silently for decades before manifesting as Chagas cardiomyopathy, a severe condition characterized by conduction abnormalities, arrhythmias, heart failure, and sudden cardiac death. Early identification of individuals with Chagas disease is critical for managing progression and improving long-term outcomes[2–4]. Currently, diagnosis primarily bases on serological testing, which, while sensitive, is often limited by cost, infrastructure, and access, particularly in resource-constrained settings[5]. As Chagas cardiomyopathy affects the electrical conduction system of the heart, the electrocardiogram (ECG) offers a promising, widely accessible, and non-invasive modality for detecting disease-related abnormalities[6, 7]. As a part of the PhysioNet Challenge 2025, this work is dedicated to automatic detection of Chagas disease based on the ECG. Thereto, we leverage a residual neural network originally trained to estimate ECG-derived biological age and adapt it to the task of Chagas detection using Low-Rank Adaptation (LoRA)[8, 9]. This approach, combined with regularization and ensembling strategies to address label noise, aims to improve the model's generalization and robustness, ultimately contributing to scalable and accessible screening tools for Chagas disease.

2. Methods

2.1. Data preparation

A public dataset combining ECG recordings from 3 different cohorts was used for hyperparameter optimization, training and local testing. The dataset contains ECG recordings originating from 3 different cohorts. The CODE15 dataset [8] is a dataset containing 343,424 ECG recordings from the Telehealth network of Minas Gerais. The labels regarding the existence of Chagas disease have been self-reported by patients and are therefore not validated. The Sami-Trop dataset is formed by a Brazilian cohort of patients with chronic chagas disease confirmed by serological tests [10]. From this cohort, 1,631 ECG recordings are provided. The PTB-XL dataset is a German dataset of 21,799 ECG recordings from patients with various cardiac disorders but presumably without Chagas disease [11]. As illustrated on Figure 1, a first split, strat-

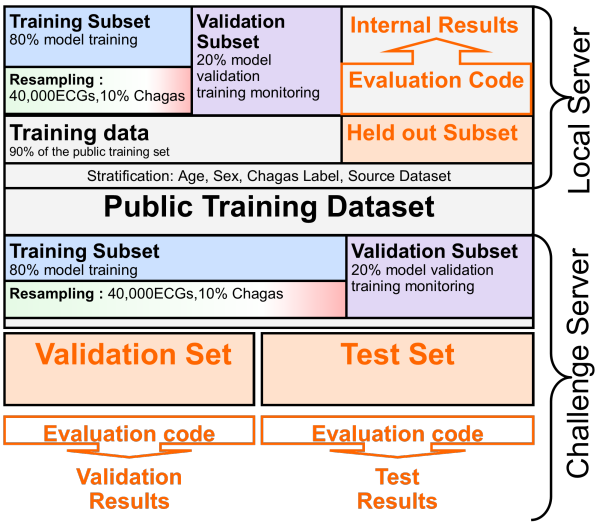

Figure 1. Splitting of the public training dataset for both training or testing our models locally, and assessing them on the challenge servers and private datasets.

ified by Age, Sex, dataset of origin and Chagas label was performed and 10 % of the data was kept apart as an internal held out subset. This set served to compare the different approaches presented in this work. The 90% remaining data were further split into a 80% training subset, and a 20% validation subset for monitoring the models training and optimizing the different model hyperparameters. For training the final model on the challenge server, the whole public dataset was used with 80% of the recordings serving to train the models while the 20% remaining data served for monitoring the model's loss and evaluation metrics during training.

Several preprocessing steps were applied to ensure uniform ECG inputs independently from the database they were retrieved from. All the twelve leads were first bandpass filtered via a 3rd order Butterworth filter between 0.1 Hz and 30 Hz. As the powerline frequencies differed between several databases, 2nd order notch filters were used to remove 50 Hz and 60 Hz frequencies. A resampling of the recordings was then performed to 400 Hz. Signals longer than 4096 sampled were cropped to the 4096 first samples, while those shorter were zero-padded to 4096 samples on the right side. To account for different acquisition settings between the sub-datasets, such as analogue preprocessing, sampling frequency, and quantization, several preprocessings were applied to the ECG recordings. A white noise was added set so that the signal to noise ratio was to 20 dB. To mitigate the difference in recordings resolution, signals were rescaled to the range $[0, 1]$ and a white noise of standard deviation 1/1024 was added. Signals were then centered to zero-mean afterwards.

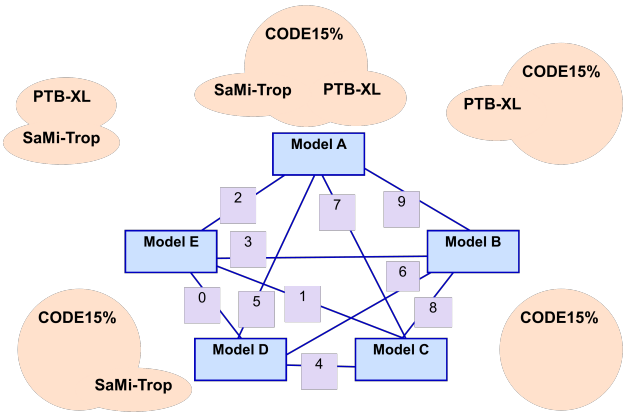

Figure 2. Graph representing the models created using different sub-datasets of the public training dataset. The edges correspond to the models trained on these subdatasets while the vertices are obtained by averaging the parameters of the models of the edges.

The different data sets that make up the public training dataset differ in terms of the reliability of the labels and the population that compose the cohorts. Several models were trained by restricting the training and validation set to different subsets of origin sources. Therefore, besides training a model using the whole local training and validation dataset, we trained several additional models by restricting these datasets to all the avaiblable pairs of subdomains, that is CODE15 with PTB-XL, CODE15 with Sami-Trop, PTB-XL with Sami-Trop. As only CODE15 contains both positive and negative Chagas labels, another model is also trained using only this cohort. As all these subdatasets and their combinations were of different size, the training set in each case was resampled with replacement so that 40,000 ECGs served as training data with among them 10% recordings having positive Chagas labels. This results in set of five models fine-tuned to detect Chagas disease from ECG recordings. Another set of 9 models was then built by averaging the parameters of all pairs of those five models. This lead to 15 trained neural network as described on Figure2. The final prediction made by our model consisted then in a soft voting of those 15 neural networks taking the average of the networks output probabilities.

2.2. Model

The base neural network used to train our model was the ResNet proposed by Lima et. al. [8], depicted on Figure 3. The model was originally used to predict biological age from ECG recordings. We used the parameters of this trained model as a starting point for our fine-tuning approach, keeping them frozen during training. The pa-

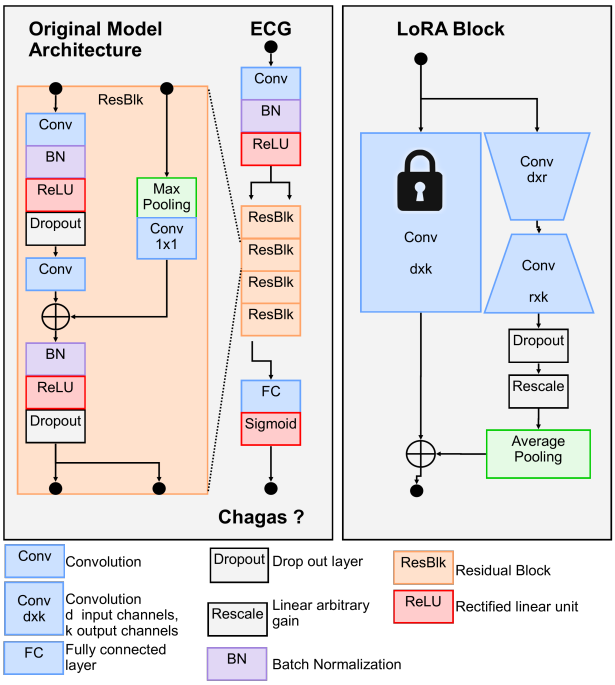

Figure 3. Left: Architecture of the pretrained neural network as described by Lima et. al. Right: Implementation of LoRA block added to each convolution layer of the original architecture.

rameters of the last fully connected layer were, however, randomly initialized via He uniform initializationand kept free for fine-tuning.

To fine tune the model, we applied Low-Rank Adaptation (LoRA). The approach uses low-rank decomposition of matrices to constrain the weights of the original model to be changed only in a low dimensional space [9]. More formally, given a weight matrix $W_0 \in \mathbb{R}^{d\times k}$, a new weight matrix $W \in \mathbb{R}^{d\times k}$ is built during the fine-tuning for the new task by adding a residual branch :

$$W = W_0 + \frac{\alpha}{r}BA \tag{1}$$

where $B \in \mathbb{R}^{d\times r}$, $A \in \mathbb{R}^{r\times k}$, and $r << min(d, k)$. This ensures that the matrix BA is low-rank. α is a rescaling coefficient. To reduce the hyperparameters' search space, we set $\alpha = 2r$ as specified in [9].

We adapted the LoRA method to convolutional layers as illustrated on Figure 3. To each convolution with d input filters and k output filters, we added a residual block consisting of a convolution of kernel-size 1, d input filters and r output filters, followed by a second convolution of kernel-size 1 of r input filters and k output filters. The output of these convolutions was followed by a dropout and rescaled by $\frac{\alpha}{r} = 2$. An average pooling using the kernel size, stride and padding size of the original convolution block was used so that the size of the output of the residual block matches the output size of the original convolution.

Table 1. Total number of trainable parameters for each model, and training parameters used for the model training during comparison. (NA stands for Not Applicable)

Parameter	LoRA	Linear	Full
Total trainable parameters	343,195	5,121	6,924,705
r	64	NA	NA
Dropout	0.1	NA	NA
α	0.5	0.75	0.25
γ	2	2	1.5
l_r	1.10^{-3}	5.10^{-4}	1.10^{-3}

Table 2. Challenge metric, AUROC, F1-score, PPV reached by all trained models on the held out subset using the different fine-tuning strategies.

Local Validation	CM	AUROC	F1-Score	PPV
LoRA	0.263	0.73	0.12	0.07
Linear	0.260	0.71	0.11	0.06
Full	**0.267**	**0.74**	**0.13**	**0.08**

2.3. Training

As Chagas disease has a low prevalence, the neural network was trained to minimize the focal loss to address the class imbalance of our dataset [12]. An Adam optimizer with decoupled weight decay regularization was used to adjust the neural networks weights [13]. For the different fine-tuning schemes compared in this work, hyperparameters were found via grid-search to maximize the challenge metric on the internal validation subset.

The best hyperparameter combination found during the local grid-search was then used to train the neural network on the challenge server.

3. Results

We compared the LoRA approach to two other finetuning schemes on the local held out set. A first one, denoted Linear, was linear probe, consisting in training only the last classification layer of the pretrained network described in Figure 3 without using any LoRA block. The second approach, denoted Full, was a full training of neural network initialized with the pretrained weights from [8], without using any LoRA block. Table 1 indicates the number of parameters trained by each finetuning scheme and summarizes the best hyperparameter combinations retained to compare the different finetuned models. It can be noticed that LoRA trains only 10 percents of parameters compared to the Full tuning approach. Table 2 summarizes the performance of the different models on the local test set. We report the metric used for the PhysioNet challenge 2025, which consists in the Sensitivity of a model after assigning a positive label to the top 5% of the samples for which

Table 3. Challenge Metric of the LoRA model on the external validation and test sets.

Validation Set	Test Set	Team Ranking
0.259	0.202	23

the model predicted the highest probability of Chagas disease while the remaining are assigned a negative label. We also report the model's positive predictive value (PPV) and F1-score to summarize sensitivity and positive predictive value. All the models obtained a negative predictive value of 0.99 on the held out subset. As reported on Table 3, the LoRA model with Majority voting obtained a score of 0.259 on the validation set of the challenge and 0.202 on the challenge test set, leading our entry to be ranked 23rd among 40 successful submissions.

4. Discussion

We used Low-Rank adaptation to finetune a residual network that was originally trained for another task on a similar population to the one of the training set. Compared to the linear approach, LoRA allowed to tune also the lower layers but required significantly less parameters than the Full fine-tuning approach. The model from [8] was trained on the Telehealth Network of Minas Guerais, from which the CODE15% dataset is extracted. While the Full approach may lead to forget knowledge that could have been encoded within the training of the original model, LoRA can still keep this information as the weights are only updated through the residual blocks introduced as on Figure 3. To address the differences in label certainty of the different sub datasets forming the public train set, we proposed a voting scheme combining several models that were trained only on combinations of these different sub datasets. We increased the number of participants to the voting by averaging the parameters of model pairs [14].

While the models seem to perform better than random classifiers according to the AUROCs, the binary classification metrics remain low. Part of this highlights the difficulty of training models to detect rare diseases even from a big dataset. As the challenge metric was independent of the binary model's classification decision threshold, no calibration was performed to adapt it. While the high negative predictive values of all the developed models would rather suggest using such an automated approach to rule out patients who would probably not have Chagas disease, the clinical implementation of this tool will be limited regarding its low F1-score and positive predictive value.

References

[1] World Health Organization. Chagas disease (American trypanosomiasis). URL https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis).

[2] Bern Caryn. Chagas' disease. New England Journal of Medicine ;373(5):456–466. Publisher: Massachusetts Medical Society.

[3] Rassi A, Rassi A, Marin-Neto JA. Chagas disease. The Lancet April 2010;375(9723):1388–1402. ISSN 0140-6736.

[4] Nunes MCP, et.al. Chagas cardiomyopathy: an update of current clinical knowledge and management: a scientific statement from the American Heart Association. Circulation September 2018;138(12):e169–e209. Publisher: American Heart Association.

[5] Piron M, et.al. Development of a real-time PCR assay for trypanosoma cruzi detection in blood samples. Acta Tropica September 2007;103(3):195–200. ISSN 0001-706X.

[6] Reyna MA, Koscova Z, Pavlus J, Weigle J, Saghafi S, Gomes P, Elola A, Hassannia MS, Campbell K, Bahrami Rad A, Ribeiro AH, Ribeiro AL, Sameni R, Clifford GD. Detection of Chagas disease from the ECG: the George B. Moody PhysioNet Challenge 2025. Computing in Cardiology 2025;52:1–4.

[7] 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.

[8] Lima EM, et.al. Deep neural network-estimated electrocardiographic age as a mortality predictor. Nature Communications August 2021;12(1):5117. ISSN 2041-1723.

[9] Hu EJ, et.al. LoRA: low-rank adaptation of large language models. In International Conference on Learning Representations. 2022; .

[10] Cardoso CS, et.al. Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile. BMJ Open May 2016; 6(5):e011181.

[11] Wagner P, et.al. PTB-XL, a large publicly available electrocardiography dataset. Scientific Data May 2020;7(1):154. ISSN 2052-4463.

[12] Lin TY, et.al. Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence July 2018;PP:1–1.

[13] Loshchilov I, Hutter F. Decoupled weight decay regularization. In International Conference on Learning Representations. 2019; .

[14] Wortsman M, et.al. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. March 2022.

Address for correspondence:

Pierre Aublin
Eichleitnerstr. 30, 86159 Augsburg, Germany
pierre.aublin@uni-a.de