# Lightweight Deep Neural Network for Chagas Disease Screening Using 12-Lead ECG Signals

Quenaz B Soares[1], Diego A C Cardenas[1], Felipe M Dias[1], Estela Ribeiro[1], Jose E Krieger[1], Marco A Gutierrez[1],

[1]Heart Institute, Clinics Hospital, University of Sao Paulo Medical School, Brazil

## Abstract

*As part of the George B. Moody PhysioNet Challenge 2025, we employed two lightweight convolutional neural network architectures based on modified VGG and ResNet models to detect Chagas disease from 12-lead electrocardiogram (ECG) signals. Our preprocessing pipeline included resampling to 128 Hz, fixed-length cropping or zero-padding, bandpass filtering (1–40 Hz), and channel-wise normalization. To address pronounced class imbalance, we employed a balanced sampling strategy during training and used Monte Carlo Dropout at inference for uncertainty estimation. Evaluated via 5-fold cross-validation, the best model (LiteVGG-11) achieved an AUROC of 0.842±0.009 and a recall of 0.725±0.018, but low precision (0.066±0.002), resulting in a challenge score of 0.410±0.009. These results highlight both the potential and the challenges of automated Chagas disease detection from ECGs in highly imbalanced datasets. It also indicates the feasibility of lightweight models for scalable and portable screening solutions in resource-limited settings. Our model received a Challenge score of 0.240 (ranked 9th out of 40 teams) on the hidden test set.*

## 1. Introduction

Automatic electrocardiogram (ECG) classification holds promise for aiding in the detection of cardiac abnormalities, including those caused by Chagas disease, a parasitic illness, which affects millions worldwide and can lead to severe cardiac complications in its chronic phase. While early infection is often asymptomatic and treatable, advanced stages can result in cardiomyopathy, arrhythmias, and heart failure [1]. Diagnosis is typically based on serological tests, which are not always accessible. Since Chagas cardiomyopathy often presents with ECG abnormalities, automated ECG analysis offers a promising tool for screening.

The George B. Moody PhysioNet Challenge 2025 [1–3] called for open-source algorithms to identify likely cases of Chagas disease from 12-lead ECGs, aiming to support early detection and triage. This paper describes our team's approach to this challenge.

## 2. Methods

### 2.1. Data

We used the 12-lead ECG dataset ensemble provided by the challenge [1]. This ensemble includes the CODE-15% [4], SaMi-Trop [5], and PTB-XL [6] datasets. The CODE-15% dataset contains more than 300,000 12-lead ECG recordings collected in Brazil, with self-reported Chagas disease labels. The SaMi-Trop dataset contains 1,631 ECGs from Chagas-positive patients confirmed by serological tests. The PTB-XL dataset comprises 21,799 ECGs from individuals in Europe, who were presumably Chagas-negative.

### 2.2. Pre-Processing

All ECG recordings were resampled to 128 Hz and either cropped or zero-padded to a fixed duration of 10 seconds to ensure uniform input length. A fourth-order Butterworth bandpass filter with cutoff frequencies of 1–40 Hz was applied to remove baseline wander and high-frequency noise. Finally, z-score normalization was performed channel-wise, resulting in signals where each lead had zero mean and unit variance.

### 2.3. Models

Two lightweight CNN architectures were employed in this work: LiteVGG [7] and LiteResNet [8]. Both are adaptations of computer vision CNNs for 1D signal processing, specifically designed for ECG classification. Several modifications were introduced to reduce computational complexity: (i) replacement of standard convolutional layers with depthwise separable convolutions; (ii) reduction in the number of filters per convolutional layer;

(iii) use of global pooling layers instead of traditional flattening between the convolutional and dense components; and (iv) reduction in the number of hidden units in the dense layers.

Figure 1 illustrates the LiteVGG-11 variant. As in the original VGG architecture, LiteVGG models differ by the number of convolutional layers per block. The tested configurations were:

- **LiteVGG-11**: blocks with [1, 1, 2, 2, 2] layers;
- **LiteVGG-13**: blocks with [2, 2, 2, 2, 2] layers;
- **LiteVGG-16**: blocks with [2, 2, 3, 3, 3] layers.

Figure 2 shows the LiteResNet-34 variant. LiteResNet models differ by the number of residual blocks per module:

- **LiteResNet-18**: modules with [2, 2, 2, 2] blocks;
- **LiteResNet-34**: modules with [3, 4, 6, 3] blocks.

## 2.4. Training Scheme

The dataset was highly imbalanced, with a strong predominance of the non-Chagas class. To address this, we implemented a balanced sampling strategy using a TensorFlow input pipeline. The training data was split into two subdatasets, one for Chagas and one for non-Chagas samples. During training, batches were constructed by randomly sampling from each sub-dataset with equal probability (50% each). Both sub-datasets were shuffled and repeated indefinitely, with shuffling applied prior to each repetition. This approach ensured consistent class balance throughout training, regardless of differences in sub-dataset sizes.

The models were trained using the AdamW optimizer with a learning rate of $1 \times 10^{-4}$ and a binary cross-entropy loss. We used a batch size of 64 and trained for 128 epochs. No data augmentation techniques were applied.

## 2.5. Inference Strategies

As the goal of this work is to develop an algorithm for prioritizing patients for confirmatory testing of Chagas disease, we go beyond interpreting the model output as a probability. Instead, model uncertainty is incorporated into the inference process, favoring high-confidence predictions and penalizing those with greater uncertainty.

To account for model uncertainty during inference, we evaluated several strategies based on Monte Carlo Dropout (MCD). These strategies aim to improve the robustness of

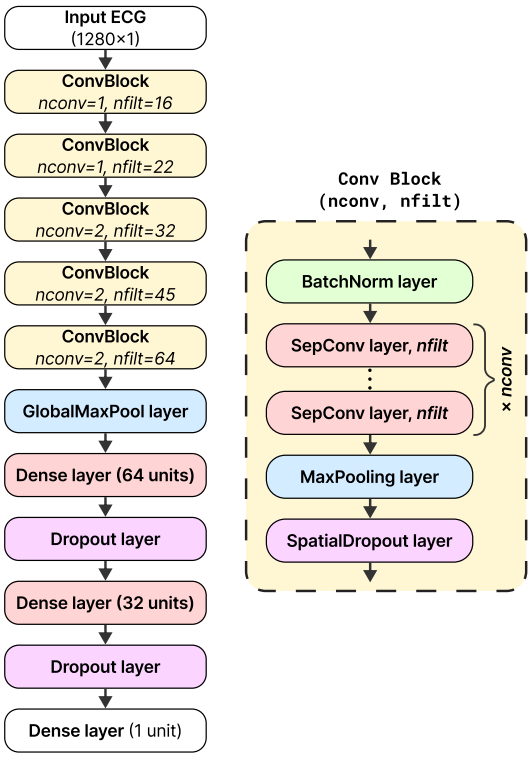


Figure 1. LiteVGG-11 architecture.

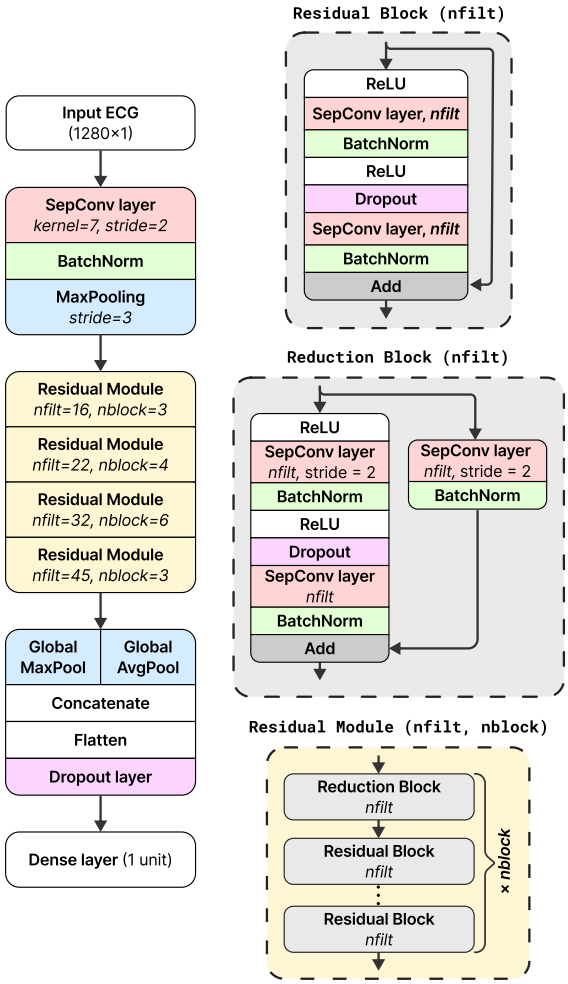


Figure 2. LiteResnet-34 architecture.

| | Equation | Description |
|---|---|---|
| **I** | $y = z_{full}$ | Deterministic output from a single forward pass with dropout disabled (baseline). |
| **II** | $y = \overline{\mathbf{z}}$ | Mean of the $T$ stochastic outputs obtained via MCD. |
| **III** | $y = \frac{1}{2}\left(z_{full} + \overline{\mathbf{z}}\right)$ | Average between the deterministic output and the MCD mean. |
| **IV** | $y = \overline{\mathbf{z}} - 0.5\sigma$ | Lower confidence estimate based on the MCD mean penalized by half its standard deviation. |
| **V** | $y = z_{full} - 0.5\sigma$ | Lower confidence estimate using the deterministic output in place of the MCD mean. |

Table 1. Inference strategies using MCD. Each strategy defines a different way to compute the final logit $y$ based on the deterministic model output $z_{\text{full}}$, the MCD mean $\overline{\mathbf{z}}$, and its standard deviation $\sigma$.

the predicted probabilities by incorporating uncertainty estimates derived from stochastic forward passes.

Given an input sample $X$, we perform $T = 10$ stochastic forward passes through the model with dropout enabled, resulting in $T$ outputs in logit space:

$$\{z_t\}_{t=1}^{T}, \quad z_t \in \mathbb{R} \tag{1}$$

From these $T$ Monte Carlo samples, we compute the empirical mean and standard deviation of the logits:

$$\overline{\mathbf{z}} = \frac{1}{T}\sum_{t=1}^{T} z_t \tag{2}$$

$$\sigma = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(z_t - \overline{\mathbf{z}}\right)^2} \tag{3}$$

We also consider the deterministic output of the model, obtained via a single forward pass with dropout disabled. This output is denoted as $z_{\text{full}}$.

Table 1 summarizes the inference strategies evaluated. Strategy **I** corresponds to the deterministic baseline, while the remaining strategies incorporate uncertainty information derived from MCD to adjust the output, either by reducing the confidence of uncertain predictions or combining uncertain and deterministic views.

## 3. Results

Table 2 shows the challenge scores from a stratified 5-fold cross-validation. Balanced sampling was applied only to training folds, while test folds retained their original imbalance to simulate real-world conditions. LiteVGG-11 consistently outperformed other models across all folds.

Table 3 provides detailed evaluation metrics for LiteVGG-11 with strategy II—the best official submission—under the same cross-validation setup, offering further insight into predictive performance.

Table 4 summarizes challenge scores for the selected submission across 5-fold cross-validation (public training set), repeated scoring on the hidden validation set, final scoring on the hidden test set, and the team's final ranking.

| Model | Strategy | Challenge Score |
|---|---|---|
| LiteResNet-18 | I | 0.365 ± 0.006 |
| | II | 0.379 ± 0.009 |
| | III | 0.375 ± 0.010 |
| | IV | 0.378 ± 0.012 |
| | V | 0.366 ± 0.007 |
| LiteResNet-34 | I | 0.359 ± 0.017 |
| | II | 0.370 ± 0.013 |
| | III | 0.368 ± 0.013 |
| | IV | 0.371 ± 0.012 |
| | V | 0.363 ± 0.017 |
| LiteVGG-11 | I | 0.408 ± 0.012 |
| | II | 0.410 ± 0.009 |
| | III | 0.411 ± 0.012 |
| | IV | 0.406 ± 0.010 |
| | V | 0.407 ± 0.011 |
| LiteVGG-13 | I | 0.396 ± 0.008 |
| | II | 0.401 ± 0.012 |
| | III | 0.402 ± 0.010 |
| | IV | 0.401 ± 0.010 |
| | V | 0.397 ± 0.009 |
| LiteVGG-16 | I | 0.387 ± 0.009 |
| | II | 0.398 ± 0.014 |
| | III | 0.396 ± 0.011 |
| | IV | 0.398 ± 0.013 |
| | V | 0.391 ± 0.010 |

Table 2. Cross validation results of the challenge score for each combination of model and strategy.

| Metric | Value |
|---|---|
| Challenge Score | 0.410 ± 0.009 |
| AUROC | 0.842 ± 0.001 |
| AUPRC | 0.167 ± 0.007 |
| Accuracy | 0.786 ± 0.012 |
| Precision | 0.066 ± 0.002 |
| Recall | 0.725 ± 0.018 |
| F1-score | 0.120 ± 0.003 |
| MCC | 0.173 ± 0.002 |

Table 3. Metrics achieved for the best submission, LiteVGG-11 model with inference strategy II, at 5-fold cross validation on the public training dataset.

| Training | Validation | Test | Ranking |
|---|---|---|---|
| $0.410 \pm 0.009$ | 0.372 | 0.240 | 9/40 |

Table 4. Challenge scores for our selected entry (team AIMED), including the ranking of our team on the hidden test set. We used 5-fold cross validation on the public training set, repeated scoring on the hidden validation set, and one-time scoring on the hidden test set.

## 4. Discussion

The LiteVGG-11 model combined with strategy II achieved a challenge score of 0.403 in cross-validation, demonstrating robust performance across folds. An AUROC of 0.836 suggests good overall discrimination between Chagas-positive and negative cases across different thresholds.

However, low precision (0.063) and AUPRC (0.164) highlight the model's difficulty in correctly identifying positives without high false positive rates. This is consistent with the dataset's strong class imbalance, where non-Chagas samples dominate. Although recall was relatively high (0.723), the low precision and F1-score (0.116) point to limited reliability at the default 50% threshold. The MCC of 0.168, which considers all confusion matrix terms, also indicates modest overall performance, emphasizing the challenge of detecting Chagas disease from ECG signals under imbalanced conditions.

The drop in challenge scores from training and validation to the hidden test set highlights limited generalization capability. This performance gap suggests that the model may be overfitting to the training distribution or failing to capture features that generalize well to unseen data.

Improving performance and generalization may require better handling of class imbalance through advanced sampling, loss functions, or threshold tuning. Incorporating clinical metadata, applying data augmentation, or exploring richer model architectures may also enhance robustness and precision.

## 5. Conclusion

Our approach achieved a challenge score of 0.240 in the hidden test set, ranking 9th out of 40 teams of the George B. Moody PhysioNet Challenge 2025.

These results highlight the potential of lightweight CNN architectures for detecting Chagas disease from ECG signals. Despite encouraging AUROC and recall values, the low precision and AUPRC underscore the need to reduce false positives and improve the reliability of positive predictions. Furthermore, the lightweight architecture of our model facilitates its integration into portable ECG devices, offering a promising avenue for scalable Chagas disease screening in resource-limited settings.


## Acknowledgments

This research received support from Foxconn Brazil, and the Zerbini Foundation as part of the research initiative titled "Machine Learning in Cardiovascular Medicine".


## References


[1] 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.

[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 AL, 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] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet. Circulation 2000;101(23):e215–e220.

[4] Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MPS, Andersson CR, Macfarlane PW, Meira W, Schön TB, Ribeiro ALP. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications April 2020;11(1). ISSN 2041-1723.

[5] Cardoso CS, Sabino EC, Oliveira CDL, de Oliveira LC, Ferreira AM, Cunha-Neto E, Bierrenbach AL, Ferreira JE, Haikal DS, Reingold AL, Ribeiro ALP. Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile. BMJ Open May 2016; 6(5):e011181. ISSN 2044-6055.

[6] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. Scientific Data May 2020; 7(1). ISSN 2052-4463.

[7] Soares QB, Monteiro R, Jatene FB, Gutierrez MA. A lightweight unidimensional deep Learning model for atrial fibrillation Detection. In 2022 Computing in Cardiology (CinC), volume 498. IEEE, 2022; 1–4.

[8] Soares QB, Gutierrez MA. Modelos de aprendizado profundo de baixo custo computacional para classificação de ritmo cardíaco. Master's thesis, Escola Politécnica, Universidade de São Paulo, April 2024. Available at: https://www.teses.usp.br/teses/disponiveis/3/3154/tde-12072024-083227/pt-br.php.



Address for correspondence:

Quenaz Bezerra Soares
Heart Institute (InCor) - Av. Dr. Enéas Carvalho de Aguiar, 44 - Cerqueira César, São Paulo - SP, Brazil
quenaz.soares@hc.fm.usp.br