

HEART - Hybrid ECG Analysis for Recognizing Chagas

Pinar Bisgin^{1,2}, Martin Sondermann¹, Harshini Eggoni¹, Ole Werger³, Prabhudev Bengaluru Kumar¹, Florian Lübke¹, Maximilian Fecke¹, Niklas Tschorn¹, Mostafa Kamal Mallick¹, Michael Pantförder¹, Hendrik Wöhrle^{3,4}

¹ Fraunhofer Institute for Software and Systems Engineering ISST, Dortmund, Germany

² TU Dortmund University, Dortmund, Germany

³ Fraunhofer Institute for Microelectronic Circuits and Systems, Duisburg, Germany

⁴ University of Duisburg-Essen, Duisburg, Germany

Abstract

Chagas disease, caused by Trypanosoma cruzi, remains a major global public health concern, particularly in endemic regions. Affecting an estimated 6.5 million individuals and causing nearly 10,000 deaths annually, the disease demands effective screening and diagnostic strategies. As part of the 2025 George B. Moody PhysioNet Challenge, we (Team DataNinyas) developed an algorithmic pipeline to prioritize patients for confirmatory testing using ECG data. Our system, HEART (Hybrid ECG Analysis for Recognizing Chagas), is a spectrogram-based convolutional neural network enhanced with a novel lead-aware attention mechanism, designed to detect subtle ECG signatures associated with Chagas disease. Our model achieved a mean Challenge Score of 19.9% during the challenge, placing our team 24th out of 41 participants. This performance demonstrates its effectiveness in ranking Chagas-positive patients among the top cases. These findings highlight the potential of algorithmic approaches to enhance diagnostic accuracy and reduce the morbidity associated with Chagas disease.

1. Introduction

We participated in the 2025 George B. Moody PhysioNet Challenge, which invited teams to develop automated, open-source algorithms for identifying cases of Chagas disease from electrocardiograms (ECGs). Our submission focuses on a spectrogram-based convolutional neural network with a lead-aware attention mechanism to prioritize patients for confirmatory serological testing. All data were provided by the Challenge organizers and are described in detail in [1–3] and the associated datasets [4–6]. Here we outline our methodology, present the official Challenge results, and discuss the strengths and limitations of our approach.

2. Methods

2.1. Data Introduction

The Challenge used the CODE-15% dataset [4], the SaMi-Trop dataset [6], the PTB-XL dataset [5], and multiple private datasets from Chagas-endemic areas. The CODE-15% dataset contains over 300,000 12-lead ECG recordings collected in Brazil between 2010 and 2016, with self-reported binary Chagas labels. The SaMi-Trop dataset includes 1,631 validated Chagas-positive ECGs collected in Brazil between 2011 and 2012. The PTB-XL dataset provides 21,799 ECGs from non-Chagas patients in Europe between 1989 and 1996. All datasets differ in collection and validation procedures but together represent realistic clinical data. The training set combines CODE-15% with weak labels and SaMi-Trop and PTB-XL with strong labels, while validation and test sets consist of strongly labeled data, reflecting the prevalence rates in endemic countries. In contrast, the validation and test sets—used for official Challenge evaluation—consisted exclusively of unseen, strongly labeled data from endemic regions and were curated to reflect realistic Chagas prevalence rates. These unseen datasets formed the basis for score computation in the official and final phases of the Challenge, ensuring a robust assessment of model generalization and clinical relevance.

2.2. Data Preprocessing

Since the ECG recordings varied in sampling frequency and duration, we transform all signals to a fix samplerate of 400 Hz. The length is standardized to a length of 10 seconds using truncation or padding on both ends, resulting in 4,000 samples per lead. Each of the 12 leads was normalized using per-lead Z-score normalization, with noise injection applied to flat signals to avoid numerical instability. From the preprocessed signals, spectrogram repre-

sentations were extracted, primarily with the Short-Time Fourier Transform (STFT, window size of 256 samples, hop length of 48, Hann window), while Mel-frequency cepstral coefficients (MFCC) and Stockwell transforms were also evaluated as alternatives. This ensured a consistent and physiologically relevant time–frequency representation of the ECG data, as illustrated in Fig. 1.

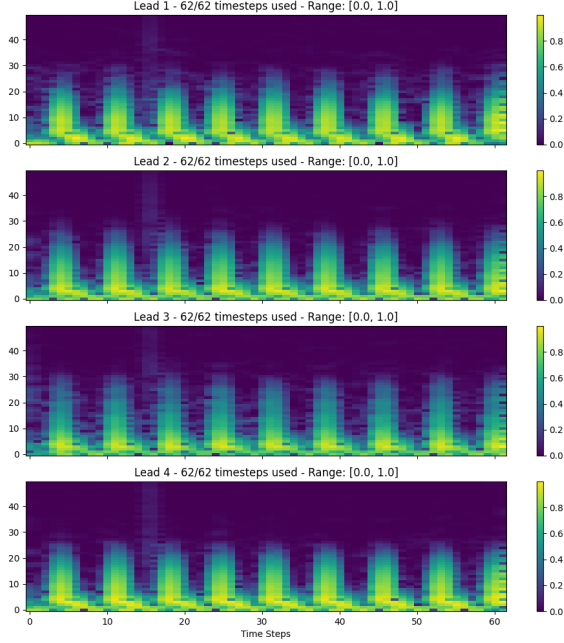


Figure 1. STFT features from four ECG leads, showing time (x-axis), coefficients (y-axis), and feature intensity (color), used for Chagas detection.

2.3. Data Augmentation

To mitigate overfitting and enhance model generalization, spectrogram-based augmentation techniques were applied. Specifically, frequency masking (up to 25 frequency bins with probability 0.4) and time masking (up to 30 frames with probability 0.4) were performed on the spectrograms during training. These augmentations simulated variability in signal acquisition and reduced dependency on narrow spectral features. In contrast, raw-signal augmentations such as time stretching, amplitude scaling, or Gaussian noise injection were disabled to preserve data consistency with the caching strategy.

2.4. Model Structure

The proposed architecture Fig. 2 is a spectrogram-based convolutional neural network enhanced with a lead-aware attention mechanism. Each ECG lead is first pro-

cessed independently through a dedicated feature extractor consisting of three convolutional blocks. These blocks combine convolutional layers with batch normalization, SiLU activations, and max-pooling operations, followed by an adaptive average pooling layer. This design captures local time–frequency characteristics of the spectrogram while progressively reducing dimensionality. The resulting per-lead feature maps are then passed to a lead attention module, which assigns an adaptive weight to each lead through a linear transformation and sigmoid activation. This mechanism emphasizes clinically relevant leads and down-weights noisy or less informative signals, improving both robustness and interpretability. The weighted features from all leads are concatenated and fed into a fully connected classification head. This module consists of dense layers with batch normalization, SiLU activations, and dropout regularization, ensuring effective hierarchical feature learning and reducing overfitting. The final output layer provides a binary prediction indicating the presence or absence of Chagas disease. Compared to conventional CNNs, the incorporation of lead-aware attention enables the model to handle missing or corrupted leads more effectively while highlighting leads strongly associated with Chagas-specific ECG abnormalities.

2.5. Training Details

The proposed model was trained using the *AdamW optimizer* [7] with an initial learning rate of 0.003 and no additional L1/L2 weight regularization. To stabilize convergence, a One-Cycle learning rate scheduler was employed, with a warm-up fraction of 0.3 and a minimum learning rate of 1×10^{-6} . The scheduler dynamically adjusted the learning rate to improve generalization and avoid premature convergence. For loss optimization, we adopted a composite loss function combining Focal Loss ($\gamma = 2.5$) weighted at 0.8 with a Top- k True Positive Rate (TPR) objective. This design directly aligns the training process with the Challenge evaluation metric, while also addressing the strong class imbalance inherent in Chagas datasets. A decision threshold of 0.5 was applied for binary classification during inference. The model was trained over 20 epochs with a batch size of 256. To evaluate model stability and generalization, we conducted 5-fold stratified cross-validation, preserving the natural distribution of Chagas-positive cases across folds. In each fold, 80 % of the data were used for training and 20 % for validation. All three datasets were incorporated during development to build a robust and generalizable model. To improve signal quality, ECG inputs were preprocessed with a Butterworth bandpass filter (0.5–45 Hz, order 2) and a notch filter at 50/60 Hz to suppress powerline interference. Gradient clipping at 0.7 was applied to prevent instability from exploding gradients.

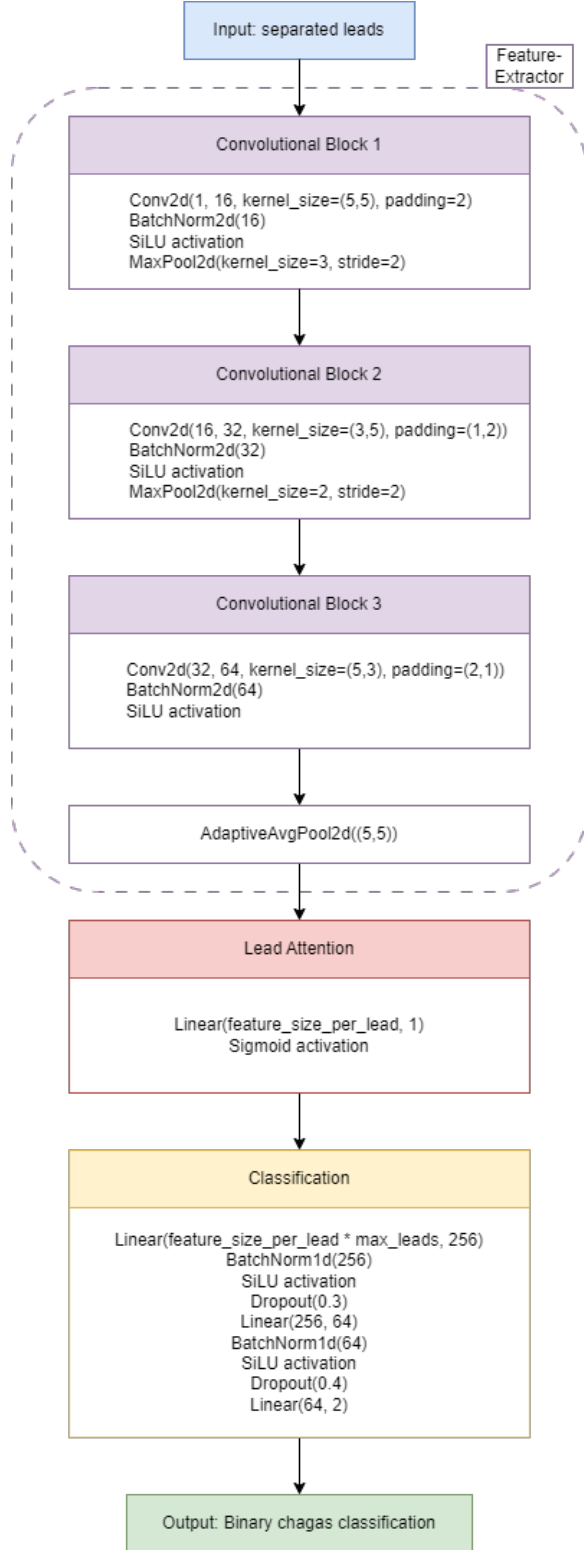


Figure 2. Architecture of the ECG-Spectrogram Classifier.

2.6. Evaluation Metric

The primary evaluation metric for the Chagas Challenge is the Challenge Score, which assesses algorithms' ability to prioritize patients for serological confirmation based on ECG data. In regions with limited testing capacity, this metric reflects the clinical utility of computational models by measuring their effectiveness in identifying high-risk individuals.

The Challenge Score is defined as the proportion of confirmed Chagas-positive patients ranked within the top 5 % of the patient cohort, emphasizing early and accurate detection to allocate limited tests to the most probable positive cases. Models are evaluated on a hidden test set, with the highest score determining the Challenge winner. Each algorithm assigns a probability of Chagas disease to every ECG record, and patients are ranked accordingly. The score is computed as the fraction of true positives in the top-ranked subset relative to the total number of positives. Ties are resolved randomly, and the expected score value is reported. This evaluation scheme optimizes algorithms for both classification accuracy and prioritization efficiency, essential for real-world clinical deployment in resource-constrained settings.

3. Results

Figure 3 illustrates the average ROC curve obtained from the validation set after 20 training epochs using 5-fold cross-validation. The mean curve (purple line) demonstrates consistent classification performance across folds, with the shaded area representing the standard deviation. The classifier achieves a mean Area Under the Curve (AUC) of 0.824, indicating strong discriminative ability between the two classes. The performance clearly surpasses the random baseline (red dashed line), which corresponds to an AUC of 0.5. In the context of the PhysioNet Challenge on Chagas disease detection, our proposed model achieved a training score of 0.509 on the training subset, and a validation score of 0.303 on a held-out portion of the development set. These results were obtained after training the model for 20 epochs using a fixed data split without cross-validation. The trained configuration was then submitted unchanged to the official challenge evaluation phase, where it achieved a test score of 0.199 on an unseen dataset provided by the organizers. These results demonstrate the model's ability to identify patients at high risk of Chagas disease using only ECG signals. Notably, it successfully ranked a substantial fraction of Chagas-positive individuals within the top 5% of the cohort, a capability particularly valuable in low-resource healthcare settings for prioritizing diagnostic testing. During the live challenge phase in Brazil, our model's test score of 0.199 placed team *DataNinjas* 24th out of 41 par-

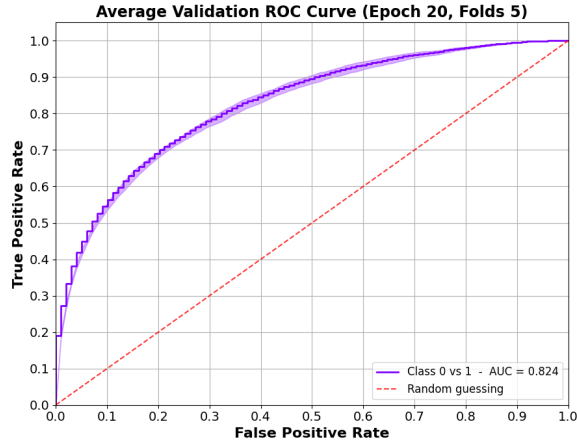


Figure 3. Average validation ROC curve after 20 epochs across 5 folds.

Training	Validation	Test	Ranking
0.509	0.303	0.199	24/41

Table 1. Challenge scores for our submitted model (team *DataNinjas*), including final ranking on the hidden test set.

ticipants. The overall performance across all stages of the challenge is summarized in Table 1.

4. Discussion

The performance gap between internal validation and official evaluation underscores the model’s sensitivity to distributional shifts in ECG data. While relevant features were successfully learned during training, generalization to unseen data remains limited—likely due to dataset homogeneity, class imbalance, and differences in acquisition protocols.

A further challenge stems from label quality. The CODE-15% dataset contains weak labels derived from self-reported Chagas status, which are less reliable than clinically validated diagnoses. In contrast, the official validation and test sets consisted exclusively of strongly labeled data. This mismatch likely impacted performance and illustrates the risk of relying on noisy supervision in medical applications. The lead-aware attention mechanism contributed to robustness by mitigating the impact of missing or corrupted leads and improving interpretability through adaptive lead weighting. However, generalization remains a bottleneck.

Future work should explore ECG-specific data augmentation, domain adaptation, and the integration of patient metadata. Increasing label quality and dataset diversity will be essential for reliable real-world deployment.

5. Conclusion

The proposed spectrogram-based CNN with lead-aware attention shows promising potential for ECG-based Chagas disease screening. While strong performance was achieved on internal data, generalization to unseen datasets remains challenging. Future efforts should target improved robustness through advanced data augmentation, domain adaptation, and incorporation of clinical metadata. Expanding diverse training data will be essential to enable reliable real-world application in resource-limited settings.

References

- [1] Reyna MA, Koscova Z, Pavlus J, Weigle J, Saghaei 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.
- [2] Reyna MA, Koscova Z, Pavlus J, Weigle J, Saghaei 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. *arXiv preprint* 2025; URL <https://arxiv.org/abs/2510.02202>.
- [3] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysiToolKit, and Physionet: Components of a new research resource for complex physiologic signals. *Circulation* 2000;101(23):e215–e220.
- [4] Ribeiro AH, Paixao GM, Lima EM, Horta Ribeiro M, Pinto Filho MM, Gomes PR, Oliveira DM, Meira Jr W, Schon TB, Ribeiro ALP. CODE-15%: a large scale annotated dataset of 12-lead ECGs, June 2021.
- [5] Wagner P, Strodthoff N, Bousselet RD, Kreiseler D, Lunze F, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset, 2020. URL <https://publica.fraunhofer.de/handle/publica/264664>.
- [6] Ribeiro ALP, Ribeiro AH, Paixao GM, Lima EM, Horta Ribeiro M, Pinto Filho MM, Gomes PR, Oliveira DM, Meira Jr W, Schon TB, Sabino EC. SaMi-Trop: 12-lead ECG traces with age and mortality annotations, June 2021. URL <https://doi.org/10.5281/zenodo.4905618>.
- [7] Loshchilov I, Hutter F. Decoupled weight decay regularization, 2019. URL <https://arxiv.org/abs/1711.05101>.

Address for correspondence:

Pinar Bisgin
 Speicherstraße 6, 44147 Dortmund, Germany
 pinar.bisgin@isst.fraunhofer.de