Classification of Chagas Disease Using Convolutional Neural Network and Random Forest

Noor Qaqos[1], Ekenedirichukwu N Obianom[2], Shamsu Idris Abdullahi[1], Fan Feng[1], Abdulhamed M Jasim[1], Zaid A Abod[3], Abdulmalik Koya[2], G André Ng[2,4], Xin Li[1,4]

[1]University of Leicester, School of Engineering, Leicester, UK
[2]University of Leicester, Department of Cardiovascular Sciences, Leicester, UK
[3]University of Leicester, School of Computing and Mathematical Sciences, Leicester, UK
[4]National Institute for Health Research Leicester, UK

Abstract

This paper presents a hybrid method for detecting the Chagas disease from the ECG signals in the George B. Moody PhysioNet Challenge 2025 by proposing two computational approaches, machine learning and deep learning. A multi-scale one-dimensional convolutional neural network (CNN) was developed, incorporating three convolutional blocks, global average pooling, and focal loss to address class imbalance, achieving robust binary classification across 12-lead ECGs normalized to 5000 samples. Additionally, a random forest (RF) classifier was trained on extracted features, including morphological and time-series attributes from P, QRS, and T waves, after denoising with a bandpass filter. Both methods were evaluated on a Physionet dataset, with the CNN demonstrating high accuracy and confidence calibration. These scalable techniques provide promising tools for automated Chagas disease detection, supporting precision cardiology and enhancing clinical outcomes. Moreover, it highlights the effectiveness of the CNN model in detecting Chagas disease using the focal loss function. Our team, Leicester Fox, had a rank of 18 and a challenge score of 0.258 on the validation set and 0.218 on the test set of data for an official phase using the CNN classification model.

1. Introduction

Chagas disease, caused by Trypanosoma cruzi, affects an estimated 6 to 7 million people worldwide. Approximately 30% of infected individuals develop chronic Chagas cardiomyopathy (CCC), which can lead to arrhythmias, heart failure, and sudden cardiac death [1]. Among these complications, atrial fibrillation (AF) is one of the most common and prognostically significant, being strongly associated with stroke and mortality in CCC patients [2]. Electrocardiography (ECG) remains the primary tool for CCC diagnosis due to its low cost and non-invasive nature, especially in resource-limited regions. However, its sensitivity for early-stage myocardial injury is suboptimal. Studies show that up to 40% of Chagas patients exhibit ECG abnormalities, including right bundle branch block and AF, even in early disease stages [3].

To address these limitations, this study proposes two automated ECG-based classification methods. The first is a multi-scale convolutional neural network (CNN) designed to extract hierarchical features directly from raw 12-lead signals. The second is a classical random forest (RF) model that leverages 29 handcrafted morphological and temporal ECG features for classification. Both models are designed to eliminate reliance on manual interpretation and to enhance scalability and objectivity.

Furthermore, we employed focal loss to address class imbalance and incorporated confidence calibration to improve model reliability in clinical settings [4, 5]. These strategies ensure that both models can deliver robust and reliable predictions, particularly when identifying Chagas-positive patients in large or imbalanced datasets.

Overall, these data-driven approaches offer scalable and practical tools for early Chagas detection. Their implementation could significantly benefit clinical workflows and public health screening, particularly in areas with limited diagnostic capacity [6].

2. Materials and Methodology

The dataset used in this study is derived from three resources (referred to as Code-15 [7], SaMi-Trop [8], and PTB-XL datasets [9]), which were utilized in the George B. Moody PhysioNet Challenge 2025 [10, 11]. In this work, we propose two methods: machine learning (based on feature extraction and a random forest model) and deep learning (based on 1D-CNN model) to classify Chagas and non-Chagas disease based on ECG signals. 30% of the data from each of the three resources was used to train the CNN model, while samples from the Code-15 and SaMi-Trop

datasets were used for both training and testing the random forest classifier. Figure 1 shows the pipeline of the proposed methods. These approaches are illustrated as follows:

2.1. Convolutional Neural Network

We developed a CNN for the automatic classification of ECGs. The architecture of the model is based on a multi-scale one-dimensional CNN. The architecture consists of three convolutional blocks followed by global average pooling and fully connected (FC) layers for binary classification. We implemented a balanced dataset loader that oversampled Chagas cases and standardized ECG signals across 12 leads by normalizing and interpolating them to a fixed length of 5000 samples, thereby addressing class imbalance. Moreover, a combination of focal loss was applied to mitigate the class imbalance problem, and a confidence calibration loss was used to boost reliable positive predictions. The Adam optimizer was used to train and test the CNN model, incorporating weight decay and a dynamic learning rate with an early stopping to prevent overfitting. Fine-tuning was performed using pre-trained weights that were loaded optionally from a previous submission.

2.2. Focal Loss Function

The focal loss (FL) function was used to address the problem of imbalanced data (Chagas vs. non-Chagas classes), as is the case in this challenge. This function helps the model pay more attention to the minority (Chagas) cases. Focal loss ensures the CNN model is not dominated by majority class signals during the learning process. Focal loss for binary classification is formulated as:

$$p_t = \begin{cases} p & if\ y = 1 \quad (chagas\ class) \\ 1-p & if\ y = 0\ (non-chagas\ class) \end{cases} \quad 1$$

$y \in \{0,1\}$ represents the ground truth label, $p \in [0,1]$ represents the predicted probability for positive class (Chagas), and p_t is the probability of the true class. The standard binary cross-entropy loss (BCE) is formulated as:

$$BCE(p_t) = -\log(p_t) \quad 2$$

The loss is small when the model predicts correctly with high confidence, and is large when the model predicts incorrectly. Therefore, focal loss modifies BCE algorithm by adding two factors to address the problem of imbalanced data. The factor $\alpha \in [0,1]$, which represents the class balancing factor for handling class imbalance explicitly by giving more weight to the minority class (chagas), and $\gamma \geq 0$, which represents the focusing parameter, which controls how much to down-weight easy cases.

$$FL(p_t) = -\alpha\,(1-p_t)^{\gamma}\log(p_t) \quad 3$$

For $\gamma = 0$, the FL reduces to BCE, and for γ larger than 0, it focuses more on hard or misclassified cases (Chagas class). Minority classes usually have low p_t, so $(1-p_t)^{\gamma}$ is large, making their loss contribution larger [12].

2.3. RF Classifier via Feature Extraction

The ECGs were denoised using a bandpass filter (0.5 to 50Hz). Time-series and morphological features were then extracted for training and evaluating purposes. A total of 31 features were extracted Lead II (4 features from patient details and other features extracted using morphology and time intervals of P, QRS, and T waves of each ECG signal). Table 1 shows the feature names that are used in this method.

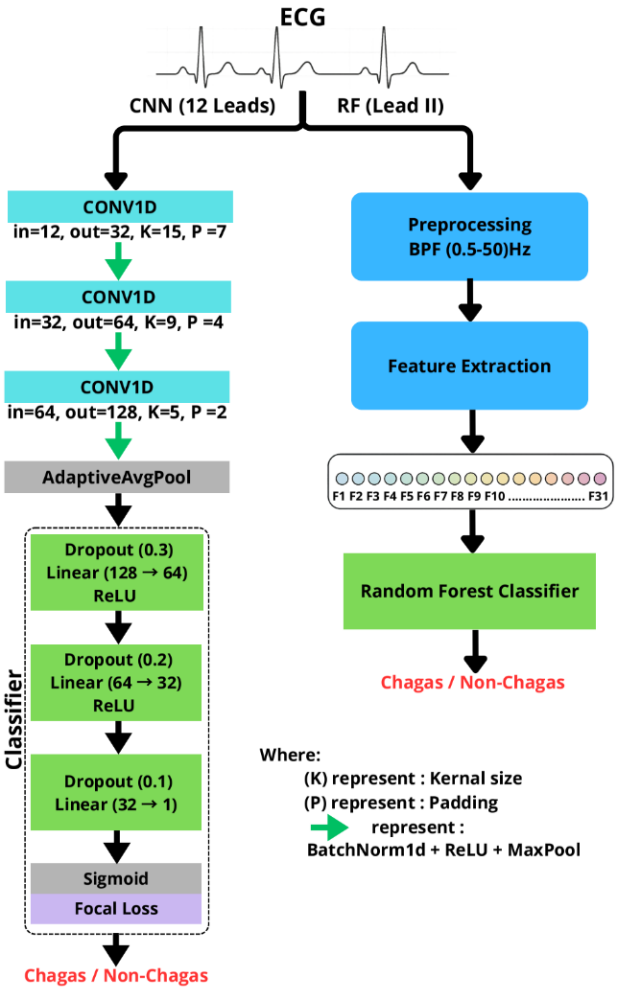

Figure 1. The pipeline of the proposed methods

Table 1. Features list extracted from ECG signals

Patient details (4 features)	P, QRS, T (15 features)	Time intervals and polarity (12 features)
Age, sex (3 features)	P average maximum amplitude P average minimum amplitude P average amplitude range P average duration P average energy QRS average maximum amplitude QRS average minimum amplitude QRS average amplitude range QRS average duration QRS average energy T average maximum amplitude T average min amplitude T average amplitude range T average duration T average energy	RR interval Heart rate PR interval QT interval QTC interval PR segment ST segment P presence T presence J wave presence P polarity T polarity

3. Results

Two methods were used to detect Chagas disease using ECG signals. The CNN model achieved the highest performance locally, with a challenge score of 0.45, an accuracy of 90.70%, an F-measure of 26.90%, an AUROC of 0.87, and an AUPRC of 0.223, as well as a challenge score of 0.258 on the validation set and 0.218 on test set of data on the official phase. In contrast, RF with a feature extraction method had lower performance locally, with a challenge score of 0.11 in the unofficial phase, an accuracy of 74.70%, an F-measure of 75.80%, an AUROC of 0.834, an AUPRC of 0.819, and an unofficial phase score of 0.11. Figure 2 shows the comparisons between the two methods locally. Table 2 illustrates the challenge score and the team rank for the official phase.

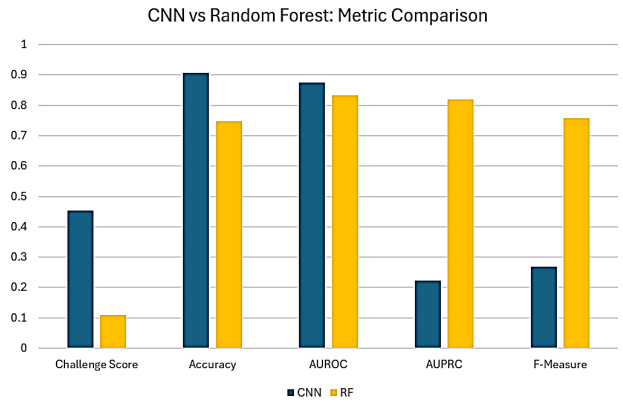

Figure 2. Results of the two models

Table 2. Official phase results for CNN model.

Task	Score on validation set	Score on test set	Rank
Classification	0.258	0.218	18/41

4. Discussion and Conclusions

This study compared two techniques (CNN and RF) in ECG feature extraction for chagas cardiomyopathy. The result demonstrates that CNN achieved a higher Challenge score (0.453 vs 0.110) and accuracy (0.907 vs 0.747) compared to RF though at a cost of longer training time (two weeks vs. one hour). This aligns with the Challenge primary objective of maximizing recall within the top 5% of predicted probabilities. This metric focuses on how many Chagas-positive cases are captured in the highest-ranked predictions, simulating real-world constraints where only a small portion of patients can be tested. The CNN's strong AUROC (0.874) further supports its ability to rank positive cases above negatives across the entire dataset. Our findings demonstrate the superior performance of CNNs in AUROC and accuracy, which is consistent with the proposal by Zubair et al. Zubair et al. proposed that CNNs are well-suited for ECG feature extraction due to their ability to learn hierarchical features from raw signals [13]. Similarly, Xu et al. have demonstrated 98% accuracy of deep residual networks in arrhythmia detection [14]. In contrast, we have shown RF's competitive performance in F-measure, reflecting its strength in handling class imbalances and providing stable predictions across heterogeneous datasets. However, due to heavy dependence on heavily engineered features, it may not capture the temporal complexity of ECG signals as effectively as CNNs. Similar to our report, prior comparative studies have shown that CNN-based methods outperform traditional machine learning models, including RF, in ECG classification [15]. Although there is a dearth of studies specifically on using these models for Chagas diagnosis, our results show that the RF model displayed a much better ability to recognize positive cases overall. Its AUPRC of 0.819 and F-measure of 0.758 show that RF consistently identifies Chagas-positive patients and maintains a strong precision-recall balance. Although its AUROC (0.834) was slightly lower than that of CNN, and its Challenge score (0.11) fell well below that of CNN, these results suggest that RF would be more effective in scenarios where more than 5% of patients can be tested or where a balanced approach to detection is required. Another key consideration is computational efficiency. The CNN required approximately two weeks to train. In

contrast, the Random Forest completed training in about one hour, making RF a much more practical choice for rapid deployment and iterative development.

In conclusion our findings have some limitations, the Challenge score represents performance only at the extreme top of the ranking, not across the entire probability spectrum. When broader performance metrics are considered, the CNN performed poorly in consistently identifying Chagas-positive cases outside the top 5%. Its AUPRC (0.223) and F-measure (0.269) are both very low, indicating that while the CNN assigns high probabilities when it does identify positives, it misses many others and struggles to balance precision and recall. The CNN is optimal for the Challenge because the evaluation metric emphasizes ranking positives in a tiny segment of the population. However, in real-world applications where the goal is to detect as many positive cases as possible across a broader range of thresholds, the Random Forest model may be the more reliable and efficient choice. These results are based on a controlled dataset within the data provided. When the CNN is tested on the unseen dataset, it performs worse in terms of the challenge score, indicating the need for further research into a more effective pipeline or model for classifying Chagas cases based on ECG. Future work could explore hybrid approaches or ranking-optimized learning methods that combine CNN's strong top-tier ranking performance with RF's ability to maintain precision and recall across the full distribution.

Acknowledgments

We acknowledge our sponsors (1) Higher Committee for Education Development, Iraq, (2) Petroleum Technology Development Fund (PTDF), Nigeria, (3) Future100 Scholarship Scheme, University of Leicester, UK, (4) The British Heart Foundation.

References

[1] A. Silvestre de Sousa, F. de Souza Nogueira Sardinha Mendes, P. Jordà, and A. García-Álvarez, "Diagnosis of Chagas disease: are clinical definitions of heart involvement accurate enough?, " *Chagas Disease: A Neglected Tropical Disease*, pp. 95-106: Springer, 2020.

[2] P. Benchimol-Barbosa, and J. Barbosa-Filho, "Mechanical cardiac remodeling and new-onset atrial fibrillation in long-term follow-up of subjects with chronic Chagas' disease," *Brazilian Journal of Medical and Biological Research,* vol. 42, pp. 251-262, 2009.

[3] L. Z. Rojas, M. Glisic, L. Pletsch-Borba, L. E. Echeverría, W. M. Bramer, A. Bano, N. Stringa, A. Zaciragic, B. Kraja, and E. Asllanaj, "Electrocardiographic abnormalities in Chagas disease in the general population: a systematic review and meta-analysis," *PLoS neglected Tropical Diseases,* vol. 12, no. 6, pp. e0006567, 2018.

[4] A. M. Hasslocher-Moreno, R. M. Saraiva, T. L. d. Silva Júnior, S. S. Xavier, and A. S. d. Sousa, "Exploring the historical background and clinical implications of electrocardiogram in the context of chagas disease research," *Revista da Sociedade Brasileira de Medicina Tropical,* vol. 56, pp. e0506-2023, 2023.

[5] A. L. P. Ribeiro, M. S. Marcolino, R. J. Prineas, and M. F. Lima-Costa, "Electrocardiographic abnormalities in elderly Chagas disease patients: 10-year follow-up of the Bambui cohort study of aging," *Journal of the American Heart Association,* vol. 3, no. 1, pp. e000632, 2014.

[6] S. Kann, G. A. C. Mendoza, M. Hartmann, H. Frickmann, and L. Kreienbrock, "Chagas disease: medical and ECG related findings in an indigenous population in Colombia," *Tropical Medicine and Infectious Disease,* vol. 8, no. 6, pp. 297, 2023.

[7] A. H. Ribeiro, M. H. Ribeiro, G. M. Paixão, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. Ferreira, C. R. Andersson, P. W. Macfarlane, and W. Meira Jr, "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nature Communications,* vol. 11, no. 1, pp. 1760, 2020.

[8] C. S. Cardoso, E. C. Sabino, C. D. L. Oliveira, L. C. de Oliveira, A. M. Ferreira, E. Cunha-Neto, A. L. Bierrenbach, J. E. Ferreira, D. S. A. Haikal, and A. L. Reingold, "Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile," *British Medical Journal Open,* vol. 6, no. 5, pp. e011181, 2016.

[9] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "PTB-XL, a large publicly available electrocardiography dataset," *Scientific Data,* vol. 7, no. 1, pp. 1-15, 2020.

[10] M. A. Reyna, Z. Koscova, J. Pavlus, S. Saghafi, J. Weigle, A. Elola, S. Seyedi, K. Campbell, Q. Li, A. B. Rad, A. H. Ribeiro, A. L. P. Ribeiro, R. Sameni, and G. D. Clifford, "Detection of chagas disease from the ECG: The George B. Moody Physionet Challenge 2025," *arXiv preprint arXiv:2510.02202*, 2025.

[11] M. A. Reyna, Z. Koscova, J. Pavlus, S. Saghafi, J. Weigle, A. Elola, S. Seyedi, K. Campbell, Q. Li, A. B. Rad, A. H. Ribeiro, A. L. P. Ribeiro, R. Sameni, and G. D. Clifford, "Detection of Chagas disease from the ECG: The George B. Moody PhysioNet Challenge 2025," *Computing in Cardiology Conference,* vol. 52, pp. 1-4, 2025.

[12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection. " pp. 2980-2988.

[13] S. Nurmaini, R. Umi Partan, W. Caesarendra, T. Dewi, M. Naufal Rahmatullah, A. Darmawahyuni, V. Bhayyu, and F. Firdaus, "An automated ECG beat classification system using deep neural networks with an unsupervised feature extraction technique," *Applied Sciences,* vol. 9, no. 14, pp. 2921, 2019.

[14] X. Xu, and H. Liu, "ECG heartbeat classification using convolutional neural networks," *IEEE Access,* vol. 8, pp. 8614-8619, 2020.

[15] S. Reznichenko, J. Whitaker, Z. Ni, and S. Zhou, "Comparing ECG lead subsets for heart arrhythmia/ECG pattern classification: convolutional neural networks and random forest," *Canadian Journal of Cardiology Open,* vol. 7, no. 2, pp. 176-186, 2025.

Address for correspondence:

Noor Qaqos, Xin Li
University of Leicester, UK, LE1 7RH
nnq2@leicester.ac.uk; xin.li@leicester.ac.uk