Deep Learning Approach for Binary Classification of 12-Lead ECGs in Chagas Disease Detection

José-Adrián Chávez-Olvera[1], Arián Villalba-Tapia[2], Blanca Tovar-Corona[2], René Luna-García[1]

[1]Centro de Investigación en Computación, Mexico City, México
[2]Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas, Mexico City, México

Abstract

Chagas disease is a condition that can cause significant cardiac complications. Early detection through automated electrocardiogram (ECG) analysis may facilitate timely intervention and treatment. In this work, we present a deep learning pipeline for the binary classification of 12-lead ECG recordings into Chagas and non-Chagas categories. All recordings were resampled to 100 Hz, when necessary, and bandpass filtered to remove baseline drift and high-frequency noise. Each lead was decomposed using the Discrete Wavelet Transform (DWT) up to level 4 with a Daubechies 3 mother wavelet. The approximation and detail coefficients from the fourth level were concatenated and used as input to a one-dimensional Convolutional Neural Network (1D CNN) designed to capture local spatial and temporal patterns across the multi-lead input.

The team was unable to be scored on the hidden data; therefore, no official score or ranking is reported to avoid misleading interpretation of the model's performance. Despite this limitation, the proposed approach demonstrates potential for improving automatic detection of Chagas-related cardiac abnormalities in future studies.

1. Introduction

The detection of Chagas disease using 12-lead ECG signals is a critical task in early diagnosis and patient care. In this work, we present a deep learning approach for binary classification of 12-lead ECGs, detailing our methodology and contributions within the Challenge [1,2].

1.1. Chagas Disease

Chagas disease, caused by the protozoan Trypanosoma cruzi, is a complex parasitic zoonosis that remains a significant public health challenge in the Americas. It is primarily transmitted by over 150 species of triatomine insects, as well as through congenital transmission, contaminated food, blood transfusions, and organ transplants. Despite progress in controlling non-native vectors and blood transmission, native vectors maintain active transmission cycles. Emerging issues such as oral transmission and vertical infection, combined with underreporting and limited healthcare access, hinder accurate burden assessment. Addressing Chagas disease requires sustained, adaptive strategies that integrate vector control, improved case detection, and comprehensive patient care [3, 4].

Chagas disease is a leading cause of infectious cardiomyopathy worldwide, with approximately 1.2 million people affected and high associated mortality [5]. Although it remains endemic in Latin America, its global relevance has increased. The disease has an acute phase, often presenting with nonspecific viral-like symptoms, and a chronic phase, typically asymptomatic but potentially progressing to severe cardiac complications.

Chagas cardiomyopathy represents the most severe clinical manifestation of the disease and is characterized by a wide range of electrocardiographic (ECG) abnormalities [6]. Heart failure is a common and serious clinical outcome of Chagas cardiomyopathy and is associated with a worse prognosis compared to other etiologies [7]. Echocardiography is the mainstay imaging modality for assessing cardiac function, guiding management, and stratifying patient risk [8].

Detecting patients with Chagas disease through ECG analysis remains a significant challenge due to the complexity and variability of the arrhythmias associated with the condition. The disease often leads to structural cardiac abnormalities, including progressive enlargement of the heart (cardiomegaly), which impairs its ability to pump blood efficiently and contributes to the development of heart failure.

2. Methods

2.1. Wavelet Transform

The wavelet transform is a signal processing technique used to analyze localized variations within a signal by shifting and scaling a basic function known as the mother wavelet (Ψ). It enables efficient evaluation of transient or small-scale features and is widely applied in areas such as

data compression, noise reduction, and pattern recognition. Two main types of wavelet transforms are commonly used: the Continuous Wavelet Transform (CWT), which analyzes signals at various scales to capture detailed frequency components, and the Discrete Wavelet Transform (DWT), which provides a more computationally efficient, multiresolution representation of signals [9].

Unlike the Fourier and short-time Fourier transforms, which have limited time localization due to their fixed window size, the wavelet transform provides adaptive multiresolution analysis, allowing precise detection of transient events and frequency changes [10].

Daubechies wavelets, denoted as dbN, are among the most used due to their ability to represent complex patterns efficiently. Higher-order Daubechies wavelets capture more intricate signal characteristics, making them especially valuable in medical signal processing such as ECG analysis [11].

2.2. Signal Filtering

In biological signal processing, filters are essential for enhancing signal quality and enabling accurate analysis. ECG signals are typically affected by noise, and filtering helps isolate relevant frequency components while suppressing unwanted artifacts. Filters are used to isolate the relevant frequency components of these signals, enhancing meaningful physiological information while suppressing noise and artifacts.

Common types include low-pass, high-pass, and band-pass filters, which remove high-frequency noise, low-frequency drift, or focus on specific frequency bands, respectively, as in EEG analysis of alpha, beta, and theta rhythms.

Filters can be implemented using analog or digital techniques, with digital filters providing greater flexibility and precision in modern applications.

2.3. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are computational models inspired by the animal visual cortex, designed to process structured data such as images [12]. They consist of convolutional and fully connected layers, where convolutional kernels automatically extract relevant features. These kernels are optimized during training using backpropagation and gradient descent to minimize prediction errors.

CNN architectures have been successfully applied to cardiac signal analysis. One-dimensional CNNs can jointly perform feature extraction and classification for ECG signals, achieving high accuracy rates of up to 96.72% in patient-specific classification tasks due to their ability to learn highly discriminative temporal patterns [13].

2.4. Database

The study utilizes three publicly available 12-lead ECG datasets included in the Challenge training set: CODE-15% [14], SaMi-Trop [15], and PTB-XL [16]. The CODE-15% dataset, collected in Brazil, contains over 300,000 ECG recordings, while the SaMi-Trop dataset includes 1,631 ECGs from Chagas patients, all clinically confirmed by serological tests. Both datasets feature signals of 7.3–10.2 seconds duration sampled at 400 Hz. The PTB-XL dataset comprises 21,799 ECG recordings of 10 seconds at 500 Hz.

2.5. Preprocessing of the ECG Signals

A band-pass Butterworth filter with a frequency range of 0.5–150 Hz was implemented to suppress noise and remove unwanted components that could distort relevant features of the biological signals. After filtering, the Discrete Wavelet Transform (DWT) was applied using a db3 mother wavelet, producing three detail coefficients and one approximation coefficient per signal. This process was performed independently for each of the 12 ECG channels. The detail coefficients from all channels were then concatenated to form matrices corresponding to each detail level, capturing relevant temporal–frequency information for subsequent model input. Figure 1 illustrates the original ECG signal, its filtered version, and the three corresponding wavelet detail levels.

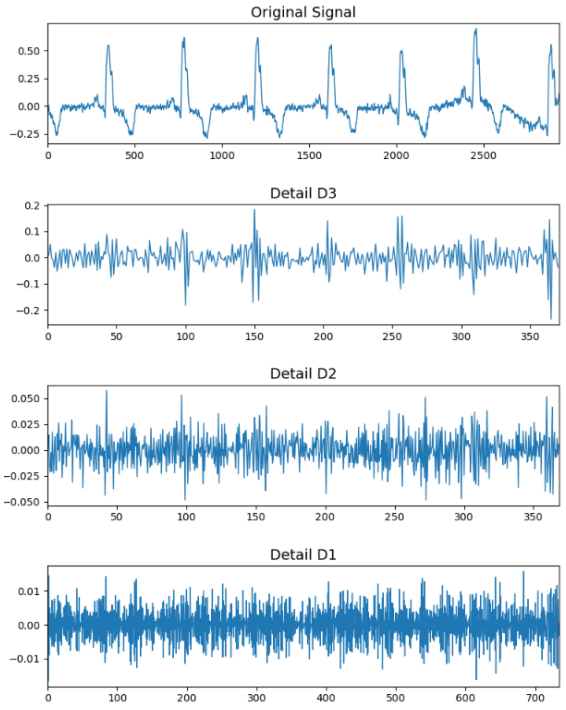


Fig 1. Filtered signal with its corresponding three detail components and one approximation, obtained using the discrete wavelet transform with a Daubechies 3 (db3) wavelet.

2.6. Propose Model

We propose a deep learning architecture based on 1D Convolutional Neural Networks (CNNs) designed to process 3-lead electrocardiogram (ECG) signals. The model takes as input multichannel ECG data with a fixed signal length (e.g., 4096 samples) and learns to extract meaningful temporal features for binary classification tasks, such as detecting the presence of Chagas disease.

The model architecture consists of parallel convolutional branches, one for each detail (3 in total). Each branch receives a one-dimensional input signal passes it through two convolutional blocks. The first block applies a 1D convolution with 32 filters, a kernel size of 3, ReLU activation, and same padding, followed by max pooling with a pool size of 2. The second block increases the number of filters to 64, maintaining the same kernel size and activation, and again applies max pooling. This structure allows each branch to capture local temporal features specific to its corresponding lead.

After processing all 3 channels independently, the outputs from each branch are concatenated into a single merged feature map. This merged representation is then flattened and passed through a fully connected dense layer with 64 units and ReLU activation. The final layer is a single-unit dense layer with sigmoid activation, outputting the probability of the target class.

By using a multi-branch design, the model captures lead-specific patterns while also learning cross-lead interactions after merging. This modular architecture is well-suited for patient-level ECG classification tasks and enables scalable adaptation to input signals with varying numbers of leads. As we can see in Fig 2.

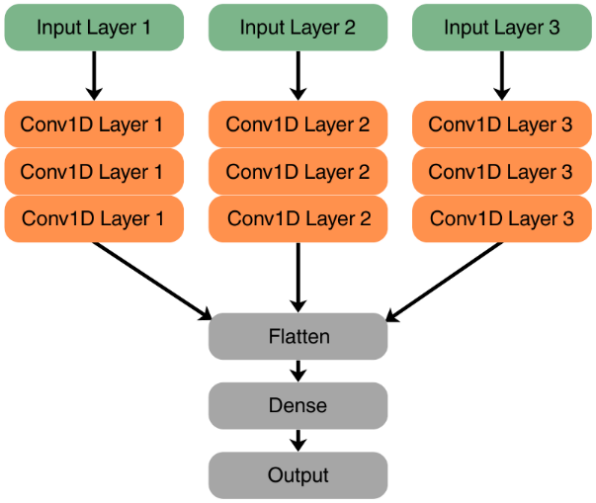


Fig 2. Architecture of the proposed multi-branch 1D Convolutional Neural Network for 3-lead ECG signal classification. Each lead is processed independently through two convolutional and pooling layers. The outputs are then concatenated, flattened, and passed through dense layers to perform binary classification.

3. Results

To evaluate the performance of the proposed model, the dataset was first preprocessed and all ECG recordings from the different sources were combined into a single dataset. The signals were appropriately labeled according to the presence or absence of Chagas disease. This unified dataset was then divided into training and testing subsets, ensuring a proper balance between classes. The model was trained using the training (Sami-Trop and PTB-XL) subset for 50 epochs and 10% of the data for validation, allowing it to learn and refine the lead-specific features necessary for accurate classification. After training, the model was evaluated on the test subset (5% of the CODE-15 data) to assess its generalization capability and classification performance. The results were obtained in the unofficial phase.

Table 1. Table summarizes the results obtained.

Challenge score	AUROC	AUPRC	Accuracy	F-measure
0.050	0.510	0.021	0.742	0.038

The experimental results of the proposed model are summarized in the following figures. Figure 3 presents the training and validation accuracy curves. Figure 4 displays the training and validation loss curves. Figure 5 shows the confusion matrix.

The proposed approach demonstrates that improvement is necessary in calculating the characteristics that are given to the model, changing the mother wavelet, or using parallel networks for approximation and details.

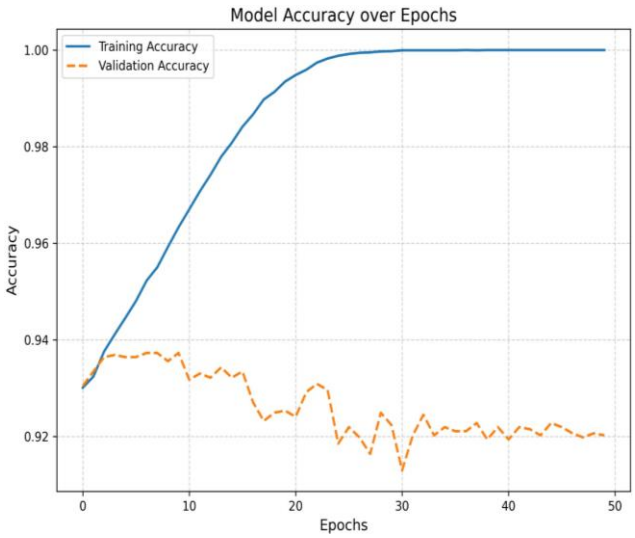


Fig 3. Training and validation accuracy and loss curves, which indicate the learning progress of the model across epochs.

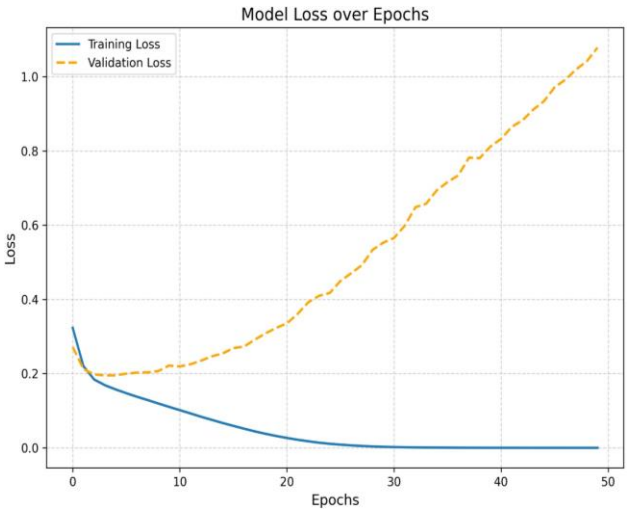


Fig 4. Training and validation loss curves, illustrating the convergence behavior during optimization and the degree of generalization achieved by the model.

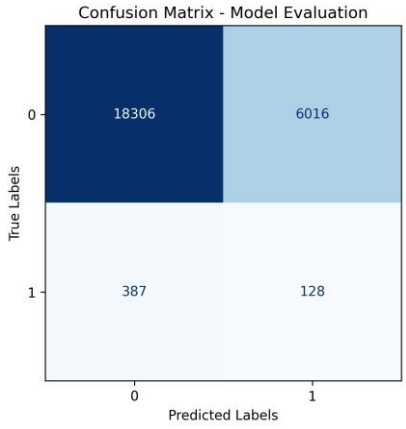


Fig 5. shows the confusion matrix, providing a detailed view of the classification performance and identifying the classes where misclassifications are more frequent. In this matrix, class 0 corresponds to False (non-Chagas) cases, while class 1 corresponds to True (Chagas) cases.

References

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

[2] Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., and Stanley, H. E., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," Circulation, vol. 101, no. 23, pp. e215–e220, 2000.

[3] Michael Tibayrenc and J. Telleria, American Trypanosomiasis: Chagas Disease One Hundred Years of Research. Elsevier, 2010.

[4] Veeranoot Nissapatorn and H. S. Oz, Chagas disease: Basic investigations and challenges. London: IntechOpen, 2018.

[5] C. Bern, "Chagas' disease," New England Journal of Medicine, vol. 373, no. 5, pp. 456–466, Jul. 2015.

[6] A. L. Ribeiro et al., "Electrocardiographic abnormalities in Trypanosoma cruzi seropositive and seronegative former blood donors," PLoS Neglected Tropical Diseases, vol. 7, no. 2, pp. e2078–e2078, Feb. 2013.

[7] M. M. Barbosa, A. Luiz, M. Amorim, and O. Costa, "Predictors of mortality in patients with dilated cardiomyopathy: Relevance of Chagas disease as an etiological factor," Revista Española de Cardiología (English Edition), vol. 63, no. 7, pp. 788–797, Jan. 2010.

[8] M. C. P. Nunes et al., "Multimodality imaging evaluation of Chagas disease: An expert consensus of Brazilian Cardiovascular Imaging Department (DIC) and the European Association of Cardiovascular Imaging (EACVI)," European Heart Journal – Cardiovascular Imaging, vol. 19, no. 4, pp. 459–460, May 2017.

[9] M. G. M. Milani, P. E. Abas, and L. C. De Silva, "A critical review of heart sound signal segmentation algorithms," Smart Health, vol. 24, p. 100283, Apr. 2022.

[10] A. P. Dhawan, "Wavelet transform and its applications in medical image analysis," in World Scientific eBooks, 2008, pp. 437–454.

[11] G. Lindfield and J. Penny, "Analyzing data using discrete transforms," in Elsevier eBooks, 2018, pp. 383–431.

[12] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," Biological Cybernetics, vol. 36, pp. 193–202, 1980.

[13] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 664–675, Aug. 2015.

[14] A. H. Ribeiro et al., "CODE-15%: A large scale annotated dataset of 12-lead ECGs," Zenodo (CERN European Organization for Nuclear Research), Jun. 9, 2021.

[15] A. L. P. Ribeiro et al., "Sami-Trop: 12-lead ECG traces with age and mortality annotations," Zenodo (CERN European Organization for Nuclear Research), Jun. 7, 2021.

[16] P. Wagner et al., "PTB-XL: A large publicly available electrocardiography dataset," Scientific Data, vol. 7, no. 1, May 2020.

Address for correspondence:

José-Adrián Chávez-Olvera
Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico City, México.
jchavezo1603@alumno.ipn.mx