Using a TensorFlowLite LSTM Deep Learning Model to Screen for Chagas Disease

Zoe E Boysen[1], Ron Yang[1], Emilia V Fallman[1]

[1]Ephemeris Labs, United States of America

Abstract

As part of the George B. Moody PhysioNet Challenge 2025 under the team name Ephemeris Labs, we developed a deep learning algorithm to detect late-stage Chagas disease in electrocardiogram signals (ECG). The use of deep learning-based models is a rapidly growing research topic in the medical field for classification and diagnostics. In addition, equipment is not designed for low-cost materials, open source code, and data sets to ensure better accessibility. We create a long- and short-term memory network to tackle this problem using the TensorFlow library on the George B. Moody PhysioNet Challenge open-source dataset. Under the challenge, our model scored 0.05 (ranked 348 out of 374). In our development outside the official scoring, we refined our model for live classification on a Raspberry Pi 4 and an evaluation of our TensorFlow Model and TensorFlow Lite FlatBuffers to demonstrate their minimal run-time requirements. By focusing on open-source and portable solutions, our approach offers a scalable method prioritizing the testing and treatment of Chagas disease, particularly in resource-limited settings.

1. Introduction

Individuals with late-stage Chagas Disease will often develop some type of cardiomyopathy, which can lead to heart failure and increase the likelihood of death. The George B. Moody 2025 Challenge invited teams to develop automated, open-source algorithms for identifying cases of Chagas disease from electrocardiograms (ECG) [1–4]. ECGs remain a gold standard for measuring heartbeats and diagnosing arrhythmia. They are applicable in assessing the severity of a disease, monitoring patients in clinical trials, and screening individuals for high-risk occupations [5].

In the following sections, we highlight the development of a Long Short-Term Memory (LSTM) deep-learning framework to classify Chagas disease in a patient's ECG dataset. With further development, the TensorFlow Lite model could be an important tool in arrhythmia detection and diagnosis on an embedded system.

2. Methods

The goal of this study is to build a working model and workflow of an ECG arrhythmia classification on a Raspberry Pi and Arduino system. We present a TensorFlow and TensorFlow Lite LSTM models able to label if a patient's ECG signal shows signs of Chagas disease or not from the provided datasets of the challenge [6–10]. Each data set contains a padded raw 12-lead ECG single (DI, DII, DIII, AVR, AVL, AVF, V1, V2, V3, V4, V5, V6) with a sampling frequency of 400 Hz and duration of 10s/7s. TensorFlow's LSTM method was chosen for its benefits in time series forecasting and anomaly detection. We utilized the provided data to build a dataset of equal parts Chagas and non-Chagas ECG samples, passing in data with the shape (length of processed signal, number of ECG leads).

The features used in the model are the sex and approximate age of the patient, and a raw ECG 12-lead signal. The sex is represented by a one-hot encoded 1 by 3 matrix of whether the patient is female, male, or other. Since raw ECG signals often have missing data or corrupted leads, we check for finite signals and perform some signal processing before passing them into the model.

Furthermore, our model can be converted to a form used on microcontroller systems, such as the Raspberry Pi Pico, for future implementations of a low-cost device capable of flagging irregular signals. All materials and methods are open to the public through GitHub. In the following, we highlight the methods for our official competition submission and the updated version submitted for the hackathon.

2.1. Official Competition Methods

In addition to the features extracted above, the 12-lead signal is processed into a 1 by 2 matrix. Across all 12 leads, the mean and standard deviations are taken as a form of processing to help with training speed and detection of patterns.

2.1.1. Model Architecture

Our model contains 3 LSTM layers and one dense layer, shown in Figure 1.

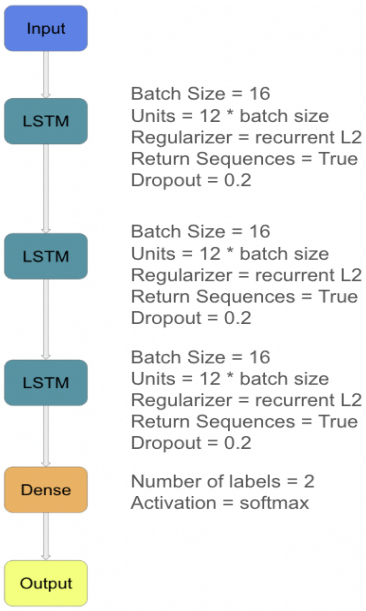


Figure 1. LSTM Model Architecture for the detection of Chagas disease

The LSTM layers run with a batch size of 16, a number divisible by the training, validation, and testing data set lengths. It also has an input shape of a maximum data length of 1 and a dropout of 0.2, common to prevent overfitting by randomly deactivating 20% of neurons during training. Each of these layers returns a set of vectors with a dimension of the batch size. The final layer of the model is a dense layer that fully connects the previous layers and changes the output dimension to match the number of units specified [15]. As shown in Figure 1, the dimensions of each layer are present. This illustrates that the input shape is 3-dimensional and the final output is a 2-dimensional representation of whether the signal indicates the presence of Chagas disease.

The input shape is given by a batch size of 16 (a number divisible by the length of the signal), the time step of 7, and the input dimension, number of features of 1. The time step was determined by the features extracted ([age], one-hot encoding sex, [signal mean, signal standard deviation]) and is the length of the longest ECG signal. These values influence the sizes of the subsequent layers, where the final layer outputs a shape of batch size and number of labels (2).

Within each LSTM layer, there are 192 LSTM cells- the value of time steps (t)- that run the specific functions and matrix math of the model. This is exhibited in Figure 1. We used the standard sigmoid functions and tanh activation functions as they are the most commonly used. This activation function is common for applications like this and can help with the common gradient issue. To capture the full complexity of the dataset, the units value was set to 192. This increases the number of neurons at each gate's functions within each LSTM cell. This increases the number of parameters at each layer, making the overall model more complex.

The model ran with L2 recurrent regularization of value 0.01 to help with overfitting. In testing, we saw that adding the typical ReLU activation layer caused an issue of the loss equaling infinity- exploding gradient problem, so we stuck with the tanh function. The model ran for 10 epochs for training. Upon compilation, the sparse categorical cross-entropy loss function and Adam optimizer were used. Both are typical methods for LSTM networks, but this particular loss function helps with numerical and multiple matching.

2.2. Hackathon Methods

To improve the model, we pulled data from a more balanced data set of equal parts Chagas and no Chagas. To address any memory allocation problems, the input features and labels are passed into the built-in TensorFlow function to build a dataset and slice it across the first dimension. Each slice becomes an element of the resulting dataset. In contrast to the previous method, to increase model speed, the number of units was reduced to 96.

In addition, instead of taking the mean and standard deviation of all leads, we take them on each lead. Further signal processing, such as padding to the largest signal, down-sampling by a factor of 4, performing a sym 3 wavelet transfer, and z-score normalization, was also taken to improve visibility and bring our key features related to Chagas detection.

2.2.1. Denoise

Wavelet transforms are commonly used in signal processing to denoise signals, thereby improving signal visualization and readability, especially in arrhythmia detection [11]. From previous work, researchers have found that daubechies-3 and symlets-3 wavelet transforms performed best in removing low-frequency baseline wanders [12]. These instances are commonly caused by patient movements(breathing, body motion, etc.) as well as environmental noise [13]. In our work, we implemented a symlets-3 wavelet transform as our primary method for denoising.

2.2.2. Z-score normalization

We implemented a z-score normalization across all 12 leads after applying the wavelet transform[14], as seen in Figure 2. The figure shows the ECG signal after processing. Looking at equation 1, it describes the standard z-

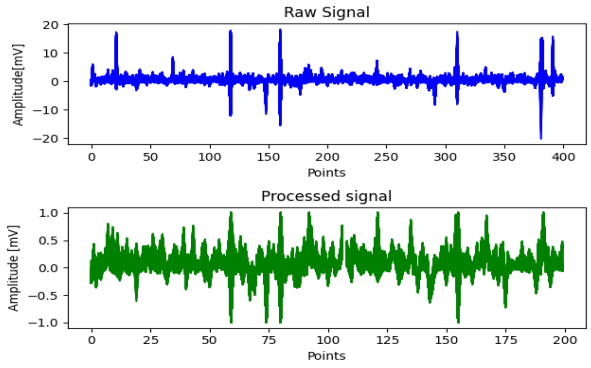


Figure 2. Example ECG signal where the top image shows the raw signal and the bottom image is after symlet-3 wavelet transform and z-score normalization

score, it normalizes each signal value by its mean and standard deviation. Here, x is the observed value, μ is the mean of the sample, and σ is the standard deviation of the sample.

$$Z = \frac{x - \mu}{\sigma} \tag{1}$$

This is performed on each sample (patient) to eliminate any patient-to-patient deviations. One study found that ECG wave amplitude is different based on ethnic origin [15]. Another study found that ECG amplitude can be lower in obese individuals [16]. The z-score normalization step helps in eliminating signal deviations based on individualistic variables.

2.3. Model Output

The model output comes from the final layer in our LSTM model architecture. We used a dense layer with softmax activation in order to obtain a binary output with the predicted probability of each label.

2.4. TensorFlow Lite

Our model, is built as TensorFlow, but for future applications, we convert it to a TensorFlow Lite for model size reduction, as seen in the following Table 1. The compression of the model is pertinent for an embedded system, model performance over a live ECG signal, and faster computations on the CPU. In order to integrate into a live running model on embedded Linux-based systems, the model must be small enough to work with limited memory. The

TensorFlow	TensorFlow Lite
453.76	155

Table 1. Comparison of model size (KB) between a normal TensorFlow model and the condensed Lite version

Raspberry Pi 5 and Pico boards are limited to RAM of 16 GB and 264 KB, and a GPU of 800 MHz and flash storage of 2 MB, respectively. Thus, converting our model to TensorFlow Lite would be a potential option for incorporation on these devices.

3. Results

In this section, we present the results of our model for the detection of Chagas disease for both methods highlighted above. Using LSTM, we were placed 342nd in the official phase of the George B. Moody Challenge. We received a final score of 0.05 as shown in Table 2, where the score represents the true positive rate out of the total number of Chagas cases.

Training	Validation	Test	Ranking
0	0.05	0.05	348/374

Table 2. Challenge scores for our selected entry (Ephemeris Labs)

Our hackathon submission was ranked 10th out of 10 submissions with a score of 0.05 for the test set. The performance of each method is illustrated in the following figure, containing the accuracy and F-measure for each dataset.

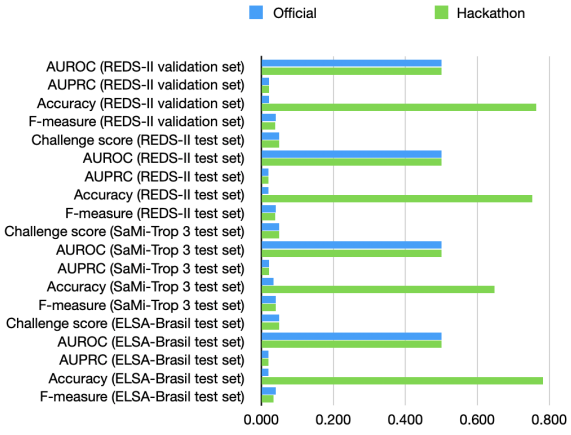


Figure 3. Official Vs Hackathon Results

4. Discussion and Conclusions

In this paper, an LSTM-based classification algorithm was proposed to label a 12-lead ECG signal as having Chagas disease or not. With our model score of 0.05, there is an improvement needed for this model to successfully identify Chagas.

As shown in Figure 3, our model's accuracy improved greatly, but the F-measure still remained low. This indicates an imbalance between precision and recall, meaning the model isn't great at identifying Chagas while minimizing false positives and false negatives. In future work there is a need to examine the data sets further, as an imbalance could be a likely cause. It may also be beneficial to alter the model parameters to best suit the input. This type of deep learning model, while effective for sequential data such as the ECG signals it is very sensitive to the parameters. As LSTM has proven to be effective in handling signals well in previous studies, we expect LSTM models can be successful for ECG signal classification, but further refinement will be needed on this model for successful detection.

While our model didn't perform as originally expected, we highlight the importance of a TensorFLow Lite Model. We were able to reduce the model size to about 1/3 of its original size, from 453.76 KB to 155 KB. This demonstrates a potential for the model to run live or record signals on a Raspberry Pi and Arduino device. In this process, we wanted to thoroughly test the model for its use on a low-cost, simple computing device. The end program could take in any ECG signal and flag abnormalities on a wearable device. This would be accessible to most and easy to use, a great advantage for the medical field.

Acknowledgments

We declare no conflicts of interests and want to thank all the mentors for this competition journey.

References

[1] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[2] Reyna MA, Koscova Z, Pavlus J, Weigle J, Saghafi S, Gomes P, et al. Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025. Computing in Cardiology 2025;52:1–4.

[3] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet. Circulation 2000;101(23):e215–e220. URL https://www.ahajournals.org/doi/abs/10.1161/01.CIR.101.23.e215.

[4] Reyna MA, Koscova Z, Pavlus J, Saghafi S, Weigle J, Elola A, et al. Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025, 2025. URL https://arxiv.org/abs/2510.02202.

[5] Sattar Y, Chhabra L. Electrocardiogram. In StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, 2025; .

[6] Ribeiro A, Ribeiro M, Paixão G, Oliveira D, Gomes P, Canazart J, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications 2020;11(1):1760.

[7] Cardoso C, Sabino E, Oliveira C, de Oliveira L, Ferreira A, Cunha-Neto E, et al. Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile. BMJ Open 2016;6(5):e0011181.

[8] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, et al. PTB-XL, a large publicly available electrocardiography dataset. Scientific Data 2020;7:154.

[9] Nunes M, Buss L, Silva J, Martins L, Oliveira C, Cardoso CS BB, et al. Incidence and predictors of progression to chagas cardiomyopathy: Long-term follow-up of trypanosoma cruzi-seropositive individuals. Circulation 2021; 144(19):1553–1566.

[10] Pinto-Filho M, Brant L, Dos Reis R, Giatti L, Duncan B, Lotufo P, et al. Prognostic value of electrocardiographic abnormalities in adults from the Brazilian longitudinal study of adults' health. Heart 2021;107(19):1560–1566.

[11] Apandi ZFM, Ikeura R, Hayakawa S. Arrhythmia detection using MIT-BIH dataset: A review. In 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA). IEEE, 2018; 1–5.

[12] Chen CC, Tsui FR. Comparing different wavelet transforms on removing electrocardiogram baseline wanders and special trends. BMC Medical Informatics and Decision Making 2020;20:1–10.

[13] Polachan K, Chatterjee B, Weigand S, Sen S. Human body-electrode interfaces for wide-frequency sensing and communication: A review. Nanomaterials 2021;11(8):2152.

[14] Pandey SK, Shukla A, Bhatia S, Gadekallu TR, Kumar A, Mashat A, et al. Detection of arrhythmia heartbeats from ECG signal using wavelet transform-based CNN model. International Journal of Computational Intelligence Systems 2023;16(1):80.

[15] Mansi IA, Nash IS. Ethnic differences in electrocardiographic amplitude measurements. Annals of Saudi Medicine 2004;24(6):459–464.

[16] Hassing G, Van der Wall H, Van Westen G, Kemme M, Adiyaman A, Elvan A, et al. Body mass index related electrocardiographic findings in healthy young individuals with a normal body mass index. Netherlands Heart Journal 2019; 27:506–512.

Address for correspondence:

Zoe Boysen
231 Park Drive, Boston, Massachusetts, 02215, United States
zoe.elizabeth.boysen@gmail.com