Vector Signals from the Heart: A Foundation Model Approach to Chagas Detection

Sehun Kim[1,*]

[1]Samsung Medical Center, Seoul, South Korea
[*]Work performed while at Seoul National University, Seoul, South Korea

Abstract

In this paper, we tested the efficacy of frozen electrocardiogram (ECG) representation vectors from a foundation model in detecting Chagas disease, as part of the George B. Moody PhysioNet Challenge 2025. Our team, Seoul Mates, utilized a pre-trained ECG foundation model (ECG-JEPA) which was trained using self-supervised learning on various ECG datasets. This approach learns robust ECG representations by predicting masked portions of the signal in a latent feature space, avoiding the pitfalls of reconstructing noisy raw signals. For the challenge, we applied a linear evaluation protocol on the features extracted from the pre-trained model without any fine-tuning. Interestingly, the feature representations can recover important ECG parameters, such as QRS duration and heart rate, suggesting the model's potential as an off-the-shelf screening tool for Chagas disease. Using these fixed representation vectors, our model achieved the challenge score score of 0.217 on the hidden test set, ranking 18th out of 41 teams.

1. Introduction

For the 2025 George B. Moody PhysioNet Challenge, we explored the application of a foundation model for the detection of Chagas disease from 12-lead electrocardiograms (ECGs) [1–3]. Since definitive serological testing is limited by resource constraints, automated ECG analysis serves as a vital, scalable screening tool. Our work is built on the hypothesis that frozen representation vectors from a self-supervised ECG foundation model contain rich, clinically relevant information sufficient for this screening task. Leveraging the large, publicly available datasets that underpin this competition [4–6], our approach tests this idea by using off-the-shelf features from the pre-trained ECG-JEPA [7] model. To classify the provided signals, we trained only a single linear layer on these fixed representations, which was then evaluated on external validation and test sets. This linear evaluation protocol was deliberately chosen to assess the raw power of the learned features while minimizing the risk of overfitting.

2. Methods

2.1. Datasets and Preprocessing

Our model was trained and evaluated using the datasets provided for the 2025 George B. Moody PhysioNet Challenge. The publicly available training data is a composite of three distinct cohorts: CODE-15% [4], a large-scale dataset of over 300,000 12-lead ECGs (400 Hz) from Brazil with self-reported labels; SaMi-Trop [5], containing 1,631 ECGs (400 Hz) from serologically confirmed Chagas-positive patients; and PTB-XL [6], providing 21,799 ECGs (500 Hz) from a European population assumed to be Chagas-negative. Final model assessment was performed on separate, non-public datasets from Chagas-endemic areas, which were kept private for hidden validation and testing.

All ECG recordings were standardized to ensure consistent input format across the dataset. We converted 12-lead ECGs to a uniform 8-lead representation by selecting leads I, II, and V1-V6. This lead selection strategy leverages the mathematical relationships established by Einthoven's triangle, where four leads (III, aVR, aVL, aVF) can be reconstructed from the eight independent leads:

$$\begin{aligned} \mathrm{III} &= \mathrm{II} - \mathrm{I}, & \mathrm{aVR} &= -(\mathrm{I} + \mathrm{II})/2, \\ \mathrm{aVL} &= (\mathrm{I} - \mathrm{II})/2, & \mathrm{aVF} &= (\mathrm{II} - \mathrm{I})/2. \end{aligned} \tag{1}$$

Each recording was standardized to 10 seconds duration and resampled to 250 Hz, resulting in 2500 time steps per lead. For recordings longer than 10 seconds, the middle portion was extracted to preserve the most stable cardiac cycles, while shorter recordings were zero-padded symmetrically. Importantly, no additional signal preprocessing was applied, allowing the pre-trained model to work with minimally processed ECG signals as encountered in clinical practice. The resampled signals were then conceptually organized as patches of 50 time steps each, creating 50 patches per lead, resulting in a 2D representation

$x = \{x_{c,i}\}$, where $1 \leq c \leq 8$ represents lead index and $1 \leq i \leq 50$ represents patch index with $x_{c,i} \in \mathbb{R}^{50}$.

2.2. ECG Representation Vector

We utilized a pre-trained, transformer-based foundation model (ECG-JEPA) developed through self-supervised learning on diverse ECG datasets. The model employs a standard Transformer architecture adapted for electrocardiographic signals, featuring a 768-dimensional embedding space, 12 transformer layers, and 16 attention heads. The architecture processes ECG patches using multi-head self-attention mechanisms to capture both intra-lead temporal patterns and inter-lead spatial relationships.

During feature extraction, each ECG patch is first projected into the 768-dimensional embedding space via a learned linear transformation. To preserve spatiotemporal information, two-dimensional sinusoidal positional encodings are added to the patch embeddings. The transformer layers then process these embeddings through alternating self-attention and feed-forward operations, with residual connections and layer normalization applied throughout. This process yields 50 patch-level representation vectors, each with a dimension of 768.

To obtain a single global representation for each ECG recording, we applied global average pooling across all 50 patch embeddings from the final transformer layer.

As reported in [7], these representation vectors are versatile enough for both diagnostic classification and the extraction of important ECG features, such as heart rate and QRS duration, suggesting they capture both high-level semantic and low-level physiological information. Given their demonstrated richness, we use these frozen representation vectors directly for the Chagas disease detection task.

2.3. Linear Classification Framework

To evaluate the diagnostic utility of the frozen representation vectors for Chagas disease detection, we implemented a linear probing methodology. This approach allows us to assess the discriminative power of the pre-trained features without introducing additional complexity through feature fine-tuning or deep classification heads.

Our classification architecture consists of a simple linear transformation mapping the 768-dimensional representation vectors directly to binary predictions, which is summarized in Figure 1. Specifically, the probability p of having a chagas disease given a resampled ECG $x \in \mathbb{R}^{8\times2500}$ is:

$$p = \sigma\left(W \cdot f(x) + b\right) \tag{2}$$

where f is a frozen pre-trained network, $f(x) \in \mathbb{R}^{768}$ represents the frozen ECG representation vector, $W \in \mathbb{R}^{1\times768}$ is the learned weight matrix, $b \in \mathbb{R}$ is a learnable bias term, and σ is the sigmoid function.

During training, we precomputed all ECG representation vectors using the frozen encoder to improve computational efficiency. This two-stage approach first extracts features from all ECG recordings in large batches, then trains the linear classifiers on these fixed representations, significantly reducing training time and memory requirements.

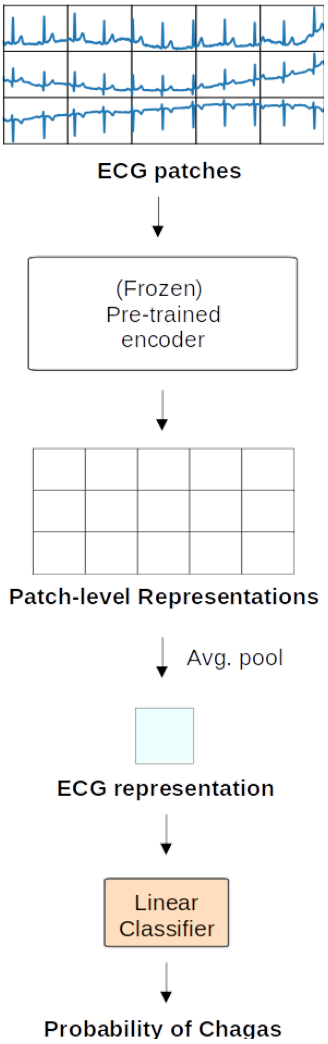

Figure 1. An overview of our framework. An input ECG is divided into patches and processed by a frozen, pre-trained encoder to generate patch-level representations. These are globally averaged to produce a single ECG representation vector, which is then fed into a learnable linear classifier to predict the probability of Chagas disease.

2.4. Cross-Validation and Ensemble Strategy

We implemented a 5-fold stratified cross-validation scheme that serves dual purposes: enabling the utilization of all training samples for model development and providing robust performance evaluation. The stratification preserves the class distribution within each fold, which is particularly important given the inherent imbalance in Chagas disease prevalence within the dataset.

For each cross-validation fold, we trained an independent linear classifier using the corresponding training partition while evaluating on the held-out validation set. This approach ensures that every sample in the training dataset contributes to model training across different folds, maximizing the utilization of available data. The process generates five distinct classifiers, each optimized on different subsets of the training data while maintaining identical architectures and hyperparameter configurations.

During inference, we combine predictions from all five classifiers through ensemble averaging, which helps prevent overfitting by reducing the variance of individual model predictions and improving generalization performance:

$$p = \frac{1}{5}\sum_{i=1}^{5} p_i \tag{3}$$

where p_i represents the probability output from the i-th classifier. The ensemble strategy leverages the diversity of models trained on different data partitions, leading to more robust and reliable predictions than any single classifier. The final binary classification decision applies a 0.5 threshold to this ensemble probability.

2.5. Training Protocol and Optimization

Linear classifiers were trained using the AdamW optimizer with a learning rate of 0.025. We employed a cosine annealing schedule with warm restarts, beginning with a 3-epoch linear warmup phase where the learning rate gradually increases from 0.0025 to the base rate, followed by cosine decay to a minimum of 0.0025 over the remaining epochs. This scheduler was adopted as a default from our training framework; while not essential for a linear-only protocol, its effect on the final performance was found to be negligible compared to a constant learning rate. Training was conducted for a maximum of 20 epochs with early stopping based on validation performance using the challenge score with patience epoch 5. We monitored the official Challenge score on the validation set and terminated training if no improvement was observed for 5 consecutive epochs. The binary cross-entropy loss function was used for optimization:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{4}$$

where N is the batch size, y_i is the true label, and p_i is the predicted probability.

To address the natural class imbalance present in Chagas disease detection, we used weighted random sampling during training. Sample weights were computed as the inverse of class frequencies, ensuring equal representation of both positive and negative cases in each training batch. Specifically, if n_0 and n_1 represent the number of negative and positive samples respectively, the weights are calculated as:

$$w_0 = \frac{1}{n_0}, \quad w_1 = \frac{1}{n_1} \tag{5}$$

This rebalancing strategy prevents the model from developing a bias toward the majority class.

Training	Validation	Test	Ranking
0.406 ± 0.007	0.306	0.217	18/41

Table 1. Challenge scores for our selected entry (team Seoul Mates), 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.

	5-fold CV.
Accuracy	0.764 ± 0.049
F1 Score	0.126 ± 0.016
AUROC	0.841 ± 0.005
AUPRC	0.176 ± 0.004

Table 2. Additional performance metrics from 5-fold cross-validation on the public training set. Values are reported as mean ± standard deviation across the folds

3. Results

Our official Challenge results are summarized in Table 1. The table presents the mean Challenge score from our 5-fold cross-validation on the public training set, alongside the official score on the hidden validation set as evaluated by the challenge organizers.

For a more detailed analysis of our model's performance during internal validation, supplementary metrics from the 5-fold cross-validation are provided in Table 2. These include standard classification metrics such as Accuracy, F1 score, AUROC, and AUPRC, which offer further insight into the model's behavior on the training data.

4. Discussion

Our automated Chagas detection approach achieved a challenge score of 0.217 on the hidden test set, which suggests ECG-based screening could serve as a valuable first-line tool in endemic regions, particularly where serological testing is limited. The method's minimal computational requirements after feature extraction and compatibility with standard 12-lead ECGs make it practical for resource-constrained healthcare settings.

4.1. Linear Evaluation vs. Fine-tuning

We conducted additional experiments with full model fine-tuning. Unlike linear probing, where we freeze the encoder and train only the linear classifier, fine-tuning updates both the encoder and linear classifier. As expected, fine-tuning the entire encoder achieved a notably higher 5-fold cross-validation score of 0.497 on the training data. However, its performance on the hidden validation set dropped sharply to 0.262. This stark difference in generalization behavior validates our methodological choice: the fixed feature approach captures generalizable cardiac patterns while avoiding overfitting to dataset-specific artifacts. The result suggests that for Chagas detection, where cardiac manifestations vary across populations and disease stages, simpler models demonstrate superior robustness.

5. Conclusion

We demonstrated that Chagas disease screening from ECG signals can be effectively performed using linear classification on pre-extracted features. The approach's simplicity, computational efficiency, and robust generalization make it suitable for deployment in resource-limited endemic regions. Our finding that linear evaluation outperforms fine-tuning in generalization highlights the challenge of capturing diverse Chagas cardiac manifestations. This work provides a practical foundation for developing accessible screening tools for Chagas disease, contributing to broader efforts in combating neglected tropical diseases through automated diagnostics.

Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2024-004370388), and a grant of the National Research Foundation of Korea (NRF), funded by the Korean government (MSIP) (Grant number: 2023R1A2C1005562). The authors also wish to thank Otto van Koert for providing the initial knowledge and introduction to ECG.

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. In Computing in Cardiology 2025, volume 52. 2025; 1–4.

[3] 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. arXiv preprint arXiv251002202 2025;.

[4] 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.

[5] 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.

[6] 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.

[7] Kim S. Learning general representation of 12-lead electrocardiogram with a joint-embedding predictive architecture, 2024. URL https://arxiv.org/abs/2410.08559.

Address for correspondence:

Sehun Kim
81 Irwon-ro, Gangnam-gu, Seoul, South Korea
sh.math.kim@gmail.com