Automated 12-Lead ECG Chagas Disease Detection: HuBERT and Multireceptive Field CNN Hybrid Approach

Monika Kisieliūtė[1], Vladislav Kolupayev [2]

[1]Aerospace Data Center, Antanas Gustaitis' Aviation Institute, Vilnius Gediminas Technical University, Vilnius, Lithuania

[2]Institute of Informatics, Faculty of Mathematics and Informatics, Vilnius University, Vilnius, Lithuania

Abstract

Detection of Chagas disease from ECG is largely understudied in computational modeling literature. Moreover, there is a need to detect Chagas from ECG data, as this method is widely accessible and affordable for target populations. Our team, HeartGoesOut2U, proposes a hybrid transformer-CNN model, composed of a pre-trained HuBERT-ECG backbone and Multireceptive Field CNN head. We trained such a model on a merged training dataset, consisting of records from PTB-XL, Code-15%, and SaMi-Trop databases (all downsampled to 100 Hz). The following training pipeline was employed: selected ECG signals were cleaned by removing baseline wander using a Butterworth high-pass filter, utility frequency using a Notch filter, and additional noise using a 4-level DWT. To address the imbalanced nature of Chagas prevalence, positive class weighted Focal Loss was utilized. We conducted an internal evaluation using 5-fold cross validation on public training data. Here, a Challenge score of 0.364, an accuracy score of 0.774, and an F-1 measure of 0.109 were observed. Furthermore, AUROC of 0.785, and AUPRC of 0.143 were observed. Additionally, on the hidden test set of George B. Moody PhysioNet Challenge 2025, the proposed method achieved a Challenge score of 0.204, placing us 22nd out of 41 teams.

1. Introduction

Electrocardiography (ECG) is a widely used, non-invasive data collection method that can inform about various heart disorders, including Chagas disease. While ECG data alone is not sufficient to confirm Chagas, it can supplement confirmatory serological testing, which is not always available to target populations. However, automatic detection of Chagas from ECG has been widely underexplored, which is why the 2025 George B. Moody PhysioNet Challenge [1–3] addresses the need for developing such algorithms.

Using deep learning (DL) algorithms, and in particular, transformer and convolutional neural network (CNN) based models, has shown promise in detecting cardiovascular disorders from ECG data. Moreover, using pre-trained models for downstream tasks has been a widely known strategy for improving performance. Therefore, for the 2025 PhysioNet Challenge, we propose a hybrid transformer-CNN based network that leverages the generalization properties of using pre-trained backbone models.

2. Methods

2.1. Data and Preprocessing

To train and evaluate our model, 5 databases were utilized: Code-15%, SaMi-Trop, PTB-XL, REDS-II, and ELSA-Brasil [4–8]. Our training dataset was composed of records from three 12-lead ECG sources: PTB-XL, Code-15%, and SaMi-Trop. Upon analyzing records of each database, a conclusion was reached that signals in these databases were of differing quality, for instance, different degrees of baseline wander. Because of this, we adopted a rigorous exclusion and cleaning pipeline that was designed to mitigate some of the data quality issues and unify the records to the greatest extent feasible.

Firstly, we excluded records shorter than 3.75 seconds. This cut-off was established to minimize the exclusion of valid records (especially ECG leads that are from target demographics), while maintaining a sufficient time interval to capture multiple cardiac cycles.

Then, downsampling to 100 Hz was performed, as relevant cardiological features show up in 0.5-47 Hz frequency and 100 Hz sampling is enough to capture the relevant bandwidth. After downsampling, three supplementary filters were utilized: Butterworth high-pass filter to remove baseline wander (around 0.5 Hz), Notch filter to remove 50 Hz utility frequency. Subsequently, we employed 4-level discrete wavelet transform (DWT). The latter was done to

ISSN: 2325-887X DOI: 10.22489/CinC.2025.197

minimize remaining noise but still retain essential ECG morphological features. When employing DWT, we used soft thresholding that was applied only on DWT detail coefficients.

$$\hat{c}_i = \mathrm{sign}(c_i)\ \max\big(|c_i| - T,\ 0\big). \tag{1}$$

In soft thresholding (equation 1), $\hat{c}_i$ denotes new thresholded DWT coefficient, c_i denotes the i^{th} detail coefficient from DWT, and T denotes universal threshold [9]:

$$T = \sigma\sqrt{2\log n} \tag{2}$$

Above, σ is the noise standard deviation and n - number of samples (in this case, coefficients). The Daubechies family order 4 wavelet was chosen as our mother wavelet, as it is suitable for ECG denoising [10].

We excluded ECG leads that were outside IQR range of signal peak-to-peak amplitudes (PPA) for each ECG lead. PPA was calculated after downsampling and filtering, with our main motivation being that signals beyond this range need an exceptional level of cleaning and/or reconstruction, however, this is beyond the objectives of this study. We did not implicitly exclude records with missing channel data. After all of these steps, the proportion of Chagas labels was similar to the unprocessed combined dataset proportion (2.24 % before exclusion and 2.14 % after).

As a last preprocessing step, we standardize all examples to a 5 second window. If the resampled training ECG is longer than 5 seconds, we use random cropping along the time axis with a probability of 0.5, and otherwise we truncate the signal to 5 seconds. If the ECG is shorter, then we pad it with zero padding to keep recordings at the same length. Afterward, signals are per-channel normalized to be in the range of [-1, 1]. Finally, channel-wise concatenation is performed, as our model takes in 1D inputs.

2.2. Model Architecture

For our model design, we chose a unique approach of combining Hidden-Unit BERT (HuBERT) [11] based feature extractor with custom designed Multireceptive Field (MRF) CNN head.

HuBERT (visualized in figure 1) was originally designed for speech representation learning. It combines a CNN waveform encoder to create sequences of features that are passed to a bidirectional encoder transformer. HuBERT utilizes self-supervised pre-training by clustering signal frames into labels, which are used in masked representation learning, allowing the model to learn meaningful latent features and global temporal relations. This approach proves to be a significant improvement in a variety of downstream speech tasks.

Furthermore, HuBERT-ECG adapts this methodology to ECG signals with a pre-training dataset of 9.1 million 12-lead ECGs composed of 11 cardiological datasets and shows promising results on downstream clinical applications [12]. HuBERT-ECG incorporates an adjustment to the convolutional encoder in the original HuBERT model and slightly modifies the masking strategy. In order to employ the strong foundation in generalization of ECG signal representation, we instantiate our model with pre-trained HuBERT-ECG Base model weights to start transfer learning on Chagas disease identification. Besides, transformer models incorporate attention masks to attend to relevant parts of the ECG leads, thus enabling inputting signals that

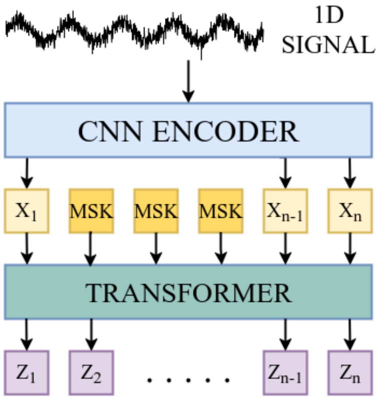

Figure 1. HuBERT model. MSK notation denotes mask embedding, X_i denotes i^{th} signal frame embedding, Z_i denotes i^{th} hidden unit.

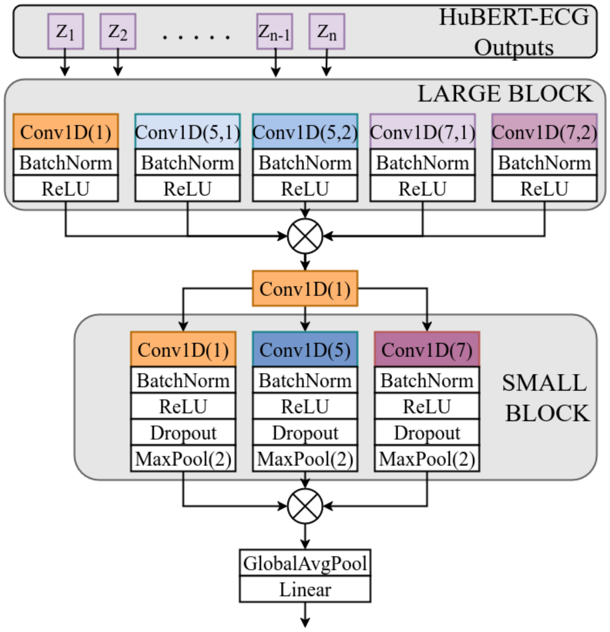

Figure 2. MRF head design. Note, that first number in brackets near layers indicates kernel size, and second one (if applicable) - dilation rates.

are shorter or have missing data. This way, more data can be utilized during training.

Our MRF head, inspired by [13], consists of two main parts: a large block and a small block (see figure2). Both blocks are composed of convolutions with varying kernel sizes (1, 5, and 7). Layers in the large block also have varying dilation rates (1 and 2) for kernel sizes 5 and 7. The input for our head is the last hidden state of HuBERT-ECG model, which is an information-rich feature representation, yet it's still local in time-domain. For this reason, our large block is intended to capture multi-scale temporal context, while the small block refines and compresses it. After each convolution layer, we include batch normalization (BatchNorm) and ReLU activation function to improve stability. In the small block, as a slight regularization setting, we add dropout (the default value for dropout in our head is 0.1), followed by max pooling with size 2. Then, global average pooling is applied along the time axis of the feature representation, followed by a linear layer to get the final model predictions.

2.3. Training Process

For training our models, we utilize a positive class weighted Focal Loss [14], outlined in equation (5). Focal loss introduces γ modulation term to reduce the relative contribution of easy-to-classify ECG signals, thereby emphasizing harder examples. This is particularly important in our setting, where the majority of ECG traces correspond to healthy patients or otherwise easily distinguishable cases. By contrast, recordings from patients with heart conditions, including those overlapping with clinical cardiac features of Chagas disease and those exhibiting distinct pathological patterns, are given greater weight during training. Moreover, positive class weighting (equation (3)) accounts for the high class imbalance by further increasing the contribution to the loss for positive Chagas cases.

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases} \tag{3}$$

$$w_t = \begin{cases} w & \text{if } y = 1 \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

$$L_{\text{focal}}(p_t, w_t) = -w_t(1 - p_t)^{\gamma} \log p_t \tag{5}$$

The parameters we set for the loss are $\gamma = 1.5$ and $w = 45$. In other words, the positive class weight scales up the loss 45 times for positive Chagas cases.

Furthermore, to train our submission model, we employ two stages and a gradual unfreezing approach[15]. In the first stage, stratified data splits of 80% of the data for the training set and 20% for the validation set are created from the public training dataset. By utilizing the validation fold, we can determine the epochs corresponding to peak validation performance, which aids in selecting the most effective model and mitigating overfitting.

First, the HuBERT-ECG feature extractor is frozen and only the MRF model head is trained for 10 epochs or until the model does not improve its validation Challenge metric for 2 epochs in a row. The optimizer used for training is AdamW with a learning rate (LR) of 5e-4. The rest of the optimizer hyper-parameters are left as their default value.

Following model head training, we set the model state to the best performing model weights at the best performing epoch, then we unfreeze the feature extractor and proceed to train the model for 35 epochs with an early stopping patience of 5 epochs. However, this time we apply discriminative LRs: 1e-6 LR for model layers preceding transformer encoder blocks (TEB); 4e-6, 6e-6, 8e-6, 1e-5, 2e-5, 4e-5 LRs for successive pairs of TEBs; and 5e-4 for the MRF head. The rest of the AdamW parameters were set to default values, except for model batch normalization, layer normalization, and bias parameters have their weight decay set to 0. This training procedure avoids catastrophic forgetting and produces better model convergence.

In the second stage, the model is reset, training and validation sets are combined, and the gradual unfreezing process is repeated. This training procedure is performed with the number of epochs found to be best in the first stage.

3. Results

Results of our internal investigation, where a stratified 5-fold cross-validation on training data was used, are shown in table 1. Here, our method achieved 0.785 AUROC and 0.774 accuracy. AUPRC and F1 scores were on the lower side, averaging at 0.143 and 0.109, respectively. All observed metrics have relatively low standard deviations.

	AUROC	AUPRC	Accuracy	F1
mean	0.785	0.143	0.774	0.109
std, ±	0.011	0.014	0.056	0.012

Table 1. AUROC, AUPRC, accuracy, and F1 scores for our model. Evaluated with stratified 5-fold cross validation on the training data.

Training	Validation	Test	Ranking
0.364 ± 0.018	0.372	0.204	22/41

Table 2. Challenge scores for our selected entry, including the ranking of our team on the hidden test set. Stratified 5-fold cross-validation scoring was used on the public training set.

Table 2 showcases our Challenge score on training, validation, and test sets. We achieved a Challenge score of

0.364 and 0.372 on training and validation sets, respectively. On the hidden test set, our method scored 0.204.

4. Discussion and Conclusions

Our proposed method leverages a pre-trained HuBERT-ECG backbone combined with a custom designed MRF head, which is capable of classifying short-duration records and records with missing channel data. This allows applications in real-world clinical and mobile settings, where recordings are often brief, incomplete, or corrupted by noise. The Challenge score, for which our methodology was optimized, demonstrates evidence of potential. However, a trade-off in other metrics can be noted. Specifically, low AUPRC and F1 scores on the training set indicate suboptimal performance in the highly unbalanced data setting. Nonetheless, our model still performs much better than random guessing, as the floor of AUPRC is the proportion of the positive cases in the evaluation dataset. Consequently, high accuracy and AUROC show that the model can perform fairly well on non-Chagas cases. Overall model stability is indicated by low standard deviations for all training set metrics. However, overfitting is observed - test set scores are much lower than training and validation scores.

Future improvements can be made in several areas. To begin with, high quality data is needed in order to train any DL model well. Although our data preprocessing included denoising steps, a systematic evaluation of its efficacy was not performed, nor did we optimize its associated parameters (e.g., choice of wavelets, thresholding functions). Finally, our proposed method is resource heavy ($\sim$200M parameters), which can lead to overfitting and hinder deployment on devices with limited computational resources.

Acknowledgments

This research was partially funded by Antanas Gustaitis' Aviation Institute, Vilnius Gediminas Technical University.

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, 2025. URL https://arxiv.org/abs/2510.02202.

[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] 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.

[8] 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.

[9] Donoho DL, Johnstone IM. Ideal spatial adaptation by wavelet shrinkage. Biometrika 1994;81(3):425–455.

[10] Balasubramaniam D, Nedumaran D. Implementation of ECG signal processing and analysis techniques in digital signal processor based system. In 2009 IEEE International Workshop on Medical Measurements and Applications. IEEE, 2009; 60–63.

[11] Hsu WN, Bolte B, Tsai YHH, Lakhotia K, Salakhutdinov R, Mohamed A. Hubert: Self-supervised speech representation learning by masked prediction of hidden units. IEEE ACM Transactions on Audio Speech and Language Processing 2021;29:3451–3460.

[12] Coppola E, Savardi M, Massussi M, Adamo M, Metra M, Signoroni A. HuBERT-ECG as a self-supervised foundation model for broad and scalable cardiac applications. medRxiv 2024;2024–11.

[13] Feyisa DW, Debelee TG, Ayano YM, Kebede SR, Assore TF. Lightweight Multireceptive Field CNN for 12-Lead ECG Signal Classification. Computational Intelligence and Neuroscience 2022;2022(1):8413294.

[14] Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision. 2017; 2980–2988.

[15] Howard J, Ruder S. Universal language model fine-tuning for text classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018; 328–339.

Address for correspondence:

Monika Kisieliūtė
Linkmenų str. 28-4, Aerospace Data Center, Antanas Gustaitis' Aviation Institute, Vilnius Gediminas Technical University, Vilnius, Lithuania. E-mail: monikakslt@gmail.com