Res2Net-Transformer Network Framework Integrating Curriculum Learning and Hard Sample Mining for Chagas Disease Detection from ECGs

Jinxiang Huang[1], Xuemin Yu[1], Jingming Zhang[1], Changsen Zi[1], Yusi Zhu[1]

[1]School of Physics and Electronic Information, Yunnan Normal University, Kunming, China

Abstract

The 2025 George B. Moody PhysioNet Challenge focuses on detecting Chagas disease from standard 12-lead electrocardiogram (ECG) recordings, where the main difficulties include heterogeneous data sources, weak versus strong labeling, and serious class imbalance. Our team, Med_YNNU, proposed a deep learning framework based on a Res2Net-Transformer backbone that integrates curriculum learning and hard sample mining to improve both convergence and sensitivity to rare cases. The architecture consists of a main classification head for Chagas prediction, auxiliary branches predicting demographic and physiological variables trained with proxy constants acting as weak regularizers, and an adaptive threshold head that refines the decision boundary. Training follows a curriculum that progresses from easy to hard samples, while hard sample mining dynamically emphasizes high-loss cases to strengthen discriminability under borderline conditions. Finally, our team, Med_YNNU, achieved a mean Challenge score of 0.221 on the hidden test set, ranking 15st out of 40 teams. The detailed test performance includes a Challenge score of 0.304 on the REDS-II test set, 0.276 on the SaMi-Trop 3 test set, and 0.082 on the ELSA-Brasil test set.

1. Introduction

Chagas disease, caused by Trypanosoma cruzi and transmitted mainly by triatomine insects, is a neglected tropical disease that remains a major health concern in Central and South America, with an estimated 6.5 million cases and nearly 10,000 deaths annually [1,2]. After an initial acute phase, many patients progress to a chronic stage that may lead to severe cardiomyopathy, including conduction delays, arrhythmias, and heart failure. Early treatment with antiparasitic drugs can prevent progression, but reliable diagnosis relies on serological testing, which is often unavailable or infeasible for large-scale screening in endemic regions. Electrocardiograms (ECGs) provide a low-cost, widely accessible alternative, since Chagas-related abnormalities are frequently reflected in waveform morphology and temporal conduction patterns [3].

The 2025 George B. Moody PhysioNet Challenge addresses this gap by inviting open-source algorithmic approaches to detect Chagas disease from standard 12-lead ECGs [4,5]. The problem is technically challenging due to the coexistence of large weakly labeled cohorts and smaller strongly labeled datasets, as well as the severe class imbalance between positive and negative cases. These conditions complicate optimization and hinder model generalization across populations.

To overcome these challenges, we investigate a Res2Net-Transformer framework that incorporates curriculum learning and hard sample mining. Curriculum learning offers stable convergence by exposing the model to progressively harder training samples, while hard sample mining improves sensitivity to rare and borderline cases that are clinically significant. In addition, we introduce a multi-head architecture with auxiliary branches designed to predict demographic and physiological features and an adaptive threshold head to refine decision boundaries.

2. Methods

2.1. Datasets and Data Preprocessing

The Challenge data consist of standard 12-lead electrocardiogram recordings collected from multiple cohorts across Central and South America, complemented with basic demographic information and binary labels for Chagas disease. The publicly available training data are derived from three main sources: (i) CODE-15% dataset [6]: more than 300,000 ECGs from Brazil (weak self-reported Chagas labels). (ii) SaMi-Trop dataset [7]: 1,631 ECGs from Brazil with strong serological labels (all positive). (iii) PTB-XL dataset [8]: 21,799 European ECGs (all assumed negative). All recordings are 12-lead, 7.3–10.2 s long, sampled at 400-500 Hz.

The preprocessing pipeline consists of several steps. First, each raw 12-lead ECG recording is truncated or zero-padded to a fixed length of 4096 samples to ensure uniform input size. Second, z-score normalization is

applied on a per-recording basis to reduce baseline drift and amplitude variability. Finally, the processed signals are stored in a tensor format of shape (12,4096), which is used as the model input.

2.2. Model Architecture

The proposed network integrates convolutional, residual, and self-attention modules with approximately 3.2 million trainable parameters, as illustrated in Figure 1. A 1D convolutional front-end (kernel size 15, stride 1, 12 input leads to 64 channels) extracts local temporal features. A subsequent Res2Net stage with squeeze-and-excitation (SE) modules increases channel capacity to 256, and enabling multi-scale representation learning with residual splits (scale = 2, base width = 16) [9]. The temporal resolution is then reduced by pooling to a sequence length of 128, which serves as the input to a 2-layer Transformer encoder (embedding dimension 256, 4 attention heads, feed-forward dimension 512, dropout 0.1) that models long-range dependencies across the ECG sequence.

The shared representation is fed into multiple heads: (i) A main classification head (Linear 256 → 64 → 1) for Chagas prediction. (ii) Four auxiliary heads (age, sex, heart rate, and QRS duration; each implemented as Linear layers from 256-d to 1) that were trained with proxy constant labels and act as weak regularizers. (iii) A threshold head (Linear 256 → 1 with Sigmoid) that produces an adaptive decision boundary. The auxiliary heads are enclosed in dashed boxes in Figure 1, indicating that they are optional components and can be removed without affecting the core framework.

In summary, this design directly aligns model capacity with clinically meaningful ECG characteristics, supporting more accurate disease detection.

2.3. Curriculum Learning and Hard Sample Mining

To stabilize training in the presence of imbalanced and noisy labels, we adopt a curriculum learning strategy combined with hard sample mining. Following the principles of curriculum learning [10] and hard sample mining [11], we define the training schedule as follows (Eq. (1) – (5)). For each training sample $\boldsymbol{x_i}$ with label $\boldsymbol{y_i}$, we define its difficulty as the prediction loss (Eq. (1)):

$$d(x_i) = L(f_\theta(x_i), y_i) \quad (1)$$

Where f_θ is the model with parameters θ. During epoch t, samples are selected if their difficulty is below a threshold τ_t (Eq. (2)), customized formula for curriculum scheduling:

$$S_t = \{x_i | d(x_i) \leq \tau_t\} \quad (2)$$

The threshold increased linearly with training epoch (Eq. (3)):

$$\tau_t = \tau_0 + \Delta \cdot t \quad (3)$$

With initial easy ratio $\tau_0 = 0.3$ and increment $\Delta = 0.1$.

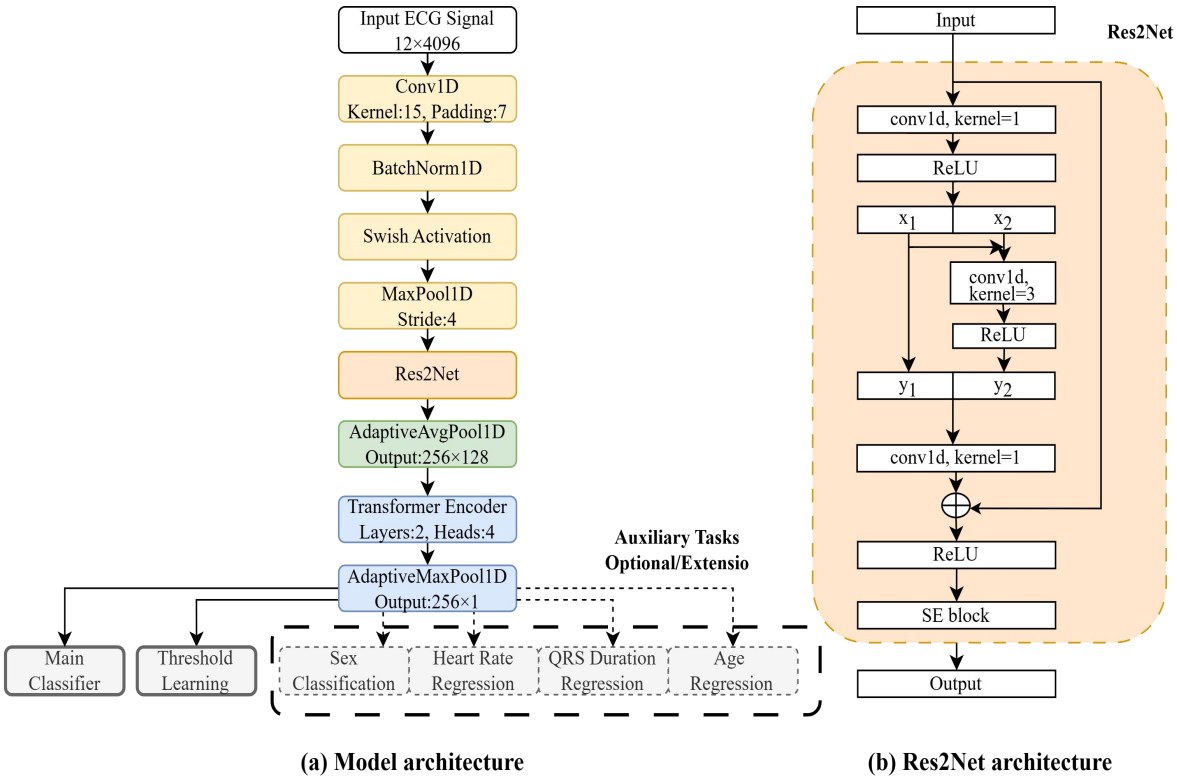

Figure 1. The architecture diagram of our proposed approach. (a) Model architecture. (b) Res2Net architecture.

This ensures that training starts from easy samples and gradually incorporates more challenging ones. Hard sample mining is further applied by dynamically collecting samples with loss exceeding a margin δ (Eq. (4)):

$$H_t = \{x_i | L(f_\theta(x_i), y_i) > \delta\}, \quad \delta = 0.5 \tag{4}$$

The final mini-batch is formed by combining easy and hard sets (Eq. (5)):

$$B_t = S_t \cup H_t \tag{5}$$

This combined strategy ensures stable early-stage optimization while emphasizing borderline and misclassified samples during later stages, directly enhancing performance on rare but clinically significant Chagas cases.

2.4. Loss Function

The overall objective combined main classification loss, auxiliary task losses, ranking loss, and threshold regularization (Eq. (6)), the total loss was formalized as:

$$L_{total} = \alpha_{main}L_{main} + \alpha_{aux}L_{aux} + \alpha_{rank}L_{rank} + \alpha_{thr}L_{thr} \tag{6}$$

Where α_{main},α_{aux},$\alpha_{rank}$and $\alpha_{thr}$are weighting coefficients for the main, auxiliary, ranking, and threshold losses, respectively. In this work, we set $\alpha_{main} = 1$, $\alpha_{aux} = 0.1$, $\alpha_{rank} = 0.1$,and $\alpha_{thr} = 0.01$.

Main classification loss

We used binary focal loss [12] to address severe class imbalance (Eq. (7)):

$$L_{main} = -\alpha(1 - p_t)^\gamma log(p_t) \tag{7}$$

Where $\alpha \in [0,1]$ is a balancing factor between positive and negative classe($\alpha = 0.25$ in our experiments)s, $\gamma \geq 0$ is the focusing parameter that down-weights easy samples and emphasizes hard ones($\gamma = 2$ in this work), and p_t denotes the predicted probability of the corresponding class label .

Auxiliary task loss

The auxiliary heads were trained with proxy constant labels. For binary auxiliary tasks, we optimize the auxiliary binary classification heads using binary cross-entropy with logits (Eq. (8)), following the standard formulation [13]:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i log\sigma(z_i) + (1 - y_i)log(1 - \sigma(z_i))] \tag{8}$$

Where *N* is the number of samples in a mini-batch, $y_i \in \{0,1\}$ is the ground-truth label of the *i*-th sample, z_i is the raw logit output of the auxiliary head before activation, $\sigma(\cdot)$ denotes the sigmoid function that maps logits to probabilities. For regression-type tasks, mean squared error [14] was applied (Eq. (9)):

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{9}$$

Where $\hat{y}_i$ is the corresponding regression output predicted by the auxiliary head. The auxiliary loss combined all branches (Eq. (10)). Our customization, supported by multitask learning theory [15]:

$$L_{aux} = \sum_j \lambda_j L_j \tag{10}$$

Where λ_j balances contributions from each auxiliary task.

Ranking loss

To align optimization with the Challenge metric, we adopted a pairwise margin-based ranking loss [16] (Eq. (11)):

$$L_{rank} = \frac{1}{|\mathcal{P}||\mathcal{N}|}\sum_{n\in\mathcal{N}} max\left(0, m - \left(s_p - s_n\right)\right) \tag{11}$$

Where $\mathcal{P}$ and $\mathcal{N}$ are sets of positive and negative samples, s_p and s_n are their predicted scores, and m is a margin hyperparameter.

Threshold regularization

Inspired by prior work on adaptive decision boundaries [17], we regularized the adaptive threshold head toward a prior mean of 0.5 using an L2 penalty (Eq. (12)):

$$L_{thr} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\tau}_i\text{-}\mu)^2 \tag{12}$$

Where $\hat{\tau}_i$ is the adaptive threshold predicted for the *i*-th sample and constrained to [0,1] by a sigmoid activation, $\mu = 0.5$ is the prior mean threshold representing a neutral decision boundary.

2.5. Model Training

The model was trained using stratified 5-fold cross-validation to preserve class balance across folds. We optimized the model's parameters with AdamW (initial learning rate 2×10−4, weight decay 1×10−2), mini-batch size was set to 96, and a cosine annealing learning rate schedule. Early stopping was applied with a patience of 5 epochs, using a joint criterion that monitored both validation loss and the Top-5% TPR metric to balance stability and clinical relevance. Curriculum learning was applied by starting from 30% "easy" samples and progressively incorporating harder ones, while hard sample mining emphasized samples with binary cross-entropy loss greater than 0.5. Dropout (0.2) and auxiliary task losses provided additional regularization, and the adaptive threshold head was trained jointly with a prior mean of 0.5 to stabilize decision boundaries. For inference, we ensembled the 5 fold-specific models by averaging their output probabilities, which provided a robust final prediction and reduced variance.

3. Results

We evaluated our proposed algorithms using five-fold cross-validation on the public training set with the official

Challenge evaluation metric. The final submitted model achieved a mean Challenge score of 0.221 on the hidden test set, ranking 15th out of 40 teams on the official leaderboard. The detailed Challenge scores on the public training, hidden validation, and test sets are summarized in Table 1

Training	Validation	Test	Ranking
0.203±0.05	0.299	0.221	15/40

Table 1. Summary of official Challenge results for team Med_YNNU.

4. Discussion and Conclusions

The proposed Res2Net–Transformer model achieved a Challenge score of 0.299 on the REDS-II validation set and a mean score of 0.221 on the hidden test set. Across individual test cohorts, the model obtained 0.304 on REDS-II, 0.276 on SaMi-Trop 3, and 0.082 on ELSA-Brasil [4,5]. These results indicate that the model generalizes well to datasets with characteristics similar to the training domain, such as REDS-II and SaMi-Trop 3, which contain serologically confirmed labels and comparable acquisition settings. In contrast, the lower score on ELSA-Brasil suggests limited transferability to more heterogeneous populations with lower Chagas prevalence and differing recording systems. The observed variation highlights the influence of cohort diversity and data distribution shift on model generalization. Overall, these findings demonstrate the potential of deep learning–based approaches for ECG-based Chagas screening, while emphasizing the need for improved cross-domain adaptation and data harmonization to achieve consistent performance across diverse clinical populations.

Acknowledgments

This work is supported by the Scientific Research Foundation Project Funded by Education Department of Yunnan Province under Program 2024J0134 and Yunnan Fundamental Research Projects under Grant 202401CF070028.

References

[1] Cucunubá Z M, Gutiérrez-Romero S A, Ramírez J D, et al. The epidemiology of Chagas disease in the Americas. *The Lancet Regional Health–Americas*, 2024, 37.

[2] de Andrade G N, Bosch-Nicolau P, Nascimento B R, et al. Prevalence of Chagas disease among Latin American immigrants in non-endemic countries: an updated systematic review and meta-analysis. *The Lancet Regional Health–Europe*, 2024, 46.

[3] Jidling C, Gedon D, Schön T B, et al. Screening for Chagas disease from the electrocardiogram using a deep neural network. *PLoS Neglected Tropical Diseases*, 2023, 17(7): e0011118.

[4] Reyna MA, Koscova Z, Pavlus J, Weigle J, Saghafi S, Gomes P, Elola A, Hassannia, MS, Campbell, K, Bahrami Rad A, Ribeiro AH, Ribeiro ALP, Sameni R, Clifford GD. Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025; In *Computing in Cardiology 2025*, volume 52: 2025, 1-4.

[5] Reyna MA, Koscova Z, Pavlus J, Weigle J, Saghafi S, Gomes P, Elola A, Hassannia, MS, Campbell, K, Bahrami Rad A, Ribeiro AH, Ribeiro ALP, Sameni R, Clifford GD. Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025; *arXiv Preprint*, 2025. DOI:10.48550/arXiv.2510.02202.

[6] Ribeiro A H, Ribeiro M H, Paixão G M M, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature Communications*, 2020, 11(1): 1760.

[7] Cardoso C S, Sabino E C, Oliveira C D L, et al. Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile. *BMJ Open*, 2016, 6(5): e011181.

[8] Wagner P, Strodthoff N, Bousseljot R D, et al. PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, 2020, 7(1): 1-15.

[9] Gao S H, Cheng M M, Zhao K, et al. Res2net: A new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 43(2): 652-662.

[10] Bengio Y, Louradour J, Collobert R, et al. Curriculum learning. *Proceedings of the 26th Annual International Conference on Machine Learning*. 2009: 41-48.

[11] Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard sample mining. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016: 761-769.

[12] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*. 2017: 2980-2988.

[13] Goodfellow I, Bengio Y, Courville A, et al. Deep learning. *Cambridge: MIT Press*, 2016.

[14] Bishop C M, Nasrabadi N M. Pattern recognition and machine learning. *New York: Springer*, 2006.

[15] Caruana R. Multitask learning. *Machine Learning*, 1997, 28(1): 41-75.

[16] Burges C, Shaked T, Renshaw E, et al. Learning to rank using gradient descent. *Proceedings of the 22nd International Conference on Machine Learning*. 2005: 89-96.

[17] Menon A, Narasimhan H, Agarwal S, et al. On the statistical consistency of algorithms for binary classification under class imbalance. *International Conference on Machine Learning*. PMLR, 2013: 603-611.

Address for correspondence:

Yusi Zhu
School of Physics and Electronic Information, Yunnan Normal University, No. 1, Yuhua, Chenggong District, Kunming, China.
zhuyusi16@mails.ucas.edu.cn.