ResNet-BiGRU with Conditioned Query-Based Cross-Attention and Weighted Loss for Automated Chagas Disease Detection from 12-Lead ECG

Hyuno Im[1], Nahyun Lee[1], Taeyoung Kang[1], Taehwan Kim[1], Donggun Kim[2], Donggyu Lee[2], Seungsang Oh[2], Wuming Gong[3] and Il-Youp Kwak[1]

[1] Chung-Ang University, Seoul, Korea
[2] Korea University, Seoul, Korea
[3] University of Minnesota, Lillehei heart institute, Minneapolis, MN, United States

Abstract

Our team CAUETUMN, a participating team in the 2025 PhysioNet/CinC Challenge, investigates whether integrating physiologically interpretable features with deep sequence representations, enhanced by conditioned Query-based cross-attention, can improve Chagas disease detection from standard 12-lead ECGs. We combine three information streams: (i) a 4-layer 1D ResNet backbone for local morphology extraction; (ii) a bidirectional GRU with gated attention for long-range temporal context; and (iii) handcrafted R-peak morphology feature and demographic features (age, sex). The auxiliary features are projected into a high-dimensional query space to conditionally attend over sequence embeddings, enabling selective integration of relevant temporal patterns. Raw ECGs undergo baseline-wander removal with an OC/CO morphological filter. The model is trained on a combination of SaMi-Trop, PTB-XL, and CODE-15% dataset using a loss function that incorporates both class-specific weights to address label imbalance and group-specific weights to account for dataset-level distribution differences. Our final Challenge score on the hidden test set was 0.218, ranking 17th among the 41 eligible participating teams.

1. Introduction

Chagas disease, caused by *Trypanosoma cruzi*, remains a public-health concern in endemic regions and is associated with chronic cardiac manifestations detectable from the electrocardiogram (ECG). Automated detection from routine 12-lead ECGs would enable large-scale screening and earlier intervention. Recent ECG classification work has shown that deep models (CNNs, RNNs, Transformers) excel at morphology and rhythm recognition, while physiologically motivated handcrafted features bring interpretability and robustness across heterogeneous datasets.

The George B. Moody PhysioNet Challenge 2025 addresses the automated detection of Chagas disease from standard 12-lead ECGs, aiming to promote scalable, low-cost screening strategies for populations in endemic regions. The Challenge provides a unified evaluation framework based on heterogeneous ECG datasets collected from multiple international cohorts, including SaMi-Trop, PTB-XL, and CODE-15%, which differ in acquisition conditions and label reliability. This setting encourages the development of algorithms that are not only accurate but also interpretable and robust to domain shifts. [1–3]. Recent ECG studies have shown that combining raw-waveform deep representations with demographic and physiologically meaningful handcrafted features can enhance both robustness and interpretability of diagnostic models.

In this work, we ask whether (1) fusing interpretable R-peak/QRS morphological features and demographics with learned sequential embeddings and (2) using an auxiliary-conditioned, query-based cross-attention mechanism can improve automated Chagas detection from 12-lead ECGs aggregated across multiple sources (SaMi-Trop, PTB-XL, CODE-15%). To this end, we adopt a hybrid ResNet–BiGRU architecture in which a 1D ResNet with four residual stages extracts local morphological cues, a bidirectional GRU captures long-range temporal context, and auxiliary features (R-peak/QRS descriptors, age, sex) are projected into a query space that conditionally attends over sequence embeddings. To mitigate the practical challenges of label imbalance and distributional shift across heterogeneous datasets, we employ a source-aware weighted BCE loss with both class-specific and group-specific weights. Within the 2025 PhysioNet/CinC Challenge framework, we evaluate the approach on held-out test data, obtaining a score of 0.218, and ranking 17th among the 41 eligible participating teams. These findings support the hypothesis that compact, clinically meaningful features complement deep sequence representations, and that source-aware reweighting facilitates effective use of large but weakly labeled corpora for scalable Chagas dis-

ease screening from standard 12-lead ECGs.

2. Methods

2.1. Datasets and Preprocessing

We trained our models on data curated for the George B. Moody PhysioNet Challenge 2025, which includes ECGs from the SaMi-Trop, PTB-XL, and CODE-15% datasets [4–6]. The SaMi-Trop dataset provides high-confidence positive labels from 12-lead ECGs of Brazilian patients with serologically confirmed Chagas disease (2011–2012) [6]. The PTB-XL dataset contains ECGs from European patients in non-endemic regions and was used as a source of Chagas-negative controls [5]. The CODE-15% dataset, sampled from the broader Brazilian CODE repository [4], adds volume but relies on less reliable, patient-reported diagnostic labels.

All ECG recordings were standardized using a preprocessing pipeline that performed a sequence of deterministic operations. To ensure cross-dataset compatibility, sampling rate, signal duration, and lead order were aligned to the PTB-XL convention. First, all signals were resampled to 500,Hz to achieve a uniform temporal resolution and to match PTB-XL. The channels were then reordered into the canonical 12-lead configuration (I, II, III, aVR, aVL, aVF, V1–V6). To reduce inter-subject and inter-device variability, we applied per-lead median-based amplitude normalization by subtracting each lead's median and scaling by its interquartile range (IQR). We preferred median/IQR to mean-based scaling because means are more sensitive to outliers and can introduce amplitude distortions. Baseline wander was suppressed using a morphological open–close/close–open (OC/CO) filter [7], which attenuates low-frequency drifts while preserving the morphology of the QRS complex. Finally, all recordings were clipped or zero-padded to a fixed duration of 10 seconds (5,000 samples). No external data were used for model training, validation, or testing beyond these datasets.

2.2. Feature Extraction

We extracted a compact set of interpretable features from lead II using R-peak localization and QRS delineation. Specifically, we computed the median QRS width (in milliseconds) using a delineation routine that estimates onset and offset of each QRS complex within a ± 200 ms window centered around the detected R peaks; the mean QRS electrical axis (in degrees) obtained from the integrated areas under leads I and aVF within the delineated QRS intervals; an RSR' flag ratio in V1 defined as the fraction of beats exhibiting RSR' morphology (serving as an indicator of conduction disturbance); and a wide-S ratio computed from leads I and V6 as the fraction of beats with a deep/wide S wave that meets predefined amplitude and duration thresholds. All four morphology-derived features were standardized across the dataset. These features were then concatenated with demographic encodings (age and one-hot encoded sex) to form the auxiliary feature vector supplied to the model.

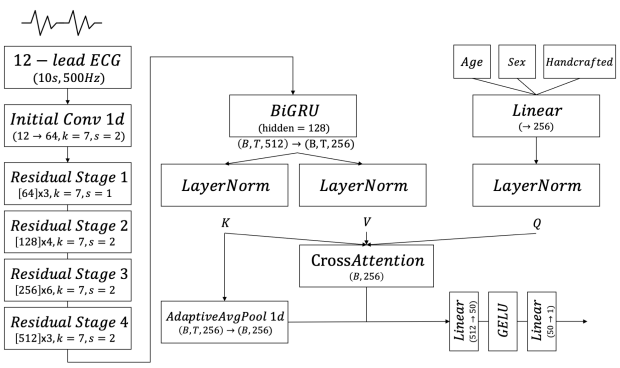

Figure 1: ResNet-BiGRU architecture

2.3. Data Augmentation

To enhance the robustness of the model, we applied lightweight, on-the-fly signal augmentations during training via the dataset loader. Each 12-lead ECG recording was subjected to random amplitude scaling, wherein the entire signal was multiplied by a scalar drawn from $\mathrm{Uniform}(0.85, 1.15)$. For temporal augmentation, long recordings were cropped by randomly shifting the start position by up to 0.5 s (implemented as an integer offset of up to 200 samples at 500 Hz), introducing mild temporal jitter without altering beat morphology. Additional augmentations (e.g., additive Gaussian noise, lead dropout) were implemented in the codebase but were not used in the principal experiments reported here.

2.4. Model Architecture

The proposed model architecture (Figure 1) combines convolutional and recurrent encoders with attention-based fusion. A 1D ResNet backbone first extracts local morphological features from 12-lead ECG signals. It begins with a Conv 1d layer that maps the 12 input channels to 64, followed by four residual stages composed of $\{3, 4, 6, 3\}$ blocks and output dimensions $\{64, 128, 256, 512\}$, using a fixed kernel size of 7 and strided convolutions for progressive temporal downsampling. Each block includes batch normalization and ReLU activations.

The output of the ResNet is reshaped to $(B, T, 512)$ and passed to a bidirectional GRU (hidden size 128 per direction), resulting in a 256-dimensional embedding per

time step that captures long-range temporal dependencies across the ECG signal. In parallel, the demographic variables (age and sex) and handcrafted ECG features are concatenated and projected into a 256-dimensional auxiliary token via a linear layer and layer normalization. Simultaneously, the BiGRU outputs are layer normalized and used as *Key* and *Value* in a cross-attention module, where the auxiliary token acts as the *Query*. A Multi-Head Attention (4 heads, 256-d embedding) enables the auxiliary features to selectively attend to relevant temporal patterns.

Finally, the temporally encoded sequence is compressed using adaptive average pooling to yield a fixed-size global representation. This pooled representation is concatenated with the output of the cross-attention module, forming a 512-dimensional fused feature vector. The fused vector is passed through a 2-layer MLP with dimensions 512→50→1, producing a single logit representing the predicted probability of Chagas disease.

2.5. Loss: Weighted BCE

To address both class imbalance and dataset-level reliability variations, we designed a training objective that combines two weighted binary cross-entropy (wBCE) terms: a positive-class weighted BCE (BCE_α) to handle label imbalance, and a source-aware BCE ($\text{BCE}_\alpha^{\text{source}}$) that modulates the contribution of each sample based on its origin.

$$\text{BCE}_\alpha(\hat{y}_i, y_i) = -\left[\alpha\, y_i \log \sigma(\hat{y}_i) + (1 - y_i) \log\left(1 - \sigma(\hat{y}_i)\right)\right]$$

$$\text{BCE}_\alpha^{\text{source}}(\hat{y}_i, y_i) = w_{s(i)} \cdot \text{BCE}_\alpha(\hat{y}_i, y_i)$$

In these equations, $y_i \in \{0, 1\}$ is the ground-truth label for sample i, $\hat{y}_i$ denotes the predicted logit, $\sigma(\hat{y}_i)$ is the corresponding sigmoid output, and $\alpha > 0$ is the global positive-class weighting factor (in our model, $\alpha = 5$). By increasing the loss on misclassified positive examples, this formulation helps mitigate class imbalance. Furthermore, to incorporate dataset source reliability, we introduce a dataset-specific weighting factor $w_{s(i)} > 0$, where $s(i)$ denotes the source of sample i. This allows the model to, for example, emphasize positive labels from high-confidence datasets like SaMi-Trop, while down-weighting contributions from noisier sources such as CODE-15%.

The final loss is a convex combination of the positive-class weighted and source-aware BCE terms, weighted by a mixing coefficient $\lambda \in [0, 1]$. All caling factors (α, $w_{s(i)}$) and the interpolation factor λ were treated as hyperparameters and selected based on performance on a held-out validation set.

$$\text{wBCE}(\hat{y}_i, y_i) = \frac{\lambda}{N} \sum_{i=1}^{N} \text{BCE}_\alpha(\hat{y}_i, y_i) + \frac{1-\lambda}{N} \sum_{i=1}^{N} \text{BCE}_\alpha^{\text{source}}(\hat{y}_i, y_i)$$

3. Experiments

3.1. Experimental Setup

The development dataset was constructed by merging SaMi-Trop, PTB-XL, and CODE-15%. For early-stage experiments and ablations, we used a pseudo-training partition comprising one-sixth of the training set, including only 10% of CODE-15% due to its lower label reliability. In later experiments with source-aware weighted loss, the full CODE-15% set was reinstated.

Evaluation was performed using two held-out configurations. The primary evaluation set included only SaMi-Trop and PTB-XL samples to avoid noisy labels, while the secondary evaluation set included a subset of CODE-15%. To better reflect endemic screening conditions, the prevalence of Chagas-positive cases in the primary evaluation set was adjusted to approximately 2.5%.

Models were trained with the Adam optimizer (initial learning rate 1×10^{-4}, batch size 36) using cost-sensitive learning with a fixed positive-class weight $\alpha = 5$. For model selection, we used a stratified 90/10 training split with early stopping (10-epoch patience) based on validation loss. The checkpoint with the lowest validation loss was retained for final evaluation, with training capped at 100 epochs.

3.2. Experimental Results

Table 1 presents the results across model variants. The "Eval (w/o CODE-15%)" column reflects performance on the same held-out partition after excluding CODE-15%, and "Challenge score" denote the validation score obtained during the competition phase through submissions to the official PhysioNet/CinC Challenge evaluation server.

The ResNet–BiGRU baseline was trained without additional handcrafted features. Because CODE-15% constitutes a large fraction of the aggregated corpus and its Chagas labels were judged to be less reliable than those in SaMi-Trop and PTB-XL, only 10% of the available CODE-15% examples were used when training the baseline. Under these conditions the baseline achieved 0.388 on our held-out eval set, 0.576 on the held-out eval set excluding CODE-15%, and 0.305 on the challenge validation set.

Augmenting the baseline with R-peak–derived features by simple concatenation, along with the data augmentation, improved performance on our internal evaluation (0.424) and on the held-out eval set without CODE-15% (0.736), indicating that morphology-based features give useful complementary information. Replacing concatenation with the proposed cross-attention fusion of R-peak features and demographic features (age and sex) led to further gains: the cross-attention variant achieved a Challenge

Model variant	Challenge score	Eval	Eval (w/o CODE-15)
ResNet-BiGRU	0.305	0.388	0.576
+ New features concatenate + Data augmentation	N/A	0.424	0.736
+ cross-attention	0.317	0.460	0.760
+ wBCE($\lambda : 0.6, w_{CODE} : 0.2, w_{SaMi} : 3.5, w_{PTB} : 1.5$)	N/A	0.480	0.672
+ wBCE($\lambda : 0.5, w_{CODE} : 0.2, w_{SaMi} : 3.5, w_{PTB} : 7.5$)	0.337	0.504	0.808
+ wBCE($\lambda : 0.6, w_{CODE} : 0.8, w_{SaMi} : 2.5, w_{PTB} : 7.5$)	0.347	0.512	0.88

Table 1: Performance of model variants. "Our evalset (w/o CODE-15)" excludes CODE-15; "N/A" means no official submission; "Challenge score" is the validation score from the challenge server during the competition phase.

Training	Validation	Test	Ranking
0.512	0.347	0.218	17/41

Table 2: Challenge scores for our selected entry (team CAUETUMN), including the ranking of our team on the hidden test set.

score of 0.317 and internal scores of 0.460 and 0.760 for the eval set with and without CODE-15%, respectively.

We also applied a positive-class weighted and source-aware BCE loss (wBCE), which permits inclusion of the entire CODE-15% corpus by down-weighting its contribution via source-specific scaling factors. Training with the full CODE-15% set under wBCE improved internal performance across the reported experiments, indicating that source-aware reweighting can harness the representational benefit of the large CODE-15% corpus, while mitigating label noise. This approach yielded consistent improvements, achieving our best Challenge validation score of 0.347. However, wBCE variants incurred substantially longer runtimes; as a result, we were limited to nine training epochs, since extending to twelve epochs would have exceeded the 72-hour wall-time limit.

Table 2 summarizes the official Challenge scores obtained using our proposed ResNet-BiGRU framework with conditioned query-based cross-attention and source-aware wBCE loss. Our model achieved a score of 0.347 on the hidden validation set and 0.218 on the hidden test set, ranking 17th among 41 eligible participating teams.

4. Conclusions

We presented a hybrid deep learning architecture combining a ResNet–BiGRU backbone with a query-based cross-attention mechanism for automated Chagas disease screening using 12-lead ECGs. The model integrates learned sequential embeddings with physiologically interpretable R-peak/QRS morphology and patient demographics features. The model achieves strong internal discrimination and, with source-aware training, outperforms a feature-agnostic baseline; our official submission reached a score of 0.218. These results indicate that concise clinical cues enhance deep representations and that source-aware loss reweighting allowed the model to leverage large-scale but weakly labeled datasets.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00341075, RS-2023-00208284)

References

[1] 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.

[2] 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.

[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] Ribeiro AH, Paixao GM, Lima EM, Horta Ribeiro M, Pinto Filho MM, Gomes PR, et al. Code-15%: a large scale annotated dataset of 12-lead ecgs, June 2021.

[5] 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).

[6] Ribeiro ALP, Ribeiro AH, Paixao GM, Lima EM, Horta Ribeiro M, Pinto Filho MM, et al. Sami-trop: 12-lead ecg traces with age and mortality annotations, June 2021.

[7] Sun Y, Chan KL, Krishnan SM. Ecg signal conditioning by morphological filtering. Computers in Biology and Medicine 2002;32(6):465–479. ISSN 0010-4825.

Address for correspondence:

Il-Youp Kwak

84, Heukseok-ro, Dongjak-gu, Seoul 06974, South Korea

ikwak2@cau.ac.kr