Chagas Disease Detection from 12-Lead ECG by Combining Self-Attention Enhanced Residual Networks with Customed Joint Loss

Zijie Zhu[1], Shuang Qiu[1], Chen Xia[1], Jinyu Wang[1], Hedi Li[1], Xiyuan Wang[1], Yan Du[1], Pan Xia[1]

[1]Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China

Abstract

The George B. Moody PhysioNet Challenge 2025 focused on detecting potential cases of Chagas disease from standard 12-lead electrocardiograms (ECGs). Our team, Kust_MeAI, proposed an approach to predict Chagas disease by combining an 18-layer residual neural network and a joint loss consisting of focal loss and precision loss. Firstly, we retain all samples from the SaMi-Trop and PTB-XL datasets and select a portion of samples from the CODE-15% dataset as the training set. All selected recordings then underwent the same series of preprocessing steps including denoising, cropping and normalization. Secondly, we built an 18-layer residual network with multi-head self-attention to extract the Chagas disease's complex pathological patterns from ECG records. Thirdly, to alleviate the severe class imbalance problem and achieve accurate detection of Chagas disease, we perform mixup data augmentation operation on all samples and then jointly optimize the proposed model using focal loss and a custom precision loss. Preprocessed 12-lead ECG segments were used as model inputs for end-to-end training, and the prediction probabilities of positive and negative categories were produced as model outputs. Finally, our proposed approach received the Challenge score of 0.195 (ranked 26th out of 40 teams) on the hidden test set.

1. Introduction

Chagas disease is a tropical parasitic disease caused by Trypanosoma cruzi and is transmitted mainly by triatomine bugs, also known as “kissing bugs”. It affects an estimated 6.5 million people in endemic countries, and causes nearly 10,000 deaths annually [1]. Serological testing is effective for diagnosing individual patients, but its limited capacity makes large-scale screening impractical. The electrocardiogram (ECG) provides a low-cost, non-invasive alternative, as symptoms of Chagas cardiomyopathy are often visible on it, offering a valuable opportunity to prioritize high-risk individuals for confirmatory testing by pre-screening large populations [2].

George B. Moody PhysioNet Challenge 2025 Invites teams to develop approaches that leverage ECGs to prioritize patients for confirmatory testing for Chagas disease, advancing scalable screening and more efficient allocation of limited diagnostic resources [3-5]. However, creating an effective algorithm is non-trivial, primarily due to the substantial data-centric hurdles of severe class imbalance from the disease's low prevalence and significant label noise from large, weakly-labeled data sources. These challenges can mislead standard models, suppressing their ability to detect the rare positive cases that matter most.

In this work, we present a tailored deep learning approach based on a residual network with multi-head self-attention, which achieves robustness against difficult data conditions through a joint loss function that mitigates noisy labels and handles class imbalance.

2. Methods

2.1. Datasets and Preprocessing

The Challenge train data were compiled from three sources, including Central and South American datasets (CODE-15% and SaMi-Trop), the PTB-XL dataset, and several private datasets from Chagas-endemic regions [6-8]. The Challenge data comprise of the public training set, the hidden validation set and the hidden test set. The data contains standard 12-lead ECG recordings, basic demographic variables, and binary labels denoting Chagas disease status [3]. The public training set includes more than 323,000 recordings drawn from three sources CODE-15% dataset with weak labels, SaMi-Trop dataset, and PTB-XL dataset, where the latter two datasets have strong labels that have been manually verified.

We counted 8,190 Chagas-positive and 357,995 negative cases in the public training set (366,185 total). The positive prevalence is only 2.24% (1:43.7), indicating that there is a serious long-tail class imbalance problem in the public training set. Considering that the CODE-15% dataset is too large and too many negative samples may

mislead model training, we randomly sampled 6,000 positive and 6,000 negative records from CODE-15%. We retained all samples from the other two databases, PTB-XL dataset and SaMi-Trop dataset. We apply the following data pre-processing procedures for all ECG recordings. Firstly, we resample all selected ECGs to 100 Hz, and then apply a 3rd-order Butterworth band-pass filter (0.5-45 Hz) for per lead to suppress baseline and electromyography (EMG) artifact. Secondly, the sliding average filter with a window of N=Fs/50 samples is used to suppress the power frequency noise contained in the ECGs. From each recording, the first 10 seconds segment were retained, and segment shorter than 10 seconds is zero-padded to 1,000 samples to ensure uniform length. The 12 leads are stacked to form a (12, 1000) array. Thirdly, each record is z-score standardized over all leads and time points within the record, with any NaN/Inf values replaced by 0.01.

2.2. Model Architecture

Deep residual networks [9] introduce skip connections that allow gradients and information to propagate more directly through very deep networks, thereby improving feature learning efficiency. Their effectiveness has been extensively studied and validated [10]. An 18-layer residual network was used in our approach. The overall structure of the model is shown in Figure 1. Preprocessed ECG signals were fed into the network after the *mixup* data augmentation operation. The input shape of main network was 12*1000, the input data was first fed into a 1D convolution with a (1, 15) kernel and a stride of 2. 8 residual blocks were stacked to form the backbone of network. We use a single multi-head self-attention (MHA) layer to perform dynamic feature aggregation in the full temporal domain and model inter-lead relationships [11]. The MHA is based on dot-product attention is given in Eq. 1:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

$$MultiHead(Q,K,V) = Concat(head_1,\dots head_h)W^o \quad (2)$$

$$where\ head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

Where the projections are parameter matrices $W_i^Q \in \mathbb{R}^{d_{model}\times d_k}$,$W_i^K \in \mathbb{R}^{d_{model}\times d_k}$,$W_i^V \in \mathbb{R}^{d_{model}\times d_k}$,$W^O \in \mathbb{R}^{hd_k\times d_{model}}$.

In this work, we use self-attention ($Q = K = V = X$). We reshape features to $X \in \mathbb{R}^{L\times d_{model}}$, d_{model}=12 × 32=384. And we employ $h = 8$ parallel attention heads. For each head, we use d_k=d_v=$\frac{d_{model}}{h}$ = 48.

After GAP, the 384-dimensional features is passed to a single classification head. The classification head converts

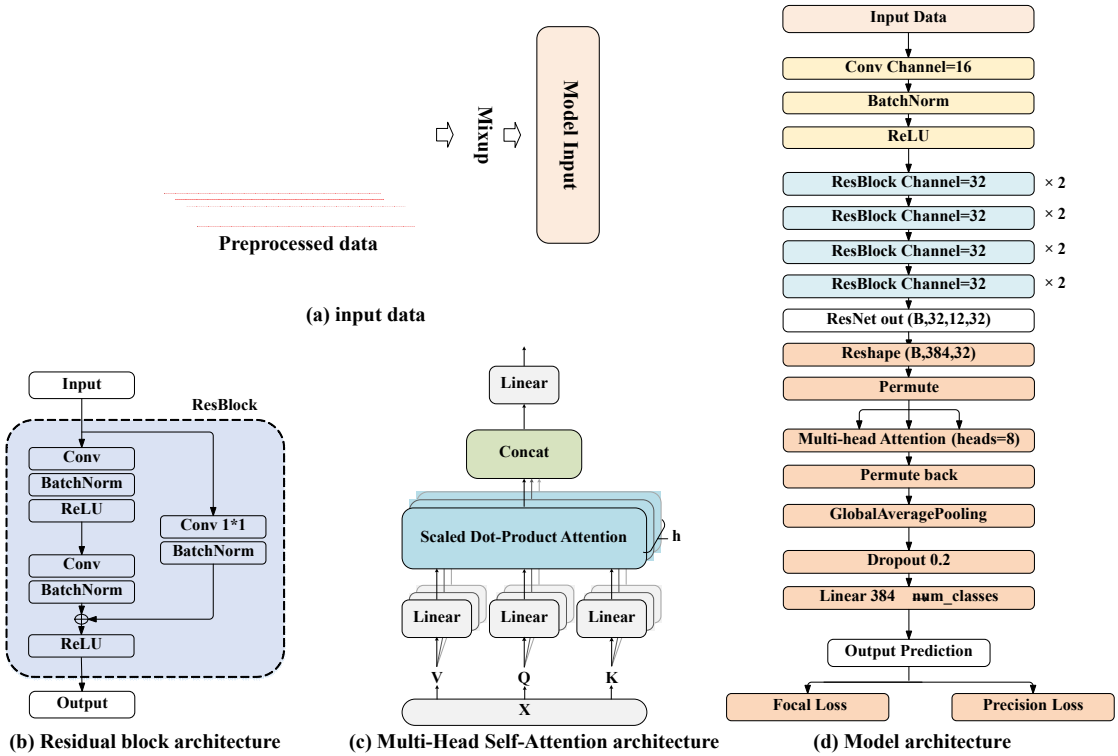

Figure 1. The architecture diagram of our proposed approach. (a) Input data for the modified residual neural network. (b) Residual block architecture. (c) Multi-Head Self-Attention architecture. (d) The model architecture of modified residual neural network.

the 384-dimensional feature vector to 2 (number of categories) by using linear operations with dropout rate 0.2. And it is directly converted into category probability through *Softmax* in the forward propagation. Finally, the probability distribution output by the classification head is passed to the focal loss function (to alleviate category imbalance and strengthen the learning of difficult-to-classify samples) and the precision loss function (to increase the probability of positive samples being predicted as positive), and the above two loss functions are added together as the total loss of the proposed model.

To improve the generalization ability and robustness of our proposed model, *Mixup* is adopted in our approach to construct more patterns of samples to expand sample diversity. *Mixup* trains on virtual examples constructed as the linear interpolation of two random examples from the training set and their labels [12]. Assuming that $(x_i,\ y_i)$ and $(x_j,\ y_j)$ are two examples drawn at random from our training data in one batch, the mixup samples $(\tilde{x}, \tilde{y})$ are created as defined in Eq. (4) and Eq. (5).

$$\tilde{x} = \omega x_i + (1-\omega)x_j \quad (4)$$

$$\tilde{y} = \omega y_i + (1-\omega)y_j \quad (5)$$

Where mixing coefficient $\omega \in [0, \varphi]$ is sampled from a Beta $(\varphi,\ \varphi)$ distribution, with the φ set to 0.1. The model is trained using the generated *mixup* samples $(\tilde{x}, \tilde{y})$.

2.3. Loss Function

We combine focal loss with precision loss for model training. The total loss formulation is given in Eq. 6.

$$L = l_1 + \lambda * l_2 \quad (6)$$

Where the λ denotes a hyper parameter for balancing the two loss functions. In this work, $\lambda = 1$.

Focal Loss: In the subset actually loaded for training, we used 27,799 negatives and 7,631 positives (35,430 total). Among them, 6,000 negatives and 6,000 positives were appended from CODE-15%, which provides weak labels with severe label noise. Although adding weakly labeled sample from CODE-15% alleviated class imbalance, training remains dominated by easy negatives. This loss imbalance suppresses gradients from hard/rare positives, limiting recall and AUPRC.

Therefore, we adopt the focal loss to handle this problem by down-weighting easy examples and emphasizing hard examples [13]. It is defined in Eq. (7).

$$l_1 = -\alpha_t(1-p_t)^{\gamma} * log(p_t) \quad (7)$$

Where p_t denotes the model's corresponding prediction probability for the true label y_t. The weighting factor α and the tunable focusing parameter γ are set to 0.25 and 4, respectively, during training to more strongly down-weight easy examples, while γ is set to 2 for validation, with α kept at 0.25. The total average is considered as the final loss.

Precision Loss: In Chagas screening, confirmatory serology is scarce excessive false positives (FP) would overload limited testing capacity, misallocate care, and increase patient anxiety. To explicitly curb such errors, we introduce a batch-level precision penalty (Eq. 8)

$$l_2 = 1 - Precision \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

Where TP denotes the number of patients with true good labels who are predicted as good. Where FP denotes the number of patients with true poor labels who are predicted as good.

As above, we combine this precision loss with focal loss. The focal component maintains high sensitivity to likely Chagas cases, while the precision term prevents over-triage of healthy patients, yielding a balanced objective suited to resource-constrained screening.

2.4. Model Training

Each model is trained 50 epochs with a batch size of 64 using a NVIDIA GeForce RTX 5090. Adam with an initial learning rate of 0.0005 was applied for model optimization. We used a custom dynamic learning-rate schedule. Starting at 0.0005, the learning rate was updated at the end of every epoch according to a predefined decay table to encourage stable convergence and improve generalization. Model training is stopped when the model's score on the validation set does not improve after 13 epochs.

The category decision threshold is set to 0.5 for classification task, respectively. Other hyper-parameter of the network (convolution kernel size, dropout rate, number of convolution layer, etc.) were adjusted according to the model 5-fold cross validation performance on the public training dataset to achieve optimal performance.

3. Results

We evaluated our proposed algorithms through 5-fold cross-validation on the public training set with the Challenge evaluation metric. The Challenge scores on both the public training set, hidden validation set, and hidden test set that our final selected entry obtained were shown in Table 1.

Training	Validation	Test	Ranking
0.216±0.003	0.201	0.195	26/40

Table 1. Official Challenge scores for our final selected entry (team Kust_MeAI), 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.

4. Discussion and Conclusions

Although our model achieved strong performance when

trained on the PTB-XL and SaMi-Trop datasets with reliable labels, the performance substantially deteriorated after incorporating 6,000 negative and 6,000 positive samples from the CODE-15% dataset. We consider that the weak labels of CODE-15% introduced significant label noise during training. Although we attempted label cleaning to mitigate this issue, the improvements were negligible. In addition, the distributional differences between CODE-15% and the strongly labeled datasets (PTB-XL and SaMi-Trop) may also have contributed to the poor performance of our model.

A broader comparison to other Challenge entries contextualizes our results. In contrast to our end-to-end trained residual network, the winning team (Biomed-Cardio) used a Vision Transformer pre-trained on over 400,000 ECGs for superior generalization [14]. Furthermore, while we used a custom joint loss function to mitigate data challenges, other leading teams implemented more explicit noise-handling strategies. The second-place team (DlaskaLabMUI) employed self-supervised pre-training with noise-robust fine-tuning [15], while the third-place team (AIChagas) used a soft-labeling strategy to manage uncertainty in self-reported labels and a combined loss of binary cross-entropy (BCE) and margin ranking loss to mitigate overfitting on potential outliers [16]. This indicates that the key to excelling in challenging tasks lies in data-centric solutions that are custom-tailored to address specific problems.

In this paper, we proposed a novel approach to detect potential cases of Chagas disease from standard 12-lead ECGs by combining an 18-layer residual neural network with multi-head self-attention and a composite loss that couples focal loss with precision loss, trained end-to-end with *mixup* to handle severe class imbalance and curb false positives. Finally, our model received Challenge score of 0.195 (ranked 26th out of 40 teams) on the hidden test set.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant NO. 62503206) and Yunnan Fundamental Research Projects (Grant NO. 202401BE070001-029).

References

[1] Cucunubá ZM, Gutiérrez-Romero SA, Ramírez J-D, Velásquez-Ortiz N, Ceccarelli S, Parra-Henao G, et al. The epidemiology of Chagas disease in the Americas. *Lancet Regional Health - Americas* 2024;37:100881.

[2] Jidling C, Gedon D, Schön TB, Oliveira CDL, Cardoso CS, Ferreira AM, et al. Screening for Chagas disease from the Electrocardiogram using a deep neural network. *PLOS Neglected Tropical Diseases* 2023;17(7):1–17.

[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; In *Computing in Cardiology 2025*, volume 52: 2025, 1-4.

[4] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation* 2000;101(23):e215–e220.

[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; 2025: URL: https://arxiv.org/abs/2510.02202, DOI: 10.48550/arXiv.2510.02202.

[6] Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature Communications* 2020;11:1760.

[7] Cardoso CS, Sabino EC, Oliveira CDL, de Oliveira LC, Ferreira AM, 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):e011181.

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

[9] He K, Zhang X, Ren S, and Sun J. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016;770–778.

[10] Strodthoff N, Wagner P, Schaeffter T, and Samek W. Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *IEEE Journal of Biomedical and Health Informatics* 2021;25(5):1519–1528.

[11] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Advances in Neural Information Processing Systems* 2017;30:5998–6008.

[12] Zhang H, Cisse M, Dauphin Y N, et al. Mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412, 2017.

[13] Lin T-Y, Goyal P, Girshick R, He K, and Dollár P. Focal loss for dense object detection. *2017 IEEE International Conference on Computer Vision (ICCV)* 2017;2999–3007.

[14] Santvliet LV, Nguyen PX, Vandenberk B, Vos MD. Detecting Chagas disease using a Vision Transformer-based ECG foundation model. *Computing in Cardiology* 2025;52:1-4.

[15] Nicolson A, Lunelli R, Proell SM, Gruber N, Bauer A, Dlaska C. Transformer–xLSTM ensembles for ECG-based Chagas disease detection. *Computing in Cardiology* 2025;52:1-4.

[16] V´azquez SG, Chou C. Managing label uncertainty in the detection of Chagas disease from the ECG. *Computing in Cardiology* 2025;52:1-4.

Address for correspondence:

Yan Du and Pan Xia
Faculty of Information Engineering and Automation, Kunming University of Science and Technology, No. 727, Jingming South Road, Chenggong District, Kunming, China.
yandu@kust.edu.cn; xiapan17@mails.ucas.edu.cn