Hybrid Convolution-Transformer Network Integrating Time and Frequency Domains from 12-Lead ECG for Chagas Disease Detection

HyeonWoo Lee[1], Junmo An[1]

[1]Philips, Cambridge, Massachusetts, USA

Abstract

Chagas disease is a leading cause of non-ischemic cardiomyopathy, and its long asymptomatic phase makes timely case detection challenging. For the George B. Moody PhysioNet 2025 Challenge, we participated under the team name PFAP and developed an end-to-end deep learning model for automated screening from 12-lead ECGs. We compared two inputs: ECG-only and a dual-domain input that combines ECG with per-lead power spectral density (PSD) to capture complementary frequency information. We also propose a lightweight multi-scale Convolutional Neural Network (CNN)–Transformer model that couples local morphology with longer-range temporal patterns, and we apply age- and sex-aware negative sampling to mitigate severe class imbalance. On an internal balanced validation dataset, the dual-domain model outperformed the ECG-only baseline across all metrics. On an internal training dataset with the original prevalence, it achieved AUROC 0.92, AUPRC 0.77, F1 0.68, and a Challenge score of 0.23, while the official test phase reported 0.205. These results show that PSD provides complementary features and that targeted negative sampling enhances generalization, while the compact CNN-Transformer supports efficient large-scale screening.

1. Introduction

Chagas disease, caused by the protozoan *Trypanosoma cruzi*, poses a major public health challenge in Latin America and is increasingly observed worldwide due to migration [1]. Following an acute phase, many patients enter a prolonged asymptomatic period, yet a substantial proportion eventually develop chronic Chagasic cardiomyopathy, a progressive condition that can lead to arrhythmias, heart failure, and sudden cardiac death [2]. The long latency before clinical manifestation makes early detection of cardiac involvement particularly difficult.

The electrocardiogram (ECG) remains an essential, non-invasive, and affordable tool for identifying cardiac abnormalities, such as right bundle branch block, left anterior fascicular block, and ventricular extrasystoles, which are frequently associated with chronic Chagasic cardiomyopathy [3]. However, manual ECG interpretation is prone to variability and can be time-intensive, underscoring the value of automated diagnostic approaches.

The George B. Moody PhysioNet Challenge 2025 [4], [5], [6] addresses this need by providing a large annotated dataset of 12-lead ECGs from Chagas-positive and Chagas-negative individuals, encouraging the development of AI-driven screening solutions. This study presents our PFAP team's approach to the challenge, detailing data preprocessing, model architecture, and evaluation, and discusses the potential of these methods for scalable early detection of Chagas-related cardiac disease.

We propose a practical deep learning-based ECG classifier that augments time-domain 12-lead signals with frequency-domain power spectral density (PSD), yielding a dual-domain input representation. PSD captures band-limited oscillatory power that is relatively robust to noise and reflects physiological state that may not be fully expressed in the raw time series. Prior findings in neurophysiology demonstrate that behavioral and reward states can significantly modulate spectral power and spike-field coupling, as quantified using Welch-based PSD analyses, suggesting that spectral features not only capture ongoing neural activity but are also highly sensitive to underlying latent dynamics in the motor cortex [7], [8]. To evaluate the contribution of PSD, we compared two variants: an ECG-only model and an ECG+PSD model. Both employ a lightweight multi-scale one-dimensional convolutional neural network (1D CNN)–Transformer backbone with top-down feature fusion. The CNN blocks capture local temporal features, while shallow Transformer encoders model long-range temporal dependencies.

Given the rarity of Chagas-positive cases, we address severe class imbalance using age- and sex-aware on-the-fly negative sampling during training. We evaluate both variants on a balanced validation set and on a held-out test set that preserves the original prevalence. The results demonstrate the feasibility of this approach for scalable detection of Chagas-related cardiac disease.

2. Methods

2.1. Dataset and Preprocessing

Our dataset consists of ECG records from CODE-15% [9], SaMi-Trop [10], and PTB-XL [11]. We use 12-lead ECGs along with associated header metadata (e.g., age and sex). Each record is resampled to 300 Hz and either zero-padded or truncated to 10 seconds (3,000 samples). A 4th-order Butterworth band-pass filter (0.5–50 Hz) is applied to reduce baseline wander, motion artifacts, muscle noise, and high-frequency interference, and per-lead Z-score normalization standardizes amplitudes.

For the frequency domain, we compute per-lead power spectral density (PSD) using Welch's method (nperseg = 64, 50% overlap). The resulting PSD traces are then resampled to match the target time length (3,000 samples), ensuring alignment between the time- and frequency-domain tensors. The PSDs are normalized separately and subsequently concatenated channel-wise with the time-domain leads, resulting in a 24-channel input (12 ECG + 12 PSD).

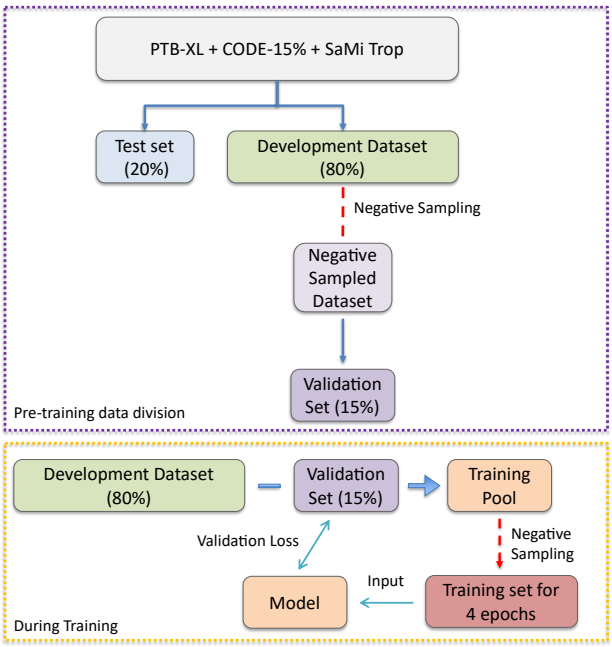


Figure 1. Negative data sampling and dataset splits used before and during training.

2.2. Data Augmentation and Sampling

Data augmentations are applied only to the time-domain signals, and the PSDs are computed from the augmented ECGs to naturally augment the PSDs, which directly correspond to the augmented signals. With a probability of 0.9 per batch, we randomly select and apply 1–4 of the following operations: lead inversion (per-lead flip), baseline-wander injection via a sine wave (fs = 300 Hz; frequencies ∈ {0.05, 0.1, 0.15, 0.2, 0.5} Hz; amplitude = 0.1), magnitude scaling (sampled uniformly within a predefined range), and temporal masking (~10% of a contiguous segment zeroed) [12], [13], [14]. Augmentations are disabled during the final 10 training epochs to promote stable convergence.

For evaluation under the natural Chagas distribution, the full dataset is randomly divided into a development set (80%) and an internal test set (20%). Within the development set, we construct a negative-sampled pool to address class imbalance: negative samples are grouped by age bins (<30, 31–50, 51–70, ≥70) and sex, then sampled evenly to achieve a 1 (positive) to 3 (negative) ratio. From this pool, a stratified internal validation set (15%) is fixed and used exclusively for model checkpoint selection. The training pool comprises the remaining records. During training, the negative sampler refreshes the training set every four epochs by drawing from the training pool using the same age- and sex-aware procedure, thereby maintaining both class balance and demographic coverage over time. The overall procedure is illustrated in Figure 1.

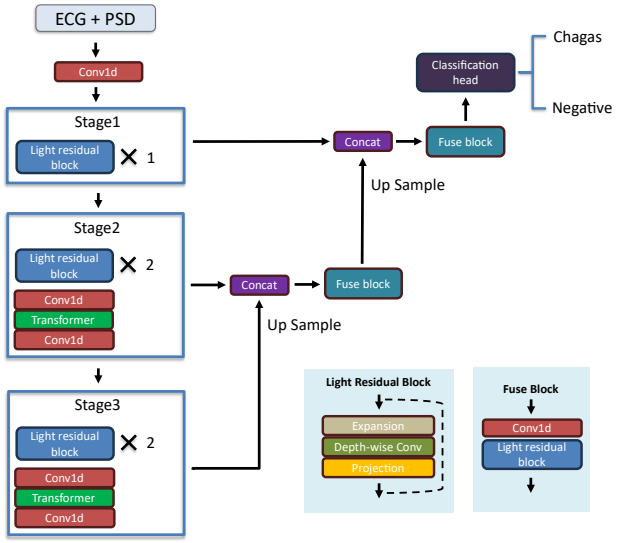


Figure 2. The proposed architecture of the end-to-end hybrid convolution-transformer with top-down feature fusion.

2.3. Deep Learning Model

We study two variants: ECG-only (12 channels) and ECG+PSD (24 channels). In the dual-domain setting, the model takes 12 time-domain leads and 12 per-lead PSD traces, concatenated along the channel axis. A 1D convolutional stem extracts low-level features in both variants. Subsequent stages use depthwise separable convolutions with residual connections to model local morphology, followed by shallow Transformer encoders to capture longer-range temporal dependencies [15], [16].

Lightweight residual blocks follow the MobileNetV2 inverted-residual pattern: a 1×1 expansion, a depthwise convolution, and a 1×1 projection. A skip connection is applied only when the input and output channel numbers

are identical and the stride is 1; when channels increase or temporal down-sampling is applied, the skip connection is omitted [15]. The network is organized into three stages. Stage 1 uses a light residual block without a skip connection. Stages 2 and 3 each apply a no-skip block for down-sampling or channel increase, followed by a skip-enabled block when the input and output shapes match. A shallow Transformer encoder with learned positional embeddings follows the pair in each stage. A top-down fusion pathway up-samples deeper features and merges them with earlier stages to recover temporal resolution (Figure 2). The fused representation is aggregated via global average pooling and mapped to a single logit for binary classification of Chagas disease.

2.4. Implementation

We trained for 60 epochs (batch size 28) under a 64-hour wall-clock limit, using approximately 75% of the available steps per epoch to reliably meet the competition constraint. Data augmentations were applied consistently until the final 10 epochs, and the sampled training subset was refreshed every four epochs.

Two variants (ECG-only and ECG+PSD) were trained to assess the contribution of PSD. The architecture uses stride-2 convolutions with Batch Normalization and GELU activation, followed by three feature stages: Stage 1 (64 channels, residual blocks), Stage 2 (128 channels, Transformer block with d_model = 128, 4 heads), and Stage 3 (256 channels, Transformer block with d_model = 256, 8 heads). Top-down fusion merges upsampled higher-level features with lower-level features via 1×1 projections and residual refinement. The classifier applies global average pooling, BatchNorm, GELU, dropout (0.2), and a final linear layer. Data augmentations were applied to ECG signals, and for dual-domain runs, PSDs were recomputed from the augmented waveforms. Optimization used AdamW (weight decay = 0.001) with a one-cycle learning rate schedule (max_lr = 5e-4, warm-up ≈ 5%) and binary cross-entropy loss (positive weight = 3), with early stopping (patience = 10).

To address class imbalance and stabilize model selection, we fix a stratified validation set comprising 15% of a carefully balanced pool and refresh the training set every four epochs using age- and sex-aware negative sampling to maintain a 1:3 positive-to-negative ratio. Checkpoints are selected via a validation threshold sweep from 0.001 to 0.999 to identify the optimal classification threshold that maximizes the F1 score. We report the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPRC), accuracy, F1, and the official Challenge score on both the fixed validation set and the held-out test set.

2.5. Metrics and Model Selection

On the fixed validation set, we compute the Challenge's sensitivity-centric score, AUROC, AUPRC, accuracy, and F1. For model selection, we use a composite metric:

$$Selection\ score = F1 + 2 \times Challenge\ Score,$$

reflecting the competition's emphasis on sensitivity while moderating false positives. The checkpoint with the highest selection score is retained for validation. To avoid threshold bias, thresholds are swept from 0.001 to 0.999, and the threshold that maximizes F1 is selected.

3. Results

We validated each checkpoint on a fixed validation set from a sex- and age-balanced pool. For each configuration, we swept the binarization threshold and selected the checkpoint with the highest selection score on the validation set. Table 1 reports the metrics for the selected checkpoints. The optimal validation thresholds were 0.462 for ECG-only and 0.57 for ECG+PSD. The ECG+PSD variant outperformed ECG-only across all metrics.

Table 1. Performance on the *internal* balanced validation set for models using ECG-only and ECG+PSD inputs.

Input	AUROC	AUPRC	Acc.	F1	Challenge Score	Opt. Threshold
ECG Only	0.86	0.67	0.76	0.62	0.16	0.46
ECG+PSD	0.87	0.69	0.80	0.65	0.18	0.57

We then evaluated the same checkpoints on the internal test set that preserves the original data distribution. Each model's validation-optimal threshold was applied to the test set: 0.462 for ECG-only and 0.57 for ECG+PSD, a 23.4% relative increase. Table 2 summarizes the results. All metrics improved relative to the validation set, and the dual-domain model outperformed the ECG-only model across all metrics. The best ECG+PSD checkpoint achieved a Challenge score of 0.317 on the hidden validation set and 0.205 on the official held-out test set.

Table 2. Model performance on the internal test set with the original data prevalence.

Input	AUROC	AUPRC	Acc.	F1	Challenge Score	Opt. Threshold
ECG Only	0.86	0.68	0.76	0.63	0.2	0.46
ECG+PSD	0.92	0.77	0.86	0.68	0.23	0.57

4. Discussion and Future Work

Our work demonstrates that adding per-lead PSD to time-domain ECG improves Chagas screening performance using a compact CNN-Transformer hybrid model. Across both the validation and held-out test sets, the dual-domain model outperformed the ECG-only

baseline on AUROC, AUPRC, F1, and the Challenge score. This improvement arises from frequency-domain information that is not fully expressed in the waveform alone, which is captured by the PSD domain. The results also indicate that the multi-scale backbone and top-down fusion preserve the temporal details necessary for accurate classification. The increase in all metrics on the internal test set with the original prevalence suggests that the training strategies and thresholds selected on a balanced validation pool generalize well to the natural prevalence.

Age- and sex-aware negative sampling appears to stabilize learning under extremely imbalanced datasets, such as the Chagas dataset. Refreshing the sampled pool during training reduces overfitting to a narrow subset of negatives and promotes better generalization. The gap between the held-out test score and the official phase test score suggests remaining calibration differences and a possible distribution shift. Thresholds tuned on a balanced pool may underweight sensitivity relative to the Challenge metric, and the current selection rule optimizes F1 rather than a surrogate of the official score.

There are several limitations. First, PSD parameters were fixed, and a systematic sweep (e.g., window length, overlap, frequency range) is needed to ensure that relevant spectral information is captured. Second, our sampling relies on under-sampling of negatives, which may reduce data diversity, and we did not compare against reweighting or class-balanced loss functions. Third, we did not perform ablation studies to isolate the contributions of PSD, Transformer depth, or the fusion pathway.

Future work can focus on optimization and robustness. Prediction confidence scores can be calibrated to expected prevalence, and decision thresholds adjusted for high sensitivity or positive predictive value (PPV). Performance can be re-evaluated under different prevalences. Controlled ablation studies can quantify PSD, Transformer, and fusion contributions. The data strategy can shift from hard under-sampling to dynamic reweighting to preserve negative diversity and demographic balance. Incorporating uncertainty estimation and reporting subgroup metrics by age and sex would enhance transparency and fairness. Profiling latency on edge hardware can confirm feasibility for ECG devices. Interpretable analyses, such as saliency mapping over leads and frequency bands, could support clinical review and error analysis.

References

[1] Z. M. Cucunubá *et al.*, "The epidemiology of Chagas disease in the Americas," *The Lancet Regional Health–Americas*, vol. 37, 2024.

[2] E. C. Sabino, M. C. P. Nunes, J. Blum, I. Molina, and A. L. P. Ribeiro, "Cardiac involvement in Chagas disease and African trypanosomiasis," *Nature Reviews Cardiology*, vol. 21, no. 12, pp. 865–879, 2024.

[3] C. Jidling *et al.*, "Screening for Chagas disease from the electrocardiogram using a deep neural network," *PLoS Neglected Tropical Diseases*, vol. 17, no. 7, p. e0011118, 2023.

[4] M. A. Reyna *et al.*, "Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025," *arXiv preprint arXiv:2510.02202*, 2025.

[5] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[6] M. A. Reyna *et al.*, "Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025," *Computing in Cardiology 2025*, no. 52, pp. 1–4, 2025.

[7] J. An, T. Yadav, J. P. Hessburg, and J. T. Francis, "Reward expectation modulates local field potentials, spiking activity and spike-field coherence in the primary motor cortex," *eNeuro*, vol. 6, no. 3, 2019.

[8] J. An, T. Yadav, M. B. Ahmadi, V. S. A. Tarigoppula, and J. T. Francis, "Near perfect neural critic from motor cortical activity toward an autonomously updating brain machine interface," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2018, pp. 73–76.

[9] A. H. Ribeiro *et al.*, "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nature Communications*, vol. 11, no. 1, p. 1760, 2020.

[10] C. S. Cardoso *et al.*, "Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile," *BMJ open*, vol. 6, no. 5, p. e011181, 2016.

[11] P. Wagner *et al.*, "PTB-XL, a large publicly available electrocardiography dataset," *Scientific Data*, vol. 7, no. 1, pp. 1–15, 2020.

[12] J. An, R. E. Gregg, and S. Borhani, "Effective data augmentation, filters, and automation techniques for automatic 12-lead ECG classification using deep residual neural networks," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2022, pp. 1283–1287.

[13] J. An, B. Bailey, and R. E. Gregg, "Paradigm Shift from Feature-Based Machine Learning to End-to-End Deep Residual Neural Networks for Pediatric Age Classification from 12-Lead ECG," in *2023 Computing in Cardiology (CinC)*, IEEE, 2023, pp. 1–4.

[14] J. An, R. E. Gregg, B. Bailey, Y.-H. Zhang, and D. J. Dzikowicz, "Left Bundle Branch Block Detection in 12-Lead ECG Using End-to-End Deep Learning with Explainability," in *2024 Computing in Cardiology (CinC)*, IEEE, 2024, pp. 1–4.

[15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.

[16] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

Address for correspondence:

Junmo An, PhD
222 Jacobs St,
Cambridge, MA 02141, USA
junmo.an@philips.com