Two-Stage Domain Adversarial Learning to Identify Chagas Disease from ECG and Patient Demographic Data

Xiaoyu Wang[1], Aron Syversen[1], Zixuan Ding[2], James Battye[1], Sharon Yuen Shan Ho[2], David C Wong[1]

[1]University of Leeds, Leeds, United Kingdom
[2]University of Cambridge, Cambridge, United Kingdom

Abstract

Large-scale automated ECG screening can combat the widespread underdiagnosis of Chagas disease due to limited serological test coverage. To this end, our team, CinCo Amigos, developed a computational approach to detect Chagas disease from electrocardiograms (ECGs) - a two-stage domain-adversarial training process to address key issues of significant label noise, extreme class imbalance, and substantial domain shift.

Our framework first pre-trains a custom neural network on a large, noisy dataset. Early Learning Regularization (ELR) and Domain-Adversarial Neural Network (DANN) were integrated to mitigate label errors and encourage domain-invariant features respectively. To handle class imbalance, we employed a objective combining Focal Loss (LMFLoss) and Label-Distribution-Aware Margin (LDAM) Loss. In the second stage, the model was fine-tuned on high-quality datasets using feature distillation.

Our model achieved an official Challenge score of 0.250 (ranked 7 of 40 teams), and was the best performing on one of the three test sets. This work suggests that our integrated approach provides a robust framework for automated ECG-based diagnosis and can improve generalisation in challenging real-world scenarios.

1. Introduction

This paper presents our entry to the 2025 PhysioNet Challenge [1–3] on automated, open-source ECG algorithms for Chagas disease detection. Training data were made available from several distinct collections [4–6].

The primary difficulties are: significant label noise, as the largest dataset has unreliable, self-reported labels (CODE-15) while smaller datasets provide reliable annotations; an extreme class imbalance with 2% positive class prevalence; and a significant domain shift, evidenced by a stark performance drop between internal testing versus public scoring metrics.

To address these, we developed an approach that combines a customised convolutional neural network with noise-robust learning, domain-adversarial techniques, and advanced class-imbalance handling.

2. Methods

Our approach begins with robust data preprocessing and augmentation, followed by training a novel model architecture for ECG analysis. The training strategy first involves pre-training on a large, noisy dataset to learn generalisable features, and subsequently fine-tuning on smaller, high-quality datasets for Chagas disease detection.

2.1. Data Preprocessing

All 12-lead ECG signals were resampled to 500 Hz, and filtered with a bandpass filter (1 Hz - 30 Hz) to remove baseline wander and high-frequency noise, and notch filters at 50 Hz and 60 Hz to eliminate powerline interference. Finally, each recording underwent z-score normalisation to standardise the signal distribution.

A diverse set of augmentation methods were applied during training, including adding Gaussian noise, random scaling, temporal shifting, dropping and cutting out of signal segments, lead mixing [7], and time warping.

2.2. Model Architecture

Our model adopts a modular architecture composed of a unified encoder and two parallel classifier heads, as depicted in Fig. 1. As we later show, the encoder can be switched for any common backbone encoder.

The **encoder** generates a domain-invariant feature representation from multi-modal inputs. It consists of two sub-modules: 1) an *ECGNeXt*, serving as the backbone, which is adapted from ECGFounder [8] with refinements from ConvNeXt [9] to capture temporal patterns; and 2) a *Meta Net*, a multi-layer perceptron that processes demographic covariates (age and sex). Features from both are

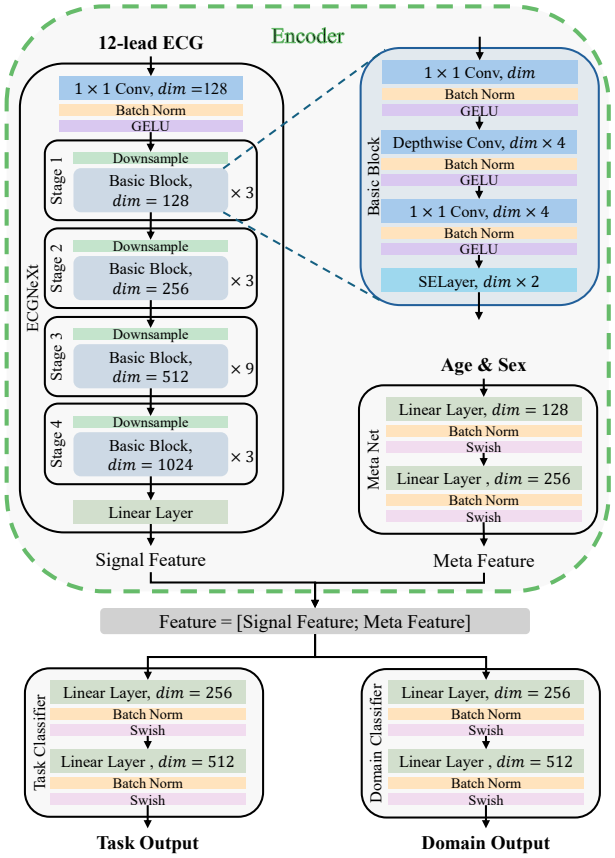

Figure 1. Architecture of the proposed network. The unified encoder, comprising the ECGNeXt and Meta Net, produces a shared feature representation. This representation is then fed into two separate heads: a task classifier for Chagas disease prediction and a domain classifier for adversarial training.

concatenated to form a unified representation.

This shared representation is then fed into two classifier heads. The *task classifier* performs the final binary prediction for Chagas disease. Concurrently, the *domain classifier*, integral to our adversarial training, learns to identify the data's source domain, compelling the encoder to produce more generalisable, domain-agnostic features.

2.3. Training Strategy

To address the key challenges, we devised the two-stage training paradigm illustrated in Fig. 2. The objective is a composite of several specialised losses, activated dynamically across the two stages.

Stage 1: Pre-Training with Noise and Domain Adaptation. The goal of this stage is to learn robust, domain-invariant features from the large-scale, noisy CODE15 dataset, by combining three techniques.

First, to handle class imbalance, we used LMFLoss, a weighted combination of *Focal Loss* (which focuses on hard-to-classify examples) and *Label-Distribution-Aware Margin (LDAM) Loss* (which enforces a larger margin for the minority class). The combined objective is defined as:

$$\begin{aligned} L_{\text{LMF}} = & -\alpha(1-p_y)^{\gamma}\log(p_y) \\ & -\beta\left[d_y \log(\sigma(s(z_y-\Delta_y))) \right. \\ & \left. +(1-d_y)\log(1-\sigma(s(z_y-\Delta_y)))\right] \end{aligned} \tag{1}$$

where p_y is the predicted probability, $d_y \in \{0,1\}$ is the true binary label, z_y is the original logit, $\sigma(\cdot)$ is the sigmoid function, Δ_y is the class-dependent margin, and s is the scale parameter from LDAM which adjusts the logits to control the steepness of the loss landscape.

Second, to counteract label noise, we integrated Early Learning Regularization (ELR) [10]. It adds a regularisation term to the standard Binary Cross-Entropy (BCE) loss, preventing the model from memorising incorrect labels by regressing towards its historical consensus. The regulariser is an MSE loss between the current prediction and an exponential moving average (EMA) of past predictions.

Third, for domain generalisation, we employed a Domain-Adversarial Neural Network (DANN) which we have previously used for ECGs [11,12]. To create a diverse set of domains, we incorporated several external datasets (e.g., CSPC, PTB from PhysioNet 2021 [13]). Each of these datasets, along with the primary CODE15 data, was treated as a distinct domain. Crucially, the diagnostic labels from these external datasets were discarded.

The DANN framework involves two competing objectives. The *domain classifier* is trained to predict the source domain out of K possible domains. Its objective is to minimise the standard multi-class cross-entropy loss, L_{D}:

$$L_{\text{D}} = \sum_{k=1}^{K} -[d_k \log(\hat{d}_k) + (1-d_k)\log(1-\hat{d}_k)] \tag{2}$$

where d is the one-hot true domain label and $\hat{d}$ is the predicted domain probability distribution.

Concurrently, the *encoder* is trained to fool the classifier by minimising a confusion loss, L_{C}. This loss encourages a uniform prediction from the domain classifier, which corresponds to high prediction entropy. We therefore define it as the negative entropy of the classifier's output:

$$L_{\text{C}} = \sum_{k=1}^{K} \hat{d}_k \log(\hat{d}_k) \tag{3}$$

This adversarial process forces the encoder to learn domain-agnostic representations.

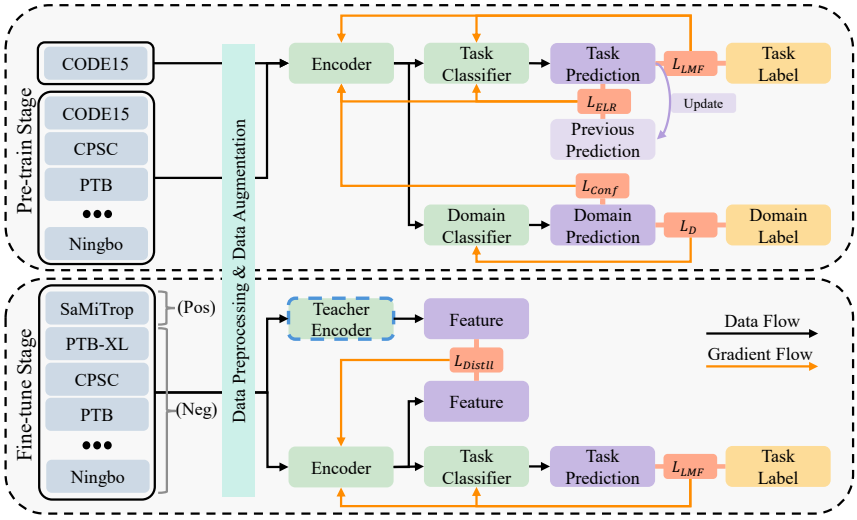

Figure 2. The proposed two-stage training strategy. Stage 1 (Pre-training) focuses on learning domain-invariant features from large-scale noisy data using DANN and ELR. Stage 2 (Fine-tuning) adapts the model to high-quality data using feature distillation to retain generalisability.

The optimization objectives for the encoder (θ_e), task classifier (θ_t), and domain classifier (θ_d) are defined as follows:

$$\theta_e^* = \arg\min_{\theta_e}(L_{\text{LMF}} + \lambda_{\text{ELR}} L_{\text{ELR}} + \lambda_{\text{DANN}} L_{\text{C}}) \quad (4)$$

$$\theta_t^* = \arg\min_{\theta_t}(L_{\text{LMF}} + \lambda_{\text{ELR}} L_{\text{ELR}}) \quad (5)$$

$$\theta_d^* = \arg\min_{\theta_d} L_{\text{D}} \quad (6)$$

Stage 2: Fine-Tuning with Preservation of Domain Generalisation. In this stage, the model is adapted using smaller, high-quality datasets. The key challenge is to specialise the model for the target task without forgetting the robust, domain-invariant features learned during pre-training. We consider two approaches to this.

The first employs feature distillation [14]. The encoder pre-trained with DANN, which excels at producing domain-agnostic representations, acts as a frozen "teacher". The fine-tuning encoder ("student") is then guided by minimising an MSE loss (L_{Distill}) between its feature outputs ($\phi_{\text{student}}(x)$) and those of the teacher ($\phi_{\text{teacher}}(x)$). This process ensures the model retains its ability to generalise across different data domains. The primary task was still optimised using LMFLoss. The optimisation objectives for this stage were:

$$\theta_e^* = \arg\min_{\theta_e}(L_{\text{LMF}} + \lambda_{\text{Distill}} L_{\text{Distill}}) \quad (7)$$

$$\theta_t^* = \arg\min_{\theta_t} L_{\text{LMF}} \quad (8)$$

Here, the λ terms are hyperparameters balancing the different loss components.

The second, simpler, approach freezes the encoder and only trains the classifier head using the fine-tuning data.

3. Results

We trained the models using the AdamW optimiser. The learning rate followed a cosine annealing schedule with a warm-up period. We used early stopping on the validation set with a 5-epoch patience. We experimented with different model architectures, such as changing the encoder to SEResNet [11]. We also experimented with different fine-tuning regimes.

3.1. Challenge Results

Our final model used an SEResnet18 encoder with frozen weights during fine tuning. It was evaluated on the hidden test sets for the official phase of the PhysioNet Challenge 2025. As detailed in Table 1, our approach achieved a mean challenge score of 0.250, ranking 7th out of 40 competing teams.

Table 1. Official Challenge Results

Metric	Score	Rank
Overall Performance		
Mean Challenge Score	0.250	7/40
Performance on Hidden Test Sets		
Challenge Score (REDS-II test set)	0.339	7/40
Challenge Score (SaMi-Trop 3 test set)	0.247	21/40
Challenge Score (ELSA-Brasil test set)	0.164	1/40

We also evaluated several model configurations to assess

the impact of different components, shown in Table 2.

Table 2. Challenge Score on validation and test datasets.

Encoder	λ_{DANN}	λ_{Distill}	Local	Valid	Test
SEResNet18	0.8	0	0.793	0.231	–
SEResNet18	0.8	0.01	0.760	0.234	–
SEResNet18	**0.8**	**N/A**	**0.627**	**0.338**	**0.250**
ECGNeXt	0.3	0	0.827	0.230	–
ECGNeXt	0.8	0	0.807	–	–
ECGNeXt	0.8	0.05	0.793	0.294	–
ECGNeXt	0.8	N/A	0.713	0.254	–
ECGNeXt	1.0	0.05	0.620	–	–

N/A: The encoder was frozen during fine-tuning.

4. Discussion

Our approach used a two-stage training strategy to tackle the key challenges of this competition: noisy labels, data imbalance, and shifts between different data sources.

Our method's effectiveness is shown in the final challenge results (Table 1). Notably, our model ranked 1st of 40 teams on the ELSA-Brasil test set. While all test data was unseen, the ELSA-Brasil data came from a source completely new to all participants.

Our internal experiments (Table 2) also gave us two major insights. First, model simplicity proved more robust. While complex models like ECGNeXt excelled locally, they tended to overfit domain-specific features, leading to poor validation performance. The simpler SEResNet18 generalised better and achieved higher validation scores.

Second, adversarial training and distillation were crucial. A stronger DANN signal improved scores by forcing the model to learn domain-invariant features. Feature distillation then effectively preserved this knowledge during fine-tuning, outperforming other adaptation strategies like encoder freezing.

While our preliminary study highlights the benefits of DANN and feature distillation, a more exhaustive set of studies, including more hyperparameter search, is required in future work to better quantify the impact of each aspect of the training strategy. In conclusion, our results show that combining domain-adversarial training to *learn* generalisable features and feature distillation to *preserve* them offers a robust framework for mitigating domain shift.

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. Computing in Cardiology 2025;52:1–4.

[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] Kim KG, Lee BT. Graph structure based data augmentation method. Biomedical Engineering Letters 2025;15(2):283–289.

[8] Li J, Aguirre A, Moura J, Liu C, Zhong L, Sun C, et al. An electrocardiogram foundation model built on over 10 million recordings with external evaluation across multiple domains. arXiv preprint arXiv241004133 2024;.

[9] Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A convnet for the 2020s. In Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. 2022; 11976–11986.

[10] Liu S, Niles-Weed J, Razavian N, Fernandez-Granda C. Early-learning regularization prevents memorization of noisy labels. Advances in Neural Information Processing Systems 2020;33:20331–20342.

[11] Shang Z, Zhao Z, Fang H, Relton S, Murphy D, Hancox Z, et al. Deep discriminative domain generalization with adversarial feature learning for classifying ecg signals. In 2021 Computing in Cardiology (CinC), volume 48. IEEE, 2021; 1–4.

[12] Shang Z, Zhao Z, Fang H, Relton S, Murphy D, Hancox Z, et al. Deep discriminative domain generalization with adversarial feature learning for classifying ecg signals. In 2021 Computing in Cardiology (CinC), volume 48. IEEE, 2021; 1–4.

[13] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, et al. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021;48:1–4.

[14] Heo B, Kim J, Yun S, Park H, Kwak N, Choi JY. A comprehensive overhaul of feature distillation. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019; 1921–1930.

Address for correspondence:

Xiaoyu Wang, wmqn2362@leeds.ac.uk

Level 10, Worsley Building, University of Leeds, Leeds, United Kingdom