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 ECG preliminary screening can efficiently identify high-risk individuals for targeted confirmatory testing, combating the widespread under-diagnosis of Chagas disease due to limited serological test coverage. It is this potential that provides the fundamental motivation for developing automated ECG screening. We developed a computational approach to detect Chagas disease from electrocardiograms (ECGs). Our team, CinCo Amigos, developed a two-stage domain-adversarial training process to address the key issues of significant label noise, extreme class imbalance, and substantial domain shift.

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

Our model achieved a Challenge score of 0.338 on the validation set. Our official Challenge score was – (ranked – out of – teams) on the hidden test set. 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 details our entry for the 2025 George B. Moody PhysioNet Challenge, which focused on the development of automated, open-source algorithms for detecting Chagas disease from electrocardiograms (ECG) [1, 2]. Training data for this work were made available through the challenge from several distinct collections [3–5].

The primary difficulties of this task are: significant label noise, as the largest dataset has unreliable, self-reported labels (CODE15) while smaller datasets provide reliable annotations; an extreme class imbalance with positive cases accounting for only 2% of the data; 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 designed 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 to specialise the model 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.

To improve generalisation and robustness to dataset variability, we applied a diverse set of augmentation methods during training. These included adding Gaussian noise, random scaling, random temporal shifting, random dropping and cutting out of signal segments, lead mixing [6], and time warping.

2.2. Model Architecture

Our model adopts a modular architecture composed of a unified encoder and two parallel classifier heads, as de-

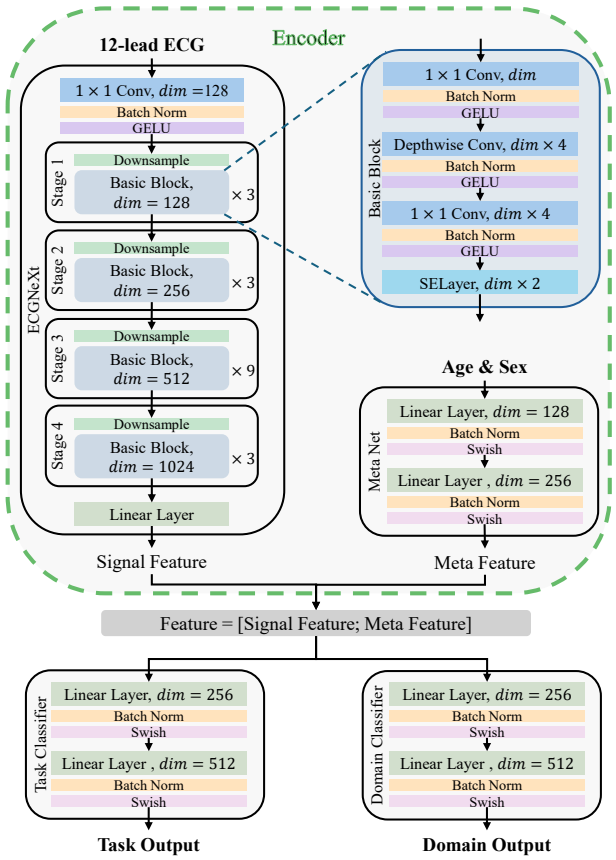

Figure 1. Architecture of the proposed network. The unified encoder, comprising the ECGNeXt and Meta Net, produces a shared feature representation. This representation is then fed into two separate heads: a task classifier for Chagas disease prediction and a domain classifier for adversarial training.

picted 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 [7] with refinements from ConvNeXt [8] to capture temporal patterns; and 2) a *Meta Net*, a multi-layer perceptron that processes demographic covariates (age and sex). Features from both are 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) [9]. 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 [10,11]. To create a diverse set of domains, we incorporated several external datasets (e.g., CSPC, PTB from PhysioNet 2021 [12]). 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

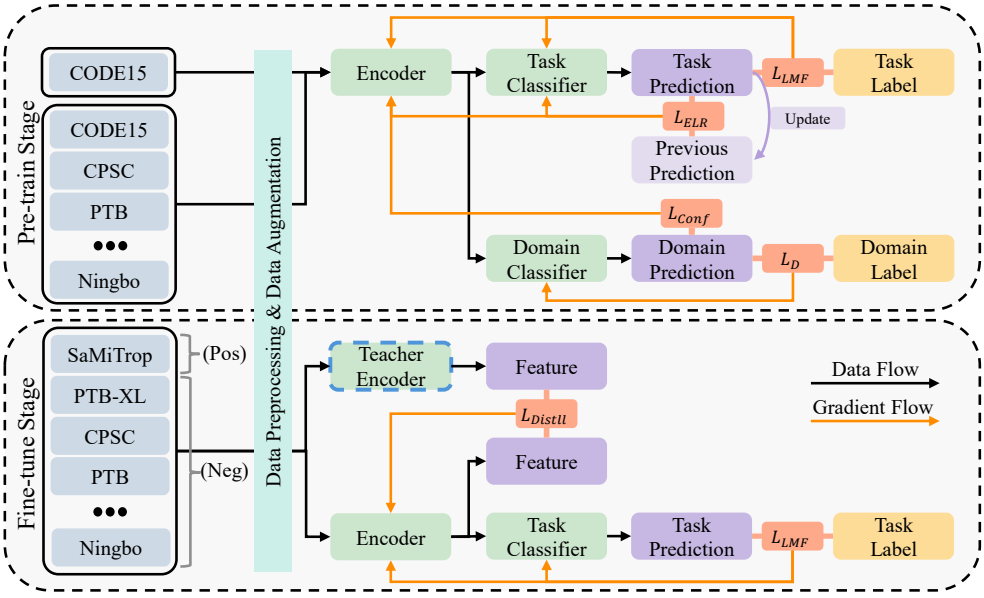

Figure 2. The proposed two-stage training strategy. Stage 1 (Pre-training) focuses on learning domain-invariant features from large-scale noisy data using DANN and ELR. Stage 2 (Fine-tuning) adapts the model to high-quality data using feature distillation to retain generalisability.

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.

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

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

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

$$\theta_d^* = \arg\min_{\theta_d} L_{\text{D}} \tag{6}$$

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 [13]. 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 ($\phi_{\text{teacher}}(x)$).

$$L_{\text{Distill}} = \text{MSE}(\phi_{\text{student}}(x), \phi_{\text{teacher}}(x)) \tag{7}$$

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:

$$\theta_e^* = \arg\min_{\theta_e}(L_{\text{LMF}} + \lambda_{\text{Distill}} L_{\text{Distill}}) \tag{8}$$

$$\theta_t^* = \arg\min_{\theta_t} L_{\text{LMF}} \tag{9}$$

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 with mini-batch stochastic gradient descent. The learning rate was managed with a warm-up period followed by a cosine annealing schedule. To prevent overfitting, we employed early stopping based on the validation set performance, with a patience of 5 epochs for pretraining and fine-tuning. We experimented with different settings, such as changing the encoder to SEResNet [10]. We also tested changing the settings of the fine-tuning stage, such as freezing the encoder.

3.1. Challenge Results

We evaluated several model configurations to assess the impact of different components, shown in Table 1.

Table 1. 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**	**–**
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 to the challenge integrated a two-stage training strategy with ELR for noise robustness, LMFLoss for imbalance, and domain-adversarial training with feature distillation for domain generalisation.

The results in Table 1 highlight two major observations. First, models that performed strongly on the local dataset (0.7–0.8) exhibited a drop on the Challenge validation set (0.23–0.34). Moreover, this effect was amplified by model complexity: while ECGNeXt achieved high local scores, it failed to generalise across domains, whereas the simpler SEResNet18 achieved the best validation score (0.338). These findings highlight that higher capacity models are prone to overfitting domain-specific features, and that simplicity may sometimes offer better robustness.

Second, adversarial training and distillation played a crucial role in improving robustness. A stronger adversarial signal ($\lambda_{\text{DANN}} = 0.8$) achieved markedly higher scores than weaker settings, confirming the role of DANN in learning domain-invariant features. Feature distillation further regularised fine-tuning, outperforming both encoder freezing and unconstrained adaptation, by preserving transferable knowledge from pre-training while still adapting to high-quality target data.

While our preliminary ablation study highlights the benefits of DANN and feature distillation, a more exhaustive set of studies, including more detailed hyperparameter search, is required in future work to better quantify the impact of each aspect of the training strategy.

In conclusion, our results strongly suggest that the combination of 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] 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.

[4] 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.

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

[7] 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;.

[8] 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.

[9] 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.

[10] 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.

[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] 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.

[13] 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
Level 10, Worsley Building, University of Leeds, Leeds, United Kingdom
wmqn2362@leeds.ac.uk