Knowledge Distillation from General Multi-Objective ECG Classification Model for Chagas Disease Detection in 12-Lead ECGs

Petr Nejedly[1], Martina Dankova[1], Radovan Smisek[1], Ivo Viscor[1], Pavel Jurak[1], Filip Plesinger[1]

[1] Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czech Republic

Abstract

We present our solution (team ISIBrno-AIMT) for the 2025 George B. Moody PhysioNet Challenge, which utilizes a teacher-student model architecture. A key component of our approach is a generalized multi-objective teacher model based on the U-Net architecture, which was pre-trained on publicly available 12-lead ECG databases (containing over 1 million ECG recordings) for both the segmentation task and the classification task of 28 cardiac pathologies. The student models were specifically trained to distill knowledge from the teacher's classification outputs, as well as to predict Chagas disease based on public challenge datasets containing Chagas disease labels. To improve the reliability of our predictions, we employed an ensemble of five student models, averaging their outputs during the inference stage. Our model achieved 4th place in the challenge, with an overall challenge score of 0.271 .

1. Introduction

Chagas disease (CD), caused by the protozoan Trypanosoma cruzi, is a tropical disease, affecting an estimated 6 to 7 million people worldwide, primarily in Latin America. After the acute phase, about 20–30% of infected individuals develop chronic Chagas cardiomyopathy, the most severe and life-threatening form of the disease [1–3]. Chronic Chagas cardiomyopathy involves progressive myocardial damage leading to conduction abnormalities and arrhythmias, which manifest on the ECG as bundle branch blocks, fascicular blocks, or ventricular ectopy [4]. Traditionally, Chagas disease is diagnosed by serological testing for antibodies against the parasite using techniques such as enzyme-linked immunosorbent assay (ELISA) or immunofluorescent antibody test (IFAT). However, these methods can be time-consuming, require specialized equipment, and are often expensive. Therefore, there is a need for prescreening methods allowing the rapid and inexpensive identification of individuals who are likely to be infected, enabling resources to be focused on those who need more definitive diagnostic testing. This is particularly important in resource-limited settings where traditional serological tests may not be readily available. Electrocardiography (ECG) is used as a preliminary screening tool to identify individuals who may have cardiac involvement due to chronic Chagas disease [4]. Abnormal ECG results help identify patients who should undergo more specific and costly serological testing to confirm infection and determine appropriate therapy.

2. Methods

For this challenge, we proposed a deep learning pipeline based on a teacher–student model architecture. The teacher model was pretrained to perform ECG segmentation and classify 28 cardiac pathologies based on the 2021 Physionet Challenge datasets [5] and other publicly available datasets (Table 1). The architecture was adapted from our winning solution in the 2021 Physionet Challenge and rebuilt into U-net architecture, allowing for ECG segmentation [6, 7]. Student models were designed based on our 2021 ResNet architecture, accompanied by multi-head outputs, allowing students to learn from the teacher's outputs in a multi-objective manner and estimate the probability of Chagas disease. To improve prediction stability and overall performance, we trained an ensemble of five student models, whose outputs were averaged during inference.

In the preprocessing stage, we followed our Challenge 2021 pipeline [6]. First, resample the data to a fixed frequency of 500 Hz. We apply a bandpass filter (1–47 Hz) to suppress frequency components unrelated to cardiac activity, such as baseline wandering and powerline interference at 50 Hz or 60 Hz. This is especially relevant since datasets SAMITROP and CODE-15 originated from Brazil (60 Hz powerline), while PTB-XL originates from Germany (50 Hz powerline), where the incidence of CD is negligible.

Some teams in the challenge's unofficial phase achieved unusually high scores, nearing a perfect 1.0. This suggested possible data bias, as the results significantly outperformed typical state-of-the-art methods [8]. This finding highlighted the need to focus on building more robust models that aren't overly reliant on these dataset-specific biases.

To augment the training data and increase model robustness, we introduced several types of controlled noise during preprocessing. Gaussian noise was added with amplitude scaled relative to each channel's signal magnitude, simulating natural variability in ECG recordings. Artificial pauses—segments of zeroed signal across all channels—were used to mimic transient signal dropouts, such as those caused by electrode detachment or motion artifacts. In addition, we occasionally inserted impulse noise to simulate abrupt high-amplitude disturbances. The signal is then z-score normalized by subtracting the mean and dividing by the standard deviation per channel. Any NaN values are replaced with zeros to maintain numerical stability. To ensure consistent input size, we crop or zero-pad all recordings to a fixed length of 8,192 samples (approximately 16 seconds) to accommodate the standard duration of 10-second strips.

2.1. Multi-objective teacher model training

The multiobjective teacher model for generalized ECG classification and segmentation was developed using publicly available ECG datasets. The model is designed as U-net architecture with classification and regression multihead outputs from the bottleneck layer. The classification output consists of 28 classes from PhysionetChallenge 2021, and the regression output involves SNR estimation, achieved by adding variable noise to the input data as an augmentation technique. The architectural design is depicted in Figure 1.

The teacher loss function is defined as:

$$L_{total} = L_{cls} + \lambda_{seg} L_{seg} + \lambda_{snr} L_{snr} \quad (1)$$

, where L_{cls} is a binary cross-entropy loss for a classification task, L_{seg} is a cross-entropy loss for ECG segmentation (P, PQ, QRS, ST, T, TP) intervals, and L_{snr} is a mean-square error regression for SNR estimation. Lambda is a task-specific coefficient to balance the contributions for individual tasks.

2.2. Student model training

The five models forming the ensemble were trained using 5-fold cross-validation with a shared pretrained teacher model and pretrained Chagas disease model backbone locally trained for 12 epochs. The server-side training was continued for 1 epoch, where the dataset was split into five folds; in each iteration, four folds were used for training and one for validation. This procedure yielded five independently trained student models, each trained on a different subset. During inference, all five models were applied to the test set, and their outputs were averaged to produce the final predicted probability of Chagas disease for each patient.

The student loss is defined [9] as:

$$L_{student} = L_{chagas}(t, p_{chagas}) + L_{cls}(p_{teacher}, p_{student}) \quad (2)$$

, where L_{chagas} is cross-entropy loss for Chagas disease classification task, and $L_{cls}(p_{teacher}, p_{student})$ is binary-cross-entropy loss for 28 pathologies classes, where $p_{teacher}$ and $p_{student}$ represent probabilities per given class.

Table 1. Publicly available datasets used for generalized teacher model training

Dataset	**Records**
CPSC 2018 [10]	6 877
MIMIC-4-ECG [11]	800 000
Chapman and Ningbo [12]	45 152
Georgia ECG database [13]	10 300
HEFEI	40 000
CODE-15 [14]	345 779
SAMITROP [15]	1 631
PTB-XL [16]	21 837
Shandong Hospital [17]	25 770
Segmentation task	
LUDB [18]	200
PTB-XL (v1.0.1) Soft-Seg [19]	-

3. Results

The Challenge score is calculated by first identifying the subset of records with the highest predicted probabilities for Chagas disease, where the size of this subset is set by the coefficient $\alpha = 0.05$, meaning the top 5% of records with the highest predicted probability for Chagas are selected for evaluation. The true positive rate is then calculated exclusively on this α-subset.

This specialized evaluation method addresses the challenges associated with the limited availability of serological testing for Chagas disease. By focusing on a small, high-confidence subset of predictions, the Challenge score provides a more relevant measure of the model's performance in scenarios where diagnostic resources are scarce. This approach mirrors real-world applications where only a fraction of the population can be tested, prioritizing the model's ability to accurately identify the most probable cases.

4. Discussion

Our strategy was initially inspired by the clinical diagnostic process, specifically focusing on the detection of common ECG abnormalities typically observed in cardio-

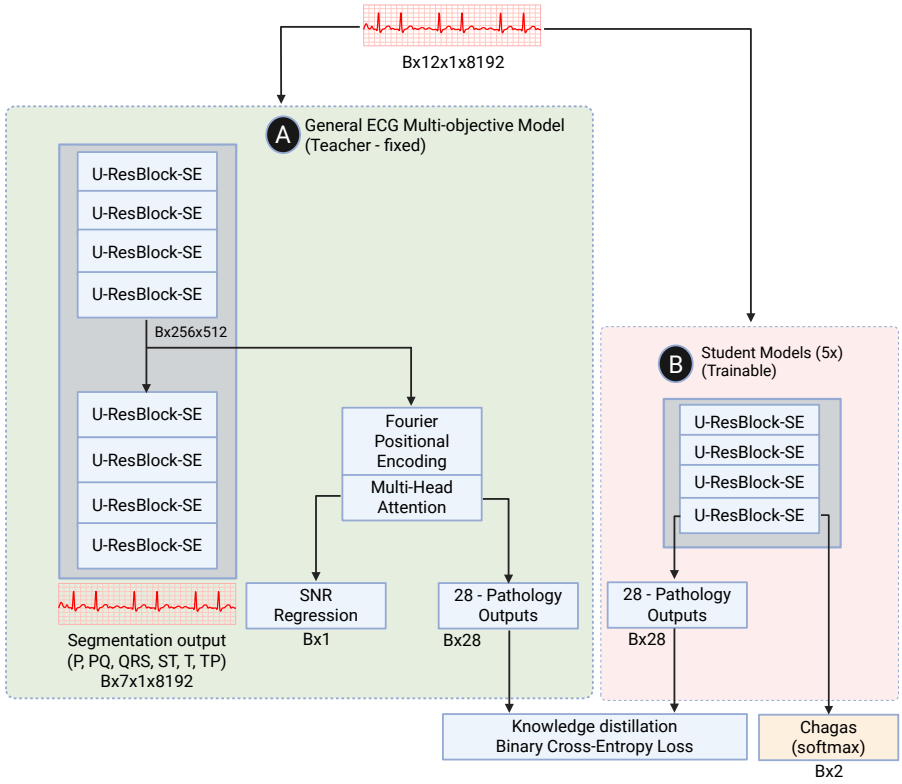

Figure 1. This figure illustrates the training pipeline for the Chagas disease classification model. The teacher model (A), a generalized multi-objective ECG model, generates pseudo-labels for the student models. The student models (B) are then trained using these pseudo-labels, enabling them to specialize in classifying specific ECG pathologies, as well as Chagas disease.

Table 2. Performance metrics for the challenge.

Dataset	Score	AUROC	AUPRC
Local Valid	0.558	0.847	0.499
Valid (REDS-II)	0.458	0.774	0.302
Test (REDS-II)	0.354	0.752	0.165
Test (SAMI-TROP-3)	0.342	0.763	0.167
Test (ELSA)	0.118	0.579	0.063
Test (overall)	0.271	-	-

vascular disorders, such as conduction abnormalities, arrhythmias, bundle branch blocks, fascicular blocks, and ventricular ectopy. This approach reflects the reasoning of a cardiologist, who begins by evaluating prevalent ECG pathologies before forming a diagnostic hypothesis. To this end, we utilized publicly available datasets to train generalized multi-objective ECG models. The classification outputs for 28 pathologies from these models were then input into logistic regression.

However, during the validation phase, we discovered that training a student model to distill knowledge from a generalized teacher model resulted in significantly improved challenge scores compared to the naive approach of simulating the clinician's decision-making process by combining individual clinical findings. Consequently, we adopted the teacher-student framework as our preferred solution.

Building on our 2021 challenge strategy, we applied data augmentations to simulate ECG artifacts (channel dropouts, impulse noise, and saturation). To reduce training variance, we pretrained models locally for 12 epochs, then performed one additional epoch of cloud training per ensemble student model.

5. Conclusion

In this work, we presented method based on distillation learning teacher–student architecture, combined with extensive data augmentation and an ensemble of models trained using 5-fold cross-validation. The proposed solution achieved 4th place in the 2025 Challenge, with an overall score of 0.271 demonstrating the effectiveness of the designed pipeline in detecting Chagas disease from

ECG data.

Acknowledgments

This work was supported by projects by the Czech Academy of Sciences RVO: 68081731 and the Czech Technological Agency FW06010766.

References

[1] Rassi Jr A, Rassi A, Marin-Neto JA. Chagas disease. The Lancet 4 2010;375(9723):1388–1402. [Online; accessed 2025-08-25].

[2] Benck L, Kransdorf E, Patel J. Diagnosis and management of chagas cardiomyopathy in the United States. Current Cardiology Reports oct 11 2018;20(12). [Online; accessed 2025-08-25].

[3] Marin-Neto JA, Cunha-Neto E, Maciel BC, Simões MV. Pathogenesis of chronic chagas heart disease. Circulation mar 6 2007;115(9):1109–1123. [Online; accessed 2025-08-25].

[4] Rojas LZ, Glisic M, Pletsch-Borba L, Echeverría LE, Bramer WM, Bano A, Stringa N, Zaciragic A, Kraja B, Asllanaj E, Chowdhury R, Morillo CA, Rueda-Ochoa OL, Franco OH, Muka T. Electrocardiographic abnormalities in Chagas disease in the general population: A systematic review and meta-analysis. PLOS Neglected Tropical Diseases jun 13 2018;12(6):e0006567. [Online; accessed 2025-08-25].

[5] Reyna MA, Sadr N, Alday EAP, Gu A, Shah AJ, Robichaux C, Rad AB, Elola A, Seyedi S, Ansari S, Ghanbari H, Li Q, Sharma A, Clifford GD. Will two do? Varying dimensions in electrocardiography: The physionet/computing in cardiology challenge 2021. In 2021 Computing in Cardiology (CinC). IEEE, sep 13 2021; 1–4. [Online; accessed 2025-08-25].

[6] Nejedly P, Ivora A, Smisek R, Viscor I, Koscova Z, Jurak P, Plesinger F. Classification of ECG using ensemble of residual cnns with attention mechanism. In 2021 Computing in Cardiology (CinC). IEEE, sep 13 2021; 1–4. [Online; accessed 2025-08-25].

[7] Nejedly P, Ivora A, Viscor I, Koscova Z, Smisek R, Jurak P, Plesinger F. Classification of ECG using ensemble of residual CNNs with or without attention mechanism. Physiological Measurement apr 28 2022;43(4):044001. [Online; accessed 2025-08-25].

[8] Jidling C, Gedon D, Schön TB, Oliveira CDL, Cardoso CS, Ferreira AM, Giatti L, Barreto SM, Sabino EC, Ribeiro ALP, Ribeiro AH. Screening for Chagas disease from the electrocardiogram using a deep neural network. PLoS Neglected Tropical Diseases jul 3 2023;17(7):e0011118.

[9] 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;52:1–4.

[10] Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, Liu Y, Ma C, Wei S, He Z, Li J, Yin Kwee EN. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. Journal of Medical Imaging and Health Informatics sep 1 2018;8(7):1368–1373. [Online; accessed 2025-08-25].

[11] Gow B, Pollard T, Nathanson LA, Johnson A, Moody B, Fernandes C, Greenbaum N, Waks JW, Eslami P, Carbonati T, Chaudhari A, Herbst E, Moukheiber D, Berkowitz S, Mark R, Horng S. Mimic-IV-ECG: Diagnostic electrocardiogram matched subset. https://physionet.org/content/mimic-iv-ecg/, sep 15 2023.

[12] Zheng J, Zhang J, Danioko S, Yao H, Guo H, Rakovski C. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Scientific Data feb 12 2020;7(1). [Online; accessed 2025-08-25].

[13] Alday EAP, Gu A, Shah A, Liu C, Sharma A, Seyedi S, Rad AB, Reyna M, Clifford G. Classification of 12-lead ecgs: The PhysioNet/Computing in Cardiology Challenge 2020. https://physionet.org/content/challenge-2020/1.0.2/, jul 29 2022.

[14] Ribeiro AH, Paixao GM, Lima EM, Ribeiro H, Filho P, Gomes PR, Oliveira DM, Jr M, Schon TB, Ribeiro ALP. Code-15 https://zenodo.org/records/4916206, jun 9 2021. [Online; accessed 2025-08-25].

[15] Ribeiro ALP, Ribeiro AH, Paixao GM, Lima EM, Ribeiro H, Filho P, Gomes PR, Oliveira DM, Jr M, Schon TB, Sabino EC. Sami-Trop: 12-lead ECG traces with age and mortality annotations. https://zenodo.org/records/4905618, jun 7 2021.

[16] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. Ptb-XL, a large publicly available electrocardiography dataset. Scientific Data may 25 2020;7(1). [Online; accessed 2025-08-25].

[17] Liu H, Chen D, Chen D, Zhang X, Li H, Bian L, Shu M, Wang Y. A large-scale multi-label 12-lead electrocardiogram database with standardized diagnostic statements. Scientific Data jun 7 2022;9(1). [Online; accessed 2025-08-25].

[18] Kalyakulina AI, Yusipov II, Moskalenko VA, Nikolskiy AV, Kosonogov KA, Osipov GV, Zolotykh NY, Ivanchenko MV. Ludb: A new open-access validation tool for electrocardiogram delineation algorithms. IEEE Access 2020;8:186181–186190. [Online; accessed 2025-08-25].

[19] Wagner P, Mehari T, Haverkamp W, Strodthoff N. Ptb-XL (v1.0.1) soft segmentations (delineation). https://zenodo.org/record/7610235, feb 6 2023. [Online; accessed 2025-08-25].

Address for correspondence:

Petr Nejedly, Ph.D.
Institute of Scientific Instruments of the CAS, v. v. i.
Kralovopolska 147
612 00 Brno
Czech Republic
nejedly@isibrno.cz