A ResNet with CBAM Attention Module Ensemble to Detect Chagas Disease in the ECG: Age, Sex and Database Biases

Jorge Sanchez[1], Sandra Perez-Herrero[1], Jose Gallego-Navarro[1], Guadalupe Garcia-Isla[1]

[1] Centro de Investigacion e Innovacion en Bioingenieria (Ci2B), Universidad Politecnica de Valencia, Valencia, Spain

Abstract

We present ElectroDoChagas's team contribution to the George B. Moody PhysioNet Challenge 2025 challenge to detect Chagas disease on 12-lead ECG signals. We focused on mitigating database, age, and sex-dependence biases by creating an ensemble model trained on data subsets where these features were stratified. This stratification was also performed for validation and local holdout-test splitting. The model was a deep Residual Network (ResNet) with 8 residual blocks enhanced with Convolutional Block Attention Modules (CBAM). While CBAM applied lead and time attention mechanisms, residual connections allowed a deeper model and more complex feature extraction. We explored the performance of a single model trained on the full database (model 1), an ensemble model composed of 3 ResNets trained on 3 different subsets, respectively (model 2), and a combination of 2 ensemble models trained each on one age group, setting 60-years-old as threshold for group sub-division (model 3). Model 1 obtained a challenge score (CS) of 0.351 and 0.217 on the hidden validation and test sets, respectively.

1. Introduction

In this paper we present our contribution to the 2025 George B. Moody PhysioNet Challenge, which consisted on developing algorithms to detect Chagas disease from 12-lead electrocardiograms (ECG) [1–3]. We proposed an ensemble of Residual Network (ResNet) models enhanced with Convolutional Block Attention Modules (CBAM) attention mechanisms to detect Chagas disease from 12-lead ECG recordings. We specifically evaluate the impact of age, sex, and database heterogeneity on model performance, aiming to develop a diagnostic tool that is not only accurate, but also generalizable across diverse populations.

2. Methodology

Three databases were provided by the challenge organizers [4–8]. We designed a pipeline that combined data preparation, a bias-aware splitting strategy, and a deep learning (DL) architecture tailored to imbalanced and heterogeneous datasets trained under a composite loss function designed to improve sensitivity in positive cases. We presented 3 different models with a shared core DL architecture trained on different data shares.

2.1. Data Preparation

2.1.1. Preprocessing

Signals were resampled to 400 Hz and filtered with a zero-phase band-pass filter between 0.1 Hz and 50 Hz to remove high-frequency noise and slow drifts while preserving clinically relevant frequency components. Baseline wander was corrected by spline-based detrending, which adaptively estimated and subtracted low-frequency fluctuations. Finally, signals were standardized and either zero-padded or truncated to 3000 samples to ensure uniform input length for the DL architecture.

2.1.2. Augmentation

To increase robustness against acquisition variability and prevent overfitting, data augmentation was performed. It included the addition of Gaussian noise, random lead inversion, and permutation, each applied with probabilities between 0.02 and 0.06. As the positive class (Chagas-labeled ECGs) was underrepresented (379,794 negative vs. 7,374 positive), oversampling was performed replicating by three-fold positive cases.

2.2. Model

2.2.1. Architecture

The proposed model was a deep ResNet architecture with CBAM attention mechanism as described in [9]. Residual connections allowed the training of a deeper network without suffering from vanishing gradient problems, thereby enabling the extraction of complex hierarchical

features. It was composed of 8 residual blocks, each containing an identity and a convolutional block as described in Figure 1. The 1D convolutional layers had a kernel size

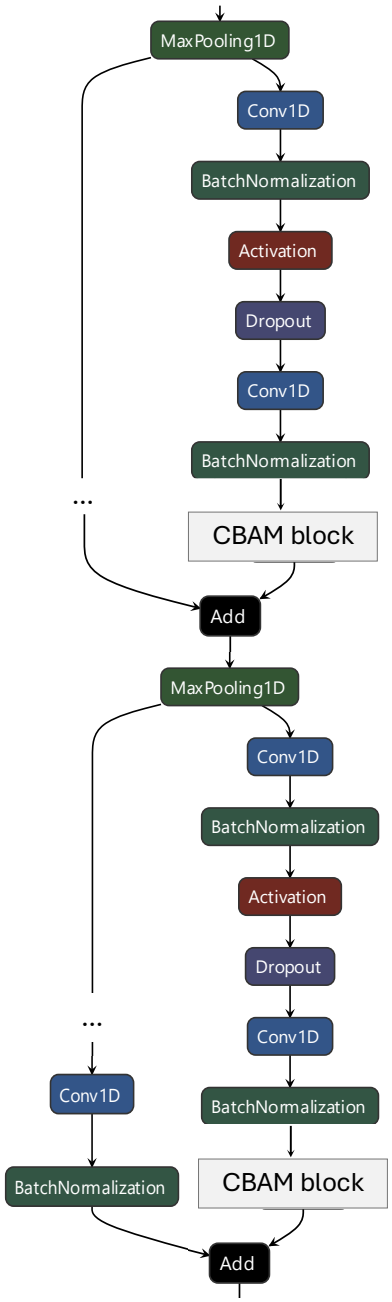


Figure 1. Residual block ResNet architecture, including a concatenation of and identity and convolutional block.

of 15 at outer and 7 at deeper layers with L2 regularization and 64 filters that doubled every 2 blocks, enhancing the model's ability to capture both low- and high-level temporal features from input signals. A dropout layer of 0.3 was added at each residual block. As described in Figure 2, CBAM modules were integrated into residual blocks. They sequentially apply channel and spatial attention mechanisms, which adaptively recalibrate feature maps by emphasizing informative leads and time segments while suppressing irrelevant information.

The final layers included global feature aggregation followed by a fully connected layer with sigmoid activation to output class probabilities. The architecture was designed to balance depth, robustness, and clinical utility, providing a reliable framework across diverse populations. A single instance of the described model will be referred as model 1.

2.2.2. Training

For local training and test we performed a data splitting for 25% validation, 20% local holdout-test set. To mitigate the age, sex and database biases, they were stratified during every data split together with the target label. To address class imbalance, model training employed a composite loss function that combined focal loss [10] with a differentiable surrogate of the true positive rate as described in equation 1.

$$\mathcal{L} = \mathcal{L}_{\text{focal}} - \lambda_{\text{tpr}} \, \text{TPR}_{\rho}^{\text{soft}} \tag{1}$$

Where $\mathcal{L}_{\text{focal}}$ addressed class imbalance focusing on hard examples and $\text{TPR}^{\text{soft}}\rho$ encouraged the model to rank positive cases within the top-ρ N predictions of each batch. The parameter λ_{tpr} controlled the strength of the TPR surrogate term. This encouraged the network not only to handle imbalanced distributions but also to directly optimize for recall of positive cases.

2.3. Ensembles

To further mitigate the underepresentation of chagas positive patients without dismissing data, we divided the chagas-negative class into 3 subsets, maintaining the database, sex and age stratification. Each subset was paired with the full number of chagas-positive signals. Each of the resulting subsets was used to train independently a single ResNet model. Independent models were combined into an ensemble through a major voting of their final prediction. The ensemble model as described in Figure 3 and it is referred to as model 2 throughout the text. Nevertheless, a more complex composition was explored (model 3). Model 3 combined two ensemble models (x2 model 2) emphasizing an age-distinction. A 60-year-old threshold was set to divide the data set. Then each of the age groups was used to train a different ensemble as described above. The output prediction for each record was given by the ensemble model corresponding to the age-group of the patient.

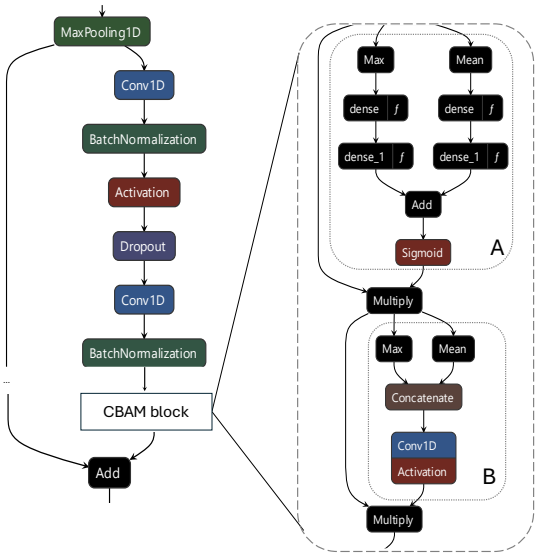


Figure 2. CBAM attention module implemented in ResNet block. (A) Channel attention module (B) Spatial attention module.

3. Results

Table 1 summarizes the performance obtained by the different models proposed: 1 a single ResNet, 2 an ensemble formed by 3 ResNets and 3 a combination of 2 ensembles each trained for an age group evaluated for 10 and 50 epochs of training. Unfortunately, due to computational time constrains, no score was provided by the challenge organizers for more than 10 training epochs, which led to an under-training of the models.

4. Discussion

Chagas disease detection on the ECG entails major challenges. Its effects are multiple, encompassing both structural and functional alterations that manifest as a wide spectrum of abnormalities. Furthermore, the provided databases entailed some challenges too as they were imbalanced both at the database and class level. All hard-labeled Chagas cases originated from a single database, so database-stratified splitting was implemented to prevent models from exploiting source-specific artifacts. At the class level, positive (Chagas) cases were substantially underrepresented; this was addressed by oversampling positives within each training subset to balance class distributions and by assigning class weights. In addition, multiple ResNet–CBAM models trained on different stratified subsets were combined into a major-voting ensemble, which reduced variance and improved generalization. Training employed a composite loss that integrated focal loss with a surrogate of the true positive rate, encouraging the model to prioritize recall in Chagas-positive cases. Nevertheless, computational constraints limited the number of full ensemble configurations that could be trained and submitted during the official challenge phase, restricting the exploration of more extensive architectures. Despite these limitations, the final submission achieved a close alignment between local validation scores and the official hidden test score, indicating that our bias-aware stratification and validation pipeline provided a good estimation of the CS and a reliable evaluation framework We proposed 3 models of increasing complexity, whose performance is gathered in Table 1.

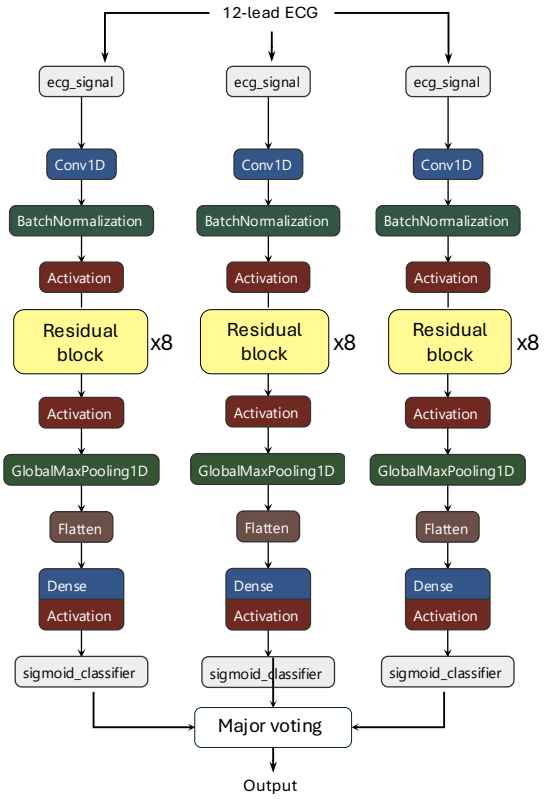


Figure 3. Ensemble model (model 2) integrating 3 ResNet networks through a major voting strategy.

The model-3 age split assumed age-related fibrosis and other pathologies increase comorbidities, complicating Chagas diagnosis. We used 60 years to secure enough hard labels, though an ideal split would separate young, middle-age, and seniors. Comparing the three models is difficult: computational limits prevented training for enough epochs. Locally, model 2 gave the highest CS and model 3 the highest F1. Model 3 didn't improve CS, likely because the 60-year threshold merged young and middle-age. With

Model	Epoch	Local		Hidden			
		Holdout Test		Validation		Test	
		F1-score	Challenge Score	Challenge Score	Ranking	Challenge Score	Ranking
1	10	0.243	0.361	0.351	46/371	0.217	18/40
	50	0.251	0.406	-	-	-	-
2	10	0.213	0.343	-	-	-	-
	50	0.249	**0.412**	-	-	-	-
3	10	0.227	0.356	0.345	53/371	-	-
	50	**0.263**	0.411	-	-	-	-

Table 1. Performance of proposed models for local holdout test and hidden validation and test sets.

more data, we would test a refined split with an independent young group. Computational power availability was a bottleneck in our submission process. Solutions could include undersampling of the majority class or decreasing the number of residual blocks of the ResNet model.

5. Conclusion

We present a model for Chagas disease detection from 12-lead ECGs. We incorporated age, sex, and database heterogeneity into both the training and evaluation protocol, addressing key sources of bias that often limit the generalizability of AI solutions. The proposed method showed competitive performance while maintaining robustness across populations. These findings underscore the importance of bias-aware design for diagnostic tools.

Acknowledgments

This work was also partially supported by the Dirección General de Política Científica de la Generalitat Valenciana (CIPROM/2023/14) and Juan de la Cierva 2022 grant from Plan Estatal de Investigación Científica y Técnica y de Innovación 2021-2023, funded by NextGenerationEU.

References

[1] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. 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, Elola A, Hassannia MS, Campbell K, Bahrami Rad A, Ribeiro AH, Ribeiro AL, Sameni R, Clifford GD. 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, Seyedi S, Campbell K, Li Q, Rad AB, Ribeiro AH, Ribeiro ALP, Sameni R, Clifford GD. Detection of Chagas Disease from the ECG: The George B. Moody PhysioNet Challenge 2025, 2025.

[4] Ribeiro A, Ribeiro M, Paixão G, Oliveira D, Gomes P, Canazart J, Ferreira M, Andersson C, Macfarlane P, Meira WJ, Schön T, Ribeiro A. 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, Bierrenbach A, Ferreira J, Haikal D, Reingold A, Ribeiro A. 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, Schaeffter T. PTB-XL, a Large Publicly Available Electrocardiography Dataset. Scientific Data 2020;7:154.

[7] Nunes M, Buss L, Silva J, Martins L, Oliveira C, Cardoso CS BB, Ferreira A, Oliveira L, Bierrenbach A, Fernandes F, Busch M, Hotta V, Martinelli L, Soeiro M, Brentegani A, Salemi V, Menezes M, Ribeiro A, Sabino E. Incidence and Predictors of Progression to Chagas Cardiomyopathy: Long-Term Follow-Up of Trypanosoma Cruzi-Seropositive Individuals. Circulation 2021;144(19):1553–1566.

[8] Pinto-Filho M, Brant L, Dos Reis R, Giatti L, Duncan B, Lotufo P, da Fonseca M, Mill J, de Almeida M, MacFarlane P, Barreto S, Ribeiro A. Prognostic Value of Electrocardiographic Abnormalities in Adults From the Brazilian Longitudinal Study of Adults' Health. Heart 2021;107(19):1560–1566.

[9] Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV). Springer, 2018; 3–19.

[10] Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV). IEEE, 2017; 2980–2988.

Address for correspondence:

Jorge Sánchez Arciniegas
Camino de Vera s/n, 46022 Valencia (Spain)
jorsana4@upv.edu.es