Self-Supervised ECG Foundation Models for Enhanced Detection of Paroxysmal Atrial Fibrillation

Sara Artal, Antonio Miguel, Juan Pablo Martínez, Julia Ramírez

Aragon Institute of Engineering Research, University of Zaragoza
Zaragoza, Spain

Background: This study systematically investigates the impact of key design choices in ECG foundation models (ECG-FMs) on representation quality and downstream performance. These models, pre-trained via self-supervised learning (SSL) on large unlabeled datasets, are evaluated in the context of paroxysmal atrial fibrillation (PAF) detection, an intermittent supraventricular arrhythmia that remains challenging to detect in low-label settings.

Materials and Methods: First, a supervised model based on a five-layer convolutional neural network (CNN) was developed to detect PAF from ECGs in sinus rhythm. Next, a large-scale collection of publicly available ECG datasets was assembled, resulting in 1,784,071 ECG recordings from 513,329 patients. Using this dataset, restricted to lead I and wearable ECG signals, six ECG-FMs were pre-trained in a self-supervised manner using contrastive predictive coding (CPC). The variations across the six implemented CPC models included dataset scale (N), optimization strategies (learning rate (Lr) and batch size (BS)), and data augmentations. Finally, each CPC model was used as a frozen feature extractor, and the resulting representations were combined with simplified CNN architectures for downstream PAF detection.

Results: The supervised baseline achieved an area under the curve (AUC) of 0.606 and a specificity (Sp) of 0.407 for a sensitivity (Se) of 0.750. In contrast, the best-performing CPC model, CPC-5, which combined all modifications (larger N, lower Lr, larger BS, and augmentations), together with a simplified single-layer CNN, markedly improved PAF detection performance to an AUC of 0.716 and a Sp of 0.573 at the same Se level.

PAF detection measuring AUC and Sp (Se=0.750).
HP=hyperparameters, CL=convolutional layers.

Model	HP	5 CL	1 CL
Sup.	–	0.606 / 0.407	0.598 / 0.361
CPC-0	Baseline	0.653 / 0.457	0.651 / 0.432
CPC-1	Larger N	0.622 / 0.398	**0.680 / 0.473**
CPC-2	Lower Lr	0.675 / 0.489	**0.705 / 0.568**
CPC-3	Larger BS	0.601 / 0.332	0.651 / 0.455
CPC-4	Augments	0.489 / 0.045	0.519 / 0.343
CPC-5	All	0.660 / 0.473	**0.716 / 0.573**

Conclusions: The use of larger unlabeled datasets and lower learning rates yields the greatest improvements for the detection of PAF. These findings underscore the importance of systematically evaluating training design choices, as SSL methods are highly sensitive to them.