Electrocardiogram Foundation Models for Ambulatory Signal Quality Assessment in the Presence of Ventricular Arrhythmias

Lorenzo Bachi[1], Magda Costi[2], Lucia Billeci[1]

[1]Institute of Clinical Physiology, National Research Council of Italy (CNR-IFC), Pisa, Italy; [2]Cardioline S.p.A., Trento, Italy

The signal processing algorithms for the electrocardiogram (ECG) are rapidly being reshaped by artificial intelligence. Due to the considerable availability of annotated resting ECG databases, large foundation models (FMs) have recently been developed. However, the availability of annotated ECG databases from exercise or ambulatory settings is reduced. Nonetheless, representations learned from resting ECG may still be useful in the accurate processing of ambulatory or exercise ECG, despite different operating conditions and lead configurations.

In this study, the ECG FM originally proposed by McKeen *et al.* (2025) is evaluated on a previously illustrated annotated database that includes normal ECG, ventricular arrhythmias, and noise episodes from multiple ambulatory devices. The database consisted of 34505 10s windows of single-lead ECG originating from 1507 subjects. Each 10s window was annotated with a binary signal quality outcome Y and a secondary outcome V used to flag episodes of ventricular arrhythmias. Zero-shot inference was performed using the ECG-FM's built-in quality and diagnostic labels. Additionally, several supervised classifiers were trained on ECG-FM embeddings and benchmarked against a feature-based pipeline, using 5 cross-validation stratified by subject.

Zero-shot inference using single-lead ECG did not yield acceptable results, with AUROC = 0.412 for the "Poor data quality" label and AUROC = 0.712 for the "Ventricular arrhythmia" label. Conversely, classifiers trained on the FM's embeddings attained higher performances, with logistic regression reaching a Matthew's correlation coefficient (MCC) of 0.868 ± 0.032 over 5 cross-validated folds, outclassing the best model of the feature-based machine learning pipeline (MCC = 0.851 ± 0.039). Results indicate that embeddings encode relevant information that meaningfully translates to ambulatory recordings. However, zero-shot quality assessment shows domain-dependent limitations, suggesting that labels calibrated on resting ECG do not generalize to Holter contexts.

Model	Classifiers trained on embeddings (MCC)	Feature-based machine learning pipeline (MCC)
Logistic regression	0.868 ± 0.032	0.754 ± 0.082
Support vector machine	0.867 ± 0.037	0.828 ± 0.042
XGBoost	0.861 ± 0.038	0.859 ± 0.037
Multilayer perceptron	0.865 ± 0.029	0.851 ± 0.039

Noise detection results for classifiers trained on 1) FM's embeddings or 2) ECG features.