Cross-Dataset Comparison to Identify Short-Term Atrial Fibrillation During Sinus Rhythm

Paul Kapust*, Beatriz Cosculluela, Massimo Walter Rivolta, Fabio Badilini, Roberto Sassi, Javier Ruiz López, Borja Gonzálvez

Smart Solutions Technologies S.L. (NUUBO), Madrid, Spain

Aim: While atrial fibrillation (AF) is the most common arrhythmia, it is rarely detected in routine clinical electrocardiograms (ECGs). Previous studies have suggested that patients with AF can be distinguished from those without AF based on sinus rhythm electrocardiograms using machine learning. Despite this, it remains uncertain how well such models generalize across datasets acquired from different Holter devices. In this study, we investigated the ability to distinguish individuals with and without paroxysmal AF within a 24-hour window using heart rate variability (HRV) features derived from beat-to-beat (RR) intervals.

Methods: Two independent datasets were utilized, collected with distinct 2-lead Holter systems, differing in original recording duration, country of origin and lead positioning. From each 24-hour window, multiple RR segments of 500 consecutive beats were extracted at least one hour away from AF episodes. A total of 27 HRV features were computed spanning time, frequency, non-linear, fragmentation, and recurrence plot-based domains. A random forest classifier was trained using nested cross-validation with greedy forward feature selection.

Results: When evaluated separately, the model trained on the smaller short-term Holter dataset achieved higher performance than the model trained on the larger long-term Holter dataset (ROC-AUC: 0.81 vs 0.75). However, during cross-dataset evaluation, the model trained on the larger dataset generalized better while the model trained on the smaller dataset showed reduced performance (ROC-AUC: 0.85 vs 0.67).

Discussion: These results highlight the potential of HRV-based approaches, while demonstrating generalization across different Holter devices. This is supported by the consistent selection of HRV features, including SODP Q1 and SD1/SD2, among datasets and studies. Overall, these findings emphasize the importance of dataset diversity for robust cross-dataset generalization.