Deep Anomaly Detection of Bipolar Electrograms for Characterizing the Atrial Substrate

Filippo Uslenghi[1], Lorenzo Gigli[2], Massimo W Rivolta[1], Roberto Sassi[1]

[1] Dipartimento di Informatica, Università degli Studi di Milano, Milan, Italy
[2] De Gasperis Cardio Center, Electrophysiology Unit, Niguarda Hospital, Milan, Italy

Aims: This study aimed to develop a deep-learning-based anomaly detection approach using normalizing flows (NF) to model bipolar electrograms (EGMs) from patients with normal atrial substrate and an early disease profile, and to detect EGMs from increasingly compromised atrial fibrillation (AF) phenotypes as anomalies. Because EGM interpretation is central to atrial substrate characterization, deep learning could provide an objective and scalable way to quantify signal abnormalities and support substrate assessment.

Methods: Bipolar EGMs were collected during routine left atrial mapping at Niguarda Hospital (Milan, Italy) over four years (133 patients). Activations were sampled at 1 kHz, bandpass filtered (30–300 Hz), downsampled to 700 Hz and z-normalized, yielding $138,816$ segments 90 samples long (~ 0.13 s). Patients were stratified into three cohorts: C1 (paroxysmal AF, normal substrate, mapped in sinus rhythm; n=75), C2 (persistent AF, abnormal substrate or AF mapping; n=46), and C3 (persistent AF, abnormal substrate, AF mapping; n=12, subset of C2). A 16-layer affine-coupling NF was trained only on C1 (with an internal train/validation/test split) to learn the distribution of EGMs from normal substrate and then evaluated on the C1 test set and the held-out C2–C3 cohorts to assess separability. Patient-level anomaly scores were derived from latent-space embeddings by mapping the norm of each NF-transformed sample to a probability-based anomaly score.

Results: Receiver Operating Characteristic analysis based on the anomaly score showed cohort separability that increased with phenotype severity: area under the curve was 0.79 (95% CI: 0.67–0.89) for C1 vs C2 and 0.95 (95% CI: 0.85–1.00) for C1 vs C3. Confidence intervals were estimated by bootstrap resampling (10,000 trials). These results suggest NF provides an effective approach for patient-level EGM anomaly scoring in AF patients.