Reconstruction of Complete 12-Lead ECGs from Incomplete Digitized ECG Printouts

Mohammadsina Hassannia[1], Matthew A. Reyna[1], Gari D. Clifford[1,2], and Reza Sameni[1,2]

[1]Department of Biomedical Informatics, Emory University, USA
[2]Department of Biomedical Engineering, Georgia Institute of Technology, USA

The electrocardiogram (ECG) remains an accurate and widely accessible tool for cardiovascular screening and diagnosis. Despite the prevalence of digital systems, paper printouts and scanned ECG images remain common in clinical workflows and archival datasets. Recent open-source efforts, including the PhysioNet Challenge 2024 and subsequent Kaggle competitions, have enabled reliable digitization of ECG time series from such images. However, standard ECG printouts interleave leads, typically providing only 2.5 s from most leads and up to 10 s from one or a few rhythm leads, leaving missing temporal segments after digitization. This limits the direct use of most automated ECG analysis tools that require complete multi-lead recordings for measurements and algorithmic diagnosis.

We address this challenge by developing a noise-robust method to reconstruct full 10 s, 12-lead ECG signals from partially observed digitized printouts. The approach uses the available 10 s rhythm lead (commonly Lead II) to detect R-peaks, which are then used to temporally align and segment beats across all leads. Robust weighted averaging is applied to estimate lead-specific average beat templates. These templates are replicated at detected R-peak locations to reconstruct missing segments, with overlapping regions combined using a Bayesian-inspired weighting scheme. A final post-processing step enforces physical constraints and known inter-lead relationships to improve consistency.

The method was evaluated on the PTB-XL dataset (21,837 recordings, sampling frequency 500 Hz) under simulated partial observation conditions reflecting standard $3 \times 2.5\,\mathrm{s} + 1$ rhythm lead layouts, and on real digitized ECGs from a recent Kaggle challenge. Reconstruction quality, measured by SNR over missing segments, reached 12.28 dB (Lead I), 13.12–13.53 dB (V1–V3), and 13.71–14.18 dB (V4–V6), with standard deviations of 4.56–5.63 dB. Errors had minimal impact for post hoc biomarker estimation, including QT-interval measurement.

This framework enables reliable recovery of complete ECG signals from partial printouts, supporting clinical measurements and downstream machine learning on digitized ECG archives.