On the role of the tokenizer in ECG transformer models

Jiawei Li, Fabio Bonassi, Johan Sundström, Thomas B. Schön, Antônio H. Ribeiro

Uppsala University, Uppsala, Sweden

Background: Despite the success of Transformer architectures in general time-series tasks, their application to ECG analysis often yields suboptimal results. We hypothesize that this performance bottleneck originates from the tokenization process. Specifically, conventional fixed-length patching adapted from Vision Transformers disrupts the intrinsic morphological integrity and rhythmic patterns essential to cardiac electrophysiology.

Methods: We evaluated a diverse set of models, including CNNs (e.g., ResNet, InceptionTime), State-Space Models (SSM), and Transformer-based time-series models, on the CPSC2018 dataset, which comprises records from 6.8K patients across nine rhythm classes: atrial fibrillation, first-degree atrioventricular block, left bundle branch block, right bundle branch block, premature atrial contraction, premature ventricular contraction, ST-segment depression, and ST-segment elevation. To isolate the impact of signal partitioning, we replaced the standard fixed-length patching with two physiologically-aware strategies: the Heartlang Tokenizer, which treats each heartbeat as a standalone token, and a Median Beat Tokenizer, which represents the ECG signal as a single representative heartbeat, with each individual time point serving as a token.

Results: Our findings quantify a significant inductive bias in standard Transformers. While vanilla models with patching tokenizers initially underperformed compared to CNN baselines (e.g., iTransformer macro AUC: 0.705 vs. LSTM: 0.965), the adoption of the Heartlang and Median Beat Tokenizer resulted in a substantial performance recovery across the board. Notably, the iTransformer's AUC improved by over 30%, while Informer and Reformer saw gains of over 14%, effectively bridging the gap with state-of-the-art CNNs and SSMs.

This study suggests that the bottleneck in current ECG Transformers is the tokenizer. By leveraging the Heartlang and Median Beat Tokenizer, we show that preserving the morphological integrity of heartbeats is crucial for these models to overcome their inherent lack of translation invariance. Our results suggest that future ECG Transformer research should prioritize domain-specific tokenization to fully unlock the potential of large-scale architectures.

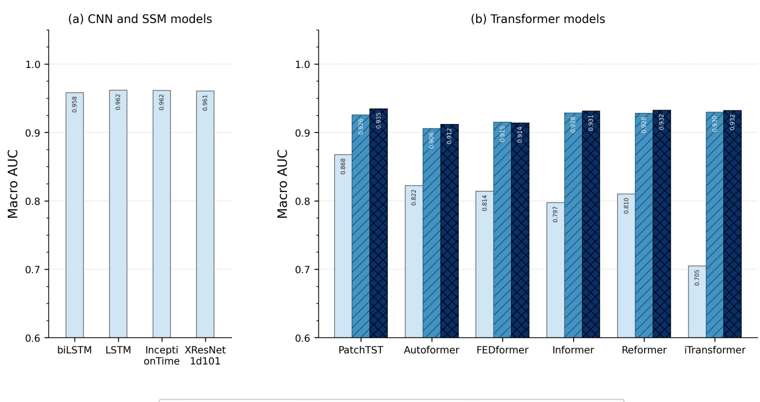

Figure 1. Macro AUC scores on the CPSC2018 dataset using CNN, SSM, and Transformer architectures with various tokenizers.