

# A Fast Algorithm for Facilitating Heartbeat Annotation in Long-Term ECG Signals

Ana Santos Rodrigues\*, Mantas Lukoševičius, Vaidotas Marozas

Biomedical Engineering Institute, Kaunas University of Technology, Kaunas, Lithuania

**Aim:** Machine-learning models for automatic interpretation of ECG signals require properly labeled datasets. However, manual annotation is arduous and time-consuming, particularly in long-term recordings. We propose a fast semi-automatic algorithm for facilitating long-term ECG analysis and annotation, free from fatigue-caused errors or exorbitantly long computations.

**Methods:** Heartbeats are compressed to short strings using Symbolic Aggregation Approximation (SAX). Every unique string becomes a pre-cluster of all the beats represented by equal strings, dramatically reducing the amount of memory and computations required. Next, hierarchical clustering of signal-averaged beats of every pre-cluster is performed. Clusters can then be conveniently analyzed and annotated by the researcher.

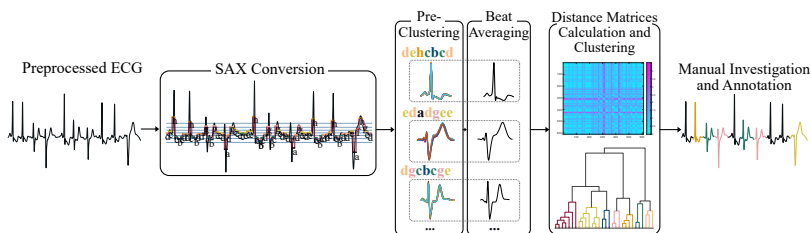


Figure 1. Employment of SAX and clustering of heartbeats by similarity.

**Results:** We tested the accuracy of our approach on the Physionet's MIT-BIH database with known classes. On average, beat misclassification was  $<0.9\%$  after SAX, and  $<3.3\%$  over 13 clusters. Misclassified beats are beats that ended up in clusters where the dominant class is different.

To test efficiency, we employed our algorithm to annotate our long-term ECG recordings (two to three days long). In one recording, more than 480 000 beats were reduced to less than 30 000 unique strings and further to 30 clusters, which were then annotated as four classes plus noise. The computations took less than an hour on an AMD Ryzen™ 5 3.6 GHz CPU with six cores (12-threads) and 16 GB of RAM.

**Conclusions:** The proposed algorithm is accurate and efficient enough to facilitate practical long-term ECG signal beat analysis and annotation.