KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Varon C., Moeyersons J., Vandenberk B., Caicedo A., De Cooman T., Lazaro J., Van Huffel S., Bailón R. (2017), **Random Forest Classification of Single-Lead ECG segments using Morphology and Rhythm Changes** 44th annual Computing in Cardiology Conference
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	NA
Journal homepage	https://www.cinc2017.org/
Author contact	Carolina.varon@esat.kuleuven.be +32 16326417
IR	NA

(article begins on next page)

Random Forest Classification of Single-Lead ECG segments using Morphology and Rhythm Changes

Carolina Varon*, Jonathan Moeyersons, Bert Vandenberk, Alexander Caicedo, Thomas De Cooman, Jesús Lázaro, Sabine Van Huffel

KU Leuven, Leuven, Belgium

This work tackles the problem of classification of normal, AF, other rhythms, and too noisy single-lead ECG segments by extracting 3 morphology features and 3 rhythm features from the ECG and feeding them into a random forest classifier. The morphology features are:

1) The standard deviation of the widths of all QRS complexes in a segment. This feature is able to identify too noisy segments.
2) The power of the ECG signal after removing the QRS complexes, in the band from 3 to 7 Hz, where the information of the P and T waves is expected.
3) The percentage of variance contained in the second principal component of the T-waves. This represents the T-wave modulation and is larger when the main modulator of the amplitude of the ECG is not the respiration.

The rhythm features are derived from the tachogram. They are the mean heart rate, the Shannon entropy, and the slope of the phase rectified signal averaging curve.

The dataset was split into training (80%) and test (20%). The training set was first selected using the fixed-size algorithm, which is based on the maximization of the Rényi entropy, and then used to train a random forest classifier with 1000 bootstrap replicates. Afterwards, the classifier was tested on the 20% remaining data points, and an overall F1 score of 0.66 was achieved, with 0.86, 0.68, 0.55, and 0.53 for the normal, AF, other, and noisy classes, respectively. When testing the model on the hidden set, an overall performance of 0.48 was obtained. There are, however, several possible improvements. For instance, the detection of contaminated parts within each segment, and the calculation of the features using a moving window approach.