

## Comparison of Noise Indexes for an Ambulatory Electrocardiogram Database with Ventricular Arrhythmias

Lorenzo Bachi<sup>1</sup>, Maurizio Varanini<sup>1</sup>, Magda Costi<sup>2</sup>, David Lombardi<sup>2</sup>, Lucia Billeci<sup>1</sup>

<sup>1</sup>Institute of Clinical Physiology, National Research Council of Italy (CNR-IFC), Pisa, Italy

<sup>2</sup>Cardioline S.p.A., Trento, Italy

Automatic analysis of the electrocardiogram (ECG) plays a key role in ambulatory recordings. While noise episodes can be detected using noise indexes, scientific literature on this matter has been focused on normal ECG vs noise classification. However, arrhythmias of ventricular origin exhibit a different pattern compared to normal or otherwise supraventricular ECG.

In this study, four ECG databases were combined into a single database containing normal ECG, noise episodes, and ventricular arrhythmias. Each ECG signal of the database was split into single-lead, 2 seconds-long windows. The signal quality of each of the 155643 windows was annotated by an expert human observer using a binary outcome:  $Y = 1$  for noise,  $Y = 0$  for clean ECG. For each 2 s window, twelve noise indexes were computed, abbreviated as *kur*, *skew*, *rpow*, *bas*, *ior*, *msqi*, *snr*, *wave*, *hos*, *se*, *edp*, *inv*. These noise indexes included popular metrics as well as two novel indexes we introduced (*edp*, *inv*). Univariate analysis of each noise index was first performed. The relative usefulness of each index was then gauged by training a decision tree model with 10-fold cross-validation and assessing the feature importance. Each decision tree was trained using all noise indexes.

Cohen's *d* analysis highlighted a large effect for only six out of the twelve noise indexes (*kur*, *msqi*, *snr*, *se*, *edp*, *inv*). Moreover, Matthew's Correlation Coefficient (MCC) of logistic regression between each noise index and the outcome was above 0.5 only for two indexes (*edp*, *inv*). The 10 best decision trees resulting from 10-fold cross-validation reached an average MCC of 0.74, sensibility of 0.74, specificity of 0.96, and feature importance was highest for the *wave*, *hos*, *se*, *edp*, and *inv* indexes on all folds. The average specificity on records of clean ECG with ventricular arrhythmias was also computed and amounted to 0.95.

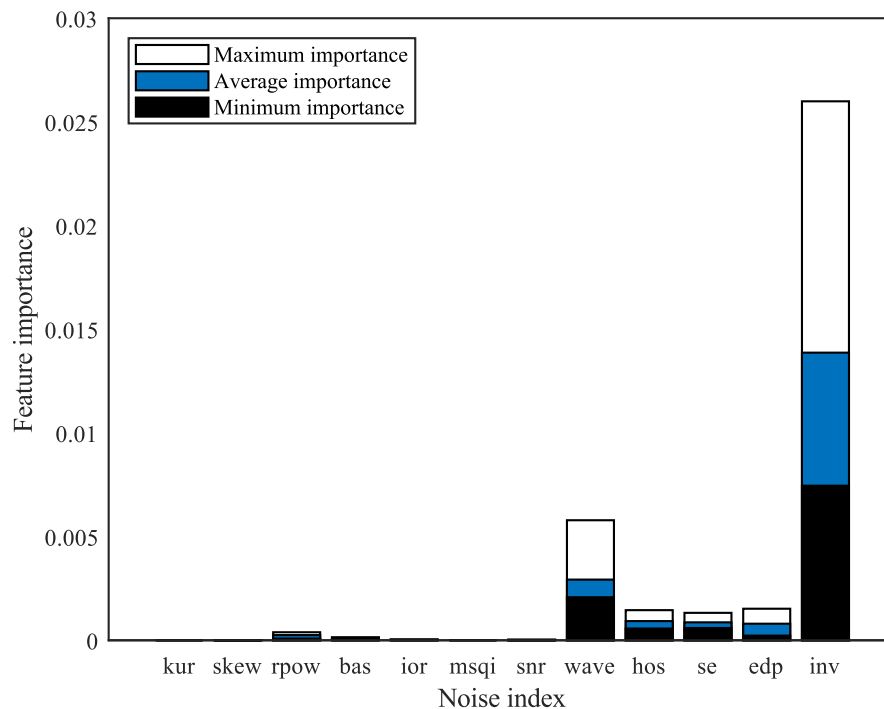


Figure 1. Maximum (white), average (blue), and minimum (black) feature importance for each noise index resulting from training a decision tree classification model on all noise indexes with 10-fold cross-validation.