

Multi-Class ECG Feature Importance Rankings: Cardiologists vs. Algorithms

Philip J. Aston*, Temesgen Mehari, Alen Bosnjakovic, Peter M. Harris, Ashish Sundar, Steven E. Williams, Axel Loewe, Claudia Nagel, Nils Strodthoff, Olaf Dössel

National Physical Laboratory/University of Surrey
Teddington, UK

Introduction Cardiologists diagnose over 150 conditions from an electrocardiogram (ECG) based on interval, amplitude and timing features. For each pathology, conditions on specific features are well documented.

Conversely, there are numerous algorithms that rank feature importance for a particular classification task. However, different algorithms often give quite different rankings. Therefore, we compared the feature importance ranking obtained by various algorithms with the features that cardiologists use for diagnosis.

Methods In previous work, we considered feature importance rankings for binary classifications of Normal vs. a single pathology. Here we consider a multi-class classification with classes of Normal, first degree atrioventricular (AV) block, right bundle branch block (RBBB) and left branch bundle block (LBBB). Cardiologists use only the PR interval to diagnose first degree AV block, and 7 or 14 features to diagnosis RBBB or LBBB respectively. For Normal class diagnosis, we took the union of features used for diagnosis of the three pathologies, which gave a total of 18 features.

We obtained four feature importance rankings for binary classifications of one class vs. all the other classes. A scoring algorithm was devised to compare rankings with the features used by cardiologists.

Results As with the single pathology rankings, most methods ranked the PR interval first for AV block. For the other classes, our scoring algorithm generally gave high scores for RBBB, middling scores for Normal, and low scores for LBBB. Including features that correlate with the features used for diagnosis increased the scores, in some cases significantly. Methods based on Random Forests generally gave the best scores while ReliefF performed poorly.

Conclusions In this multi-class classification problem, the feature rankings are quite similar to those for the corresponding binary classifications with scores for RBBB being higher than for LBBB. Scores generally decrease for conditions that involve many features for diagnosis.