Explaining black-box automatic electrocardiogram classification to cardiologists

Derick M Oliveira; Antônio H. Ribeiro; João A. O. Pedrosa; Gabriela MM Paixão; Antonio L Ribeiro; Wagner Meira Jr.
Universidade Federal de Minas Gerais, Brazil

Introduction: Deep neural networks trained on large datasets have demonstrated the ability to provide extremely accurate automated analysis of the electrocardiogram (ECG). These models use the raw signal as an input to the classifiers, being called "end-to-end" approaches. There are, however reported case studies for which the neural network learns to solve the task in unwanted ways or for which small perturbations may have a huge impact on the model prediction outcome.

Methods: In this work, we developed a method to understand the ECG "end-to-end" classifiers and report a close-to-cardiologist interpretation of the model output. The proposed method is used to analyse the deep learning ECG classification models from [Ribeiro, A.H. et al., 2019]. ECG segmentation method Neurokit [Makowski, Dominique et al., 2020] was applied to ECG tracings, to obtain (1) Intervals, Segments and Axis; (2) Rate and (3) Rhythm. Noise was added to the signal to disturb the ECG features in a realistic way. The method is tested using Monte Carlo simulation and the feature impact is estimated by the change in the model prediction averaged over 499 executions. A feature is defined important if the mean value change the result of the classifier.

Results: Table 1 display our results for the 6 classes of ECG abnormalities of the original study, showing which perturbation is related to each ECG abnormality.

Table 1. Relevance of interpretable features modified by random noise for the automatic diagnosis of each ECG class using the deep learning classifier.

		Measurements							
	#Exams	P Wave	PR Interval	QRS Complex	QT Interval	T Wave	A/V Rate	Rhythm	Random
1dAVb	30	0.00	0.30	0.78	0.26	0.00	0.00	0.26	0.00
RBBB	38	0.00	0.00	0.52	0.00	0.21	0.48	0.64	0.00
LBBB	30	0.00	0.00	0.86	0.00	0.07	0.14	0.25	0.00
SB	18	0.00	0.00	0.00	0.00	0.00	0.86	1.00	0.00
AF	10	0.00	0.30	0.20	0.00	0.00	0.70	1.00	0.00
ST	38	0.00	0.00	0.03	0.00	0.00	0.69	1.00	0.00

Conclusions and future works: We developed a method of explaining an automatic ECG classifier based in deep learning, using features that are understandable by any cardiologist. Further researcher is needed to evaluate the accuracy of the present method, and how it can be adopted in the clinical practice.