ECG Quality Assessment via Deep Learning and Data Augmentation

A. Huerta, A. Martínez-Rodrigo, J.J. Rieta, R. Alcaraz

Research Group in Electronic, Biomedical and Telecommunication Engineering, University of Castilla-La Mancha Cuenca, Spain

Background and Aim. Quality assessment of ECG recordings acquired with wearable systems is essential to avoid misdiagnosis of some cardiac disorders, such as atrial fibrillation (AF). For this purpose, novel deep learning algorithms have been recently proposed. However, training of these methods require large amount of data, and only reduced public databases with annotated samples are nowadays available. Within this context, the present work aims at validating usefulness of a well-known data augmentation approach in the context of ECG quality assessment.

Methods. As a reference, 2,000 5 s-length ECG excerpts from the PhysioNet/CinC Challenge 2017 database were used to train and test a deep learning algorithm via 5-fold cross-validation. Half of the ECG segments were extracted from noisy recordings and the other half from high-quality signals, which presented normal sinus rhythm (NSR), AF and other rhythms (OR) in similar proportions. Making use of a well-known approach based on time and pitch shifting, noise addition and time-scale modification of the original noisy ECG experts, 1,000 additional low-quality intervals were generated. Then, the algorithm was again trained and tested with these surrogated noisy ECG segments and the original high-quality ones. Results for both cases were finally compared using a McNemar test. The used classifier was a pre-trained 2-D AlexNet structure which is fed with scalograms of the ECG intervals, obtained through a continuous Wavelet transform.

Results. For both analyses, no statistically significant differences (p-value > 0.05) were noticed in terms of accuracy (Acc), sensitivity (Se) and specificity (Sp), such as the table below shows. Similarly, no differences in the rate of NSR ($\mathcal{R}_{NSR}$), AF ($\mathcal{R}_{AF}$) and OR ($\mathcal{R}_{OR}$) episodes correctly classified were also observed. For all cases, values about 90% were always seen, with standard deviation among the five conducted iterations being lower than 6%.

Conclusions. The analyzed data augmentation approach could be an interesting tool to increase the number of samples available for training and testing of deep learning algorithms designed for ECG quality assessment, thus resulting in more robust and reliable methods.

Dataset	Acc	Se	Sp	$\mathcal{R}_{NSR}$	$\mathcal{R}_{AF}$	$\mathcal{R}_{OR}$
Original	0.88±0.02	0.89±0.02	0.87±0.045	0.92±0.03	0.88±0.03	0.87±0.04
Surrogated	0.90±0.01	0.90±0.03	0.90±0.04	0.91±0.02	0.89±0.06	0.89±0.05