Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240547

A Novel Method and Python Library for ECG Signal Quality Assessment

Charles BERGER[a,1], Hugues TURBÉ[b,1,2], Mina BJELOGRLIC[b] and Christian LOVIS[b]
[a] *Institute of Bioengineering/Center for Neuroprosthetics, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*
[b] *Division of Medical Information Sciences, Geneva University Hospitals, Geneva, Switzerland and Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland*

Abstract. Electrocardiogram (ECG) is one of the reference cardiovascular diagnostic exams. However, the ECG signal is very prone to being distorted through different sources of artifacts that can later interfere with the diagnostic. For this reason, signal quality assessment (SQA) methods that identify corrupted signals are critical to improve the robustness of automatic ECG diagnostic methods. This work presents a review and open-source implementation of different available indices for SQA as well as introducing an index that considers the ECG as a dynamical system. These indices are then used to develop machine learning models which evaluate the quality of the signals. The proposed index along the designed ML models are shown to improve SQA for ECG signals.

Keywords. Electrocardiogram, Signal Quality Assessment (SQA), Time Series Dimension (TSD)

1. Introduction

Electrocardiogram (ECG) is an important diagnostic tool, serving as an inexpensive, non-invasive, quick, and safe tool to acquire several important heart physiology's parameters. Nowadays, ECG signals are often interpreted by automatic diagnostic tools [1] which mostly rely on machine learning (ML) or Deep Learning (DL) techniques. An important challenge encountered by these methods is signal interactions with external and undesirable sources. ECGs are often exposed to undesirable sources, which we will refer to as artefacts. These alterations affect the automatic classification and detection of heart pathologies by raising false alarms leading to misdiagnosis [2]. The most common artefacts encountered are [3]: i) Baseline Wander (BW) : A low-frequency artefact caused by the patient's respiration or body motions as well as poor electrode contact. ii) Power Line Interference (PLI): A bandpass artefact due to the ambient sector power line and iii) Electromyographic artefact (or muscle artefact): A wide range frequency artefact caused by ambient muscle activity (e.g., body movement). BW and PLI can be removed by using sets of digital signal filtering methods (such as Bandpass or Butterworth filters) [4]. Muscle artefact is the most troublesome as its frequency range overlaps greatly with

[1] Equal contribution
[2] Corresponding Author: Hugues Turbé; E-mail: hugues.turbe@unige.ch.

that of the ECG signal [3]. In that case, common filtering methods would distort the ECG signal [3].

As the different artefacts cannot be removed without altering the ECG signal, various Signal Quality Assessment (SQA) methods have been developed [5,6] to assess the quality of the signal and prevent misdiagnosis on very noisy signal. However, a lack of clarity in the methods used has been identified in the literature [7]. In addition, most SQA methods leverage only frequency bands or correlation across signals not considering the dynamical nature of the ECG signal. The contributions of the presented research are three-fold: i) the evaluation of a range of different indices commonly found in the literature with an open-source implementation of the indices in Python, ii) a new dynamical index used for the first time in the context of ECG SQA, iii) ML based models for SQA of ECG signals.

2. Methods

The presented research is based on data from the “PhysioNet 2011 challenges” (referred to as Cinc2011) [8]. This dataset is still the most used for the evaluation of SQA methods on ECG signals [5,7]. It contains 12-leads ECG records for 1000 patients. The signal quality of each ECG record was assessed by multiple annotators, as either “acceptable” (773 ECGs), “unacceptable” (225 ECGs), or "undetermined" (2 ECGs).

The following indices from the literature were included in the developed SQA methods:

- Wavelet Probability Mass Function (*wPMF*): Ratio of the energy contained in sub-band frequency to the entire signal's energy, measured using wavelet decomposition.
- Signal-to-Noise Ratio for ECG (*SNR*): Ratio of the signal's Power Density Spectrum (PSD) in ECG frequency range (2-40 Hz) to the PSD found outside this range.
- Flatline detection (*F*): Percentage of the lead that is flat.
- Heart Rate (*HR*): Heart rate measured by calculating the RR interval using the Pan-Tompkins algorithm.
- Intralead Correlation Coefficient (r_{intra}): Pearson correlation between each QRS complex in a given lead and the average QRS signal in this lead.
- Interlead correlation coefficient (r_{inter}): Pearson correlation between each lead.

The main caveat of indices presented above is that most of them are not able to make an efficient separation between the signal and artefact sources. Most of them rely only on frequency bands to separate the ECG signal from the noise. The Time Series Dimension (*TSD*) [9] index considers the ECG signal as a measurement of a dynamical system. It corresponds to a measure of signal’s predictability. This index determines the fractal dimension of a time series using the Higuchi method as follows:

$$D^q = \frac{\log \overline{\overline{L_q}}(\theta_1) - \log \overline{\overline{L_q}}(\theta_2)}{\log 2}, \text{ with q= Lead index} \quad (1)$$

where L_q represent the mean length of lead q, using a sampling time of $\theta_k = k \times \Delta t$. Here, $D^q \in [1,2]$ is the fractal dimension of the signal with 1 indicating a signal with only deterministic components, 1.5 only stochastic components and 2 being equal to white noise. This index was implemented with slight modifications to consider flatline cases. It was found that the Higuchi method can be unstable if the signal is close to a horizontal line [10]. The value of the TSD was set to 2 if any flatlines were present in the signal. Further studies analyzing sensibility of this method to different noise types can be found on the GitHub page[3].

2.1. Individual index evaluation

The indices were first evaluated individually on the 2011 PhysioNet dataset. Different metrics are commonly reported in the literature to evaluate SQA tasks but the Matthew's correlation coefficient (MCC) has recently been recommended for this type of study [7]. This metric was used to select the best model as it best summarizes a model performance for two-class unbalanced classification tasks. In this study, the positive class was designed as the class corresponding to ECG deemed of an unacceptable quality. Considering the context of developing an SQA method as part of a pipeline for automatic ECG classification, we are particularly interested in finding signal with non-acceptable quality. The metrics presented next therefore make more sense with this convention. A five-fold approach was used to determine the optimal threshold applied to transform each index into a binary label. For each fold, we determined the optimal threshold by taking the value maximizing the MCC metric. At the end of the stratified cross validation process, the metrics were averaged across all folds. All the indices and models are publicly available on GitHub[3].

2.2. Model development and evaluation

Following the individual evaluation of the indices, the latter were used as input features for the development of ML-based classification models. An initial feature selection process was performed to extract the most relevant indices. The mean of each index across the 12 leads was used as the input feature for all feature except the heart rate. For the heart rate, the minimal value computed across the lead was used. Three different feature selection methods: were tested: i) a backward model selection based on the p-value with the Akaike Information Criterion (AIC) used as the stopping criterion, ii) Logistic regression model with L2-regularisation, iii) Mutual Information (MI) filter-based algorithm [11]. The Historical JMI developed by Gocht et al. [12] was used to select the optimal set of features based on this last method.

Prior to this step, all indices were normalized between zero and one, with one indicating perfect signal quality. Two models were then evaluated using the features selected by the different feature selection methods: i) a logistic regression and ii) a light Gradient Boosted Machine (GBM) based tree algorithm [13]. The logistic regression model was adapted to classify an ECG as bad quality if an unacceptable heart rate was measure, either below 24 BPM or over 420 BPM. The light Gradient Boosted Machine decision tree (LGBM) was trained with a focal loss [14]. Hyperparameter optimization was performed to determine the optimal learning rate, number of leaves, maximum tree

[3] https://github.com/CBergerEPFL/ecg_evaluation

depth, the maximum bin number, the parameter for L1 and L2 regularization, the minimum gain to split, the bagging fraction and frequency.

3. Results

Results for the individual evaluation of the indices as well as the developed models are presented in Table 1. The features selected by the three different selection methods are presented in Table 2.

Table 1. Classification metrics along the derived optimal threshold for each of the eight individual indices as well as the developed ML models built using these indices. For the developed models, the first three columns refer to a logistic classification model trained with features selected using three different feature selection methods. A_{ROC} and A_{PR} stand respectively for area under the ROC and PR curve. Prec. stands for precision, Rec. for recall, Spec. for specificity, Acc for accuracy, MCC for Matthew's correlation coefficient and T for threshold.

	Individual index							Developed Models			
	r_{inter}	*F*	r_{intra}	*wPMF*	*HR*	*SNR*	*TSD*	L2-reg	HJMI	p-value	LGBM
A_{ROC}	0.82	0.72	0.87	0.87	0.82	**0.88**	0.67	0.93	0.93	0.94	**0.96**
A_{PR}	0.79	0.66	0.81	0.77	**0.83**	0.78	0.62	0.86	0.88	0.87	**0.91**
Prec.	0.93	**0.97**	0.95	0.96	0.93	0.86	0.95	0.82	0.85	0.86	**0.93**
Rec.	0.63	0.45	0.56	0.48	**0.66**	0.59	0.41	0.73	**0.75**	0.75	0.71
F1	0.75	0.62	0.7	0.64	**0.77**	0.7	0.58	0.77	**0.80**	**0.80**	**0.80**
Spec.	0.99	1	0.99	0.99	0.99	0.97	0.99	0.95	0.96	0.97	**0.98**
Acc.	**0.91**	0.87	0.89	0.88	**0.91**	0.88	0.86	0.9	0.91	**0.92**	**0.92**
MCC	0.72	0.61	0.67	0.63	**0.73**	0.65	0.57	0.71	0.74	0.75	**0.76**
T	0.6	0.45	0.38	0.89	1	0.52	0.56	0.36	0.36	0.29	0.52

Table 2. Selected features using the 3 different features selection methods.

Feature selection model	Selected features
Backward model selection (p-value)	r_{inter} , *wPMF, HR, TSD*
HJMI	r_{inter} , *SNRECG, TSD*, r_{intra}
L2-regularization (C=1)	r_{inter} , *SNR, HR*, r_{intra} , *wPMF*

4. Discussion and Conclusions

The study aimed to evaluate different indices both individually and as features for ML-based classification models to classify the quality of ECG signals. In addition, the TSD index was for the first time evaluated in the context of ECG SQA. This study was motivated by the lack of literature that rigorously assessed the performance of the indices previously developed. Indeed, many studies relabeled ECGs according to their personal requirement preventing a fair comparison across studies. With this work, we introduce a thorough evaluation of different common indices on an important benchmark as well as a novel index not previously used in the context of SQA methods. Another important aspect is the lack of open implementation in Python in the literature. With this study, we also released an open-source implementation of all indices on Github.

The *HR* and r_{inter} are the indices showing the best individual classification performance as shown in Table 1. The TSD index showed relatively low performance when considered by itself but was interestingly found to be included by two (best performing) out of the three feature selection methods tested (see Table 2). The proposed index therefore seems to bring new information for the quality assessment not captured by the other indices commonly found in the literature. The TSD is the only index considering the ECG signal as a dynamical system and can therefore capture source of noise which lies in the same frequency bands as the signal itself. Combining indices as part of a ML based model was found to perform better than the individual index as shown in Table 1. Across most of the metrics and especially when considering the MCC, the LGBM model was found to outperform the logistic classification models.

References

[1] Gliner V, Keidar N, Makarov V, Avetisyan AI, Schuster A, Yaniv Y. Automatic classification of healthy and disease conditions from images or digital standard 12-lead electrocardiograms. Sci Rep. 2020;10(1):16331. doi:10.1038/s41598-020-73060-w.

[2] Au-Yeung WTM, Sahani AK, Isselbacher EM, Armoundas AA. Reduction of false alarms in the intensive care unit using an optimized machine learning based approach. NPJ Digit Med.2019;2:86. doi:10.1038/s41746-019-0160-7.

[3] Satija U, Ramkumar B, Manikandan MS. An automated ECG signal quality assessment method for unsupervised diagnostic systems. Biocybern Biomed Eng. 2018;38(1):54-70. doi:10.1016/j.bbe.2017.10.002..

[4] Li H, Boulanger P. An Automatic Method to Reduce Baseline Wander and Motion Artifacts on Ambulatory Electrocardiogram Signals. Sensors. 2021;21(24):8169. doi:10.3390/s21248169.

[5] Kramer L, Menon C, Elgendi M. ECGAssess: A Python-based toolbox to assess ECG lead signal quality. Front Digit Health. 2022;4:847555. doi:10.3389/fdgth.2022.847555.

[6] Neri L, Gallelli I, Dall'Olio M, Lago J, Borghi C, Diemberger I, et al. Validation of a New and Straightforward Algorithm to Evaluate Signal Quality during ECG Monitoring with Wearable Devices Used in a Clinical Setting. Bioengineering. 2024;11(3):222. doi:10.3390/bioengineering11030222.

[7] van der Bijl K, Elgendi M, Menon C. Automatic ECG Quality Assessment Techniques: A Systematic Review. Diagnostics. 2022;12(11):2578. doi:10.3390/diagnostics12112578.

[8] Silva I, Moody GB, Celi L. Improving the quality of ECGs collected using mobile phones: The Physionet/Computing in Cardiology Challenge 2011. In: 2011 computing in Cardiology. IEEE; 2011. p. 273-6.

[9] Sase T, Ramírez JP, Kitajo K, Aihara K, Hirata Y. Estimating the level of dynamical noise in time series by using fractal dimensions. Phys Lett A. 2016;380(11):1151-63. doi:10.1016/j.physleta.2016.01.014.

[10] Liehr L, Massopust P. On the Mathematical Validity of the Higuchi Method. Phys Nonlinear Phenom. 2020;402:132265. doi:10.1016/j.physd.2019.132265.

[11] Bennasar M, Hicks Y, Setchi R. Feature selection using Joint Mutual Information Maximisation. Expert Syst Appl. 2015;42(22):8520-32. doi:10.1016/j.eswa.2015.07.007.

[12] Gocht A, Lehmann C, Schone R. A New Approach for Automated Feature Selection. In: 2018 IEEE International Conference on Big Data (Big Data). Seattle, WA, USA: IEEE; 2018 p. 4915-20. doi:10.1109/BigData.2018.8622548

[13] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al. Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2017.

[14] Ross TY, Dollár G. Focal loss for dense object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 2980-8