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

Lorenzo Bachi[1], Maurizio Varanini[1], Magda Costi[2], David Lombardi[2], Lucia Billeci[1]

[1] Institute of Clinical Physiology, National Research Council of Italy (CNR-IFC), Pisa, Italy
[2] Cardioline S.p.A., Trento, Italy

Abstract

Noise detection plays a key role in the automatic analysis of ambulatory electrocardiogram (ECG). The scientific literature on this matter has been focused on normal ECG vs noise classification. In this study, four ECG databases were combined into a single database containing normal ECG, noise episodes, and ventricular arrhythmias. Twelve noise indexes were considered, two of which were developed in this study (edp, inv). Univariate analysis of each noise index shows that some indexes lacked a strong association to the signal quality outcome. The relative usefulness of each index was gauged by training a decision tree model with 10-fold cross-validation and assessing the feature importance. Five indices were found to be most predictive of noise, including edp and inv, with average classification test performances of MCC = 0.743, sensitivity = 0.744, specificity = 0.965, and average specificity on ventricular arrhythmias = 0.952.

1. Introduction

Ambulatory ECG is an established technique in clinical practice, where ECG is recorded for a time duration ranging from several hours to up to one week in length. Its purposes include detection of rare arrhythmia episodes, investigation of symptoms, monitoring the effects of medication, and assessing the response to different conditions like sleeping or free-living activity. The ambulatory ECG recording is typically analyzed by a set of algorithms tasked with measuring several parameters related to the status of the subject's heart. A trained professional assesses the analysis report and eventually provides the patient with a diagnosis and therapeutic indications.

Ambulatory ECG is notoriously subject to several sources of noise and interference[1], and excessive noise is discarded with dedicated noise detection algorithms. This task is particularly challenging due to the ample variability shown by abnormal ECG compared to normal ECG. Noise episodes could be mistaken as clinically relevant ECG and rare ECG events could also be incorrectly labeled as noise.

Noise detection has historically been performed through noise indices, *i.e.* measures of the effect of noise on the ECG record. These are also known as feature-based approaches. Typically, these noise detection models feature multiple noise indices that attempt to capture multiple sources of interference [2]. More recently, the advent of machine learning popularized non-feature-based methods, where signal quality is assessed right on the ECG [3]. Machine learning has also proven to be an essential tool for feature-based approaches, as complex relations between noise indices can be modeled by performing model training on large enough databases.

Many ECG signal quality assessment approaches have been proposed [3]. However, few of them have been specifically targeted at evaluating noise detection in presence of abnormal ECG. In this study, several noise indices are evaluated on ambulatory ECG signals with ventricular arrhythmias. Ventricular arrhythmias differ substantially in both pattern and rhythm compared to normal ECG, and thus noise indices originally developed on normal ECG databases may be less predictive of noise presence in arrhythmic ECG.

2. Methods

2.1. Data

Four databases were included in this study:

- The Physionet/Computing in Cardiology Challenge 2011 Database [4], which features clean 12-lead ECG, flat signals, electrode placement errors, and noise.
- The MIT-BIH Malignant Ventricular Ectopy Database, [5], a collection of 2-lead Holter ECG records which includes noise and annotated episodes of ventricular arrhythmias: ventricular tachycardia, fibrillation, or flutter.
- Cardioline's Private R&D Database, composed of 25 records of 3-lead or 12-lead Holter ECG records, 20 minutes long each. This database includes clean ECG, both normal and abnormal, as well as numerous noise episodes.

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.287

- A simulated database generated with the simulator described in [6]. This database was included to increase the prevalence of ventricular arrhythmias and noise episodes. A total of 60 12-lead ECG records were simulated. Each record is 1 minute long and is sampled at 1000 Hz. Half of these records (30) feature noise episodes, which were simulated as a mixture of muscle noise and motion artifacts. The other 30 records feature clean ECG with ventricular tachycardia episodes.

An expert human data curator labeled the signal quality of leads I, II, and V1 to V6 of each database. Each 2 second single-lead ECG window was labeled independently of the other leads. A binary outcome **Y** was assigned to each 2 s ECG window k, with $Y_k = 0$ for clean ECG and $Y_k = 1$ for noise. This procedure was computed using a custom MATLAB graphic interface. The data curator discarded ECG windows of uncertain quality that could be labeled either way. Windows with missing signals were also discarded. Additionally, a second label **V** was assigned to each 2 s ECG window, where $V_k = 1$ for each 2 s window of clean ECG with ventricular arrhythmias, and $V_k = 0$ for all other records.

The labeled records were merged in a single database composed of 148502 records, with 30358 (20.4 %) records of noise and 118144 (79.6 %) records of clean ECG. Among the clean ECG records, 10768 (7.25 %) featured ventricular arrhythmias.

2.2. Noise Indices

Ten noise indices from relevant scientific literature were selected for this study. These indices are kurtosis (abbreviated as *kur*) [2], skewness (*skew*) [2], relative power in the QRS complex (*rpow*) [2], relative power in the baseline (*bas*) [2], in-band to out-of-band spectral power ratio (*ior*) [7], modulation spectrum quality index (*msqi*) [8], signal-to-noise ratio computed with wavelet decomposition (*wave*) [9], high order statistic (*hos*) [10], sample entropy (*se*) [11]. Additionally, two novel noise indices were also included: the ECG derivative pattern index (*edp*) and signal slope inversions over a minimum threshold (*inv*). The *edp* index was defined as

$$edp = \frac{mDs - mDn - mDhn + \epsilon}{mDs + mDn + mDhn + \epsilon} \quad (1)$$

where *mDs* is the average of the unsigned derivative maxima in consecutive 0.67-second windows, *mDn*, *mDhn* are the average of the unsigned derivative maxima in consecutive 0.067-second windows, and ϵ is a small constant. The derivative used for *mDs* and *mDn* was computed through the band-pass FIR filter with impulsive response $h_1 = [-1, 0, 1]$, while the derivative used for *mDhn* was computed through the high-pass FIR filter with impulse response $h_2 = [-1, 1]$. This noise index models the quasi-periodic nature of the ECG signal, characterized by a sequence of high-energy events.

The *inv* index was defined as the number of sign changes in the slope of ECG that are tied to non-negligible amplitude variations. Calculating this index requires the identification of all the time points t_n where the sign of the signal's first derivative changes:

$$tn : \text{sgn}\left(x_{t_n+1} - x_{t_n}\right) \neq \text{sgn}\left(x_{t_n} - x_{t_n-1}\right) \quad (2)$$

The difference in amplitude between each pair of consecutive tn points is then computed to form set D

$$D = \left\{x_{t_2} - x_{t_1}, x_{t_3} - x_{t_2}, \ldots, x_{t_N} - x_{t_N-1}\right\} \quad (3)$$

Neighboring elements of D are summed if their absolute value is lower than a set threshold M, set at 100 µV. *E.g.*, if $|x_{t_2} - x_{t_1}| < M$, D is updated to set D':

$$D' = \left\{x_{t_3} - x_{t_1}, \ldots, x_{t_N} - x_{t_N-1}\right\} \quad (4)$$

This process is repeated from the first element of the set until no amplitude difference is lower than threshold M, obtaining set D^*. The *inv* index is the size of D^*:

$$\text{inv} = |D^*| \quad (5)$$

Each noise index was computed for each of the 148502 records of the annotated database, resulting in a data matrix with 148502 rows by 12 columns.

2.3. Experimental Setup

Each noise index was first analyzed in an univariate setting, to investigate the ability of each measure to distinguish between clean ECG and noise. The strength of the relationship between each noise index and outcome **Y** was assessed with Cohen's d, investigating the difference between the means of the two groups ($Y_k = 0$ and $Y_k = 1$). Additionally, logistic regression (LR) between each index and signal quality outcome **Y** was fitted. The relative usefulness of each noise index relative to the others was also gauged in a multivariate setting, where a decision tree was fitted in 10-fold cross-validation. No data from the same human patient was put into different folds to avoid data leakage. The feature importance of each of the ten decision trees was adopted as a measure of the predictive power of each index when considering all indices together. To counter the effects of imbalanced data, different sample weights were used in model training. Noise records ($Y_k = 1$) were assigned a weight w_n, clean ECG records without ventricular arrhythmias ($Y_k = 0$, $V_k = 0$) were assigned a weight w_c, and clean ECG records with ventricular arrhythmias ($Y_k = 0$, $V_k = 1$) were assigned a weight w_v. These weights were defined so that $w_n n_n =$

$w_c n_c = w_v n_v$, where n_n, n_c, and n_v are the number of records of the three groups. This device ensured a high sample weight for clean ECG with ventricular arrhythmias records. In turn, this ensured a higher loss for misclassification of these records. Model training was carried out in MATLAB.

The performances of logistic regression and decision tree models were measured with confusion matrix metrics. In the confusion matrix noise instances were considered as positives, while clean ECG records were considered as negatives. Performances of logistic regression were measured with Matthew's correlation coefficient (MCC). Decision tree classification performances were evaluated with MCC, sensitivity (Se), and specificity (Sp). Additionally, specificity on records of clean ECG with ventricular arrhythmias (SpV) was also introduced to highlight model performance on the most critical subclass.

3. Results

3.1. Univariate Analysis

Table 1 shows Cohen's d values and the MCC of logistic regression fit for each noise index. Large effects of Cohen's d, *i.e.* $|d| > 0.8$, were found for only six of the twelve noise indices: *kur*, *msqi*, *snr*, *se*, *edp*, and *inv*. The MCC of logistic regression was higher for the *se*, *edp*, and *inv* indices, while other indices struggled with binary classification. These results suggest that some noise indices are more strongly related to the quality outcome than others, especially *se*, *edp*, and *inv*.

Index	Cohen's d	LR MCC
kur	0.806	0.041
skew	0.229	0.007
rpow	0.401	0.012
bas	0.094	0.000
ior	0.136	0.000
msqi	1.057	0.191
snr	-0.841	0.200
wave	0.616	0.279
hos	0.329	0.000
se	-1.345	0.431
edp	1.760	0.603
inv	-2.201	0.621

Table 1. Cohen's d and Matthew's correlation coefficient (MCC) of logistic regression (LR) fit for each noise index.

3.2. Multivariate Analysis

Figure 1 shows the average, minimum, and maximum feature importance for the classification trees trained on labeled data with 10-fold cross-validation. Notably, despite variation in the importance values, all classification trees favored five indices: *wave*, *hos*, *se*, *edp*, and *inv*. The information content of other indices appeared to be overshadowed by these five features. In Table 2 the average, minimum, and maximum MCC, Se, Sp, and Spv of the decision trees are shown. These metrics were computed on the ten test folds. Due to the high sample weights defined for clean ECG records with ventricular arrhythmias, each decision tree was biased toward lower sensitivity to noise (0.743 on average) and higher specificity on ventricular arrhythmias (0.952 on average). The wide variation in MCC is likely to be attributed to some test folds containing data that was not represented in the corresponding training folds.

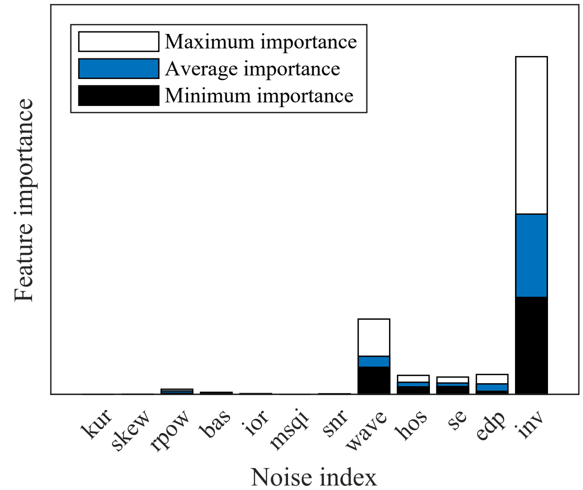

Figure 1. Feature importance of the ten decision trees trained on the labeled dataset with 10-fold cross-validation.

Metric	Average (min - max)
MCC	0.747 (0.545 - 0.887)
Se	0.743 (0.452 - 0.913)
Sp	0.965 (0.926 - 0.992)
SpV	0.952 (0.871 - 1.000)

Table 2. Average, minimum, and maximum test performances of the decision trees trained on the labeled data with 10-fold cross-validation.

4. Discussion and conclusions

Ambulatory ECG maintains a fundamental role in clinical practice, and its relevance for on-demand monitoring through commercial devices increases as wearable and portable ECG devices become popular. Thus, accurate and efficient ECG automatic analysis can benefit both patients and healthcare professionals. In this study, the ability of several noise indices to separate clean ECG from noise was evaluated on a database that included ventricular arrhythmias. Ventricular arrhythmias are cardiac events of utmost importance, which may entail severe risk of cardiovascular disease and sudden cardiac death.

The univariate analysis results show that some of the considered noise indices were not strongly related to the signal quality label. One such example is kurtosis, which, despite having a large effect size, reached a very low MCC in logistic regression. Thus, even with a large difference in the means of the clean ECG and the noise groups, the significant variability in the individual observations complicated binary classification. Kurtosis is one of the most popular ECG noise indices. However, in this ambulatory ECG that includes ventricular arrhythmias, *kur* appeared to be less predictive of the presence of noise.

When considering all noise indices together, five of them emerged as the most important for decision tree classification. Of note, *se*, *edp*, and *inv* showed both the highest feature importance as well as the largest effect in Cohen's d analysis. The *edp* and *inv* indices were introduced in this study, and these findings suggest that these two signal quality measures are promising and warrant further evaluation and development. Classification performances resulted in an average MCC of 0.743, average sensitivity of 0.744, average specificity of 0.965, and average specificity on ventricular arrhythmias of 0.952. The bias toward specificity is owed to the high sample weights assigned to ventricular arrhythmias as well as the heterogeneity of noise episodes. This bias is not undesired, as misclassification of ECG as noise should be avoided.

Further developments to accurate noise detection in ambulatory ECG would require more labeled data of both noise and pathologic ECG such as ventricular arrhythmias. Additionally, different classification models could reach higher performances, both feature-based and otherwise.

Acknowledgments

The authors thank Cardioline S.p.A. for supporting this study and granting exclusive access to the private ECG database.

References

[1] Sörnmo L, Laguna P. Chapter 7 - ECG Signal Processing. In Sörnmo L, Laguna P (eds.), Bioelectrical Signal Processing in Cardiac and Neurological Applications, Biomedical Engineering. Burlington: Academic Press. ISBN 978-0-12-437552-9, January 2005; 453–566.

[2] Clifford GD, Behar J, Li Q, Rezek I. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. Physiological Measurement August 2012;33(9):1419. ISSN 0967-3334. Publisher: IOP Publishing.

[3] van der Bijl K, Elgendi M, Menon C. Automatic ECG Quality Assessment Techniques: A Systematic Review. Diagnostics November 2022;12(11):2578. ISSN 2075-4418. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[4] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet. Circulation June 2000;101(23):e215–e220. Publisher: American Heart Association.

[5] Greenwald SD. The development and analysis of a ventricular fibrillation detector. PhD Thesis, Massachusetts Institute of Technology, 1986.

[6] Bachi L, Halvaei H, Pérez C, Martín-Yebra A, Petrėnas A, Sološenko A, Johnson L, Marozas V, Martínez JP, Pueyo E, Stridh M, Laguna P, Sörnmo L. ECG Modeling for Simulation of Arrhythmias in Time-Varying Conditions. IEEE Transactions on Biomedical Engineering December 2023; 70(12):3449–3460. ISSN 1558-2531. Conference Name: IEEE Transactions on Biomedical Engineering.

[7] Clifford GD, Azuaje, Francisco, McSharry, Patrick. ECG Statistics, Noise, Artifacts, and Missing Data. Advanced Methods and Tools for ECG Data Analysis 2006;6(1):55–99.

[8] Tobón V. DP, Falk TH, Maier M. MS-QI: A Modulation Spectrum-Based ECG Quality Index for Telehealth Applications. IEEE Transactions on Biomedical Engineering August 2016;63(8):1613–1622. ISSN 1558-2531. Conference Name: IEEE Transactions on Biomedical Engineering.

[9] Smital L, Haider CR, Vitek M, Leinveber P, Jurak P, Nemcova A, Smisek R, Marsanova L, Provaznik I, Felton CL, Gilbert BK, Holmes III DR. Real-Time Quality Assessment of Long-Term ECG Signals Recorded by Wearables in Free-Living Conditions. IEEE Transactions on Biomedical Engineering October 2020;67(10):2721–2734. ISSN 1558-2531. Conference Name: IEEE Transactions on Biomedical Engineering.

[10] Rahman S, Karmakar C, Natgunanathan I, Yearwood J, Palaniswami M. Robustness of electrocardiogram signal quality indices. Journal of The Royal Society Interface April 2022;19(189):20220012. Publisher: Royal Society.

[11] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology Heart and Circulatory Physiology June 2000;278(6):H2039–H2049. ISSN 0363-6135. Publisher: American Physiological Society.

Address for correspondence:

Lorenzo Bachi
Institute of Clinical Physiology, National Research Council of Italy (IFC-CNR), via Moruzzi 1, 56124, Pisa, Italy
lorenzobachi@cnr.it