Automatic Real-Time Quality Assessment of a 12-Lead ECG Recording

Reza Firoozabadi, Richard E Gregg, Beth Zengo, Saeed Babaeizadeh

Advanced Algorithm Research Center, Philips Healthcare, Andover, MA, USA

Abstract

In this paper, we present a novel algorithm to evaluate the quality of ECG recordings. Our algorithm is designed to help clinicians in rapid selection of good quality ECG segments from long recordings collected by an ECG monitoring device such as a 12-lead bedside monitor. With some adjustments, we used the Computing in Cardiology Challenge 2011 database in order to compare the performance of our algorithm to the published results. The challenge was aimed to develop near real-time algorithms in mobile phones and provide feedback on quality of the ECGs for interpretation to the users who are mostly laypersons with little knowledge of ECG interpretation. Our algorithm generates a noise score which is a combination of two parameters: a high-frequency noise measure which accounts for the muscle noise and other fast changing artifacts, and a baseline wander noise measure quantifying the low-frequency noise. The training dataset (set A) with reference quality assessments was used to determine an optimum threshold on the ROC curve for classification of acceptable and unacceptable segments. The algorithm was then evaluated on the test dataset (set B) with undisclosed annotations. Our method achieved maximum accuracy of 93.9% on the training dataset and an accuracy of 90.2% on the test dataset, placing itself among the top 10 performers who participated in the challenge.

1. Introduction

It is critical in electrocardiogram (ECG) recording and monitoring devices to evaluate the level of artifact on the waveforms. Accuracy of ECG reading depends heavily on the signal-to-noise ratio of the recording, especially for computerized ECG analysis. Therefore, there is a need to detect low-noise segments for the analysis and avoid segments with high noise levels. Furthermore, advanced mobile technology has made it feasible for almost everyone possessing a mobile phone to record and transmit the ECGs remotely for the analysis by cardiologists in large city medical centers [1]. Since ECG recording is likely to be performed by a non-expert layperson in rural areas, artifacts resulting from different sources and other issues may happen which need to be identified rapidly and prompt the operator to repeat the acquisition process while the patient is still available.

This paper addresses the issue of poor-quality ECG acquisition by providing an automatic algorithm which reflects the quality level in a single-number measure which can be used to either exclude the noisy segments from automated ECG analysis, or give rapid feedback to the user to repeat the ECG acquisition.

2. Methods

2.1. Database

In order to evaluate our algorithm performance, we used the CinC/Physionet challenge 2011 database which was collected by the Sana project and is freely available on PhysioNet website [1]. An introduction to the challenge and a summary of the results along with the discussion of their implications are provided by Silva et al [2].

The database contains standard 12-lead 10-second ECG recordings with diagnostic bandwidth (0.05 to 100Hz) which were collected at 500 samples-per-second and 5μV amplitude resolution.

The database was recorded by people with a varying range of expertise including nurses, technicians and volunteers with minimal experience to investigate the operation of laypersons. Recordings were manually annotated for signal quality as acceptable or unacceptable by a group of annotators with varying level of ECG analysis expertise.

The part of the database which was disclosed for the challenge includes 1500 10-second recordings which were split into training dataset (n=1000) for which the reference annotations are provided to participants and test dataset (n=500) where the annotations are withheld.

2.2. Algorithm

We define a noise score as a combination of high- and low-frequency (baseline wander) noise measures. The 10-second ECG recordings from the database are divided into 1-second 12-lead segments on which both noise measures are calculated. The total noise score in each segment is a combination of these two noise measures averaged on all 10 segments. The recordings with at least one disconnected lead (flat line) in a segment were considered unacceptable.

a. High-frequency noise measure

The high-frequency noise measure is estimated as the median of standard deviations of ECG at the short-term low activity intervals of the segment [3]. The activity function at each sample across all leads in an ECG segment is defined as:

$$A_i = \sum_{leads} (ECG_i - ECG_{i-1})^2 \qquad (1)$$

The values are then low-pass filtered and the local minima are detected by a short-term sliding window.

For each lead, the standard deviations of the samples are calculated in a window around the local minima:

$$\sigma_{lead} = \sqrt{\mathrm{E}[(ECG_{lead} - E[ECG_{lead}])^2]} \qquad (2)$$

where the operator E denotes the average or expected value.

The high frequency noise measure for each lead is the median of these standard deviations across the segment. The total high frequency noise measure in each segment is the average of the lead high frequency noise measures.

$$N_{HF} = \mathrm{E}[median(\sigma_{lead})] \qquad (3)$$

b. Low-frequency (baseline wander) noise measure

Baseline wander is a low-frequency noise identified by the changes to the segment baseline levels. We calculate the lead baseline in each 1-second segment by averaging the ECG segment samples. Baseline wander noise measure per lead is defined as the sum of absolute values of the baseline difference between the current segment and each of the two previous segments. The total baseline wander noise measure per segment is the average of all lead values:

$$N_{BW} = \mathrm{E}\big[|BW_{1,lead}| + |BW_{2,lead}|\big] \qquad (4)$$

where

$$BW_{1,lead} = \begin{cases} \mathrm{E}[ECG_k] - \mathrm{E}[ECG_{k-1}]\,, & k \neq 1 \\ 0\,, otherwise \end{cases} \qquad (5a)$$

$$BW_{2,lead} = \begin{cases} \mathrm{E}[ECG_k] - \mathrm{E}[ECG_{k-2}]\,, & k \neq 1,2 \\ 0\,, otherwise \end{cases} \qquad (5b)$$

c. Noise score

Noise score is defined as the empirically scaled sum of high-frequency and baseline wander noise measures in each 1-second segment, averaged on all leads:

$$Noise\ Score = \mathrm{E}[\alpha N_{HF} + \beta N_{BW}] \qquad (6)$$

This measure is compared against a classifier threshold to determine whether the quality of segment is acceptable or unacceptable. An ECG recording is identified acceptable only if all segments have a noise score lower than the classifier threshold.

Our algorithm was run on the training dataset and the outcomes were compared to the reference annotations. The ROC curve was plotted by changing the classifier threshold for the training dataset and the optimum threshold with maximum accuracy was identified. Figure 2 shows the ROC curve and the the maximum accuracy point.

With the selected classified threshold from training dataset, we evaluated the algorithm performance by entering the test dataset results in Physionet/CinC Challenge 2011 webpage for event 1.

3. Results

Table 1 summarizes the algorithm performance on training and test datasets. On the training dataset, our method achieved a maximum accuracy of 93.9%, with 84.9% sensitivity and 96.5% specificity. On the test dataset, the accuracy was 90.2% which is among the top ten scores in the challenge. Figures 3 and 4 illustrate examples of running the algorithm on recordings from training dataset. Figure 3 is an example of true positive case, where an acceptable annotated recording is correctly identified acceptable by the algorithm. Figure 4 illustrates an example of a true negative case, where the recording is identified unacceptable by both the challenge annotators and the algorithm.

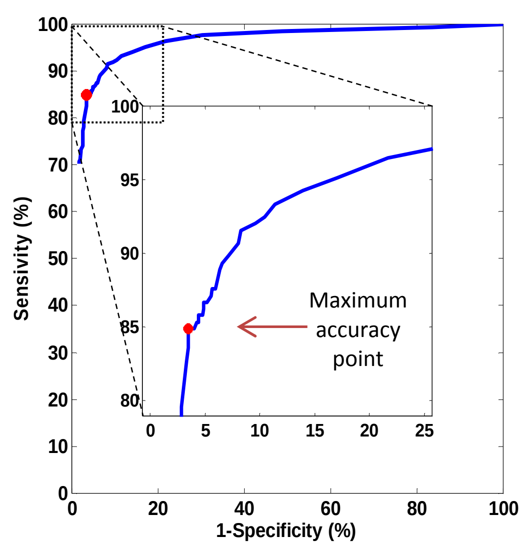

Figure 2. The ROC curve for the training dataset. The point of maximum accuracy, 93.9%, is marked by a dot. The area under curve (AUC) is 95.6%.

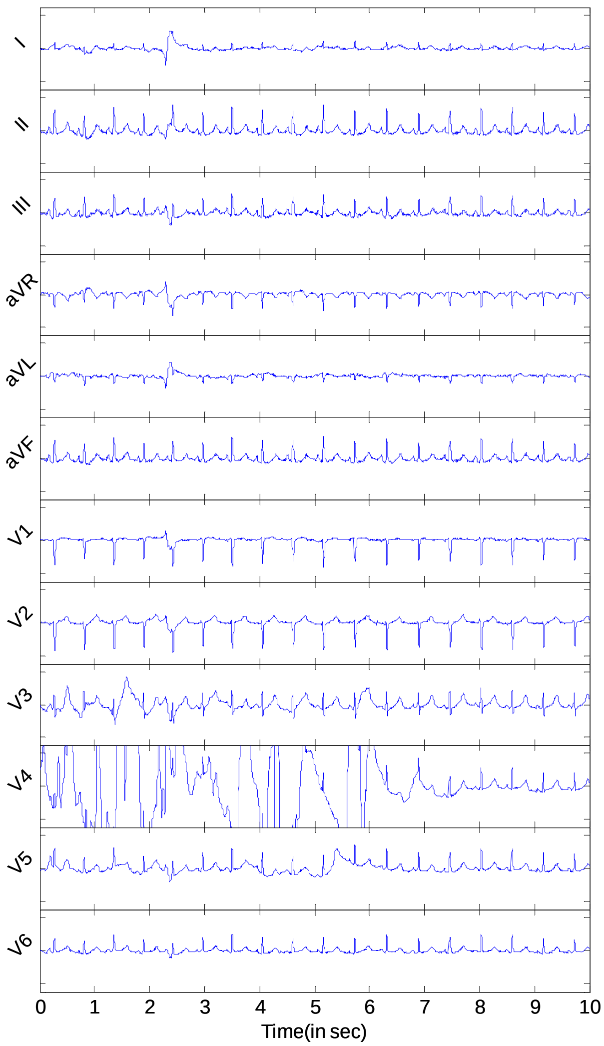

Figure 3(a)

High Frequency Noise

Lead	1	2	3	4	5	6	7	8	9	10
I	72	32	48	32	40	32	32	48	48	32
II	136	120	112	128	144	72	128	112	112	88
III	128	88	120	96	80	56	88	96	80	72
aVR	80	72	104	80	80	32	80	72	72	48
aVL	104	96	88	112	120	72	104	112	96	72
aVF	64	32	64	40	32	24	40	48	32	40
V1	40	32	32	40	32	24	16	24	16	24
V2	56	40	96	56	48	64	40	32	88	64
V3	192	272	136	96	112	104	72	88	104	120
V4	760	1672	712	704	800	808	232	96	72	80
V5	96	88	96	104	96	112	80	80	64	64
V6	64	72	64	64	56	48	64	56	48	48

Segment

Figure 3(b)

Baseline Wander Noise

Lead	1	2	3	4	5	6	7	8	9	10
I	0	0	36	32	25	24	10	11	4	4
II	0	15	27	22	16	13	7	9	11	7
III	0	14	14	11	9	10	3	2	7	3
aVR	0	7	31	26	19	19	9	10	9	6
aVL	0	14	14	7	6	2	2	2	8	4
aVF	0	6	22	20	17	17	6	8	3	1
V1	0	5	15	20	15	14	7	17	9	9
V2	0	14	14	24	18	16	21	30	19	13
V3	0	45	47	46	34	15	39	27	20	12
V4	0	129	351	275	701	302	370	252	149	54
V5	0	5	29	36	94	131	95	70	21	19
V6	0	6	18	12	11	5	0	2	9	10

Segment

Figure 3(c)

Figure 3. Example of a true positive acceptable ECG (a) 12-lead ECG, (b) high-frequency noise measure, (c) baseline wander noise measure.

The 12-lead ECG waveforms and per-segment per-lead noise measures are plotted in figures. Light gray and dark gray cells indicate moderate and high noise levels, respectively, and are both considered unacceptable by the algorithm.

4. Discussion and conclusions

The performance of our algorithm is among the top ten scores from the participants in the Physionet/CinC challenge 2011. Forty nine teams or persons participated in the challenge and developed algorithms to classify the ECGs using their quality measures in three defined events. In event 1 (closed source, open data set B), the top score was reported 93.2% by Xia et al [4] and followed by 9 more participants who scored 89.6% or more [2]. Our algorithm achieved an accuracy of 90.2% on the same dataset, ranking 9th. However, accuracy of our algorithm on training dataset was 93.9% which puts it in second place of reported scores.

Our algorithm was initially designed to flag visually poor-, moderate-, and high-quality segments in hours-long ECG recordings. The challenge annotation used a single threshold while our visual system graded quality in 3 levels. As a result, additional work was performed to adapt our algorithm to the application defined by the challenge. We reduced the quality levels to two, acceptable and unacceptable, as defined by the challenge.

Our ECG quality assessment algorithm has many applications including the one specified by the challenge to reject the poor quality ECGs in mobile devices, and also selecting the good quality segments in a time interval when the time scale is very compressed in the case of full disclosure. Full automation is also possible for nurse-charting request with 12-lead snapshots at any time.

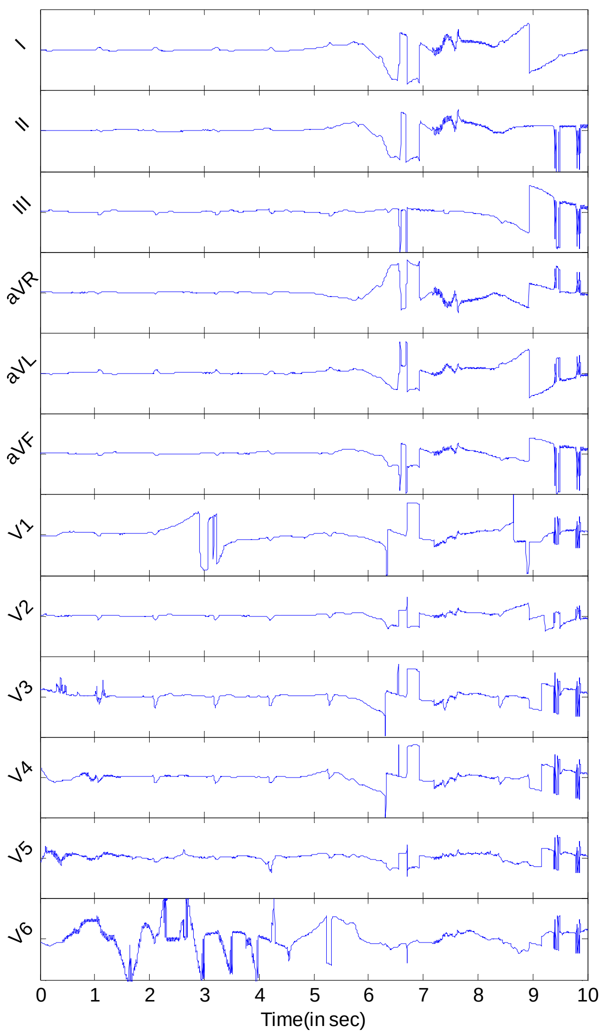

Figure 4(a)

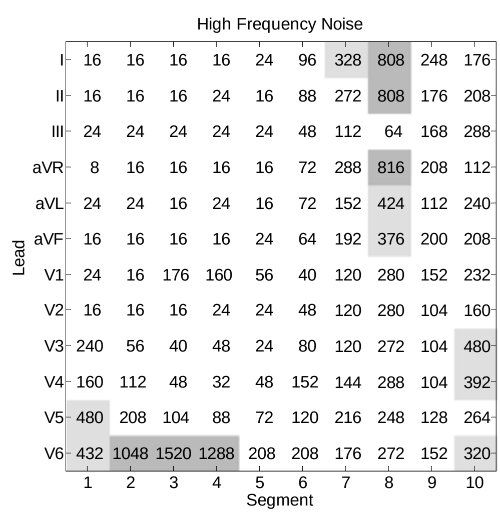

Figure 4(b)

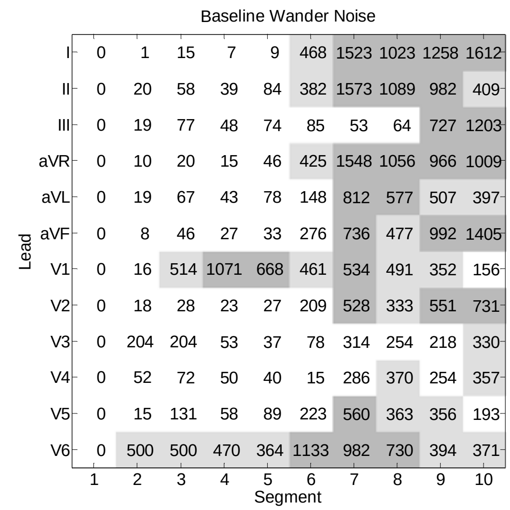

Figure 4(c)

Figure 4. Example of a true negative unacceptable ECG (a) 12-lead ECG, (b) high-frequency noise measure, (c) baseline wander noise measure.

Table 1. Summary of results on training and test datasets. Recordings are identified by the algorithm as either acceptable or unacceptable. Recordings containing flat line segments (disconnected leads) are also unacceptable. The results correspond to the selected classifier threshold. Individual annotations and hence the sensitivity and specificity are not available for the test dataset.

Dataset	Training	Test
Number of recordings	1000	500
Acceptable recordings	782	388
Recordings with flat line	168	89
Accuracy (%)	93.9	90.2
Sensitivity (%)	84.9	-
Specificity (%)	96.5	-

References

[1] PhysioNet/Computing in Cardiology Challenge 2011. http://www.physionet.org/challenge/2011

[2] Silva I, Moody G, Celi B. Improving the quality of ECGs collected using mobile phones: The Physionet/Computing in Cardiology Challenge 2011. Computing in Cardiology 2011;38:273−276.

[3] Gregg R, Lindauer J, Helfenbein E, Shao M, Zhou S. A new method for measuring high-frequency noise. Journal of Electrocardiology 2005;38:141.

[4] Xia H, McBride J, Sullivan A, Bock TD, Bains J, Worthham D, Zhao XA multistage computer test algorithm for improving the quality of ECGs. Computing in Cardiology 2011;38:369−372.

Address for correspondence.

Reza Firoozabadi: reza.firoozabadi@philips.com