Assessment of ECG Quality on an Android Platform

Lars Johannesen

Division of Pharmacometrics, Office of Clinical Pharmacology, Office of Translational Sciences, Center for Drug Evaluation and Research, FDA, Silver Spring, MD, USA

Abstract

An algorithm to determine quality of ECGs can enable inexperienced nurses and paramedics to record ECGs of sufficient diagnostic quality.

We propose an algorithm that can assess the quality of an ECG designed for an Android-based platform. The algorithm is based on previously established ECG quality metrics for quantifying ECG quality but designed in a way to make it efficient to run on a mobile platform.

Using the training data set the proposed algorithm obtained a sensitivity of 91% and a specificity of 85%. Testing against the test data sets, resulted in a score of 0.88 (events 1 and 2) and 0.79 (events 3).

The proposed algorithm discriminates between ECGs of good and bad quality, which could help diagnose patients earlier and reduce associated treatment costs.

1. Introduction

Cardiovascular diseases are the number one cause of death in the world according to the World Health Organization. The majority of these deaths take place in low- and middle-income countries. [1] The lack of adequate primary care results in cardiovascular disease being diagnosed late and follow-up treatments the patients cannot afford. In some low- and middle-income countries the large population spread combined with an increase in cardiovascular disease is a public health concern. To solve this problem a student-managed open-source mobile telemedicine group at MIT called Sana has joined forces with one of India's leading healthcare providers Narayana Hrudayalaya.

The goal of the collaboration is to take advantage of the growing number of mobile phones to overcome the problem with the large population spread, by allowing inexperienced nurses or paramedics to record an ECG in the field and transmit the recorded ECG to a cardiologist that can use the ECG to diagnose the patient. So far, Sana has successfully created the software for transmitting and receiving ECGs. In order to facilitate the ECG recording an algorithm for determining if the quality of the recorded ECG is sufficient is required.

Currently, there is no gold standard to assess ECG quality, but several ECG quality metrics have been proposed [2, 3]. Mainly, the different algorithms focus on different ways of assessing low- and high-frequency noise. These techniques have only been applied to ECGs from clinical studies, where the quality of the ECGs might differ compared to those recorded using a mobile phone.

We propose a rule-based algorithm based on the previously proposed quality metrics and evaluate it on noisy recordings provided in this year's Physionet challenge. As the secondary objective of this year's challenge is to implement the algorithm in Java, to enable it to run on an Android-based mobile platform, the algorithm will be designed and optimized to facilitate this.

2. Methods

2.1. Data

This year's Physionet challenge consists of three datasets, one training set and two test sets. All datasets contain standard ECG recordings, 10-second long, with 12 leads recorded simultaneously at 500 Hz with a 16-bit resolution. The signals contain a full diagnostic bandwidth of 0.05 Hz to 100 Hz. The data were recorded by nurses, technicians and volunteers with varying amounts of training. [4]

The training set contains 1000 recordings which are labeled as either acceptable, unacceptable or indeterminate. As there is no gold standard for assessing ECG quality the quality labeling is based on the individual labeling of 3 to 18 annotators working independently. Each ECG is given a score of 0.95, 0.85, 0.75, 0.60 or 0 as Excellent, Good, Adequate, Poor or Unacceptable quality respectively. Subsequently, the scores are averaged and if the average score is 0.70 or more, and at most one annotator graded the ECG as being of unacceptable quality, the ECG signal is labeled as acceptable. If the average grade was less than 0.70, it was labeled as unacceptable. Signals not falling into either category were labeled as indeterminate.

The first test set consists of 500 ECGs and the second test set consists of an unknown number of ECGs. The second test set is not released to the participants of the challenge. The performance of the algorithm is the fraction of ECGs classified correctly. The performance is evaluated either on the first test data set allowing any programming language and closed-source (event 1), open-source Java code (event 2) or on the second data set running on an Android reference phone (event 3). In event 3 the performance of the algorithm is also partly based on execution time.

2.2. Algorithm

At each step in the algorithm ECGs are grouped into two groups depending on a set of ECG features. The features are: lead-fail in all leads, global high frequency noise, leads with noise causing QRS detection problems, global low frequency noise and low and high frequency noise in the beats of sinus origin. The algorithm is shown in figure 1.

2.2.1. Lead fail

ECG recordings with lead fail in all leads are defined as noisy ECGs. Lead fail is defined as a constant derivative of zero for all samples of a lead.

2.2.2. Global high-frequency noise

The QRS detector, which is described in the following, is sensitive to high frequency noise such as power line noise; therefore ECGs were first filtered with a 10th order low pass FIR filter with a cutoff of $50\,\mathrm{Hz}$. If the average RMS across leads of the difference was above $0.5\,\mathrm{mV}$ the record was classified as a noisy record.

2.2.3. QRS detection

As no single lead can be assumed to be noise free, we detected QRS complexes in each lead independently, using the U3 algorithm [5]. U3 uses a detection threshold that is adjusted over time; as we only have $10\,\mathrm{s}$ available, another detection threshold was required. We defined the threshold as 0.7 of the RMS of the U3 of a given lead.

To reduce false positives only QRS complexes present in at least 6 leads were detected. Presence across leads was defined using a $90\,\mathrm{ms}$ time-window. Using the synchronized QRS complexes the average RR was defined, which allowed for determination of missed beats in a fashion similar to [6]. The second threshold was defined as 0.2 in order to trigger a detection of a QRS complex, which must be present in at least 6 leads.

Leads that contain less than $50\,\%$ of the synchronized QRS complexes are classified as noisy leads. If a record contains 3 or more noisy leads it is labeled as a noisy recording.

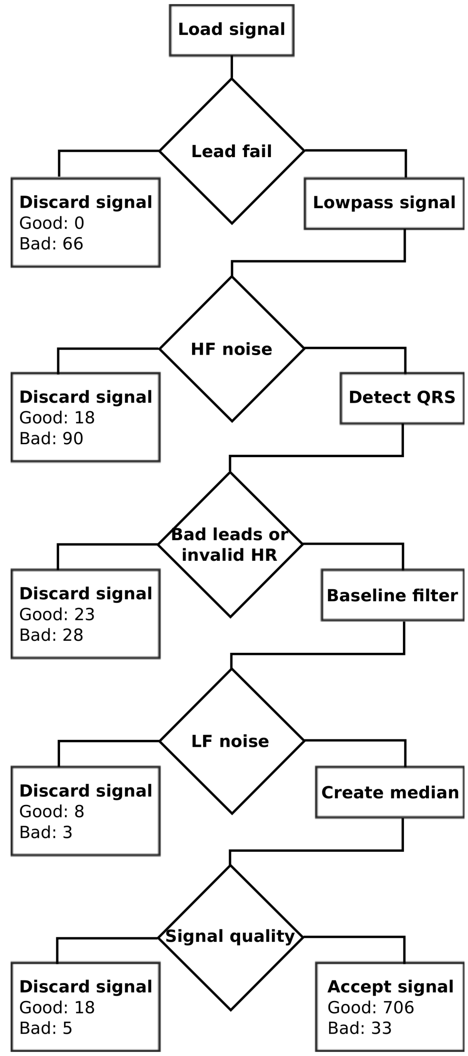

Figure 1. A flow-chart of the proposed algorithm. After each pre-processing step a rule is applied based on a heuristically determined threshold to determine if the recording is noisy. See text for a more in-depth explanation. The numbers indicate how many ECGs that were removed at a given step. Good indicates ECGs labeled with high quality and bad with low quality.

2.2.4. Global low-frequency noise

The global low-frequency noise is assessed by first detecting the onset of the QRS complexes. For each QRS detection in each lead the onset is defined as the first sample where the value is below 0.15 times the U3 value of the synchronized QRS complex [5]. If the onset is not defined within $70\,\mathrm{ms}$ of the R-peak the onset is defined as the minimal U3 value within the $70\,\mathrm{ms}$. The onset for the R-peak is the average of all onsets across leads.

The PQ point is assigned $10\,\mathrm{ms}$ to the left of the onset. For each lead the value in a window of $6\,\mathrm{ms}$ to either side of the PQ point is used together with the PQ-point to correct for baseline wander, using cubic spline interpolation [7]. The RMS of the cubic spline curve is the global low-frequency noise of a given lead [2]. If the average for all leads is more than $175\,\mu\mathrm{V}$, a recording is labeled as noisy.

2.2.5. Average beat quality

The low- and high-frequency is assessed by first calculating the average beat by lead and subsequently, subtracting this from beats of sinus origin, determined by the cross-correlation coefficient between the beat and the average beat. The result of this subtraction is then defined as the residuum, which is used to quantify noise [3]. The standard-deviation of the residuum defines the low-frequency noise and the RMS of the differentiated residuum defines the high-frequency noise. If the average across leads of either of these values are above $250\,\mu\mathrm{V}$ the recording is labeled as noisy.

2.3. Implementation

The algorithm was implemented in Java, using the Android API supplied by Physionet [4].

3. Results

3.1. Training

The results of applying the algorithm on the training data were assessed by the sensitivity and specificity, where sensitivity is defined as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

where TP is defined as the number of records the algorithm correctly determined as adequate quality or better, and FN is the number of records where the algorithm incorrectly determined the quality as being inadequate. Similarly, specificity is defined as:

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

where TN is defined as the number of records where the algorithm correctly determined the record as inadequate and FP as the number of records where the algorithm incorrectly determined the record as being of adequate quality.

On the training data a sensitivity of $91\,\%$ sensitivity and $86\,\%$ specificity were obtained.

3.2. Test

In events 1 and 2 the proposed algorithm obtained a score of 0.88, and in event a score of 0.79.

4. Discussion and conclusions

Accurate determination of quality and guidance to inexperienced nurses or paramedics can help diagnose patients at risk for developing cardiovascular diseases earlier which can reduce associated costs and therefore make treatment options more affordable.

The results shown in this paper indicate that by using a simple rule-based approach, it is possible to determine the quality of ECGs which can help getting high quality ECGs to the cardiologist earlier. However, the results of applying the heuristically determined thresholds to the test data sets indicate that the thresholds might have been over-fitted to the training data. The scoring of event 3 does however include execution time, but as the algorithm had a low execution time on the author's own android mobile phone (less than $500\,\mathrm{ms}$), it is more likely due to the thresholds.

It is however important to recognize that there is no gold standard for determining the quality of ECGs and simply estimating quality by looking at high and low-frequency noise and how many of the leads are usable is a reasonable technique worthwhile exploring.

Additionally, another trait of the U3 method for detection is that it can be used to estimate the onset and offset of the QRS complex, which we used for estimating and correcting baseline wander. However, one could also use the offset to calculate the QRS angle and use the value of the angle and integral value for the QRS complex by lead to determine if there is a lead interchange.

The last three evaluated quality values for a signal, baseline, low- and high-frequency noise could be used in a continuous, rather than binary, ECG quality score; this could be used to guide the end-user in how good the quality of the ECG is and help eliminate the potential problems associated with fixed thresholds.

Overall, methods such as the one proposed in this paper can be useful to increase the quality of recorded ECGs and help diagnose cardiovascular diseases earlier and over time decrease deaths caused by cardiovascular diseases.

Acknowledgements

The author would like to thank Jens Stampe Sørensen and Dr. Norman Stockbridge for scientific collaboration and discussion pertaining to the manuscript, and Dr. Loriano Galeotti for ideas and discussion about the methodology.

Disclaimer

The paper is a result of independent research and does not represent the findings of the U.S. Food and Drug Administration.

References

[1] World Health Organization. Cardiovascular diseases. http://www.who.int/cardiovascular_diseases/en/.

[2] Vaglio M, Isola L, Gates G, et al. Use of ECG Quality Metrics in Clinical Trials. Comput Cardiol 2010;37:505–508.

[3] Batchvarov V, Hnatkova K, Malik M. Assessment of Noise in Digital Electrocardiograms. PACE 2003;25(4):499–503.

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

[5] Paoletti M, Marchesi C. Discovering dangerous patterns in long-term ambulatory ECG recordings using a fast QRS detection algorithm and explorative data analysis. Comput Methods Programs Biomed 2006;82(1):20–30.

[6] Marchesi C, Paoletti M. ECG Processing Algorithms for Portable Monitoring Units. Internet J Med Tech 2004;1(2).

[7] Meyer C, Keiser H. Electrocardiogram baseline noise estimation and removal using cubic splines and state-space computation techniques. Comput Biomed Res 1977;10(5):459–470.

Address for correspondence:

Lars Johannesen
10903 New Hampshire Avenue
Building 51
Silver Spring, Maryland, 20993
USA
lars.johannesen@fda.hhs.gov