Automated RR Interval Detection and Quality Assessment in Telehealth Electrocardiograms

Sharon Yuen Shan Ho[1], Florian Kristof[2], Jonathan Mant[1], Peter H Charlton[1]

[1] University of Cambridge, Cambridge, United Kingdom
[2] University of California, Berkeley, United States

Abstract

Mobile devices like smartwatches and handheld ECG recorders which can record single-lead ECGs which could be used to detect atrial fibrillation. A key step in using these ECGs is extracting RR-intervals. We aimed to develop an algorithm to automatically: (i) derive RR-intervals from ECGs; and (ii) predict whether they are accurate enough for diagnosis. The TELE ECG Database was used, containing 250 ECGs recorded using a handheld ECG device with manual QRS annotations. An algorithm was designed to: (i) detect QRS complexes and extract RR-intervals using a primary QRS detector; and (ii) assess the reliability of each RR-interval through comparison with QRS complexes detected by a secondary QRS detector. Performance was assessed using each combination of three primary and 18 secondary open-source QRS detection algorithms. The algorithm with the lowest mean absolute error (MAE) in RR-intervals deemed to be reliable used 'unsw' (primary) and 'rpeak' (secondary) detectors, achieving a mean MAE of 24ms with 33% of RR-intervals deemed reliable. The algorithm with the lowest mean MAE, whilst deeming at least 90% of RR-intervals reliable, used 'unsw' (primary) and 'nk' (secondary) detectors, achieving a mean MAE of 33ms with 92% of RR-intervals deemed reliable. These algorithms accurately derive RR-intervals from telehealth ECGs.

1. Introduction

Atrial fibrillation (AF) is a common heart arrhythmia which often goes undiagnosed, increasing stroke risk. Mobile health devices like smartwatches and handheld ECG recorders can record single-lead ECGs, and could be used to screen for AF. A key step in using these ECGs to detect AF is to extract RR-intervals to determine whether the heart rhythm is regular or irregular. However, ECGs from mobile health devices are often of low quality. Therefore, robust signal processing algorithms are required to identify periods which are of sufficient quality for analysis, and to accurately extract RR-intervals from them.

Several approaches have been proposed to assess ECG signal quality [1]. Typically, these consist of extracting features from an ECG (*e.g.* features measured from fiducial points or statistical features), and using these to classify ECGs according to quality (using heuristic rules or machine learning methods) [1]. One approach that appears particularly promising for determining whether RR-intervals can be accurately extracted from ECG signals is the *bSQI* approach [2–4]. This involves using two separate QRS detection algorithms to detect QRS complexes, and assessing the signal quality from the proportion of time for which the two detectors are in agreement. A previous study [5] generalised this approach by using two or more QRS detectors, with best performance obtained by using four QRS detectors (which we refer to as *u3*, *unsw*, *dom*, and *two-avg* where *two-avg* corresponds to 'OKB' in the original publication). The current study investigates *bSQI* methods using two QRS detectors. It builds on previous work by: (i) assessing the suitability of *bSQI* methods for RR-interval extraction; (ii) including recently developed QRS detection algorithms (*e.g.* the *neurokit* beat detector [6]); and (iii) using open-source algorithms [7].

The aim of this study was to develop an open-source algorithm to automatically: (i) derive RR-intervals from telehealth ECGs; and (ii) predict whether the derived RR-intervals are sufficiently accurate to inform diagnosis.

2. Methods

2.1. Dataset

We used the publicly available TELE ECG Database, containing 250 30-second single lead ECG recordings from home-dwelling patients with chronic obstructive pulmonary disease and/or congestive heart failure [8, 9]. As described in [7], ECGs were acquired by patients using the TeleMedCare Health Monitor (TeleMedCare Pty. Ltd. Sydney, Australia), which records an ECG from the hands using dry metal electrodes [8, 9]. The dataset includes 221 ECGs from 120 patients, and a further 29 specifically poor-

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.084

quality ECGs. It also contains manual annotations of QRS complexes, agreed by three annotators [9]. One ECG was truncated to 30s for this study [7].

2.2. QRS Detection

A total of 18 open-source QRS detection algorithms were used in this study. These consisted of 16 of the 18 algorithms used in [7] (we excluded the *gamb* and *mart* algorithms as they previously performed particularly poorly). In addition, two further QRS detectors were implemented for this study (*u3* [10] and *dom* [11]), ensuring that all the top performing QRS detectors reported in [5] were included in this study. In summary, the QRS detectors used in this study were: *christ*, *dom*, *engz*, *fnvg*, *fwhvg*, *gqrs*, *hamilt*, *jqrs*, *kali*, *nab*, *nk*, *pan-tomp*, *rdeco*, *rpeak*, *two-avg*, *u3*, *unsw*, and *wqrs*. The reader is referred to [7] for descriptions of the 16 QRS detectors used previously.

2.3. RR-Interval Extraction and Quality Assessment

RR-intervals were calculated as the time delays between consecutive QRS complexes detected using a high-performance, ‘primary’ QRS detector. The quality of the ECG signal was assessed for each RR-interval using a modified version of the *bSQI* approach, as follows. A ‘secondary’ QRS detector was used to detect QRS complexes. Then, the locations of QRS complexes detected by the primary and secondary QRS detector were compared. Each QRS complex detected by the primary QRS detector was deemed to be accurately detected if the secondary QRS detector detected a single QRS complex within ± 150ms of it. An RR-interval was considered high-quality if the QRS complexes marking its start and end were both deemed accurate. Otherwise, it was considered low-quality.

In this study, the three most accurate QRS detectors identified in [7] were used as primary QRS detectors: *nk*, *unsw*, and *two-avg*. All 18 QRS detectors were used as secondary detectors. Every combination of primary and secondary QRS detectors was assessed. Results are also provided when using the same algorithm as both primary and secondary QRS detectors, corresponding to using one QRS detector with no quality assessment.

2.4. Performance Evaluation

The evaluation was performed as follows. The algorithm-detected and reference RR-intervals, and the quality of the detected RR intervals, were linearly interpolated to obtain two uniformly sampled time series at 500 Hz. The time series were truncated to have the same start and end times by only considering the region of overlap between the detected and reference RR-interval time series. The proportion of time for which RR-intervals were classified as high-quality was calculated for each ECG. The mean MAE between detected and reference RR-intervals was calculated from the errors at each point in time between the interpolated time series for each ECG, in units of ms. Periods of RR-intervals classified as low-quality were excluded from this calculation. These results were summarised for each pair of QRS detectors as the mean proportion and mean MAE across all 250 ECGs.

We used two approaches to select the best-performing pair of QRS detectors. First, we selected the pair with the lowest mean MAE - this could be suitable for applications requiring highly accurate RR-intervals where a high proportion of data can be discarded. Second, we selected the pair with the lowest mean MAE with a mean proportion of time for which RR-intervals were classified as high-quality of ≥90% - this could be suitable for applications requiring accurate RR-intervals with only limited data available.

3. Results

The results for the mean MAE and mean proportion of time for which RR-intervals were classified as high-quality are presented in Figures 1 and 2 respectively.

Figure 1 shows that there was a wide range of mean MAEs for different pairs of QRS detectors, ranging from 24ms to 99ms. Furthermore, it shows that the RR-intervals were generally most accurate when using *unsw* as the primary QRS detector, followed by *nk*, then *two-avg* (this was the case for 13 out of 18 secondary QRS detectors).

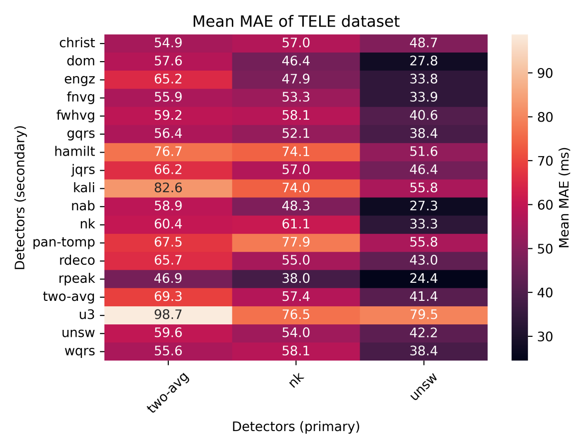

Figure 1. Heatmap showing the mean MAEs for all pairs of QRS detectors.

Considering our first approach to selecting the best-performing pair of QRS detectors: the most accurate RR-intervals were obtained when using [*unsw*, *rpeak*] as [primary, secondary] QRS detectors, with a mean MAE of 24ms. Whilst this pair provided the lowest mean MAE, it

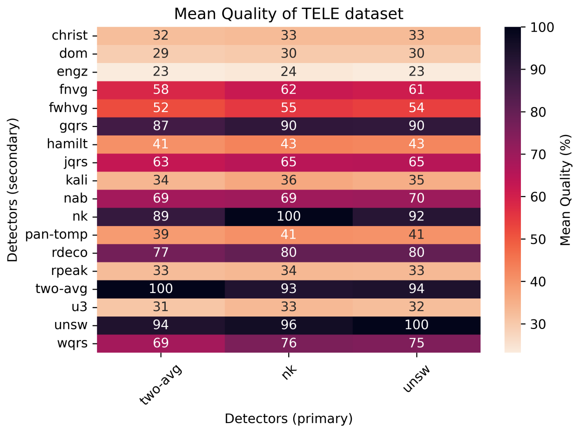

Figure 2. Heatmap showing the mean proportion of time for which RR-intervals were classified as high-quality for all pairs of QRS detectors.

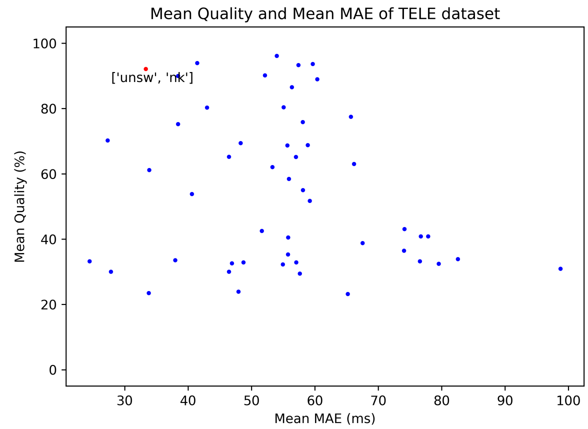

Figure 3. Scatter plot showing the mean proportion of time for which RR-intervals were classified as high-quality, against the mean MAE, for all pairs of different QRS detectors. The only pair with mean MAE < 40 ms and quality $> 90\%$ is marked in red.

only classed 33% of RR-intervals as high-quality. This was one of nine pairs which achieved mean MAEs of <40ms (eight of which used *unsw* as the primary QRS detector).

The results in Figure 1 also show the benefit of using two QRS detectors for quality assessment, rather than using one QRS detector alone to extract RR-intervals with subsequent quality assessment. For each primary QRS detector, several pairs of different QRS detectors achieved lower mean MAEs (*i.e.* including the quality assessment step) than when using the same QRS detector as both primary and secondary (*i.e.* without quality assessment).

Figure 2 shows that the mean proportion of time for which RR-intervals were classified as high-quality was relatively low for most QRS detector pairs (*i.e* <90%). Only seven of 51 pairs classified $\geq 90\%$ of ECG signals as high-quality (disregarding pairs using the same primary and secondary QRS detector). Five of these used two of the three QRS detectors previously identified as the best-performing detectors for telehealth data (*two-avg*, *nk*, and *unsw*), and the remainder used the *gqrs* detector. The [*nk*, *unsw*] pair provided the highest mean proportion of time for which RR-intervals were classified as high-quality (96%).

Figure 3 is a plot of the mean proportion of time for which RR-intervals were classified as high-quality, against the mean MAE. Each point represents a pair of different QRS detectors. Considering our second approach to selecting the best-performing pair of QRS detectors: the most suitable pairs are those in the top left corner of the scatter plot, *i.e.* which classified a high proportion of RR-intervals as high-quality, and achieved a relatively low mean MAE. Based on our requirement of providing $\geq$ 90% high-quality RR-intervals, [*unsw*, *nk*] was the best-performing pair, classifying RR-intervals as high-quality for 92% of the time, and achieving a mean MAE 33ms.

4. Discussion

4.1. Summary of Findings

This study shows that it is possible to use pairs of peak detectors to derive RR-intervals from telehealth ECGs, and to predict whether the RR-intervals obtained are accurate. The pair of QRS detectors which produced the lowest mean MAE was [*unsw*, *rpeak*], which achieved a mean MAE of 24ms, which deemed a low proportion (33%) of RR-intervals to be high-quality. This may have utility in applications such as AF screening where multiple ECGs are acquired from each patient and it is possible to discard high proportions of data. For applications that require RR-intervals to be obtained for a high proportion of the time, [*unsw*, *nk*] was identified as the best pair, which achieved a mean MAE of 33 ms and classified 92% of RR-intervals as high-quality. This may have utility in applications where only a single ECG is obtained from the patient, and we would like to obtain accurate RR-intervals in a high proportion of ECGs to inform clinical decision making.

4.2. Comparison with Existing Literature

This study builds on previous work which has: (i) assessed different combinations of QRS detectors for assessing ECG signal quality [5]; and (ii) assessed the accuracy of QRS detectors [7]. The current study investigated applying the *bSQI* signal quality assessment approach to RR-interval extraction, which to our knowledge has not been reported before. We believe our finding that the [*unsw*, *nk*] pair performed best is novel since previous work on using combinations of QRS detectors for quality assessment has not included the *nk* QRS detector [5].

4.3. Strengths and Limitations

The key strength of this study is that it assessed a large set of 18 open-source QRS detection algorithms, including the recently developed *nk* algorithm. Furthermore, the data and code used in this study are publicly available, allowing others to reproduce the results and build on the work.

However, this study only used a single dataset, which was used to develop the *unsw* QRS detector, so it is not clear whether the results are generalisable. The dataset only contains QRS annotations after resolution of disagreements between annotators, so we could not assess agreement between annotators. In addition, the *unsw* algorithm is currently implemented in MATLAB, whereas *nk* is in Python, which may pose difficulty for utilisation.

4.4. Implications

The algorithms developed in this study may be useful for extracting RR-intervals from telehealth ECGs, such as those acquired in AF screening [12]. In the future, it would be helpful to assess their performance on other datasets, in the presence of arrhythmias, and potentially with ECG signals acquired using wearable as opposed to hand-held devices. Future studies could also investigate the execution time of algorithms to understand how practical they are for use in mobile or wearable devices [7]. Finally, if the *unsw* QRS detector performs well on other datasets, then it could be beneficial to implement it in Python.

5. Conclusion

A simple algorithm using two QRS detectors can derive RR-intervals from telehealth ECGs, and predict whether the RR-intervals are accurate. The [*unsw*, *nk*] pair provided accurate RR-intervals for most ECGs collected by home-dwelling patients using a hand-held device.

Acknowledgments

This study is funded by the British Heart Foundation [FS/20/20/34626]. For the purpose of open access, the author(s) has applied a Creative Commons Attribution (CC BY) license to any Accepted Manuscript version arising.

References

[1] Satija U, Ramkumar B, Sabarimalai Manikandan M. A review of signal processing techniques for electrocardiogram signal quality assessment. IEEE Reviews in Biomedical Engineering 2018;11:36–52.

[2] Li Q, Mark RG, Clifford GD. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a kalman filter. Physiological Measurement 2008;29(1):15–32.

[3] Johnson AE, Behar J, Andreotti F, Clifford GD, Oster J. R-peak estimation using multimodal lead switching. In Proc CinC, volume 41. 2014; 281–284.

[4] Behar J, Johnson A, Clifford GD, Oster J. A comparison of single channel fetal ECG extraction methods. Annals of Biomedical Engineering 2014;42(6):1340–1353.

[5] Liu F, Liu C, Zhao L, Jiang X, Zhang Z, Li J, Wei S, Zhang Y. Dynamic ECG signal quality evaluation based on the generalized bSQI index. IEEE Access 2018;6:41892–41902.

[6] Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SHA. NeuroKit2: A Python toolbox for neurophysiological signal processing. Behavior Research Methods 2021;53(4):1689–1696.

[7] Kristof F, Kapsecker M, Nissen L, Brimicombe J, Cowie MR, Ding Z, Dymond A, Jonas SM, Lindén HC, Lip GYH, Williams K, Mant J, Charlton PH. QRS detection in single-lead, telehealth electrocardiogram signals: Benchmarking open-source algorithms. PLOS Digital Health 2024;3(8):e0000538.

[8] Redmond SJ, Xie Y, Chang D, Basilakis J, Lovell NH. Electrocardiogram signal quality measures for unsupervised telehealth environments. Physiological Measurement 2012; 33(9):1517–1533.

[9] Khamis H, Weiss R, Xie Y, Chang CW, Lovell NH, Redmond SJ. QRS detection algorithm for telehealth electrocardiogram recordings. IEEE Transactions on Biomedical Engineering 2016;63(7):1377–1388.

[10] 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.

[11] Yeh YC, Wang WJ. QRS complexes detection for ECG signal: the difference operation method. Comput Methods Programs Biomed 2008;91(3):245–254.

[12] Mant J, Modi RN, Charlton P, Dymond A, Massou E, Brimicombe J, Freedman B, Griffin SJ, Hobbs FDR, Lip GYH, McManus RJ, Williams K. The feasibility of population screening for paroxysmal atrial fibrillation using handheld ECGs. EP Europace 2024;euae056.

Address for correspondence:

Peter H Charlton
Department of Public Health and Primary Care, University of Cambridge; Strangeways Research Laboratory; 2 Worts' Causeway; Cambridge, CB1 8RN, UK.
pc657@cam.ac.uk