Accurate Identification of Actionable ECG Data Using a Signal Quality Assessment Algorithm

Jordan Diven[1], Rebecca Hand[1], Jonathan Francey[1], Holly Easlea[1], Andrew Miller[1], David Burke[2]
[1]B-Secur Ltd, City Quays, Belfast, N. Ireland
[2]Beacon Hospital, Dublin, Ireland

Abstract

Accurate identification of high-quality ECG data is crucial for reliable cardiac rhythm assessment, particularly in ambulatory settings. This study presents a machine learning-based signal quality assessment algorithm designed to address the challenges of ECG signal classification in diverse and real-world environments. The algorithm prioritises sensitivity, effectively identifying actionable ECG segments while minimising the risk of discarding critical clinical data. Trained and validated on a wide range of datasets, including publicly available and proprietary ECG recordings, the algorithm maintains a consistently high sensitivity (>98%) across various device types, patient demographics, and recording environments. Our results highlight the algorithm's utility as a generalisable signal quality triage tool, with the potential to reduce burden on clinicians and enhance the efficiency of ECG analysis in clinical practice.

1. Introduction

Single-lead ECGs, utilised in Holter monitors, wearable patches, and handheld devices, are favoured for cardiac rhythm assessment in the ambulatory setting due to their portability and convenience [1]. However, persistent patient movement and environmental interference can lead to exacerbated noise in out-of-hospital ECG monitoring, degrading signal quality and making it difficult to identify the variations in waveform timing and sequence that are indicative of arrhythmia. As a result, clinicians must manually scan ECG recordings to locate segments of sufficient quality to reliably inform clinical decisions; a laborious task that adds substantial burden to the diagnostic pathway. In cases where high-quality data does not coincide with the timing of paroxysmal events, interpreters face a difficult scenario where they must either diagnose based on poor-quality data, risking diagnostic errors, or disregard the data, potentially leading to missed or delayed diagnoses.

Traditionally, ECG noise has been addressed by employing various filters across multiple time and/or frequency domains, with more recent methods incorporating adaptive, data-driven filtering approaches [2]. While denoising strategies have proven effective at recovering data even in the presence of moderately high noise levels, significant proportions of ambulatory recordings are often so severely corrupted that they cannot be restored without significantly distorting the signal's morphology. The indiscriminate application of filters to entire ambulatory recordings without clarifying the quality of the input signal is therefore problematic as distorted signals can lead to interpretive error for both clinicians and automated ECG software [3,4].

An alternative approach to manage noise-corrupted data is to supplement ECG denoising with signal quality (SQ) assessment algorithms that can detect and triage regions with sufficient quality for subsequent analysis. This approach not only reduces the review burden of low-quality data but also minimises the risk of erroneous interpretations that could arise from noisy signals. The development of ECG SQ assessment methods gained considerable momentum following the *PhysioNet Computing in Cardiology (CinC) Challenge 2011*, which focused on developing real-time quality assessment methods for mobile applications [5]. Research in this area has advanced considerably in recent years, exploring a variety of parameters, including statistical, morphological, nonlinear, or time-frequency domain features, among others, to serve as indicators of signal quality [6]. Early classifier models primarily relied on rule-based systems, which segmented signal quality based on predefined thresholds. However, more recently, machine learning (ML) techniques have introduced more sophisticated, data-driven models that possess superior performance, particularly in the presence of noise [7].

While these advancements are promising, the efficacy of ML-based SQ algorithms is heavily dependent on the diversity of the datasets used in development. Many studies have employed training datasets with limited size, often derived from homogeneous populations or controlled environments, which limits the generalisability of these models [8]. Such limited datasets fail to capture the variability in ECG signals due to differences in demographics, pathology, and recording environments, leading to models that perform well in controlled settings but struggle when applied to real-world populations.

In this study, we present an ML-based SQ algorithm specifically designed to address the challenges of identifying actionable ECG data in diverse and noisy ambulatory settings.

2. Methods

2.1. ECG Pre-Processing

Signals were processed with HeartKey®, a cloud-based ECG processing platform that employs a series of iterative, logic-based digital filters for denoising, including: a mains subtraction filter with adaptive harmonic estimation to cancel interference at 50/60 Hz, a low-pass filter to remove noise above the standard 40 Hz ambulatory cut-off, and a baseline and smoothing filter featuring dynamic components to address non-stationary noise interference.

2.2. Feature Extraction

The model utilises a total of 26 statistical and spectral features derived from both raw and denoised ECG signals for quality assessment. The statistical measures include the mean, mean absolute deviation, median absolute deviation, standard deviation, kurtosis, skewness, and the skewness-kurtosis ratio, among others [9]. In the frequency domain, the model analyses power distribution across six characteristic frequency bands: 0-1 Hz (baseline wander), 1-5 Hz (low-frequency motion or P/T waves), 5-15 Hz (QRS complex and P/T waves), 15-25 Hz (QRS complex), 25-40 Hz (high-frequency muscle noise), and 40-100 Hz (high-frequency noise and powerline interference).

2.3. Classification

The extracted features are provided to an ML model that utilises a Bagged Classification Tree method, consisting of 30 trees with a maximum of 50,000 splits per tree. The model employs feature extraction and comparative statistical analysis of both raw and processed signals to classify each consecutive two-second ECG segment as either *high* or *low* signal quality (Figure **1**).

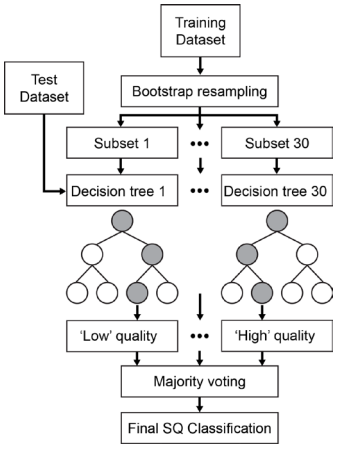

Figure **1**. ML signal quality classification model using Bagged Classification Tree method.

2.4. Training Datasets

The model was trained using both statistical and spectral features on: i) several publicly available Physionet datasets (BRNO-QDB, BRNO P-wave, LUDB, MIT-BIH NST, MIT-BIH Long); ii) a number of proprietary ECG databases acquired on a variety of medical (Holter, patch) and consumer (smartwatch, handheld, chest strap) devices, and iii) single-lead, dry electrode records from the PhysioNet/CinC Challenge 2017 [10] corrupted with various levels of synthetic ECG noise (i.e., baseline wander, electrode motion, muscle artefact and powerline interference). To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to ensure an equal number of samples within each class [11].

2.5. Validation Databases

Algorithm performance was evaluated using three diverse datasets, representing a range of devices, lead types, cardiac pathologies, and varying intensities and combinations of ECG noise: i) leads I-III of the PhysioNet Computing in Cardiology (CinC) 2011 Challenge dataset (998 x 10 s files), ii) a proprietary database of lead III Holter ECGs with known periods of leads ON and OFF (80 x 10 min files), and iii) a database of modified lead III (MLIII) and sternum-lead recordings collected by the Beacon Hospital (80 x 5-30 min files) (Table **1**).

Table **1**. Validation datasets overview.

Dataset	Leads	Patients (Files)	Total duration	Low quality (%)
PCINC 2011	Leads I-III	998 (998)	166 min	13.7
Holter Leads ON/OFF	Lead III	17 (30)	245 min	45.1
Beacon Hospital	MLIII/ Sternum	70 (80)	800 min	2.5
Total		1,085 (1,108)	1,211 mins	20.4

2.6. Performance Evaluation

Reference quality annotations were manually generated for signals from each individual lead in two-second segments on the PhysioNet/CinC (PCINC) 2011 dataset, and continuously for the Holter Leads ON/OFF and Beacon Hospital datasets, by two ECG analysts, with

discrepancies resolved by a third, more-experienced analyst, using the following criteria: *Low* signal quality: QRS complexes cannot be detected reliably. *High* signal quality: All QRS complexes are clearly identifiable. An example of the signal quality classifications provided by HeartKey is shown in Figure **2**.

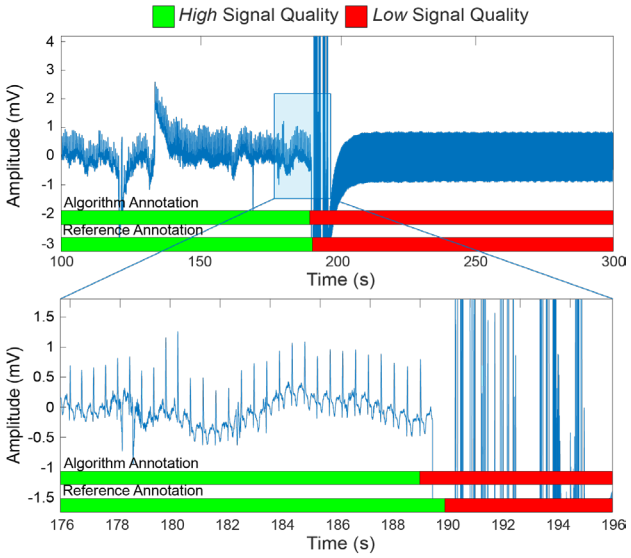

Figure **2**. An example of HeartKey Signal Quality classification relative to manual reference annotation on a signal from the Holter Leads ON/OFF database.

The algorithm's ability to correctly classify *high* signal quality was assessed using duration sensitivity (*Se*) and specificity (*Sp*) metrics, calculated relative to reference annotations. Duration statistics were chosen over episode statistics because they more effectively reflect the length of an ECG recording as being clinically actionable. This is crucial when using SQ algorithms as a triage tool on extended ambulatory recordings given that the duration of a single noise event can vary significantly.

$$Duration\ Se\ = \frac{(TP\ Duration)}{(FP\ Duration) + (TP\ Duration)}$$

$$Duration\ Sp\ = \frac{(TN\ Duration)}{(FP\ Duration) + (TN\ Duration)}$$

Where true positive (TP) = correct classification of a *high* quality segment; false positive (FP) = incorrect classification a *low* quality segment; true negative (TN) = correct classification of a *low* quality segment; false negative (FN) = incorrect classification of *high* quality segment.

3. Results

The *high* signal quality classification performance results are shown in Table **2**. The HeartKey SQ algorithm maintained a *high* signal quality duration *Se* >98% on the three test databases. A mean duration *Sp* of 79% was achieved across leads I-III of the PhysioNet CinC Challenge 2011 database, and 92% on the Holter Leads ON/OFF database. A lower *Sp* of 53% was obtained on the Beacon Hospital database, which was attributed to the low proportion of *low* quality data within the dataset (2.5%).

Table **2**. *High* signal quality classification performance on validation databases.

Dataset	Duration *Se* (%)	Duration *Sp* (%)
PCINC 2011 (Lead I)	99	82
PCINC 2011 (Lead II)	98	83
PCINC 2011 (Lead III)	98	72
Holter Leads ON/OFF	98	92
Beacon Hospital	99	53
Average	98.4	76.4

A consistently high duration *Se* ensured that instances where the algorithm erroneously classified *high* quality data as *low* signal quality were minimal. Analysis showed that most instances of *high* quality misclassifications were due to short bursts of noise. Figure **3** shows that in the Holter Leads ON/OFF and MLIII/Patch databases, 63.1% of *high* quality misclassifications were on events of 2-seconds or less, with only 17.5% and 1.8% of misclassifications on events longer than 5- or 10-seconds, respectively.

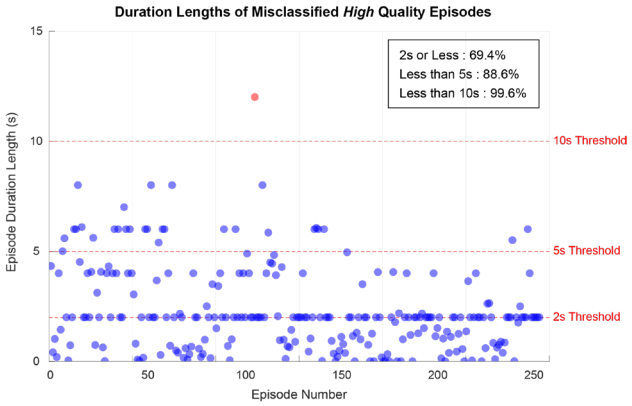

Figure **3**. Scatter plot showing the duration of *high* quality ECGs misclassified as *low* quality in the Holter Leads ON/OFF and Beacon Hospital datasets.

4. Discussion

A key challenge in the development of ECG SQ assessment methods is achieving the optimal balance between sensitivity – an algorithm's ability to correctly identify *high* quality segments – and specificity, which ensures the accurate exclusion of noise-corrupted data. The HeartKey SQ algorithm is designed with an intentional bias towards higher sensitivity, making it particularly effective as a triage tool. By prioritising sensitivity, the algorithm minimises the risk of discarding actionable data while filtering out clearly unusable segments, thereby reducing the *low* quality ECG review burden on interpreters.

The emphasis on sensitivity does result in a lower, though still acceptable, specificity, where some *low* quality signals may be incorrectly classified as suitable for review. In practice this trade-off is justified by the algorithm's primary goal of minimizing the risk of missing crucial clinical data, such as paroxysmal arrhythmias, through erroneous exclusion of actionable data. This sensitivity-focused design contrasts with other multi-level classification algorithms, which are primarily used as confidence measures to distinguish between artefact and ECG on a beat-to-beat level, where a more balanced performance between sensitivity and specificity is required.

The training and validation of the HeartKey SQ algorithm on diverse ECG datasets, encompassing various recording environments, device types, and patient demographics, ensures that the algorithm can reliably perform across different real-world scenarios. In combination with the tailored balance in performance, the diversity in training and validation confirms HeartKey's utility as a generalisable triage tool for ECG signal quality assessment that can maintain a consistent performance with minimal likelihood of errors in varied clinical contexts.

5. Conclusions

The HeartKey SQ algorithm offers a solution for ECG signal quality assessment in ambulatory settings, where noise and artefacts are common. By prioritising sensitivity, the algorithm effectively identifies *high*-quality ECG segments, reducing the risk of missing crucial clinical data like paroxysmal arrhythmias and easing the review burden on clinicians. While this sensitivity focus results in a lower specificity, the trade-off is justified by the algorithm's role as a triage tool. Crucially, the algorithm's training and validation on diverse ECG datasets ensure reliable performance across various recording environments, device types, and patient demographics, making it a versatile and generalisable tool for real-world clinical applications.

6. References

[1] Abdou A, Krishnan S. Horizons in Single-Lead ECG Analysis From Devices to Data. *Front. Sig. Proc.* 2022;2:866047.

[2] Chatterjee S, Thakur RS, Yadav RN, Gupta L, Raghuvanshi DK. Review of noise removal techniques in ECG signals. *IET Signal Process.* **2020**;14: 569-590.

[3] Apandi ZFM, Ikeura R, Hayakawa S, Tsutsumi S. An Analysis of the Effects of Noisy Electrocardiogram Signal on Heartbeat Detection Performance. *Bioengineering (Basel).* **2020**;7(2):53.

[4] McKenna S, McCord N, Diven J, Fitzpatrick MP, Easlea H, Gibbs A, Mitchell ARJ. Evaluating the impacts of digital ECG denoising on the interpretive capabilities of healthcare professionals. *Eur Heart J Dig Health*. 2024; ztae063 (preprint).

[5] Silva I, Moody GB, Celi L. Improving the Quality of ECGs Collected Using Mobile Phones: The PhysioNet/Computing in Cardiology Challenge 2011. *Computing in Cardiology*. 2011;38:273-276.

[6] Kuetche F, Alexendre N, Pascal NE, Colince W, Thierry S. Signal quality indices evaluation for robust ECG signal quality assessment systems. *Biomed Phys Eng Express*. 2023;9(5).

[7] Fu F, Xiang W, An Y, Liu B, Chen X, et al. Comparison of Machine Learning Algorithms for the Quality Assessment of Wearable ECG Signals via Lenovo H3 Devices. *J Med Biol Eng.* 2021;41, 231–240.

[8] Bijl KVD, Elgendi M, Menon C. Automatic ECG Quality Assessment Techniques: A Systematic Review. *Diagnostics*. 2022; 12(11): 2578.

[9] Rahman S, Karmakar C, Natgunanathan I, Yearwood J, Palaniswami M. Robustness of electrocardiogram signal quality indices. *J R Soc Interface*. 2022;19:20220012.

[10] Clifford GD, Liu C, Moody B, Li-Wei HL, Silva I, Li Q, Johnson AE, Mark RG. AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017. *Computing in Cardiology*. 2017;44:1-4.

[11] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J. Artif. Int. Res.* 2002;16(1):321–357.