HeartView: An Extensible, Open-Source, Web-Based Signal Quality Assessment Pipeline for Ambulatory Cardiovascular Data

Natasha Yamane(✉), Varun Mishra, and Matthew S. Goodwin

Khoury College of Computer Sciences and Bouvé College of Health Sciences, Northeastern University, Boston, MA 02115, USA
{yamane.n,v.mishra,m.goodwin}@northeastern.edu

Abstract. Wearable sensing systems enable peripheral physiological data to be collected repeatedly in naturalistic settings. However, the ambulatory nature of wearable biosensors predisposes them to common signal artifacts that researchers must address before analysis. Signal quality assessment procedures are time-consuming and non-standardized across research teams, and transparent reporting of custom, closed-source pipelines needs improvement. This paper presents HeartView, an extensible, open-source, web-based signal quality assessment pipeline that visualizes and quantifies missing beats and invalid segments in heart rate variability (HRV) data obtained from ambulatory electrocardiograph (ECG) and photoplethysmograph (PPG) signals. We demonstrate the utility of our pipeline on two datasets: (1) 34 ECGs recorded with the Actiwave Cardio from children with and without autism, and (2) 15 sets of ECGs and PPGs recorded with the RespiBAN and Empatica E4, respectively, from healthy adults in the publicly available WESAD dataset. Our pipeline demonstrates interpretable group differences in physiological signal quality. ECGs of children with autism contain more missing beats and invalid segments than those without autism. Similarly, PPG data contains more missing beats and invalid segments than ECG data. HeartView has a graphical user interface in the form of a web-based dashboard at https://github.com/cbslneu/heartview.

Keywords: Signal Quality Assessment · Data Pipelines · Ambulatory Cardiovascular Data · Electrocardiography · Photoplethysmography

1 Introduction

Signal quality assessment (SQA) involves detecting and evaluating outliers, artifacts, and missingness in signal-based data using expected signal morphology and dynamics. This procedure is an increasingly important step during and after data collection, as wireless ambulatory technologies are gaining popularity for their ability to monitor physiological states continuously and unobtrusively in both research and clinical settings [1–4]. Many wearable devices that capture peripheral physiological signals in free-living contexts are

D. Salvi et al. (Eds.): PH 2023, LNICST 572, pp. 107–123, 2024.
https://doi.org/10.1007/978-3-031-59717-6_8

commercially available and becoming progressively smaller and lighter. However, due to their size and ambulatory nature compared to traditional stationary systems, modern wearable system signals are more susceptible to artifacts, increasing missing or distorted data. Common sources of ambulatory signal artifacts include powerline interference, baseline wander, muscle activity, physical movement, and pressure disturbance [5, 6]. Figure 1 illustrates examples of ambulatory signal corruption by different artifacts.

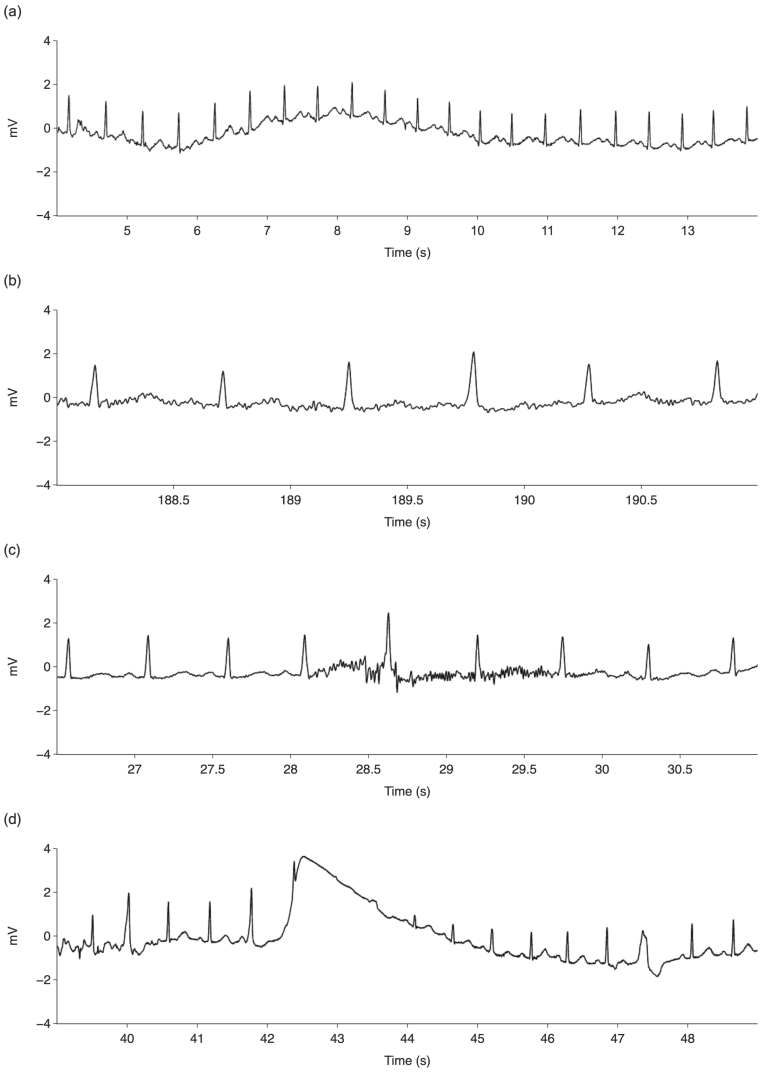

Fig. 1. Ambulatory electrocardiograph signals corrupted with (a) baseline wander, (b) powerline interference, (c) muscle activity, and (d) pressure disturbance.

Wearable devices use multiple biosensors to capture peripheral physiological signals. In the case of cardiovascular activity, photoplethysmography (PPG), to measure blood volume pulse (BVP), and electrodes, to record electrocardiograph (ECG) data, are the most common. PPG is an optical method of measuring volumetric changes in blood profusion [7]. Pulse rate, a proxy for heart rate (HR), is a function of the changes in light absorbed or reflected by blood flowing through a particular measurement site. ECG is a technique for observing and recording changes in heart electrical activity.

Before deriving summary statistics and making inferences based on PPG and ECG data, artifacts should be inspected using the expected morphological characteristics and dynamics of PPG and ECG signals. In a PPG waveform, each dominant peak represents a change in the absorption or reflection of light due to increased blood flow during systole (i.e., heart contraction). Similarly, the ECG signal contains three waveforms—the P, QRS, and T—corresponding to independent events in the cardiac cycle [8]. The most dominant wave is the QRS complex, which represents the depolarization of the heart's ventricles as a contraction begins. Correct detection of R peaks ensures that QRS complexes are captured, thus confirming valid heartbeats [9].

The primary goal of physiological SQA is to identify outliers, signal artifacts, and missingness to increase the reliability and validity of physiological measurements; however, the process varies across research teams. There is no unified approach to assessing signal quality in biosensor data [5], as standardized and transparent reporting of custom data preprocessing procedures is lacking [10]. In addition, most SQA procedures exist in closed-source pipelines, limiting reproducibility and uniformity across studies [11]. We developed HeartView to increase the reproducibility of and accessibility to SQA procedures typically performed only by trained researchers with computational backgrounds.

HeartView is a Python-based, open-source, extensible SQA pipeline and dashboard that visualizes and summarizes segment-by-segment quantification of missing and invalid beats in ambulatory cardiovascular data obtained in research contexts. SQA of a signal's basic quality and physiological feasibility is essential for making informed decisions about further data cleaning and processing procedures [5]. SQA of basic quality addresses whether beats are identifiable for reliable HR and heart rate variability (HRV) calculation. At the same time, physiological feasibility describes whether HR and inter-beat interval (IBI) values are valid. We demonstrate the utility of our pipeline in assessing the quality of physiological signals on two datasets covering different use cases: ECG data collected from children with and without autism spectrum disorder (ASD) [12], and a publicly available dataset containing PPG and ECG data collected from healthy adults [13].

2 Related Work

The level of SQA one performs depends on research integrity and clinical purpose. For instance, additional algorithmic development may be necessary in clinical contexts to assess specific waveform characteristics—e.g., whether the P, QRS, or T waves are identifiable—to diagnose conditions like myocardial ischemia [14] and heart disease [15]. Indeed, most work on SQA uses clinical datasets to derive and evaluate signal quality

indices (SQIs) to detect signal artifacts. SQIs are statistical or machine learning-based measures that heuristically describe characteristics and acceptability of signal waveforms [14–17] and are thus binary in most cases (i.e., "acceptable" or "unacceptable") [5]. Statistical SQIs may include kurtosis [18], skewness [16], signal-to-noise ratio (SNR) [19], and signal power [20]. Present machine learning-based SQIs commonly include measurements derived with support vector machine classifiers [21, 22] and neural networks [23, 24].

Several open-source data processing software and libraries [25–31] are available and can be applied to SQA of cardiovascular data. Some of the most popular data processing Python packages, such as NeuroKit2 and pyphysio, also support the computation of common statistical SQIs, including kurtosis and SNR [25, 28]. Other physiological data processing libraries are available for assessing basic quality and physiological feasibility checks, including filtering and visualizing signals and deriving peaks. BioSPPy, for example, provides a library of standard biosignal processing functions and feature extraction algorithms, including filtering, QRS complex detection, and visualization [26]. NeuroKit2, a community-driven Python package, contains functions to derive different types of peaks, filter signals, and compute HR [28]. Table 1 presents several software packages and libraries and their available features and functions relevant to the SQA of ambulatory PPG and ECG data.

Table 1. Overview of popular cardiovascular data processing software and libraries.

Package	FR	CE	SF	BD	V	CA	PPG	ECG	GUI
ANSLab	◐	●	●	●	●	◐	●	●	●
BioSPPy	○	○	●	●	●	●	●	●	○
ECGAssess	●	○	●	●	●	●	○	●	●
HeartPy	○	○	●	●	●	●	●	●	○
HRVTool	●	○	○	●	●	●	○	●	●
NeuroKit2	○	○	●	●	●	●	●	●	○
pyphysio	○	○	●	●	○	●	●	●	○
HeartView	●	⊕	●	●	●	●	●	●	●

○ Not existing ◐ Partially existing ● Completely existing ⊕ In development

Acronyms: *FR* = File reader; *CE* = Configuration exporter; *SF* = Signal filtering; *BD* = Beat detection; *V* = Visualization; *CA* = Code available; *PPG* = Photoplethysmography; *ECG* = Electrocardiography; *GUI* = Graphical user interface

In an informal survey of user needs distributed by one of our co-authors to 421 researchers and engineers[1] who process and analyze physiological data, 78% of respondents favored using well-documented, open-source software with user-friendly graphical user interfaces (GUIs). Based on our audit of existing popular open-source tools, only

[1] The survey sample comprised 31% researchers and 69% engineers from the Society of Psychophysiological Research (SPR), the IEEE International Machine Learning for Signal Processing (MLSP) workshop, and snowball sampling using personal contacts and social media.

three implement a GUI. ECGAssess is Python-based software that performs automated SQA and binary classification of the acceptability of multi-lead ECG data collected in clinical contexts for medical diagnosis [27]. ANSLab [31] provides both open-source and licensed, closed-source MATLAB-based software options through OpenANSLab and ANSLab Professional, respectively. The software suite contains modules that allow physiological data pre-processing on text files, artifact editing, and analysis. Additional functionalities in ANSLab Professional include batch processing, HRV analysis, and reading of multiple file types. To our knowledge, ANSLab is the only other GUI-based software suite with functions to process PPG data. Additionally, it is the only software suite capable of generating and exporting configuration files, a functionality that we plan to add to a future iteration of HeartView. Although open-source, OpenANSLab does not run as a standalone executable application and thus requires an installation of MATLAB, which is not free and therefore inaccessible to many outside of academia. In contrast, HRVTool is a standalone MATLAB application that performs ECG data processing and HRV feature extraction [30].

While the abovementioned open-source tools are valuable, uses are generally limited to those with programming skills. Further, some possess functionalities restricted by paywalls. In contrast, we propose a data pre-processing pipeline with an accompanying free, GUI-based solution for researchers without programming skills to perform necessary preliminary checks for basic quality and physiological feasibility of both ECG and PPG data. Our approach delivers an open-source, well-documented, and extensible web interface intended to increase efficiency and accessibility to a broader range of researchers who may not otherwise be able to conduct rigorous SQA on their ambulatory cardiovascular data.

3 HeartView Overview

We developed HeartView in Python 3.9 and its accompanying web-based dashboard using Plotly's Dash framework (version 2.8.1). Dash is an open-source Python framework built on Flask, Plotly.js, and React [32].

3.1 Data Processing Pipeline

The HeartView pipeline performs three main procedures (see Fig. 2) before outputting summary information on the dashboard: (1) data pre-processing (i.e., transformation and cleaning); (2) peak extraction; and (3) SQA metric computation.

Data Pre-processing. HeartView begins by reading and transforming raw ECG, PPG, and accelerometer data into Pandas data frames using device-specific file reading functions. Acceptable file types include European Data Format (EDF) files from the Actiwave Cardio and archive files from the Empatica E4. In addition, HeartView uses Pandas to read and pre-process comma-separated value (CSV) files generated from these devices, as well as the RespiBAN.

ECG. HeartView extracts timestamps and raw ECG values in units of millivolts from each EDF or CSV file of the Actiwave Cardio. Next, optional filters are applied to

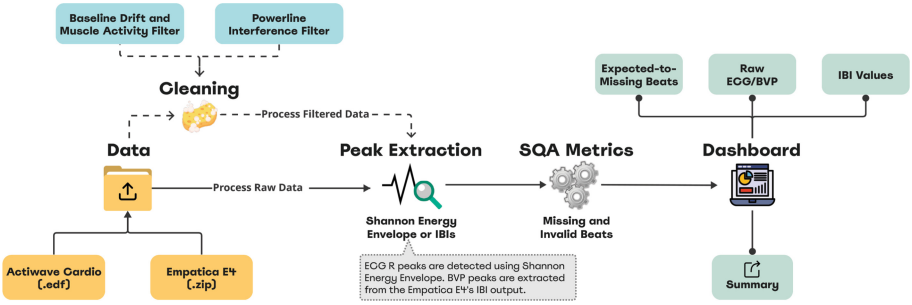

Fig. 2. HeartView pipeline architecture.

eliminate noise from the ECG data. For example, baseline wander and muscle noise can be eliminated using a bandpass filter between 0.5 and 45 Hz, and powerline interference at 60 Hz can be removed using a notch filter with a quality factor of 20.

PPG. Each archive file of the Empatica E4 comprises a set of CSV files containing raw and pre-processed data for HRV analysis. HeartView reads pre-processed inter-beat interval (IBI) values for later peak extraction and raw BVP values with timestamps for plotting and visual inspection purposes.

Acceleration. HeartView can also extract and process acceleration data from EDF, CSV, and archive files of the Actiwave Cardio, Empatica E4, and RespiBAN. The pipeline contains functions for smoothing raw data, converting from g-force to meters per second squared, and computing area under the curve (AUC) of acceleration magnitude. AUC is a commonly used proxy for movement over a given time window [33, 34], particularly when complex time or frequency-domain features are not under consideration at the SQA stage. In the present context, a higher AUC value indicates greater motion.

Peak Extraction. HeartView identifies peak locations from filtered data using the algorithm by Manikandan and Soman [35] for ECG data and from pre-processed IBI output from the Empatica E4 for PPG data[2]. R peaks from the ECG waveform are automatically detected using a Shannon Energy Envelope (SEE) estimator, peak-finding logic based on the Hilbert-transform [40], and zero-crossing point detection [35, 41]. This R peak detection algorithm has been validated using ambulatory ECG recordings from the MIT-BIH arrhythmia database [42, 43], demonstrating 99.8% average detection accuracy, 99.9% sensitivity, and 99.8% positive prediction. Peak locations in the PPG waveform are identified using timestamps from the pre-processed IBI file output provided by Empatica. Although multiple PPG peak detection algorithms exist [24, 44–46], HeartView uses the IBI time series provided by Empatica, given the device's widespread use [47] and the fact that existing algorithms have not been validated on datasets that are standardized to assess the performance of PPG peak detection algorithms.

[2] As our primary intent is to introduce a tool for researchers to perform initial assessment checks in their SQA, HeartView currently includes only one of several possible algorithms for ECG beat detection [36–39]. Future users could select and implement additional or alternative state-of-the-art algorithms.

Signal Quality Metrics. HeartView performs SQA on the basic quality (i.e., whether peaks are identifiable for reliable HR and IBI extraction) and physiological feasibility of a signal (i.e., whether the number of extracted peaks is valid) across segments of a user-customizable length. Thus, the pipeline measures signal quality based on the number of missing peaks per segment and invalid signal segments.

Missing Peaks. HeartView determines the number of missing peaks against an expected number of peaks. The pipeline derives this expected number of peaks by computing the median of all second-by-second HR values observed within each segment. We chose to use the median of all second-by-second HR values given its robustness to outliers. Second-by-second HR values are derived with the following steps. First, for each calculated IBI value, the pipeline computes a HR value by dividing the IBI value from 60,0000. Next, each second-by-second HR value is calculated using the harmonic mean of HRs (i.e., the reciprocal of the mean of the reciprocals of HRs) observed in a 2-s window based on Graham's approach [48].

$$\bar{x} = \frac{n}{\sum {}^{1}/_{x_i}} \tag{1}$$

In (1), n represents the number of HR values, and x_i represents a HR value at a timepoint i within the 2-s window. Thus, $\bar{x}$ represents the harmonic mean HR at one second, and the expected number is set to the median of all observed $\bar{x}$ in one segment. If the number of detected peaks is greater than the expected number of peaks in a segment, the missing number of peaks in that segment is set to zero; otherwise, the pipeline derives the number of missing peaks by calculating the difference between the number of detected peaks and the expected number of peaks.

Invalid Peaks. After peak extraction, HeartView counts invalid segments based on whether the number of detected peaks in a segment falls outside the range of [30:220] bpm. Based on prior work [2, 47, 49], the upper bound is set to the maximum human HR value, and the lower bound is set conservatively to a value close to the lower bound of the human physiological range.

3.2 Dashboard

We developed the HeartView dashboard using the open-source Dash framework, which consists of a Flask server that communicates with front-end React components [32]. Multiple callback functions with user input and state arguments are mapped to Dash core components, including a file upload component, data segmentation field, and checklist buttons corresponding to filter types (baseline wander, muscle activity, and powerline interference). These functions are then called separately to output interactive charts and a summary table. All user input components are contained within an off-canvas menu that can be toggled to appear or disappear from the left side of the screen.

As illustrated in Fig. 3, the main dashboard contains three separate panels: (1) a data summary panel, including information about the uploaded data file, computed signal quality metrics, and an export button to save all pre-processed data and signal quality information; (2) a bar chart with overlaying bars corresponding to the expected and

missing numbers of beats per segment; and (3) a line chart of the raw and filtered cardiovascular data with an overlaying scatterplot of detected peak locations. Two separate buttons are provided to access additional line charts of the corresponding IBI series and raw accelerometer data. We also include a range slider tied to a callback function that takes user-selected segment values and outputs a filtered view of the line charts. For example, in the top-right view in Fig. 3, the bottom panels contain line charts displaying raw and filtered ECG data from the Actiwave Cardio with points denoting locations of detected peaks within the first 20 s of segment 5.

Fig. 3. Multiple panels of the HeartView dashboard. *Top left*: Launch view of the off-canvas containing user input elements; *Top right:* Dashboard view of electrocardiograph (ECG) signal quality assessment and visualization of raw and filtered ECG with peaks; *Bottom left*: Dashboard view with a visualization of inter-beat interval (IBI) series; *Bottom right:* Dashboard view with a visualization of raw acceleration data.

4 Methods

In the following subsections, we discuss the datasets and data pre-processing and analysis procedures used in our assessment of the HeartView pipeline.

4.1 Datasets

Two datasets were used to assess the utility and incipient internal and external validity of HeartView. We leverage cardiovascular data from two studies, enTRAIN and Wearable

Stress and Affect Detection (WESAD). All adult participants and children's caregivers provided written informed consent to participate in the research before data collection.

The enTRAIN study [12] was carried out to investigate socio-affective behavior in 23 typically developing (TD) children and 11 children with autism spectrum disorder (ASD) while collecting interpersonal physiological data [50] during a series of standardized social-emotional regulation tasks. Children's cardiovascular data in the enTRAIN dataset comprise a total of 82.6 h ($M = 2.4$, $SD = 0.6$) of raw ambulatory ECG recorded at 1024 Hz with the chest-worn Actiwave Cardio by CamNtech, which has demonstrated good reliability and validity with gold-standard cardiovascular measures [51]. Children in the study had a mean age of 4.0 years ($SD = 1.1$).

WESAD [13] is a publicly available dataset featuring physiological and motion data recorded from 15 healthy adults during 'neutral,' 'stress,' and 'amusement' affective states. We used 24.1 h ($M = 1.6$, $SD = 0.2$) of ECG recorded at 700 Hz with the chest-worn RespiBAN and 29.7 h ($M = 2.0$, $SD = 0.2$) of PPG simultaneously recorded at 64 Hz from the Empatica E4. Participants in the WESAD study had a mean age of 27.5 years ($SD = 2.4$).

4.2 HeartView Assessment

HeartView outputs metrics regarding the basic quality (i.e., whether beats are identifiable) and physiological feasibility (i.e., whether IBI values are valid) of cardiovascular data and visualizes them on a web-based dashboard. We evaluate the HeartView pipeline by assessing whether group differences in signal quality can be captured in each dataset. Specifically, we tested differences in the numbers of missing peaks per segment and invalid segments between TD children and children with ASD in enTRAIN and between PPG and ECG recordings in WESAD. We hypothesized the following:

H_1: Our pipeline can capture group differences in the number of missing peaks and invalid segments between (a) ECG recordings of TD children and children with ASD and (b) PPG recordings from the Empatica E4 and ECG recordings from the RespiBAN devices.

Our rationale for H_{1a} is based on the observation that children with ASD display increased motor stereotypies [52], wandering behaviors [53], and symptoms of attention-deficit/hyperactivity disorder [54] compared to typically developing peers. As a result, these increased movement behaviors are likely to introduce more frequent signal artifacts into data collected from this population in psychophysiological experiments [55].

Our rationale for H_{1b} is based on previous work demonstrating that ECG devices tend to provide higher-quality data than PPG devices, which are subject to motion artifact [56, 57]. This discrepancy is likened to differences in sampling rate and mechanical configuration (e.g., optical sensors versus electrodes). Thus, we expect the data quality of PPG recordings to be impacted by signal artifacts more than that of ECG recordings.

Because data distributions were found to be non-normal for all groups with quantile-quantile plots and the Shapiro-Wilk test, we tested group differences with the Mann-Whitney test in the enTRAIN dataset and with the Wilcoxon signed-rank test for paired PPG and ECG recordings in the WESAD dataset. PPG and ECG data from the WESAD dataset is time-synchronized using the timestamps recorded by the RespiBAN upon initialization. In all ECG recordings, we first apply filters to eliminate baseline wander

and muscle noise, followed by powerline interference filters. All enTRAIN and WESAD recordings were segmented into 60-s windows and then trimmed to the start of the first experimental condition and the end of the last experimental condition for each participant before testing for differences.

We also present a potential use case of HeartView's signal quality metrics with data from one randomly selected participant with ASD and one randomly selected TD participant from the enTRAIN dataset. Considering previous findings [52–55] suggesting that ambulatory physiological data from children with ASD may be noisier than that of TD children due to increased physical activity, we sought to demonstrate that HeartView can reveal the relationship between ECG signal quality and physical motion in children. For each enTRAIN participant, the Actiwave Cardio device also contained a 3-axis accelerometer that collected acceleration data at 32 Hz. First, we smooth the acceleration data using a quarter-second moving average filter and then calculate acceleration magnitude values. We then compute second-by-second AUC values of acceleration magnitude using Riemann sums and normalize each value using min-max normalization so that each normalized AUC value is scaled to the range of [0, 1]. Next, we aggregate all AUC values into 60-s sliding windows with 15-s steps. Here, we use a sliding window approach to account for any time lags between physical motion and signal artifact onsets. Therefore, in each 60-s window, the AUC value of acceleration falls in the range of [0, 60]. Finally, we compute Spearman's rank correlation coefficients to evaluate correlations between AUC values and proportions of missing R peaks across all sliding windows.

5 Results

Below, we present the results of our assessment of the HeartView pipeline and demonstrate a use case of HeartView's signal quality assessment metrics with two randomly selected subjects from the enTRAIN dataset.

5.1 Group Differences

Overall, the results of our analyses substantiate our hypotheses. Among the ECG recordings of children with ASD in the enTRAIN study, we found greater proportions of invalid segments with a moderate effect size ($U = 156.0$, $p = .03$, $r = 0.32$) and average proportions of missing peaks per segment with a large effect size ($U = 210.0$, $p = .001$, $r = 0.53$) compared to the ECG recordings of TD children (see Fig. 4a).

Across all WESAD participants, Wilcoxon signed-rank tests revealed that PPG data recorded contained significantly more proportions of invalid segments ($z = -4.17$, $p < .001$, $r = 0.76$) and average proportions of missing peaks per segment ($z = -4.17$, $p < .001$, $r = 0.76$) than ECG data with large effect sizes (see Fig. 4b).

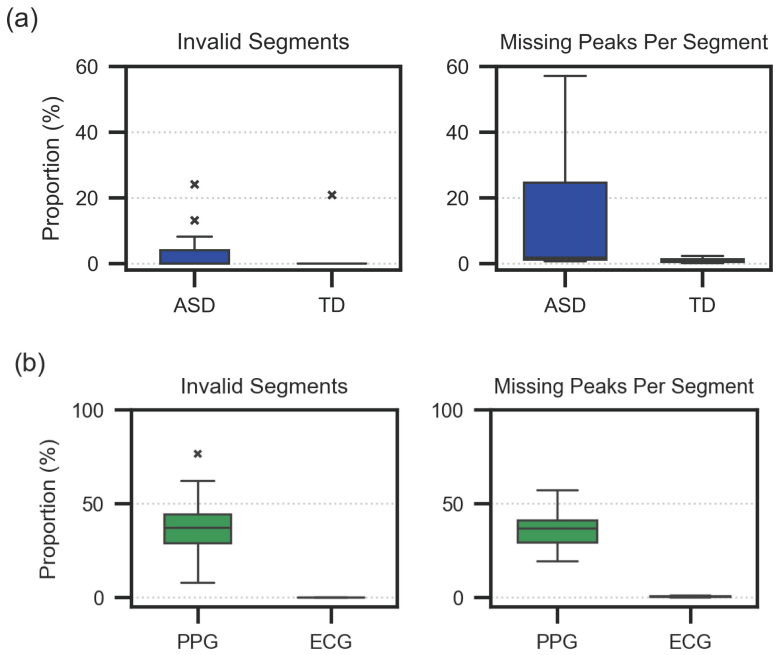

Fig. 4. Comparison of signal quality metrics between data of: (a) children with autism spectrum disorder (ASD) and typically developing (TD) children in the enTRAIN dataset; (b) photoplethysmograph (PPG) and electrocardiograph (ECG) devices in the WESAD dataset.

5.2 Cases: 'A03' and 'T10'

We randomly selected cases 'A03' and 'T10' in the enTRAIN dataset to demonstrate a use case for HeartView's signal quality metrics informing data processing procedures. Overall, across $n = 173$ sliding segments, 'A03's ECG recording contained an average of 3.47% ($SD = 6.88\%$) missing peaks per segment. Spearman's rank correlation coefficients for 'A03's proportions of missing peaks and normalized AUC values of acceleration magnitude also reveal a moderate, positive relationship between physical motion and data missingness ($\rho = .52$, $p < .001$). Across $n = 149$ sliding segments in 'T10's ECG recording, we found an average of 0.94% ($SD = 4.64\%$) missing peaks per segment and a non-significant correlation between physical motion and data missingness ($\rho = .12$, $p = .14$). Figure 5 illustrates the relationships between ECG data missingness and physical motion for 'A03' and 'T10.'

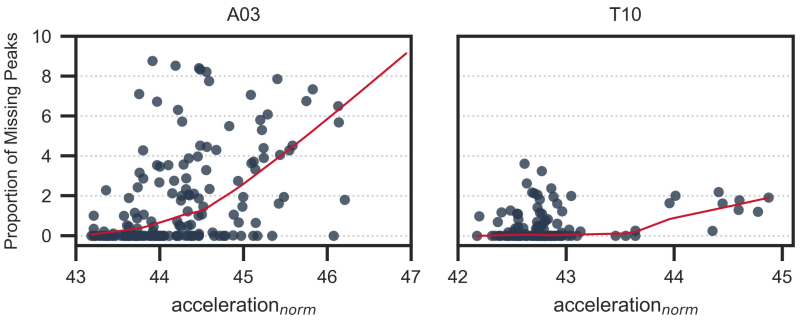

Fig. 5. Scatterplots of proportions of missing peaks per segment and normalized area-under-the-curve acceleration magnitude values in 'A03' and 'T10.' The red line represents locally weighted/estimated scatterplot smoothing curves.

6 Discussion and Future Work

HeartView is an extensible, open-source signal quality assessment pipeline with a web-based dashboard. With the HeartView dashboard, we provide researchers a tool to visually inspect ambulatory cardiovascular data's signal quality (i.e., numbers of missing beats and invalid segments) on a web browser using open-source Python libraries and frameworks. Assessments of basic quality and physiological feasibility of cardiovascular signals from wearable systems can improve the reliability of their measurements—an issue that continually precludes more widespread adoption in clinical contexts [58–60]. Our evaluation of the HeartView pipeline revealed group differences in signal quality within two datasets that support its incipient internal and external validity. We also present and discuss results from two randomly selected cases to demonstrate HeartView's ability to inform decisions about subsequent data processing procedures.

Within the enTRAIN dataset, the HeartView pipeline captured signal quality differences between ECG recordings of TD children and children with ASD. Specifically, ECG recordings of children with ASD contained significantly more invalid segments and missing peaks per segment than those of TD children. Our results are congruent with previous work by Kylliäinen and colleagues [55], who report more artifactual data collected from children with ASD or other developmental delays than from TD children, due to stereotypic behaviors and sensory differences in children with ASD. Further visual inspection of raw ECG signals with high percentages of missing and invalid data also revealed many segments containing motion artifacts and relatively weak or flat line signals. Indeed, our correlation analyses for two enTRAIN cases suggest that data missingness can be attributable to various noise sources. As observed in the case of 'A03,' data missingness can result from increased physical motion. However, it can also result from sources unrelated to physical motion, as observed in the case of 'T10.' In such instances, researchers may benefit from corroborating physiological data quality with ground truth labels, such as video-based annotations of gross motor movements (e.g., running around the observation room, fiddling with the ambulatory recording device). Additionally, researchers may decide to devote more time to cleaning data from children

with ASD, as well as applying specific inclusionary and exclusionary criteria (e.g., tolerating certain fidgeting movements in the ASD group, analyzing the same number of trials and time points in both groups) to pre-processing procedures [55]. Motion and behavioral data can be used to formulate these criteria if available. Analyses of HeartView's signal quality metrics with motion data also support decision-making about excluding specific data points or applying additional signal-cleaning procedures on ambulatory cardiovascular data. Further, HeartView's signal quality metrics can support researchers during feasibility testing of experimental protocols involving young children or children with behavioral challenges and wearable cardiovascular devices in pilot studies.

Within the WESAD dataset, our pipeline captured differences in data missingness and the number of invalid segments between PPG and ECG recordings from two different devices. Compared to ECG data from the RespiBAN, PPG data from the Empatica E4 contained more invalid segments and missing peaks per segment. This finding is consistent with those of previous studies comparing the data quality of cardiovascular signals derived from ECG and PPG devices [56, 57]. However, such discrepancies in signal quality could be due to a couple of confounding reasons. First, because PPGs from the Empatica E4 were recorded using fewer inputs than ECGs from the RespiBAN (i.e., a single input versus three leads, respectively), the Empatica E4 had an inherently greater likelihood of collecting data with missing peaks or invalid segments. Second, while the RespiBAN is meant to be worn on the chest, the Empatica E4 is worn on the wrist. Differences in sensor and electrode placement sites result in different levels and types of artifacts. For example, Zhang, Zhou, and Zeng [61] found that ECG data collected from sites on the upper arm were more susceptible to motion artifacts and muscle noise than ECG data collected from the chest. While we did observe differences between wrist-derived PPG and chest-derived ECG data, a limitation of the WESAD dataset is that recordings are limited to only one placement site per signal type. Future work should evaluate HeartView based on differences in signal quality using at least two placement sites per device.

The current functions and feature set of the HeartView dashboard limit its data processing and quality assessment to cardiovascular data collected with the Actiwave Cardio, Empatica E4, and RespiBAN. To increase HeartView's functional generalizability across devices, future iterations of the pipeline and dashboard will incorporate a data transformer to streamline data pre-processing by translating data from various devices by brands commonly used in research (e.g., Polar, Bittium, Shimmer, etc.) into a single format. Further, we plan to add more data processing algorithms and SQA functions specific to different devices and sensor types. For example, researchers may benefit from various algorithmic options at different pipeline steps to evaluate their relative performance when assessing signal quality. Such algorithms may incorporate acceleration measures to corroborate cardiovascular data quality. Given differences in device capabilities to output higher-level values such as HR and IBI (or lack thereof) [62], we also plan to add an IBI detection algorithm for raw PPG data. This need for additional functionality specific to devices and sensor types highlights the importance of an extensible, community-driven approach to software development, whereby researchers and developers use their own datasets to contribute to open-source code and provide valuable feedback on usability, utility, and reproducibility.

7 Conclusion

This paper presents HeartView, an extensible signal quality assessment pipeline with a web-based visualization dashboard for ambulatory cardiovascular data. We developed HeartView using open-source libraries and frameworks in Python. We assessed our pipeline using an ECG dataset collected from children with and without autism and the publicly available WESAD dataset. Our tool has a singular advantage over most extant cardiovascular signal pre-processing and quality assessment approaches. We offer an open-source pipeline with a user-friendly web interface that summarizes signal quality metrics through interactive visualizations and a summary table. A free and well-documented user interface can increase accessibility to signal quality assessment procedures historically only available to researchers with computer science and electrical engineering backgrounds. As a result, our signal quality assessment dashboard can contribute to more methodological transparency, reproducibility, and rigor that empowers researchers from diverse methodological backgrounds to make more informed decisions about the reliability and validity of their data when ambulatory biosensor data collection systems are used.

References

1. Goodwin, M.S.: Passive telemetric monitoring: novel methods for real-world behavioral assessment. In: Handbook of Research Methods for Studying Daily Life. pp. 251–266. Guilford Press, New York (2012)
2. Mishra, V., et al.: Continuous detection of physiological stress with commodity hardware. ACM Trans. Comput. Healthcare **1**, 1–30 (2020)
3. Weenk, M., et al.: Continuous monitoring of vital signs using wearable devices on the general ward: Pilot study. JMIR Mhealth Uhealth **5**, 1–12 (2017)
4. Pantelopoulos, A., Bourbakis, N.G.: A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **40**, 1–12 (2010)
5. Orphanidou, C.: Signal Quality Assessment in Physiological Monitoring: State of the Art and Practical Considerations. Springer, Cham (2018)
6. Fine, J., et al.: Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Biosensors **11**, 1–36 (2021)
7. Madhavan, G.: Plethysmography. Biomed. Instrum. Technol. **39**, 367–371 (2005)
8. Kusumoto, F.: ECG Interpretation. Springer Nature, Cham (2020)
9. Elgendi, M.: Optimal signal quality index for photoplethysmogram signals. Bioengineering (Basel) **3**, 21 (2016)
10. Quintana, D.S., Alvares, G.A., Heathers, J.A.: Guidelines for reporting articles on psychiatry and heart rate variability (GRAPH): recommendations to advance research communication. Transl. Psychiatry **6**, 1–10 (2016)
11. Stupple, A., Singerman, D., Celi, L.A.: The reproducibility crisis in the age of digital medicine. NPJ Digital Med. **2**, 3 (2019)
12. Yamane, N., Mishra, V., Goodwin, M.S.: Developing an open-source web-based data quality assessment pipeline for analysis of ambulatory cardiovascular data in individuals with autism. In: Paper presented at the International Society for Autism Research Annual Meeting (INSAR 2023), Stockholm, Sweden, 3 May 2023 (2023)

13. Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., Van Laerhoven, K.: Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In: Paper presented at the 20th ACM International Conference on Multimodal Interaction (ICMI 2018), New York, NY, 16 October 2018 (2018)
14. Quesnel, P.X., Chan, A.D.C., Yang, H.: Real-time biosignal quality analysis of ambulatory ECG for detection of myocardial ischemia. In: 2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1–5 (2013)
15. Bae, T.W., Kwon, K.K.: ECG PQRST complex detector and heart rate variability analysis using temporal characteristics of fiducial points. Biomed. Signal Process. Control **66**, 1–21 (2021)
16. Clifford, G., Behar, J., Li, Q., Rezek, I.: Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. Physiol. Meas. **33**, 1419–1433 (2012)
17. Daluwatte, C., Johannesen, L., Galeotti, L., Vicente, J., Strauss, D.G., Scully, C.G.: Assessing ECG signal quality indices to discriminate ECGs with artefacts from pathologically different arrhythmic ECGs. Physiol. Meas. **37**, 1370–1382 (2016)
18. Sharma, L.N., Dandapat, S., Mahanta, A.: Kurtosis based multichannel ECG signal denoising and diagnostic distortion measures. In: TENCON 2009 - 2009 IEEE Region 10 Conference. pp. 1–5. IEEE, New York (2009)
19. Zaunseder, S., Vehkaoja, A., Fleischhauer, V., Hoog Antink, C.: Signal-to-noise ratio is more important than sampling rate in beat-to-beat interval estimation from optical sensors. Biomed. Signal Process. Control **74**, 103538 (2022)
20. Castiglioni, P., Parati, G., Faini, A.: Cepstral analysis for scoring the quality of electrocardiograms for heart rate variability. Front. Physiol. **13**, 1–13 (2022)
21. He, R., et al.: Reducing false arrhythmia alarms in the ICU using novel signal quality indices assessment method. Paper presented at the 2015 Computing in Cardiology Conference, New York, NY, 6 September 2015 (2015)
22. Orphanidou, C., Bonnici, T., Charlton, P., Clifton, D., Vallance, D., Tarassenko, L.: Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. IEEE J. Biomed. Health Inf. **19**, 832–838 (2014)
23. Behar, J., Oster, J., Li, Q., Clifford, G.D.: A single channel ECG quality metric. In: 2012 Computing in Cardiology, pp. 381–384. IEEE, Kraków (2012)
24. Kazemi, K., Laitala, J., Azimi, I., Liljeberg, P., Rahmani, A.M.: Robust PPG peak detection using dilated convolutional neural networks. Sensors **22**, 1–22 (2022)
25. Bizzego, A., Battisti, A., Gabrieli, G., Esposito, G., Furlanello, C.: Pyphysio: a physiological signal processing library for data science approaches in physiology. SoftwareX **10**, 1–5 (2019)
26. Carreiras, C., Alves, A.P., Lourenço, A., Canento, F., Silva, H.P. da, Fred, A.: BioSPPy: Biosignal processing in Python (2015). https://biosppy.readthedocs.io/
27. Kramer, L., Menon, C., Elgendi, M.: ECGAssess: a python-based toolbox to assess ECG lead signal quality. Frontiers in Digital Health. **4**, 1–9 (2022)
28. Makowski, D., et al.: NeuroKit2: a python toolbox for neurophysiological signal processing. Behav. Res. Methods **53**, 1689–1696 (2021)
29. van Gent, P., Farah, H., van Nes, N., van Arem, B.: HeartPy: a novel heart rate algorithm for the analysis of noisy signals. Transport. Res. F: Traffic Psychol. Behav. **66**, 368–378 (2019)
30. Vollmer, M.: HRVTool—an open-source MATLAB toolbox for analyzing heart rate variability. In: 2019 Computing in Cardiology (CinC), pp. 1–4. IEEE, Singapore (2019)
31. Blechert, J., Peyk, P., Liedlgruber, M., Wilhelm, F.H.: ANSLAB: integrated multichannel peripheral biosignal processing in psychophysiological science. Behav. Res. Methods **48**, 1528–1545 (2016)
32. Plotly: Dash Python user guide. https://dash.plotly.com/

33. Jenkins, J.L., Valacich, J.S., Williams, P.: Human-computer interaction movement indicators of response biases in online surveys. In: ICIS 2017 Proceedings, pp. 1–16. Association for Information Systems (2017)
34. Varni, G., et al.: Interactive sonification of synchronisation of motoric behaviour in social active listening to music with mobile devices. J. Multimodal User Interfaces **5**, 157–173 (2012)
35. Manikandan, M.S., Soman, K.P.: A novel method for detecting R-peaks in electrocardiogram (ECG) signal. Biomed. Signal Process. Control **7**, 118–128 (2012)
36. Pan, J., Tompkins, W.J.: A real-time QRS detection algorithm. IEEE Trans. Biomed. Eng. BME **32**, 230–236 (1985)
37. Christov, I.I.: Real time electrocardiogram QRS detection using combined adaptive threshold. Biomed. Eng. Online **3**, 28 (2004)
38. Elgendi, M., Eskofier, B., Dokos, S., Abbott, D.: Revisiting QRS detection methodologies for portable, wearable, battery-operated, and wireless ECG systems. PLoS ONE **9**, 1–18 (2014)
39. Qin, Q., Li, J., Yue, Y., Liu, C.: An adaptive and time-efficient ECG R-peak detection algorithm. J Healthc, Eng. **2017**, 5980541 (2017)
40. Feldman, M.: Hilbert transforms. In: Braun, S., Ewins, D., Rao, S.S. (eds.) Encyclopedia of Vibration, pp. 642–648. Elsevier Ltd., Amsterdam (2001)
41. Sunami, N.: Shannon energy R peak detection (2020). https://github.com/nsunami/Shannon-Energy-R-Peak-Detection
42. Goldberger, A.L., et al.: PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation **101**, E215–E220 (2000)
43. Moody, G.B., Mark, R.G.: The impact of the MIT-BIH arrhythmia database. IEEE Eng. Med. Biol. Mag. **20**, 45–50 (2001)
44. Chen, L., Reisner, A.T., Reifman, J.: Automated beat onset and peak detection algorithm for field-collected photoplethysmograms. In: Paper presented at the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, 3 Sept 2009 (2009)
45. Shin, H.S., Lee, C., Lee, M.: Adaptive threshold method for the peak detection of photoplethysmographic waveform. Comput. Biol. Med. **39**, 1145–1152 (2009)
46. Kuntamalla, S., Reddy, L.R.G.: An efficient and automatic systolic peak detection algorithm for photoplethysmographic signals. Int. J. Comput. Appl. **97**, 1–6 (2014)
47. Mishra, V., et al.: Evaluating the reproducibility of physiological stress detection models. Proc. ACM Interact. Mobile Wearable Ubiq. Technol. **4**, 1–29 (2020)
48. Graham, F.K.: Constraints on measuring heart rate and period sequentially through real and cardiac time. Psychophysiology **15**, 492–495 (1978)
49. Tanaka, H., Monahan, K.D., Seals, D.R.: Age-predicted maximal heart rate revisited. J. Am. Coll. Cardiol. **37**, 153–156 (2001)
50. Palumbo, R.V., Marraccini, M.E., Weyandt, L.L., Wilder-Smith, O., Liu, S., Goodwin, M.S.: Interpersonal autonomic physiology: a systematic review of the literature. Pers. Soc. Psychol. Rev. **21**, 99–141 (2016)
51. Kristiansen, J., et al.: Comparison of two systems for long-term heart rate variability monitoring in free-living conditions: a pilot study. Biomed. Eng. **10**, 1–14 (2011)
52. Goldman, S., Wang, C., Salgado, M.W., Greene, P.E., Kim, M., Rapin, I.: Motor stereotypies in children with autism and other developmental disorders. Dev. Med. Child Neurol. **51**, 30–38 (2009)
53. Rice, C.E., et al.: Reported wandering behavior among children with autism spectrum disorder and/or intellectual disability. J. Pediatr. **174**, 232-239.e2 (2016)
54. Sinzig, J., Walter, D., Doepfner, M.: Attention deficit/hyperactivity disorder in children and adolescents with autism spectrum disorder: symptom or syndrome? J. Atten. Disord. **13**, 117–126 (2009)

55. Kylliäinen, A., Jones, E.J.H., Gomot, M., Warreyn, P., Falck-Ytter, T.: Practical guidelines for studying young children with autism spectrum disorder in psychophysiological experiments. Rev. J. Autism Dev. Disord. **1**, 373–386 (2014)
56. McCarthy, C., Pradhan, N., Redpath, C., Adler, A.: Validation of the empatica E4 wristband. In: 2016 IEEE EMBS International Student Conference (ISC), pp. 1–4 (2016)
57. Charlton, P.H., et al.: Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: technical and physiological determinants. Physiol. Meas. **38**, 669 (2017)
58. Bonnici, T., Orphanidou, C., Vallance, D., Darrell, A., Tarassenko, L.: Testing of werable monitors in a real-world hospital environment: what lessons can be learnt? In: 2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks, pp. 79–84 (2012)
59. Joshi, M., et al.: Perceptions on the use of wearable sensors and continuous monitoring in surgical patients: interview study among surgical staff. JMIR Form. Res. **6**, e27866 (2022)
60. Areia, C., et al.: Experiences of current vital signs monitoring practices and views of wearable monitoring: a qualitative study in patients and nurses. J. Adv. Nurs. **78**, 810–822 (2022)
61. Zhang, Q., Zhou, D., Zeng, X.: Highly wearable cuff-less blood pressure and heart rate monitoring with single-arm electrocardiogram and photoplethysmogram signals. Biomed. Eng. Online **16**, 23 (2017)
62. Nelson, B.W., Low, C.A., Jacobson, N., Areán, P., Torous, J., Allen, N.B.: Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. npj Dig. Med. **3**, 1–9 (2020)