

# RespWatch: Robust Measurement of Respiratory Rate on Smartwatches with Photoplethysmography

Ruixuan Dai

Washington University in St. Louis  
Computer Science and Engineering  
St. Louis, Missouri  
dairuixuan@wustl.edu

Michael Avidan

Washington University in St. Louis  
Department of Anesthesiology  
St. Louis, Missouri  
avidanm@wustl.edu

Chenyang Lu

Washington University in St. Louis  
Computer Science and Engineering  
St. Louis, Missouri  
lu@wustl.edu

Thomas Kannampallil

Washington University in St. Louis  
Department of Anesthesiology  
St. Louis, Missouri  
thomas.k@wustl.edu

## ABSTRACT

Respiratory rate (RR) is a physiological signal that is vital for many health and clinical applications. This paper presents *RespWatch*, a wearable sensing system for robust RR monitoring on smartwatches with Photoplethysmography (PPG). We designed two novel RR estimators based on signal processing and deep learning. The signal processing estimator achieved high accuracy and efficiency in the presence of moderate noise. In comparison, the deep learning estimator, based on a convolutional neural network (CNN), was more robust against noise artifacts at a higher processing cost. To exploit their complementary strengths, we further developed a hybrid estimator that dynamically switches between the signal processing and deep learning estimators based on a new Estimation Quality Index (EQI). We evaluated and compared these approaches on a dataset collected from 30 participants. The hybrid estimator achieved the lowest overall mean absolute error, balancing robustness and efficiency. Furthermore, we implemented *RespWatch* on commercial Wear OS smartwatches. Empirical evaluation demonstrated the feasibility and efficiency of *RespWatch* for RR monitoring on smartwatch platforms.

## CCS CONCEPTS

• **Applied computing** → **Consumer health**; • **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**.

## KEYWORDS

Smartwatch, mobile sensing, deep learning, hybrid model, respiratory rate

## ACM Reference Format:

Ruixuan Dai, Chenyang Lu, Michael Avidan, and Thomas Kannampallil. 2021. *RespWatch: Robust Measurement of Respiratory Rate on Smartwatches with Photoplethysmography*. In *International Conference on Internet-of-Things Design and Implementation (IoTDI '21)*, May 18–21, 2021, Charlottesville, VA, USA. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3450268.3453531>

## 1 INTRODUCTION

Respiratory rate (RR) is an important physiological variable associated with serious health conditions such as cardiopulmonary arrest [10]. In addition to the clinical applications, RR is important for ascertaining driving safety [24, 54], assessing sleep quality [4], monitoring stress [17] and even detecting opioid overdose [30]. However, unobtrusive monitoring of RR outside of laboratory and hospital settings is difficult. Traditional approaches for RR measurements rely on the use of the specialized equipment, e.g. capnography system and nasal/oral pressure transducers [6]. These approaches are not suitable for "free-living" or long-term measurement outside controlled clinical environments. Robust RR measurements with a popular commercial device can renovate the approaches to the real-time detection and long-term monitoring of respiration-related health conditions.

In this paper, we address the problem of robust RR monitoring using photoplethysmography (PPG) sensors on commercial smartwatches. The adoption of wearable devices, and smartwatches in particular, has increased exponentially over the past decade [5]. PPG sensors have been commonly embedded in smartwatches to measure heart rate and detect various health conditions, such as atrial fibrillation [46] and sleep apnea [23]. And smartwatches have the potential to enable unobtrusive longitudinal RR monitoring outside clinical environments with the PPG sensor.

However, RR monitoring on smartwatches with PPG faces several challenges. First, many previous studies [22, 26, 37] focused on PPG sensors for measuring light signals transmitted through fingertips, whereas smartwatch PPG sensors measure signals reflected from the wrist, which degrades signal quality and introduces noise artifacts [41]. As such, it is essential to develop robust approaches to extract RR from noisy PPG signals [42], and to investigate the feasibility of reliable RR measurements on off-the-shelf smartwatches.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*IoTDI '21*, May 18–21, 2021, Charlottesville, VA, USA

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8354-7/21/05.

<https://doi.org/10.1145/3450268.3453531>

Second, previous research on RR monitoring with PPG usually targeted use cases with minimum or no motion (e.g., a patient wearing a pulse oximeter in an Intensive Care Unit (ICU) bed) [22, 38, 50], whereas we aim for RR monitoring in the presence of some user motions and noise artifacts. It is inevitable to the motions with a wrist-worn smartwatch in the unconstrained settings. Therefore, smartwatch-based RR measuring system must be consistently robust for the longitudinal monitoring. Finally, smartwatches have limited computational resources and power. For any real-time and long-term RR monitoring system running on the smartwatch, data processing pipelines and algorithms should be highly efficient and capable of continuous execution on the resource-constrained platform.

Towards this end, we present *RespWatch*, a wearable sensing system for robust RR measurements with built-in PPG sensors on commercial smartwatches. *RespWatch* provides end-to-end processing pipelines from the raw PPG signals to RR measurements that can maintain high accuracy in the presence of some noise and motion artifacts. We explore and compare both signal processing and deep learning approaches, and develop a hybrid approach to combine their complementary strengths. Furthermore, *RespWatch* is capable to run completely on commercial smartwatches which allows for non-obtrusive RR monitoring. Specifically, the main contributions of this research are as follows:

- A *signal processing estimator* with fine-grained elimination of noise artifacts, which achieves efficiency and accuracy in the presence of moderate noise artifacts;
- A *deep learning estimator* for extracting RR from noisy PPG signals, which exhibits robustness in the presence of increasing noise artifacts;
- A *hybrid approach* which dynamically switches between signal processing and deep learning based on a novel Estimation Quality Index (EQI), achieving both robustness and efficiency;
- A comparative evaluation of the RR estimation approaches on a dataset including 30 participants of various activities, which demonstrates the complementary strengths of the signal processing and deep learning estimators and the advantage of combining both approaches in the hybrid estimator;
- Implementation and experimentation of *RespWatch* on commercial Wear OS smartwatches, which demonstrates the feasibility and efficiency of RR monitoring on smartwatch platforms.

## 2 RELATED WORK AND BACKGROUND

### 2.1 Non-contact RR measurement

Recently, non-contact sensing approaches have been developed for measuring RR. Techniques based on radio frequency (RF) detect respiration based on changes in RF signals caused by inhalation and exhalation motions. RR has been estimated using Frequency Modulated Carrier Waves (FMCW) [52] and Doppler radar [55]. WiFi signals have also been adopted to estimate RR based on the received signal strength (RSS) [35] and channel state information (CSI) [53]. Other non-contact sensing techniques for RR measurement exploit energy spectrum density (ESD) of acoustic signals [54] and ground movement from geophones [20]. As these non-contact approaches

rely on external devices in the environment, they are constrained to instrumented environments and cannot provide monitoring when users leave such environments.

### 2.2 IMU-based RR measurement

Smartwatches provide a portable platform with built-in sensors that can be utilized for unobtrusive sensing. Previous research [13, 16, 25, 49] on RR measurement with smartwatches exploited the inertial measurement unit (IMU) to capture subtle motions owing to respiration. However, this micro-motion is easily overwhelmed by motion artifacts [25] during normal activities. Hence, IMU-based RR monitoring is usually limited to constrained settings with minimum motion. For instance, Sun et al. [49] designed a total variation filter to extract respiratory signals from accelerometer data captured by smartwatch during sleep. Similarly, Hao et al. [13] developed the *MindfulWatch* to monitor respiratory during meditation, using a similar filtering approach. To extend the RR measurements in daily living activities, Liaqat et al. [25] proposed to identify accurate sensor readings with respiration information using a machine learning model, and extract RR only from those accurate sensor readings. However, since the micro-motions associated with respiration could be of the same order of magnitude as the sensor noise [13] and several orders of magnitude lower than other body motions, the signal-to-noise ratio (SNR) can often drop below the threshold for valid measurements.

### 2.3 PPG-based RR measurement

PPG is an optical sensing technology that detects pulsatile blood volume changes in tissues [44]. Compared to IMU, PPG sensor readings are less vulnerable to motion artifacts, as PPG measures the optical changes that are not directly impacted by the motions. As illustrated in Figure 1, a PPG sensor consists of a light-emitting diode (LED) to illuminate the tissue and a photodiode (PD) to measure the light transmitted through or reflected by the tissue. Transmission-mode PPG is commonly used in fingertip pulse oximeters, whereas reflectance-mode PPG is usually used on wrist or forehead for heart rate monitoring. The mode and placement of the PPG sensor has impacts on its sensing accuracy and waveform shape [41].

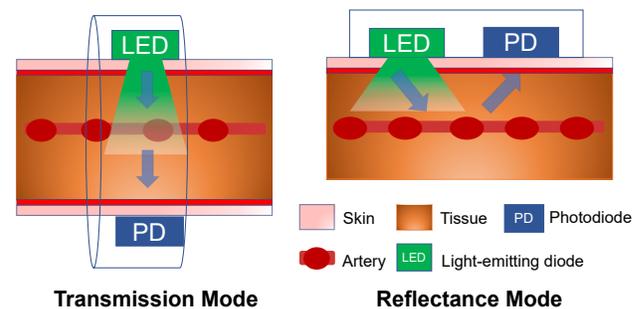
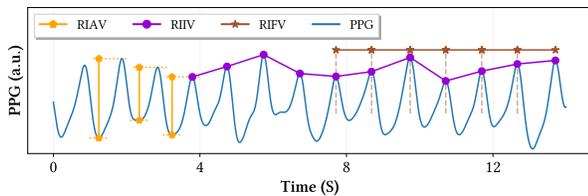


Figure 1: Two modes of the PPG sensor [3].

RR measurement is based on the fact that the PPG waveform is modulated by the respiration process. As illustrated in Figure 2, the PPG waveform contains three types of respiratory-induced

variations caused by amplitude, intensity and frequency modulation [22, 28].

- Amplitude modulation leads to *respiratory-induced amplitude variation (RIAV)*, which is related to changes in peripheral pulse strength [22]. RIAV is reflected in the different amplitudes of the peaks and corresponding valley for each pulse in the PPG waveform, and can be extracted as a time-series of the vertical distances from the peak to the valley for each pulse.
- Intensity modulation leads to *respiratory-induced intensity variation (RIIV)*, which is related to the intrathoracic pressure variation [22]. RIIV is reflected in a baseline wander [28] in the PPG waveform, and can be extracted as a time-series of the peak heights.
- Frequency modulation leads to *respiratory-induced frequency variation (RIFV)*, which is related to an autonomic response to respiration. RIFV, also referred as respiratory sinus arrhythmia (RSA) [22], is reflected in different inter-beat intervals, and can be extracted as a time-series of the horizontal distances between the successive peaks in the PPG waveform.



**Figure 2: PPG waveform and respiratory-induced variations [22]. RIAV: respiratory-induced amplitude variation; RIIV: respiratory-induced intensity variation; RIFV: respiratory-induced frequency variation.**

RR can be estimated in two general steps [7]: (1) extracting the respiratory variation signals, and (2) estimating of RR from the variation signals. Karlen et al. [22] used the Incremental-Merge Segmentation method to detect artifacts and extract the three respiratory-induced variations (RIIV, RIAV, RIFV). RR can then be obtained from the variations by Fast Fourier Transform (FFT) with smart fusion. Pimentel et al. [38] improved the reliability of RR measurements with multiple autoregressive (AR) models for determining the dominant respiratory frequency in the three variations. Compared to the fusion method from [22], the AR models can retain more data windows. The aforementioned studies on PPG-based RR measurements [7, 22, 26, 38] focused on fingertip sensors in clinical settings or during sleep with limited motion. Since the signal admission control in those studies discards entire sampling windows affected by noise artifacts, the approaches may lead to low data yield in the presence of user activities. How to robustly distill respiratory information from the raw PPG signal remains challenging especially in the presence of noise.

Video-based PPG has also been explored to measure RR with smartphone cameras[45]. Due to its reliance on video taken by cameras, this approach is not suitable for long-term and non-obtrusive RR monitoring during daily activities.

Recent studies [19, 27, 50] applied similar signal processing approaches to measure RR with reflective PPG sensors. Jarchi et al. [19] and Longmore et al. [27] explored measuring RR at different body positions (including wrist) with reflective PPG sensors. They found that upper-body positions (e.g., head and neck) produced the best respiration signals. Trimpop et al. [50] demonstrated a system on commercial smartwatches for RR monitoring during sleep and evaluated it on four users, but without revealing the details of the methodology. As those studies [19, 27, 50] adopted similar signal processing approaches to those developed for fingertip PPG, they did not address the more significant noise artifacts with user activities and the reflectance mode of PPG sensor. In contrast, we present novel signal processing techniques specifically designed to robustly estimate RR in the presence of noise artifacts and user activities. Moreover, we explore deep learning to further enhance the robustness of RR measurements against noise and motion artifacts, and integrate both approaches to balance robustness and efficiency of RR monitoring on commercial smartwatches.

## 2.4 Deep learning on smartwatch

Deep learning with wearable data has drawn great attentions in activity recognition [2, 40], Parkinson Disease monitoring [12, 51], atrial fibrillation detection [36, 46] and other mobile health applications [25, 54]. Ravichandran et al. [43] had proposed a dilated residual inception model to regress the respiration waveform from the PPG waveform. But their study was limited to the fingertip PPG signals collected in the intensive care unit (ICU), and cannot estimate the respiratory rate directly. Those application-driven studies have demonstrated that deep learning is capable to handle some sophisticated problems with the wearable data. However, further empirical evaluations of the deep learning models are required to test their capability of running on the wearable devices in real life.

## 3 DESIGN OF RESPWATCH

Towards RR measurements outside the clinic settings, our RR monitoring system shoots the following goals:

- **Accuracy.** The system should produce accurate RR measurements.
- **Robustness.** The system should maintain accuracy and data yield in the presence of noise artifacts.
- **Efficiency.** The system should have light-weight and efficient processing pipeline on smartwatches.

We exploited both signal processing and deep learning approaches to the RR estimations. In this section, we first design a signal processing estimator that achieves efficiency and accuracy in the presence of moderate noise artifacts. We then build a deep learning estimator that is more robust against increasing noise artifacts while incurring higher processing cost. Finally, we develop a hybrid estimator that balances robustness and efficiency by dynamically switching between the signal processing and deep learning estimators.

### 3.1 Signal Processing Estimator

The signal processing estimator employs digital signal processing techniques, which are training-free and allow for efficient processing on a commercial smartwatch. We designed a signal processing pipeline comprising three stages (see Figure 3).

1) In the *preprocessing* stage, we use a bandpass filter to eliminate noise outside the cardiac and respiratory bands from the raw signals, and then divide the signal waveform into 60-second windows.

2) In the *artifact elimination and pulse peak finding* stage, we employ a fine-grained technique to remove data points corrupted by noise artifacts within the cardiac and respiratory bands. We then use a novel *PPG pattern detector* to find the pulse peaks in the remaining data points. To facilitate finding the pulse peak, the PPG pattern detector employs a novel forward-backward highpass filter to remove respiratory band information while preserving the time-domain features. This allows the identified pulse peaks to be mapped to the original waveform with the respiratory information.

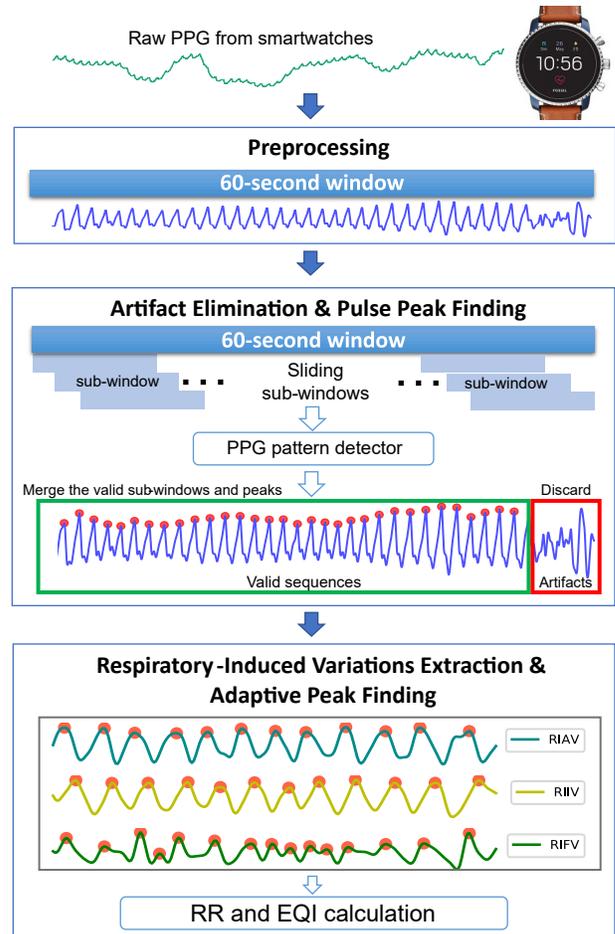
3) Finally, we extract the respiratory-induced variation signals, using the pulse peak positions found by the PPG pattern detector. Then, we mapped the variation signals to the RR estimations with an adaptive peak finding method. In addition, we introduce a novel *estimation quality index (EQI)* to assess the accuracy of our RR measurements, which enables the hybrid estimator to dynamically switch between signal processing and deep learning estimators, maintaining high accuracy and efficiency.

In the following we detail the design of each stage.

**3.1.1 Preprocessing.** The sampling rates of the PPG sensors vary across smartwatches and are usually higher than 100 Hz. In order to minimize the effect of the differences of sampling rates, we re-sample the collected PPG data at a fixed rate of 50 Hz based on data timestamps. The raw PPG waveform contains many noisy frequency components. A sixth-order Butterworth bandpass filter is first applied to remove the unwanted noisy components with cut-off frequency of 0.14Hz and 3Hz, only keeping the information from respiratory band to cardiac band [31]. The preprocessing does not remove noise artifacts within the ranges of respiratory and cardiac bands, which is handled in the next stage. The signal is then re-scaled to the range from -1 to 1. We divide the PPG data into 60-second windows for RR estimation in following stages. There is no overlap between the adjacent windows. The 60-second length has been used for RR studies in previous literature [22, 38].

**3.1.2 Artifact Elimination and Pulse Peak finding.** Noise artifacts are inevitable on smartwatches due to the wrist movement and poor contact between the sensor and skin. Previous research [7, 22, 26, 38] has often discarded any data window containing noise artifacts. Although this approach can help in avoiding noise artifacts, it leads to significant drop in RR measurement yield, especially during user activities. To support long-term RR monitoring in the presence of noise artifacts and improve data yield, we introduce a *sliding sub-window* technique to discard noise artifacts at a finer granularity while preserving the valid data samples in the same 60-second data window. The sliding sub-window has a size of 10 seconds and a step size of 2 seconds. Each 10-second sub-window from the 60-second window is passed through the PPG pattern detector to evaluate whether it is free from noise artifacts, and to identify valid pulse peaks simultaneously. The entire procedures for artifact elimination and pulse peak finding are summarized in Algorithm 1.

In the PPG pattern detector, we first filter out the respiratory band information in the sub-window, as the respiratory band can impact the accuracy of finding the pulse peaks. A novel design of



**Figure 3: Architecture of the signal processing estimator in RespWatch**

the PPG pattern detector is the adoption of a second-order forward-backward highpass Butterworth filter with cut-off frequency of 0.6 Hz. The forward-backward filter is a *zero-phase* filter in which the phase response slope is zero at all frequencies. It achieves the zero-phase response by filtering the input data twice, first in the forward direction and then in the reverse direction. Hence, the order of the filter is doubled, and the filter is non-causal due to the reverse filtering [48]. Since the processing of sub-windows is performed once we have the 60-second large window, there is no requirement of the causality of the filter.

The forward-backward filter is a key component of the PPG pattern detector. A significant benefit of the zero-phase filter is that it is able to preserve important time-domain features in the filtered signal. Specifically, the pulse peaks in the filtered waveform appear at the same positions as the pulse peaks in the unfiltered waveform in the time domain [8]. This allows us to directly map the pulse peaks found in the filtered waveform back to the unfiltered waveform. Consequently, we can find the pulse peaks in the unfiltered waveform in the 60-second window when iterating through the sub-windows containing the filtered signals. The pulse peaks in the

**Algorithm 1: Artifact Elimination & Pulse Peak Finding**

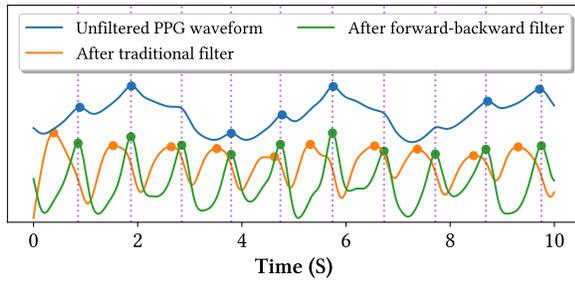

---

**Data:** 60-second preprocessed PPG waveform

- 1 Sliding sub-windows with size of 10s and step of 2s;
- 2 **for** each sub-window **do**
  - /\* begin of PPG pattern detector \*/
  - 3 2nd-order highpass **forward-backward filtering**;
  - 4 **Re-scale** the waveform with range of [-1,1];
  - 5 **Find** the peaks higher than 0;
  - 6 **Calculate** heart rate, peak intervals, peak-to-valley distances;
  - // PPG pattern matching
  - 7 **if** heart rate  $\leq 180$  **and** heart rate  $\geq 40$   
 $STD(\text{peak intervals}) < 0.4s$  **and**  
 $STD(\text{peak-to-valley distances}) < 0.4$  **then**
  - 8 | mark the sub-window as valid ;
  - 9 **end**
- 10 **end**
- 11 Merge the consecutive valid sub-windows into valid sequences;
- 12 Merge the peaks from the valid sequences into peak lists;

**Result:** Valid sequences and corresponding peak lists

---



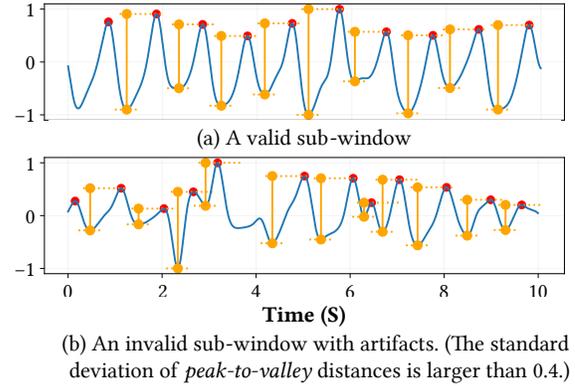
**Figure 4: Comparison of preprocessed PPG waveform with traditional filter and forward-backward filter.**

unfiltered waveform will be used to extract the respiratory-induced variation time-series in the next stage.

Figure 4 shows the advantage of the forward-backward filter with the real PPG data collected as part of our study. The blue curve shows the unfiltered waveform containing the respiratory band components. It has a large respiratory-induced baseline wander. Additionally, the pulses are not clearly distinguishable, making it difficult to find pulse peaks using standard peak finding methods [21, 47]. The green curve shows the waveform after it is processed using our forward-backward filter. Clearly, the peaks are distinguishable making it easier to check whether the waveform contains valid PPG patterns with a pattern matching method, and identify each peak by detecting local maxima above a certain threshold. It is important to note that these pulse peaks can be mapped back to the unfiltered curve (see the vertical dashed lines in Figure 4). In contrast, the orange curve is the filtered waveform after being processed by a traditional fourth-order Butterworth filter with the same cut-off frequency. We can observe variable time shifts in the peak positions

after the traditional filter, making it impossible to map the pulse peaks back to the blue curve. As a result, the forward-backward filter not only removes the respiratory band to facilitate PPG pattern matching and peak finding, but also allows the mapping of pulse peaks back to the unfiltered PPG waveform containing the raw respiratory information.

After forward-backward filtering, we re-scale the signal to a range of [-1, 1], and identify all the peaks whose amplitude are higher than 0. Then, we implement our PPG pattern matching method derived from [32] to detect whether the PPG signal is valid using three rules: (1) extracted heart rate based on the peaks should be within 40 and 180 bpm; (2) the standard deviation of the peak intervals should be less than 0.4s; (3) the standard deviation of the peak-to-valley distance should be less than 0.4, where the peak-to-valley distance is the vertical distance from the peak to its previous valley. We find the valleys via the local minimum between two adjacent peaks. Only those sub-windows satisfying all three rules are marked as valid.



**Figure 5: Examples of PPG pattern matching.**

Figure 5 shows real examples of a valid sub-window and an invalid sub-window with artifacts identified by the PPG pattern detector. The valid sub-window contains pulses satisfying the aforementioned rules, whereas the invalid sub-window has the standard deviation of the peak-to-valley distances larger than 0.4, not satisfying the third rule.

Once we have iterated through all the sub-windows with the PPG pattern detector, consecutive valid sub-windows are merged into larger valid sequences, and the peaks in the valid sub-windows are also merged into longer lists of peaks (see Figure 3). Here, the valid sequences mark the start and end points of a preprocessed PPG waveform free of noise artifacts, and the peak lists contain timestamps of the pulse peaks in the corresponding valid sequences. A 60-second preprocessed PPG waveform can have multiple valid sequences and valid peaks lists, if the invalid sub-windows appear in the middle of the 60-second period.

**3.1.3 Respiratory-induced Variations Extraction and Adaptive Peak Finding.** As described in Section 2.3, the respiratory-induced amplitude variation (RIAV) is the time-series of the vertical distances from the peak to the valley for each pulse. The respiratory-induced

intensity variation (RIIV) is the time-series of the height of each pulse peak. The respiratory-induced frequency variation (RIFV) is the time-series of the horizontal distances between successive pulse peaks in the time domain. All the three variation time-series are closely related to pulse peaks that need to be extracted from the preprocessed PPG waveform containing the respiratory band information. We directly adopt the peak lists from the last stage, and map the pulse peaks from the filtered waveform to the preprocessed PPG waveform. The valleys in the preprocessed waveform are then obtained by finding the local minimum between two adjacent peaks. The adoption of peak lists from the second stage improves the accuracy of extracting the three time series, because we avoid directly finding pulse peaks in the preprocessed PPG waveform in which the respiratory band information can degrade the accuracy of pulse peak finding. And it also saves us from the pulse peak finding twice for the PPG pattern matching and respiration signal extraction, making the system more energy-efficient.

The time-series of RIAV, RIIV and RIFV are not equally sampled in time domain, so we re-sample the three at  $f_s = 5\text{Hz}$  with linear interpolation. We also employ a bandpass filter to keep only the respiratory band (0.14-0.9Hz) information. To robustly detect all the respiratory peaks in the RIAV, RIIV and RIFV, we apply an adaptive peak finding method derived from [21]. The method starts with initial thresholds for the distance between two adjacent peaks. We find all the peaks whose amplitude are higher than 0, and calculate the horizontal distance from the current peak to the last peak. If the distance is below the lower threshold, the current peak will be discarded and the lower threshold decreases. If the distance is beyond the higher threshold, a virtual "peak" will be inserted in the middle of the current peak and the last peak, and the higher threshold increases. The initial thresholds and the adjusting rates are set based on [31]. Adaptive peak finding method can handle the cases in which the artifacts cause a spurious peak or obliterate a possible peak in the signals, based on the assumption that the RR is constant within a short period of time.

After we get the respiratory peaks in the three variations, the respiratory rate (RR) is calculated for each valid sequence:

$$RR_{RIXV,i} = \frac{60}{\text{MEAN}(\text{peak\_intervals}_{(i)})/f_s} \quad (1)$$

where  $RIXV$  is one of the three respiratory-induced variations,  $\text{MEAN}(\cdot)$  is the average value of  $\cdot$ ,  $f_s$  is the sampling rate,  $i$  is the index of the valid sequences, and  $\text{peak\_intervals}_{(i)}$  is the respiratory peak intervals detected by the adaptive peak finding method for the  $i^{\text{th}}$  valid sequence. The final RR measurement  $\text{RespWatch}_{RIXV}$  is the length-weighted average of  $RR_{RIXV,i}$ :

$$\text{RespWatch}_{RIXV} = \frac{\sum_i RR_{RIXV,i} \cdot \text{seq\_length}_{(i)}}{\sum_i \text{seq\_length}_{(i)}} \quad (2)$$

**3.1.4 Estimation Quality Index (EQI).** Furthermore, we introduce an *estimation quality index (EQI)* as a novel metric to estimate how accurate our RR measurement is. EQI is based on the two intuitions: (1) the respiration is rhythmic, so the standard deviation of the respiration peak intervals should be small, and (2) RR measurement is more accurate on the longer sequence. Specifically, the EQI of

each valid sequence is formulated as:

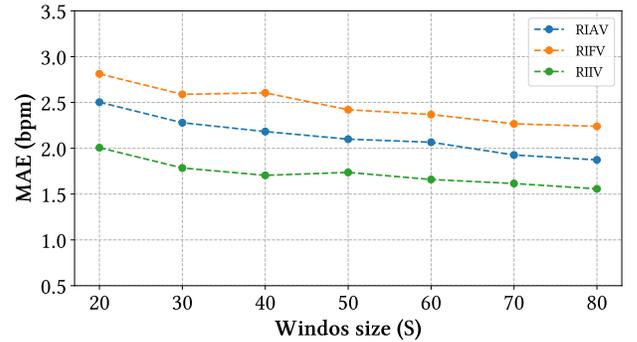
$$EQI_{RIXV,i} = \alpha \cdot \frac{\text{STD}(\text{peak\_intervals}_{(i)})}{\text{seq\_length}_{(i)}} \quad (3)$$

where  $\alpha$  is a fixed scaling factor,  $\text{STD}(\cdot)$  is the standard deviation of  $\cdot$ ,  $\text{seq\_length}_{(i)}$  is the length of the  $i^{\text{th}}$  valid sequence. The final  $EQI_{RIXV}$  is the sum of  $EQI_{RIXV,i}$  for each valid sequence:

$$EQI_{RIXV} = \sum_i EQI_{RIXV,i} \quad (4)$$

EQI offers several important advantages. First, most prior studies only focus on the RR estimation without providing an accuracy estimation. Lack of confidence of the measurement could lead to wrong decision in some practical cases. For example, the inaccurate high RR measure could give a false alarm of respiration conditions. Second, although past studies focused on motion as the main factor influencing PPG-based sensing accuracy [42, 46], other factors (e.g., light conditions and sweat) may also affect accuracy. EQI therefore captures noise artifact in a more comprehensive manner than motion artifacts alone. Finally, as EQI utilizes only the characteristics of the RR estimation process itself, it does not require external inputs (e.g., motion intensity, light, and sweat).

To assess our assumption that the accuracy of RR measurement improves with larger data windows, Figure 6 shows the mean absolute error (MAE) of RR measurements with different window sizes. From data collected from our user study (see Section 4), we randomly sampled 100 data windows at each window size, and all the sampled data windows were free from artifacts. We can observe that the MAE decreases with larger window size, which supports our assumption.

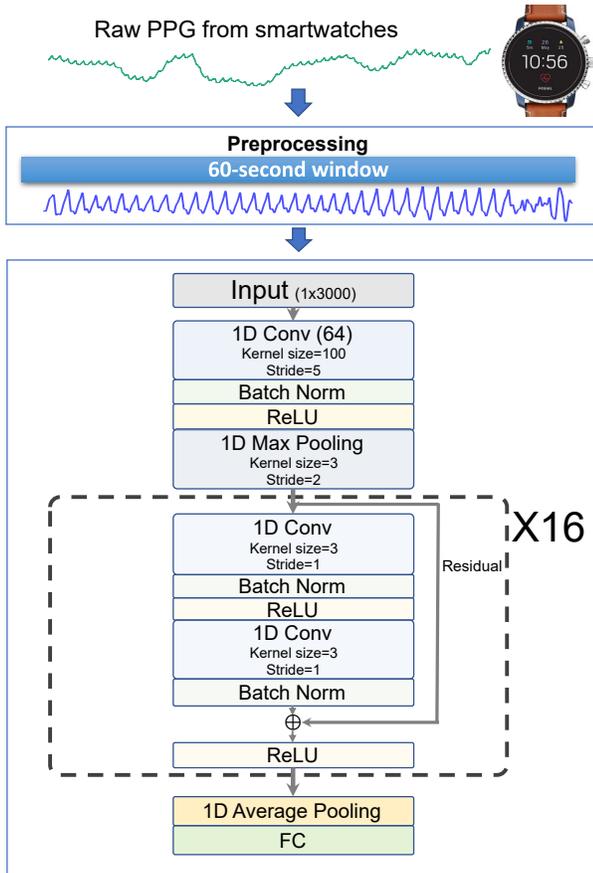


**Figure 6: MAE vs. window sizes. All the data windows are free from artifacts**

## 3.2 Deep Learning Estimator

This section presents the deep learning estimator for RR measurement. Our work was inspired by recent success of deep learning in processing smartwatch data [25, 46]. Particularly, Shen et al. [46] showed that a convolutional neural network (CNN) model with residuals was robust in the presence of motion artifacts for detecting atrial fibrillation with the smartwatch PPG. Building upon this, we designed the deep learning estimator with a CNN model. After

some basic preprocessing steps, our CNN model can directly output the estimation of RR using the PPG waveform. To the best of our knowledge, our deep learning estimator is the first deep neural network to estimate RR with wrist PPG on smartwatches. The high-level architecture of the deep learning estimator is illustrated in Figure 7.



**Figure 7: Architecture of the deep learning estimator in RespWatch**

**3.2.1 Preprocessing.** Although the CNN model can directly learn from raw signals, preprocessing is still needed to account for issues associated with the PPG signals. Specifically, raw PPG signals exhibit different ranges and amplitudes under different conditions, and the noise outside the respiratory and cardiac bands can lead to the overfitting of the CNN model. To standardize data and reduce noise, we re-sampled the PPG signal at 50Hz, applied the same bandpass filter used in the signal preprocessing estimator, and normalized the signals to a zero mean and a unit variance.

**3.2.2 CNN Model.** As shown in Figure 7, we developed our deep learning approach based on the residual neural network. Unlike previous classification tasks with PPG [46], our aim is to output the RR estimation with continuous values. As the PPG sensor on smartwatches contains only one channel, we adopted 1D convolutional layers across the network.

An initial 1D convolutional layer with kernel size of 100 is adopted to down-sample the input and reduce the computation complexity. Then, 16 basic blocks sharing the same topology with residuals bypass and 1D convolutions are applied. Each basic block contains 2 convolutions and a shortcut connection. The shortcut connections can optimize the training by allowing information to propagate in deep neural networks [15] and make the optimization process tractable. Batch normalization (Batch Norm) and a rectified linear unit (ReLU) activation layer are also employed after each convolutional layer. The 16 basic blocks are grouped into 4 stages consisting of 3, 4, 6 and 3 blocks, and the number of output channels for each stage is 64, 128, 256 and 512, respectively. The spatial map of the signal is down-sampled while the channels are incremented stage by stage. After the last stage of the basic blocks, we append a 1D average pooling layer and a fully connected layer. The fully connected layer performs the regression tasks of the final RR estimation. We employed the mean squared loss and the stochastic gradient descent optimizer with the momentum. During training, we ensured there is no overlap between the training and testing signal. All the convolutional layers were initialized with He initialization [14], and batch normalization layers were initialized with weight of 1, bias of 0.

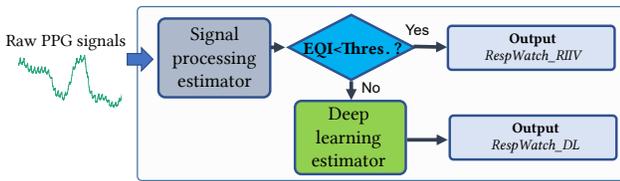
### 3.3 Hybrid Estimator

A key finding in our experimental results (see Sections 5 and 6) is that the signal processing and deep learning estimators have complementary strengths in efficiency and robustness, respectively. Specifically, the signal processing estimator achieves higher accuracy in the presence of moderate noise artifact. It also incurs lower processing cost on smartwatch platforms. In contrast, the deep learning estimator exhibits more robustness against increasing noise artifact. To maintain accuracy, robustness, and efficiency under varying noise artifact, we developed the hybrid estimator to combine the strengths of both the signal processing and deep learning estimators. Under increasing noise artifact, the hybrid estimator automatically switches from signal processing to deep learning to take advantage of its higher level of robustness. Conversely, the hybrid estimator switches back to signal processing when noise artifact diminishes to benefit from its higher efficiency and accuracy.

The key to the design hybrid estimator is to identify the metric used to make the switching decision. We explored motion intensity and EQI as two alternative metrics used to choose between the two estimators. Motion intensity is defined as the standard deviation of the magnitude of the tri-axial acceleration in a 60-second window [46]. It can be obtained from the IMU sensor in smartwatches. In comparison, as defined in Section 3.1.3, EQI characterizes the estimation quality that may be influenced by noise artifact in general, which may include both motion and other sources of noise (e.g., poor sensor contact).

For either motion intensity or EQI, we applied a grid search to find the best switch threshold that leads to the lowest mean absolute error (MAE) in RR measurements. We found experimentally (see evaluation in Section 5) that EQI outperforms motion intensity in accuracy and efficiency. In addition, EQI is derived from the PPG

signal itself, which does not rely on external sensors such as the IMU.

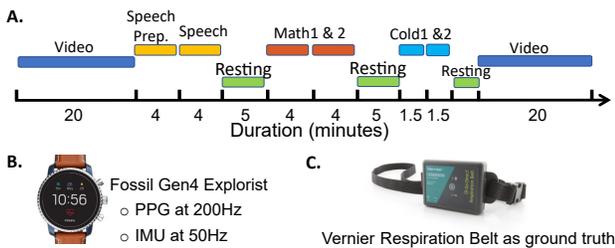


**Figure 8: Architecture of hybrid estimator. *RespWatch\_RIIV* is the output from signal processing estimator with RIIV; *RespWatch\_DL* is the output from deep learning estimator.**

The EQI-based hybrid estimator is illustrated in Figure 8. The EQI is first generated from the signal processing estimator. Since the signal processing method is highly efficient, the execution of the signal processing estimator incurs minor overhead for the whole system. After we get the RR measurement and EQI from signal processing estimator, if the EQI is below the switching threshold, the hybrid estimator directly output RR measurement from signal processing estimator. Otherwise, it invokes the deep learning estimator to produce the RR measurement.

#### 4 USER STUDY

We collected PPG data through a user study involving 32 healthy volunteers. The data collected in this study was primarily used to evaluate the accuracy of the RR estimations (see Section 5). The run-time efficiency of the estimators was empirically evaluated on smartwatches in Section 6.



**Figure 9: (A). Sequence of activities of the collecting procedure. (B). Fossil Gen4 Explorist smartwatch instrumented for this study. (C). Vernier Respiratory Go Direct Respiration Belt as ground truth.**

#### 4.1 Devices

We instrumented mainstream smartwatches, Fossil Gen4 Explorist, to collect raw PPG signals used to evaluate the RR estimations from RespWatch. The ground truth was obtained with Vernier Respiratory Go Direct Respiration Belt, which was used in the previous respiration studies [9, 29]. We also collected acceleration data from the IMU on the watch to measure motion intensity during the study. During data collection, each participant was asked to wear the belt over their chest and the smartwatch on their non-dominant hand. A

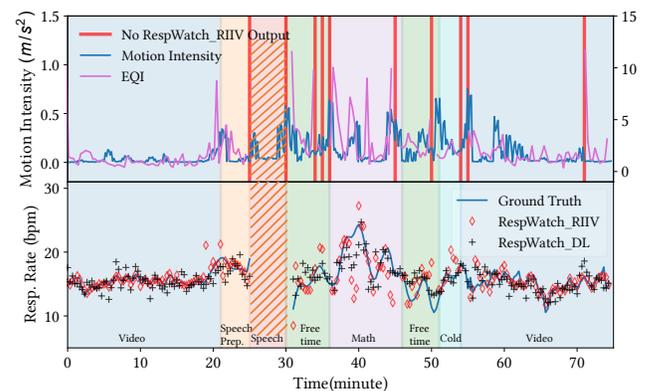
custom application was installed on the smartwatches to record the data from the PPG sensor at 200Hz and IMU at 50Hz. The data were initially stored locally on the smartwatches, and then uploaded to a secure server.

#### 4.2 Study Protocol

32 healthy volunteers were recruited through flyers posted across the campus at Washington University in St. Louis. All the participants met the inclusion criteria (between 18 and 69 years of age, with no heart disease, not pregnant at the time of recruitment, and not having an implanted pacemaker). At the end of the study, a compensation of a \$25 was provided. The institutional review board (IRB) of Washington University in St. Louis approved this study, and written consents were obtained from all participants (IRB#2019-04150). The data was collected in various scenarios, including (1) watching a video, (2) preparing and delivering a speech, (3) doing mathematical tasks on computers, and (4) holding a cold object for an extended period. All the activities involved motions to some degree. The timeline of the data collecting procedure and the devices is shown in Figure 9.

Two participants’ data were lost due to issues during data upload. As a consequence, only 30 participants’ data were used in the analysis and evaluation. Additionally, we exclude the data when the ground truth is not available, i.e., the Respiration Belt failed to acquire valid RR measurements. This occurred during some segments when participants were delivering the speech, as speaking caused unreliable RR measurements [17].

#### 4.3 Impacts of Activities



**Figure 10: RR measurements, motion intensity, and EQI of one user participating in the study.**

To show the impacts of noise or activity on RR measurements, we plot the RR time series produced by the RespWatch estimators for one user over the entire session, as shown in Figure 10 (excluding the speech activity as mentioned in the last subsection). The top graph shows the motion intensity and EQI overtime. We observed that both motion intensity and EQI have larger variations during math and free time periods. However, the correlation between the EQI and motion intensity was only around 0.17 (Pearson Correlation,  $p < 0.5$ ), which suggests that motion might not be the only

source of noise artifact for the PPG sensor. The solid red vertical lines mark the timestamp when the signal processing estimator failed to produce RR measurements when a data window contains no valid sequence after the artifact elimination. In the bottom graph of Figure 10, we show the ground truth and the output from the signal processing and deep learning estimators. Since estimations from RIIV outperforms those from RIAV and RIFV (see Section 5), we only displayed the output with RIIV here. During the video period, we had fewer motions with lower EQI, and the signal processing outputs (RespWatch\_RIIV) were closer to the ground truths than the deep learning outputs (RespWatch\_DL). This shows that the signal processing estimator achieved high accuracy in the presence of moderate noise artifact (as indicated by the low motion intensity and EQI). However, when the motion or EQI increased, e.g., during math or free time periods, the signal processing estimator produced larger errors than the deep learning estimator. This demonstrates that deep learning estimator is more robust against higher level of noise, and highlights the advantages of our hybrid approach that utilizes the signal processing for high accuracy when EQI is low, and deep learning for robustness when EQI is high.

## 5 EVALUATION OF RESPWATCH

This section presents an evaluation of the three estimators supported by RespWatch. The accuracy of RR measurements was assessed using the mean absolute error (MAE) in breaths per minute (bpm), defined as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_{ref,i}| \quad (5)$$

where  $n$  is the number of data windows,  $\hat{y}_i$  is the estimated RR and  $y_{ref,i}$  is the reference RR from ground truth. Moreover, we analyze the trade-off between accuracy and yield of RR measurements.

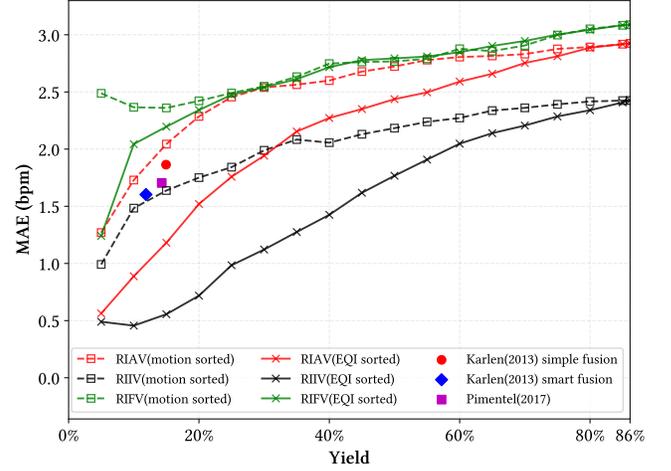
### 5.1 Signal Processing Estimator

We compared the performance of our signal processing estimator to existing methods for measuring RR based on PPG. We implemented three state-of-the-art methods from [22, 38] as the baselines for performance evaluation. Those methods were previously evaluated on large data sets, and have been adapted to work with wrist-worn PPG [27]. The first two baseline methods were the *simple fusion* and *smart fusion* methods from [22], which utilized Fourier transform and fusion techniques. The third baseline method [38] utilized autoregressive (AR) models. All the three baseline methods have the data admission controls, which discard an entire data window that are found to contain noise artifacts. Table 1 shows the performance of the baseline methods on our dataset. We observed that a large portion of data windows were discarded due to the admission control. Unlike the baseline methods, our signal processing estimator employs fine-grained artifact elimination and can estimate RR even with some artifacts in a data window. We only discarded 13.86% of the data windows that contained no valid sequence after the artifact elimination. Hence, our signal processing estimator achieved a significant higher yield of 86.14% than the baseline methods.

We analyzed the trade-off between accuracy and yield of RR measurements. Figure 11 shows the MAE-yield curves for the signal processing estimator. As motions are previously used as a metric

**Table 1: Baseline methods on our dataset**

Method	Yield	MAE (bpm)
Karlen (2013) (Simple Fusion)	14.95%	1.876
Karlen (2013) (Smart Fusion)	11.87%	1.603
Pimentel(2017) (AR models)	14.29%	1.704



**Figure 11: MAE vs. Yield. Different colors represent the estimations from RIAV, RIIV and RIFV, respectively. The line styles indicate different sorting criterion (Motion, EQI). The baselines are illustrated as dots with different shapes and colors.**

to reject the PPG measurements [33] and the proposed EQI is also capable for the same propose, we rank the data windows based on the corresponding motion intensity and EQI, respectively, to investigate the accuracy at different yields. For motion intensity, we computed the motion intensity for each PPG data window, and sorted the data windows by ascending motion intensity. Then, we calculated the MAE of RR measurements with motion intensity in the lowest  $\alpha$ -th percentile, ranging from 5% to 100%. Similarly, we also calculated the MAE for different yields using EQI as the metric for sorting the data windows. In Figure 11, Different colors distinguish the estimations from different respiratory-induced variation signals (RIA, RIIV, RIFV), and different line styles distinguish the different ranking criterion (Motion intensity or EQI). For example, the dashed lines with square markers are the MAE curves with increasing motion intensity. Each point  $(\alpha, e)$  on the dashed lines indicates the MAE of  $e$  on the subset of data windows whose motion intensities are in the lowest  $\alpha$ -th percentile, corresponding to the yield of  $\alpha$ %. The max yield of the signal processing estimator is 86.14% because we discarded 13.86% of the data windows that contain no valid sequence. In contrast, as the baseline methods have fixed data yield due to their data admission control policies, the results of the baseline methods are displayed as three discrete data points in the figure.

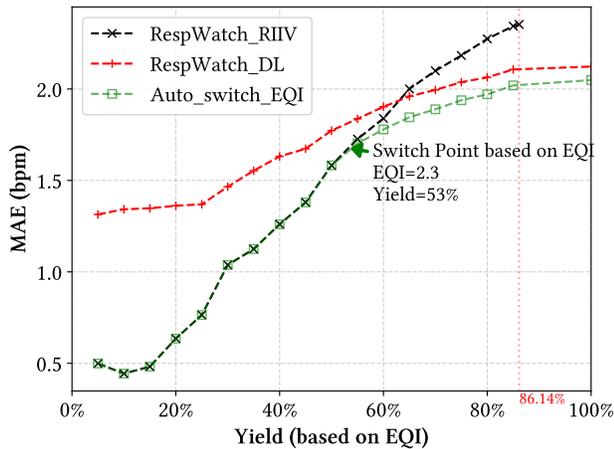


Figure 12: MAE vs. Yield based on EQI ranking.

Observing the dashed lines with square marks in Figure 11, we found that all the three outputs from the signal processing estimator, RIAV(motion sorted), RIFV(motion sorted) and RIIV(motion sorted), had an increasing trend, suggesting that the motions indeed have a negative influence on the estimation accuracy. In practice, we may select the motion intensity threshold to achieve the desired trade-off between accuracy and yield of RR measurements based on Figure 11. RIIV(motion sorted) outperforms all the baselines when at the same yield level. And RIIV(motion sorted) has the lowest MAE at any yield level among the three outputs from the signal processing estimator. This suggests that the RIIV is the most suitable respiratory-induced variation signal to estimate RR from the smartwatch PPG signals.

Next, we investigate the relations between the EQI and the accuracy of RR measurements. We note that RIIV(EQI sorted) significantly outperformed the three baseline methods, achieving around 3-fold decrease in MAE for the same yield, and also around 3-fold increase in yield for the same MAE. The solid curves in Figure 11 shows that the MAE also increases with EQI. However, the three solid lines are below the corresponding dashed lines with same color, especially for the RIIV and the RIAV. This indicates that when targeting the same yield, using EQI to reject PPG data can have lower MAE than using motion intensity. In another viewpoint, when targeting at the same accuracy, using EQI as the criterion to reject data can have a higher yield. Therefore, EQI is a more accurate indicator of measurement quality than motion intensity, as noise artifacts may be caused by sources other than motion.

The above evaluations demonstrated that our signal processing estimators can provide the flexibility to balance accuracy and yield according to the application requirements. Since the RIIV shows the best result, we focused on the signal processing estimator with RIIV for the following evaluations.

## 5.2 Deep Learning Estimator

In this subsection, we compared the deep learning estimator and the signal processing estimator. The deep learning estimator directly

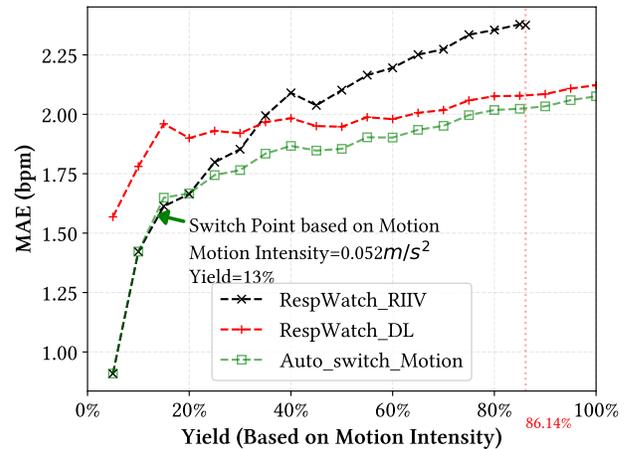


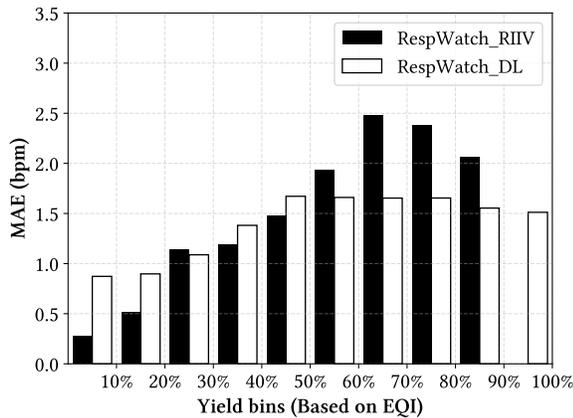
Figure 13: MAE vs. Yield based on Motion ranking.

learns from the waveform and does not rely on any admission control or artifact elimination, so it produced estimations for all data windows, achieving 100% yield. We employed a 5-fold sample-based cross validation (CV) scheme to train and test our deep learning estimator. We ensured there is no PPG waveform overlap between the training and testing set. The out-of-sample error is reported in the evaluation. Figure 12 and 13 plot the MAE-yield curves of the different estimators when the RR measurements are sorted based on EQI and motion intensity, respectively. The EQI is from the signal processing estimator with RIIV. For those 13.86% of data windows that signal processing estimator cannot estimate RR, we assigned an EQI of infinity. We observe that the signal processing estimator (RIIV) achieved lower MAE than the deep learning estimator when the EQI or motion intensity are lower. However, as EQI or motion intensity increases, the deep learning estimator becomes more accurate, suggesting a higher level of robustness against noise artifacts. The MAE dynamic range of deep learning is also not as large as it of signal processing, indicating deep learning is less sensitive with varying noise artifacts. The crossing point of the signal processing and deep learning curves in Figure 12 is at yield of 63%, while it is only at yield of 37% in Figure 13. And the deep learning curve in Figure 12 is relatively smooth compared to it in Figure 13. These once again show that the EQI can indicate the accuracy for RR measurements more accurately and smoothly than the motion intensity even for deep learning estimations.

## 5.3 Hybrid Estimator

For the evaluation of our hybrid estimator, we report the outputs dynamically chosen from signal processing and deep learning based on the best switching point. The best switching point of either EQI or motion intensity was obtained offline via the grid search. For real use, the hybrid estimator automatically switches between the signal processing and deep learning according to the best switching point without human efforts.

We first evaluated the EQI as the switching criterion. The green curve in Figure 12 shows the results of the hybrid estimator with EQI



**Figure 14: MAE in different yield bins with the EQI sorting criterion.**

(*Auto\_switch\_EQI*). It achieves the best MAE (2.017 bpm) compared to both deep learning and signal processing. The best switching point based on the grid search is around  $EQI = 2.3$ , corresponding yield of 53%, which means that the hybrid estimator automatically chooses signal processing when  $EQI \leq 2.3$ , and chooses deep learning when  $EQI > 2.3$ . We further investigated the relationship between MAE and EQI for the signal processing and deep learning estimators, as shown in Figure 14. The data windows were still sorted with the increasing EQI. Each bin contains the data windows with EQI from  $\alpha$ -th to  $(\alpha + 10)$ -th percentiles. We observe that the MAE of *RespWatch\_RIIV* becomes significant higher than *RespWatch\_DL* from the sixth bin, corresponding EQI of range [2.01, 2.45]. The grid search was in a finer granularity, so we found the best switch point of EQI at round 2.3.

Besides, we explored the hybrid estimator switching with motion intensity, using the same grid search approach to find the best switching point of motion intensity. The green curve in Figure 13 shows the results. The hybrid estimator switching with motion intensity demonstrates slightly higher MAE (2.076 bpm vs. 2.017 bpm) than switching with EQI. And it jumped to deep learning earlier at the yield of 13%, which utilizes significantly more times of deep learning outputs. This makes the hybrid estimator with the motion intensity less efficient. As a result, EQI is a better switching criterion in terms of both the accuracy and efficiency. So, in real implementation of RespWatch, we developed our hybrid estimator with the  $EQI = 2.3$  as the switching threshold.

## 6 SYSTEM EXPERIMENTATION

### 6.1 Implementation on Smartwatches

We have implemented RespWatch on Wear OS in mainstream smartwatches. Wear OS [11] is a version of Android operating system tailored for smartwatches and other wearables. For the CNN model in the deep learning estimator, we first trained the model on the

<sup>1</sup><https://www.qualcomm.com/products/snapdragon-processors-wear-2100>

<sup>2</sup><https://www.pixart.com/products-detail/27/PAH8011ES-IN>

<sup>3</sup><https://www.qualcomm.com/products/snapdragon-wear-3100-platform>

**Table 2: Information of the testing smartwatches**

Device	Platform	RAM	System	PPG Sensor
Fossil Gen4	Wear 2100 <sup>1</sup>	512MB	H	PAH8011 <sup>2</sup> (200Hz)
Fossil Sport	Wear 3100 <sup>3</sup>	512MB	H	PAH8011 (100Hz)

server in PyTorch [34] framework, and then transcribed the model into mobile version[39] on Wear OS. For the hybrid estimator, we chose the switching scheme of  $EQI = 2.3$  based on our results in Section 5.3.

### 6.2 Empirical Evaluation

Two smartwatches were used in our empirical evaluation, shown in Table 2. Each experiment was repeated 500 times, and the average of running time and resource usage are reported in Table 3 and 4.

The signal processing estimator was highly efficient with a total running time less than 50 ms (see Table 3), whereas the deep learning estimator had a total running time higher than 6000 ms (as shown in Table 4). The average energy consumption and average CPU utilization were acquired through the Android Profiler [1]. The signal processing estimator consumes less energy with lower CPU utilization. For the deep learning estimator, the CNN model consumed about 98% of the total time, suggesting the need for optimization in the future. For the hybrid estimator based on the best switching threshold on our dataset, the expected running time of switching with EQI was significantly lower than switching with Motion Intensity.

Our results established the feasibility to run RespWatch locally on smartwatches for RR monitoring. Even though the deep learning estimator takes about more than 6 seconds to run per RR measurement, it only needs to be executed every 1 minute for a RR sampling rate of once per minute. The results also highlight our hybrid approach, which only invokes the deep learning estimator in the presence of significant noise artifacts with high EQI. The hybrid method drastically lowers the running time and saves energy while maintaining high accuracy.

## 7 DISCUSSION

We have demonstrated our RespWatch in this study, which outperforms the state-of-the-art baseline methods. The empirical evaluations also quantitatively reveal the execution efficiency and capability of running on the commercial smartwatches. Nevertheless, there are still some room for future improvement.

**RR measurement with excessive motions.** Our RespWatch system has been tested in various activity scenarios involving motions, but it did not cover all scenarios in our daily life. The evaluations in this paper shows the feasibility and high accuracy of RespWatch in situations, like working in front of computer, resting and others activities with motions to some degree. The applicability of RespWatch, especially in scenarios with excessive motions (e.g. running) still needs to be evaluated.

**Inter-individual Difference.** For both the signal processing estimator and deep learning estimator, we applied the same parameters for all the subjects. The differences between individuals (e.g., skin tone and wearing habits) may have impacts on PPG signals. Earlier

**Table 3: Profile of Signal Processing Estimator**

Devices	Preprocessing	Art. Elim.* & Pulse Peak Finding	RIXV* Extraction & Adaptive Peak Finding	Total Time	Ave. CPU (%)	Ave. Energy Consumption
Fossil Gen4 (H)	5.836ms	19.139ms	19.919ms	44.895ms	53.53%	Light to Medium
Fossil sport (H)	5.385ms	16.058ms	16.621ms	38.064ms	50.25%	Light to Medium

\*Art. Elim.: Artifact Elimination

\*RIXV: Respiratory-Induced Variations (RLAV, RIIV, RIFV).

**Table 4: Profile of Deep Learning and Hybrid Estimator**

Devices	Preprocessing	CNN model	Deep learning Total Time	Ave. CPU (%)	Ave. Energy	Hybrid with EQI*	Hybrid with Motion Intensity*
Fossil Gen4 (H)	8.856ms	6504.262ms	6592.828ms	85.34%	Above Medium	2879.811ms	5780.655ms
Fossil sport (H)	8.472ms	7934.962ms	7943.434ms	70.23%	Around Medium	3453.740ms	6948.851ms

\*The running time of hybrid estimator is the expected running time based on our dataset with the corresponding best switching threshold.

studies [18] show the personalized models may mitigate the impact. A potential research direction is to leverage personalized models while minimizing the burden of the personalizing process.

**Detection of Potential Respiratory Diseases.** Currently, our evaluation is limited to healthy volunteers. Further studies are needed to test the feasibility of wearable RR monitoring for patients and exploit the technology to detect respiratory conditions.

## 8 CONCLUSION

RespWatch is a wearable sensing system for robust RR monitoring on smartwatches with PPG. We explored signal processing, deep learning and hybrid approaches to measure RR based on PPG signals. We collected a large dataset from 30 participants who performed multiple activities with the smartwatches. The signal processing estimator achieved higher accuracy in the presence of moderate noise artifacts, while the deep learning estimator was more robust to significant noise artifacts. Given the complementary strengths, we developed a novel hybrid estimator that can automatically switch between the signal processing and deep learning based on the EQI. The hybrid estimator not only achieved the best accuracy but also leveraged the high efficiency.

## ACKNOWLEDGMENTS

This research was supported in part by a grant-in-aid from the Division of Clinical and Translational Research of the Department of Anesthesiology, Washington University School of Medicine, St. Louis, Missouri; and in part, by the Healthcare Innovation Lab and the Institute for Informatics at BJC HealthCare and Washington University School of Medicine. Special thanks to Linda Yun, who was the study coordinator and helped to collect all the data.

## REFERENCES

- [1] Android Developers 2020. *Inspect energy use with Energy Profiler*. Retrieved 2020-06-12 from <https://wearos.google.com/#hands-free-help>
- [2] Sourav Bhattacharya and Nicholas D Lane. 2016. From smart to deep: Robust activity recognition on smartwatches using deep learning. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, 1–6.
- [3] A Bilgaiyan, R Sugawara, F Elsamnah, C Shim, Md Affiq, and Reiji Hattori. 2018. Optimizing performance of reflectance-based organic Photoplethysmogram (PPG) sensor. In *Organic and Hybrid Sensors and Bioelectronics XI*, Vol. 10738. International Society for Optics and Photonics, 1073808.
- [4] Daniel J Buysse, Charles F Reynolds, Timothy H Monk, Susan R Berman, David J Kupfer, et al. 1989. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry res* 28, 2 (1989), 193–213.
- [5] Marta E Cecchinato, Anna L Cox, and Jon Bird. 2015. Smartwatches: the Good, the Bad and the Ugly?. In *Proceedings of the 33rd Annual ACM Conference extended abstracts on human factors in computing systems*. 2133–2138.
- [6] Peter H Charlton, Drew A Birrenkott, Timothy Bonnici, Marco AF Pimentel, Alistair EW Johnson, Jordi Alastruey, Lionel Tarassenko, Peter J Watkinson, Richard Beale, and David A Clifton. 2017. Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review. *IEEE reviews in biomedical engineering* 11 (2017), 2–20.
- [7] Peter H Charlton, Timothy Bonnici, Lionel Tarassenko, David A Clifton, Richard Beale, and Peter J Watkinson. 2016. An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. *Physiological measurement* 37, 4 (2016), 610.
- [8] Ioannis Dologlou and George Carayannis. 1989. Pitch detection based on zero-phase filtering. *Speech Communication* 8, 4 (1989), 309–318.
- [9] Victoria B Egizio, Michael Eddy, Matthew Robinson, and J Richard Jennings. 2011. Efficient and cost-effective estimation of the influence of respiratory variables on respiratory sinus arrhythmia. *Psychophysiology* 48, 4 (2011), 488–494.
- [10] John F Fieselmann, Michael S Hendryx, Charles M Helms, and Douglas S Wakefield. 1993. Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients. *Journal of general internal medicine* 8, 7 (1993), 354–360.
- [11] Google 2020. *Wear OS by Google Smartwatches*. Retrieved 2020-06-12 from <https://wearos.google.com/#hands-free-help>
- [12] Nils Yannick Hammerla, James Fisher, Peter Andras, Lynn Rochester, Richard Walker, and Thomas Plötz. 2015. PD disease state assessment in naturalistic environments using deep learning. In *Twenty-Ninth AAAI conference on artificial intelligence*.
- [13] Tian Hao, Chongguang Bi, Guoliang Xing, Roxane Chan, and Linlin Tu. 2017. MindfulWatch: A smartwatch-based system for real-time respiration monitoring during meditation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–19.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*. 1026–1034.
- [15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [16] Javier Hernandez, Daniel McDuff, and Rosalind W Picard. 2015. Biowatch: estimation of heart and breathing rates from wrist motions. In *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. IEEE, 169–176.
- [17] Alberto Hernando, Jesus Lazaro, Eduardo Gil, Adriana Arza, Jorge Mario Garzón, Raül Lopez-Anton, Concepcion de la Camara, Pablo Laguna, Jordi Aguiló, and Raquel Bailón. 2016. Inclusion of respiratory frequency information in heart rate variability analysis for stress assessment. *IEEE journal of biomedical and health*

- informatics* 20, 4 (2016), 1016–1025.
- [18] Sinh Huynh, Rajesh Krishna Balan, JeongGil Ko, and Youngki Lee. 2019. VitaMon: measuring heart rate variability using smartphone front camera. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*. 1–14.
- [19] Delaram Jarchi, Dario Salvi, Lionel Tarassenko, and David A Clifton. 2018. Validation of instantaneous respiratory rate using reflectance PPG from different body positions. *Sensors* 18, 11 (2018), 3705.
- [20] Zhenhua Jia, Amelie Bonde, Sugang Li, Chenren Xu, Jingxian Wang, Yanyong Zhang, Richard E Howard, and Pei Zhang. 2017. Monitoring a person's heart rate and respiratory rate on a shared bed using geophones. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. 1–14.
- [21] Walter Karlen, J Mark Ansermino, and Guy Dumont. 2012. Adaptive pulse segmentation and artifact detection in photoplethysmography for mobile applications. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 3131–3134.
- [22] Walter Karlen, Srinivas Raman, J Mark Ansermino, and Guy A Dumont. 2013. Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering* 60, 7 (2013), 1946–1953.
- [23] Chandan Karmakar, Ahsan Khandoker, Thomas Penzel, Christoph Schöbel, and Marimuthu Palaniswami. 2013. Detection of respiratory arousals using photoplethysmography (PPG) signal in sleep apnea patients. *IEEE journal of biomedical and health informatics* 18, 3 (2013), 1065–1073.
- [24] Boon Giin Lee, Jae-Hee Park, Chuan Chin Pu, and Wan-Young Chung. 2015. Smartwatch-based driver vigilance indicator with kernel-fuzzy-C-means-wavelet method. *IEEE sensors journal* 16, 1 (2015), 242–253.
- [25] Daniyal Liaqat, Mohamed Abdalla, Pegah Abed-Esfahani, Moshe Gabel, Tatiana Son, Robert Wu, Andrea Gershon, Frank Rudzicz, and Eyal De Lara. 2019. Wear-Breathing: Real World Respiratory Rate Monitoring Using Smartwatches. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 1–22.
- [26] Yue-Der Lin, Ya-Hsueh Chien, and Yi-Sheng Chen. 2017. Wavelet-based embedded algorithm for respiratory rate estimation from PPG signal. *Biomedical Signal Processing and Control* 36 (2017), 138–145.
- [27] Sally K Longmore, Gough Y Lui, Ganesh Naik, Paul P Breen, Bin Jalaludin, and Gaetano D Gargiulo. 2019. A comparison of reflective photoplethysmography for detection of heart rate, blood oxygen saturation, and respiration rate at various anatomical locations. *Sensors* 19, 8 (2019), 1874.
- [28] David J Meredith, D Clifton, P Charlton, J Brooks, CW Pugh, and L Tarassenko. 2012. Photoplethysmographic derivation of respiratory rate: a review of relevant physiology. *Journal of medical engineering & technology* 36, 1 (2012), 1–7.
- [29] K Mostov, E Liptson, and R Boutchko. 2010. Medical applications of shortwave FM radar: Remote monitoring of cardiac and respiratory motion. *Medical physics* 37, 3 (2010), 1332–1338.
- [30] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Jacob E Sunshine. 2019. Opioid overdose detection using smartphones. *Science translational medicine* 11, 474 (2019).
- [31] Shamim Nemati, Atul Malhotra, and Gari D Clifford. 2010. Data fusion for improved respiration rate estimation. *EURASIP journal on advances in signal processing* 2010, 1 (2010), 926305.
- [32] Christina Orphanidou, Timothy Bonnici, Peter Charlton, David Clifton, David Vallance, and Lionel Tarassenko. 2014. Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring. *IEEE journal of biomedical and health informatics* 19, 3 (2014), 832–838.
- [33] Jaeyeon Park, Woojin Nam, Jaewon Choi, Taeyeong Kim, Dukyong Yoon, Sukhoon Lee, Jeongyeup Paek, and JeongGil Ko. 2017. Glasses for the third eye: Improving the quality of clinical data analysis with motion sensor-based data filtering. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. 1–14.
- [34] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems* 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.). Curran Associates, Inc., 8024–8035. <http://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- [35] Neal Patwari, Joey Wilson, Sai Ananthanarayanan, Sneha K Kasera, and Dwayne R Westenskow. 2013. Monitoring breathing via signal strength in wireless networks. *IEEE Transactions on Mobile Computing* 13, 8 (2013), 1774–1786.
- [36] Marco V Perez, Kenneth W Mahaffey, Haley Hedlin, John S Rumsfeld, Ariadna Garcia, Todd Ferris, Vidhya Balasubramanian, Andrea M Russo, Amol Rajmane, Lauren Cheung, et al. 2019. Large-scale assessment of a smartwatch to identify atrial fibrillation. *New England Journal of Medicine* 381, 20 (2019), 1909–1917.
- [37] Marco AF Pimentel, Alistair EW Johnson, Peter H Charlton, Drew Birrenkott, Peter J Watkinson, Lionel Tarassenko, and David A Clifton. 2016. Toward a robust estimation of respiratory rate from pulse oximeters. *IEEE Transactions on Biomedical Engineering* 64, 8 (2016), 1914–1923.
- [38] Marco AF Pimentel, Alistair EW Johnson, Peter H Charlton, Drew Birrenkott, Peter J Watkinson, Lionel Tarassenko, and David A Clifton. 2016. Toward a robust estimation of respiratory rate from pulse oximeters. *IEEE Transactions on Biomedical Engineering* 64, 8 (2016), 1914–1923.
- [39] PyTorch 2020. *PyTorch Mobile*. Retrieved 2020-06-12 from <https://pytorch.org/mobile/home/>
- [40] Valentin Radu, Nicholas D Lane, Sourav Bhattacharya, Cecilia Mascolo, Mahesh K Marina, and Fahim Kawsar. 2016. Towards multimodal deep learning for activity recognition on mobile devices. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. 185–188.
- [41] Satu Rajala, Harri Lindholm, and Tapio Taipalus. 2018. Comparison of photoplethysmogram measured from wrist and finger and the effect of measurement location on pulse arrival time. *Physiological measurement* 39, 7 (2018), 075010.
- [42] M Raghu Ram, K Venu Madhav, E Hari Krishna, Nagarjuna Reddy Komalla, and K Ashoka Reddy. 2011. A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter. *IEEE Transactions on Instrumentation and Measurement* 61, 5 (2011), 1445–1457.
- [43] Vignesh Ravichandran, Balamurali Murugesan, Vaishali Balakarthikeyan, Keerthi Ram, SP Preejith, Jayaraj Joseph, and Mohanasankar Sivaprakasam. 2019. RespNet: A deep learning model for extraction of respiration from photoplethysmogram. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 5556–5559.
- [44] Andrew Reisner, Phillip A Shaltis, Devin McCombie, and H Harry Asada. 2008. Utility of the photoplethysmogram in circulatory monitoring. *Anesthesiology: The Journal of the American Society of Anesthesiologists* 108, 5 (2008), 950–958.
- [45] Bersain A Reyes, Natasa Reljin, Youngsun Kong, Yunyoung Nam, and Ki H Chon. 2016. Tidal volume and instantaneous respiration rate estimation using a volumetric surrogate signal acquired via a smartphone camera. *IEEE journal of biomedical and health informatics* 21, 3 (2016), 764–777.
- [46] Yichen Shen, Maxime Voisin, Alireza Aliamiri, Anand Avati, Awni Hannun, and Andrew Ng. 2019. Ambulatory Atrial Fibrillation Monitoring Using Wearable Photoplethysmography with Deep Learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1909–1916.
- [47] Hang Sik Shin, Chungkeun Lee, and MyoungHo Lee. 2009. Adaptive threshold method for the peak detection of photoplethysmographic waveform. *Computers in biology and medicine* 39, 12 (2009), 1145–1152.
- [48] Julius Orion Smith. 2007. *Introduction to digital filters: with audio applications*. Vol. 2. Julius Smith.
- [49] Xiao Sun, Li Qiu, Yibo Wu, Yeming Tang, and Guohong Cao. 2017. Sleepmonitor: Monitoring respiratory rate and body position during sleep using smartwatch. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–22.
- [50] John Trimppop, Hannes Schenk, Gerald Bieber, Friedrich Lämmel, and Paul Burggraf. 2017. Smartwatch based respiratory rate and breathing pattern recognition in an end-consumer environment. In *Proceedings of the 4th international Workshop on Sensor-based Activity Recognition and Interaction*. 1–5.
- [51] Terry T Um, Franz MJ Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulic. 2017. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 216–220.
- [52] Siying Wang, Antje Pohl, Timo Jaeschke, Michael Czaplík, Marcus Köny, Steffen Leonhardt, and Nils Pohl. 2015. A novel ultra-wideband 80 GHz FMCW radar system for contactless monitoring of vital signs. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 4978–4981.
- [53] Xuyu Wang, Chao Yang, and Shiwen Mao. 2017. TensorBeat: Tensor decomposition for monitoring multiperson breathing beats with commodity WiFi. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 1 (2017), 1–27.
- [54] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. Breathlistener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. 54–66.
- [55] Mari Zakrzewski, Antti Vehkaoja, Atte S Joutsen, Karri T Palovuori, and Jukka J Vanhala. 2015. Noncontact respiration monitoring during sleep with microwave Doppler radar. *IEEE Sensors Journal* 15, 10 (2015), 5683–5693.