*Application of Information Technology* ■

# A Method for Automatic Identification of Reliable Heart Rates Calculated from ECG and PPG Waveforms

Chenggang Yu, PhD, Zhenqiu Liu, PhD, Thomas McKenna, PhD, Andrew T. Reisner, MD, Jaques Reifman, PhD

**Abstract** **Objective:** The development and application of data-driven decision-support systems for medical triage, diagnostics, and prognostics pose special requirements on physiologic data. In particular, that data are reliable in order to produce meaningful results. The authors describe a method that automatically estimates the reliability of reference heart rates (HRr) derived from electrocardiogram (ECG) waveforms and photoplethysmogram (PPG) waveforms recorded by vital-signs monitors. The reliability is quantitatively expressed through a quality index (QI) for each HRr.

**Design:** The proposed method estimates the reliability of heart rates from vital-signs monitors by (1) assessing the quality of the ECG and PPG waveforms, (2) separately computing heart rates from these waveforms, and (3) concisely combining this information into a QI that considers the physical redundancy of the signal sources and independence of heart rate calculations. The assessment of the waveforms is performed by a Support Vector Machine classifier and the independent computation of heart rate from the waveforms is performed by an adaptive peak identification technique, termed ADAPIT, which is designed to filter out motion-induced noise.

**Results:** The authors evaluated the method against 158 randomly selected data samples of trauma patients collected during helicopter transport, each sample consisting of 7-second ECG and PPG waveform segments and their associated HRr. They compared the results of the algorithm against manual analysis performed by human experts and found that in 92% of the cases, the algorithm either matches or is more conservative than the human's QI qualification. In the remaining 8% of the cases, the algorithm infers a less conservative QI, though in most cases this was because of algorithm/human disagreement over ambiguous waveform quality. If these ambiguous waveforms were relabeled, the misclassification rate would drop from 8% to 3%.

**Conclusion:** This method provides a robust approach for automatically assessing the reliability of large quantities of heart rate data and the waveforms from which they are derived.

■ **J Am Med Inform Assoc.** 2006;13:309–320. DOI 10.1197/jamia.M1925.

Decision-support algorithms that automatically interpret streaming physiologic time-series data are valuable tools for a broad range of medical surveillance applications. Examples of such applications include acute monitoring of patients in intensive care, home care, and ad hoc monitoring to continuously assess the health status of personnel, such as firefighters and soldiers, who are at risk of sudden injury.[1] Advances in vital-signs monitoring software/hardware, miniaturization, storage capacity, wireless transmission, and computational power now allow recording and analysis of large quantities of physiologic data in a timely fashion. These data are invaluable for the development of triage, diagnostic, and prognostic algorithms. However, collection of time-series vital-signs data is subject to many factors that affect the quality of the data. In particular, because vital-signs data are mostly collected in a noninvasive fashion, sensor motion artifact is of significant concern when the subject is moving or being transported. Other factors that may degrade data quality include electrical interference, sensor/monitor malfunction, and poor sensor placement on the subject. If valid decision-support algorithms are to be developed, and subsequently used to monitor patients, it is critical that reliable data be distinguished from artifact. Moreover, the process of distinguishing reliable from unreliable data must be automated since the sheer volume of collected time-series vital-signs data makes post hoc manual assessment an overwhelming task, while real-time streaming data cannot be manually evaluated at all.

Heart rate (HR) is a critical vital sign that is continuously monitored during transport of trauma patients from the scene

Affiliations of the authors: Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, Fort Detrick, MD (CY, ZL, TM, JR); Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA (ATR); Dr. Liu is currently with the Division of Biostatistics, Greenebaum Cancer Center and Department of Epidemiology and Preventive Medicine, University of Maryland Medical Center.

The work presented here was supported by the U.S. Army Medical Research and Materiel Command, Fort Detrick, MD.

The authors express their gratitude to Col. John Holcomb and Dr. Jose Salinas of the U.S. Army Institute of Surgical Research, Fort Sam Houston, San Antonio, TX, who provided the trauma patient data.

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or the U.S. Department of Defense.

Correspondence and reprints: Jaques Reifman, PhD, U.S. Army Medical Research and Materiel Command MRMC/TATRC, 504 Scott Street, Ft. Detrick, MD 21702-5012; e-mail: <jaques.reifman@us.army.mil>.

Received for review: 08/01/05; accepted for publication: 01/16/06.

of injury to the hospital. It is used as an input for existing prehospital trauma severity scores, such as the prehospital index,[2,3] and may be used for future triage scoring systems. Also, studies of heart rate variability (HRV) suggest that decreasing HRV may be associated with worsening patient status. Unfortunately, we have observed that randomly imposed noise spikes are sometimes counted as heart beats by a vital-signs monitor. These sorts of data corruption can mislead diagnosis and compromise the development and application of inductive algorithms based on the synthesis of time-series physiologic data. Therefore, it is imperative that validated HRs be available for clinical use and development of advanced automated monitoring systems.

Automated HR calculation is usually based on the identification of heart beat signals, which could be taken from the QRS complex or simply the R waves in electrocardiogram (ECG) waveforms, or the pulse waves in photoplethysmogram (PPG) waveforms,[4–6] and dependent on the count of heart beats over a period of time. Given noisy waveforms, however, true heart beat signals may be masked or noise artifacts may resemble and be counted as true heart beats. Therefore, the quality of the HR calculated from the waveform depends on the quality of the waveform, making the qualification of waveforms a necessary step in validating HRs provided by a vital-signs monitor. Here, we refer to the monitor-calculated HRs as reference HRs (HRr). Accordingly, such HRr can be categorized as unreliable when the associated waveform is determined to be of suboptimal quality. For a conservative validation method, a high standard for good-quality waveforms is preferred to minimize the possibility that bad-quality HRs are falsely categorized as good. However, an overly stringent threshold is not advisable since it will increase the chance that good-quality HRs are falsely categorized as bad and, for post hoc data analysis, will considerably reduce the amount of available good-quality HR for the development of data-driven, decision-support algorithms.

In this paper, we present an approach to automatically and systematically qualify ECG HRr and PPG HRr provided by a vital-signs monitor. We assume that the monitor also provides the corresponding waveforms from which they are derived and that the monitored individuals are alive and have been subject to a trauma injury, where arrhythmia is seldom observed. The approach numerically qualifies each sampled HRr by assigning to it a quality index (QI) that concisely expresses its reliability. The approach exploits the physical redundancy provided by ECG HRr and PPG HRr and employs an independent method for recomputing HRs from the provided waveforms. This work addresses the first and key step of automatic and systematic qualification of large amounts of time-series data of our trauma database, so that we can next address our ultimate goal: mining these data to find predictive information for some clinical outcome.

Figure 1 illustrates the three components of the approach. In the first component, we use the newly developed adaptive peak identification technique, termed ADAPIT, to independently compute HRs (HRc) from both ECG and PPG waveform segments corresponding to the HRr we wish to validate. ADAPIT is a computationally simple peak detection algorithm, yet robust in the presence of random, motion-

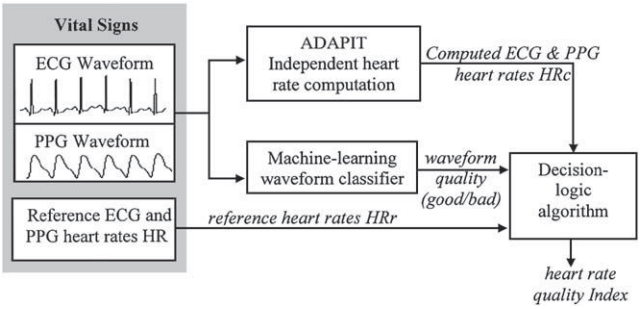


**Figure 1.** The three elements of the algorithm used to infer a quality index for reference heart rates provided by a vital-signs monitor.

accounted for, these noise spikes are likely to be counted as heart beats by the vital-signs monitor. Next, we separately qualify ECG waveform segments and PPG waveform segments as either good (excellent quality) or bad (suboptimal quality) through the use of a machine-learning algorithm in the form of support vector machines (SVMs).[7] In the third and final component, through a decision-logic algorithm, we combine the results of the two previous steps, the ADAPIT-computed ECG HRc and PPG HRc and the quality of their corresponding waveform segments, and compare them against ECG HRr and PPG HRr provided by a vital-signs monitor to infer a QI for the two HRr. A QI is inferred each time a HRr is provided by a vital-signs monitor and ranges from zero to three, with three representing the best-possible quality. In the absence of one of the waveforms, the decision–logic algorithm still provides a QI by assuming that the absent signal is present but possesses poor quality. Should additional HR sources be available, the approach could be extended by properly accounting for the quality of the new signal information and modifying the QI decision rules.

The approach is modular, self-contained, and independent of the data collection hardware. The waveform qualification algorithm (SVM), the HR recomputation algorithm (ADAPIT), and the QI decision rules are developed independently of each other and can be separately exchanged by functionally equivalent modules based on other methods. The three components form an effective, stand-alone system to validate reference HRs. Our approach is simply based on recorded time-series data from a vital-signs monitor, which is taken as a black box. From this point of view, the approach is independent of the data collection hardware.

## Methods

In this section, we briefly describe the three components depicted in Figure 1: the HR estimation via the ADAPIT algorithm, the waveform qualification via an SVM algorithm, and the QI determination. We start by describing the data that precipitated the development of these components and that are used for the synthesis and testing of our algorithms.

### Data

This study is based on physiologic time-series data collected during transport of trauma patients from the scene of injury by helicopter service to the Level I unit at the Memorial Hermann Hospital in Houston, TX.[8,9] The data were collected by ProPaq 206EL vital-signs monitors[10] on the helicopters

PPG waveform signals and their corresponding monitor-calculated HRr. The time series sampling rates are approximately 182 Hz for the ECG waveform, 91 Hz for the PPG waveform, and 1 Hz for the HRr. Complete vital-signs data for a total of 726 patients were deposited into our Physiology Analysis System,[11] which provides curated data and the ability to query and analyze discrete and time-series data over the Internet with a Web browser. The patient population is composed of 538 males and 186 females (two genders not noted), with a mean age of 37.7 years. The predominant type of injury is blunt trauma (641 patients), followed by penetrating trauma (78 patients).

### Heart Rate Estimation with the ADAPIT Algorithm

The first component of our approach is the independent estimation of ECG and PPG HRs from their corresponding high-frequency waveforms. While we acknowledge that a large body of work has been developed over the past two decades,[4–6] most of the approaches are rather involved because they are designed to accommodate irregular morphologies and irregular rhythms, even though such phenomena are rarely observed in our data set of trauma victims. Due to the ambulatory nature and dynamic environment in which trauma data are collected, the major challenge is the filtering of noise and artifacts in the waveforms. Furthermore, most approaches are limited to the estimation of ECG-derived HRs through the detection and analysis of the QRS complex,[6] while we also need to estimate PPG-derived HRs. To achieve these objectives, we developed the ADAPIT algorithm. ADAPIT is a generic algorithm that, through changes in parameter settings and one computational step, is equally applicable to the estimation of HRs from both ECG and PPG waveforms and is designed to filter out noise and artifacts so they are not counted as heart beats. ADAPIT, however, may have limited ability to compute HRs in settings of highly irregular rhythms.

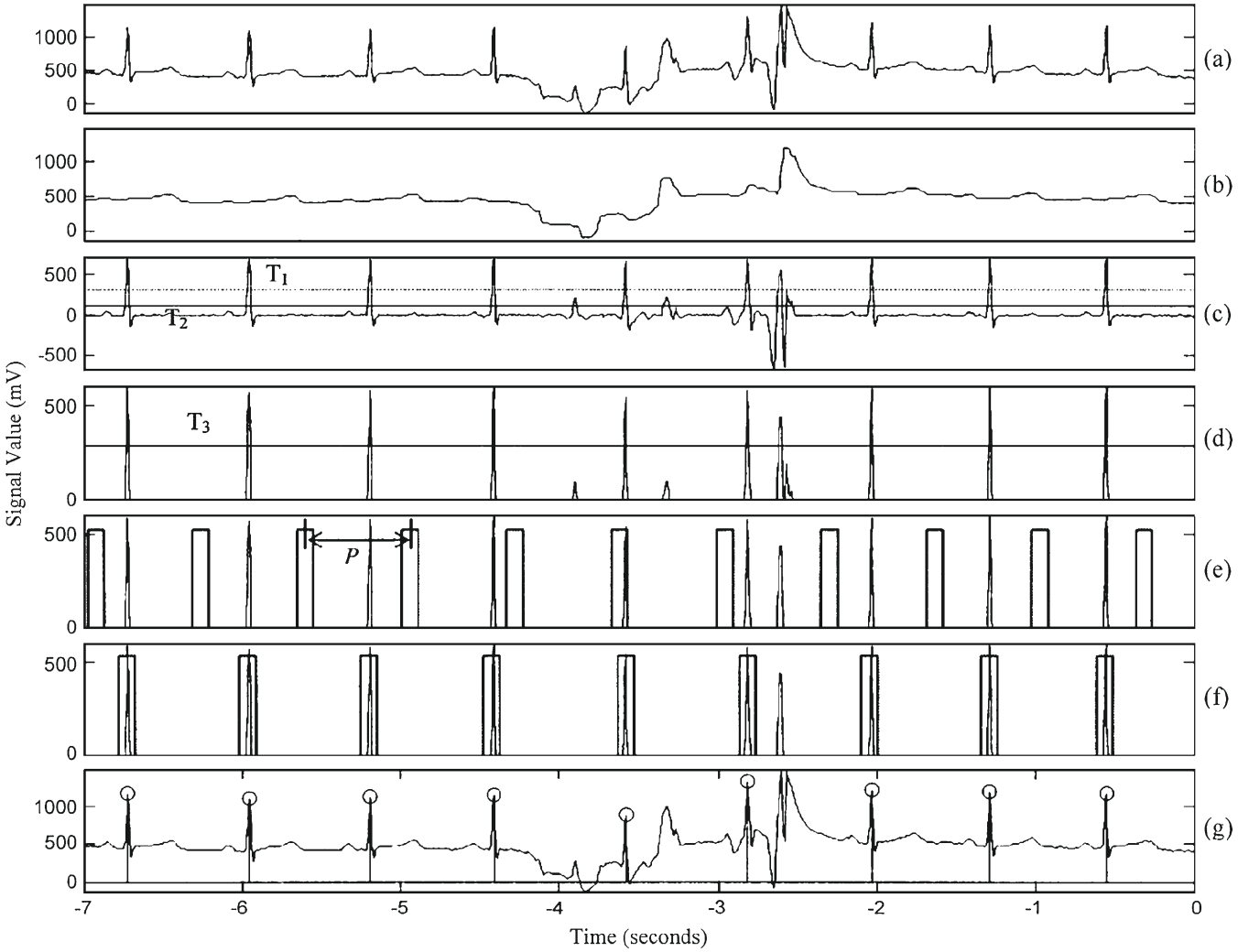


**Figure 2.** Illustration of the identification of heart beats by the ADAPIT algorithm. (*a*) Original 7-second ECG waveform segment. (*b*) Waveform after application of a median filter. (*c*) Difference of the original waveform in *a* minus the median-filtered waveform in *b*. The threshold $T_1$ defines the segment's baseline range $[-T_1, T_1]$ and the threshold $T_2$ provides a first cut on the lower limit of the peaks' magnitude. (*d*) The first estimates of the actual peaks and threshold $T_3$ (*horizontal line*) are used to eliminate small-magnitude spikes that clearly are not actual peaks. (*e*) String of markers with constant period *P*. (*f*) Best alignment

*Estimation from Electrocardiogram Waveforms*

The ADAPIT algorithm computes an HRc at each time point (i.e., each second) $t$ that a HRr is provided by the vital-signs monitor. This computation is performed based on a 7-second ECG waveform from time $t$-7 to $t$, which is approximately the same waveform length used by the vital-signs monitor,[10] to estimate one HRr. Figure 2 illustrates the four major steps of the algorithm to compute HRc at $t = 0$ (see Appendix 1 for additional technical details).

*Step 1.* ADAPIT applies a median filter (with a 55-ms window size) to the original 7-second waveform (Fig. 2a) and then subtracts the filtered signal (Fig. 2b) from the original one to yield the waveform in Figure 2c. This step de-trends the waveform, retains the amplitude of sharp R waves, and attenuates broad waves, such as the P wave and T wave.

*Step 2.* This step provides a first estimate of the actual peaks of the waveform through the sequential computation of two thresholds, $T_1$ and $T_2$. $T_1$, illustrated in Figure 2c, is taken as $2\sigma_1$, where $\sigma_1$ denotes the standard deviation of all data point values of the 7-second waveform and defines the segment's baseline range $[-T_1, T_1]$, from which the baseline standard deviation $\sigma_2$ is calculated. $T_2$, set to $3\sigma_2$, is used as a lower limit of the waveform amplitude for considering potential peaks. Peaks greater than $T_2$ are taken as the first estimate of the actual peaks (Fig. 2d).

*Step 3.* To eliminate small-amplitude spikes that clearly are not R waves, a threshold $T_3$ is defined as one half of the median amplitude of all peaks identified in Step 2 (Fig. 2d). All peaks less than $T_3$ are eliminated, as illustrated in Figure 2e.

*Step 4.* To determine actual R waves from the peaks retained in Step 3, strings of markers with period $P$ (Fig. 2e) are iteratively generated and moved along the time line to align with the retained peaks. Through this iterative process, $P$ is modified to range from lengths equivalent to HRs between 25 and 250 beats per minute (bpm). The string with the largest P aligned to the largest number of retained peaks is selected. Next, each unaligned marker of the selected string is allowed to move back and forth along the time line by as much as one half of $P$ in an attempt to line up any unaligned peak (Fig. 2f). Finally, all aligned peaks, marked with circles on the original ECG waveform in Figure 2g, are assumed to be actual R waves. It should be noted that ADAPIT computes HRc based on all markers rather than the aligned peaks because an R wave could have been dropped during data collection or filtered out during the ADAPIT four-step process.

To verify ADAPIT's capability to filter out motion-induced artifacts and correctly compute HR of ambulatory trauma victims, we had a human expert visually estimate the HR of 80 seven-second, good-quality waveform samples from our database. Considering the human's estimations as the gold standard, we compare them against ADAPIT, HRr, and a well-established QRS-based detection program termed ecgpuwave.[12]

Figure 3 shows the difference between the algorithms' and the human's estimations for each of the 80 samples. The mean differences of ADAPIT, HRr, and ecgpuwave are, respectively, −0.62, 0.78, and 1.03 bpm, and the root mean square differences are 7.1, 5.1, and 7.1 bpm, respectively. These results

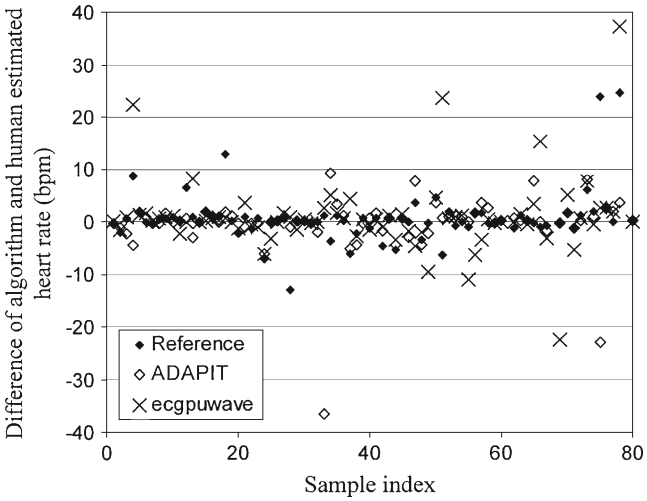


**Figure 3.** Difference in heart rates computed by three different algorithms (ADAPIT, reference heart rate [HRr], and ecgpuwave) and a human expert.

indicate that simple measures of filtering out motion artifacts

HRs, while the two other algorithms tend to overestimate them. This feature of ADAPIT is noticed, in particular, in waveforms with highly irregular rhythms (samples 33 and 76) and provides a lower bound estimate for the HRs that allows for a conservative consistency check (larger delta) between HRr and HRc.

*Estimation from Photoplethysmogram Waveforms*

ADAPIT employs the same four-step process with two small modifications in the estimation of PPG-derived HRc. First, in Step 1, the median filter window size is extended to 550 ms to preserve broad pulse waves and attenuate sharp dicrotic notches. Second, after the identification of peaks in Step 3, each peak is smoothed with a moving-average filter of window size equal to 110 ms. This additional filtering is needed to smooth out the broad and often distorted pulse waves and reduce the ambiguity in detecting the exact time of a heart beat, assumed to occur when the smoothed pulse wave reaches its maximum.

*Waveform Qualification*

This component of the approach implements our premise that the reliability of HRr is highly dependent on the quality of the underlying waveforms from which they are derived. A machine learning classifier, implemented by an SVM, automates the categorization of waveforms by attempting to mimic the performance of human experts who rely on visual inspection and the application of some implicit or explicit rules of thumb. A classifier "learns" these rules by finding coefficients that optimize the "correlations" between a set of waveform-extracted features and waveform quality obtained from manually categorized waveform samples.

Figure 4 illustrates the four steps in the development of a machine-learning classifier: (1) manually categorize sample waveform segments, (2) define candidate waveform features that distinguish good/bad waveforms, (3) select the most informative features, and (4) train and test the classifier. Once trained and given input features, the classifier categorizes waveform segments as being good or bad.

*Manual Waveform Categorization*

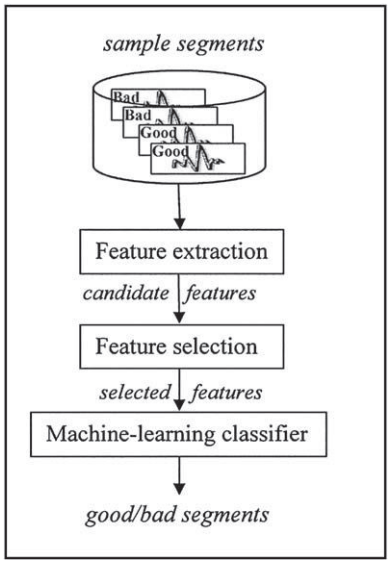


**Figure 4.** The development of machine-learning classifiers requires (1) manual categorization of good/bad waveform-segment samples, (2) definition and extraction of candidate waveform features, (3) selection of the most discriminatory features, and (4) training and testing of the machine-learning classifier. Once trained and given input features, the classifier categorizes waveform segments as being good or bad.

ECG samples and 388 PPG samples randomly selected from different patients. Of these, 194/168 ECG samples and 180/208 PPG samples were categorized as good/bad based on the following rules:

> *An ECG segment is ranked as bad (suboptimal) if more than one expected R wave is not observed or if the R wave is indistinguishable from noisy peaks. Otherwise it is ranked as good. A PPG segment is ranked as bad (suboptimal) if more than one expected pulse wave is not observed or if any one pulse wave peak cannot be distinguished from a dicrotic notch. Otherwise it is ranked as good.*

These rules express the hypothesis that if more than one heart beat signal in a 7-second waveform segment is ambiguous, the HR calculated from such segment may be inaccurately extrapolated. The rules are conservative by design so that the inductively constructed classifiers are equally conservative and attempt to ensure that even if the classifier produces occasional false good waveform evaluations, those false good waveforms will still be of sufficient quality for estimation of HRs.

### *Candidate Waveform Features*

A key phase in the development of machine-learning classifiers involves the definition and extraction of candidate features that can be used as class discriminators. For the characterization of waveforms as good or bad, we define three features in the frequency domain from ECG waveforms and three features in the time domain from ECG and PPG waveforms. Their definitions are presented in Appendix 2.

Similar to the ADAPIT algorithm, we extract features from 7-second waveform segments that immediately precede each HRr we wish to qualify. The three frequency-domain features [illegible] (HFE) [illegible] the discrete-time fast Fourier transform[13] to the ECG time-series data. These features are designed to exclude ECG frequency components that are associated with a QRS complex, while capturing high- and low-frequency component characteristics that may be attributed to noise and baseline drifts and shifts.

The first time-domain feature is the fraction of aligned waves FW, which provides a measure of temporal regularity of potential heart beat signals. The second time-domain feature is a specific signal-to-noise ratio SN, which provides a measure of the distinctiveness of potential heart beat signals above the baseline. The pulse-wave variability (PV), extracted from PPG waveform segments, is the third time-domain feature and provides a measure of the variability of the time interval between two adjacent pulse waves.

### *Feature Selection*

The goal of automatic feature selection is to choose and retain a subset of salient features from the original list of candidate features such that the process of pattern discovery by the machine-learning classifier is implemented in a reduced space without degrading its performance. The underlying philosophy is to retain features that can clearly characterize or discriminate the quality of the waveforms and eliminate features that are redundant, and hence, do not contribute additional information. Here, we employ information entropy[14,15] as a measure of discriminatory power of the features. The most discriminatory (informative) feature has the lowest entropy.

Our previously developed Rule Generator (RG) program[14,15] is used to compute entropies of candidate ECG and PPG waveform features. The RG program also defines patterns formed by these features and populated by the previously characterized samples to discriminate good/bad waveforms. The features that characterize the most discriminatory patterns, defined as the patterns that discriminate the largest number of samples, are selected as the most informative. Through this procedure, we find that HFE, FW, and SN are the most discriminatory features for ECG waveform classification and that FW and PV are the most informative features for PPG waveform classification.

### *Support Vector Machine Classifier*

In this study, we employ our previously developed version of an SVM algorithm[16] to classify ECG and PPG waveforms. The SVM, a recently proposed supervised machine-learning algorithm,[7] has been shown to be an effective classifier in a wide variety of applications, including the categorization of ECG data.[17–20] As a supervised-learning algorithm, the development (or "training") of an SVM requires a set of input/output training samples, where the inputs consist of a list of discriminatory features, such as the three ECG features and two PPG features selected in the previous section, and the outputs consist of labeled binary classes, good and bad. Once trained to implicitly "learn" the "rules" embedded in the training samples, given the values of the input features, extracted from a waveform segment that we wish to classify, the SVM automatically categorizes the segment as good or bad. An in-depth description of SVMs can be found in Vapnik.[7]

DOCKET ALARM

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

**LAW FIRMS**
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

**FINANCIAL INSTITUTIONS**
Litigation and bankruptcy checks for companies and debtors.

**E-DISCOVERY AND LEGAL VENDORS**
Sync your system to PACER to automate legal marketing.