Available online at www.sciencedirect.com
SciVerse ScienceDirect
Journal of Electrocardiology xx (2012) xxx – xxx
JOURNAL OF Electrocardiology
www.jecgonline.com

Signal quality and data fusion for false alarm reduction in the intensive care unit

Qiao Li, PhD,[a,b] Gari D. Clifford, PhD[b,*]

[a] Institute of Biomedical Engineering, School of Medicine, Shandong University, Jinan, Shandong, China
[b] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK

Received 30 March 2012

Abstract Due to a lack of integration between different sensors, false alarms (FA) in the intensive care unit (ICU) are frequent and can lead to reduced standard of care. We present a novel framework for FA reduction using a machine learning approach to combine up to 114 signal quality and physiological features extracted from the electrocardiogram, photoplethysmograph, and optionally the arterial blood pressure waveform. A machine learning algorithm was trained and evaluated on a database of 4107 expert-labeled life-threatening arrhythmias, from 182 separate ICU visits. On the independent test data, FA suppression results with no true alarm (TA) suppression were 86.4% for asystole, 100% for extreme bradycardia and 27.8% for extreme tachycardia. For the ventricular tachycardia alarms, the best FA suppression performance was 30.5% with a TA suppression rate below 1%. To reduce the TA suppression rate to zero, a reduction in FA suppression performance to 19.7% was required.

Keywords: False alarm reduction; Signal quality assessment; Genetic algorithm; Relevance vector machine; Intensive care unit

Introduction

False cardiac monitor alarm rates in the intensive care unit (ICU) are extremely frequent, and can be up to 95% for some types of alarms.[1] Since the publication by Lawless[2] on the "crying wolf" phenomenon in 1994, the unfortunate reality is that not much has changed over the intervening 15 years.[3] There are two main reasons for the high false alarm rate. One is that physiological data can be severely corrupted by artifacts, noise and missing values. The other reason is that univariate alarm algorithms and simple numeric thresholds are predominantly used in current clinical bedside monitors.[4] Moreover, alarm thresholds are often adjusted in an ad hoc manner, based on how annoying the alarm is perceived to be to the clinical team in attendance. There is little evidence that alarm thresholds are optimized for any population, particularly in any multivariate manner.

Various strategies have been employed to deal with the false alarm problem including median filtering,[5] conventional statistical signal processing and filtering,[6] multivariable fuzzy temporal profile modeling,[7] multi-parametric analysis[1,8-11] and signal quality assessment techniques.[10-12] Most of these studies however, use small number of alarms and patients. There are two studies that have used a large database and robust study design by splitting the data into a training and test data set to develop and evaluate their algorithms. Aboukhalil et al.[1] used arterial blood pressure (ABP) waveform and signal quality indices (SQIs) to suppress electrocardiogram (ECG) arrhythmia false alarms. Among five alarm categories and 5386 critical ECG arrhythmia alarms, the false alarm (FA) reduction rates were from 93.5% to 33.0% respectively and the true alarm (TA) reduction rates were 0%, except for ventricular tachycardia (VT) alarms (9.4%). Deshmane[10] used a signal quality assessment scheme for the pulse oximetry or photoplethysmogram (PPG) waveform as well as ABP and ECG to suppress false ECG critical arrhythmia alarms. Among 4012 critical ECG arrhythmia alarms, the FA reduction rates were from 68.2% to 1.6% with TA reduction rates of 4.0% (asystole), 0% (extreme bradycardia, EB), 0.8% (extreme tachycardia, ET) and 0.2% (VT). The main problem Aboukhalil et al.[1] and Deshmane[10] faced was that the VT alarm had high TA reduction rate but low FA reduction rate, as ABP and PPG waveforms did not always manifest low cardiac output, pulse pressure or sometimes, particularly

* Corresponding author. Institute of Biomedical Engineering, Old Road Campus Research Building, Off Roosevelt Drive, OX3 7DQ Oxford, UK.

recently applied a model-based filtering approach to detecting the above listed alarms in the MIMIC II database. They also quote superior FA suppression rates (except for bradycardia). However, it should be noted that all three waveform signals need to be present plus a central venous pressure (CVP) or pulmonary arterial pressure (PAP) waveform, which significantly limits the application of the system to a small subset of the population and only when the signals exhibit high quality. Moreover, the authors treated all alarms together, rather than dividing the data into independent training and testing sets.

In the work we describe a framework that learns the relationship between the occurrences of noise and signals across all the cardiovascular signals in the ICU during life-threatening ventricular arrhythmias. Features extracted from the ECG, ABP and PPG (including heart rate (HR), pulse oxygen saturation (SpO_2), signal quality indices and rates of changes in parameters) were combined in a novel data fusion framework to suppress the false arrhythmia alarms.

As the ABP is an invasive measurement, present in only about two thirds of a typical ICU population, we compared the algorithms with ABP and without ABP. First, we generated a novel PPG signal quality assessment method using dynamic time warping algorithm[14] and used it to suppress the false alarms, according to the frame which Aboukhalil et al.[1] and Deshmane[10] used. We then estimated the heart rate (HR) from the ECG, ABP and PPG separately, fused the result using a Kalman filter and SQIs,[12,14,15] and used it to suppress the false alarms. These traditional methods showed a good performance on asystole and extreme bradycardia (EB) alarms, modest on extreme tachycardia (ET) alarms, but poor performance on VT alarms. To improve the VT alarm performance, in this work we extracted 114 variables from ECG, ABP and PPG signals, including signal features and SQIs. We then used a feature selection technique, a genetic algorithm (GA), to select the optimal variable combination. The GA mimics the principles of natural selection to "breed" possible successful combinations of parameters, and "kills off" poorly performing combinations of parameters. The best feature combination are then presented to a nonlinear classifier known as a relevance vector machine (RVM), to label the alarms as true and false.

Materials and methods

Data sets

We used the same data sets as described by Deshmane[10] with minor adjustments, drawn from the multi-parameter ICU database (PhysioNet's MIMIC II database),[16-18] containing simultaneous ECG, ABP, and PPG recordings with 4107 multiple expert-annotated alarms (asystole, EB, ET, and VT) on 182 ICU admissions. The adjustments include adding one case into the data sets, eliminating the alarms when PPG is unavailable at the time the alarm occurs, and revising 41 annotations from true to false which we

dard data sets and subsets of critical ECG arrhythmia alarms: relative frequency of true and false alarms on a per-alarm basis.

e	Total alarms	Subset 1 (ECG and PPG available)								Subset 2 (ECG, ABP and PPG available)							
		Total alarms	% Of all	True alarms			False alarms			Total alarms	% of all	True alarms			False alarms		
				N	% of all	% of alarm type	N	% of all	% of alarm type			N	% of all	% of alarm type	N	% of all	% of alarm type
	644	564	15.1	36	1.0	6.4	528	14.1	93.6	285	12.4	25	1.1	8.8	260	11.3	91.2
	360	327	8.8	240	6.4	73.4	87	2.3	26.6	233	10.1	171	7.4	73.4	62	2.7	26.6
	843	729	19.5	644	17.2	88.3	85	2.3	11.7	257	11.2	220	9.6	85.6	37	1.6	14.4
	2260	2114	56.6	1133	30.3	53.6	981	26.3	46.4	1525	66.3	849	36.9	55.7	676	29.4	44.3
	4107	3734		2053	55.0		1681	45.0		2300		1265	55.0		1035	45.0	

me bradycardia, ET: extreme tachycardia, VT: ventricular tachycardia.

Table 2
Distribution of alarms in training and test sets of subset 1.

Alarm type	Training				Test			
	False	True	Total	FA rate (%)	False	True	Total	FA rate (%)
Asystole	293	19	312	93.9	235	17	252	93.3
EB	63	123	186	33.9	24	117	141	17.0
ET	29	401	430	6.7	56	243	299	18.7
VT	483	672	1155	41.8	498	461	959	51.9
All	868	1215	2083	41.7	813	838	1651	49.2

PPG was unavailable, 3734 alarms remained. The false alarm rates were 93.6% for asystole, 26.6% for EB, 11.7% for ET, and 46.4% for VT respectively, and 45.0% overall. The ICU visits were divided into two separate sets for testing and training, ensuring that the frequency of alarms in each category was roughly equal through frequency ranking and separating odd and evenly numbered signals. The data were divided into two further subsets based on signal availability: subset 1 with ECG and PPG available for 30 s before and 10 s after each alarm; and subset 2 with ECG, ABP and PPG available in the same temporal window. Table 1 details the relative frequency of each alarm category and their associated true and false rates. Tables 2 and 3 show the distribution of alarms in training, test, and combined sets of subset 1 and subset 2. Three examples of false VT alarms and one true VT alarm are shown in Fig. 1.

We took three approaches to false alarm reduction, which are now described.

False alarm reduction based on PPG

We developed a novel PPG SQI using the Dynamic Time Warping (DTW), multiple-template matching, and a heuristic fusion algorithm, which is described in Li and Clifford.[14] A PPG beat dynamic template was built by detecting and averaging the regular beats in a 30-s PPG signal window and segmenting each beat from the onset of the beat to the onset of the next beat. Beat detection was performed using *wabp.c* (an open source ABP beat detector available at www.physionet.org) with a time and amplitude threshold adjustment to fit PPG beat width and height. If no beat was found 3 s after the onset of any given beat, then the end of the beat window was truncated to 3 s. The correlation coefficient between each PPG beat and the template was calculated. However, because the morphology of beat may change in length due to changes in heart rate or cardiac output, three methods were used to fit each PPG beat with the template: (1) a direct correlation (no beat morphology changes), (2) linear interpolation of the beat with resampling to match the template, and (3) DTW, which stretches the nonlinear time-base and traces an optimal path to minimize the cumulative distance between the beat and the template. We also applied a clipping detection algorithm to quantify the percentage of samples which were saturated (to the maximum or minimum values) within the beat window. These four measures were then fused heuristically to classify each beat into excellent (E), acceptable (A), and unacceptable (U) according to Eq. (1). Taking SQI_i, $i=1,2,3,4$, as the SQIs derived from direct correlation, linear interpolation, DTW, and clipping detection, then they are fused to form qSQI by Eq. (1). The percentage of good beats (E and A) in a 17-s analysis window (13 s prior to the alarm onset and 4 s after the alarm, which was also used by Aboukhalil et al.[1] and Deshmane[10]) was set as the SQI of PPG.

$$qSQI=\begin{cases}\text{Excellent(E)} & \text{if All of the 4 } SQI_i\geq 0.9\\ \text{Acceptable(A)}\begin{cases}\text{if 3 of the 4 } SQI_i\geq 0.9 \text{ OR}\\ \text{if All of the 4 } SQI_i\geq 0.7 \text{ OR}\\ \text{if median}(SQI_1,SQI_2,SQI_3)\geq 0.8 \text{ and } SQI_1\geq 0.5 \text{ and } SQI_4\geq 0.7\end{cases} & \\ \text{Unacceptable(U)} & \text{otherwise}\end{cases} \quad (1)$$

where the coefficients 0.9, 0.8, 0.7 and 0.5 are arbitrary and empirically determined.

We set a PPG SQI threshold (SQI_{th}) for each type of alarm to accept or reject the information in the PPG. The PPG signals with $SQI\geq SQI_{th}$ (where the PPG was of sufficient high quality), were used to suppress the alarms. In order to avoid TA suppression, at first, SQI_{th} was set strictly to 1. Subsequently, the SQI_{th} was gradually decreased, ensuring that the TA suppression was always minimized.

False alarm reduction based on HR and SQI derived from PPG, ABP, and ECG

Following our previous study,[12] we estimated the HRs and SQIs from PPG, ABP, and ECG to suppress false alarms.

Table 3
Distribution of alarms in training and test sets of subset 2.

Alarm type	Training				Test			
	False	True	Total	FA rate (%)	False	True	Total	FA rate (%)
Asystole	166	14	180	92.2	94	11	105	89.5
EB	58	108	166	34.9	4	63	67	6.0

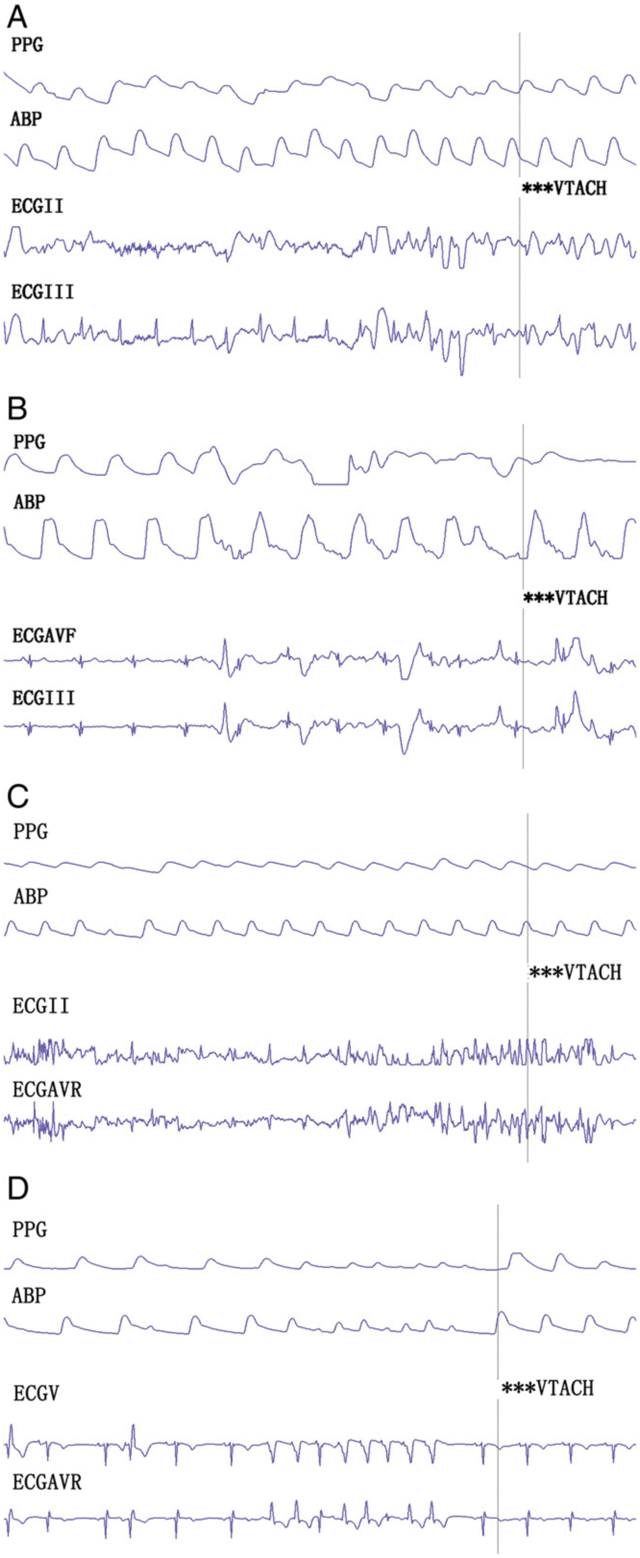

Fig. 1. Examples of false and true ventricular tachycardia alarms. Note the vertical line marks the time the alarm sounded. (A and B) False alarms and the algorithm failed to suppress them. (C) A false alarm and is suppressed correctly. (D) A true alarm and is accepted correctly by the algorithm.

A 20-s analysis window (prior to the alarm onset) was used described in our earlier work.[12] The maximum, minimum, and mean HR were also calculated for each of the seven HRs over a window that was centered on the current beat and included both neighboring beats. The resulting 21 HRs and corresponding SQIs were then used to suppress false alarms by varying the SQI thresholds to decide if the source data are trustworthy or not. Subset 2 was used to evaluate the algorithm at this step.

Machine learning for false alarm reduction

A machine learning algorithm approach was used to learn the noise and signal relationships in each true and false VT alarm condition, which are the most difficult false alarms to suppress. Therefore, an extensive set of features were defined and a genetic algorithm (GA) was used to select pertinent features which were then presented to an RVM to classify VT alarms as true or false.

Variable choice. In total 114 variables (including 87 features and 27 SQI metrics) were extracted from ECG, ABP, PPG, and SpO_2 signals within the 20-s analysis window. The features included HR (extracting from ECG, ABP, and PPG), systolic, diastolic, mean, and pulse blood pressure, SpO_2, amplitude of PPG, and area difference of beat (ADB) with the mean area under the waveform of each beat in the 20-s window of the ECG, ABP, and PPG. Each feature except ADB has five sub-features calculated over the 20-s window: including maximum, minimum, median, variance, and gradient (derived from a robust least squares fit over the entire window). The ADB has only four sub-features; the mean ADB of five beats with the shortest beat-to-beat intervals (ADB_{mean_top5}), the maximum of mean ADB of five consecutive beats (ADB_{max_mean5}), the variance ($ADB_{variance}$), and the robust least squares gradient ($ADB_{gradient}$) of beats in the 20-s window. The SQI metrics of ECG included two metrics of inter-channel and inter-algorithm comparisons of two QRS detectors, kurtosis of ECG, spectral distribution of ECG and a fusion of these four metrics.[12] The ABP SQI metrics included a signal abnormality index with its nine sub-metrics[19] and the DTW-based SQI fusion with its four sub-metrics[14] which was discussed above and was applied on the ABP signal as well. The PPG SQI metrics included the DTW-based SQIs[14] and two Hjorth parameters[10] which estimated the dominant frequency and half-bandwidth of the spectral distribution of PPG.

Feature selection. Since it is unlikely that all 114 parameters are useful (and in fact some may end up lowering the performance) a variable selection technique is required. Moreover, with a limited number of patterns from which to learn, it is important to keep the number of free parameters which we need to learn as low as possible. A genetic algorithm (GA)[20,21] was therefore used to select the optimal subset of variables for true/false alarm classification. Genetic algorithm is a general adaptive optimization search method-

Table 4
Performance of the PPG-based false alarm suppression algorithm.

Alarm Type	Data set	#True	#False	TA suppression	FA suppression	SQI threshold
Asystole	Training	19	293	0 (0%)	236 (80.5%)	0.1
	Test	17	235	0 (0%)	203 (86.4%)	
	Total	36	528	0 (0%)	439 (83.1%)	
EB	Training	123	63	0 (0%)	59 (93.7%)	0.1
	Test	117	24	0 (0%)	20 (83.3%)	
	Total	240	87	0 (0%)	79 (90.8%)	
ET	Training	401	29	0 (0%)	3 (10.3%)	1
	Test	243	56	0 (0%)	15 (26.8%)	
	Total	644	85	0 (0%)	18 (21.2%)	
VT	Training	672	483	1 (0.15%)	8 (1.66%)	0.1
	Test	461	498	1 (0.21%)	10 (2.01%)	
	Total	1133	981	2 (0.18%)	18 (1.83%)	

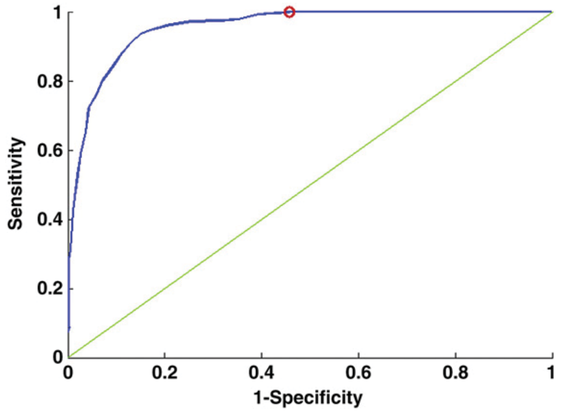

Fig. 2. ROC curves of 56 selected variables (with η=12) for training data using the RVM algorithm. The circle marks the operating point where no true alarm suppression occurs.

chromosome was defined to be a binary vector with the same length as the number of features (114 elements long in our scenario), each element (gene) representing one of the features (with a "1" indicating a feature is selected). A set of chromosomes that were created randomly made up of the original generation called a population. Then three operations, called selection, crossover and mutation, were iterated to generate next generations until acceptable results were obtained or a fixed number of generations elapsed. In the selection operation, a fitness function was used to pick up the chromosomes with better performance. In the crossover operation, pairs of chromosomes (parents) were chosen randomly to swap parts of their information (binary string) at a randomly selected locus to give birth of their children. Mutation is used to randomly flip the value of some single bits within individual strings. An operator call clone was also used to copy some parents which have good performance to the next generation without crossover or mutation. Associated with the characteristics of exploitation and exploration search, GA can deal with large search spaces efficiently, and hence has less chance to get local optimal solution than other algorithms. In our study, with a population of 50 chromosomes with 114 genes each, a 2% mutation rate, a 10% cloning rate, a 45% cull rate for crossover and a 100-generation limit, the search space of possible variable combinations was rapidly explored. The fitness function that was minimized was the root mean squared error (rMSE) of a multivariate logistic regression. The training set of subset 2 was used and was split further into training and validation sets to train and evaluate the algorithm. A bootstrapping procedure was performed by running the logistic regression on the training set and evaluating the rMSE on the validation set. The GA selection was repeated 100 times and the selected variables were sorted by the frequency of selection. This ranking was then used as the order of priority in the machine learning module. The process was repeated with and without ABP features in order to indicate the performance of the algorithm on patients when the ABP line is not required.

Machine learning algorithm choice. A Relevance Vector Machine is a sparse Bayesian model that provides probabilistic predictions through Bayesian inference.[22-24] The central idea of RVM is to map a set of input data to a high-dimensional feature space through kernel functions and construct decision boundaries that separate the labeled data into their constituent classes by predicting the posterior probabilities of their class membership. Given a training data set composed of N samples $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$ with input $\mathbf{x}_i \in R^M$ and output $y_i \in R$, the RVM algorithm aims at constructing a function as shown in equation (2).

$$y = \mathbf{w}^{\mathrm{T}} \phi(\mathbf{x}) \tag{2}$$

Table 5
Performance and variable selections based on HR and SQI derived from PPG, ABP and ECG of subset 2.

Alarm type	Data set	No. of true alarms	No. of false alarms	Variable selections	SQI threshold	TA suppression	FA suppression
Asystole	Training	14	166	HR_{ABP_mean}	0.9	0 (0%)	123 (74.1%)
	Test	11	94				66 (70.2%)
	Total	25	260				189 (72.7%)
EB	Training	108	58	$HR_{ECG\_ABP\_PPG\_mean}$	0.1	0 (0%)	55 (94.8%)
	Test	63	4				4 (100%)
	Total	171	62				59 (95.2%)
ET	Training	116	19	HR_{ECG_min}	0.6	0 (0%)	12 (63.2%)
	Test	104	18	$HR_{ABP\_PPG\_min}$	0.5		5 (27.8%)
	Total	220	37				17 (46.0%)

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.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.