Heartbeat Fusion Algorithm to Reduce False Alarms for Arrhythmias

Chathuri Daluwatte[1], Lars Johannesen[1,2], Jose Vicente[1,3,4], Christopher G Scully[1], Loriano Galeotti[1], David G Strauss[1]

[1]Division of Biomedical Physics, Office of Science and Engineering Laboratories, CDRH, US FDA, Silver Spring, MD, USA

[2]Department of Clinical Physiology, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden

[3]Division of Cardiovascular and Renal Products, Office of New Drugs, CDER, US FDA, Silver Spring, MD, USA

[4]BSICoS Group, Aragón Institute of Engineering Research (I3A), IIS Aragón, University of Zaragoza, Zaragoza, Spain

Abstract

There is a need for patient monitoring algorithms to reduce alarm fatigue by rejecting clinically irrelevant alarms. We developed an algorithm using multimodal physiological waveforms (electrocardiogram, blood pressure, photoplethysmogram) and noise classifiers to improve arrhythmia detection by reducing the incidence of false alarms while maintaining a high true alarm rate as part of the Physionet Challenge 2015. Combining information from multiple physiological signals our algorithm was able to discard 362 of 456 false alarms (true negative rate [TNR] of 79%), while correctly classifying 268 of the 294 true alarms (true positive rate [TPR] of 91%) on the training set, a score of 73.8. When applied to the test set which had 343 false alarms and 157 true alarms, we achieved a TNR of 81%, TPR of 86% and score of 70.2. Our results support the concept that false alarms can be reduced in the intensive care unit by removing noise segments in signals and combining information from multiple physiological signals.

1. Introduction

The high number of false alarms in Intensive Care Units (ICU) can lead to "alarm fatigue" [1], a top medical device hazard [2]. In the ICU, 88.8% of arrhythmia alarms have been shown to be false [3]. False alarms not only disturb patient rest and pose unnecessary burden on the caregiver staff, but also put patients at risk because caregivers become desensitized to alarms leading to delayed reaction times to clinically relevant events [4, 5]. Reducing false arrhythmia alarms is a complex problem because multiple factors such as inappropriate user settings, patient conditions, noise and motion artifacts, and algorithm performance contribute to alarm fatigue [3]. However, physiological signals from different monitors contain complementary information that is used by clinicians to recognize false alarms. Therefore information from multiple signals could be used in automatic algorithms to reduce false alarms and mitigate alarm fatigue. Here we present an algorithm using multimodal physiological waveforms (electrocardiogram [ECG], blood pressure [BP], and/or photoplethysmogram [PPG]) to improve the detection of five life threatening arrhythmias (asystole, extreme bradycardia, extreme tachycardia, ventricular flutter/fibrillation [VF], and ventricular tachycardia [VT]) by reducing the incidence of false alarms while maintaining a high true alarm rate as part of the Physionet Challenge 2015 (2015 Challenge) [6].

2. Methods

2.1. Dataset

We used 750 records from the 2015 Challenge training set for algorithm development and tested on 500 records in the 2015 Challenge test set [7]. Each record is at least five minutes long and includes ECG, BP and/or PPG signals. In each record a bedside monitor detected an arrhythmia and triggered an alarm at the end of the record, thus the onset of the arrhythmia alarm condition must be during the last 10 seconds of the record [6, 7]. Expert human annotators reviewed and labeled each of these alarms as true or false [7]. The type of alarm detected is indicated in each record. Arrhythmias included in this dataset were asystole, extreme bradycardia, extreme tachycardia, VF, and VTs as defined in Clifford et al. [6, 7]. The distribution of alarms and signals used is reported in [7].

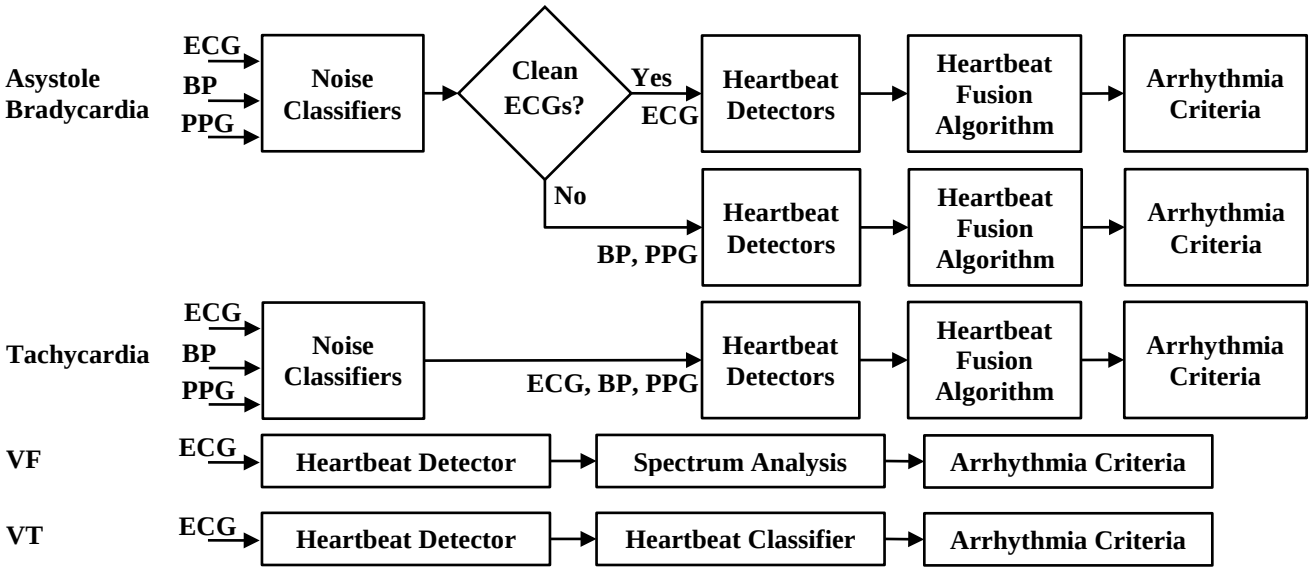

Fig. 1: Algorithm Overview

2.2. Algorithm

2.2.1. Classifying asystole, extreme bradycardia, and extreme tachycardia alarms

To classify asystole, extreme bradycardia, and extreme tachycardia alarms as true or false, we developed an algorithm based on global heartbeat annotations generated by fusing individual heartbeat detections from multiple physiological signals. Briefly, we adapted our previously developed "heartbeat fusion algorithm" [8] to identify global heartbeats from multiple signals in 2015 Challenge datasets. We then applied the following arrhythmia criteria based on definitions given by Clifford et al. [6, 7] to the global heartbeat detections to classify whether the alarm associated with the recording was true or false:

- Asystole: No QRS for at least 4 seconds
- Extreme Bradycardia (bradycardia): fewer than 5 heartbeats occur within a period of 6 s
- Extreme Tachycardia (tachycardia): Sum of RR interval for 17 consecutive heartbeats is less than 8 s

To minimize the effect of spurious heartbeat detections due to noisy signals, which can increase the false negative rate (classifying a true alarm as a false alarm), we implemented noise classifiers for each signal type to exclude signals containing artifacts. Three reviewers independently labeled the last 10 s waveform segment of each ECG, BP, and PPG channel of each asystole, bradycardia and tachycardia training set records as "clean" or "noisy" by viewing the waveform segment with marked locations of beat detections. If the reviewer considered that the heartbeat detector did not identify correct heartbeat locations in the segment, due to noise, it was labeled as noisy. A set of clean/noisy reference annotations was generated from the segments where both reviewers agreed on clean/noisy labeling. Segments that reviewers did not agree on were not considered in the development or assessment of the noise classifiers. The records with agreement between reviewers were then divided into two sets randomly: "noise-training" and "noise-validation" (Table 1). A set of signal features likely related to noise artifacts (Table 2) were selected based on previous experience.

Using these features, support vector machines (SVM) with radial basis function kernels to classify signals as "noisy" or "clean" were trained for each signal type using libSVM [9]. To evaluate the predictive ability of the selected features to correctly classify noisy and clean segments, SVMs were trained on the "noise-training" set and validated on the "noise-validation" set. The final SVMs used in the algorithm were then trained after combining both "noise-training" and "noise-validation" datasets together to be used in the 2015 Challenge test set assessment. The algorithm selects signals which are labeled as "clean" by their respective noise classifier to feed into our heartbeat fusion algorithm [8] to generate global heartbeat detections, upon which the respective arrhythmia criteria is then applied to classify whether it is a true alarm or a false alarm (Fig. 1).

Set	Reference Annotation	ECG	BP	PPG
Noise-Training	**Noisy**	102	15	64
	Clean	296	65	94
Noise-Validation	**Noisy**	47	3	33
	Clean	174	42	56

Table 1: Number of signals used to train and validate noise classifiers.

Signal	Features (calculated using last 10 s of signal)
ECG	Standard deviations of points perpendicular to the axis on Poincare plot (SD1 and SD2) 5^{th} to 95^{th} percentile of highpass filtered signal Mean of correlation coefficients for correlations between inter heart beat interval signal values and fitted straight line for the interval Upper 95^{th} percentile of first derivative Root mean square value of unfiltered signal
BP	Standard deviations of points perpendicular to the axis on Poincare plot (SD1 and SD2) Average and maximum of pulsatile signal upstroke slope Minimum and maximum inter heartbeat interval Mean systolic pressure and mean diastolic pressure Number of non-normal heartbeats Peak frequency in fast Fourier Transform Width to left side of dominant peak
PPG	Mean pulsatile signal upstroke slope Maximum inter heartbeat interval Number of non-normal heartbeats Peak frequency in fast Fourier Transform Width of dominant frequency peak Width to either side of dominant peak

Table 2: Signal features used to train SVM for noise classifiers by signal type.

2.2.2. Classifying VF alarms

Based on the VF definition given by Clifford et al. [6, 7] to classify VF alarms as true alarms, we checked for the absence of QRS detections for four seconds or if VF filter leakage (ratio between the waveform energy before and after sending through a narrow bandstop filter with a central frequency equivalent to the mean signal frequency) [10] is less than a tuned threshold. After studying the distributions of VF filter leakage in VF true alarms and VF false alarms in the training set, the threshold was determined to be 0.68.

2.2.3. Classifying VT alarms

Based on the VT definition given by Clifford et al. [6, 7] to determine true VT alarms we used a heartbeat classification method to identify ventricular ectopic heartbeats, based on QRS template-matching. Templates were formed by extracting a 100 ms window around the combined ECG annotations. Afterwards, heartbeat one is defined as class 1, then heartbeat two is matched to class 1 using cross correlation, if they are of the same class, class template is updated (median across all members of class) otherwise a new class is defined and so forth. After defining heartbeat classes, the most frequent heartbeat class is labeled as the "normal" group (likely normal sinus heartbeats), while the other classes are considered "non-normal" (likely of ventricular origin or potentially noise artifacts). Noisy signals (e.g., due to electrode movements) resulting in spurious heartbeat detection usually generate a different template for each false-heartbeat detection, thus if there is a large number of beat classes (greater than six) the signal is considered noisy and discarded. If more than five consecutive "non-normal" heartbeats with a heart rate higher than 90 bpm were detected the VT alarm is identified to be a true alarm.

2.3. Implementation

Our algorithm was implemented in MATLAB (The Mathworks, Inc., Natick, MA)/Octave [11] with some components written in C++.

2.4. Statistical Analysis

The performance of the algorithm to correctly classify the reference true/false alarm annotations was assessed by calculating three metrics: true positive rate (TPR = number of true alarms classified as true alarms/total number of true alarms), true negative rate (TNR = number of false alarms classified as false alarms/total number of false alarms), and score (score = (TP+TN)/(TP+TN+FP+5*FN)) [6, 7].

3. Results

Performances of noise classifying SVMs are reported on Table 3. Table 4 reports overall and by alarm type performance of the algorithm on 2015 Challenge training and test sets. Our algorithm was able to discard 362 of the 456 of false alarms, while correctly classifying 268 of the 294 true alarms in the training set. When applied to the test set, our algorithm discarded 277 of 343 false alarms while correctly identifying 136 of 157 true alarms.

Set	Reference Annotation	ECG	BP	PPG
Noise-Training	**Noisy**	45	100	95
	Clean	98	98	95
Noise-Validation	**Noisy**	45	67	42
	Clean	96	90	93

Table 3: Noise classifier performance (% of correctly classified signals) in the "noise-training" and "noise-validation" sets.

Alarm Type	Training Set			Test Set		
	TPR	TNR	Score	TPR	TNR	Score
Asystole	91	86	81.6	78	87	77.8
Bradycardia	98	44	69.1	74	53	43.8
Tachycardia	96	78	82.5	96	100	84.2
VF	100	83	84.8	89	92	85.5
VT	80	82	67.4	82	81	70.8
Overall	91	79	73.8	86	81	70.2

Table 4: Performance of the algorithm.

4. Discussion

Alarm fatigue is a consequence of high false alarm rates in ICU and can result in patient harm. We developed an algorithm that classifies arrhythmia alarms as true or false using robust heartbeat detections from multiple physiological signals and signal quality information.

Our algorithm manages the presence of noise in signals by using noise classifiers for each signal type. However, while the noise classifiers are very specific (high percentage of clean signals are identified correctly), are not very sensitive (percentage of noisy signal identified is low). This is due to high variety of noise artifacts present on signals. For BP and PPG noise classifiers the performance is not consistent across "noise-training" and "noise-validation" sets due to the small number of signals available to train the SVM.

Some of the records in the 2015 training set, which are from critical patient conditions, are likely mislabeled as false alarms while some do not match the definition for the given arrhythmia, hence can thus be missed by the specific alarm classifier. For example, asystole alarm a539l is annotated as a false alarm; however the ECG waveform shows signs of the ventricular arrhythmia torsade de pointes, initiated with the classical short-long-short sequence and presence of QRS complexes that twists around the isoelectric line, and later on the ECG shows sign of VF. Classifying this alarm as false can lead to missed patient care causing potential harm. This example highlights the need to look before the time of the alarm to ensure proper alarm classification, and illustrate the difficulties with alarm annotation.

Our work suggests that algorithms combining multiple physiological signals together with information about signal quality can mitigate alarm fatigue in an ICU setting by reducing the false alarms rate while maintaining high true alarms rates.

Acknowledgements

This project was supported in part by U.S. Food and Drug Administration's Medical Countermeasures Initiative, Critical Path Initiative, Office of Women's Health and appointments to the Research Participation Programs at the Oak Ridge Institute for Science and Education through an interagency agreement between the Department of Energy and FDA.

Disclaimer

The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

References

[1] Chambrin MC, et al. Alarms in the intensive care unit: how can the number of false alarms be reduced? Crit Care 2001;5(4):184–8.

[2] ECRI Institute., Top 10 Health Technology Hazards for 2014: Key safety threats to manage in the coming year [guidance article]. Health Devices 2013; 42(11):354-80.

[3] Drew BJ, et al. Insights into the problem of alarm fatigue with physiologic monitor devices: A comprehensive observational study of consecutive intensive care unit patients. PLoS ONE 2014; 9(10):e110274.

[4] Scully C, et al. Advancing Regulatory Science to Bring Novel Medical Devices for Use in Emergency Care to Market: The Role of the Food and Drug Administration. Ann Emerg Med 2015; 65(4):400-3.

[5] Bonafide CP, et al. Association between exposure to nonactionable physiologic monitor alarms and response time in a children's hospital. J Hosp Med 2015; 10(6): 345-51.

[6] Physionet. Physionet Challenge 2015. http://www.physionet.org/challenge/2015. Accessed 17 August 2015.

[7] Clifford GD, et al. The PhysioNet/Computing in Cardiology Challenge 2015: Reducing False Arrhythmia Alarms in the ICU. Comput Cardiol, 2015.

[8] Galeotti L, et al. Robust algorithm to locate heart beats from multiple physiological waveforms by individual signal detector voting. Physiol Meas 2015;36(8):1705.

[9] Chang C, et al. LIBSVM : a library for support vector machines. ACM Trans Intell Syst Technol 2011; 2(3); 2:27:1--27:27.

[10] Kuo S and Dillman R. Computer detection of ventricular fibrillation. Proc Comput Cardiol 1978; 347-9.

[11] Eaton JW, et al. 2014. GNU Octave version 3.8.1 manual: a high-level interactive language for numerical computations. CreateSpace Independent Publishing Platform. ISBN 1441413006, http://www.gnu.org/software/octave/doc/interpreter/

Address for correspondence:
Chathuri Daluwatte
10903 New Hampshire Avenue
Silver Spring, Maryland, 20993
USA
chathuri.daluwatte@fda.hhs.gov