Reduction of False Cardiac Arrhythmia Alarms through the use of Machine Learning Techniques

Miguel Caballero, Grace M Mirsky

Benedictine University, Lisle, IL, USA

Abstract

Due to the so-called "crying wolf" effect, frequent false cardiac arrhythmia alarms have been shown to diminish staff attentiveness and thus reduce the quality of care patients receive in the ICU.

The PhysioNet/Computing in Cardiology 2015 Challenge seeks to improve patient care by decreasing the number of these false cardiac arrhythmia alarms. Using a training set of 750 multi-parameter recordings organized by type of arrhythmia alarm, we developed a decision tree for each arrhythmia category. We derived the features utilized in the decision tree from the arterial blood pressure (ABP) waveform and the photoplethysmogram (PPG).

For Phase I of the challenge, our score for the real-time test set = 57.64 and retrospective test set = 61.15, resulting in an overall score of 59.39. For Phase II, our score for the real-time test set = 65.19 and retrospective test set = 72.19. In conclusion, decision trees have been shown to generate reasonable results in reducing false cardiac arrhythmia alarms; future work will involve more sophisticated machine learning algorithms to improve performance.

1. Introduction

The objective of the 2015 Computing in Cardiology / PhysioNet Challenge is to develop algorithms to reduce the frequency of false cardiac arrhythmia alarms in the Intensive Care Unit (ICU) [1,2]. Research has shown the quality of care patients receive in the ICU is adversely affected by frequent false alarms due to reduced staff attentiveness [3-4]. The ICU, as a result, tends to be a very noisy environment, which can be bothersome to patients and distracting to caregivers [5-6]. The number of false alarms tends to be elevated in the ICU since the alarm detection sensitivity is set sufficiently high so as to not inadvertently miss any true cardiac arrhythmia alarms, which as a side effect also produces a high number of false alarms [7]. Consequently, a number of approaches in artificial intelligence have been taken to reduce false alarms, including rule-based expert systems, neural networks, fuzzy logic, support vector machines, relevance vector machines, and Bayesian networks [6,8,9,10], but further work is still necessary to provide an accurate, robust, and clinically-relevant solution.

The Challenge is divided into two events [2]. The objective of event 1 is to reduce the incidence of false alarms while detecting true alarms by only using data prior to the sounding of the alarm. Each of these records is exactly five minutes in length. The objective of event 2 is the same as event 1, except that data from up to 30 seconds after the alarm may be utilized. Furthermore, the Challenge consisted of two phases, the Unofficial phase which limited participants to a maximum of five submissions and the Official phase which limited participants to a maximum of ten submissions.

2. Materials & methods

2.1. Data

A total of 1250 recordings were obtained from four hospitals in the USA and Europe, each containing an arrhythmia alarm [2]. The training set consisted of 750 multi-parameter recordings organized by alarm type. There were five types of critical arrhythmias included in the dataset: asystole, extreme bradycardia, extreme tachycardia, ventricular tachycardia, and ventricular flutter/fibrillation. The test set, consisting of 500 recordings, was hidden from participants for the duration of the Challenge. Experts reviewed all of the recordings and determined which alarms were true alarms and which were false alarms. These designations were made available to Challenge participants for the training set only.

2.2. Preprocessing

The arterial blood pressure (ABP) and photoplethysmogram (PPG) signals were obtained and resampled to 125 Hz for each patient, using scripts provided by the Challenge organizers [2]. Note that both ABP and PPG were not available for all patients, but all patients had at least one of the two signals recorded. Signals were filtered with a finite impulse response bandpass filter (0.05-40Hz) as well as with a notch filter to reject noise. Two electrocardiogram (ECG) leads were also available, but for this work, we limited our analysis to the ABP and PPG only. Examples of sample signals available for analysis for asystole, ventricular flutter/fibrillation, and ventricular tachycardia alarms are shown in Figures 1-3.

2.3. Machine learning

In searching for an appropriate machine learning algorithm to approach the Challenge, we investigated several different algorithms. Of these algorithms, we selected decision trees for several reasons. Decision trees are a frequently utilized technique for inductive inference [11]. One of their main strengths is robustness to noise in the data. Furthermore, since learned decision trees may be rewritten as if-else statements, the computational complexity is minimal once realized in hardware in an actual system. Therefore, we wanted to determine if sufficiently high accuracy in correctly identifying cardiac arrhythmia alarms could be achieved with a computationally simple solution.

We developed a decision tree for each arrhythmia category, which we combined with domain knowledge to produce a set of if/else statements. The attributes obtained from the sample set included the following for both the ABP and PPG [2]:

1. Signal quality index
2. High heart rate of seventeen sequential beats
3. Low heart rate of five sequential beats
4. Maximum heart rate
5. Maximum RR-interval
6. Median RR-interval

Separate decision trees were trained using the ABP and PPG, since we observed that the attributes varied based upon the signal type. The specific tests derived from the decision tree will be made available as our submission on the PhysioNet Challenge webpage.

3. Results

Table 1 shows the best entry for the Unofficial Phase. Scores are calculated according to the following equation, provided by the Challenge organizers [2]:

$$\text{Score} = (TP + TN) / (TP + TN + FP + 5*FN)$$

where TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

In order to improve the results from the Unofficial Phase, we iteratively pruned the decision tree and added domain knowledge about the characteristics of the arrhythmias. The best results from the Official Phase are shown in Table 2.

Table 1: Results on the test set from the best entry in the Unofficial Phase.

	TPR	TNR	Score
Asystole	50%	89%	68.59
Bradycardia	90%	76%	69.91
Tachycardia	97%	60%	86.18
Ventricular Flutter Fib	56%	88%	64.86
Ventricular Tachycardia	57%	62%	44.94
Real-time	79%	72%	58.86
Retrospective	81%	75%	61.15

Table 2: Results on the test set from the best entry in the Official Phase.

	TPR	TNR	Score
Asystole	72%	88%	76.57
Bradycardia	95%	90%	84.76
Tachycardia	97%	60%	86.18
Ventricular Flutter Fib	89%	69%	67.74
Ventricular Tachycardia	49%	87%	55.22
Real-time	79%	84%	65.19
Retrospective	84%	89%	72.19

4. Discussion

Supervised machine learning techniques are effective at predicting future events, given a representative sample set. While the decision tree performed reasonably well for the test set, it may not be ideal for all patients since it is based on hardcoded parameters. A more robust approach would be to detect changes on a per-patient basis rather than over a patient population [12]; however, this would require a much longer recording than was provided for the purposes of the Challenge.

The highest scores in the test set were achieved on Bradycardia and Tachycardia, with scores of 84.76 and 86.18, respectively. This is likely due to the straightforward nature of detecting an elevated or depressed heart rate rather than an additional change in rhythm. Ventricular tachycardia proved most difficult of the arrhythmias to accurately classify using the decision tree method, which may have resulted from overtraining. Attempts to prune the tree yielded improvements in performance on the test set, but further work is necessary to create a more generalized tree that performs better on the test set.

Alternative machine learning techniques will be investigated in future work to determine if better performance may be obtained within the constraints of the Challenge design. These methods will likely include neural networks, fuzzy logic, and support vector machines.

Acknowledgements

The authors would like to thank the Department of Mathematical & Computational Sciences and the College of Science at Benedictine University for supporting this research.

References

[1] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000;101(23):E215–E220.

[2] Reducing False Arrhythmia Alarms in the ICU: the PhysioNet/Computing in Cardiology Challenge 2015. http://physionet.org/challenge/2015/

[3] Chambrin, MC. Alarms in the intensive care unit: how can the number of false alarms be reduced?. Critical Care-London 2001;5(4):184-188.

[4] Cvach, M. Monitor alarm fatigue: an integrative review. Biomedical Instrumentation & Technology 2012;46(4): 268-277.

[5] Konkani A, Oakley B, Bauld TJ. Reducing hospital noise: a review of medical device alarm management. Biomedical Instrumentation & Technology 2012;46(6): 478-487.

[6] Schmid F, Goepfert MS, Reuter DA. Patient monitoring alarms in the ICU and in the operating room. Critical Care 2013;17(2):216.

[7] Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. AACN advanced critical care 2013;24(4):378-386.

[8] Ben Rejab F, Nouira K, Abdelwahed T. Support Vector Machines versus Multi-layer Perceptrons for Reducing False Alarms in Intensive Care Units. International Journal of Computer Applications 2012; 49(11): 41-47.

[9] Li Q, Clifford GD. Signal quality and data fusion for false alarm reduction in the intensive care unit. Journal of electrocardiology 2012; 45(6): 596-603.

[10] Aboukhalil A, Nielsen L, Saeed M, Mark RG, Clifford GD.Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. Journal of biomedical informatics 2008; 41(3): 442-451.

[11] Mitchell T. Machine Learning. McGraw Hill. 1997.

[12] Zhang Y, Szolovits P. Patient-specific learning in real time for adaptive monitoring in critical care. Journal of biomedical informatics 2008; 41(3): 452-460.

Address for correspondence.

Dr. Grace Mirsky
Department of Computer Science
Benedictine University
5700 College Road
Lisle, IL, USA 60532
gmirsky@ben.edu

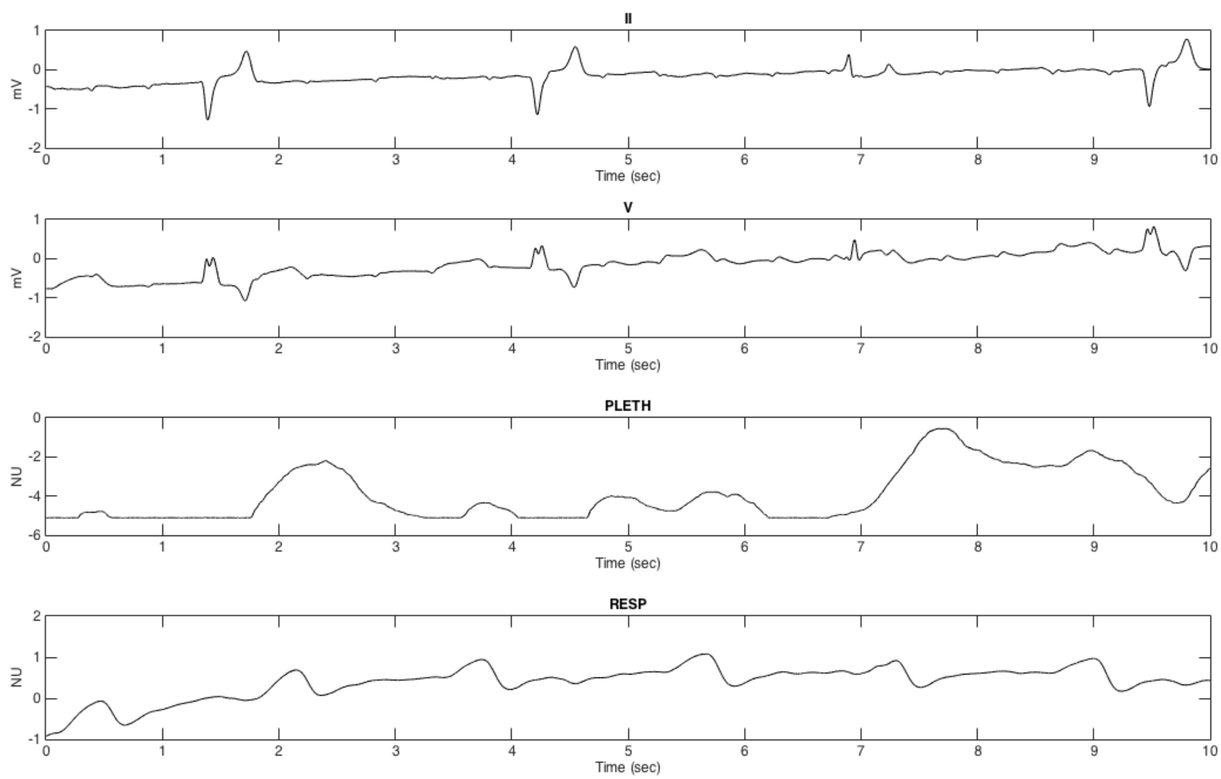

Figure 1: Example of signals containing true asystole alarm 10 seconds prior to alarm (record a145)

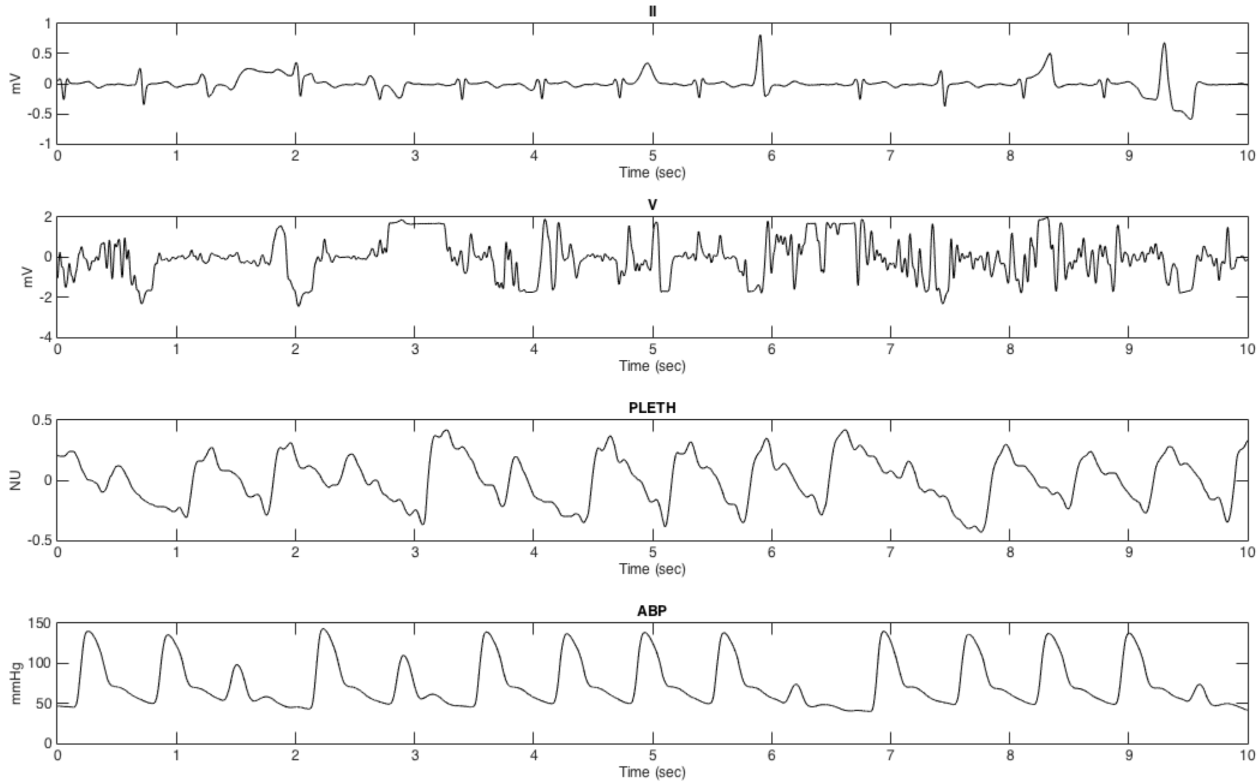

Figure 2: Example of signals containing true ventricular flutter / fibrillation alarm 10 seconds prior to alarm (record f352)

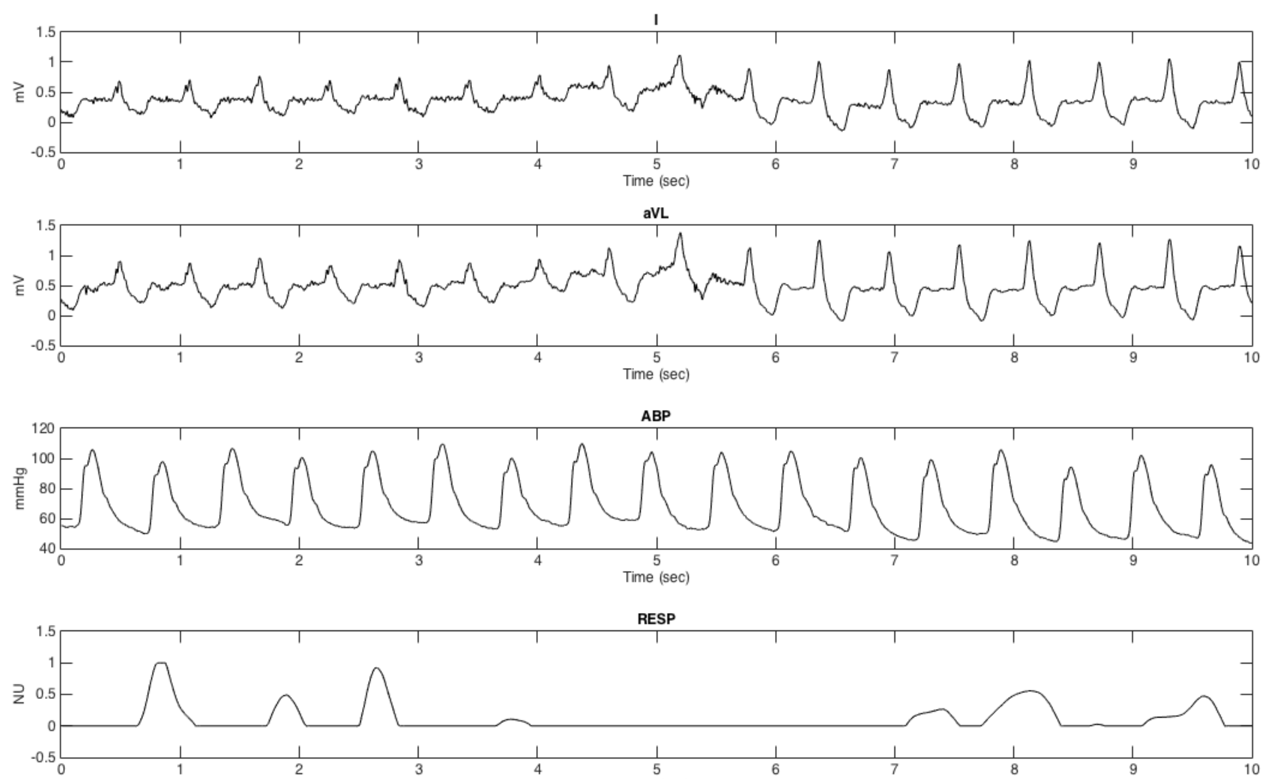

Figure 3: Example of signals containing true ventricular tachycardia alarm 10 seconds prior to alarm (record v619)