Decreasing the False Alarm Rate of Arrhythmias in Intensive Care Using a Machine Learning Approach

Linda M Eerikäinen[1], Joaquin Vanschoren[2], Michael J Rooijakkers[1], Rik Vullings[1], Ronald M Aarts[1,3]

[1] Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands
[2] Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands
[3] Department of Personal Health Solutions, Philips Research, Eindhoven, The Netherlands

Abstract

We present a novel algorithm for classifying true and false alarms of five life-threatening arrhythmias in intensive care. This algorithm was entered in the PhysioNet/Computing in Cardiology Challenge 2015 Reducing False Arrhythmia Alarms in the ICU. The algorithm performs a binary classification of the alarms for a specified arrhythmia type by combining signal quality information and physiological features from multiple sources, such as electrocardiogram (ECG), photoplethysmogram (PPG), and arterial blood pressure (ABP). Signals were selected for feature computation by first assessing the quality for available signals. Random Forest classifiers were trained separately for every type of arrhythmia with arrhythmia-specific features. Hence, the complete algorithm leverages five different predictive models. Classification sensitivities of true alarms 75–99 % (average 93 %) on the training set with cross-validation and 22–100 % (average 90 %) on the unrevealed test set. Classification specificities on the training and test set were 76–94% (average 80%) and 75–100% (average 82%), respectively. The best performance was for extreme bradycardia, whereas the poorest results were for ventricular arrhythmias. The results are for the real-time category when only information prior to the alarm is considered. The final challenge score was 75.54.

1. Introduction

Cardiac monitor algorithms are intentionally set to have high sensitivity which, as a consequence, often leads to a large number of false alarms [1]. An excessive number of false alarms may compromise patient safety. Nuisance alarms, which are annoying alarms typically not resulting an adverse condition, and/or false alarms may cause a delay in reaction time or reduce the probability of caregivers to respond [2].

The problem of false arrhythmia alarm reduction has been approached with various strategies including multi-parameter analysis [3–5] and signal quality indices (SQIs) [5–7]. A machine learning approach has been proposed to reduce the amount of false ventricular tachycardia alarms [5].

In this work we present an algorithm that consists of five arrhythmia-specific alarm classification models for five life-threatening arrhythmia alarms in the intensive care unit (ICU). Our classifiers take as input physiological features and signal quality features that were extracted from electrocardiogram (ECG), photoplethysmogram (PPG), and arterial blood pressure (ABP) after selecting the best available signals. Features were designed and selected separately for every arrhythmia type.

The paper will describe first the data and annotations followed by a description of heart beat and pulse detection in Section 2. In Section 3.2, the signal selection for feature computation is explained. Features are described in Section 3.3 and the feature selection in Section 3.4. The classification model is described in Section 3.5. Finally, results and conclusions are presented in Section 4 and 5, respectively.

2. Data

The training data of 750 records from bedside monitors in the ICU was provided by the PhysioNet/Computing in Cardiology Challenge 2015 [8]. Each record contained an alarm which was either true or false for one arrhythmia event. The arrhythmia types were asystole (ASY), extreme bradycardia (EBR), extreme tachycardia (ETC), ventricular fibrillation or flutter (VFB), or ventricular tachycardia (VTA). The test set consisted of 500 records that remained unrevealed during the challenge.

Each recording contained two leads of ECG, and at least one pulsatile waveform of either PPG or ABP. In some

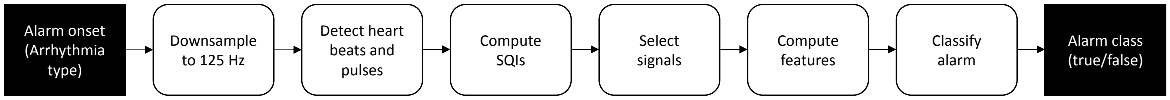

Figure 1. Flowchart of the algorithm

records, both pulsatile waveforms or a respiratory signal were present. All signals had a sample frequency of 250 Hz and had been filtered with a FIR band-pass filter with a pass band of 0.05–40 Hz and mains notch filters to remove noise.

3. Methods

In this section, we will describe the different steps of the algorithm. The overall flowchart of the algorithm is presented in Figure 1. Before proceeding to detection of heart beats and pulses, signals were downsampled to 125 Hz.

3.1. Heart beat and pulse detection

Heart rate (HR) computation was performed by first detecting beats from ECG, ABP, and PPG. Beat detection from ECG was performed with a low-complexity R-peak detector [9]. First, a convolution with a single wavelet and ECG is computed, and then the peaks are searched using an adaptive threshold. The local noise level around every peak is estimated by the detector. The R-peak detector was modified for the current data.

The pulse detection from the ABP and PPG was performed with an open-source ABP pulse detection algorithm, *wabp* [10]. The algorithm is available from PhysioNet [8].

The beat detection is not necessarily reliable if the signals are noisy. Therefore, beats detected from different channels and different signals were compared with each other. If a beat detected in a single ECG channel was detected simultaneously from another signal source, it was considered as a matched beat. Delays between ECG and ABP, and ECG and PPG were compensated in the search for matching beats.

3.2. Signal selection

As mentioned in Section 2, records contain different combinations of data. To select the most reliable signals for feature computation, an assessment of the signal quality was made after the beat detection. The quality of the signals was evaluated in a window of 16 seconds before the alarm. The best ECG channel and either ABP or PPG were selected for feature computation.

The selection of ECG channel was made based on the median local noise level around the R-peaks given by the R-peak detector. When both pulsatile waveforms were present, ABP was selected if its quality was good. If the quality of ABP was not sufficiently good, the better signal between ABP and PPG was selected.

Signal quality for ABP was computed at beat-level by the signal abnormality index (SAI) algorithm [11]. SAI extracts several features from the pulses and assigns a flag '1' if the feature does not meet the criteria of normality. In this work, feature flags were transformed to a quality index between 0 and 1, 1 corresponding to a clean signal.

The quality index of the PPG signal was computed with an open-source script available for the challenge which is based on a beat template correlation. The template is formed by averaging detected beats from the signal.

3.3. Features

Features extracted for the alarm classification can be divided into two categories: signal quality features and physiological features. Physiological features were designed for every arrhythmia type separately based on the clinical definition of the arrhythmia in question, whereas the same signal quality features were computed for every arrhythmia.

The signal quality features included seven ECG features, and the quality indices presented in the previous section for ABP and PPG. The former included five features published in previous works [7,12]. Behar et al. [7] discuss in their work that arrhythmia specific models for ECG signal quality are necessary. Therefore, several SQIs were introduced separately instead of having a combined quality measure. The SQIs included were kurtosis, skewness, spectral distribution measures of QRS complex and baseline, and pcaSQI, where pcaSQI is the ratio of the sum of the five largest eigenvalues associated with the principal components over the sum of all eigenvalues obtained by principal component analysis applied to the time aligned ECG segments in the window. In addition, the median local noise level from the R-peak detector, and the percentage of detected beats that are detected on the other ECG channel were included.

The physiological features varied between the different arrhythmia types, since every arrhythmia has different characteristics. Relying on the definitions of the arrhythmias [8], the features were designed to represent either the HR, intervals between the beats and their variation, or changes in morphological characteristics of the beats or pulses. Moreover, blood pressure variations may occur during ASY, VFB, and VTA, and therefore blood pressure features were included. The features are listed in Table 1. The complete number of features per arrhythmia varied from 12 to 17. All the features presented here were computed from a window starting 16 seconds before the alarm.

Table 1. Features for alarm classification

	Feature
SQIs	- ECG: median local noise level - ECG: match ratio with other lead - ECG: kurtosis - ECG: skewness - ECG: relative power in the QRS complex - ECG: relative power in the baseline - ECG: pcaSQI - ABP/PPG SQI
ASY	- maximum inter-beat interval - minimum of ABP
EBR	- minimum HR of 5 consecutive beats - maximum number of consecutive beats with HR under 40 bpm
ETC	- maximum HR of 17 consecutive beats - maximum number of consecutive beats with HR under 140 bpm
VFB	- maximum HR - standard deviation of normalized inter-beat intervals in the pulsatile waveform - standard deviation of normalized pulse amplitudes in the pulsatile waveform - minimum ratio of inter-beat interval and pulse amplitude in the pulsatile waveform - range of systolic blood pressure - range of diastolic blood pressure - range of pulse pressure
VTA	- maximum HR - range of systolic blood pressure - range of diastolic blood pressure - range of pulse pressure - number of ventricular beats

During VFB, a fibrillatory, flutter, or oscillatory waveform is exhibited by the heart for at least four seconds [8]. For evaluating regularity of the beats, standard deviation of inter-beat intervals and amplitudes of the pulses was computed from the pulsatile waveform. Inter-beat intervals and amplitudes were normalized in the window by subtracting the mean. Moreover, the minimum ratio of pulse amplitude and succeeding inter-beat interval was included as a feature, in expectancy of the amplitude to be smaller and succeeding interval longer, when the heart is not contracting properly.

Ventricular beats in VTA records were detected forming a template of a regular QRS complex in the ECG and computing cross-correlation between the template and detected beats. The beats that had a poor correlation with the template QRS were compared to the template in polarity, amplitude, and width. In addition, the ratio of preceding and succeeding RR intervals was computed. If at least two of the characteristics exceeded a predefined thresholds, the beat was counted as a ventricular beat.

3.4. Feature selection

Evaluating the relevance of the features and selection of the best feature combinations was made with the help of a network method that uses a penalized maximum likelihood model [13]. In this method, dependencies between features are found, and irrelevant or redundant features are removed. A feature is considered irrelevant when it does not contain information about the class label. An advantage of the method is that it compares the information in groups of features and avoids selecting less informative individual features over more informative groups of features.

3.5. Classification

The classification between true and false alarms was made with a separate Random Forest classifier for every arrhythmia. A Random Forest is a classifier that consists of a large number of tree-structured classifiers and the trees in the classifier vote for the most popular class. Having an ensemble of trees improves the classification accuracy compared to a single decision tree. The Random Forests are relatively robust to outliers and noise, and do not easily overfit on the training data when the training set is sufficiently large. [14]

In the training of the classifiers, k-fold cross-validation was used. k was 10, unless there were less than 10 samples in a class. Then k was the size of the class.

4. Results

The number of features used for the classification were reduced to simplify the classification models and to achieve better performance. Based on the relevancy of the features, the classification was made with 3–8 features depending on the arrhythmia type. For example, the alarms for ETC were classified based on only three features, which were median local noise level, maximum HR of 17 consecutive beats, and number of consecutive beats with HR above 140 bpm from ECG.

The classification performance of the algorithm was evaluated with true positive rate (TPR), true negative rate (TNR), and a challenge score [8] within the training set and the test set. Results in the training and test set are listed in Tables 2 and 3, respectively.

Table 2. Classification results in the training set

Arrhythmia	TPR (%)	TNR (%)	Score
ASY	85	88	79.60
EBR	96	79	83.05
ETC	99	89	96.38
VFB	75	94	87.29
VTA	84	74	66.34
Average	**93**	**80**	**77.65**

Table 3. Classification results in the test set

Arrhythmia	TPR (%)	TNR (%)	Score
ASY	89	92	87.12
EBR	100	86	91.75
ETC	98	100	91.60
VFB	22	94	55.81
VTA	81	75	65.88
Average	**90**	**82**	**75.54**

5. Conclusion

The algorithm with arrhythmia-specific Random Forest classifiers combining three to eight features from multi-parameter data succeeds in significantly reducing the number of false alarms of life-threatening arrhythmias in the ICU. On average, 80 % of the false alarms were detected in the training and 82 % in the test set. The best false alarm detection rate (100 %) in the test set was achieved for ETC. The classification was based only on three features. Compared to the rhythm based arrhythmias, such as ASY, EBR, and ETC, false alarms of ventricular arrhythmias were more difficult to classify resulting in a lower detection rate.

The average detection rate of true alarms was 93 % in the training and 89 % in the test set. The detection rate was 98 % or higher for ETC and EBR. Since misclassification of true alarms should be minimal, further modifications are required for improving the classification of true alarms for ASY, VFB, and VTA.

Acknowledgements

The authors would like to thank Amir Jalalirad for providing his method and help in the feature selection. In addition, the authors would like to thank the cardiologist Dr. Lukas Dekker for sharing his clinical knowledge in interpretation of the data.

References

[1] Drew BJ, Califf RM, Funk M, Kaufman ES, Krucoff MW, Laks MM, Macfarlane PW, Sommargren C, Swiryn S, Van Hare GF. Practice standards for electrocardiographic monitoring in hospital settings: An American Heart Association scientific statement from the councils on cardiovascular nursing, clinical cardiology, and cardiovascular disease in the young. Circulation 2004;110(17):2721–2746. ISSN 00097322.

[2] Graham KC, Cvach M. Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms. American Journal of Critical Care 2010;19(1):28–34. ISSN 10623264.

[3] Clifford G, Aboukhalil A, Sun J, Zong W, Janz B, Moody G, Mark R. Using the blood pressure waveform to reduce critical false ECG alarms. Computers in Cardiology 2006; 829–832. ISSN 02766574.

[4] 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;41:442–451. ISSN 15320464.

[5] 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. ISSN 00220736.

[6] Li Q, Mark RG, Clifford GD. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. Physiological measurement 2008;29:15–32. ISSN 0967-3334.

[7] Behar J, Oster J, Li Q, Clifford GD. ECG signal quality during arrhythmia and its application to false alarm reduction. IEEE Transactions on Biomedical Engineering 2013; 60(6):1660–1666. ISSN 00189294.

[8] Clifford GD, Silva I, Moody B, Li Q, Kella D, Shahin A, Kooistra T, Perry D, Mark RG. The PhysioNet / Computing in Cardiology Challenge 2015 : Reducing False Arrhythmia Alarms in the ICU. Computing in Cardiology 2015; September.

[9] Rooijakkers MJ, Rabotti C, Oei SG, Mischi M. Low-complexity R-peak detection for ambulatory fetal monitoring. Physiological Measurement 2012;33:1135–1150. ISSN 0967-3334.

[10] Zong W, Heldt T, Moody G, Mark R. An open-source algorithm to detect onset of arterial blood pressure pulses. Computers in Cardiology 2003;259–262. ISSN 0276-6547.

[11] Sun J, Reisner A, Mark R. A signal abnormality index for arterial blood pressure waveforms. Computers in Cardiology 2006;13–16. ISSN 02766574.

[12] Clifford GD, Behar J, Li Q, Rezek I. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. Physiological Measurement 2012; 33:1419–1433. ISSN 0967-3334.

[13] Jalalirad A, Tjalkens T. Using a penalized maximum likelihood model for feature selection. IEEE International Workshop on Machine Learning for Signal Processing 2015;Accepted.

[14] Breiman L. Random Forests. Machine Learning 2001; 45:5–32. ISSN 0885-6125.

Address for correspondence:

Linda Eerikäinen

Eindhoven University of Technology, Department of Electrical Engineering, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

L.M.Eerikainen@tue.nl