False Arrhythmia Alarm Reduction in the Intensive Care Unit Using Data Fusion and Machine Learning*

Qiang Zhang, *Student Member,* Xianxiang Chen, *Member,* Zhen Fang, *Member,* and Shanhong Xia, *Senior Member, IEEE*

Abstract—With aim of reducing the incidence of false critical arrhythmia alarms in intensive care units, a novel data fusion and machine learning algorithm is presented in this article. The 2015 PhysioNet/Computing in Cardiology Challenge database was used in this present algorithm, with each grouped as an asystole (AS), extreme bradycardia (EB), extreme tachycardia (ET), ventricular tachycardia (VT) or ventricular flutter/fibrillation (VF) arrhythmia alarm. A 10-second segment before the onset of the alarm was truncated from available signals, namely electrocardiogram (ECG), arterial blood pressure (ABP), and/or photoplethysmogram (PPG). By first assessing signal quality of available signals, a robust estimation of beat-to-beat intervals could then be derived. Features in heart rate variability (HRV) analysis and ECG parameters such as temporal statistical parameters, spectral analysis results, wavelet transformation coefficients, and complexity measurement etc were extracted and formed a vector. After feature selection through genetic algorithm (GA), a support vector machine (SVM) model was applied to conduct the classification of alarms for the specific arrhythmia type. The overall true positive rate (TPR) of classification algorithm is 93%, with the true negative rate (TNR) 94%. According to the method of performance evaluation in the 2015 Challenge, this algorithm achieved a gross score of 84.4.

I. INTRODUCTION

Electrocardiogram (ECG) analysis is now a routine monitoring tool for cardiovascular diseases, particularly in intensive care units (ICU). However, high false cardiac monitor alarm rates, mainly due to severe corruption by artifacts, noise and missing data, are extremely concerning. According to the publication of Lawless [1], the ICU false alarm (FA) rates might be so high as 86% for some arrhythmia types, with between 6% and 40% clinically insignificant, while only 6% alarms require immediate care. Poor performance of ICU monitoring devices results in care disruption, affecting both patients and medical staff by means of noise disturbances, desensitization to warning and longer response time. Furthermore, such disruptions have been shown to affect recovery of patients [2].

Essentially, there are three kinds of technical approaches to help reduce false alarms [3]: (1) improving signal extraction; (2) improving algorithms for alarm generation; (3) improving alarm validation. In recent years, several detection strategies have been applied to tackle this problem. For example, Aboukhalil et al. [4] and Deshmane [5] applied a multi-parameter analysis on ECG and pulsatile waveforms and signal quality assessment technology to improve algorithms for alarm generation. However, both methods met the problem that alarms of ventricular tachycardia (VT) had high true alarm (TA) suppression rate while low false alarm reduction rate because they only used morphological and timing information. The work described by Sayadi et al. [6] deployed a model-based filtering method to detecting alarms. Superior as the FA suppression rates are, this algorithm is computationally intensive. Qiao Li and Gari D. Clifford [7] extracted features from ECG, arterial blood pressure (ABP), and photoplethysmogram (PPG) and employed a machine learning approach. They achieved a ventricular tachycardia FA suppression of more than 30% with a true alarm suppression rate below 1%.

Figure 1. The diagram to check validity of an alarm.
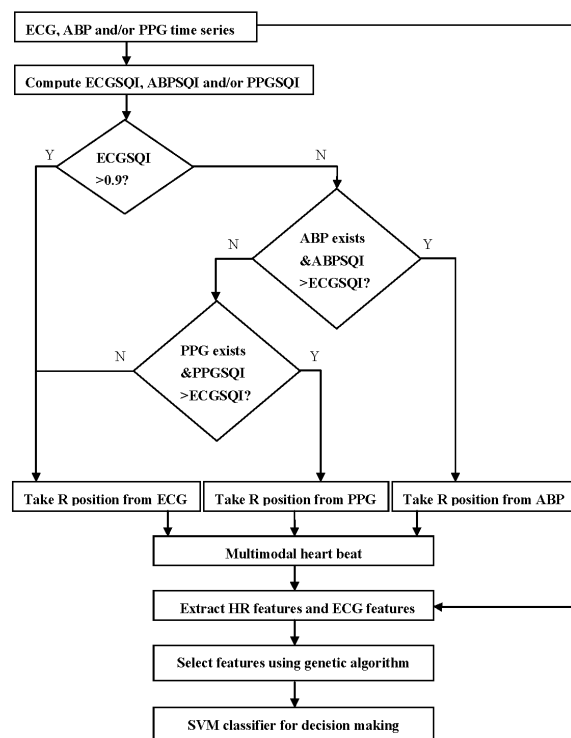
Qiang Zhang is with the Institute of Electronics, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, Beijing, China (email: zhangqiangee@hotmail.com).

Xianxiang Chen, Zhen Fang and Shanhong Xia are with the Institute of Electronics, Chinese Academy of Sciences, Beijing, China (e-mail: cxxvac@163.com; zfang@mail.ie.ac.cn; shxia@mail.ie.ac.cn). Corresponding author: Shanhong Xia (phone: +86-5888-7180; fax: +86-5888-7180).

Alarm Type	All alarms		True Alarms			False Alarms		
	Total alarms	*% of all alarms*	*N*	*% of all alarms that are true*	*% of specific alarm type that are true*	*N*	*% of all alarms that are false*	*% of specific alarm type that are false*
AS	122	16.3	22	2.9	18.0	100	13.4	82.0
EB	89	11.9	46	6.1	51.2	43	5.8	48.8
ET	140	18.7	131	17.5	93.6	9	1.2	6.4
VT	341	45.4	89	11.9	26.0	252	33.5	74.0
VF	58	7.7	6	0.8	10.3	52	6.9	89.7
ALL	750		294	39.2		456	60.8	

The present study aims to build an algorithm for detection of the asystole (AS), extreme bradycardia (EB), extreme tachycardia (ET), ventricular tachycardia (VT) and ventricular flutter/fibrillation (VF) alarms. As is shown in Fig. 1, extraction of beat-to-beat intervals was dependent on the signal quality indices (SQIs) and data fusion techniques from ECG, ABP and/or PPG. The HRV parameters were proposed to supplement rhythm-related information to some previously defined ECG features, and together they formed a feature vector. The genetic algorithm (GA) was employed to select the best feature combination. The optimal variable combinations were then taken as input to SVM classifiers, to label the alarms as true or false.

II. MATERIALS AND METHODS

A. Databases

The 2015 PhysioNet/Computing in Cardiology Challenge[8] has a multimodal database containing typical life-threatening arrhythmia alarm recordings from bedside monitors in the ICU. The arrhythmia types include AS, EB, ET, VT and VF. A total of 750 recordings are used, as is shown in Table I. Each recording consists of two leads of ECG, and at least one pulsatile waveform of either ABP or PPG. An alarm is triggered five minutes from the start of each record. A team of expert annotators reviewed all alarms and agreed to label each alarm either "true" or "false" as the "Gold Standard".

B. Heart Beat Detection

Accurate detection of heart beat may potentially contribute to suppressing false arrhythmia alarms. Therefore, a robust heart beat detection algorithm was employed to derive characteristic point positions from multiple physiological signals [9]. This method fused R peak positions detected on the ECG using an energy detector and onset positions detected on the pulsatile waveforms (ABP and/or PPG) using length transform. Detections were then merged by selecting heart beats from the highest-quality signal. This method was evaluated on the 2014 PhysioNet/Computing in Cardiology Challenge and achieved success across a variety of signal morphologies.

In this context, all signals were resampled (with anti-alias filters) to 250 Hz and noise and artifacts are removed using FIR band pass [0.05 to 40 Hz] and notch filters. We truncated a 10-second segment instantly before the alarm onset from all the ECG leads and pulsatile waveforms. An assessment of signal quality was made after the beat detection. The ECG SQI was based on the agreement of two QRS detectors: jqrs and gqrs [9]. Signal quality for ABP was computed at the beat-level by the signal abnormality index (SAI) algorithm[10]. SAI extracts 13 features from the pulsatile waveform and assign a flag of "1" if the feature does not meet the criteria of normality. The quality of PPG was computed based on correlation with a beat template that is formed by averaging beats detected from the PPG with an open-source script provide for this challenge [8]. To select the most reliable signals, the multimodal signals were compared in a 10-second window, and checked for switching every second.

C. Feature Extraction

Features for the alarm classification were all extracted from the ECG lead with a higher SQI. While HRV parameters were calculated on the whole 300-second data before the alarm onset, other published ECG features were computed in a 10-second segment. All these features can be generally classified into six major types.

1) *HR Temporal Features*
- the maximum, minimum, mean, median, variance, skewness and kurtosis of RR series.

- the number of Premature Ventricular Complexes (NPVCs) , where a beat is defined as a PVC if its RR interval is more than 0.55 mean of proceding RR intervals and less than 0.85 mean of proceding RR intervals.

- the PVCR is denoted as the ratio of PVCs over the whole number of beats

2) *Heart rate variability (HRV) parameters*
- Temporal Domain: the standard deviation of average intervals (SDANN), the standard deviation of successive interval differences (SDSD), and the square root of mean squared interval differences (RMSDD).

- Frequency Domain: the power in the range of 0.04 Hz to 0.15 Hz (PLF), the power in the range of 0.15 Hz to 0.4 Hz (PHF), and the ratio of PLF to PHF (LHR).

3) *ECG Morphological parameters*

- the standard deviation, skewness and kurtosis of the ECG amplitude.

4) *ECG Spectral parameters*

- The ratio of ECG power in 2 Hz – 9 Hz over the whole ECG power (LOCAL_PR).

5) *Complexity parameters*

- Sample entropy (SpEn) measures the similarity within an ECG timel sequence. A higher value of SpEn means less self-similarity.

- Bin Complexity measurement (BCM) is defined as the normalized value of the Lempel-Ziv complexity measurement of a binary data derived from the ECG signal segment.

- Phase space reconstruction (PSR) evaluates the system chaos of original and a time-delay ECG signal.

6) *Wavelet transform coefficient parameter*

Features derived from wavelet transform were calculated by decomposing the ECG signal with higher SQI into wavelet coefficients using a discrete wavelet transform (DWT). We employed a 4-level multiresolution analysis with use of the wavelet Daubechies 2. The majority of ECG energy signal lies in the range 0.5 Hz - 40 Hz. Hence, coefficients corresponding to A4, D4, and D3 were retained, which corresponded to 0 – 15.625 Hz, 15.625 Hz – 31.25 Hz, and 31.625 Hz – 62.5Hz, respectively. Then, the ration of power in 0 – 62.5 Hz over the total power (0 – 250 Hz) was calculated as the WAVECOEF_RP.

D. Feature Selection

As it is unlikely that all calculated features are helpful in false alarm reduction (and in fact, some features might lead to a poorer performance of alarm classifiers), a mechanism of variable selection is in need. Therefore, a genetic algorithm (GA) was employed to select the optimal combination of variables to determine validity of an alarm. In the GA, a "chromosome" is defined as a binary vector with each element "1" representing a selected feature and "0" representing not selected. By randomly "mutating" chosen chromosomes, the GA efficiently explores the space of variable combinations. A total population of 100 chromosomes was applied and the mutation rate was 5%, the cloning rate was also 5%, the cull rate for crossover was 95%, and a 100-generation limit was considered; the potential variable combinations was quickly investigated in the search space. The fitness function was the root mean squared error (rMSE) of a multivariate logistic regression, which was minimised. The GA selection procedure was conducted for 150 times, and the selected variables were sorted according to the selected times of 150 genetic algorithm repeats. Table II shows the rank of 15 variables, including PVCR, PSR, RR_MAX, PLF, RR_MIN, PHF, ECG_STD, BCM, SDANN, RR_STD, WAVECOEF_RP, LOCAL_PR, RMSDD, NPVCs, and RR_MEAN. These features were all selected more than 50% by the GA runs, and were also eventually used as input to SVM decision making.

E. Machine Learning for False Alarm Reduction

Choosing Support vector machine (SVM) classifiers (LIBSVM library) [12] over other machine learning methods is justified by the generalization capability originating from the structural risk minimization principle. All feature values were generalized with zero mean and unity standard deviation, and then input to the SVM classifies. While an SVM with a linear kernel was used for AS, EB, ET and VF alarm classification, a SVM with Gaussian (nonlinear) kernel was defined for the VT alarm. The SVM model with a Gaussian kernel has two parameters: C and γ. We used the Grid Search Algorithm to determine these two parameters: C=2 and γ=0.2. Each combination of selected features by the GA was input into the SVM model to train classifiers.

In the algorithm validation phase, models were evaluated with the use of fivefold cross validation. The fivefold cross validation divides the total recordings into five folds, with four folds used for training and the last fold for evaluation. Every fold will be selected as the evaluation fold in turn. The fivefold cross validation procedure was repeated 50 times at random, and the average performance of classifiers was referenced for evaluation.

TABLE II.   ORDERED RANKING OF SELECTED VARIABLES BY GA

Rank	Variable name	Selected times
1	PVCR	148
2	PSR	147
3	RR_MAX	147
4	PLF	140
5	RR_MIN	138
6	PHF	130
7	ECG_STD	125
8	BCM	113
9	SDANN	110
10	RR_STD	109
11	WAVECOEF_RP	102
12	LOCAL_PR	99
13	RMSDD	89
14	NPVCs	85
15	RR_MEAN	78

TABLE III.   MEAN PERFORMANCE OF FALSE ALARM SUPPRESSION ALGORITHM

Arrhythmia type	TP	FP	FN	TN	AC	SE	SP	Fivefold cross validation accuracy	Score
AS	0.156	0.049	0.024	0.771	0.959	0.864	0.940	81.967%	84.3
EB	0.494	0.034	0.022	0.450	0.944	0.957	0.929	76.404%	86.6
ET	0.936	0.000	0.000	0.064	1.000	1.000	1.000	97.857%	100
VT	0.214	0.038	0.047	0.701	0.915	0.820	0.948	85.337%	77.0
VF	0.087	0.017	0.017	0.879	0.966	0.836	0.981	93.276%	90.4
Gross	0.364	0.306	0.293	0.577	0.940	0.925	0.949	86.613%	84.4

TABLE IV.   COMPARISON OF DIFFERENT ALGORITHMS ON FALSE ALARM SUPPRESSION

Algorithm	TPR (%)	TNR (%)	Score
Plesinger et al.[a]	92	88	81.39
Kalidas & Tamil[a]	**94**	82	79.44
Krasteva et al.[a]	93	83	79.41
Couto et al.[a]	89	91	79.02
Fallet et al.[a]	**94**	77	76.11
This work	93	**94**	**84.4**

a. Top candidate entry in the 2015 PhysioNet/Computing in Cardiology Challenge

According to the 2015 PhysioNet/Computing in Cardiology Challenge, performance of FA suppression is quantified by accuracy (AC), sensitivity (SE), specificity (SP) and a score, which is calculated from the following variables: true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). More specifically, due to the severity of ignoring a real life-threatening arrhythmia alarm, the FN is multiplied by a punishment factor, i.e. 5, and the final score is implemented by the equation:

$$\text{Score} = 100 * (TP + TN) / (TP + TN + FP + 5*FN). \quad (1)$$

III. RESULTS

Using the 15 most selected features, fivefold cross validation was repeated 50 times randomly to examine the average performance of false alarm classifiers, and the results are shown in Table III.

This algorithm showed good performance on ET and VF false alarm suppression with high scores and fivefold cross validation accuracy. As for AS and EB, while classifiers achieved very high scores, both of the corresponding cross validation are moderate. The VT alarms are classified well with 82% sensitivity, 94.8% specificity and more than 85% fivefold cross validation accuracy. Overall, in comparison with the top candidate entries of the 2015 PhysioNet/Computing in Cardiology Challenge, this algorithm shows promising results, as is demonstrated by Table IV.

IV. CONCLUSION

Data fusion and signal quality provide an accurate detection of heart beat, which is proved to help in a decision-making process. Though the selected features are defined previously, their combination describes properties from various aspects and may reveal more difference between true and false alarms. Among the features, the HRV analysis parameters lift the performance of alarm classifiers to a great extent. Besides, the average time for extracting features from a recording is 1.63 second, thus, the alarm classification will be finished within 2 seconds when an alarm is triggered. Similar results are expected to present in other arrhythmias, such as atrial fibrillation, and further work will be concentrated on clinical experiments and applications.

REFERENCES

[1] Lawless, Stephen T. "Crying wolf: false alarms in a pediatric intensive care unit." Critical care medicine 22.6 (1994): 981-985.

[2] Donchin, Yoel, and F. Jacob Seagull. "The hostile environment of the intensive care unit." Current opinion in critical care 8.4 (2002): 316-320.

[3] Schmid, Felix, Matthias S. Goepfert, and Daniel A. Reuter. "Patient monitoring alarms in the ICU and in the operating room." Crit Care 17.2 (2013): 216.

[4] Aboukhalil, Anton, et al. "Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform." Journal of biomedical informatics 41.3 (2008): 442-451.

[5] Deshmane, AnaghaVishwas. False arrhythmia alarm suppression using ECG, ABP, and photoplethysmogram. Diss. Draper Laboratory, 2009.

[6] Sayadi, Omid, and Mohammad BagherShamsollahi. "Life-threatening arrhythmia verification in ICU patients using the joint cardiovascular dynamical model and a Bayesian filter." Biomedical Engineering, IEEE Transactions on 58.10 (2011): 2748-2757.

[7] Li, Qiao, and Gari D. Clifford. "Signal quality and data fusion for false alarm reduction in the intensive care unit." Journal of electrocardiology 45.6 (2012): 596-603.

[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". In Comput in Cardiol. 2015.

[9] Johnson, Alistair EW, et al. "Multimodal heart beat detection using signal quality indices." Physiological measurement 36.8 (2015): 1665.

[10] Sun, J. X., A. T. Reisner, and R. G. Mark. "A signal abnormality index for arterial blood pressure waveforms." Computers in Cardiology, 2006. IEEE, 2006.

[11] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: A library for support vector machines." ACM Transactions on Intelligent Systems and Technology (TIST) 2.3 (2011): 27.s