Reducing false arrhythmia alarm rates using robust heart rate estimation and cost-sensitive support vector machines

Qiang Zhang[1,2]**,Xianxiang Chen**[1]**,Zhen Fang**[1] **and Shanhong Xia**[1]

[1] Institute of Electronics, Chinese Academy of Sciences, Beijing, People's Republic of China

[2] University of Chinese Academy of Sciences, Beijing, People's Republic of China

E-mail: `shxia@mail.ie.ac.cn`

Abstract.
To lessen the rate of false critical arrhythmia alarms, we used robust heart rate estimation and cost-sensitive support vector machines. The PhysioNet's MIMIC II database and the 2015 PhysioNet/CinC Challenge public database were used as the training dataset and the test dataset, respectively. Each record had an alarm labeled with asystole, extreme bradycardia, extreme tachycardia, ventricular tachycardia or ventricular flutter/fibrillation. Before alarm onsets, 300-second multimodal data was provided, including electrocardiogram, arterial blood pressure, and/or photoplethysmogram. A signal quality modified Kalman filter fulfilled robust heart rate estimation. Based on that, we extracted heart rate variability features and statistical ECG features. Next, we applied a genetic algorithm (GA) to select the optimal feature combination. Finally, considering the high cost of classifying a true arrhythmia as false, we selected cost-sensitive support vector machines (CSSVM) to classify alarms. Evaluation on the test dataset showed the overall true positive rate was 92.5%, and the true negative rate was 92.3%.

PACS numbers: 87.85.Ng

Keywords: false arrhythmia alarm reduction, multimodal data, robust heart rate estimation, cost-sensitive support vector machines

1. Introduction

Electrocardiogram (ECG) analysis is now a routine monitoring tool for cardiovascular diseases in intensive care units (ICU). Because of severe corruption by artifacts and noise, high rates of false cardiac monitor alarms in ICU are very concerning. Poor performance of ICU monitoring devices leads to care disruption, affecting both patients and medical staff by noise disturbances, desensitization to warning and longer response time. Further, such disruptions have been shown to affect recovery of patients (Donchin and Seagull 2002). This problem is so severe that it was selected as the central issue of the 2015 PhysioNet/CinC Challenge (Clifford *et al* 2015).

Essentially, there are three kinds of technical approaches to help reduce false alarms (FA) (Schmid *et al* 2013): (1) improving signal extraction; (2) improving algorithms for alarm generation; (3) improving alarm validation. Reports about false arrhythmia alarm rates date back to 1994 by Lawless (Lawless 1994). He found the ICU FA rates might be as high as 86% for some arrhythmia types, with between 6% and 40% clinically insignificant, while only 6% alarms need immediate care. In recent years, several strategies have been developed to tackle this problem. For example, Aboukhalil et al. (Aboukhalil *et al* 2008) and Deshmane (Deshmane 2009) applied a multi-parameter analysis on ECG and pulsatile waveforms to improve algorithms for alarm generation. However, both methods met the problem that alarms of ventricular tachycardia had high true alarm (TA) suppression rates while low FA reduction rates. The work described by Sayadi et al. (Sayadi and Shamsollahi 2011) used a model-based filtering method. Superior as the FA suppression rates are, this algorithm is computationally intensive. In 2012, Qiao Li and Gari D. Clifford (Li and Clifford 2012) extracted features from multimodal signals and employed a machine learning approach. They reduced more than 30% of false ventricular tachycardia alarms with TA suppression rates below 1%.

The 2015 PhysioNet/CinC set the annual challenge to reduce false critical arrhythmia alarms in ICU, and attracted 38 research teams from the world. For the real-time mode (i.e., using no information after sounding of an alarm), Filip Plesinger (Plesinger *et al* 2015) was the winner with 92% true positive rate and 88% true negative rate. He used elementary algebra, descriptive statistics and fuzzy logic. Kalidas (Kalidas and Tamil 2015) and Fallet (Fallet *et al* 2015) both gained the highest true positive rate of 94%, and Coutu (Couto *et al* 2015) made the highest true negative rate 91%. Varying contestants ranked highest in each separate alarm category, indicating that there was no best general algorithm.

Solely using information before alarms, the present study built arrhythmia-specific models to decrease FA rates. Figure 1 shows the general scheme of this algorithm. Data from the MIMIC II database was used for model training and the 2015 PhysioNet/CinC Challenge public dataset (the Training Dataset) was used for testing. We fulfilled a signal quality modified Kalman filter to obtain robust heart rate estimation. We extracted features including heart rate variability (HRV) parameters and statistical ECG variables. Since it is not likely that all features are helpful to classification, we used a genetic algorithm (GA) to select the best feature combination. It means unredeemable loss to ignore a true arrhythmia alarm. Therefore, we chose cost-sensitive support vector machines (CSSVM) for decision making.

The structure of this paper is organized as follows. Section II provides a description of data source and outlines fundamental procedures to build the present model. In Section III, performance of the proposed method is examined. Finally, Section IV presents the discussion and conclusion.

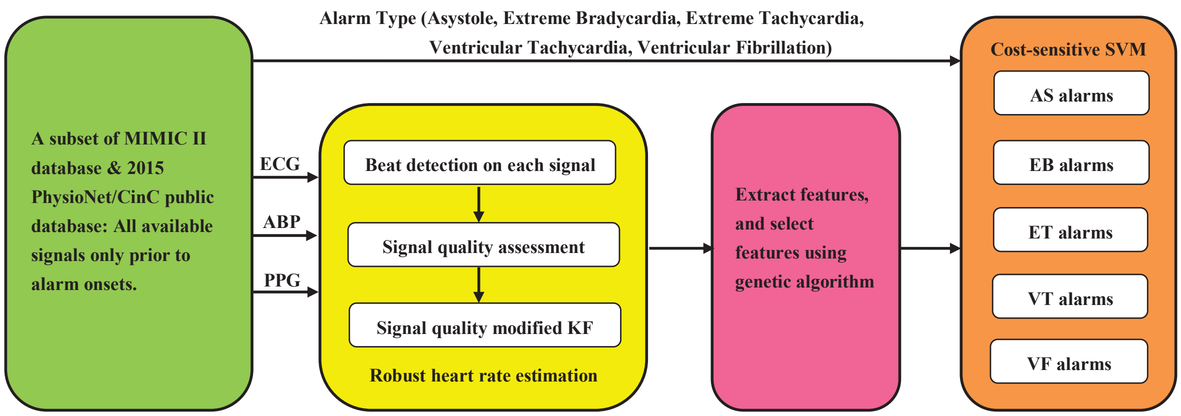

Figure 1. General scheme for reducing false arrhythmia alarms using robust heart rate estimation and cost-sensitive support vector machines.

2. Methods

2.1. Data

Major focus is on five types of life-threatening arrhythmia: asystole (AS), extreme bradycardia (EB), extreme tachycardia (ET), ventricular tachycardia (VT), and ventricular fibrillation/flutter (VF). AS is the case that no heart beats at all for four seconds or more. EB occurs when the patients heart rate is lower than 40 beats per minute for five consecutive beats. ET is detected when the heart rate is higher than 140 beats per minute for 17 consecutive beats. VT is defined as the case that five or more consecutive ventricular beats with heart rate higher than 100. VF happens when the heart presents a rapid fibrillatory, flutter, or oscillatory waveform for at least four seconds.

We selected a subset of the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II) database as our training dataset. The 2015 PhysioNet/CinC Challenge public dataset (the Training Dataset) was the test dataset. The MIMIC II database was established to promote research of ICU systems. It has experts-annotated alarms that were suitable for development of life-threatening arrhythmia alarm detection strategy. Records that satisfied following requirements were selected: one of the five kinds of arrhythmia alarms is triggered; ECG and pulsatile signals (ABP and/or PPG) are simultaneously recorded; and signals are present continuously through a 5-minute time window before the alarm onset. The final subset of MIMIC II contained 2829 labeled alarms, with similar FA rates in the test dataset.

The 2015 PhysioNet/CinC Challenge has a public dataset (the Training Dataset) that contains typical life-threatening arrhythmia alarm recordings in the ICU monitors. In total, 750 ICU recordings fall into five kinds of arrhythmia. Each recording has 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

Table 1. Distribution of selected MIMIC II data recordings

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	279	9.9	43	1.5	15.4	236	8.3	84.6
EB	402	14.2	204	7.2	50.7	198	7.0	49.3
ET	747	26.4	682	24.1	91.3	65	2.3	8.7
VT	1128	39.9	245	8.7	21.7	883	31.2	78.3
VF	273	9.7	34	1.2	12.5	239	8.4	87.5
ALL	2829		1208	42.7		1621	57.3	

Table 2. Distribution of 2015 PhysioNet/CinC public data recordings

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	

reviewed all alarms and agreed to label each alarm either true or false. Distribution of the training set and the test set with alarm types and numbers are presented in table 1 and table 2, respectively.

2.2. Robust Heart Rate Estimation

As pointed out by Clifford (Clifford *et al* 2015), key to rhythm detection is accurate heart rate estimation. Because of noise or arrhythmia, the ECG morphology can be so complex that heart rate estimation could be very difficult. Therefore, a robust heart rate estimation algorithm is in need. Generally, the heart rate estimation consists of heart beat detection and heart rate calculation. In this paper, we used a signal quality modified Kalman filter to fulfill robust heart rate estimation.

2.2.1. Heart beat detection We detected heart beats on all available signals. Noise and artifacts were removed by FIR band pass filter (0.05 to 40 Hz) (Clifford *et al* 2015). We used the 'eplimited' algorithm (Pan and Tompkins, 1985, Hamilton and Tompkins,

1986) to detect R waves, and the open source 'wabp' algorithm (Zong *et al* 2003) to detect onset positions of pulsatile waveforms (ABP and/or PPG), as shown in figure 2. Then we calculated time intervals between consecutive characteristic points as heart beat intervals.

2.2.2. Signal quality assessment After beat detection, signal quality indice (SQI) was calculated. Since different QRS detection algorithms are sensitive to different types of noise, the SQI for ECG (ECGSQI) was based on the agreement of two QRS detectors: 'eplimited' and 'wqrs' (Li *et al* 2008, Johnson *et al* 2015). With a 10-second window, ECGSQI was defined as the ratio of beats detected simultaneously by both algorithms (within an interval of 150 ms) over all detected beats. This method is routinely known as the 'bSQI'. The SQI for ABP (ABPSQI) was determined by the signal abnormality index (SAI) algorithm (Sun *et al* 2006). SAI extracts a series of features in each ABP waveform pulse, such as systolic blood pressure and beat-to-beat duration variations. We modified SAI by first assigning a flag of '1' if a parameter met the criteria of physiological normality, and then defined ABPSQI through dividing the number of flag '1' by the number of all flags. The SQI for PPG (PPGSQI) was computed based on correlation with a beat template that was formed by averaging beats detected from the PPG (Clifford *et al* 2015). When ABP or PPG is absent, then corresponding SQI will be set zero.

2.2.3. Signal quality modified Kalman filter We performed robust heart rate estimation by adapting the signal quality modified Kalman filter (Li *et al* 2008). In details, three Kalman filter extracted heart rate from ECG, ABP and PPG separately. SQI (ECGSQI, ABPSQI, or PPGSQI) was used to adjust the measurement noise covariance to more heavily weight on cleaner signal.

$$R = R \cdot \exp(1/SQI^2 - 1) \tag{1}$$

At high SQI, the exponential factor tends to be unity, pushing Kalman filter to trust the current measurement. As SQI tends low, R tends to be infinity (but in practice, R is limited to a large value), forcing Kalman filter to lower Kalman gain and trust previous value more. In addition, if current SQI is below certain threshold, i.e., SQI_{th}, then the Kalman filter will not get updated. Once the heart rate was derived from ECG, ABP or PPG separately, they were merged to produce a consensus estimation. The SQI-weighted residual of Kalman filter, r, were considered as combination weights, like

$$HR = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2+\sigma_3^2}HR_1 + \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2+\sigma_3^2}HR_2 + \frac{\sigma_3^2}{\sigma_1^2+\sigma_2^2+\sigma_3^2}HR_3 \tag{2}$$

where $\sigma_1 = ECGSQI/r_1$, $\sigma_2 = ABPSQI/r_2$ and $\sigma_3 = PPGSQI/r_3$, HR_1 is from ECG, HR_2 is from ABP and HR_3 is from PPG, r_1, r_2 and r_3 are residual of Kalman filter for ECG, ABP and PPG, respectively. By relying on Kalman filter estimation with bigger σ, the estimation of heart rate is more robust than either of that derived from individual Kalman filter.

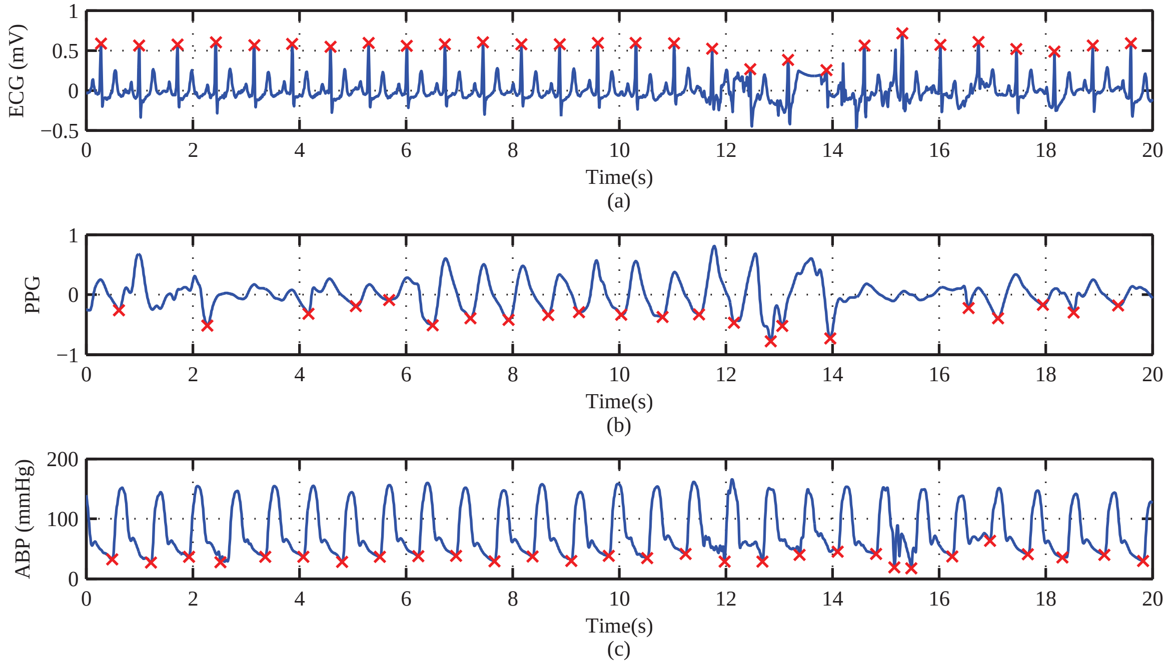

Figure 2. ECG, PPG, and ABP signal along with detected heart beats.

2.3. Feature extraction

Features for alarm classification were all extracted from the ECG lead with a higher SQI. While HRV parameters were calculated from the whole 300-second data, other ECG features were computed in a 10-second segment. All these features can be divided into five groups.

2.3.1. HR temporal features

- The maximum, minimum, mean, median, standard deviation, skewness and kurtosis of RR series are denoted as RR_MAX, RR_MIN, RR_MEAN, RR_MED, RR_STD, RR_SK, RR_KT, respectively.
- The number of premature ventricular complexes (NPVCs). The PVC is an extra electrical impulse arising from the cardiac ventricles before the next normal heart beat occurs. A beat was defined as a PVC if its RR interval was more than 0.55 mean of preceding RR intervals and less than 0.85 mean of preceding RR intervals. This parameter was ever found associated with atrial fibrillation (Sabeti *et al* 2012).
- Threshold crossing interval (TCI) (Thakor *et al* 1990) refers to the time interval between consecutive crossing-threshold pulses within 1-second ECG segments. A threshold was set to 20% of the maximum value within one second and was recalculated every second. A binary signal was generated from the ECG data according to the position of the signal above or below the threshold. Next, data analysis took place over successive one-second segments. For every segment, the TCI was the average interval between threshold crossings.

- Threshold crossing sample count (TCSC) (Arafat *et al* 2011) is the number of samples that cross a given threshold within a one-second ECG interval. The ECG signal could cross the detection threshold one or more times, and the number of pulses were counted.

2.3.2. 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 over the range of 0.04 Hz to 0.15 Hz (PLF), the power over the range of 0.15 Hz to 0.4 Hz (PHF), and the ratio of PLF to PHF (LHR).

2.3.3. ECG morphological parameters

- The standard deviation, skewness and kurtosis of the ECG amplitude are represented as ECG_STD, ECG_SK, ECG_KT, respectively.

2.3.4. ECG spectral parameters

- The ratio of ECG power between 2 Hz to 9 Hz over the whole ECG power is defined as LOCAL_PR.

- VF filter leakage (VFleak) (Kuo and Dillman 1978) measures the residue after applying a narrow-band elimination filter that is centered in the region of the mean frequency of an ECG segment.

- Median frequency (FM) (Dzwonczyk *et al* 1990) is the central frequency of the power spectrum of the ten-second ECG segment. This parameter estimated the duration of the cardiac arrest. Since it provides information on the duration of the VF episode, we included it here to analyze its discriminatory characteristics.

2.3.5. Complexity parameters

- Sample entropy (SpEn) (Li *et al* 2009) measures the similarity within an ECG time sequence. A higher value of SpEn means less self-similarity. This parameter evaluates the irregularity/complexity within the nonlinear dynamic cardiovascular system.

- Bin complexity measurement (BCM) is defined as the normalized value of Lempel-Ziv complexity measurement of a binary data. The binary data string was generated by comparing the ECG signal with a defined threshold. The Lempel-Ziv complexity and its variants are popular measures for characterizing the randomness of biomedical signals.

- Phase space reconstruction (PSR) evaluates the system chaos of original and a time-delay ECG signal. We first down sampled the 10-second ECG data to 50 Hz, so we got 500 sample points, and the embedded dimension was 500. The delay time interval in this study was set to 1 second.

2.4. Feature selection

Since it is not likely that all features are helpful in false alarm reduction, in fact, some features might lead to a poorer performance of classifiers. Therefore, a genetic algorithm (GA) was used to select the optimal variable combination (Goldberg and Holland 1988). In the GA, a chromosome is 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 chromosomes was 100, 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 after the trial period. The fitness function was the root mean squared error (rMSE) of a multivariate logistic regression, which was minimized. The GA selection procedure was conducted for 150 times, and the selected variables were sorted according to the chosen times. Table 4 shows the rank of selected variables. These features were all selected more than 50% by the GA runs, and were also eventually input to the CSSVM for decision making.

2.5. Machine learning for false alarm reduction

Based on the principle of structural risk minimization (Vapnik 2013), the support vector machine (SVM) algorithm has strong generalization capability and has been successful in areas like image retrieval, handwriting recognition and text classification. Therefore, in this work, we consider the SVM architecture. To better reflect the different cost of alarm classification results, a cost-sensitive method was used. The most widely researched approach of making SVM cost-sensitive is by introducing different penalties C_1 and C_{-1} for the positive and negative slack factors during training. The dataset can be reformed as $(x_i, y_i, \mathrm{cost}_i)$, where $x_i \in \mathrm{R}^l$, $y_i \in \{1, -1\}$, and

$$\mathrm{cost}_i = \begin{cases} \mathrm{C}_1 & \text{for } y_i = 1, \\ \mathrm{C}_{-1} & \text{for } y_i = -1. \end{cases} \tag{3}$$

Assuming the training set can be classified by the hyper plane

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b = 0 \tag{4}$$

then the standard SVM problem can be transformed into

$$\begin{aligned} &\arg_{\boldsymbol{w},\xi,b} \max \frac{1}{2}\|\boldsymbol{w}\|^2 + \mathrm{C}[\mathrm{C}_1 \sum_{\{i|y_i=1\}} \xi_i + \mathrm{C}_{-1} \sum_{\{i|y_i=-1\}} \xi_i] \\ &\text{s.t.} \qquad y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b) \geq 1 - \xi_i \end{aligned} \tag{5}$$

where $\| \boldsymbol{w} \|^2$ is the structural risk, $\mathrm{C}[\mathrm{C}_1 \sum_{\{i|y_i=1\}} \xi_i + \mathrm{C}_{-1} \sum_{\{i|y_i=-1\}} \xi_i]$ represents the empirical risk that considers different misclassification cost, and C is the slack factor that controls balance between the two kinds of risk. To solve the optimization problem of (4), we can solve the Lagrange dual form as

$$\begin{aligned} &\arg_\alpha \max \sum_i \alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \mathrm{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ &\text{s.t.} \quad \begin{aligned} &\sum_i \alpha_i y_i = 0 \\ &0 \le \alpha_i \le \mathrm{cost}_i \mathrm{C} \end{aligned} \end{aligned} \tag{6}$$

The optimal classifier can be

$$f(x) = \mathrm{sgn}(\sum(\alpha_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b)) \tag{7}$$

In this algorithm, classifying a true alarm as false was set to incur a loss of 5 (i.e., $\mathrm{C}_1 = 5$), and classifying a false alarm as true was set to incur a loss of 1 (i.e., $\mathrm{C}_{-1} = 1$). All features were generalized with zero mean and unity standard deviation, and then input to the CSSVM classifies. After trial and comparison, a CSSVM model with a linear kernel was selected for AS, EB, ET and VF alarm classification, and a Gaussian (nonlinear) kernel was defined for the CSSVM model to classify VT alarms. We used the grid search algorithm to determine two parameters: $C = 2.5$ and $\gamma = 0.3$. In the validation phase, models were evaluated with the test dataset.

2.6. Statistical measures

Statistical measures used in this paper were all calculated from primary variables, including true positive (TP), true negative (TN), false positive (FP), and false negative (FN). First, accuracy (AC) is the proportion of correctly identified with

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

true positive rate (TPR) evaluates the proportion of true alarms that have been correctly classified with

$$TPR = \frac{TP}{TP + FN} \tag{9}$$

and true negative rate (TNR) corresponds to the proportion of false alarms that have been correctly classified with

$$TNR = \frac{TN}{TN + FP} \tag{10}$$

Then according to the 2015 PhysioNet/CinC Challenge, performance of FA suppression is quantified by a score. 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:

$$Score = \frac{100 \cdot (TP + TN)}{TP + TN + FP + 5FN} \tag{11}$$

Third, one of the popular metrics to evaluate classifiers' performance is the receiver operating characteristic (ROC) curve. It depicts visual representation of a relative

Table 3. Ordered ranking of selected features

Rank	Variable name	Variable meaning	Selected times
1	RR_MAX	The maximum value of RR series	145
2	FM	Median frequency	140
3	RR_MIN	The minimum value of RR series	138
4	PLF	Power in range of 0.04 to 0.15 Hz in HRV analysis	137
5	TCI	Threshold crossing interval	130
6	RMSDD	The square root of mean squared interval differences in HRV analysis	123
7	RR_MEAN	The mean value of RR series	123
8	PSR	Phase space reconstruction variable	113
9	PHF	Power in the range of 0.15 to 0.4 Hz in HRV analysis	112
10	SDANN	Standard deviation of average intervals in HRV analysis	110
11	BCM	Bin complexity measurement	105
12	VFleak	VF filter leakage	105
13	PVCR	The ratio of PVCs over the whole number of beats	101
14	RR_STD	The standard deviation of RR series	95
15	LOCAL_PR	The ratio of ECG power in 2 to 9 Hz over the whole ECG power	82

trade-off. The number of positive examples correctly classified can be increased at the expense of introducing additional false positives. TPR is plotted on the y-axis denoting the percentage of correctly classified positive instances, and false positive rate (FPR, i.e., $1 - TNR$) is plotted on the x-axis meaning the percentage of misclassified negative instances. According to the ROC structure, point (0,1) represents a perfect classification result. Perhaps the most common metric in ROC analysis is using the area under the ROC curve (AUC) to assess overall classification performance.

3. Results

3.1. Features selection results

Table 4 shows the rank of selected variables, including RR_MAX, FM, RR_MIN, PLF, TCI, RMSDD, RR_MEAN, PSR, PHF, SDANN, BCM, VFleak, PVCR, RR_STD, and LOCAL_PR. These features were all selected more than 50% by the GA runs, and were also eventually input to CSSVM decision making.

3.2. Classifiers performance evaluation

Using the 15 most selected features, 750 recordings from the 2015 PhysioNet/CinC Challenge examined the performance of false alarm detectors, and the results are shown in table 5.

Generally, the present algorithm achieved promising performance with 0.925 TPR

Table 4. Statistical metrics for performance test of the false alarm suppression algorithm

Arrhythmia type	TP	FP	FN	TN	AC	TPR	TNR	$Score$
AS	0.156	0.041	0.025	0.779	0.934	0.864	0.950	85.1
EB	0.483	0.022	0.033	0.461	0.942	0.935	0.953	83.2
ET	0.914	0.007	0.021	0.057	0.971	0.977	0.889	89.5
VT	0.229	0.070	0.032	0.669	0.897	0.876	0.905	79.5
VF	0.069	0.052	0.034	0.845	0.914	0.667	0.942	80.3
Gross	0.363	0.047	0.029	0.561	0.924	0.925	0.923	82.7

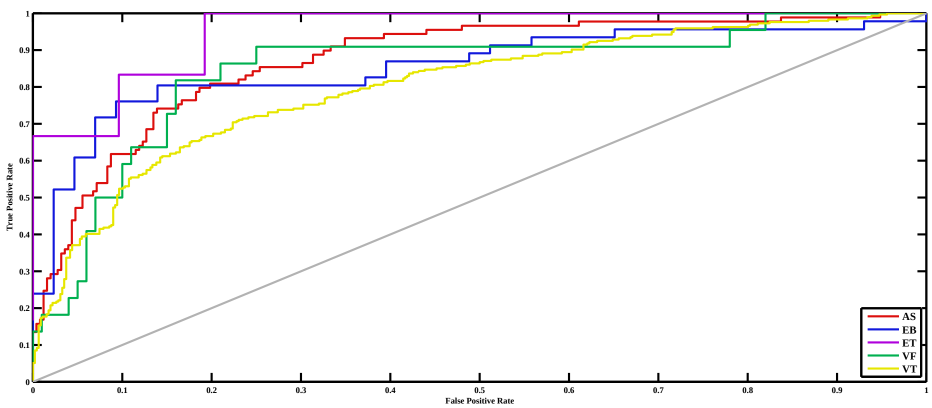

Figure 3. ROC curves of five specific arrhythmia alarm classifiers.

and 0.923 TNR. In details, this algorithm reduced FA rates for ET with the highest score (89.5) and the highest TPR (0.977), though the TNR is only 0.889. Another well-behaved classifier is for AS alarms, it obtained 0.950 TNR and 0.864 TPR. In the third place comes the classifier for EB alarms, which achieved the highest TNR (0.953) and 0.935 TPR. The algorithm got 0.942 TNR for VF alarms with an overall score of 80.3, but a low TPR (0.667). VT alarms were comparably least well classified, with the lowest score (79.5) and 0.905 TNR. Frankly, it ignored the biggest amount of true alarms, although its TPR (0.876) is not the lowest.

Figure 3 shows the ROC curve of arrhythmia alarm classifiers. In this context, the FPR should be close to 0; that is, no true alarms are suppressed. The AUC metric of classifiers for AS, EB, ET, VT, and VF are 0.8738, 0.8564, 0.9219, 0.8007, and 0.8250, respectively.

3.3. Effect evaluation of HRV parameters

All features were extracted from a 10-second segment except that HRV parameters were calculated from the whole 300-second data. The importance of adding HRV features were evaluated by classifiers scores, as shown in figure 4. HRV parameters

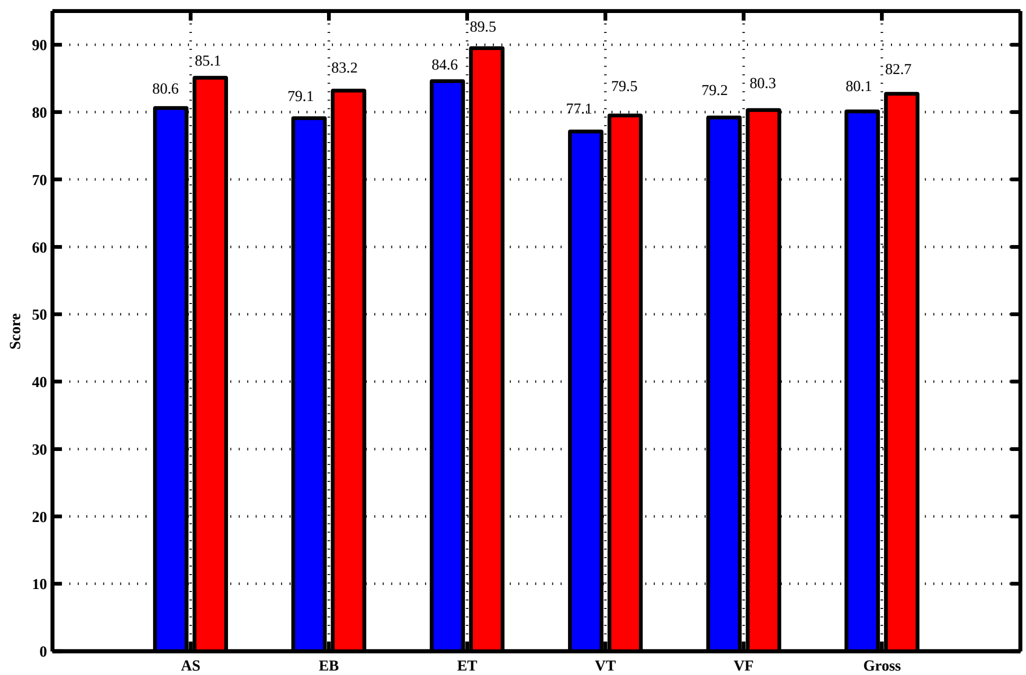

Figure 4. Effect evaluation of HRV features on test dataset. Blue bars represent trained classifiers without HRV parameters and red bars represent classifiers trained with HRV parameters.

lifted performance of all arrhythmia alarm classifiers, but the classifier for EB alarms gained the biggest lift in terms of scores.

4. Discussion and conclusions

Among three kinds of approaches to help reduce false alarms, as mentioned in Introduction Section, the present algorithm aims to improve alarm validation. The robust heart rate estimation and the cost-sensitive machine learning method showed promising performance in deciding validity of critical arrhythmia alarms.

Signal quality assessment was widely used by most of the top algorithms in the 2015 PhysioNet/CinC Challenge. Filip Plesinger (Plesinger *et al* 2015) detected invalid data slices and ignored them in following steps. Kalidas (Kalidas and Tamil 2015) used signal quality analysis to check presence of flat lines and zigzag lines. Krasteva (Krasteva *et al* 2015) defined a lead quality module for detection of spike noise, powerline interference and baseline wander. These approaches are all basic applications of quality assessment, simply for artifact detection or to ignore low-quality data. Fallet (Fallet *et al* 2015) estimated heart rate based on signal quality, but limitations lie in the absence of ECG quality assessment. Unlikely, we evaluated all available signals, and embedded SQI into robust heart rate estimation.

It has been a common problem to reduce FA rates for ventricular arrhythmia. Because of morphological changes of QRS complexes, previous studies reported that heart rate was not sufficient to classify ventricular tachycardia and ventricular fibrillation episodes (Aboukhalil *et al* 2008, Li *et al* 2014, Salas *et al* 2014) and additional features were required. Fallet (Fallet *et al* 2015) used ECG spectral purity index (SPI) to quantity morphological changes. Krasteva (Krasteva *et al* 2015) designed a beat classification module to provide prior knowledge about supraventricular beats or ventricular beats. In the present algorithm, we did not specifically design a module to predetermine the beat type, but drew support from complexity measurement parameters that has been successfully applied in describing chaos theory. In table 3, GA selection results display effectiveness of complexity features.

The present algorithm picked features solely from ECG, but other physiological parameters available in ICU, such as diastolic blood pressure extracted from ABP and blood oxygen saturation extracted from PPG, may also provide discriminative information. New physiological features can enrich this frame for other arrhythmia alarm detection, such as atrial fibrillation.

As for decision making, some studies kept decision-making straightforward by setting physiological thresholds. Fallet (Fallet *et al* 2015) did similarly, but he suggested the use of a more elaborated model, such as machine learning methods. We also noted that different costs of false negative and false positive were used in his linear discriminant analysis. Inspired by this, we adopted a cost-sensitive machine learning method for false alarm reduction.

Owing to the urgency of arrhythmia alarms in ICU, the monitor must make decision as soon as possible. In the present algorithm, the average time for extracting features from a recording is 1.63 second on the Matlab 2012b in the 32-bit Windows 7 operating system with an Intel i5 CPU and 4G RAM. Thus, the alarm classification will be finished within 2 seconds after an alarm is triggered. This ensures that alarms are validated as true or false in less than 10 seconds, according to AAMI guidelines (ANSI/AAMI EC13:2002).

In conclusion, the robust heart rate estimation and cost-sensitive machine learning method have shown promising outlook for false arrhythmia alarm suppression. HRV parameters provided discriminative information about arrhythmia alarms. In the case of medical diagnosis, a cost-sensitive approach is obviously more suitable than conventional methods.

Acknowledgments

We thank Dr. Joachim Behar and Dr. Alistair Johnson from University of Oxford for their valuable discussion about the multimodal approach to heart beat detection. This work is supported by the National High Technology Research and Development Program of China (863 program-2012AA040506, 2013AA041201), and the National Natural Science Foundation of China (No.61302033).

References

Aboukhalil A, Nielsen L, Saeed M, Mark R G and Clifford G D 2008 Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform *J. Biomed. Inform.* **41** 442-51

American National Standard (ANSI/AAMI EC13:2002) 2002 Cardiac monitors, heart rate meters, and alarms. *Arlington, VA: Association for the Advancement of Medical Instrumentation*

Arafat M A, Chowdhury A W and Hasan M K 2011 A simple time domain algorithm for the detection of ventricular fibrillation in electrocardiogram. *Signal Image Video P.* **5** 1-10

Behar J, Oster J, Li Q and Clifford G D 2013 ECG signal quality during arrhythmia and its application to false alarm reduction *IEEE Trans. Biomed. Eng.* **60** 1660-6

Clifford G D, Silva I, Moody B, Li Q, Kella D, Shahin A, Kooistra T, Perry D and Mark R G 2015 Reducing false arrhythmia alarms in the ICU: The physionet/computing in cardiology challenge *Comput. Cardiol.* **42** 273-6

Couto P, Ramalho R and Rodrigues R 2015 Suppression of False Arrhythmia Alarms Using ECG and Pulsatile Waveforms *Comput. Cardiol.* **42** 749-52

Deshmane A V 2009 False arrhythmia alarm suppression using ECG, ABP, and photoplethysmogram *PhD Thesis* Massachusetts Institute of Technology

Donchin Y and Seagull F J 2002 The hostile environment of the intensive care unit *Curr. Opin. Crit. Care* **8** 316-20

Dzwonczyk R, Brown C G and Werman H A 1990 The median frequency of the ECG during ventricular fibrillation: its use in an algorithm for estimating the duration of cardiac arrest *IEEE Trans. Biomed. Eng.* **37** 640-6

Fallet S, Yazdani S and Vesin J M 2015 A Multimodal Approach to Reduce False Arrhythmia Alarms in the Intensive Care Unit *Comput. Cardiol.* **42** 277-80

Goldberg D E and Holland J H 1988 Genetic algorithms and machine learning *Mach. Learn.* **3** 95-9

Hamilton P S and Tompkins W J 1986 Quantitative investigation of qrs detection rules using the mit/bih arrhythmia database *IEEE Trans. Biomed. Eng.* **33** 1157-65

Johnson A E, Behar J, Andreotti F, Clifford G D and Oster J 2015 Multimodal heart beat detection using signal quality indices *Physiol. Meas.* **36** 1665-77

Kalidas V and Tamil L S 2015 Enhancing accuracy of arrhythmia classification by combining logical and machine learning techniques *Comput. Cardiol.* **42** 733-6

Krasteva V, Jekova I, Leber R, Schmid R and Abacherli R 2015 Validation of arrhythmia detection library on bedside monitor data for triggering alarms in intensive care *Comput. Cardiol.* **42** 737-40

Kuo S and Dillman R 1978 Computer detection of ventricular fibrillation *Comput. Cardiol.* 347-9

Lawless S T 1994 Crying wolf: false alarms in a pediatric intensive care unit *Crit. Care. Med.* **22** 981-5

Li H, Han W, Hu C and Meng M Q 2009 Detecting ventricular fibrillation by fast algorithm of dynamic sample entropy. *IEEE Int. Conf. on Robotics and Biomimetics* pp 1105-10

Li Q and Clifford G D 2012 Signal quality and data fusion for false alarm reduction in the intensive care unit *J. Electrocardiol.* **45** 596-603

Li Q, Mark R G and Clifford G D 2008 Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a kalman filter *Physiol. Meas.* **29** 15

Li Q, Rajagopalan C, Clifford G D 2014 Ventricular Fibrillation and Tachycardia Classification Using a Machine Learning Approach IEEE Trans. Biomed. Eng. **61** 1607-13

Pan J and Tompkins W J 1985 A real-time QRS detection algorithm *IEEE Trans. Biomed. Eng.* **32** 230-6

Plesinger F, Klimes P, Halamek J and Jurak P 2015 False alarms in intensive care unit monitors: Detection of life-threatening arrhythmias using elementary algebra, descriptive statistics and fuzzy logic *Computing in Cardiology Conf.* pp 1-4

Sabeti E, Shamsollahi M B and Afdideh F 2012 Prediction of Paroxysmal Atrial Fibrillation using Empirical Mode Decomposition and RR intervals *IEEE EMBS Conf. on Biomedical Engineering and Sciences* pp 750-4

Salas B R, Bai Y, Harris P, Drew B, Hu X 2014 False ventricular tachycardia alarm suppression in the ICU based on the discrete wavelet transform in the ECG signal *J. Electrocardiol.* **47** 775-80

Sayadi O and Shamsollahi M B 2011 Life-threatening arrhythmia verification in ICU patients using the joint cardiovascular dynamical model and a Bayesian filter *IEEE Trans.Biomed. Eng.* **58** 2748-57

Schmid F, Goepfert M S and Reuter D A 2013 Patient monitoring alarms in the ICU and in the operating room *Crit. Care* **17** 216-22

Sun J X, Reisner A T and Mark R G 2006 A signal abnormality index for arterial blood pressure waveforms. *Computers in Cardiology* vol 33 pp 13-6

Thakor N V, Zhu Y S and Pan K Y 1990 Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm *IEEE Trans. Biomed. Eng.* **37** 837-43

Vapnik V 2013 *The Nature of Statistical Learning Theory* Springer

Zong W, Heldt T, Moody G and Mark R 2003 An open-source algorithm to detect onset of arterial blood pressure pulses *Proc. Computers in Cardiology (21-24 September 2003)* pp 259-62