

Chapter 27

Signal Processing: False Alarm Reduction

Qiao Li and Gari D. Clifford

Learning Objectives

Use a data fusion and machine learning approach to suppress false arrhythmia alarms.

This case study introduces concepts that should improve understanding of the following:

1. Extract relevant features from clinical waveforms.
2. Assess signal quality of clinical data, and
3. Develop a machine learning model, train and validate it using a clinical database.

27.1 Introduction

Modern patient monitoring systems in intensive care produce frequent false alarms which lead to a disruption of care, impacting both the patient and the clinical staff through noise disturbances, desensitization to warnings and slowing of response times [1, 2]. This leads to decreased quality of care [3, 4], sleep deprivation [1, 5, 6], disrupted sleep structure [7, 8], stress for both patients and staff [9–12] and depressed immune systems [13]. Intensive care unit (ICU) false alarm rates as high as 90 % have been reported [14], while only 8 % of alarms were determined to be true alarms with clinical significance [15] and over 94 % of alarms may not be clinically important [16]. There are two main reasons for the high false alarm rate. One is that physiological data can be severely corrupted by artifacts (e.g. from movement), noise (e.g. from electrical interference) and missing data (e.g. from transducer ‘pop’ leading to impedance or pressure changes and a resultant signal saturation). Figure 27.1 illustrates the bedside monitor ‘waveforms’ (or high resolution data)

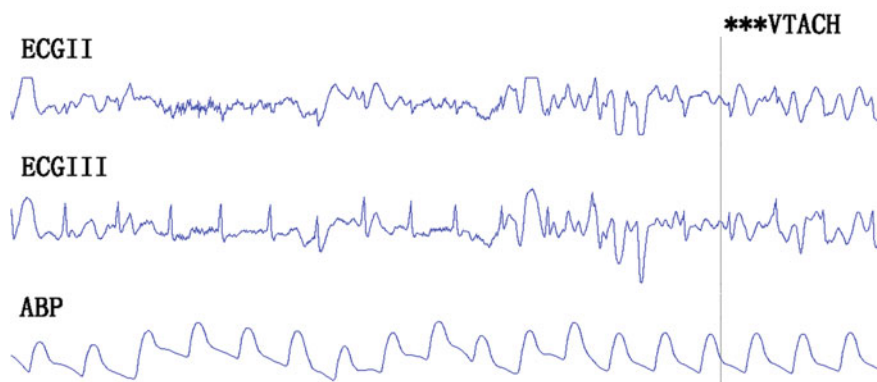


Fig. 27.1 False ventricular tachycardia alarm, ‘called’ at the point where the vertical line is placed in a 30 s snapshot of two leads of ECG (ECGII and ECGIII) and an arterial blood pressure signal (ABP). The alarm is triggered by the strong noise manifesting as high amplitude (± 2 mV) oscillations on the ECG at approximately 5 Hz beginning a little over halfway through the snapshot (and a little under 10 s from the vertical VT marker). Note that the ABP continues as normal, with no significant change in rhythm or morphology

recorded around a false ventricular tachycardia alarm (the vertical line indicates the moment at which the monitor triggered the alarm). The alarm is caused by significant noise affecting the electrocardiogram (ECG) leads. However, the regular pulsatile beats present in the arterial blood pressure (ABP) lead clearly indicate this is a false alarm (since the poor pump function during this arrhythmia should cause a significant drop in pulse amplitude and an increase in rate). The other reason for the high rate of false alarms is that univariate alarm algorithms and simple numeric thresholds are predominantly used in current clinical bedside monitors. The reason for this is an historical artifact, in that manufacturers have developed different embedded systems with bespoke hardware and single mode transducers. Univariate alarm-detection algorithms therefore consider a single monitored waveform at a time. The alarm is generally triggered when a variable (e.g. heart rate) derived from the waveform (e.g. ECG) is above or below a preset (or adjustable) threshold for a given length of time, regardless of whether the change is caused by a change in physiological state, by an artifact or by medical interventions, such as moving or positioning the patient, drawing blood and flushing the arterial line, or disconnecting the patient from the ventilator for endotracheal suctioning. Moreover, alarm thresholds are often adjusted in an ad hoc manner, based on how annoying the alarm is perceived to be by the clinical team in attendance. There is little evidence that alarm thresholds are optimized for any population or individual, particularly in a multivariate sense.

Various noise cancellation algorithms such as median filtering [17] or Kalman filtering [18] have been used to suppress false alarms. While transient noise can be removed by median filtering it is brutally non-adaptive. Kalman filtering, on the other hand, is an optimal state estimation method, which has been used to improve heart rate (HR) and blood pressure (BP) estimation during noisy periods and

arrhythmias [18]. However, alarm detection has changed little in decades, with the univariate alarm algorithm paradigm persisting. A promising solution to the false alarm issue comes from multiple variable data fusion, such as HR estimation by fusing the information from synchronous ECG, ABP and photoplethysmogram (PPG) from which oxygen saturation is derived [18]. Otero et al. [19] proposed a multivariable fuzzy temporal profile model which described a set of monitoring criteria of temporal evolution of the patient's physiological variables of HR, oxygen saturation (SpO_2) and BP. Aboukhalil et al. [14] and Deshmane [20] used synchronous ABP and PPG signals to suppress false ECG alarms. Zong et al. [21] reduced false ABP alarms using the relationships between ECG and ABP. Besides calculated physiological parameters, signal quality indices (SQI), which assess the waveform's usefulness or the noise levels of the waveforms, can be extracted from the raw data and used as weighting factors to allow for varying trust levels in the derived parameters. Behar et al. [22] and Li and Clifford [23] suppressed false ECG alarms by assessing the signal quality of ECG, ABP and PPG. Monasterio et al. [24] used a support vector machine to fuse data from respiratory signals, heart rate and oxygen saturation derived from the ECG, PPG, and impedance pneumogram, as well as several SQIs, to reduce false apnoea-related desaturations.

27.2 Study Dataset

A dataset drawn from PhysioNet's MIMIC II database [25, 26] was used in this study, containing simultaneous ECG, ABP, and PPG recordings with 4107 multiple expert-annotated life-threatening arrhythmia alarms [asystole (AS), extreme bradycardia (EB), extreme tachycardia (ET) and ventricular tachycardia (VT)] on 182 ICU admissions. A total of 2301 alarms were found by selecting the alarms when the ECG, ABP and PPG were all available. The false alarm rates were 91.2 % for AS, 26.6 % for EB, 14.4 % for ET, and 44.4 % for VT respectively, and 45.0 % overall. The ICU admissions were divided into two separate sets for training and testing, ensuring that the frequency of alarms in each category was roughly equal through frequency ranking and separating odd and evenly numbered signals. Table 27.1 details the relative frequency of each alarm category and their associated true and false alarm rates. The waveform data from 30 s before to 10 s after the alarm were extracted for each alarm to aid expert verification (since the Association for the Advancement of Medical Instrumentation (AAMI) guidelines require an alarm to respond within 10 s of the initiation of any alarm event [27]). A consensus of three experts was required to label each alarm as true or false. Only data from 10 s before the alarm to the alarm onset were used for automated feature extraction and model classification.

Since the VT alarm was considered the most difficult type of false alarm to suppress, with an associated low false alarm reduction rate and high true alarm suppression rate in literature [14, 20–23, 28], we therefore focus on reducing this

Table 27.1 Distribution of alarms in the dataset and training and test set

Alarm type	Total				Training set				Test set			
	False	True	Total	FA rate (%)	False	True	Total	FA rate (%)	False	True	Total	FA rate (%)
AS	260	25	285	91.2	166	14	180	92.2	94	11	105	89.5
EB	62	171	233	26.6	58	108	166	34.9	4	63	67	6.0
ET	37	220	257	14.4	19	116	135	14.1	18	104	122	14.8
VT	677	849	1526	44.4	306	478	784	39.0	371	371	742	50.0
All	1036	1265	2301	45.0	549	716	1265	43.4	487	549	1036	47.0

false alarm for the rest of the chapter. Interested readers are directed to Li and Clifford [23] for methods to reduce false alarms on the other types of alarms.

27.3 Study Pre-processing

In total 147 features and SQI metrics were extracted from ECG, ABP, PPG, and SpO₂ signals within the 10 s analysis window. These features were generally chosen based upon previous research by the authors and others [14, 20–24, 28–32]. The typical features included HR (extracted from ECG, ABP, and PPG), blood pressure (systolic, diastolic, mean), oxygen saturation (SpO₂), and the amplitude of PPG. Each feature had five sub-features calculated over the 10 s window: including the minimum, maximum, median, variance, and gradient (derived from a robust least squares fit over the entire window). Besides the typical features, the area difference of beats (ADB), the area ratio of beats (ARB) in the ECG, ABP and PPG and thirteen ventricular fibrillation metrics (taken from [29]) were also extracted. The area of each beat was defined to be the area between the waveform and the x-axis, from the start of the ECG beat to 0.6 times of mean beat-by-beat interval (BBi). Note the start of the ECG beat was taken as the position of R peak—0.2 * BBi. The ADB was calculated by comparing each beat to the median of the beats in the window, as shown in Fig. 27.2. The ADB used four sub-features; the mean ADB of five beats with the shortest beat-to-beat intervals, the maximum of mean ADB of five consecutive beats, the variance and gradient of ADB. The ARB used five sub-features; the ratio between the mean area of five smallest beats and five largest beats of the ECG (ARB_{ECG}), ABP (ARB_{ABP}), and PPG (ARB_{PPG}), the ratio between ARB_{ECG} and ARB_{ABP}, and the ratio between ARB_{ECG} and ARB_{PPG}. The description of the thirteen ventricular fibrillation metrics can be found in Li et al. [29], and included spectral and time domain features shown to allow highly accurate classification of VF. The ECG SQI metrics included thirteen metrics [30], based on standard moments, frequency domain statistics and the agreement between event detectors with different noise sensitivities. The ABP SQI metrics included a signal abnormality index with its nine sub-metrics [31] and a dynamic time warping

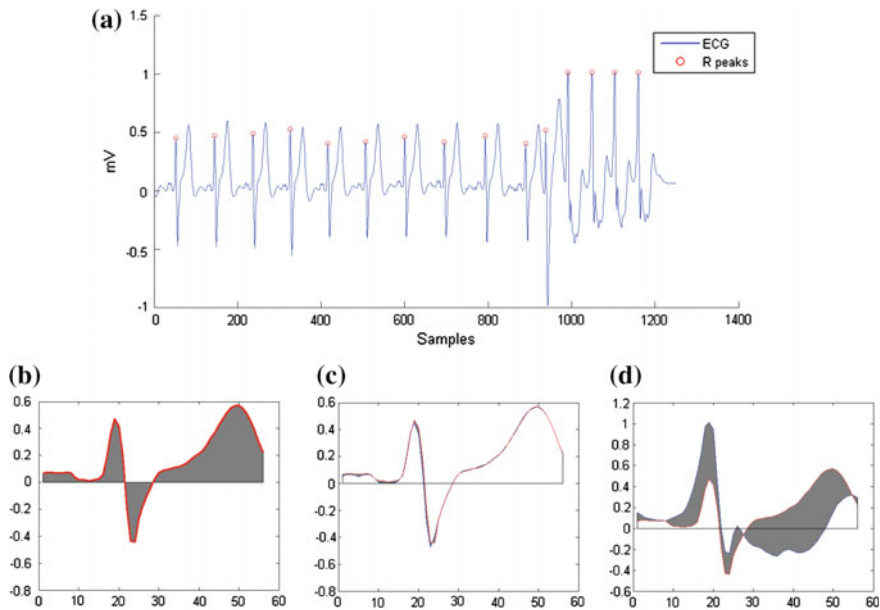


Fig. 27.2 Example of area difference of beats calculation. **a** ECG in a 10 s window. **b** The median beat of the beats in the window (*gray* area shows the area between the waveform and the x-axis). **c** ADB of a normal beat (the first beat, *gray* area shows the ADB). **d** ADB of an abnormal beat (the last beat)

(DTW) based SQI approach with its four sub-metrics [32]. The DTW based SQI resampled each beat to match a running beat template by derived using the DTW. The SQI was then given by the correlation coefficient between the template and each beat. The PPG SQI metrics included the DTW-based SQIs [32] and the first two Hjorth parameters [20] which estimated the dominant frequency and half-bandwidth of the spectral distribution of PPG. While these do not necessarily represent an exhaustive list of features, they do represent the vast majority of features identified as useful in previous studies.

27.4 Study Methods

A modified random forests (RF) classifier, previously described by Johnson et al. [33], was used. The RF [34] is an ensemble learning method for classification that constructs a number of decision trees at training time and outputs the class that is the mode of the classes of the individual trees. The basic principle is that a group of “weak learners” can come together to form a “strong learner.” RFs correct for decision trees’ defects of overfitting and adding bias to their training set. Each tree selects a subset of observations via two regression splits. These observations are

then given a contribution equal to a random constant times the observation's value for a chosen feature plus a random intercept. The contributions across all trees are summed to provide the contribution for a single "forest," where a "forest" refers to a group of trees plus an intercept term. The predicted likelihood function output (L) by the forest is the inverse logit of the sum of each tree's contribution plus the intercept term (27.1). The intercept term is set to the logit of the mean observed outcome.

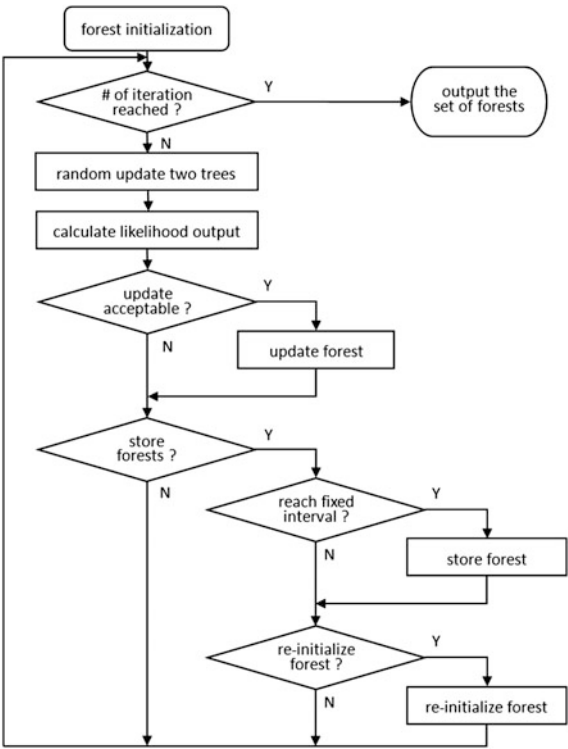
$$L = \sum_{i=1}^N ((-t_i) * \log(\text{logit}^{-1}(s_i)) - (1 - t_i) * \log(1 - \text{logit}^{-1}(s_i))) \quad (27.1)$$

where t_i is the target of the training set, s_i is the sum of tree's contribution, $i = 1 \dots N$ is the number of observations in the training set.

The core of the new RF model we used is the custom Markov chain Monte Carlo (MCMC) sampler that iteratively optimizes the forest. This sampling process constructs the Markov chain by a memoryless iteration process which selects randomly two trees from the current forests and updates their structure. The MCMC randomly samples the observation space by a large user-defined number of bootstrap iterations. After standardizing the training data to a standard normal distribution, the forest is initialized to a null model, with no contributions assigned for any observations.

At each iteration, the algorithm randomly selects two trees in the forest and randomizes their structure. That is, it randomly re-selects first two features which the tree uses for splitting, the value at which the tree splits those features, the third feature used for contribution calculation, and the multiplicative and additive constants applied to the third feature. The total forest contribution is then recalculated and a Metropolis-Hastings acceptance step is used to determine if the update is accepted. The predicted likelihood of the previous forest (L_i) and the likelihood of the forest with the two updated trees (L_{i+1}) were calculated. If $e^{(L_i - L_{i+1})}$ is greater than a uniformly distributed random real number within unit interval, the update is accepted. If the update is accepted, the two trees are kept in the forest, otherwise they are discarded and the forest remains unchanged. After a set fraction of the total number of iterations to allow the forest to learn the target distribution (generally 20 %), the algorithm begins storing forests at a fixed interval, i.e. once every set number of iterations. Once the number of user-defined iterations is reached, the forest is re-initialized as before, and the iterative process restarts. Again, after the set burn-in period, the forests begin to be saved at a fixed interval. The final result of this algorithm is a set of forests, each of which will contribute to the final model classification. The flowchart of the RF algorithm is shown in Fig. 27.3.

Fig. 27.3 The flowchart of the random forests algorithm



27.5 Study Analysis

The RF model was optimized on the training set and evaluated for out-of-sample accuracy on the test set. During the training phase, a model of 320 forests with 500 trees in each forest was established. The output of the model provides a probability between 0 and 1, which is an estimated value equivalent to a false or true alarm respectively. The receiver operating characteristic (ROC) curve was extracted by raising the threshold on the probability where we switch from false to true from 0 to 1—i.e. the probability greater than the threshold indicates a true alarm and below (or equal) indicates a false alarm. The optimal operating point was selected at the ROC curve when sensitivity equals 1 (no true alarm suppression) with the largest specificity. However, a sub-optimal operating point was also selected with acceptable sensitivity to balance specificity, e.g. sensitivity equals 99 %. (The reason for this is that anecdotally, clinical experts have indicated a 1 % true alarm suppression rate (or increase in true alarm suppression rate) would be acceptable—see discussion in study conclusions.) The model was then evaluated on the test set with the selected operating points.

In the algorithm validation phase, the classification performance of the algorithm was evaluated using 10-fold cross validation. The process sorted the study dataset into ten folds randomly stratified by ICU admissions rather than by the alarms. Then, nine folds were used for training the model and the last fold was used for validation. This process was repeated ten times as one integral procedure, with each of the folds used exactly once as the validation data. The average performance was used for evaluation. We note however, that this may be suboptimal and a voting of all folds may produce a better performance.

27.6 Study Visualizations

The ROC curve on the training set is shown in Fig. 27.4. The optimal operating point (marked by a circle) shows sensitivity 100.0 % and specificity 24.5 %, indicating we suppress 24.5 % of the false alarms without true alarm suppression. The sub-optimal operating point (marked by a star) shows a sensitivity 99.2 % and specificity 53.3 %, indicating a false alarm reduction of 53.3 % with only a 0.8 % true alarm suppression rate. When the model was used on the test set by the optimal

Fig. 27.4 ROC curve for the training set. *Circle* indicates optimal operating point (in terms of clinical acceptability) and *star* a sub-optimal operating point which may in fact be preferable

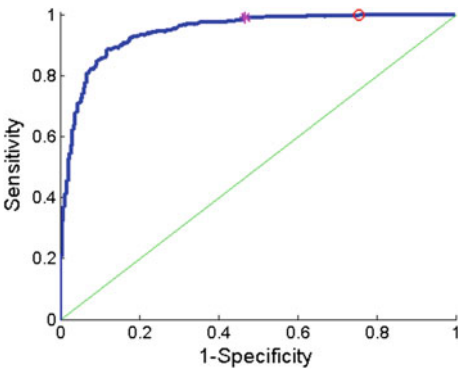


Table 27.2 Result of 10-fold cross validation of the classification model with different operating points

Operating point (by sensitivity) (%)	Training (on 9 folds)		Validation (on 1 held out fold)	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
99.00	99.06 ± 0.04	56.41 ± 5.60	95.82 ± 5.62	51.68 ± 16.88
99.50	99.56 ± 0.04	49.08 ± 5.37	96.50 ± 5.39	45.19 ± 17.94
99.60	99.66 ± 0.04	43.49 ± 6.45	98.72 ± 2.06	38.14 ± 17.25
99.70	99.75 ± 0.03	39.50 ± 7.39	98.75 ± 2.08	32.07 ± 16.19
99.80	99.87 ± 0.02	34.57 ± 9.02	98.87 ± 2.11	28.16 ± 15.80
100.0	100.0 ± 0.00	27.85 ± 6.17	99.04 ± 2.02	18.10 ± 9.87

operating point, a sensitivity of 99.7 % and a specificity of 17.0 % were achieved, with a sensitivity of 99.5 % and a specificity of 44.2 % for the sub-optimal operating point. The result of 10-fold cross validation with different options of operating points is shown in Table 27.2.

27.7 Study Conclusions

We show here that a promising approach to suppression of false alarms appears to be through the use of multivariate algorithms, which fuse synchronous data sources and estimates of underlying quality to make a decision. False VT alarms are the most difficult to suppress without causing any true alarm suppression since the ABP and PPG waveforms may have morphology changes indicating the hemodynamics changes during VT. We also show that a random forests-based model can be implemented with high confidence that few true alarms would be suppressed (although it's impossible to say 'never'). A practical operating point can be selected by changing the threshold of the model in order to balance the sensitivity and specificity. We note that the best previously reported results on VT alarms were by Aboukhalil et al. [14] and Sayadi and Shamsollahi [28] who achieved false VT alarm suppression rates of 33.0 and 66.7 % respectively. However, the TA suppression rates they achieved (9.4 and 3.8 % respectively) are clearly too high to make their algorithms acceptable for this category of alarm. Compared with our previous studies using some common machine learning algorithms such as support vector machine [22] and relevance vector machine [23], the random forests algorithm, which fused the features extracted from synchronous data sources like ECG, ABP and PPG, provided lower TA suppression rates and higher FA suppression rates. Moreover, a systematic validation procedure, such as k-fold cross validation, is necessary to evaluate the algorithm and we note that earlier works did not follow such a protocol. Without such validation, it is hard to believe that the algorithm will work well on unseen data because of overfitting. This is extremely important to note, that even a 0 % true alarm suppression is unlikely to always hold, and so a small true alarm suppression is likely to be acceptable. In private discussions with our clinical advisors, a figure of 1 % has often been suggested. In the work presented here, we show that with just half a percent of true alarms being suppressed, almost half of the false alarms can be suppressed. This true alarm suppression rate is likely to be negligible compared to the actual number of noise-induced missed alarms from the bedside monitor itself. (No monitor is perfect, and false negative rates of between 0.5 and 5 % have been reported [35].) We also note that the algorithm proposed here used 10 s of data before the alarm only, which meets the 10 s requirement of AAMI standard [27]. In recent work from the PhysioNet/Computing in Cardiology Challenge 2015, it was shown that extending this window slightly can lead to significant improvements in false alarm suppression [36]. Although the regulatory bodies would need to approve such changes, and that is often seen as unlikely, we do note that the 10 s rule is somewhat arbitrary

and such work may indeed influence the changes in regulatory acceptance. We note several limitations to our study. First, the number of alarms is still relatively low, and they come from a single database/manufacture. Second, medical history, demographics, and other medical data were not available and therefore used to adjust thresholds. Finally, information concerning repeated alarms was not used to adjust false alarm suppression dynamically based on earlier alarm frequency during the same ICU stay. This latter point is particularly tricky, since using earlier alarm data as prior information can be entirely misleading when false alarm rates are non-negligible.

27.8 Next Steps/Potential Follow-Up Studies

The issue of false alarms has disturbed the clinical patient monitoring and monitor manufacturers for many years, but the alarm handling has not seen the same progress as the rest of medical monitoring technology. One important reason is that in the current legal and regulatory environment, it may be argued that manufacturers have external pressures to provide the most sensitive alarm algorithms, such that no critical event goes undetected [4]. Equally, one could argue that clinicians also have an imperative to ensure that no critical alarm goes undetected, and are willing to accept large numbers of false alarms to avoid a single missed event. A large number of algorithms and methods have emerged in this area [4, 14, 17–24, 28, 37, 38]. However, most of these approaches are still in an experimental stage and there is still a long way to go before the algorithms are ready for clinical application.

The 2015 PhysioNet/Computing in Cardiology Challenge aimed to encourage the development of algorithms to reduce the incidence of false alarms in ICU [36]. Bedside monitor data leading up to a total of 1250 life-threatening arrhythmia alarms recorded from three of the most prevalent intensive care monitor manufacturers' bedside units were used in this challenge. Such challenges are likely to stimulate renewed interest by the monitoring industry in the false alarm problem. Moreover, the engagement of the scientific community will draw out other subtle issues. Perhaps the three key issues remaining to be addressed are: (1) Just how many alarms should be annotated and by how many experts? (see Zhu et al. [39] for a detailed discussion of this point); (2) How should we deal with repeated alarms, passing information forward from one alarm to the next?; and (3) What additional data should be supplied to the bedside monitor as prior information on the alarm? This could include a history of tachycardia, hypertension, drug dosing, interventions and other related information including acuity scores. Finally, we note that life threatening alarms are far less frequent than other less critical alarms, and by far the largest contributor to the alarm pollution in critical care comes from these more pedestrian alarms. A systematic approach to these less urgent alarms is also needed, borrowing from the framework presented here. More promisingly, the tolerance of true alarm suppression is likely to be much higher for less important alarms, and so we expect to see very large false alarm suppression rates. This is particularly

important, since the techniques described here are general and could apply to most non-critical false alarms, which constitute the majority of such events in the ICU. Although the competition does not directly address these four points (and in fact the data needed to do so remains to become available in large numbers), the competition will provide a stimulus for such discussions and the tools (data and code) will help continue the evolution of the field.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, a link is provided to the Creative Commons license and any changes made are indicated.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

1. Chambrin MC (2001) Review: alarms in the intensive care unit: how can the number of false alarms be reduced? *Crit Care* 5(4):184–188
2. Cvach M (2012) Monitor alarm fatigue, an integrative review. *Biomed Inst Tech* 46(4):268–277
3. Donchin Y, Seagull FJ (2002) The hostile environment of the intensive care unit. *Curr Opin Crit Care* 8(4):316–320
4. Imhoff M, Kuhls S (2006) Alarm algorithms in critical care monitoring. *Anesth Analg* 102(5):1525–1537
5. Meyer TJ, Eveloff SE, Bauer MS, Schwartz WA, Hill NS, Millman RP (1994) Adverse environmental conditions in the respiratory and medical ICU settings. *Chest* 105(4):1211–1216
6. Parthasarathy S, Tobin MJ (2004) Sleep in the intensive care unit. *Intensive Care Med* 30(2):197–206
7. Johnson AN (2001) Neonatal response to control of noise inside the incubator. *Pediatr Nurs* 27(6):600–605
8. Slevin M, Farrington N, Duffy G, Daly L, Murphy JF (2000) Altering the NICU and measuring infants' responses. *Acta Paediatr* 89(5):577–581
9. Cropp AJ, Woods LA, Raney D, Bredle DL (1994) Name that tone. The proliferation of alarms in the intensive care unit. *Chest* 105(4):1217–1220
10. Novaes MA, Aronovich A, Ferraz MB, Knobel E (1997) Stressors in ICU: patients' evaluation. *Intensive Care Med* 23(12):1282–1285
11. Topf M, Thompson S (2001) Interactive relationships between hospital patients' noise induced stress and other stress with sleep. *Heart Lung* 30(4):237–243
12. Morrison WE, Haas EC, Shaffner DH, Garrett ES, Fackler JC (2003) Noise, stress, and annoyance in a pediatric intensive care unit. *Crit Care Med* 31(1):113–119
13. Berg S (2001) Impact of reduced reverberation time on sound-induced arousals during sleep. *Sleep* 24(3):289–292

14. Aboukhalil A, Nielsen L, Saeed M, Mark RG, Clifford GD (2008) Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. *J Biomed Inform* 41 (3):442–451
15. Tsien CL, Fackler JC (1997) Poor prognosis for existing monitors in the intensive care unit. *Crit Care Med* 25(4):614–619
16. Lawless ST (1994) Crying wolf: false alarms in a pediatric intensive care unit. *Crit Care Med* 22(6):981–985
17. Mäkitvirta A, Koski E, Kari A, Sukuvaara T (1991) The median filter as a preprocessor for a patient monitor limit alarm system in intensive care. *Comput Meth Prog Biomed* 34(2–3):139–144
18. Li Q, Mark RG, Clifford GD (2008) Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiol Meas* 29(1):15–32
19. Otero A, Felix P, Barro S, Palacios F (2009) Addressing the flaws of current critical alarms: a fuzzy constraint satisfaction approach. *Artif Intell Med* 47(3):219–238
20. Deshmane AV (2009) False arrhythmia alarm suppression using ECG, ABP, and photoplethysmogram. M.S. thesis, MIT, USA
21. Zong W, Moody GB, Mark RG (2004) Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure. *Med Biol Eng Comput* 42(5):698–706
22. Behar J, Oster J, Li Q, Clifford GD (2013) ECG signal quality during arrhythmia and its application to false alarm reduction. *IEEE Trans Biomed Eng* 60(6):1660–1666
23. Li Q, Clifford GD (2012) Signal quality and data fusion for false alarm reduction in the intensive care unit. *J Electrocardiol* 45(6):596–603
24. Monasterio V, Burgess F, Clifford GD (2012) Robust classification of neonatal apnoea-related desaturations. *Physiol Meas* 33(9):1503–1516
25. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE (2000) Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* 101(23):e215–e220
26. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L, Moody G, Heldt T, Kyaw TH, Moody B, Mark RG (2011) Multiparameter intelligent monitoring in intensive care II (MIMIC-II): a public-access intensive care unit database. *Crit Care Med* 39(5):952–960
27. American National Standard (ANSI/AAMI EC13:2002) (2002) Cardiac monitors, heart rate meters, and alarms. Association for the Advancement of Medical Instrumentation, Arlington, VA
28. Sayadi O, Shamsollahi M (2011) Life-threatening arrhythmia verification in ICU patients using the joint cardiovascular dynamical model and a Bayesian filter. *IEEE Trans Biomed Eng* 58(10):2748–2757
29. Li Q, Rajagopalan C, Clifford GD (2014) Ventricular fibrillation and tachycardia classification using a machine learning approach. *IEEE Trans Biomed Eng* 61(6):1607–1613
30. Li Q, Rajagopalan C, Clifford GD (2014) A machine learning approach to multi-level ECG signal quality classification. *Comput Meth Prog Biomed* 117(3):435–447
31. Sun JX, Reisner AT, Mark RG (2006) A signal abnormality index for arterial blood pressure waveforms. *Comput Cardiol* 33:13–16
32. Li Q, Clifford GD (2012) Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiol Meas* 33(9):1491–1501
33. Johnson AEW, Dunkley N, Mayaud L, Tsanas A, Kramer AA, Clifford GD (2012) Patient specific predictions in the intensive care unit using a Bayesian ensemble. *Comput Cardiol* 39:249–252
34. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
35. Schapira RM, Van Ruiswyk J (2002) Reduction in alarm frequency with a fusion algorithm for processing monitor signals. Meeting of the American Thoracic Society. Session A56, Poster H57

36. Clifford GD, Silva I, Moody B, Li Q, Kella D, Shahin A, Kooistra T, Perry D, Mark RG (2006) The PhysioNet/computing in cardiology challenge 2015: reducing false arrhythmia alarms in the ICU. *Comput Cardiol* 42:1–4
37. Borowski M, Siebig S, Wrede C, Imhoff M (2011) Reducing false alarms of intensive care online-monitoring systems: an evaluation of two signal extraction algorithms. *Comput Meth Prog Biomed* 2011:143480
38. Li Q, Mark RG, Clifford GD (2009) Artificial arterial blood pressure artifact models and an evaluation of a robust blood pressure and heart rate estimator. *Biomed Eng Online* 8:13
39. Zhu T, Johnson AEW, Behar J, Clifford GD (2014) Crowd-sourced annotation of ECG signals using contextual information. *Ann Biomed Eng* 42(4):871–884