

ORIGINAL WORK



Automated False Alarm Reduction in a Real-Life Intensive Care Setting Using Motion Detection

Carl Muroi¹ , Sando Meier², Valeria De Luca², David J. Mack¹, Christian Strässle¹, Patrick Schwab³, Walter Karlen³ and Emanuela Keller^{1*}

© 2019 Springer Science+Business Media, LLC, part of Springer Nature and Neurocritical Care Society

Abstract

Background: Contemporary monitoring systems are sensitive to motion artifacts and cause an excess of false alarms. This results in alarm fatigue and hazardous alarm desensitization. To reduce the number of false alarms, we developed and validated a novel algorithm to classify alarms, based on automatic motion detection in videos.

Methods: We considered alarms generated by the following continuously measured parameters: arterial oxygen saturation, systolic blood pressure, mean blood pressure, heart rate, and mean intracranial pressure. The movements of the patient and in his/her surroundings were monitored by a camera situated at the ceiling. Using the algorithm, alarms were classified into RED (true), ORANGE (possibly false), and GREEN alarms (false, i.e., artifact). Alarms were reclassified by blinded clinicians. The performance was evaluated using confusion matrices.

Results: A total of 2349 alarms from 45 patients were reclassified. For RED alarms, sensitivity was high (87.0%) and specificity was low (29.6%) for all parameters. As the sensitivities and specificities for RED and GREEN alarms are interrelated, the opposite was observed for GREEN alarms, i.e., low sensitivity (30.2%) and high specificity (87.2%). As RED alarms should not be missed, even at the expense of false positives, the performance was acceptable. The low sensitivity for GREEN alarms is acceptable, as it is not harmful to tag a GREEN alarm as RED/ORANGE. It still contributes to alarm reduction. However, a 12.8% false-positive rate for GREEN alarms is critical.

Conclusions: The proposed system is a step forward toward alarm reduction; however, implementation of additional layers, such as signal curve analysis, multiple parameter correlation analysis and/or more sophisticated video-based analytics are needed for improvement.

Keywords: Smart alarms, Motion sensor, Alarm reduction, ICU, Alarm fatigue, False alarms

Introduction

Patients in an intensive care unit (ICU) are subject to extensive monitoring, in order to detect critical changes in vital parameters as early as possible. In daily clinical practice, monitoring systems collect signals from biosensors and trigger audio visual alarms when values go below

or beyond predefined thresholds [1]. It has been reported that for each patient alarms are triggered for almost 11 out of 24 h [2]. A previous study performed at the Neurocritical Care Unit, University Hospital Zurich, analyzing 13,938 alarms, yielded an average rate of 696.7 alarms per patient per day [3]. Contemporary monitoring systems are extremely sensitive to motion artifacts and technical errors [1, 4]. It has been reported that up to 90% of the alarms can be false alarms [5, 6]. Therefore, alarm fatigue and alarm desensitization pose a medical technology hazard [7]. Critical alarms might be missed or ignored

*Correspondence: emanuela.keller@usz.ch

¹ Neurocritical Care Unit, Department of Neurosurgery, University Hospital Zurich, 8091 Zurich, Switzerland

Full list of author information is available at the end of the article

due to the alarm overload, thereby putting patients' well-being at risk [8]. Another noteworthy aspect is the contribution of monitoring alarms to the general noise level, which can reach up to 80 dB in an ICU [9]. The exposure to noise not only affects medical staff, but also disturbs patients' sleep [10]. In order to avoid alarm desensitization and unnecessary noise, the reduction in false alarms would be beneficial for both medical staff and patients. Some efforts in this field include a simple introduction of an alarm delay [4], the use of fuzzy set classifiers [11, 12], and machine-learning approaches [1, 3] among others. Although many false alarms are motion-induced [4], no one has leveraged this information yet. We present a computer-vision approach on video recording for motion detection. The hypothesis was that a motion detection-based automated classification reduces the number of false alarms, as motion induces artifacts triggering false alarms in a greater extent.

Materials and Methods

The study was conducted between November 2016 and December 2017 at the Neurocritical Care Unit, University Hospital Zurich, Switzerland. The findings from the current motion detection-based algorithm were not shared with the clinicians treating the patients. Therefore, it had neither impact on the patient monitoring and treatment nor on the clinical decision making. The observational study, part of the project "ICU Cockpit," was approved by the ethics committee of Kanton Zurich, Switzerland.

Patient Selection

Any patient admitted to the unit was eligible for inclusion. Incoming patients were consequently assigned to the two monitored beds depending on their availability. Therefore, a random selection of patients could be assumed representing a cohort of the current ICU population. Once assigned, written consent was obtained from patients or legal representatives. Two patients were excluded from the study, as the consent could not be obtained.

Vital Parameter Collection

Commonly measured vital signs were streamed to the CNS 220 patient monitor (Moberg Research Inc., Ambler, PA, USA) and forwarded to the "ICU Cockpit," a platform for biosignal consolidation and processing (Fig. 1a). Vital patient parameters and triggered alarms were continuously recorded at 1 Hz. For the current study, the following vital parameters were chosen for analysis: peripheral oxygen saturation (SpO_2) measured by a photoplethysmography finger-clip, systolic arterial

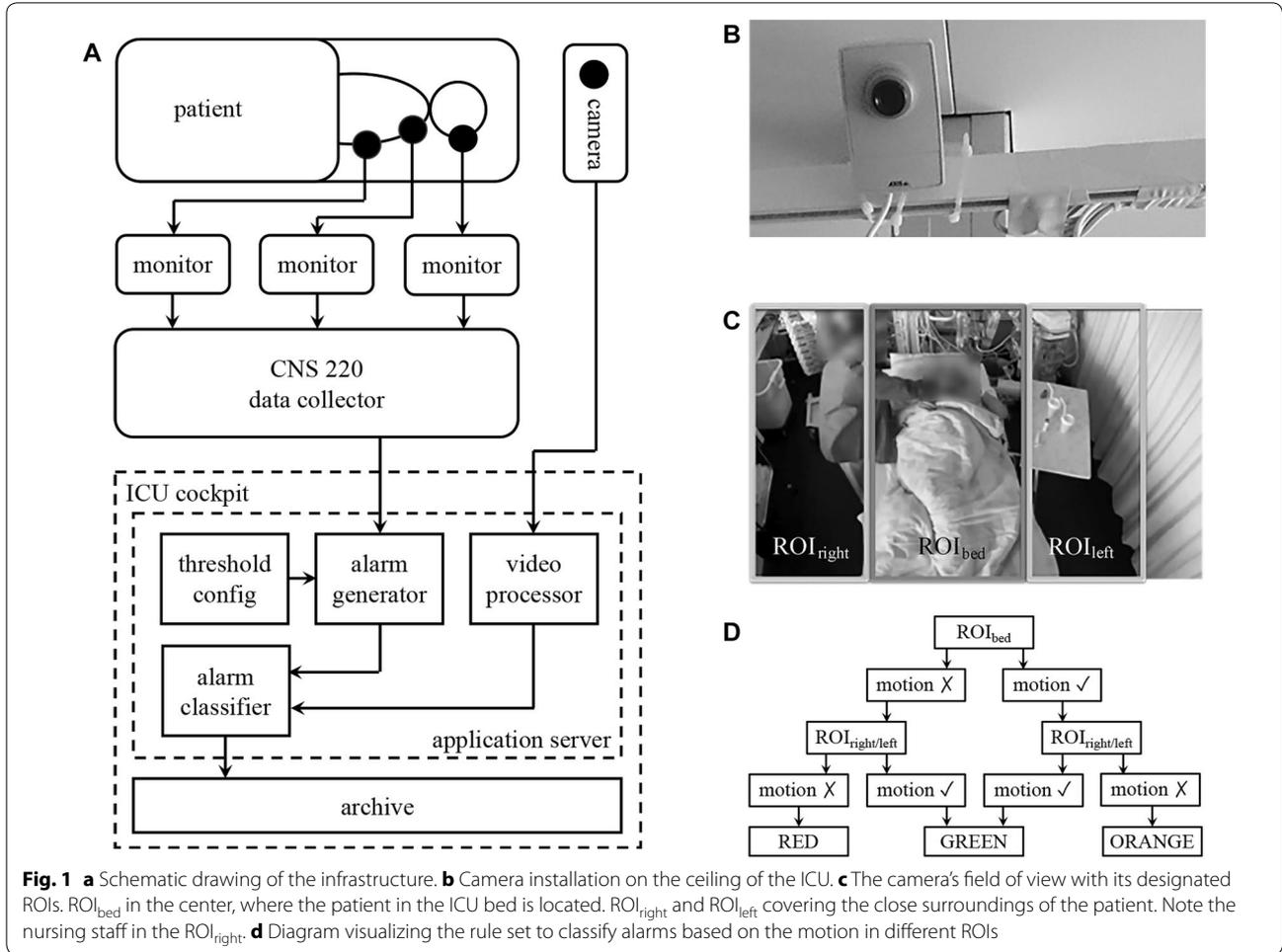
blood pressure (Art_{sys}), and mean arterial blood pressure (MAP), both continuously measured by arterial line, heart rate (HR) measured via electrocardiography and mean intracranial pressure (ICP_{mean}) measured by an ICP probe (ventricular drainage or intraparenchymal pressure transducer probe). The alarm state was defined as "too high" (for Art_{sys} , MAP, HR, ICP) if the according parameter exceeded the upper alarm threshold. Consequently, the alarm state was defined as "too low," if the parameter undercut the lower threshold (for Art_{sys} , MAP, HR, SpO_2). The thresholds that triggered alarms were set and adjusted by the nursing staff according to the treatment targets for the individual patients.

Motion Detection

Two AXIS network camera (AXIS communications AB, Lund, Sweden) were installed on the ceiling over two beds to capture video data and to monitor patient movements and the surroundings (Fig. 1a–c). The video had a spatial resolution of 1024×640 pixels per frame and temporal resolution of 13 frames per second. In the field of view, a main region of interest (ROI) was defined, where the patient (and his bed) was situated (ROI_{bed}) to monitor the patients' movement. Two additional ROIs were automatically added on both sides of the selected area: ROI_{right} and ROI_{left} (Fig. 1c). The latter ROIs captured movements around the bed, which might generate false alarms. The faces were blurred (Fig. 1c), and the video data were irreversibly discarded after computation of motion fields [13].

Optical Flow Algorithm for Motion Detection

Optical flow is the motion of every pixel between two frames taken at times t and $t + \Delta t$. A pixel can be described by (x, y, t) where x and y are the positions in the two dimensions of an image frame and t the time, and is characterized by intensity $I(x, y, t)$. To reduce the amount of processing needed, the optical flow algorithm was run on grayscale images only. Standard optical flow was based on brightness constancy constraint, i.e., a pixel can only move over consecutive frames but does not change its intensity. This assumption can be formulated as follows: $I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$, where Δx and Δy are the pixel displacement in the x and y direction, respectively, and Δt is the time that passed. Assuming that this displacement will only be small, the Taylor series can be developed that finally results in $\frac{\delta I}{\delta x} V_x + \frac{\delta I}{\delta y} V_y + \frac{\delta I}{\delta t} = 0$, where V_x and V_y are the velocity in the x and y direction, respectively, also known as the optical flow. This equation with two unknowns cannot be solved without a further constraint. This further constraint is subject to the algorithm that is applied to



calculate the optical flow. In this project, the Lukas-Kanade method [14] was used that assumes that the flow of pixel $(x; y)$ is constant in a small neighborhood around it. The optical flow equations are solved for all pixels in this patch which results in an overdetermined system of equations. This system is solved by the least squares method. The motion field was calculated on a fixed grid of 20 by 20 pixels. Once the computation of optical flow was finished, we had a vector for every point in the grid indicating the displacement of that point from within the last two frames. These vectors were then summed up as follows: $m_{bed} = \sum_{n \in ROI_{bed}} \|m_n\|$, where m_n is the n -th

motion vector within ROI_{bed} . Analogous summations were done for m_{left} and m_{right} .

Alarm Definition and Classification Using Motion Detection

The alarms were classified into three categories: RED, being a true alarm; ORANGE, possibly a false alarm, but not enough data to verify it; GREEN, being a false alarm. The basic assumption was that the handling of the patient

by medical staff mostly induces signal artifacts that trigger a false alarm [4]. Based on this, a rule set to classify alarms based on the motion in the ROIs was developed and applied (Fig. 1d). The thresholds for the cumulative motion were previously evaluated [13]. In brief, the parameter to be optimized in the motion alarm is the threshold of the cumulative motion vectors m in all ROIs. In a preliminary study, three different thresholds, namely 1.0, 0.75, and 0.5., were tested. Based on these results, a threshold of 0.75 was used for all further experiments. Using 0.5 as a threshold would not result in a big reduction in the false alarms but increase the risk of tagging a RED alarm as a GREEN alarm. An illustrative case of an automated classification is shown in Fig. 2.

Manual Alarm Reclassification

The motion detection-based automated alarm classifications were reclassified by trained neurocritical care specialists that were blinded from video and extracted motion data. Alarms were annotated as GREEN if (1) the signal curve was not visible (sensor disabled), (2)

the signal curve showed an atypical (non-physiological) shape, or (3) the numerical value derived from the signal was completely unrealistic, beyond any pathophysiological range. In case the annotating specialist was in doubt whether the alarm should be labeled as GREEN to his or her best knowledge and belief, the alarm was labeled as ORANGE. All others were labeled as RED. Additionally, a second independent review was conducted over all annotations in order to ensure the internal consistency of the set of annotations as a whole. In case of an inconsistency, a corrected label of the alarm was assigned based on a majority vote of a committee of labelers.

Statistical Analysis

Sensitivity, specificity, positive predictive value (PPV), and negative predictive value were calculated for the whole dataset and for particular subsets using classical confusion matrices. To obtain a measure of the accuracy

of the automated classification, the F_1 -score was calculated. As the F_1 -score does not take the true negatives into account, the Matthews correlation coefficient and the diagnostic odds ratio (DOR) with its standard error (SE_{DOR}) were calculated to obtain a measure of the effectiveness of the automated classification [15]. For statistical analysis, SPSS Ver 23 (IBM, Armonk, NY, USA) was used.

Results

A total of $n = 2349$ alarms from 45 patients were reclassified by clinical experts. Patients' characteristics are shown in Table 1. The distribution of the automatically generated alarms of the specific parameters is as follows: Art_{sys} 45.3% ($n = 1063$), SpO_2 30.2% ($n = 709$), Art_{mean} 19.8% ($n = 465$), ICP_{mean} 3.4% ($n = 81$), and HR 1.3% ($n = 31$). An overview of the automated motion detection-based and manual classification is shown as a contingency table (Table 2). A total of 1025 alarms (43.6%) were classified

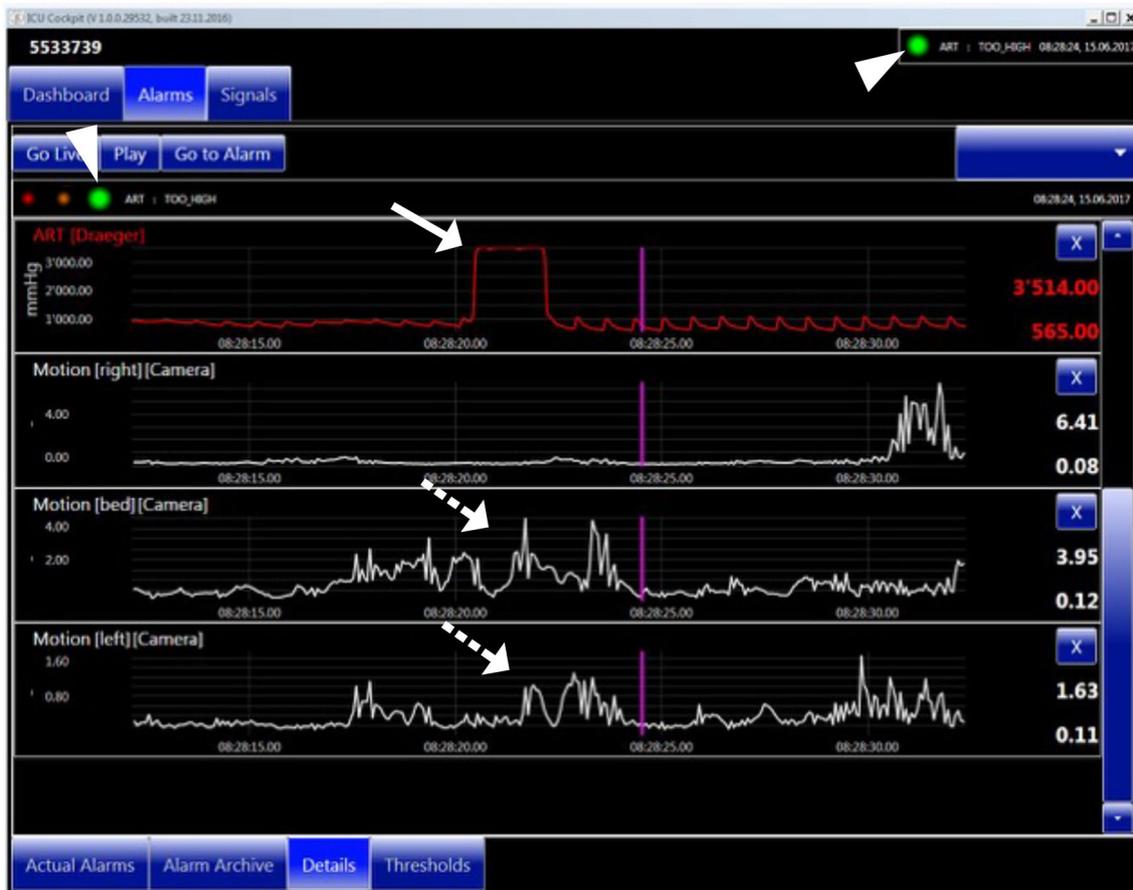


Fig. 2 Screenshot of a representative event. The top row indicates the signal curve for the Art_{sys} , the second row the movement in ROI_{right} , the third row the movement in ROI_{bed} , and the bottom row the movement in ROI_{left} , respectively. The signal of the Art_{sys} exceeded the upper alarm threshold (arrow) and triggered an alarm. As there was motion in ROI_{bed} and ROI_{left} (dotted arrows) that exceeded the threshold, the system classified the alarm as a GREEN alarm (arrowheads), i.e., an artifact due to motion

Table 1 Patients' baseline characteristics

Baseline variable	Value
Age (mean \pm SD)	59.6 \pm 14.9 years
Sex	
Female (n, %)	27 (60%)
Male (n, %)	18 (40%)
Diagnosis	
aSAH (n, %)	29 (64%)
ICH (n, %)	6 (13%)
Other (n, %)	10 (23%)
Artificial ventilation	37 (82%)

aSAH aneurysmal subarachnoid hemorrhage, ICH intracerebral hemorrhage, SD standard deviation

Table 2 Overall results of motion detection-based and manual classification

Automated classification	Manual classification			Total
	RED	ORANGE	GREEN	
RED	1377	114	425	1916
ORANGE	8	8	8	24
GREEN	198	24	187	409
Total	1583	146	620	2349

as “too high” and a total of 1324 (56.4%) as “too low.” The majority of the “too high” alarms were Art_{sys} ($n=729$, 71.1%), followed by Art_{mean} ($n=188$, 18.3%) and ICP_{mean} ($n=78$, 7.6%). The majority of the “too low” alarms were SpO₂ ($n=709$, 53.5%), followed by Art_{sys} ($n=334$, 25.2%) and Art_{mean} ($n=277$, 20.9%). A total of 797 (33.9%) manual annotations were inconsistent, and corrected labels to these alarms were subsequently assigned.

The results of the confusion matrices for parameters and subsets of parameters are summarized in Table 3. Overall, the motion-based algorithm classified 66.9% ($n=1572$) of all alarms correctly and 33.1% ($n=777$) falsely. Generally speaking, for parameters and subsets of parameters, the sensitivity for RED alarms was high (70.4–100.0%) and specificity low (17.3–55.6%). The opposite was observed for GREEN alarms, with low sensitivities (17.2–83.3%) and high specificities (73.2–100%). Considering all parameters together, the sensitivity for RED alarms was 87.0%. The specificity was low with 29.6%. For all GREEN alarms, the sensitivity was low with 30.2%, while the specificity was 87.2%, which, however, meant a false-positive rate of 12.8%. A detailed overview of the false-positive rate for GREEN alarms is given in Table 4. The F₁-score was high for RED alarms and low for GREEN alarms, while the Matthews correlation

coefficient (MCC) and DOR were similar. This pattern was more or less present in the subanalyses (Table 3).

The sensitivity for RED alarms was 88.6% for “too low” alarms and was superior to the sensitivity for “too high” RED alarms (85.4%). For GREEN alarms, the sensitivity for “too high” alarms was superior to the sensitivity for “too low” alarms. The F₁-scores, MCC, and DOR were higher for both RED and GREEN “too high” alarms (Table 3).

Parameter-stratified results showed that our system performs the best on HR-generated alarms. Yet, due to a small sample size ($n=31$), the results must be interpreted with caution. When considering classification accuracy for RED alarms only, classification based on ICP_{mean} alarms achieved the second best results. However, the sample size was equally small ($n=81$). For GREEN alarms, Art_{sys} performed the second best (Table 3).

The signals Art_{sys} and Art_{mean} were further stratified into “too high” and “too low” alarms. For RED alarms, the sensitivity was best for Art_{sys} “too low” alarms. Taking the F1-score, MCC, and DOR into account, the automated classification seemingly performed better for Art_{sys} than for Art_{mean}. Concerning GREEN alarms, the sensitivity was best for Art_{sys} “too high” alarms. Taking the F1-score, MCC, and DOR into account, the automated classification seemingly performed better for Art_{sys} than for Art_{mean}.

Discussion

A novel automated alarm classification algorithm based on motion detection with cameras has been developed and evaluated. The obtained classification results suggest a high sensitivity for RED (true) alarms, but also a high false-positive rate. Taking into account that RED alarms should not be missed under any circumstances, even at the expense of false-positive ones, the system met the requirements. However, the false-negative rate was 13%. Further improvement of the algorithm to lower the false-negative rate is desirable. The system featured low specificities and therefore performed less convincingly when it came to correctly ruling out RED alarms. Considering all alarms, the system was able to correctly rule out 29.6% of alarms. However, ORANGE is “uncertain alarms” by definition. From a medical and medico-legal point of view, one would suggest that if in doubt, an “uncertain” alarm should be regarded as “real.” Therefore, ruling out only RED alarms would bear some risks that cannot be accepted in the ICU setting. Thus, in order to reduce the number of false alarms, the target goal would be to achieve a high correct classification of GREEN alarms. Our results reveal that the current system does not meet this target. The sensitivity for all GREEN alarms was 30.2%, i.e., a 69.8%

Table 3 Detailed results of the confusion matrices

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	F ₁ -Score	MCC	DOR	SE _{DOR}
All parameters (n = 2349)								
RED	87.0	29.6	71.9	52.4	0.787	0.201	2.82	± 1.14
GREEN	30.2	87.2	45.7	77.7	0.363	0.201	2.93	± 1.12
Too high, all parameters (n = 1025)								
RED	85.4	38.5	84.5	40.2	0.850	0.243	3.67	± 1.19
GREEN	40.3	85.5	33.0	89.0	0.363	0.238	3.98	± 1.21
Too low, all parameters (n = 1324)								
RED	88.6	26.3	62.3	62.8	0.732	0.194	2.79	± 1.16
GREEN	26.8	88.8	56.6	69.1	0.364	0.200	2.91	± 1.16
SpO ₂ (n = 709)								
RED	90.7	17.3	66.1	51.2	0.765	0.118	2.05	± 1.26
GREEN	17.2	91.2	46.9	70.9	0.252	0.122	2.15	± 1.27
Art _{sys} (n = 1063)								
RED	87.1	36.5	81.5	46.7	0.842	0.258	3.87	± 1.18
GREEN	38.4	86.7	38.6	86.6	0.385	0.252	4.07	± 1.20
Art _{sys} too high (n = 729)								
RED	86.7	39.6	89.40	33.60	0.880	0.246	4.27	± 1.26
GREEN	42.1	86.4	26.45	92.76	0.325	0.234	4.61	± 1.30
Art _{sys} too low (n = 334)								
RED	88.3	34.2	63.4	69.4	0.738	0.272	3.93	± 1.33
GREEN	36.0	87.7	60.3	72.6	0.451	0.279	4.01	± 1.33
Art _{mean} (n = 465)								
RED	76.2	35.0	53.7	59.9	0.630	0.123	1.73	± 1.23
GREEN	35.6	78.6	54.3	63.0	0.430	0.157	2.03	± 1.23
Art _{mean} too high (n = 188)								
RED	70.4	40.0	61.3	50.0	0.655	0.108	1.58	± 1.36
GREEN	38.5	73.2	43.1	69.2	0.407	0.120	1.70	± 1.39
Art _{mean} too low (n = 277)								
RED	81.3	32.5	49.0	68.5	0.612	0.155	2.09	± 1.33
GREEN	34.1	83.1	63.8	59.1	0.444	0.199	2.55	± 1.33
ICP _{mean} (n = 81)								
RED	93.8	25.0	83.6	50.0	0.884	0.251	5.08	± 2.17
GREEN	22.2	93.1	28.6	90.5	0.250	0.171	3.83	± 2.52
Heart rate (n = 31)								
RED	100.0	55.6	84.6	100.0	0.917	0.686	n.a.	n.a.
GREEN	83.3	100.0	100.0	96.2	0.909	0.895	n.a.	n.a.

n.a., cannot be calculated as contingency table contains zeros

Art_{sys} arterial blood pressure, DOR diagnostic odds ratio, ICP_{mean} mean intracranial pressure, MCC Matthews correlation coefficient, NPV negative predictive value, PPV positive predictive value, SpO₂ peripheral oxygen saturation

false-negative rate. The latter may be acceptable, as labeling GREEN alarms as RED (or ORANGE) would not lead to potentially harmful consequences for the patient, and it would still roughly eliminate a third of false alarms. However, there was a 45.7% PPV and an 12.8% false-positive rate for all GREEN alarms, which means that more than half of the alarms classified as GREEN were RED (or ORANGE) alarms, and 12.8% of RED (or ORANGE) alarms were classified as GREEN

by the system. Although the results of the data subsets showed some differences, the overall trend was clear with a high sensitivity for RED and low for GREEN alarms, accompanied with an unsatisfactory PPV for GREEN alarms. To summarize, the false positives (type I error) for GREEN alarms are problematic and must be further reduced. However, as the sensitivity and specificity, among other measures, for RED and GREEN alarms are interrelated, an attempt to increase the

Table 4 False-positive rates stratified by signals

All parameters ($n = 409$)	12.8%
Too high, all parameters ($n = 188$)	15.5%
Too low, all parameters ($n = 221$)	12.2%
SpO ₂ ($n = 81$)	8.8%
Art _{sys} ($n = 189$)	13.3%
Art _{sys} too high ($n = 121$)	13.6%
Art _{sys} too low ($n = 68$)	12.3%
Art _{mean} ($n = 127$)	21.4%
Art _{mean} too high ($n = 58$)	26.8%
Art _{mean} too low ($n = 69$)	16.9%
ICP _{mean} ($n = 7$)	6.9%
Heart rate ($n = 5$)	0.0%

Art_{sys}, arterial blood pressure, ICP_{mean}, mean intracranial pressure, SpO₂, peripheral oxygen saturation

sensitivity for GREEN alarms by lowering the thresholds of the cumulative motion would not only lead to a decrease in sensitivity for RED alarms [13], but also increase the false-positive rate for GREEN alarms. The results suggest that improved algorithms are needed to solve this issue.

Another problem challenge is the “uncertain” alarm, represented by the ORANGE alarms. Interestingly, by manual classification, the ORANGE alarms annotations were 8 times higher than by automated classification, which reflects the difficulty to distinguish a real alarm from a false alarm in a real-life ICU setting. As already indicated, from a medical and medico-legal point of view, ORANGE alarms should be regarded as a “real” alarm. Hence, only the results of the GREEN alarms have to be taken into consideration. Concerning automated classification, an alarm was labeled ORANGE when motion of the patient, but no motion in ROI_{right/left} was present. The rationale behind this is that patient movement would less likely produce artifacts triggering an alarm. However, a real alarm is still possible while the patients move. It can be assumed that improving the automated classification and resolving type I errors for GREEN alarms might be furthered by the same means, resulting in the elimination of ORANGE alarms.

The shortcoming of the current analysis is the fact that it could not be judged whether falsely GREEN tagged RED/ORANGE alarms could have serious clinical consequences or not. To note, even a real (i.e., not motion artifact triggered) alarm might not be of any concern in a given clinical context. However, as the primary objective was to detect motion triggered false alarms, the detailed clinical context of every single alarm was not analyzed.

As options for improvement, a change of the motion threshold would unlikely be the solution, as previously mentioned. An improvement of the vision-based motion detection might be considered. The ROI_{bed} might be constructed more specific, to enable the system to identify, track, and identify movement of different body parts, such as head, torso, and extremities. Having information regarding on which body part the respective sensor is located, the system might provide better results. Furthermore, the combination with other rules—such as decreasing alarm thresholds—and algorithms acting on biosignals independently from video analytics might improve the quality of performance [3]. In other words, the currently developed motion detection-based algorithm could serve as a basic layer, while further layers can be added. One possibility is to add an automated signal curve analysis, which would identify atypical shapes, or optimally, typical shapes for artifacts. Alternatively, a more complex set of rules, which takes multiple parameters and their interrelations into account, could be added. Such a system has been proposed and evaluated for post-coronary artery bypass graft patients with promising results [12]. In this particular setup, respiration rate, blood pressure, SpO₂, and HR were classified using fuzzy sets. In another study, applied fuzzy logic was able to detect hypovolemia by monitoring HR, blood pressure, and pulse volume [11]. Last but not least, the addition of a smart alarm system based on machine learning might be considered. A machine-learning approach using a novel multitask network architecture that utilizes distant supervision through multiple related auxiliary tasks—in order to reduce the number of expensive labels required for training—has very recently been developed by our group with promising results [3].

Conclusion

In conclusion, a novel motion detection-based alarm classification algorithm has been developed and evaluated in a real neurocritical care setting for the parameters SpO₂, Art_{sys}, MAP, HR, and ICP_{mean}, demonstrating an 87% sensitivity to detect true alarms for all parameters. However, the sensitivity to detect artifacts, i.e., false alarms, was 30%. Although a reduction in nearly one third of false alarms would already be beneficial, the 13% false-positive rate, i.e., type I error, is problematic and should be improved. This would likely be achieved by adding further layers, which might be: (1) an automated analysis of the shape of signal curve, (2) the application of a complex set of rules, which takes multiple parameters and their interrelation into account, and (3) the addition of a smart alarm system based on machine learning. In

this context, the current system may serve as a robust basis for a complex automated false alarm detection algorithm.

Author details

¹ Neurocritical Care Unit, Department of Neurosurgery, University Hospital Zurich, 8091 Zurich, Switzerland. ² Computer Vision Lab, ETH Zurich, 8092 Zurich, Switzerland. ³ Mobile Health Systems Lab, Institute of Robotics and Intelligent Systems, Department of Health Sciences and Technology, ETH Zurich, 8092 Zurich, Switzerland.

Acknowledgments

The authors thank Dr. Christof Bühler, Susanne Suter, and Rudolf Limacher, Supercomputing Systems AG for the development of the graphical user interface (GUI) and coding. The authors thank Mrs. Anna Wang for English proofreading.

Author contribution

VDL, WK, and EK contributed to the conception and study design. SM, DJM, and CS contributed to the data collection. CM, SM, DJM, and PS contributed to the data analysis and interpretation. CM and EK contributed to the manuscript preparation.

Source of support

The study was financially supported by the Swiss National Science Foundation, the Gebert RUF Foundation, the USZ Foundation, and the Commission for Innovation and Technology, Switzerland.

Compliance with ethical standards

Conflicts of interest

All authors have no conflict of interest to declare.

Ethical Approval/Informed Consent

All research contained here in was approved by ethics committee of Kanton Zurich, Switzerland.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published online: 09 July 2019

References

1. Li Q, Clifford GD. Signal quality and data fusion for false alarm reduction in the intensive care unit. *J Electrocardiol.* 2012;45(6):596–603.
2. Roederer A, Dimartino J, Gutsche J, et al. Clinician-in-the-loop annotation of ICU Bedside Alarm Data. In: 2016 IEEE first international conference on connected health: applications, systems and engineering technologies (CHASE); 2016. p. 229–37.
3. Schwab P, Keller E, Muroi C, et al. Not to cry wolf: distantly supervised multitask learning in critical care. In: Proceedings of 35th international conference on machine learning research (ICML) 2018; PMLR(80):4518–27.
4. Gorges M, Markewitz BA, Westenskow DR. Improving alarm performance in the medical intensive care unit using delays and clinical context. *Anesth Analg.* 2009;108(5):1546–52.
5. Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. *Anesth Analg.* 2006;102(5):1525–37.
6. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. *AACN Adv Crit Care.* 2013;24(4):378–86.
7. Cvach M. Monitor alarm fatigue: an integrative review. *Biomed Instrum Technol.* 2012;46(4):268–77.
8. Drew BJ, Harris P, Zegre-Hemsey JK, et al. Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. *PLoS ONE.* 2014;9(10):e110274.
9. Balogh D, Kittinger E, Benzer A, Hackl JM. Noise in the ICU. *Intensive Care Med.* 1993;19(6):343–6.
10. Xie H, Kang J, Mills GH. Clinical review: the impact of noise on patients' sleep and the effectiveness of noise reduction strategies in intensive care units. *Crit Care.* 2009;13(2):208.
11. Baig MM, Gholamhosseini H, Harrison MJ. Fuzzy logic based smart anaesthesia monitoring system in the operation theatre. *WSEAS Trans Circuits Syst.* 2012;11(1):21–32.
12. King A, Fortino K, Stevens N, et al. Evaluation of a smart alarm for intensive care using clinical data. In: Conference proceedings of IEEE Engineering in Medicine and Biology Society; 2012. p. 166–9.
13. Meier S. Bachelor thesis: Alarm detection and classification in neurointensive care. Bachelor Thesis 2016, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland.
14. Lucas BD, Kanade T. *An iterative image registration technique with an application to stereo vision.* In: Proceedings of the 7th international joint conference on artificial intelligence. vol 2. Vancouver, BC, Canada: Morgan Kaufmann Publishers Inc.; 1981.
15. Glas AS, Lijmer JG, Prins MH, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol.* 2003;56(11):1129–35.