

Feasibility of Automated Vital Sign Instability Detection in Children Admitted to the Pediatric Intensive Care Unit

Georg Seidel^{1,2}, Srinivas Murthy², Cheryl Peters², Philipp Rostalski¹, Matthias Görges²

¹Universität zu Lübeck, Lübeck, Germany

²The University of British Columbia, Vancouver, Canada

Abstract

Children admitted to a Pediatric Intensive Care Unit (PICU) are at risk of deterioration, which can lead to a cardiac arrest if undetected. Outcomes after pediatric cardiac arrest remain poor, even for witnessed, in-hospital events. Thus, early detection of deterioration is paramount; ideally long before the risk of harm increases. Vital signs trends of patients admitted to the PICU at BC Children's Hospital were extracted from local outcomes registries (n=96). A rule-based algorithm (RBA) for detecting vital signs instabilities was developed; we did so in the expectation of enhancing clinician trust compared to black box approaches such as deep neural networks. Two PICU physicians provided expert classifications for episodes indicative of vital signs instability or their absence. The RBA's best result generated 91.6% correct, 6% false negatives, and 3% false positives on the test data (n=29) showing promise for eventual application in a clinical setting. Future research is needed to refine the algorithm and implement it clinical practice.

1. Introduction

Patients admitted to a Pediatric Intensive Care Unit (PICU), are critically ill or injured and have an increased risk of rapid deterioration leading to cardiac arrest. Cardiac arrests occur in approximately 2-6% of children admitted to PICU [1], but deterioration happens more frequently. If left undetected or interventions fail, deterioration may result in cardiac arrest as a final step. Hence, early detection represents an enormous source of preventable morbidity, mortality and cost [2], [3].

Kennedy *et al.* observed that "subtle drops may be noted as many as 20 minutes before an arrest" [4]. Pollack *et al.* demonstrated that early detection of high-risk patients suffering a cardiac arrest is possible if four hours of clinical and monitoring data are available [5]. A previous in-house case review of vital sign trend data by PICU physicians identified that signs of instabilities preceded cardiac arrest; thus, it was deemed possible that there is a relation between vital sign variability and the critical instabilities that

precede cardiac arrest.

The major goal of this research is to develop a reliable tool for early detection of critical instabilities that provides PICU clinicians with support in their triaging task. Decision support systems may help by identifying the need for closer monitoring and treatment adjustments. However, we must also acknowledge a major problem in health care: the risk of alarm fatigue due to high false alarm rate [6]. Current patient monitors use adjustable threshold alarms to notify clinicians if a patient exceeds pre-specified limits. Artifacts from both sensor noise and patient movement are common causes of false alarms; thus, more sophisticated approaches to reduce false alarm rates are required.

PICU patients' variation by age and condition is more significant than it is in adults; the range from newborn to 18 years is associated with significant (and expected) changes in all routinely measured physiological variables. This poses a major obstacle to the use of absolute threshold values for generating clinically relevant alerts. This study aims to address this issue by using relative thresholds and other vital sign change features.

2. Material and Methods

We aim to demonstrate that short-term instabilities can be detected by a rule-based algorithm (RBA), and to test its performance against a physician's expert classification (EC). A RBA was deemed preferable to Deep or Machine Learning in this context, as its functioning is more transparent to clinicians and it does not need huge datasets.

With approval from the University of British Columbia and Children's & Women's Health Centre of British Columbia Research Ethics Board (H17-01627), this retrospective study used a database of previously collected vital signs data, with waiver of patient consent. Vital signs trend data of patients, admitted to the BC Children's Hospital PICU between January 1st 2016 and June 30th 2017 were analyzed. These data were previously captured from MP70 IntelliVue Patient Monitors (Philips Healthcare, Best, Netherlands) using in-house developed software., Children who had experienced at least one episode of cardiopulmonary resuscitation (CPR) were

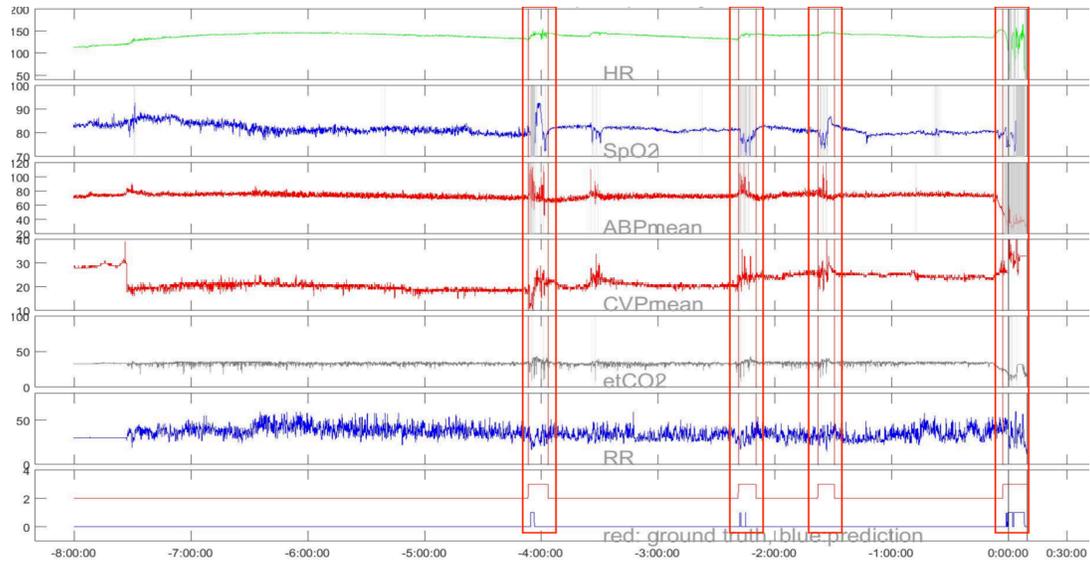


Figure 1. CPR cohort vital sign plot. Instability annotations fall within red boxes. Vital signs: heart rate (HR), oxygen saturation (SpO_2), mean arterial blood pressure (ABP_{mean}), mean central venous pressure (CVP_{mean}), end-tidal carbon dioxide ($etCO_2$), and respiratory rate (RR).

identified from a local outcomes registry. Patients in the CPR cohort had been admitted with different diagnoses, primarily congenital cardiac disease or pulmonary/respiratory problems such as pneumonia. Two control cohorts (i.e. without cardiac arrests) were created: from patients with complex admitting diagnoses as outlined above (cohort A) and relatively healthy patients admitted to the PICU following tonsillectomy or spinal surgery (cohort B). Some patients had additional and/or multiple underlying etiologies.

MATLAB R2016b (MathWorks, Natick, USA) was used for preprocessing and data analysis. Data were imported as CSV-files and preprocessed; this included resampling data at 0.2 Hz (one sample every five seconds), removal of outliers and error-state values. Data were selected from a predefined time frame: for CPR cases, this was 8 hours before CPR episode, plus data until the end of the CPR episode; for control cases, this was a random 8 hours from the available set. Data were plotted (Fig. 1) and reviewed by PICU intensivists to identify periods of vital signs instabilities.

During the first review, only cases from the CPR cohort and control cohort A were presented to physician A. Due to inconsistencies in the first EC, where similar patterns in some cases were marked as an event while others were not [7], physician A attempted a second review, in which RBA performance was reviewed, using vital signs, A's former annotations and the RBA's identified events. This method raised concerns about biasing the result; thus, it was the only version in which the RBA's annotations were available. Physician B reviewed data without additional information available, such as physician A's annotations, RBA's prediction, or knowledge whether the presented

case ended in CPR or not. Finally, as conflicts were found between the two physicians' indications, physician A was asked to resolve these conflicts (Table 1). Physicians provided their annotations by drawing markings on a tablet PC on which the plots (e.g. Fig. 1) were presented.

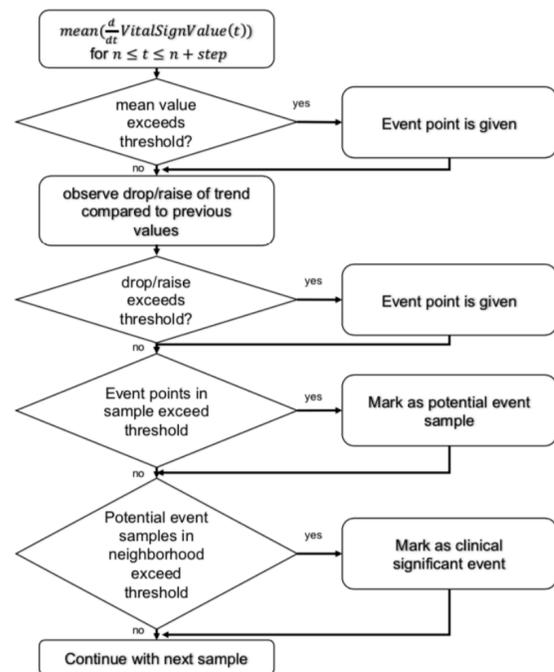


Figure 2. Flow chart of the Rule Based Algorithm

The dataset was split into training and test patients, whereby 67/96 (70%) patients were used for

training/optimizing the parameters, and 29/96 (30%) for testing. The proposed RBA (Fig. 2) evaluated variability of the heart rate (HR), oxygen saturation (SpO₂), respiratory rate (RR), end-tidal carbon dioxide concentration (etCO₂), and mean arterial blood pressure. The invasive arterial blood pressure (ABP) was used when available; otherwise, the non-invasive blood pressure (NBP) was used.

Table 1. Expert Classification review by two different physicians (pediatric intensivists) across 4 rounds

Review	physician	question	patients
1.	A	'events a physician wants to know about'	65
2.	A	check the algorithm's performance on first classification and for the plots, which have not been reviewed	96
3.	B	'events a physician wants to know about'	96
4	A	Conflicts between second and third classification 'should we aim for detecting these events or not?'	96

Thirteen tuning parameters allowed for optimization:

- 2 delays - minimal number of samples needed as a reference before starting detection
- 5 factors - vital sign variability thresholds, one for each of HR, RR, etCO₂, SpO₂, BP; if exceeded the vital sign in the sample is flagged
- 3 percent threshold of vital sign deviations - one for each of HR, SpO₂, BP; if exceeded the vital sign in the sample is flagged
- number of event points having to occur in one sample to yield a potential event
- number of potential event samples in a predefined window which have to be present to identify this event set as a clinically relevant instability

Initially, the tuning parameters were chosen randomly. Next, different combination sets of parameters were tested systematically to optimize the algorithm's performance by exploring the parameter space using an automated test algorithm. Finally, validation of the algorithm and selected parameters were performed on the test dataset; best performance was chosen by applying a cost function, which included a double weighted false positive (FP) rate, and a single weighted false negative (FN) rate.

A window of 15 minutes was used to calculate the FP, FN, true positive (TP), and true negative (TN) episodes. This was done in part to address the potential delayed detection or tagging of an event. Clinically relevant events detected by the RBA were counted during each 15-minute window. Finally, the number of expected and detected events were compared. Thus, event classification correctness was performed once every 15 minutes, and hence the total number of episodes is less than the number of all samples.

3. Results

The CPR cohort (n=30) included 14 patients diagnosed with congenital cardiac diseases and 3 with acute respiratory distress syndrome; all patients in this cohort had at least one episode of CPR. Control cohort A (n=45) included only patients with congenital cardiac diseases, who did not require CPR. Control cohort B (n=21) included 11 patients post-tonsillectomy and 10 patients post-spine surgery.

We obtained different results for the three different versions of the EC (1, 2, 4) used for training and testing of the algorithm. Preliminary results, based on EC1, were previously reported [7] (Table 2).

Table 2: Performance on expert classification 1 (n=65)

	training data	test data		
False Positive Rate	10.0%	8.4%		
False Negative Rate	51.8%	66.7%		
Accuracy	85.7%	83.3%		
data	FP	FN	TP	TN
Training	9.0%	5.3%	4.9%	80.8%
Test	7.2%	9.5%	4.7%	78.6%

Due to inconsistencies in EC1, physician A reviewed the data a second time as described above. With EC2, the algorithm performed with an accuracy of 91.6% on the test data (Table 3).

Table 3: Performance on expert classification 2 (n=96)

	training data	test data		
False Positive Rate	5.1%	3.2%		
False Negative Rate	31.9%	40.8%		
Accuracy	92.4%	91.6%		
data	FP	FN	TP	TN
Training	4.6%	2.9%	6.4%	86.1%
Test	2.7%	5.6%	8.1%	83.6%

The conflicts identified after consulting a second physician were resolved by physician A. However, due to the focus on resolving conflicts, a slightly reworded question was asked at this time: 'Should we aim for detecting these events or not'; physician A was also shown the original annotations at this time. This change of emphasis yielded significantly different annotations from EC1 and EC2 as, instead of identifying new events, the physician aimed to reduce false alarms. This resulted in different performance scores for EC 4 (Table 4).

Table 4: Performance on expert classification 4 (n=96)

	training data	test data		
False Positive Rate	7.7%	7.7%		
False Negative Rate	31.5%	42.1%		
Accuracy	90.9%	90.0%		
data	FP	FN	TP	TN
Training	7.3%	1.7%	3.6%	87.4%
Test	7.2%	2.7%	3.8%	86.3%

4. Discussion

This study has demonstrated the feasibility of automated detection of short-term instabilities using an RBA, with an accuracy of 91.6% against EC2. The RBA performed similarly on other tested ECs (83.3% and 90.0%), which support our performance assessment. Kennedy et al. [4], and Pollack et al. [5] had already demonstrated that vital signs can be a good predictor for instabilities and thus can be used to predict cardiac arrest. Hence, detecting these short-term instabilities may be the first step toward establishing a more sophisticated algorithm for cardiac arrest prediction.

4.1. Limitations

Establishing a reliable EC was a significant challenge. Our physicians' decision-making may have been biased if it was known that a case belonged to the CPR or control cohorts, how the specific questions to the experts were posed, and whether a previous EC or machine prediction (MP) annotation was available. With notable differences especially between EC 1-2 vs. EC4, it is difficult to derive clinical implications from the current data. To reduce bias in future tests, physicians should review cases blinded to cohort (CPR or control) and previous EC or MP annotations. Furthermore, consulting a greater number of physicians should allow inconsistencies between different ECs to be analyzed and resolved.

The number of patients with episodes of CPR was limited (n=30), and consequently, the number of instabilities identified by the PICU physicians was quite low. This generated a large number of no-event samples, which made the dataset strongly unbalanced and may be responsible for our high accuracy. As a countermeasure, the amount of stable patients and thereby no-event samples could be reduced, or event episode windows could be resampled to increase their frequency.

4.1. Future research

Strategies for long-term deviation detection should be developed and evaluated, as they may allow for prediction of deterioration earlier. Next, RBA performance may be improved through the introduction of clinically relevant transformations, such as the shock index and an estimated cardiac output. These variables are typically only available from entirely separate monitoring devices, suggesting that they would not always be available and that some integration would be required to incorporate them into a single RBA. However, some useful transformations can be calculated by combining routinely available measures, such as HR and BP, for example. Next, normalizing vital signs through the use of Z-scores may allow comparison of

vital signs with absolute thresholds. This has the potential to minimize the age-dependence that makes developing models for use in the pediatric setting so challenging. Finally, results should be verified by cross-validation, especially as the data were highly unbalanced.

4.1. Conclusion

Automated detection, by a RBA, of vital signs patterns that represent potential instabilities preceding a cardiac arrest was shown to be feasible. Given the challenges in reliably identifying the reference labels, working with heavily unbalanced data, and the lack of long-term trend detection features, additional work is needed before attempting clinical implementation.

Acknowledgments

We wish to thank Gordon Krahn for extracting patient data from the registry, and Nicholas West for editorial assistance. No authors report any conflict of interests.

References

- [1] Berg MD, Nadkarni VM, Zuercher M, Berg RA. In-hospital pediatric cardiac arrest. *Pediatrics Clinics of North America* 2008;55(3):589–604.
- [2] Schein RM, Hazday N, Pena M, Ruben BH, Sprung CL. Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest* December 1990;98(6):1388–1392.
- [3] Al-Qahtani S, Al-Dorzi HM. Rapid response systems in acute hospital care. *Annals of thoracic medicine* January 2010;5(1):1–4.
- [4] Kennedy CE, Aoki N, Mariscalco M, Turley JP. Using time series analysis to predict cardiac arrest in a picu. *PCCM* November 2015;16(9):332–339.
- [5] Pollack MM, Holubkov R, Berg RA, Newth CJ, Meert KL, Harrison RE, et al. Predicting cardiac arrests in pediatric intensive care units. *Resuscitation* 2018;133:25–32.
- [6] Borowski M, Gorges M, Fried R, Such O, Wrede C, Imhoff M. Medical device alarms*. *Biomed Tech* 2011;56:73–83.
- [7] Seidel G, Rostalski P, Gorges M. Feasibility of automated vital sign instability detection in children admitted to the pediatric intensive care unit using variability and vital sign changes indices. *Student Conference Proceedings 2019 of 8th Conference on Medical Engineering Science* March 2019;199–202.

Address for correspondence:

Matthias Gorges, BC Children's Hospital Research Institute, Digital Health Innovation Lab, V3-324 - 948 28th Ave W, Vancouver, BC V5Z 4H4, Canada
mgoerges@bcchr.ca