# Reduction of Clinical Alarms in the Paediatric Intensive Care Unit

## A Machine Learning Approach

C.A. van de Ruit

October 2025

Master of Science in
Technical Medicine

*This page is intentionally left blank*

# Reduction of Clinical Alarms in the Paediatric Intensive Care Unit: A Machine Learning Approach

**Christie van de Ruit**
Student number: 4865618
October 2025

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

## Technical Medicine

Leiden University ; Delft University of Technology ; Erasmus University Rotterdam

Master thesis project (TM30004 ; 35 ECTS)
Dept. of Biomechanical Engineering, TUDELFT
*March 2025 – October 2025*

Supervisors:
Prof.dr.ir. A.C. Schouten, TU Delft
Dr. J.W. Kuiper, Erasmus MC
Dr. R.C.J. de Jonge, Erasmus MC
M. Kalden, Erasmus MC
E. van Twist, Erasmus MC
B. van Winden, Erasmus MC

Thesis committee members:
Dr. J.W. Kuiper, MD, PhD, Erasmus MC (chair)
Prof.dr.ir. A.C. Schouten, Professor, TU Delft
Dr.ir. T.G. Goos, PhD, Erasmus MC

An electronic version of this thesis is available at http://repository.tudelft.nl/.

# Preface

After seven years of study, I am proud to finally call myself a Technical Physician. Throughout my degree I often reflected on the direction I wished to pursue. I enjoy working on a broad, overarching level with medical technology while focusing on healthcare professionals and how they adapt to change. I particularly value being the link between healthcare and technology and exploring how innovation can enhance the job satisfaction of healthcare professionals.

I found all of these elements come together in my graduation project. Conversations with nurses helped me identify their real needs and consider the wider context of integrating new technology into daily practice. As a final challenge, I also undertook a Python project, something I had not anticipated when I began my thesis.

I would like to express my sincere gratitude to everyone who supported me during this journey. First of all, **Eris** and **Brian**: it was invaluable to be able to stop by each day for thoughtful advice or simply to share a conversation about Brian's cycling adventures or Eris's running achievements. Your prompt feedback and efficient organisation of data allowed me to progress smoothly.

Many thanks to **Jan Willem** and **Rogier** for your enthusiasm, your valuable medical perspectives on the project and your guidance on clinical practice in the paediatric intensive care unit.

My appreciation also goes to **Melissa** for her guidance in machine learning. It was inspiring to discover new techniques and to realise how much there is still to explore in this field, as well as the challenge of applying these methods in clinical practice.

Thank you to **Alfred** for your technical advice. Preparing presentations for you was both helpful and enjoyable, and I appreciated the opportunity to exchange ideas.

To my **fellow students** and **housemates**: thank you for your encouragement, the many coffees at the *Koffiemannetje*, the fun activities and the countless conversations about our Technical Medicine graduation projects.

Finally, I extend my heartfelt thanks to **my family** and **boyfriend** for their unwavering support and encouraging words. During my studies I experienced the sudden loss of my father, which gave me a deeper understanding of what families experience in an intensive care unit. With your support I was able to persevere and complete my degree successfully.

In this thesis I aim to provide a clear picture of alarm fatigue and alarm reduction in the paediatric intensive care unit, highlighting both the technical interventions that can help and the perspectives of healthcare professionals. I hope you enjoy reading it as much as I enjoyed carrying out the project!

**Christie van de Ruit**
*Rotterdam, October 2025*

# Abstract

**Background:** Continuous monitoring in paediatric intensive care units (PICUs) generates frequent clinical alarms, 87-97% of which are nonactionable, contributing to alarm fatigue and patient safety risks. At the Erasmus MC Sophia Children's Hospital PICU, improved alarm management is needed. Machine learning offers a promising strategy to distinguish actionable from nonactionable alarms.

**Objectives:** The primary objective was to develop a machine learning algorithm to classify actionable alarms from multimodal vital signs. Secondary objectives were to characterise the PICU alarm burden, capture stakeholder perspectives and evaluate model feasibility for clinical application.

**Methods:** Retrospective Dräger monitoring data and alarms from 2,582 PICU patients (Nov 2021 – Oct 2024) were analysed. A machine learning algorithm was developed using ART M, HR, RESP and $SpO_2$ data from 26 patients, with alarms annotated using clinical interventions, small signal deviations and temporal associations with other parameters. Logistic regression, decision tree, random forest and XGBoost were evaluated with nested cross-validation using pre-alarm features. Performance was assessed by sensitivity, specificity, balanced accuracy, with AUROC and F1-score as complementary metrics. Semi-structured interviews with four nurses and one psychologist explored alarm experiences, impacts and reduction strategies.

**Results:** The best model, a decision tree, showed limited performance (sensitivity 0.36-0.49, specificity 0.51-0.70, balanced accuracy ≈0.50). No discriminative features were identified, and substantial overlap and outliers limited classification. Most alarms involved $SpO_2$ desaturation, with unit variation and temporal patterns linked to ward activity. Interviews highlighted overstimulation and desensitisation, but also that nonactionable alarms can serve as early warnings, with interpretation requiring clinical context.

**Discussion:** This study provides an evaluation of machine learning-based alarm classification in the PICU and integrates clinician perspectives to guide future interventions. The machine learning model was not suitable for clinical use because of retrospective labelling, the small annotated dataset and the absence of clinical context. Future research should focus on prospective annotation, larger and more diverse datasets and complementary strategies. Alarm dashboards and daily reviews are recommended to reduce alarm burden and mitigate alarm fatigue while maintaining patient safety.

# Table of Contents

# List of Abbreviations

| | |
|---|---|
| **AI** | Artificial intelligence |
| **ART M** | Mean invasive arterial blood pressure |
| **AUROC** | Area under the receiver operating characteristic curve |
| **BP** | Blood pressure |
| **DT** | Decision tree |
| **ECG** | Electrocardiogram |
| **ECMO** | Extracorporeal membrane oxygenation |
| **$FiO_2$** | Fraction of inspired oxygen |
| **HR** | Heart rate |
| **ICU** | Intensive care unit |
| **IQR** | Interquartile range |
| **LR** | Logistic regression |
| **NaCl** | Normal saline |
| **NICU** | Neonatal intensive care unit |
| **PEEP** | Positive end-expiratory pressure |
| **PICU** | Paediatric intensive care unit |
| **PIM 3** | Paediatric index of mortality 3 |
| **PRISM III** | Paediatric risk of mortality III |
| **PLS** | Pulse rate derived from pulse oximetry ($SpO_2$) |
| **Q1** | First quartile (25$^{th}$ percentile) |
| **Q3** | Third quartile (75$^{th}$ percentile) |
| **RESP** | Respiratory rate |
| **RF** | Random forest |
| **SASICU** | Smart and silent ICU |
| **SDC** | Service-oriented device connectivity |
| **$SpO_2$** | Oxygen saturation |
| **$TV_i$** | Inspiratory tidal volume |
| **XGBoost** | eXtreme Gradient Boosting |

# AI Disclosure

Generative artificial intelligence (AI) was used in this thesis only to improve the clarity and structure of the writing and to support Python code generation. It was not employed for research design, analysis, interpretation or the development of arguments. These were independently conceived and written by the author.

# 1

# Introduction

# 1 Introduction

Paediatric intensive care units (PICUs) are dynamic and high-pressure clinical environments where continuous patient monitoring and intensive care are essential. Children admitted to these units are critically ill and at risk of acute, life-threatening conditions, necessitating constant observation and life support and treatment. (1)

To facilitate timely clinical assessment and intervention, bedside monitors are employed to continuously track vital parameters such as heart rate (HR), blood pressure (BP) and oxygen saturation ($SpO_2$). (1) These monitors are designed to generate both audible and visual alarms when parameters deviate from predefined thresholds, irrespective of the signal quality or the cause of deviation. (2, 3) Continuous monitoring is widely adopted in critical care settings, as it has been shown to significantly enhance patient safety by ensuring that healthcare professionals are immediately alerted to physiologic changes, enabling prompt and effective response to important deterioration events. (1, 2, 4-8)

In order to minimise the probability of monitors missing indications of patient deterioration, alarm algorithms and default parameters are frequently configured to maximise sensitivity, often at the expense of specificity. (3) Consequently, this results in a high number of nonactionable alarms being generated by monitors. Previous studies have reported that 87–97% of alarms in PICU settings are nonactionable. (9-13)

The responsibility for distinguishing between true, actionable alarms and false, nonactionable alarms is typically delegated to clinicians, most often nurses. In the majority of clinical settings, nurses are required to continuously assess whether a response is required. (7) Excessive noise, including frequent alarms, has been shown to increase stress among staff, impairing concentration and potentially compromising the care of critically ill patients. (14) Frequent exposure to nonactionable alarms can lead to alarm fatigue, a state of desensitisation and diminished responsiveness, which has been internationally recognised as a major patient safety concern. (15)

Clinical alarms not only impact healthcare providers but also affect patients and their families. In the PICU, alarm-related noise is considered one of the most disruptive environmental factors, contributing to sleep disturbance and increased anxiety. (1, 16, 17) This disruption can hinder patient recovery, as noise exposure has been linked to various negative physiological responses, including elevated HRs, increased respiratory rates (RESP), and sleep deprivation. (18, 19) Furthermore, continuous monitoring of patients may capture regular physiological fluctuations, often resulting in unnecessary diagnostic workups and interventions, which may contribute to longer hospital stays. (20)

Given the impact of alarm fatigue, primarily driven by the high prevalence of nonactionable alarms, on both clinicians and patient outcomes, addressing this issue within the PICU of the Erasmus MC Sophia Children's Hospital has become a priority. Accordingly, the literature review concluded that the development of a machine learning algorithm constitutes an appropriate next step in reducing nonactionable alarms in the PICU. (21)

## 1.1 Objectives

The primary objective of this thesis is to develop a machine learning algorithm capable of classifying actionable alarms, with the aim of reducing the frequency of nonactionable alarms by utilising multimodal vital sign data. It is important to emphasise that the goal is not to predict the occurrence of alarms, but rather to classify alarms at the moment they are generated as either actionable or nonactionable.

The secondary objectives are: firstly, to characterise the current alarm burden in the PICU; secondly, to conduct semi-structured interviews with relevant stakeholders in order to explore their experiences with alarm systems, the impact of alarms and their perspectives on potential interventions for alarm reduction; and thirdly, to evaluate the feasibility of the proposed machine learning model and to outline a strategy for its integration into clinical practice.

# 2

# Background

# 2 Background

Within the PICU of Erasmus MC Sophia Children's Hospital, the high prevalence of nonactionable alarms has created an urgent need to mitigate alarm fatigue. This chapter provides an overview of the alarm environment and outlines approaches for addressing this challenge. Section 2.1 describes the sources of alarms in the PICU, Section 2.2 outlines the classification of nonactionable alarms, Section 2.3 examines the recognition of alarm fatigue as a significant patient safety concern and Section 2.4 reviews existing interventions aimed at reducing clinical alarms. Finally, Sections 2.5 and 2.6 consider the current and future alarm environments, with the latter addressing the proposed integration of a machine learning model into the alarm pathway.

## 2.1 Sources of Alarms in the PICU

In addition to physiological monitoring, numerous medical devices commonly utilised in PICUs, such as ventilators, extracorporeal membrane oxygenation (ECMO) systems, intravenous infusion pumps, feeding pumps and hospital beds, possess their own integrated alarm systems. These devices may indicate a range of issues, including technical malfunction, sensor disconnection and treatment interruption, further increasing the number of alarms in the clinical environment. (2, 22)

## 2.2 Classification of Nonactionable Alarms

Nonactionable alarms are defined as both false alarms, which do not reflect the patient's true status, and nuisance alarms, which reflect the true patient status but do not require clinical attention or intervention. (1, 23) False alarms may arise from various factors, including motion artefacts and technical or equipment-related issues. (1) In contrast, nuisance alarms are characterised by deviations that are clinically insignificant. For instance, a patient may experience a desaturation event, falling just below the minimum acceptable pulse oximeter level for a brief period and subsequently return to the established parameters without intervention. (7) In addition, redundant alarms represent a specific subset of nuisance alarms. While these alarms are technically accurate, they appear to duplicate information already communicated by other monitoring devices or alarm sources. For instance, a bedside monitor and a ventilator may both generate alarms in response to the same desaturation event. Despite their clinical validity, these alarms offer no additional information and typically do not necessitate further clinical intervention. (24)

## 2.3 Recognition of Alarm Fatigue as a Patient Safety Concern

The phenomenon of alarm fatigue is supported by the findings of Bonafide et al.(9), who demonstrated that nurses exposed to frequent nonactionable alarms exhibit slower response times to subsequent alarms. Alarm fatigue delays recognition of critical changes in patient status, thereby compromising safety and increasing the likelihood of missed alarms.(9, 10, 14, 15, 25-29)

Alarm fatigue has been a major patient safety concern for the Joint Commission in the United States since 2013, when a report documented 80 alarm-related deaths between 2009 and 2012. (30, 31) In response, the 2014 Joint Commission National Patient Safety Goal required hospitals to enhance alarm system safety by developing policies and procedures to mitigate alarm fatigue. (32) Since then, clinical alarm safety has remained a key priority, consistently appearing in the Joint Commission's National Patient Safety Goals and frequently ranking among the 'Top 10 Health Technology Hazards', thereby highlighting the ongoing challenges in effectively addressing this issue. (33, 34). In the Netherlands, the Dutch Society for Medical Physics (NVKF) has similarly identified alarm fatigue as a serious alarm-related risk, as outlined in its guideline *Leidraad medische bewakings- en alarmeringssystemen*, which highlights the considerable challenges healthcare providers face due to the high prevalence of clinically irrelevant alarms. (35)

## 2.4 Interventions to Reduce Clinical Alarms

A previous literature review (21) examined interventions aimed at reducing clinical alarms in paediatric hospitals, as well as relevant practices from the process industry, to identify potential lessons for the PICU context. The review identified several effective interventions, including the optimisation of data-driven thresholds, adjustment of graduated time delays, daily review of alarm parameters by alarm dashboards, daily electrode replacement with proper skin preparation and the implementation of machine learning algorithms. A multimodal approach is recommended over single-parameter models. This aligns with strategies commonly employed in the process industry, such as multimodal integration, alarm grouping and correlation analysis. Based on these findings, the review concluded with a recommendation to develop a machine learning algorithm aimed at reducing nonactionable using multimodal vital sign data.

If this approach is found to be impracticable, attention will shift to the simpler intervention, such as the adjustment of alarm thresholds and the implementation of alarm delays. The reduction of ambient noise levels will not be considered as an intervention within the scope of this research. Although this strategy is frequently reported in the literature and may contribute to an improved ward environment, it addresses the impact of alarms rather than its underlying causes.

## 2.5 Overview of the Existing Alarm Environment

The department comprises four distinct units, each with a specific clinical focus. The first unit accommodates patients requiring short-term admission and primarily functions as a high-dependency care unit. The second unit is designated for cardiothoracic surgical patients, and the third unit admits non-cardiothoracic surgical patients as well as those requiring neuromonitoring; both units provide care for critically ill patients. The fourth unit, also a high-dependency care unit, is intended for patients requiring long-term admission; although these patients are not acutely ill, they nonetheless require substantial care, for example, due to chronic illness.

Each unit consists of a six-bed ward and two single-patient rooms. Every bed is equipped with a Dräger Infinity® Acute Care System (M540 and Medical Cockpit; Dräger, Lübeck, Germany; Figure 1), which continuously displays the patient's vital signs, waveforms and alarms generated by threshold deviations or artefacts. A central nursing station is located in the middle of each ward and includes a Dräger Infinity® CentralStation (Figure 1) providing real-time access to monitoring data and active alarm notifications for all patients in the unit. A schematic floor plan is shown in Figure 2.

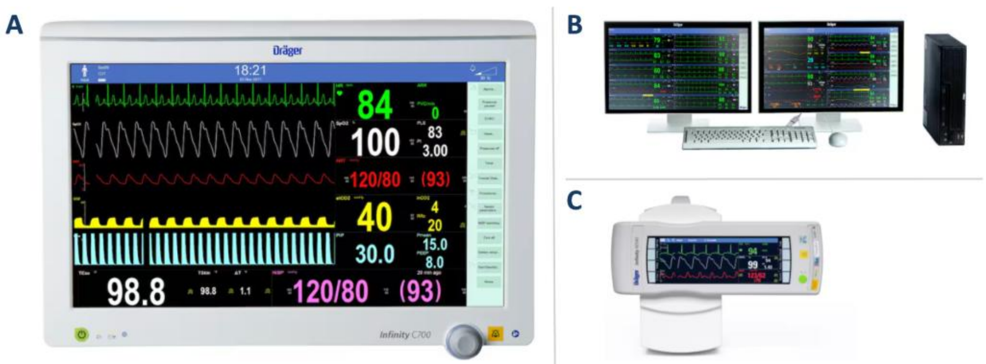



**Figure 1.** Dräger monitoring system components: A. Dräger Infinity® Acute Care System Medical Cockpit (bedside monitor) (36); B. Dräger Infinity® CentralStation (central nursing workstation) (37); C. Dräger Infinity® Acute Care System M540 (portable monitor) (38).

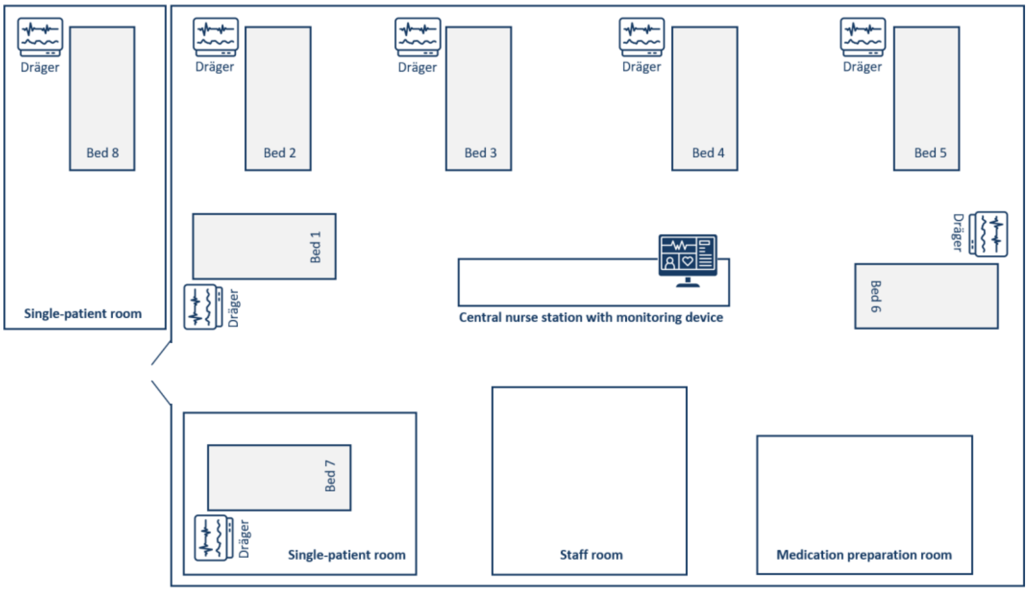



**Figure 2.** A schematic floor plan of a PICU unit at Erasmus MC Sophia Children's Hospital.

Each alarm condition is assigned one of three priority levels: high (life-threatening), medium (serious) or low (informational). (39) The corresponding priority level is conveyed through both visual and auditory alarm signals:

- **High-priority alarms** are typically associated with physiological conditions that may be life-threatening and require immediate clinical intervention (e.g. asystole). High-priority alarms are indicated by a **red** signal.
- **Medium-priority alarms** also pertain to physiological conditions. Although they require prompt attention, they are not considered immediately life-threatening. For instance, an alarm may be initiated when the RESP exceeds a predefined threshold. These alarms are indicated by a **yellow** signal.
- **Low-priority alarms** most often relate to technical issues that could compromise the system's ca to monitor the patient effectively. A typical example is the disconnection of electrodes. Depending on the model of the Dräger monitor, these alarms may either lack a visual signal or be displayed in **turquoise**.

In addition to visual differentiation, each priority level is characterised by a distinct acoustic pattern. Higher priority levels are associated with greater alarm volume, an increased number of audible tones and a higher repetition frequency. In contrast to high-priority alarms, medium- and low-priority alarms can often be reduced through human factors interventions or system-level improvements. Such interventions may include the customisation of alarm limits or the implementation of intelligent algorithms capable of prioritising or suppressing alarms. (40)

## 2.6 Future Alarm Environment and Technological Integration

Erasmus MC Sophia Children's Hospital is scheduled to relocate to a new facility in which the PICU will transition from multi-patient wards to single-patient rooms. Previous research by Kalden et al. (41) has demonstrated that both the physical care environment and the method of alarm transmission contribute to alarm burden. In wards, nursing staff must maintain constant visual and auditory contact with patients, which can constrain workflow and increase their exposure to alarms intended for other colleagues. The transition to single-patient rooms is expected to enhance acoustic conditions, patient privacy and infection control. However, this change may also reduce nurses' bedside presence, potentially compromising their situational awareness. (41-44) To address this, handheld devices will be introduced to support safe alarm management. These devices have the capability to modify alarm pathways, thereby enabling the implementation of effective alarm management interventions aimed at reducing the number of nonactionable alarms.

Erasmus MC has developed a conceptual model describing the anticipated future configuration of the clinical alarm chain. As illustrated in Figure 3 , a machine learning algorithm, or an alternative alarm processing method, could be integrated into the alarm pathway connecting the Dräger monitor to handheld devices. A machine learning algorithm, as an application of artificial intelligence (AI), may be embedded within the alarm processing component of the Intelligent Alarm Services.

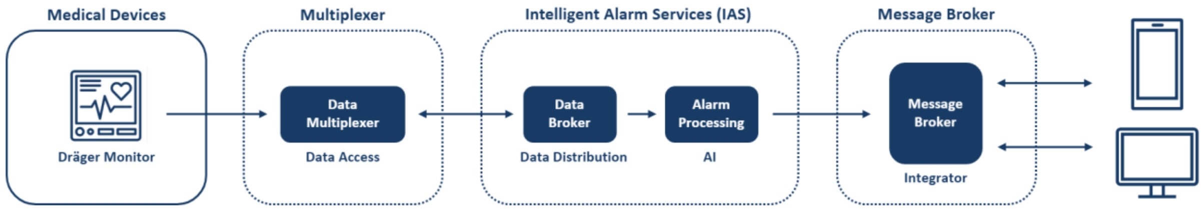



**Figure 3.** Conceptual model of the anticipated clinical alarm chain, showing the flow of alarm data from the Dräger monitor to handheld and desktop devices and illustrating how a machine learning algorithm (AI) could be integrated into the alarm processing pathway.

To implement the proposed alarm management strategy at the future PICU of Erasmus MC, service-oriented device connectivity (SDC) will provide interoperability between medical devices and handheld devices. As part of the IEEE 11073 standard, SDC will facilitate a silent ICU by enabling seamless data exchange among medical devices from different manufacturers. (35, 45) This connectivity allows related alarms to be aggregated and forwarded to handheld devices. Nurses can acknowledge or reject alarms, and their responses are relayed to the monitor, creating a closed-loop system. When a stable connection exists between the remote alarm notifiers and the medical devices, audible bedside alerts remain silent and only visual indicators appear on the display, while audio signals sound where caregivers are located. To ensure safety, the system exchanges bidirectional confirmations between the medical devices and remote notifiers; if these fail, the devices automatically resume both audible and visual alarms. This capability supports a quieter and safer PICU environment.

# 3

# Methods

# 3 Methods

This study employed a three-part methodology:

- Retrospective alarm data were analysed to characterise the current alarm burden in the PICU (Section 3.2);
- A supervised machine learning algorithm was developed to classify actionable alarms using multimodal vital sign data (Section 3.3);
- Semi-structured interviews with clinical stakeholders were conducted to explore alarm management challenges and their perspectives on strategies for alarm reduction (Section 3.4).

All quantitative analyses were performed using Python version 3.12.9. Ethical approval for this study was granted by the Erasmus Medical Ethics Committee (MEC-2021-0937), with a waiver of consent.

## 3.1 Data Acquisition

The database used in this study to characterise the alarm environment (Section 3.2) and develop the machine learning model (Section 3.3) consists of retrospective bedside data collected from Dräger monitors (Dräger, Lübeck, Germany). It includes physiological monitoring data for all vital parameters recorded during each patient admission, associated alarm settings (including parameter-specific threshold changes) and the alarm events generated by the Dräger monitors. The monitoring data were stored at a sampling frequency of 1 Hz. The database comprises records from 2,582 patients admitted to the PICU at Erasmus MC Sophia Children's Hospital between November 2021 and October 2024.

## 3.2 Problem Definition

To characterise the clinical alarm problem in the PICU, alarm events generated by Dräger monitors from 2,582 patient admissions were analysed. The analysis focused on several aspects of the current alarm burden. First, the proportions of high-, medium-, and low-priority alarms were calculated. Second, the most frequently observed alarm parameters and alarm messages were identified. Third, the median number of alarms per patient-day was determined for each PICU unit to enable comparisons between units. Finally, the median hourly distribution of alarms was assessed both by unit and by priority level to identify temporal patterns in alarm frequency and the types of alarms most frequently occurring at different times of day. This approach provided a comprehensive characterisation of the overall alarm burden in the PICU.

## 3.3 Machine Learning Model

This section outlines the development of the machine learning algorithm designed to classify actionable alarms using multimodal vital sign data. An overview of the development process is presented in Figure 4. This overall procedure comprised three main phases: Data Handling, Model Building and Model Assessment. The components of these phases are described in the following subsections.

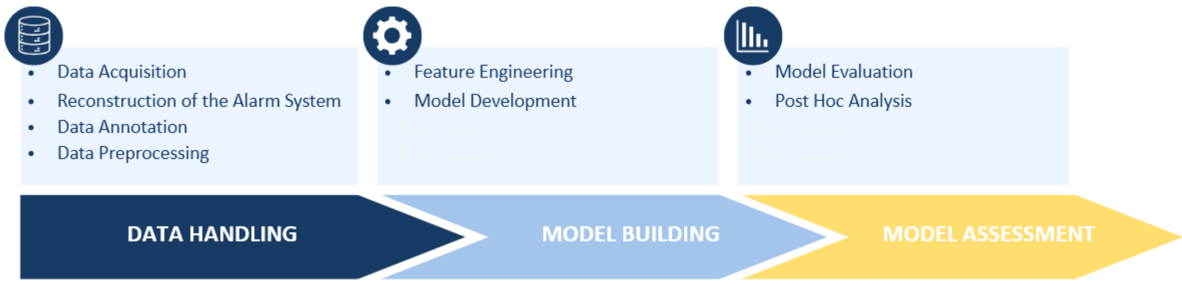



**Figure 4.** Methodological pipeline from data acquisition to post hoc analysis.

### 3.3.1 Data Acquisition

For the development of the machine learning model, data from the mean invasive arterial blood pressure (ART M), HR, RESP and $SpO_2$ were used. Model development employed data from a randomly selected subset of the 2,582 PICU patients, including only those for whom all four physiological parameters were recorded during admission.

Gender, age, reason for admission and severity of illness were recorded to characterise the included patients. Severity of illness was considered relevant because patients with lower severity are generally more awake and active, which may increase movement artefacts and, consequently, the occurrence of nonactionable alarms.

Illness severity was quantified using the paediatric risk of mortality III (PRISM III) and the paediatric index of mortality 3 (PIM 3) scores. Both scores estimate the risk of mortality: PRISM III is derived from the most abnormal physiological and laboratory values recorded during the first 12-24 hours following admission to the PICU, whereas PIM 3 is based on a smaller set of physiological variables and diagnostic information collected within the first hour of admission. (46) The PRISM III score was available from the electronic patient record, while the PIM 3 score was calculated using the official calculator provided by the European Society of Paediatric and Neonatal Intensive Care (47).

### 3.3.2 Reconstruction of the Alarm System

The alarm system was reconstructed based on physiological monitoring data from four parameters: ART M, HR, RESP and $SpO_2$. This reconstruction incorporated parameter-specific threshold settings that varied over time, together with configuration details from *the Dräger Infinity Acute Care System User Manual* (39). Reconstruction was required to correct inconsistencies in time synchronisation between stored alarm events and the monitoring data.

#### 3.3.2.1 Detection of Alarm Events

Alarm events were identified through two approaches: firstly, by comparing signal values with dynamic threshold ranges; and secondly, by interpreting categorical flags embedded within the dataset. The threshold-based alarms were configured per patient and per parameter to determine whether a signal exceeded either the upper or lower predefined range, resulting in an alarm classified as either *HIGH* or *LOW*. In parallel, categorical alarm flags (***, +++, --- and APN) were treated as qualitative indicators of non-numeric alarm status, such as artefacts or clinical conditions. These were mapped to corresponding alarm types based on definitions from the *Dräger Infinity Acute Care System User Manual* (39), with the classifications summarised in Table 1.

To minimise redundant detection of alarms occurring in rapid succession, multiple alarms detected within a 1-second interval were considered as a single alarm episode. It is important to note that the sampling frequency of the Dräger system is higher than the sampling frequency of 1 Hz at which data could be obtained for this study. As a result, the Dräger system is likely to detect a greater number of alarms. However, this has no implications for the present study, as the reconstructed system is used as the reference standard.

| Categorical Flags | Description | Alarm Type |
|---|---|---|
| *** | No parameter values are available, potentially due to patient movement, poor electrode contact, expired electrodes or a poor signal-to-noise ratio caused by external equipment | ARTEFACT |
| +++ | The measured parameter value exceeds the upper limit of the monitor's measurement range | OUT OF RANGE HIGH |
| --- | The measured parameter value falls below the lower limit of the monitor's measurement range | OUT OF RANGE LOW |
| APN | An apnoeic event (cessation of breathing) has been detected | APNOEA |

**Table 1.** Overview of categorical flags and corresponding alarm types used in the reconstructed alarm system.

#### 3.3.2.2 Implementation of Alarm Delays

In the Dräger monitoring system, an alarm validation function, also referred to an alarm delay, is implemented. This mechanism ensures that alarm signals are only generated if the alarm condition persists for a specified duration. This function helps suppress false alarms caused by transient or self-resolving threshold violations.

Although the *Dräger Infinity Acute Care System User Manual* (39) specifies certain alarm delay times, it remained unclear whether these delays were active in the Dräger monitors used in the PICU at Erasmus MC Sophia Children's Hospital. The delay times listed in the manual are presented in Table 2. To investigate this, a comparison was conducted between alarms generated by the reconstructed system and those produced by the Dräger system in the PICU. The temporal differences between corresponding alarm onset times were used to estimate the actual delay times. These estimated delays were then incorporated into the reconstructed system to ensure its behaviour more accurately reflected that of the Dräger monitoring system.

| **Parameter** | **Lower Limit (s)** | **Upper Limit (s)** |
|---|---|---|
| ART M | 4 | 10 |
| HR | 6 | 6 |
| RESP | 14 | 14 |
| $SpO_2$ | 10 | 6 |

*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Table 2.** Alarm delay times specified in the *Dräger Infinity Acute Care System User Manual* (39) for the vital parameters of this study.

### 3.3.3 Data Annotation

To establish a gold standard for training and evaluating the machine learning algorithm, all alarm events were annotated as actionable or nonactionable. The annotation strategy was developed in consultation with two clinicians from the PICU and is summarised in Table 3. An alarm was considered actionable if it was followed by a clinical intervention or reflected clinically relevant trends in monitoring data.

Annotations were primarily based on clinical interventions and, as a first step, were applied to ART M and $SpO_2$ alarms. Informed by previous work (48-50), an alarm was classified as actionable if a change in intervention settings occurred within 30 minutes of the alarm.

For RESP alarms, clinicians noted that actionable alarms did not occur in isolation but were typically accompanied by alarms from other parameters. Accordingly, a RESP alarm was classified as actionable if another alarm occurred within one minute before or after it.

Van Kekem et al. (48) further demonstrated that actionable alarms were typically associated with significant decreases in $SpO_2$, whereas nonactionable alarms were frequently attributable to transient artefacts. On this basis, alarms were classified as nonactionable under three conditions: when the percentage deviation from the specified limit was less than 2%, when they did not meet the criteria for actionability or when they were initiated by categorical flags such as *OUT OF RANGE HIGH*, *OUT OF RANGE LOW* or *ARTEFACT*. These categorical alarms were considered nonactionable as they did not require immediate clinical intervention.

Finally, HR alarms were excluded from annotation in this study. A complete overview of the criteria used for each physiological parameter is presented in Table 3.

| Parameter | Interventions | Actionable Alarm | Nonactionable Alarm |
|---|---|---|---|
| ART M | • Noradrenaline<br>• Adrenaline<br>• Dobutamine<br>• Ringer's lactate<br>• NaCl | • An intervention occurring during the alarm period<br>• An intervention occurring within 30 minutes after an alarm | • Does not meet the criteria for classification as an actionable alarm<br>• Percentage deviation from the limit is less than 2%<br>• Categorical flags (OUT OF RANGE HIGH, OUT OF RANGE LOW, ARTEFACT) |
| $SpO_2$ | • $FiO_2$<br>• PEEP<br>• Tvi | • An intervention occurring during the alarm period<br>• An intervention occurring within 30 minutes after an alarm | • Does not meet the criteria for classification as an actionable alarm<br>• Percentage deviation from the limit is less than 2%<br>• Categorical flags (OUT OF RANGE HIGH, OUT OF RANGE LOW, ARTEFACT) |
| HR | - | - | - |
| RESP | - | • An alarm from another parameter occurring within 1 minute before or after the alarm in question | • Does not meet the criteria for classification as an actionable alarm<br>• Percentage deviation from the limit is less than 2%<br>• Categorical flags (OUT OF RANGE HIGH, OUT OF RANGE LOW, ARTEFACT) |

*ART M = mean invasive arterial blood pressure, $FiO_2$ = fraction of inspired oxygen, HR = heart rate, NaCl = normal saline, PEEP = positive end-expiratory pressure, RESP = respiratory rate, $SpO_2$ = oxygen saturation, Tvi = inspiratory tidal volume*

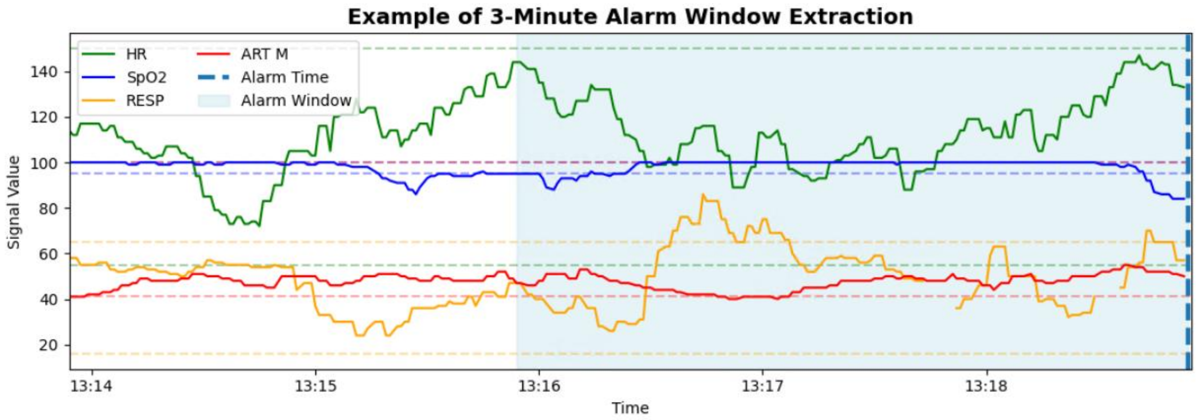
**Table 3.** Criteria for annotating actionable and nonactionable alarms for each physiological parameter.

### 3.3.4 Data Preprocessing

The monitoring data were preprocessed prior to model development. In some cases, monitoring may began before the patient was fully connected to all sensors following admission to the PICU, which resulted in empty records. These empty records at the start of the recording, as well as records extending beyond the time of patient discharge, were removed. Categorical flags (***, +++, --- and APN) were converted to NaN values and binary indicators were added to denote their presence.

As the Dräger system had already applied a certain degree of filtering, although the methodology was not documented, no additional filtering was performed. Data were retained to reflect clinical reality as closely as possible, using values obtained directly from the monitor without further manipulation.

After preprocessing, alarm data were merged with the monitoring data. For each alarm, a defined time window was extracted from the period preceding the event. A 3-minute window was used as the baseline and additional windows of 0.5, 1 and 5 minutes were evaluated to assess their impact on model performance. For illustration, Figure 5 shows Dräger monitoring signals aligned to an alarm event, with a 3-minute pre-alarm window indicated.




*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 5.** Illustration of Dräger monitoring signals aligned to an alarm event (dashed line). The shaded region indicates the 3-minute pre-alarm window used for feature extraction.

### 3.3.5 Feature Engineering

For model development, features were computed within predefined time windows to capture trends and variations in physiological parameters. In collaboration with two clinical experts, the extracted features included the median, interquartile range (IQR), variability, slope and cross-correlation between parameter pairs. Counts and proportions of categorical flags were also extracted. Feature calculation was performed for each parameter and detailed descriptions are provided in Appendix I. When more than 50% of values in a window were missing (NaN), the corresponding feature was set to NaN. Features containing NaN values were removed, as several of the classifiers evaluated in this study cannot handle missing values.

Feature extraction was based on a fixed pre-alarm time window to enable near real-time classification at alarm onset. Features were computed once at the moment the alarm was initiated, allowing rapid classification while avoiding computational overhead of sliding-window processing. A running window was not required, as the task concerned classification at alarm onset rather than prediction in advance.

To ensure comparability across features, all features were scaled to account for differences in units and magnitudes. Scaling method selection was informed by feature distributions and outlier presence, evaluated using histograms for each feature. *StandardScaler* was applied when features had approximately normal distributions without substantial outliers. *RobustScaler* was applied when distributions deviated from normality or contained significant outliers. The selected scaling method was then applied consistently to all features.

### 3.3.6 Model Development

Four machine learning models, logistic regression (LR), decision tree (DT), random forest (RF) and eXtreme Gradient Boosting (XGBoost), were evaluated for alarm classification using a 3-minute baseline window and compared with additional windows of 0.5, 1 and 5 minutes. These classifiers represent different levels of complexity, ranging from simple, interpretable linear models (LR), through transparent rule-based methods for non-linear relationships (DT), to ensemble approaches (RF, XGBoost). RF combines multiple DTs and is less susceptible to overfitting than individual DTs, while XGBoost extends gradient boosting with additional regularisation to improve performance and capture complex patterns. However, RF and XGBoost are less interpretable than LR and DT. (51, 52) Compared with deep learning approaches, these models are generally more computationally efficient and scalable, enabling near-real-time deployment and facilitating clinical implementation.

Model training was performed via nested cross-validation with ten outer folds and five inner folds, stratified and grouped at the patient level to preserve class imbalance and prevent data leakage. Within each outer training set, hyperparameter optimisation was conducted in the inner folds, using Optuna. Optuna adaptively explores the hyperparameter space and prunes unpromising trials, providing a more efficient search than conventional grid or random search methods. (53) The hyperparameter spaces for each machine learning model are provided in Appendix II. All steps in the modelling pipeline, including feature scaling and hyperparameter optimisation, were fitted exclusively on the training data, after which the resulting model was evaluated on the held-out outer test fold. Figure 6 illustrates the nested cross-validation procedure.

To address the expected class imbalance arising from the relatively low number of actionable alarms in the PICU data, class weights were set to *balanced* during model training. In addition, undersampling of the majority class, namely nonactionable alarms, was explored as an alternative strategy.

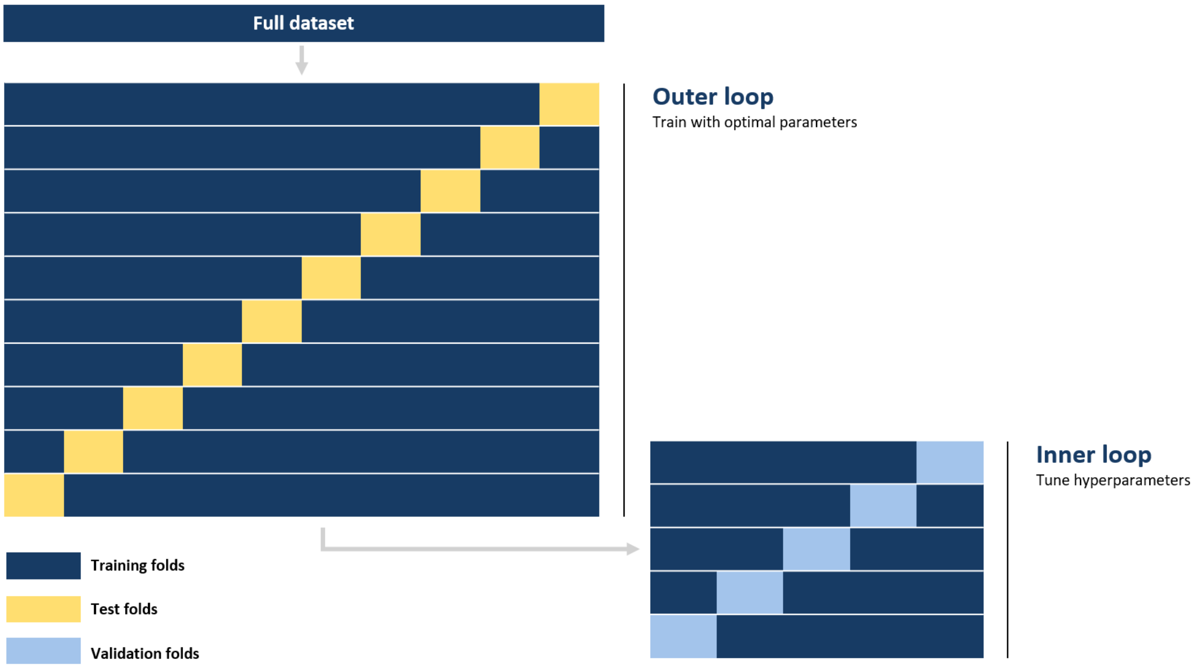



**Figure 6.** Nested cross-validation with a 10-fold split in the outer loop and 5 folds in the inner loop.

### 3.3.7 Model Evaluation

Model performance was assessed using sensitivity, specificity, balanced accuracy, area under the receiver operating characteristic curve (AUROC) and the F1-score. (54) Balanced accuracy was selected as the optimisation target, as it equally weights sensitivity and specificity and is more reliable than overall accuracy in the presence of class imbalance. Hyperparameters for each outer fold were chosen to maximise balanced accuracy in the inner cross validation.

Sensitivity and specificity quantify the ability to correctly classify actionable and nonactionable alarms, respectively. Balanced accuracy, defined as their mean, mitigates the bias of models that achieve high apparent accuracy by favouring the majority class. AUROC was reported as a complementary threshold-independent measure of a model's capacity to discriminate between actionable and nonactionable alarms, with 0.5 indicating random classification and 1.0 perfect separation. The F1-score quantifies the trade-off between actionable alarms (sensitivity) and limiting false positives (precision), though it ignores true negatives and was therefore considered alongside other metrics. As no single metric fully characterises performance, results were assessed across this set of complementary measures. Final values were reported as the median across the ten outer test folds, with the IQR (first quartile (Q1) – third quartile (Q3)) to indicate variability.

Optimisation was performed using balanced accuracy to maintain an appropriate trade-off between sensitivity and specificity, thereby limiting the false positives that contribute to alarm fatigue. In the Results section, however, the best-performing model is presented. This model achieved the highest sensitivity among all evaluated models, reducing the risk of missing critical alarms.

### 3.3.8 Post Hoc Analysis

To gain deeper insight into the contribution of features to model performance and into the ability to distinguish actionable from nonactionable alarms, a post hoc analysis was conducted on the best-performing model.

Feature importance was calculated to evaluate the relative contribution of individual features. Features with high importance rankings were assessed visually using scatterplots. Feature values for actionable and nonactionable alarms were summarised as medians with IQRs (Q1-Q3) and visualised with boxplots. A Pearson correlation matrix was computed to examine relationships between features.

Finally, the speed of alarm classification was assessed to determine whether the model could be applied in near-real time in a clinical setting. Inference time, defined as the wall-clock time required for a trained model to classify new alarms, was measured on each outer test fold. Predictions were repeated five times per fold to reduce noise and per-alarm inference time was calculated by dividing the total prediction time by the number of alarms. For each outer test fold, the median and IQR (Q1-Q3) of per-alarm inference times were obtained. To provide an overall estimate across folds, the median of these fold-level medians, together with the corresponding IQR (Q1-Q3) was reported.

## 3.4 Stakeholder Input

To investigate alarm management issues in the PICU, semi-structured interviews were conducted with four nurses and one psychologist specialising in patient stress within the PICU environment.

The nurse interviews addressed three main areas:

1. Identification of key problems related to alarm management, including the impact of alarms on nurses and the circumstances under which they do or do not respond.
2. Perspectives on alarm prediction, distinctions between actionable and nonactionable alarms and the potential use of machine learning algorithms to detect actionable alarms.
3. Recommendations for strategies to reduce the overall burden in the PICU.

The interview with the psychologist focused on the impact of alarms on patients’ stress and overall experiences within the PICU environment.

# 4

# Results

# 4 Results

This chapter presents the main findings of the study, based on the three methodological components:

- Descriptive analyses of the PICU alarm data from the Erasmus MC Sophia Children's Hospital to characterise the current alarm burden (Section 4.1);
- The development and evaluation of the machine learning model (Section 4.2);
- Perspectives from clinical stakeholders on current alarm management practices, challenges and potential strategies for reducing the impact of clinical alarms (Section 4.3).

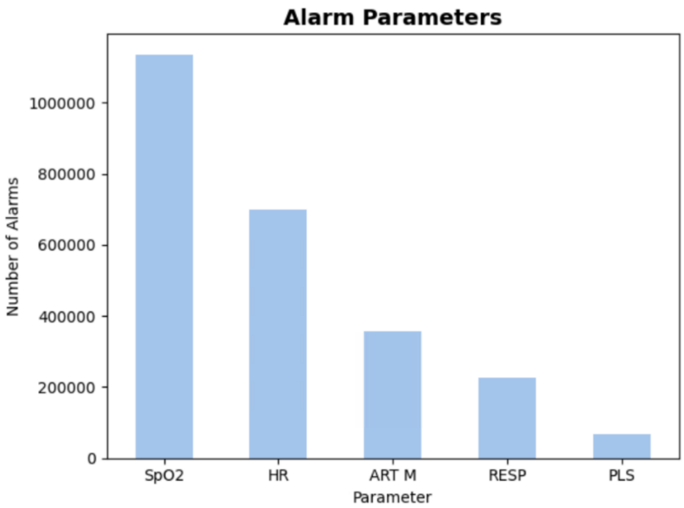
## 4.1 Problem Definition

### 4.1.1 Distribution of Alarms by Priority Level

Alarm data from a randomly selected cohort of 66 patients were analysed to determine the proportion of high-, medium- and low-priority alarms. Among all alarms, 10.8% were classified as low priority, 83.0% as medium and 6.2% as high.

### 4.1.2 Distribution of Alarm Parameters

In the complete dataset, $SpO_2$ alarms accounted for the largest proportion of clinical alarms. Alarms related to HR, ART M and RESP were also frequent. Figure 7 presents the number of alarms for the five most frequent parameters.

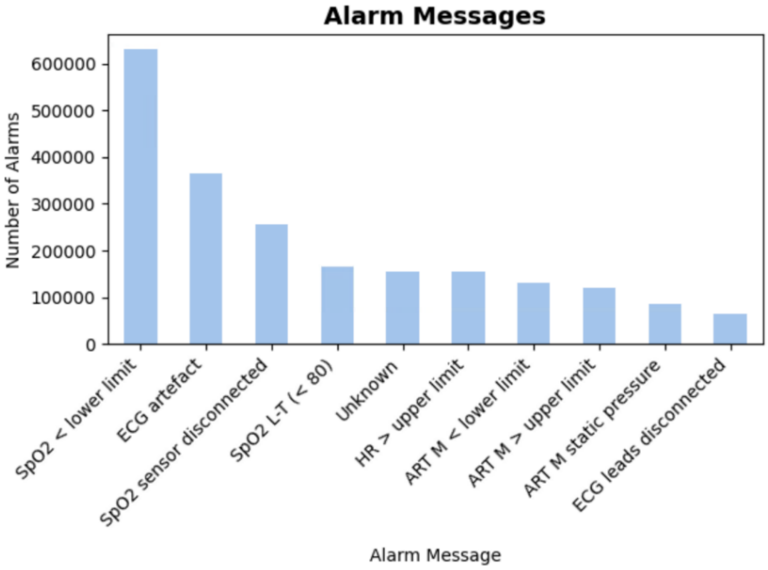



*ART M = mean invasive arterial blood pressure, HR = heart rate, PLS = pulse rate derived from pulse oximetry ($SpO_2$), RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 7.** Top five alarm parameters in the PICU at Erasmus MC Sophia Children's Hospital, November 2021 – October 2024 (n = 2,582 patients).

### 4.1.3 Distribution of Alarm Messages

Figure 8 presents the distribution of alarms by message, in contrast to Figure 7, which categorises alarms by physiological parameters. The most frequent alarm message was *$SpO_2$ < lower limit.* Technical alarm messages, including *ECG artefact* and *$SpO_2$ sensor disconnected,* were the second and third most frequent categories. Notably, 90.3% of the alarms labelled *Unknown,* indicating missing alarm messages, were attributable to *ARTEFACT* (***) alarms.

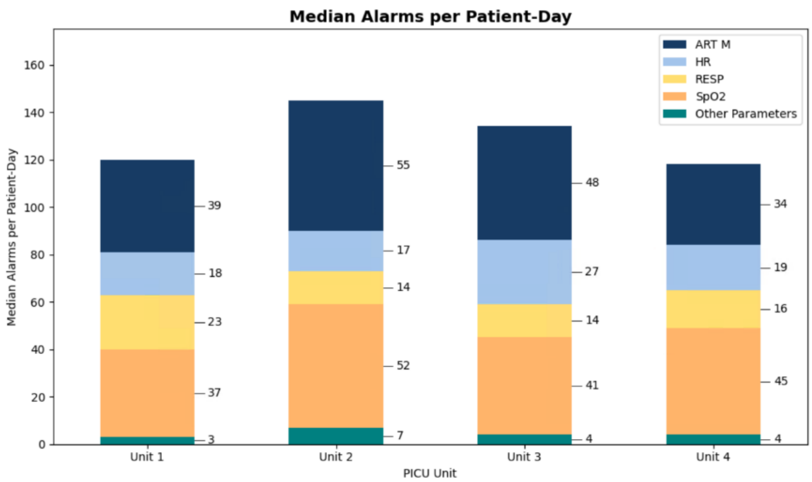



*ART M = mean invasive arterial blood pressure , ECG = electrocardiogram, HR = heart rate, $SpO_2$ = oxygen saturation*

**Figure 8.** Top ten alarm message types in the PICU at Erasmus MC Sophia Children's Hospital, November 2021 – October 2024 (n = 2,582 patients).

### 4.1.4 Distribution of Alarms per PICU unit

Figure 9 presents the median number of alarms per patient-day across the PICU units. Units 2 and 3 exhibited a higher median number of alarms per patient-day compared with units 1 and 4. Unit 2 showed a higher frequency of ART M alarms. Unit 4 showed a relatively high number of $SpO_2$ alarms. The distributions of all alarm parameters by unit are also shown in Figure 9.




*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 9.** Median number of alarms per patient-day for each PICU unit at Erasmus MC Sophia Children's Hospital, November 2021 – October 2024 (n = 2,582 patients).

### 4.1.5 Distribution of Alarms over Time

Figure 10 shows the median hourly distribution of alarms. A peak in alarm occurrence was observed at approximately 09:00, with a similar pattern observed for medium-priority alarms (Figure 11). Decreases in alarm frequency were observed around 13:00 and 19:00. The median and IQR of the hourly distribution of alarms, differentiated by PICU unit and by alarm priority, are presented in Appendix III.

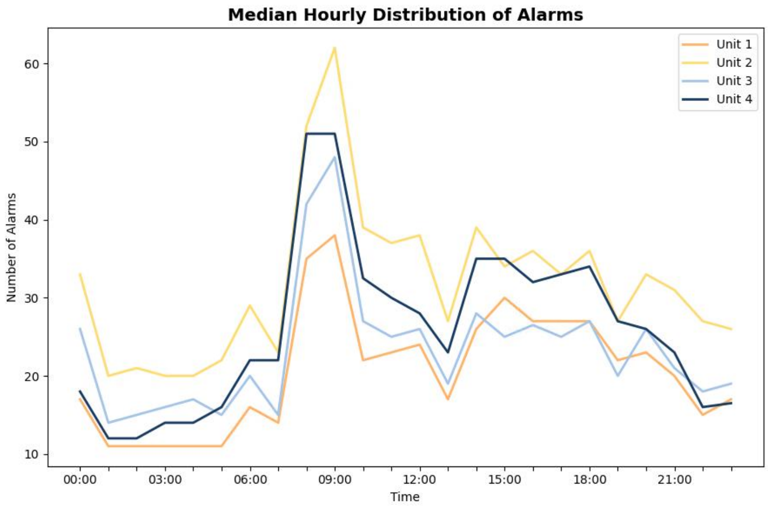



**Figure 10.** Median hourly distribution of alarms across the four PICU units at Erasmus MC Sophia Children's Hospital, November 2021 – October 2024 (n = 2,582 patients).

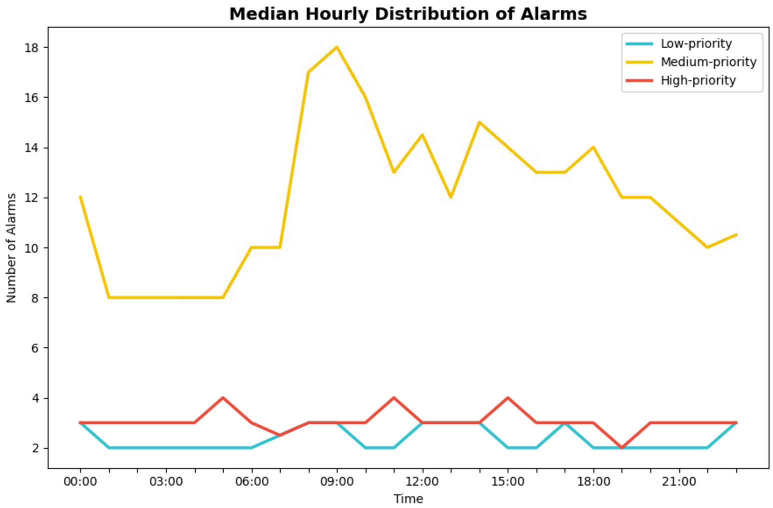



**Figure 11.** Median hourly distribution of alarms by priority level across all PICU units at Erasmus MC Sophia Children's Hospital, November 2021 – October 2024 (n = 66 patients).

## 4.2 Machine Learning Model

### 4.2.1 Research Population

A total of 26 patients were included in model development. The median age was 1.31 years. Of these, 14 (53.8%) were male and 12 (46.2%) were female. Reasons for admission and illness severity scores are summarised in Table 4.

| **Characteristic** | **Dataset (N = 26)** |
|---|---|
| Gender, N (%) | |
| Male | 14 (53.8) |
| Female | 12 (46.2) |
| Age (years), median (Q1-Q3) | 1.31 (0.07 – 14.96) |
| Reason for Admission, N (%) | Cardiovascular: 8 (30.8)<br>Craniofacial: 3 (11.5)<br>Critical Illness: 1 (3.8)<br>Gastrointestinal: 1 (3.8)<br>Neurology: 1 (3.8)<br>Orthopaedic: 6 (23.1)<br>Other Surgery: 1 (3.8)<br>Prematurity: 1 (3.8)<br>Respiratory: 2 (7.7)<br>Trauma/Injury: 2 (7.7) |
| PRISM III, median (Q1-Q3) | 15 (15 – 17) |
| PIM 3 (%), median (Q1-Q3) | 0.8 (0.2 – 2.7) |

*PIM 3 = paediatric index of mortality 3, PRISM III = paediatric risk of mortality III*

**Table 4.** Characteristics of patients included in model development.

### 4.2.2 Reconstruction of the Alarm System

Parameter-specific delay times, reported as median with IQR (Q1–Q3), were calculated by comparing reconstructed system alarms with Dräger monitor alarms. The shortest alarm delays, with a median of 1 s, were observed for the categorical flags and for $SpO_2$ L-T (<80%), whereas the longest delay, with a median of 13 s, was observed for the RESP limits. Results for lower and upper limits and categorical flags are presented in Table 5.

| **Parameter** | **Lower Limit (s)** *Median (Q1-Q3)* | **Upper Limit (s)** *Median (Q1-Q3)* | **Categorical Flags (s)** *Median (Q1-Q3)* |
|---|---|---|---|
| ART M | 3 (3 - 4) | 9 (8.5 -9) | 1 (0 - 1) |
| HR | 5 (5 - 6) | 5 (5 - 6) | 1 (0 - 1) |
| RESP | 13 (13 - 14) | 13 (13 - 14) | 1 (0 - 1) |
| $SpO_2$* | 9 (6 - 9) | 5 (4 – 6) | 1 (0 - 1) |

*$SpO_2$ L-T (<80%) has an alarm delay time of 1 s.

*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Table 5.** Estimated alarm delay times (median with IQR (Q1 – Q3) showing the temporal differences between reconstructed system alarms and corresponding Dräger alarms.

### 4.2.3 Alarm Dataset and Features

The dataset initially comprised 26,832 alarm windows across all window lengths. After excluding alarm windows with more than 50% missing values, 19,621 windows of 3 minutes, 14,817 of 0.5 minute, 17,317 of 1 minute and 20,269 of 5 minutes remained. Actionable alarms accounted for 13% of all alarms.

For each alarm window, 48 features were extracted and feature distribution histograms (Appendix IV) showed non-normal distribution with substantial outliers. Therefore, a *RobustScaler* transformation was applied to all features to enhance comparability and limit the influence of extreme values.

### 4.2.4 Model Evaluation

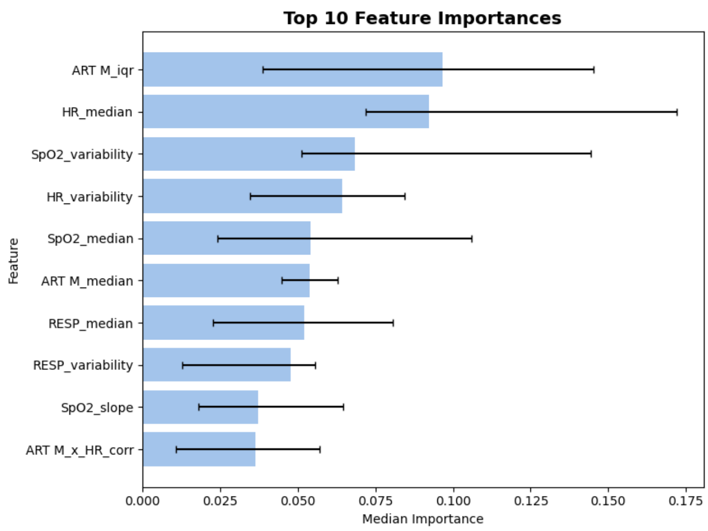
Among the evaluated algorithms, the DT achieved the highest sensitivity, with a median of 0.48 (0.40 – 0.60) at the 0.5-minute window (Table 6). The specificity ranged from 0.51 (0.43 – 0.64) to 0.70 (0.49 – 0.81) across alarm windows, with higher values observed at longer windows. Balanced accuracy and AUROC values remained close to 0.5 across all windows. F1-scores were consistently low (0.19 – 0.24). Results for the other algorithms, as well as the models incorporating undersampling, are presented in Appendix V.

| **Alarm Window** *(min)* | **Sensitivity** *Median (Q1-Q3)* | **Specificity** *Median (Q1-Q3)* | **Balanced Accuracy** *Median (Q1-Q3)* | **AUROC** *Median (Q1-Q3)* | **F1-score** *Median (Q1-Q3)* |
|---|---|---|---|---|---|
| 3 | 0.42 (0.33 - 0.55) | 0.60 (0.55 – 0.70) | 0.53 (0.49 – 0.55) | 0.53 (0.52 – 0.56) | 0.19 (0.10 – 0.27) |
| 0.5 | 0.48 (0.40 – 0.60) | 0.51 (0.43 – 0.64) | 0.53 (0.49 – 0.56) | 0.53 (0.50 – 0.58) | 0.22 (0.11 – 0.28) |
| 1 | 0.40 (0.19 – 0.67) | 0.66 (0.45 – 0.78) | 0.52 (0.48 – 0.53) | 0.55 (0.52 – 0.58) | 0.20 (0.13 – 0.27) |
| 5 | 0.36 (0.29 – 0.48) | 0.70 (0.49 – 0.81) | 0.54 (0.52 – 0.57) | 0.55 (0.54 – 0.58) | 0.24 (0.12 – 0.26) |

**Table 6.** Performance of DT, identified as the best-performing model, across different alarm windows, expressed as median with IQR (Q1 -Q3).

### 4.2.5 Post Hoc Analysis

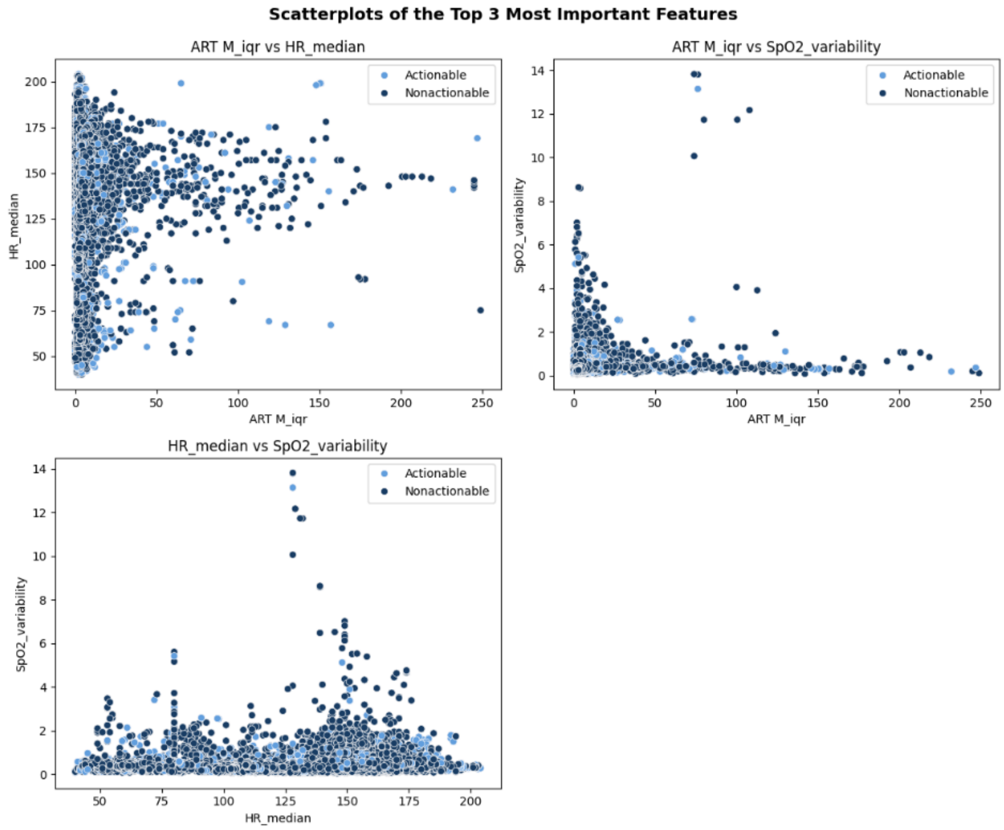
A post hoc analysis was performed on the best-performing model, the DT. Feature importance was examined, with the ten features contributing most to the model shown in Figure 12. The median feature importance did not exceed 0.10 (0.04 – 0.15). The categorical flag features contributed minimally and were therefore not considered further in the subsequent post hoc analysis.




*ART M = mean invasive arterial blood pressure, HR = heart rate, IQR = interquartile range, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 12.** Top 10 feature importances of the DT for the baseline alarm window of 3 minutes, expressed as median with IQR.

To assess whether the five most important features of the DT for the 3-minute alarm window (Figure 12) could discriminate between actionable and nonactionable alarms, scatterplots of feature pairs were generated. Figure 13 presents the scatterplots of the three features with the highest importance ('ART M IQR', 'HR Median' and '$SpO_2$ Variability'). The plots show overlapping distributions of actionable and nonactionable alarms across all three feature pairs. Scatterplots for the remaining combinations are provided in Appendix VI.




*ART M = mean invasive arterial blood pressure, HR = heart rate, IQR = interquartile range, $SpO_2$ = oxygen saturation*

**Figure 13.** Scatterplots of the three most important features of the DT for the 3-minute alarm window. The plots show the relationships between 'ART M IQR', 'HR Median' and '$SpO_2$ Variability', with actionable alarms in light blue and nonactionable alarms in dark blue.

To evaluate potential collinearity between the included features, a correlation matrix was generated (Appendix VI). Overall, the features showed low correlation with each other. The highest correlation was observed between 'RESP Variability' and 'RESP IQR' (r = 0.65).

To investigate potential differences between alarm groups for the individual features, medians and IQRs (Q1 – Q3) for the 3-minute alarm window were calculated for actionable and nonactionable alarms (Table 7). Minimal group differences were observed for several features, including 'ART M Variability', 'HR Variability', 'RESP IQR' and 'RESP Variability', with substantial overlap of the IQRs between the alarm groups. The corresponding boxplot distributions are presented in Appendix VI, which likewise demonstrate minimal inter-group differences and numerous outliers.

Finally, inference time per alarm was assessed. For the DT with the 3-minute alarm window, the median interference time across outer test folds  was 0.002 (0.002 – 0.003) ms per alarm.

| Feature | Actionable Alarms<br>*Median (Q1-Q3)* | Nonactionable Alarms<br>*Median (Q1-Q3)* |
|---|---|---|
| ART M Median | 60 (52 – 69) | 58 (50 – 68) |
| ART M IQR | 3 (2 – 6) | 3 (2 – 4.5) |
| ART M Variability | 0.92 (0.54 – 1.67) | 0.72 (0.47 – 1.25) |
| ART M Slope | 0.005 (- 0.01 – 0.03) | 0.0006 (- 0.01 – 0.02) |
| HR Median | 146 (118 – 161) | 146 (130 – 158) |
| HR IQR | 2 (1 – 5) | 2 (1 – 4) |
| HR Variability | 0.48 (0.25 – 1.09) | 0.44 (0.26 – 0.84) |
| HR Slope | 0.0017 (-0.009 – 0.02) | - 0.00002 (-0.01 – 0.012) |
| $SpO_2$ Median | 94 (79 – 97) | 94 (83 – 98) |
| $SpO_2$ IQR | 1 (1 – 3) | 1 (1 – 2) |
| $SpO_2$ Variability | 0.46 (0.33 – 0.69) | 0.4 (0.3 – 0.54) |
| $SpO_2$ Slope | -0.0027 (-0.015 – 0.0038) | - 0.0006 (- 0.0065 – 0.004) |
| RESP Median | 39 (29 – 48) | 40 (29 – 51) |
| RESP IQR | 11.5 (6 – 16.88) | 10 (4 -15) |
| RESP Variability | 3.23 (2 – 4.58) | 2.77 (1.5 – 4.27) |
| RESP Slope | 0.00034 (-0.046 – 0.049) | -0.00021 (-0.038 – 0.036) |
| ART M x HR Correlation | 0.11 (-0.15 – 0.43) | 0.2 (-0.09 – 0.5) |
| ART M x $SpO_2$ Correlation | -0.03 (0.25 – 0.18) | -0.02 (-0.22 – 0.18) |
| ART M x RESP Correlation | 0.08 (-0.16 – 0.31) | 0.04 (-0.18 – 0.27) |
| HR x $SpO_2$ Correlation | -0.09 (-0.34 – 0.14) | -0.05 (-0.27 – 0.16) |
| HR x RESP Correlation | 0.02 (-0.19 – 0.25) | 0.03 (-0.18 – 0.26) |
| $SpO_2$ x RESP Correlation | 0 (-0.2 - 0.17) | -0.004 (-0.18 – 0.17) |

*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Table 7.** Summary of physiological signal features within 3-minute alarm windows for actionable versus nonactionable alarms. Data are reported as median values with IQR (Q1–Q3).

## 4.3 Stakeholder Input

To investigate alarm management in the PICU, semi-structured interviews were conducted with four nurses and one psychologist specialised in patient stress within the PICU. Detailed findings are provided in Appendix VII, while the principal themes are summarised below.

### 4.3.1 Alarm Management Practices and Challenges

Analysis of the interviews with the four nurses identified the following themes:

- Nurses reported routinely adjusting alarm thresholds at the start of each shift, basing these adjustments on factors such as patient age, underlying pathology and familiarity with the patient's condition.
- Repeated exposure to alarm sounds was described as increasing stress and contributing to desensitisation, which in turn reduced responsiveness and encouraged a tendency to interpret alarms as non-urgent.
- Nurses emphasised that they consistently responded to all high-priority (red) alarms, regardless of whether these originated from their own patients or those under the care of colleagues.
- Duplicate alarms, such as simultaneous alerts from the ventilator and the Dräger monitor, were not generally perceived as problematic.
- Alarms were frequently described as disruptive during specific clinical situations, including clinical interventions and patient movement. Respiratory alarms were regarded as necessary only in cases of respiratory insufficiency, while infusion pump alarms for non-critical medication (e.g. paracetamol) were often considered avoidable distractions.

### 4.3.2 Perspectives on Alarm Prediction

From interviews with two nurses, the following themes emerged:

- Nurses reported that they were often able to anticipate alarms based on information provided during handovers, observations of patient restlessness or excessive movement and during clinical interventions.
- The clinical context of the patient was described as crucial for predicting alarms, with assessments extending beyond physiological signals alone.
- Distinguishing between actionable and nonactionable alarms was considered challenging, as clusters of nonactionable alarms within a short timeframe could still carry clinical significance.
- Limited confidence was expressed in the current potential of machine learning algorithms designed to transmit only actionable alarms to handheld devices. Nonactionable alarms were regarded as potentially useful precursors or early indicators of actionable events.
- Actionable events, such as episodes of oxygen saturation requiring supplemental oxygen, can be documented by nurses in HiX.

### 4.3.3 Strategies for Reducing Alarm Burden

The interviews with the four nurses highlighted several strategies:

- Many of the proposed interventions focused on reducing the impact of alarm noise.
- Nurses emphasised the importance of considering interventions directed towards patients, as they remain continuously exposed to loud alarm sounds.
- A need was identified for a foot pedal or push-button mechanism to silence alarms during clinical interventions.
- Alarm settings were considered ideally adjustable for each individual infusion pump.
- Nurses expressed a preference for the introduction of audiovisual support via handheld devices at the current stage, to allow familiarisation prior to the transition to single-patient rooms.

### 4.3.4 Patient Stress and Experience

The interview with the psychologist yielded the following insights into the impact of alarms on patients and their families:

- For parents, the principal source of stress arises not from the alarms themselves, but from the severity of their child's medical condition.
- Alarms cannot be directly associated with patient recovery or extended hospitalisation, as their frequent occurrence is often a consequence of the patient's deteriorating condition.
- Auditory stimuli resembling alarms from the PICU frequently evoke recollections of parents' experiences on the unit.
- Alarms may also contribute to a perceived sense of safety among parents, as they provide reassurance that their child is subject to continuous monitoring.
- Single-patient rooms are regarded by parents as challenging, as children are no longer continuously visible to nursing staff. However, such rooms exert a positive influence on the child's care by reducing ambient noise levels.
- A parental dashboard has been developed to provide contextual information concerning continuous monitoring and the significance of alarms.

# 5
# Discussion

# 5 Discussion

This study addressed three objectives: (1) to characterise the current alarm burden in the PICU, (2) to develop a machine learning algorithm for classifying actionable alarms using multimodal vital sign data, and (3) to explore clinicians' perspectives on alarm management.

The best-performing model, a DT, showed limited performance, with sensitivities of 0.36-0.49, specificities of 0.51-0.70 and balanced accuracies of approximately 0.50. Post hoc analysis revealed no single feature with dominant importance, minimal differences in feature distributions and a high prevalence of outliers, indicating that actionable and nonactionable alarms cannot be reliably distinguished, thereby addressing Objective 2.

Descriptive analyses demonstrated that most alarms were related to $SpO_2$ desaturations. Alarm frequency varied between units, reflecting differences in patient populations and monitoring practices, and showed distinct temporal patterns linked to ward activity. These findings address Objective 1 and support the transition to single-patient rooms or the implementation of targeted interventions at ward level.

Semi-structured interviews with nurses indicated that frequent alarms contribute to overstimulation and desensitisation, yet even nonactionable alarms are valued as early warnings of deterioration. Moreover, distinguishing actionable from nonactionable alarms was reported to require clinical judgement and contextual information beyond physiological signals, addressing Objective 3.

The following sections situate the findings within the existing literature (Section 5.1), interpret the results in light of the three methodological components (Section 5.2), discuss study limitations (Section 5.3) and provide recommendations for future research and clinical practice (Section 5.4).

## 5.1 Comparison to Literature

The performance of machine learning models is highly dependent on data quality, with accurate annotation representing a critical determinant. Reviewing how other intensive care unit (ICU) studies have addressed annotation in developing machine learning approaches for alarm reduction, as well as Dutch initiatives on alarm-burden reduction, may provide insights relevant to the PICU at Erasmus MC Sophia Children's Hospital.

Several neonatal intensive care unit (NICU) studies have relied on retrospective labelling by professional judgement or annotator consensus, but with limited methodological detail and risk of bias, underscoring the need for prospective, context-supported approaches. (55-57) Other ICU studies have adopted rule-based annotation, with expert-defined scenarios and balanced datasets of true and false alarms. (58, 59) While these approaches simplify model training, they do not reflect clinical reality, where only 3-13% of alarms are actionable, and therefore tend to overestimate performance. Targeted strategies have also been explored. Kalden et al. (60) predicted critical oxygen desaturations based on red alarms, enabling delayed presentation of yellow alarms without compromising safety, though their scope was limited. Caregiver response has likewise been used as a proxy for intervention (61), but is unreliable in open-ward PICUs where silencing may reflect either noise suppression or clinical management.

Taken together, these findings suggest that starting with narrowly defined scenarios, such as desaturations, could represent a practical first step before addressing the broader alarm burden. This is particularly relevant due to the predominance of $SpO_2$ desaturation alarms in this PICU cohort. Moreover, as most prior studies were conducted in the NICU, the present study represents an initial step in developing machine learning approaches tailored to the PICU context.

Dutch initiatives demonstrate complementary strategies for alarm reduction. Within the Smart and Silent ICU (SASICU) project (62), Boelhouwer (63) highlighted challenges in configuring personalised alarm settings and variability among nurses. A configuration system was developed to guide patient-specific alarm settings based on clinical profile and risk level. Varisco et al. (64) implemented an alarm management system in the NICU of

Máxima MC, Eindhoven, combining handheld delivery of waveform alarms with workflow changes such as sensor replacement protocols, alarm delays, pausing alarms during caregiving moments and regular review of policies and alarm settings. These interventions significantly reduced alarms per patient-day without compromising patient safety.

## 5.2 Interpretation of the Results

### 5.2.1 Problem Definition

Alarm analysis showed that $SpO_2$, HR, ART M and RESP generated the highest number of alarms and were therefore selected as model inputs. The most frequent alarm, *$SpO_2$ < lower limit,* included both true desaturations and motion artefacts, while *$SpO_2$ sensor disconnected,* the third most frequent alarm, remained clinically relevant because prompt recognition of sensor detachment is essential for reliable monitoring.

Alarm frequency varied across units. Units 2 and 3 generated more alarms per patient-day than units 1 and 4, reflecting higher illness severity, narrower alarm thresholds and greater need for intensive monitoring. Unit 2 showed more ART M alarms, consistent with its predominance of cardiothoracic surgical patients. Unit 4 demonstrated a high number of $SpO_2$ alarms, likely related to motion artefacts in more active patients.

Temporal analysis revealed a pronounced alarm peak around 09:00 across all units, coinciding with the start of bedside rounds when increased nursing interventions and physician presence heightened patient restlessness. Alarm frequency declined thereafter, with a marked reduction between 12:00 and 14:00 during scheduled rest periods, when lights were dimmed and interventions minimised. Afternoon shift handovers between 15:00 and 17:00 were followed by a gradual decline in ward activity. Evening meal breaks around 19:00 were associated with a further reduction in alarms, after which frequencies stabilised at lower levels into the night.

These findings indicate that alarm frequency is influenced by patient factors, technical issues, monitoring settings and the daily ward routines, supporting the transition to single-patient rooms to minimise ward-related disturbances.

### 5.2.2 Machine Learning Model

The model cohort consisted mainly of younger, elective surgical patients of low illness severity, who may require less intensive monitoring but generate more artefacts and nonactionable alarms.

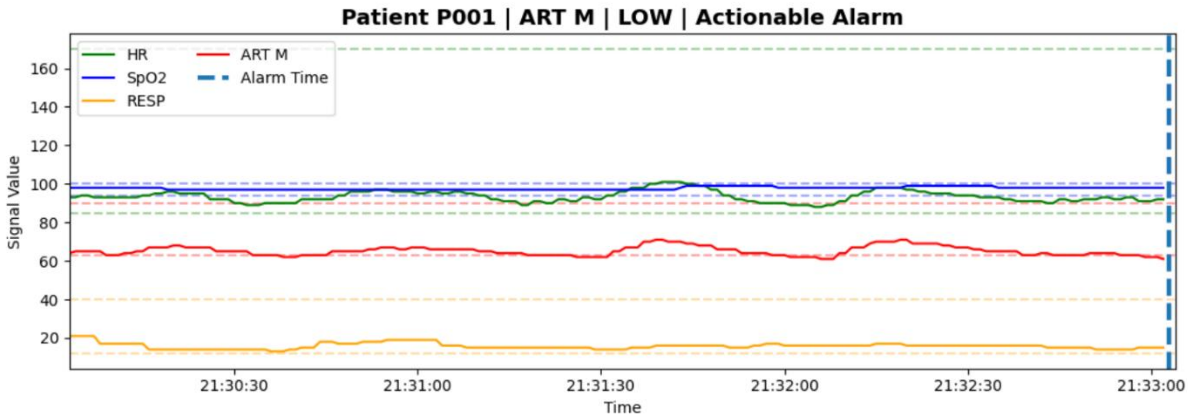
Observed alarm delays in the reconstructed system closely matched values from the *Dräger Infinity Acute Care System User Manual* (39), with minor discrepancies due to rounding. Manual values were therefore adopted, except for categorical flags and $SpO_2$ L-T (< 80%), where calculated delays were retained due to the absence of manual values.

Only 13% of alarms were actionable, creating a pronounced class imbalance that biased models toward the majority class. To preserve fidelity to the original Dräger data, upsampling was not applied. The DT achieved the highest sensitivity (36-48%), while LR and RF showed higher specificity but lower sensitivity. XGBoost yielded the lowest performance, likely due in part to limited hyperparameter tuning, but all models were ultimately constrained by the low discriminative value of the available features. Across all models, sensitivity and specificity were inadequate for clinical application, with AUROC and balanced accuracy near 0.5 and F1-scores of about 20%, indicating near-random performance. Undersampling did not improve performance, as reducing majority-class samples left insufficient data for effective model training. Given this imbalance, some evaluation metrics require cautious interpretation; for instance, the F1-score is prevalence-dependent, as precision is strongly affected when the positive class is rare. (54) Consistent with the methods, performance was therefore assessed across multiple complementary metrics.

Model performance varied with alarm-window length. The DT achieved its highest sensitivity at 0.5 minutes, whereas LR and RF peaked at 5 minutes. These shifts were model-dependent and mainly altered the sensitivity-specificity balance without improving discrimination. Balanced accuracy and AUROC remained near 0.5 across all windows, indicating no benefit from varying this parameter. Therefore, the baseline 3-minute alarm window was adopted for post-hoc analyses.

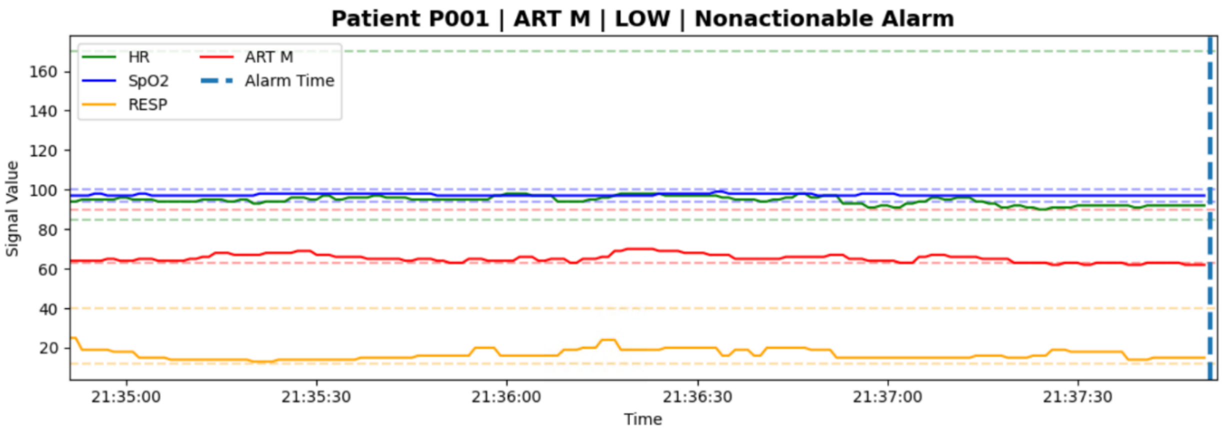
Feature-importance analysis showed uniformly low contributions, minimal influence of categorical flag features, low inter-feature correlations (< 0.7) and extensive overlap between actionable and nonactionable alarms with minimal distributional differences. These findings indicate that the features had limited discriminatory value. Inference time was negligible (0.002 ms), suggesting near-instantaneous classification. Residual latency may be attributable to data transfer, but this was not investigated in the present study.

A likely explanation for poor performance is the retrospective annotation, which depends on clinical context and subjective judgement not captured in the dataset. This limitation is evident in Figures 14 and 15, where actionable and nonactionable alarms show no clear separation, underscoring the uncertainty of labels.




*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 14.** Illustration of an actionable ART M *LOW* alarm for Patient P001. This alarm was classified as actionable, most likely because the ART M signal deviated by more than 2% below the threshold.




*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 15.** Illustration of a nonactionable ART M *LOW* alarm for Patient P001. This annotation may be explained by the fact that the ART M signal remained within 2% below the threshold.

### 5.2.3 Stakeholder Input

Interviews with nursing staff highlighted the need to balance reduced auditory overstimulation with the reliable detection of patient deterioration. Nurses expressed concern that a machine learning algorithm transmitting only actionable alarms could delay recognition of early warning signs. They also emphasised that interpretation of alarms require clinical context, which may explain the model's limited sensitivity and near-random performance of models based solely on physiological signals. Suggested interventions included reducing ambient noise, introducing a foot pedal to silence alarms during interventions and enabling adjustment of infusion-pump alarm settings.

The psychologist noted that alarms contribute to stress primarily through overstimulation but also provide a sense of safety. The transition to single-patient rooms was considered beneficial for creating a calmer environment, and parents were expected to benefit from clear explanations of alarm purposes and reassurance that not every alarm signals an emergency.

Overall, stakeholders preferred interventions that preserve nurses' control over complex algorithms that diminish it. Reflecting this preference for oversight, nurses favoured gradual adoption of handheld devices displaying waveforms and alarms to familiarise themselves with the technology before transitioning to single-patient rooms.

## 5.3 Limitations

This study has several limitations. The main limitation is the retrospective labelling of alarms as actionable if followed by an intervention within 30 minutes. This assumption is of uncertain validity, as it presumes that interventions were always initiated by the alarm, that all actionable alarms prompted responses and that alarms without intervention were nonactionable. In practice, alarms may have been missed due to alarm fatigue or misclassified because of incomplete documentation. Moreover, alarm actionability depends on clinical context, which cannot be captured solely through physiological signals and intervention data. Future improvements could include motion sensors or camera systems to detect motion artefacts, or prospective labelling methods such as annotation directly in HiX, integrated buttons in Dräger monitors or handheld devices. However, such manual approaches are resource-intensive and may be overlooked during busy clinical workflows.

Another limitation was the exclusion of HR alarms and those from other physiological parameters, including ECG, and devices such as ventilators and infusion pumps. In addition, the relatively small cohort size of 26 patients limits robustness and generalisability. Future research should therefore incorporate all PICU sources and include larger, more diverse patient populations to provide a more representative overview.

Finally, although the Dräger monitoring system was reconstructed as accurately as possible, discrepancies remain. The precise configuration, including the signal-processing filters, is not publicly documented, precluding exact replication. Moreover, the Dräger monitor operates at a higher sampling frequency than the stored data, making it more sensitive to alarm detection. Silenced alarms were not incorporated because their interpretation is ambiguous, as silencing may indicate either suppression of a nonactionable alarm or intervention for an actionable event. In addition, the $SpO_2$ L-T (< 80%) delay was incorrectly set to 1 s in the model, whereas it is actually 0 s, as this alarm is classified as high-priority alarm. As the calculated alarm delays in this study corresponded with those in the *Dräger Infinity Acute Care System User Manual* (39), time synchronisation is unlikely to pose problems. Future research should therefore rely on registered alarm events from the Dräger monitoring data to ensure comprehensive inclusion of alarms, especially the immediate activation of high-priority alarms from other parameters that was not fully captured in this study.

## 5.4 Recommendations

This section provides recommendations for future research on machine learning algorithms (Section 5.4.1) and for clinical practice (Section 5.4.2); with emphasis on the latter, as the findings highlight an immediate need for practical clinical interventions over machine learning algorithms.

### 5.4.1 Recommendations for Further Research

Future work on machine learning for alarm reduction should prioritise high-quality data annotation to produce reliable labels distinguishing actionable from nonactionable alarms. Poor annotation risks the "garbage in, garbage out" problem, where unreliable inputs yield unreliable outputs. (65) Accurate annotation would also clarify whether actionable alarms can be identified solely from physiological signals without clinical context. A practical first step could involve prospective annotation by shadowing nurses and recording annotations on handheld devices.

In addition, larger and more diverse datasets, including more severely ill patients to increase the proportion of actionable alarms, are needed to improve generalisability. The literature review (21) also identified relevant strategies from the process industry, such as alarm grouping, where related alarms are aggregated and displayed hierarchically rather than separately. Applied in clinical practice, this could help nurses recognise clusters of alarms and assess their relevance, though it would not reduce the total number of alarms generated.

### 5.4.2 Recommendations for Clinical Practices

The findings of this study support the implementation of straightforward interventions to reduce alarm burden in the PICU, with particular emphasis on threshold adjustments and alarm delays.

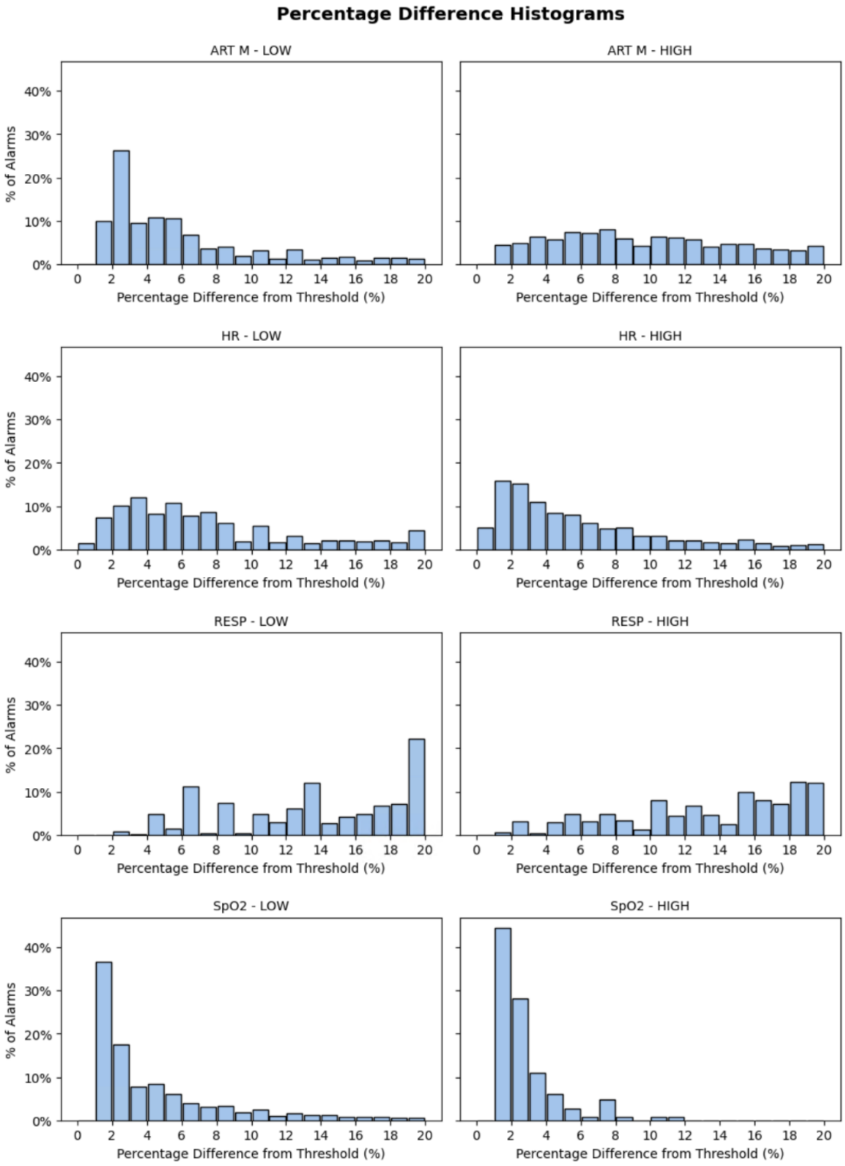
Nurses routinely adjust alarm limits at the beginning of each shift. Therefore, providing them with clear guidance on appropriate threshold adjustments could help reduce nonactionable alarms. Analysis of the reconstructed data demonstrated that $SpO_2$ alarms below the lower limit constituted the largest proportion of alarms; approximately 50% of these alarms could have been prevented by lowering the threshold by 2% (Figure 16). ART M alarms also clustered near the threshold, suggesting that modest adjustments (1-2%) may reduce their frequency, whereas HR and RESP alarms were more broadly distributed and less amenable to this strategy.

Alarm delays represent an additional intervention. The majority of alarms occurred within the first seconds, and more than 20% of $SpO_2$ alarms could potentially be avoided by extending the default Dräger delay by 2 seconds (Figure 17). However, current monitors do not allow manual configuration of delays and HR alarms are constrained by the AAMI/ANSI/IEC 60601-2-27 standard, which limits the total delay to 10 seconds. (39) For this reason, threshold adjustment should be prioritised.

To assess nurses' perspectives on this intervention, additional questions were posed to the two nurses who had been interviewed regarding alarm prediction and the implementation of the machine learning model. A 2% threshold adjustment was considered to increase the risk of missing important alarms for acutely ill or unfamiliar patients but feasible for stable and familiar patients with longer admissions. A similar perspective was expressed regarding delay extension. Nurses reported that, in current practice, thresholds are set tightly and subsequently broadened if repeated nonactionable alarms occur.

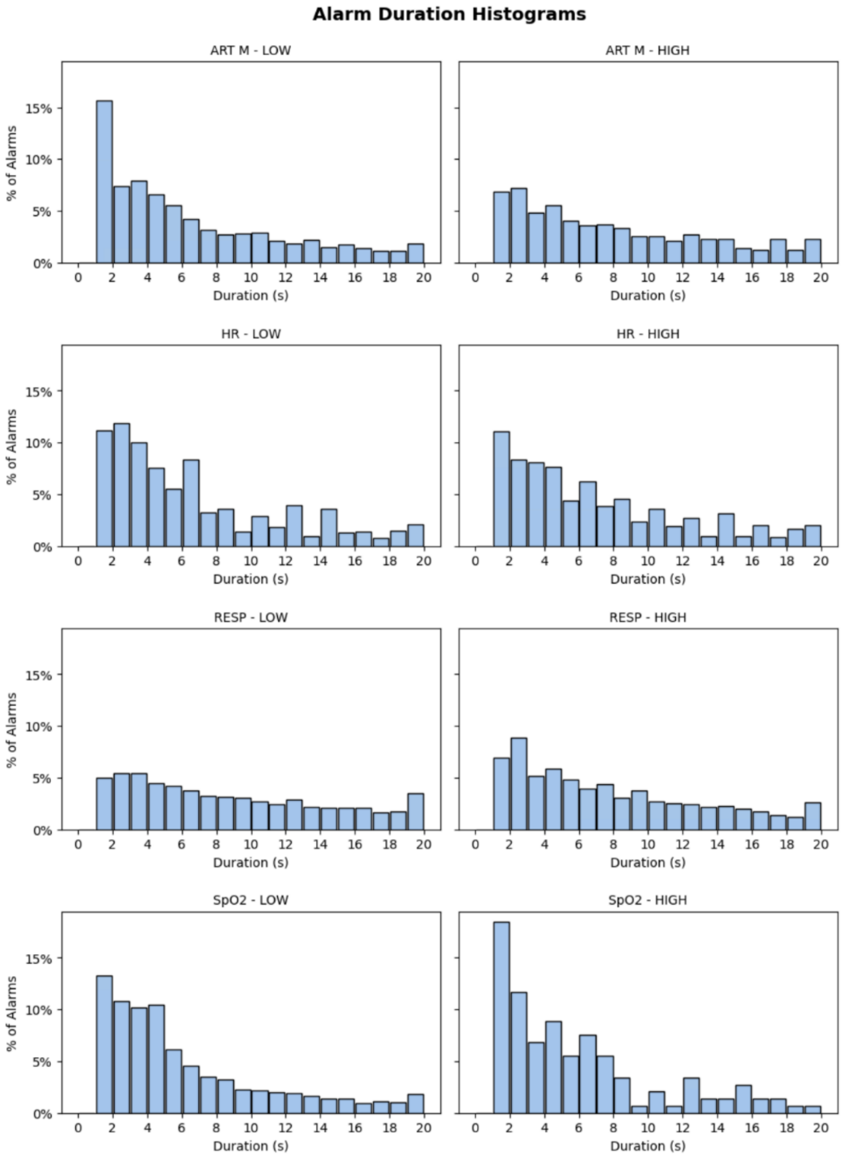
To ensure safe implementation, nurses emphasised the importance of education on the clinical consequences of threshold adjustments and structured dialogue with physicians about the impact of these changes. The introduction of alarm dashboards is recommended to enable regular multidisciplinary review during ward rounds. Such dashboards would enable identification of patterns, for example a higher number of $SpO_2$ alarms due to poor sensor attachment or inappropriate limits. Acutely ill patients could be reviewed daily and stable long-term patients weekly. It was further suggested that alarm limits be defined by the multidisciplinary team at the time of postoperative admission.

In summary, modest threshold adjustments, supported by multidisciplinary review and enhanced education, represent feasible interventions to reduce alarm burden while safeguarding patient safety and promoting a calmer ward environment. It is further recommended to draw inspiration from and collaborate with existing initiatives in the Netherlands, as described in this report, to avoid unnecessary duplication of efforts.




*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 16.** Percentage difference histograms for ART M, HR, RESP and $SpO_2$ alarms at low and high thresholds. The x-axis shows the percentage difference from the threshold, while the y-axis indicates the proportion of alarms. Data are based on n = 63 patients admitted to the PICU at Erasmus MC Sophia Children's Hospital between November 2021 and October 2024.




*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 17.** Alarm duration histograms for ART M, HR, RESP and $SpO_2$ alarms at low and high thresholds. The x-axis shows the alarm duration in seconds, while the y-axis indicates the proportion of alarms. Data are based on n = 63 patients admitted to the PICU at Erasmus MC Sophia Children's Hospital between November 2021 and October 2024.

# 6
# Conclusion

# 6 Conclusion

This study demonstrates that the machine learning algorithm developed to classify actionable alarms from multimodal vital sign data is not yet suitable for clinical application. The best-performing DT model achieved sensitivities of only 36-48%, specificities of 51-70% and balanced accuracies of about 50%, with no clear feature distinction between actionable and nonactionable alarms. These results indicate that, with the current feature set and modelling approach, the algorithm cannot reliably distinguish actionable from nonactionable alarms.

These findings can be attributed, first, to the limitations of retrospective labelling of actionable alarms based on assumptions and, second, to the inherent difficulty of distinguishing actionable alarms from physiological signals without clinical context, an observation reinforced by nurses, who also emphasised the value of nonactionable alarms as early warnings.

Future research should prioritise prospective annotation, increase the size and heterogeneity of the dataset and consider complementary strategies such as alarm grouping to support clinical decision-making. For current practice, the principal recommendation is to implement an alarm dashboard for multidisciplinary review of alarm limits and electrode placement, thereby reducing alarm burden and mitigating alarm fatigue while safeguarding patient safety.

# References


1. Herrera H, Wood D. Battling Alarm Fatigue in the Pediatric Intensive Care Unit. Crit Care Nurs Clin North Am. 2023;35(3):347-55.
2. Langga L, Oh J, López D, Blake N, McField E, Hotz J, et al. Prevalence, Proportionality, and Cause of Ventilator Alarms in a Pediatric Intensive Care Setting. Respir Care. 2021;66(4):541-50.
3. Drew BJ, Califf RM, Funk M, Kaufman ES, Krucoff MW, Laks MM, et al. Practice standards for electrocardiographic monitoring in hospital settings: an American Heart Association scientific statement from the Councils on Cardiovascular Nursing, Clinical Cardiology, and Cardiovascular Disease in the Young: endorsed by the International Society of Computerized Electrocardiology and the American Association of Critical-Care Nurses. Circulation. 2004;110(17):2721-46.
4. Dewan M, Wolfe H, Lin R, Ware E. Impact of a safety huddle–based intervention on monitor alarm rates in low-acuity pediatric intensive care unit patients. Journal of hospital …. 2017.
5. Yang JK, Su F, Graber-Naidich A, Hedlin H, Madsen N, DeSousa C, et al. Mitigating Alarm Fatigue and Improving the Bedside Experience by Reducing Nonactionable Alarms. J Pediatr. 2025;276:114278.
6. Moreno RP, Rhodes A, Donchin Y. Patient safety in intensive care medicine: the Declaration of Vienna. Intensive Care Med. 2009;35(10):1667-72.
7. Karnik A, Bonafide CP. A framework for reducing alarm fatigue on pediatric inpatient units. Hosp Pediatr. 2015;5(3):160-3.
8. Welch J, Rayo M, Kanter B, Bagian T, Jacobs K, Shanawani H, et al. Framework for Alarm Management Process Maturity. Biomed Instrum Technol. 2016;50(3):165-79.
9. Bonafide CP, Lin R, Zander M, Graham CS, Paine CW, Rock W, et al. Association between exposure to nonactionable physiologic monitor alarms and response time in a children's hospital. J Hosp Med. 2015;10(6):345-51.
10. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. Crit Care Med. 1994;22(6):981-5.
11. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. Crit Care Med. 1997;25(4):614-9.
12. Talley LB, Hooper J, Jacobs B, Guzzetta C, McCarter R, Sill A, et al. Cardiopulmonary monitors and clinically significant events in critically ill children. Biomed Instrum Technol. 2011;Suppl:38-45.
13. Rosman EC, Blaufox AD, Menco A, Trope R, Seiden HS. What are we missing? Arrhythmia detection in the pediatric intensive care unit. J Pediatr. 2013;163(2):511-4.
14. Morrison WE, Haas EC, Shaffner DH, Garrett ES, Fackler JC. Noise, stress, and annoyance in a pediatric intensive care unit. Crit Care Med. 2003;31(1):113-9.
15. Voepel-Lewis T, Parker ML, Burke CN, Hemberg J, Perlin L, Kai S, et al. Pulse oximetry desaturation alarms on a general postoperative adult unit: a prospective observational study of nurse response time. Int J Nurs Stud. 2013;50(10):1351-8.
16. Kaur H, Rohlik GM, Nemergut ME, Tripathi S. Comparison of staff and family perceptions of causes of noise pollution in the Pediatric Intensive Care Unit and suggested intervention strategies. Noise Health. 2016;18(81):78-84.
17. Korniewicz DM, Clark T, David Y. A national online survey on the effectiveness of clinical alarms. Am J Crit Care. 2008;17(1):36-41.
18. Choiniere DB. The effects of hospital noise. Nurs Adm Q. 2010;34(4):327-33.
19. Kudchadkar SR, Aljohani OA, Punjabi NM. Sleep of critically ill children in the pediatric intensive care unit: a systematic review. Sleep Med Rev. 2014;18(2):103-10.
20. Schondelmeyer AC, Dewan ML, Brady PW, Timmons KM, Cable R, Britto MT, et al. Cardiorespiratory and Pulse Oximetry Monitoring in Hospitalized Children: A Delphi Process. Pediatrics. 2020;146(2).
21. Van de Ruit CA. Reducing Clinical Alarms to Mitigate Alarm Fatigue in the Paediatric Intensive Care Unit 2025.
22. Poncette AS, Wunderlich MM, Spies C, Heeren P, Vorderwülbecke G, Salgado E, et al. Patient Monitoring Alarms in an Intensive Care Unit: Observational Study With Do-It-Yourself Instructions. J Med Internet Res. 2021;23(5):e26494.
23. Paine CW, Goel VV, Ely E, Stave CD. Systematic review of physiologic monitor alarm characteristics and pragmatic interventions to reduce alarm frequency. Journal of hospital …. 2016.

24. Dewan M, Cipriani L, Boyer J, Stark J, Seger B, Tegtmeyer K. Reducing Redundant Alarms in the Pediatric ICU. Multimodal Technologies and Interaction. 2019;3(1):11.
25. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. AACN Adv Crit Care. 2013;24(4):378-86; quiz 87-8.
26. Wilken M, Hüske-Kraus D, Klausen A, Koch C, Schlauch W, Röhrig R. Alarm Fatigue: Causes and Effects. Stud Health Technol Inform. 2017;243:107-11.
27. Turmell JW, Coke L, Catinella R, Hosford T, Majeski A. Alarm Fatigue: Use of an Evidence-Based Alarm Management Strategy. J Nurs Care Qual. 2017;32(1):47-54.
28. Rasooly IR, Kern-Goldberger AS, Xiao R, Ponnala S, Ruppel H, Luo B, et al. Physiologic Monitor Alarm Burden and Nurses' Subjective Workload in a Children's Hospital. Hosp Pediatr. 2021;11(7):703-10.
29. Lee SJ, Lee YM, Seo EJ, Son YJ. Impact of Hospital Nurses' Perception on Clinical Alarms and Patient Safety Culture on Alarm Management Practice. Int J Environ Res Public Health. 2021;18(8).
30. Medical device alarm safety in hospitals. Sentinel Event Alert. 2013(50):1-3.
31. Jones K. Alarm fatigue a top patient safety hazard. Cmaj. 2014;186(3):178.
32. Commission TJ. Alarm system safety 2013 [Available from: https://www.jointcommission.org/assets/1/18/R3_Report_Issue_5_12_2_13_Final.pdf.
33. Commission TJ. Hospital: 2025 National Patient Safety Goals 2025 [Available from: https://www.jointcommission.org/standards/national-patient-safety-goals/hospital-national-patient-safety-goals/.
34. ECRI. Top 10 Health Technology Hazards Executive Brief 2025 [Available from: https://home.ecri.org/blogs/ecri-thought-leadership-resources.
35. (NVKF) NVvKF. Leidraad medische bewakings- en alarmeringssystemen. 2023.
36. KGaA DAC. Infinity® Acute Care System n.d. [Available from: https://www.draeger.com/en-us_us/Products/Infinity-Acute-Care-System.
37. KGaA DAC. Infinity® CentralStation Wide n.d. [Available from: https://www.draeger.com/nl_nl/Products/Infinity-CentralStation.
38. KGaA DAC. Infinity® M540 [Available from: https://www.draeger.com/nl_nl/Products/Infinity-M540-monitor.
39. KGaA DAC. Gebruiksaanwijzing Infinity Acute Care System - Bewakingstoepassingen Software VG7.1. Lübeck, Germany.
40. Bostan I, Ozcan Vieira E, Gommers D, van Egmond R. Annoyance by Alarms in the ICU: A Cognitive Approach to the Role of Interruptions by Patient Monitoring Alarms. Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2022; Torino, Italy2022. p. 133-48.
41. Kalden M, Goos TG, Kalden N, Groenendaal LA, Reiss IKM, Van Bommel J, et al. The Physiological Effects of an Adjusted Alarm Architecture on a Neonatal Intensive Care Unit. Acta Paediatr. 2025.
42. Stevens DC, Helseth CC, Thompson PA, Pottala JV, Khan MA, Munson DP. A Comprehensive Comparison of Open-Bay and Single-Family-Room Neonatal Intensive Care Units at Sanford Children's Hospital. Herd. 2012;5(4):23-39.
43. Almadhoob A. Sound reduction management in the neonatal intensive care unit for preterm or very low birth weight infants. Cochrane Database of Systematic Reviews. 2020(1).
44. Szymczak SE, Shellhaas RA. Impact of NICU design on environmental noise. J Neonatal Nurs. 2014;20(2):77-81.
45. Kasparick M, Klotz T, Thürk G, Andersen B, Rother B, Golatowski F, et al. Silent ICU based on IEEE 11073 SDC. Current Directions in Biomedical Engineering. 2023;9(1):303-6.
46. Gulla K, Sachdev A. Illness severity and organ dysfunction scoring in Pediatric Intensive Care Unit. Indian Journal of Critical Care Medicine. 2016;20:27.
47. ESPNIC. Paediatric Index of Mortality 3 [Available from: www.espnic.eu/calcolatori/Paediatric%20Index%20of%20Mortality%203%20_%20Professional%20Resources%20_%20Education%20-%20European%20Society%20of%20Paediatric%20Neonatal%20Intensive%20Care%20-%20ESPNIC.html.
48. van Kekem MJ. Towards a Quieter ICU: Mapping Alarms and Exploring Opportunities for SpO2 Alarm Reduction: Delft University of Technology; 2025.
49. Chromik J, Flint AR, Prendke M, Arnrich B, Poncette A-S. Enabling machine learning models in alarm fatigue research: Creation of a large relevance-annotated oxygen saturation alarm data set. Computers in Biology and Medicine. 2024;183:109244.

50. Klopfenstein SAI, Flint AR, Heeren P, Prendke M, Chaoui A, Ocker T, et al. How to Annotate Patient Monitoring Alarms in Intensive Care Medicine for Machine Learning. Research Square. 2023.
51. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, California, USA: Association for Computing Machinery; 2016. p. 785–94.
52. Albaldawi T. An overview of machine learning classification techniques. BIO Web of Conferences. 2024;97:00133.
53. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; Anchorage, AK, USA: Association for Computing Machinery; 2019. p. 2623–31.
54. Van Calster B, Collins G, Vickers A, Wynants L, Kerr K, Barreñada L, et al. Performance evaluation of predictive AI models to support medical decisions: Overview and guidance2024.
55. Ostojic D, Guglielmini S, Moser V, Fauchere JC, Bucher HU, Bassler D, et al. Reducing False Alarm Rates in Neonatal Intensive Care: A New Machine Learning Approach. Adv Exp Med Biol. 2020;1232:285-90.
56. Tsien CL, Kohane IS, McIntosh N. Multiple signal integration by decision tree induction to detect artifacts in the neonatal intensive care unit. Artificial Intelligence in Medicine. 2000.
57. Monasterio V, Burgess F, Clifford GD. Robust classification of neonatal apnoea-related desaturations. Physiol Meas. 2012;33(9):1503-16.
58. Au-Yeung WM, Sahani AK, Isselbacher EM, Armoundas AA. Reduction of false alarms in the intensive care unit using an optimized machine learning based approach. NPJ Digit Med. 2019;2:86.
59. Hever G, Cohen L, O'Connor MF, Matot I, Lerner B, Bitan Y. Machine learning applied to multi-sensor information to reduce false alarm rate in the ICU. J Clin Monit Comput. 2020;34(2):339-52.
60. Kalden M, Goos T, Pul Cv, Bennis F. Prediction of Critical Oxygen Saturation Alarms at the Neonatal Intensive Care Unit (NICU). 2023.
61. Cabrera-Quirós L, Varisco G, Zhan Z, Long X. Listen to the Real Experts: Detecting Need of Caregiver Response in a NICU using Multimodal Monitoring Signals. … Publication of the 2021 …. 2021.
62. Smart and Silent ICU 2025 [Available from: https://www.sasicu.eu/.
63. Boelhouwer I. Silencing the Chaos: A New Approach to Alarm Configuration in Intensive Care through Interface Design. 2025.
64. Varisco G, van de Mortel H, Cabrera-Quiros L, Atallah L, Hueske-Kraus D, Long X, et al. Optimisation of clinical workflow and monitor settings safely reduces alarms in the NICU. Acta Paediatr. 2021;110(4):1141-50.
65. Koçak B, Cuocolo R, dos Santos DP, Stanzione A, Ugga L. Must-have Qualities of Clinical Research on Artificial Intelligence and Machine Learning. Balkan Med J. 2023;40(1):3-12.

# Supplementary Materials

## Appendix I. Feature Descriptions

| Feature | Description |
|---|---|
| ART M_median | Median of ART M within the alarm window |
| ART M_iqr | IQR of ART M within the alarm window |
| ART M_variability | Variability of ART M within the alarm window |
| ART M_slope | Temporal slope of ART M within the alarm window |
| ART M_missing_count | Count of ART M time points flagged as *ARTEFACT* within the alarm window |
| ART M_missing_frac | Proportion of ART M time points flagged as *ARTEFACT* within the alarm window |
| ART M_high_count | Count of ART M time points flagged as *OUT OF RANGE HIGH* within the alarm window |
| ART M_high_frac | Proportion of ART M time points flagged as *OUT OF RANGE HIGH* within the alarm window |
| ART M_low_count | Count of ART M time points flagged as *OUT OF RANGE LOW* within the alarm window |
| ART M_low_frac | Proportion of ART M time points flagged as *OUT OF RANGE LOW* within the alarm window |
| HR_median | Median of HR within the alarm window |
| HR_iqr | IQR of HR within the alarm window |
| HR_variability | Variability of HR within the alarm window |
| HR_slope | Temporal slope of HR within the alarm window |
| HR_missing_count | Count of HR time points flagged as *ARTEFACT* within the alarm window |
| HR_missing_frac | Proportion of HR time points flagged as *ARTEFACT* within the alarm window |
| HR_high_count | Count of HR time points flagged as *OUT OF RANGE HIGH* within the alarm window |
| HR_high_frac | Proportion of HR time points flagged as *OUT OF RANGE HIGH* within the alarm window |
| HR_low_count | Count of HR time points flagged as *OUT OF RANGE LOW* within the alarm window |
| HR_low_frac | Proportion of HR time points flagged as *OUT OF RANGE LOW* within the alarm window |
| SpO2_median | Median of $SpO_2$ within the alarm window |
| SpO2_iqr | IQR of $SpO_2$ within the alarm window |
| SpO2_variability | Variability of $SpO_2$ within the alarm window |
| SpO2_slope | Temporal slope of $SpO_2$ within the alarm window |
| SpO2_missing_count | Count of $SpO_2$ time points flagged as *ARTEFACT* within the alarm window |
| SpO2_missing_frac | Proportion of $SpO_2$ time points flagged as *ARTEFACT* within the alarm window |
| SpO2_high_count | Count of $SpO_2$ time points flagged as *OUT OF RANGE HIGH* within the alarm window |
| SpO2_high_frac | Proportion of $SpO_2$ time points flagged as *OUT OF RANGE HIGH* within the alarm window |
| SpO2_low_count | Count of $SpO_2$ time points flagged as *OUT OF RANGE LOW* within the alarm window |
| SpO2_low_frac | Proportion of $SpO_2$ time points flagged as *OUT OF RANGE LOW* within the alarm window |
| RESP_median | Median of RESP within the alarm window |
| RESP_iqr | IQR of RESP within the alarm window |
| RESP_variability | Variability of RESP within the alarm window |
| RESP_slope | Temporal slope of RESP within the alarm window |
| RESP_missing_count | Count of RESP time points flagged as *ARTEFACT* within the alarm window |
| RESP_missing_frac | Proportion of RESP time points flagged as *ARTEFACT* within the alarm window |
| RESP_high_count | Count of RESP time points flagged as *OUT OF RANGE HIGH* within the alarm window |
| RESP_high_frac | Proportion of RESP time points flagged as *OUT OF RANGE HIGH* within the alarm window |
| RESP_low_count | Count of RESP time points flagged as *OUT OF RANGE LOW* within the alarm window |
| RESP_low_frac | Proportion of RESP time points flagged as *OUT OF RANGE LOW* within the alarm window |
| RESP_apn_count | Count of RESP time points flagged as *APNOEA* within the alarm window |
| RESP_apn_frac | Proportion of RESP time points flagged as *APNOEA* within the alarm window |
| ART M_x_HR_corr | Correlation between ART M and HR within the alarm window |
| ART M_x_SpO2_corr | Correlation between ART M and $SpO_2$ within the alarm window |
| ART M_x_RESP_corr | Correlation between ART M and RESP within the alarm window |
| HR_x_SpO2_corr | Correlation between HR and $SpO_2$ within the alarm window |
| HR_x_RESP_corr | Correlation between HR and RESP within the alarm window |
| SpO2_x_RESP_corr | Correlation between $SpO_2$ and RESP within the alarm window |

*ART M = mean invasive arterial blood pressure, HR = heart rate, IQR = interquartile range, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 1.** Overview of the features accompanied by brief descriptions.

## Appendix II. Hyperparameter Optimisation

The following tables summarise the hyperparameter search spaces defined for each machine learning model within the Optuna optimisation framework.

### Logistic Regression (LR)

| Hyperparameter | Range/ Options |
|---|---|
| Penalty | {“l1”, “l2”, “elasticnet”} |
| C | [0.001, 100] (log scale) |
| l1_ratio | [0.0, 1.0] |

**Table 1.** Hyperparameter search space for LR.

### Decision Tree (DT)

| Hyperparameter | Range/ Options |
|---|---|
| Max depth | [2, 20] |
| Min samples split | [2, 20] |
| Min samples leaf | [1, 10] |
| Max features | {“sqrt”, “log2”, None} |
| ccp_alpha | [1e-5, 1e-2] (log scale) |

**Table 2.** Hyperparameter search space for DT.

### Random Forest (RF)

| Hyperparameter | Range/ Options |
|---|---|
| Number of estimators | [100, 1000] (step = 100) |
| Max depth | [3, 20] |
| Min samples split | [2, 20] |
| Min samples leaf | [1, 10] |
| Max features | {“sqrt”, “log2”} |
| Bootstrap | {True, False} |

**Table 3.** Hyperparameter search space for RF.

### Extreme Gradient Boosting ( XGBoost )

| Hyperparameter | Range/ Options |
|---|---|
| Number of estimators | [200, 800] (step = 100) |
| Max depth | [3, 10] |
| Learning rate | [0.001, 0.3] (log scale) |
| Subsample | [0.5, 1.0] |
| Colsample_bytree | [0.5, 1.0] |
| Gamma | [0, 5] |
| Reg_lambda | [0.001, 10] (log scale) |
| Reg_alpha | [0.001, 10] (log scale) |
| Scale pos weight | [0.5, 5.0] |

**Table 4.** Hyperparameter search space for XGBoost.

## Appendix III. Dräger Alarm Statistics

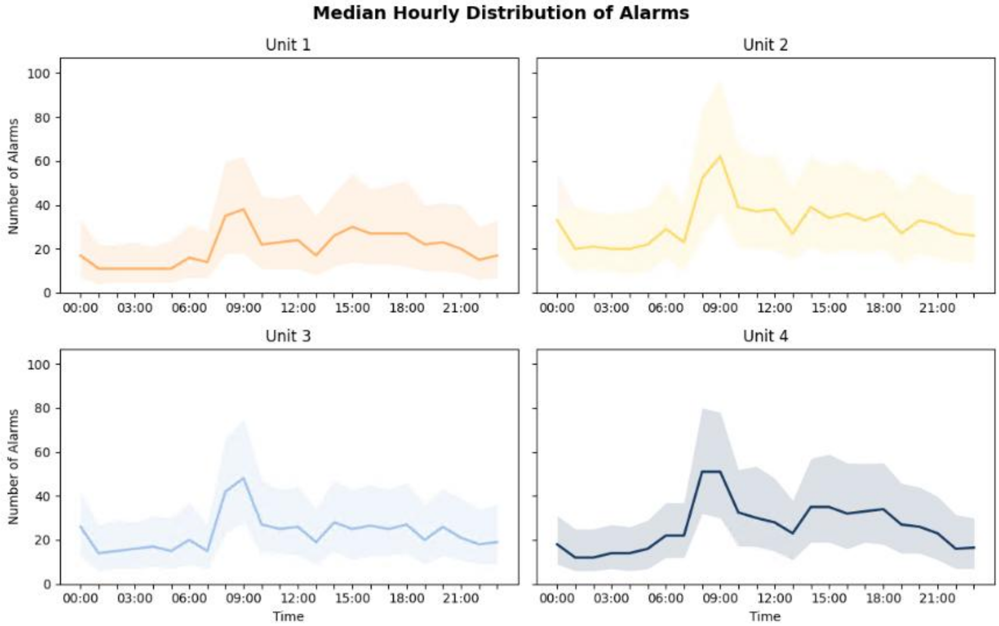



**Figure 1.** Median hourly distribution of alarms per PICU units at Erasmus MC Sophia Children's Hospital, November 2021–October 2024. Values are presented as the median with IQR. Data are based on n = 2,582 patients.

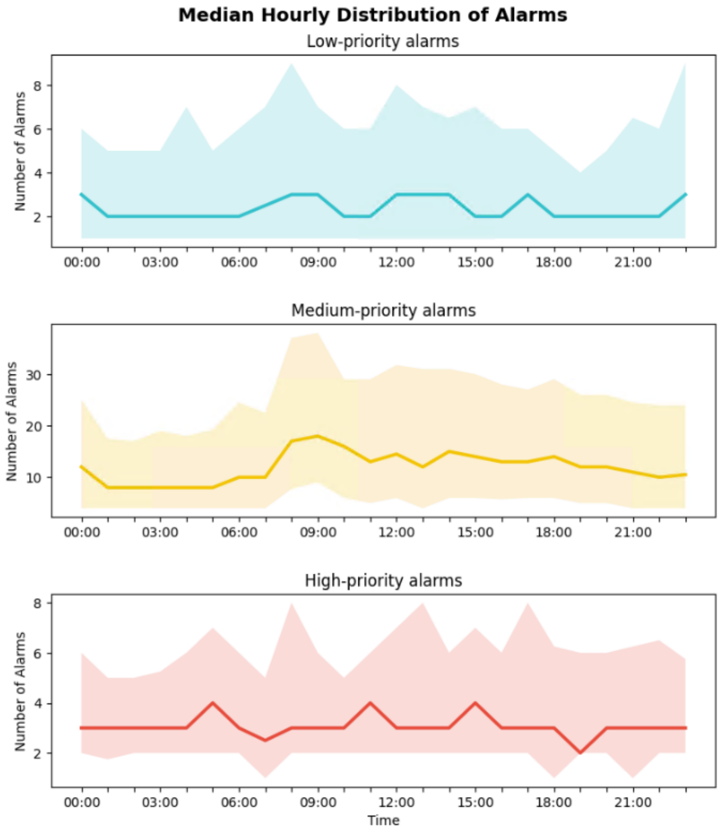
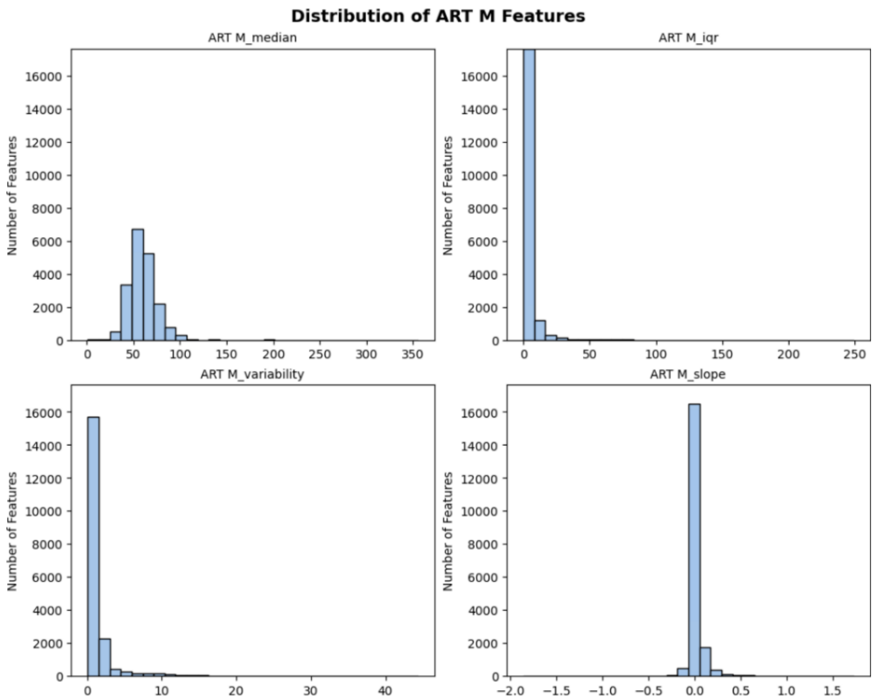


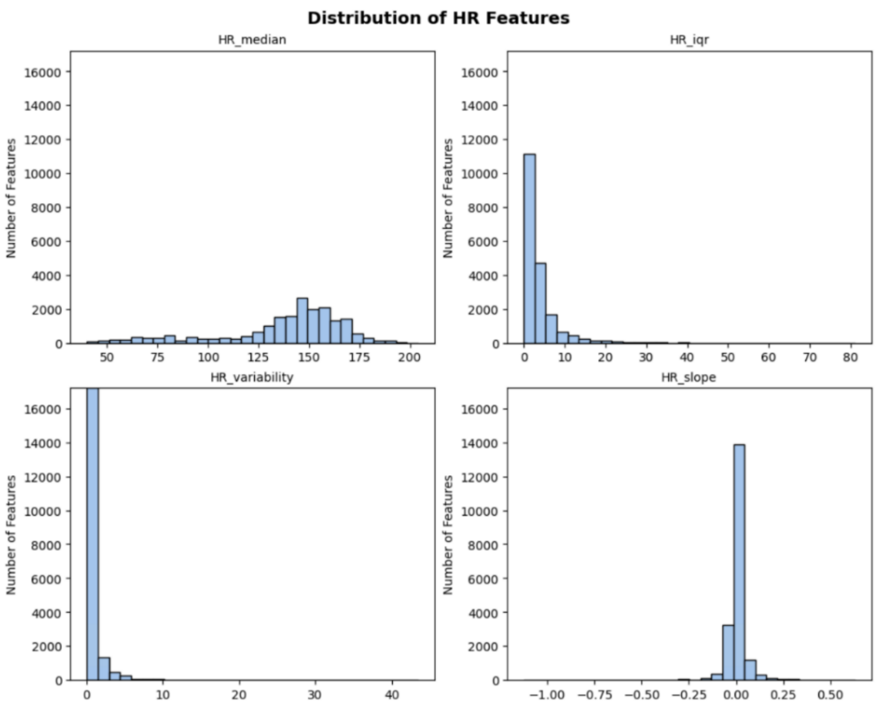

**Figure 2.** Median hourly distribution of alarms by priority level across all PICU units at Erasmus MC Sophia Children's Hospital, November 2021–October 2024. Values are presented as the median with IQR. Data are based on n = 66 patients.

## Appendix IV. Feature Distributions



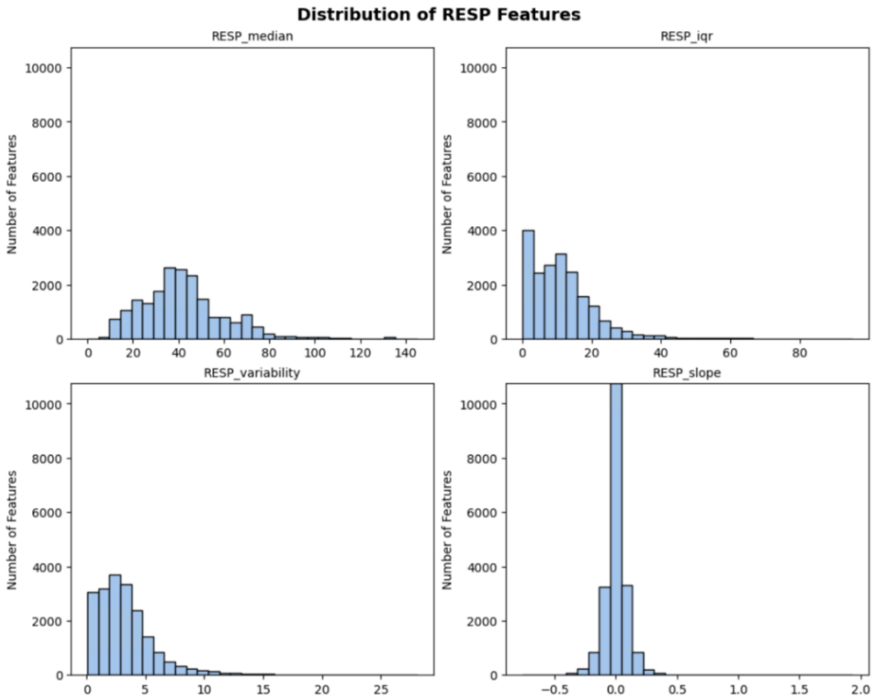

*ART M = mean invasive arterial blood pressure, IQR = interquartile range*

**Figure 1.** Distributions of the ART M features computed over the 3-minute alarm window, demonstrating their ranges and skewness to inform the choice of feature scaling for model development.



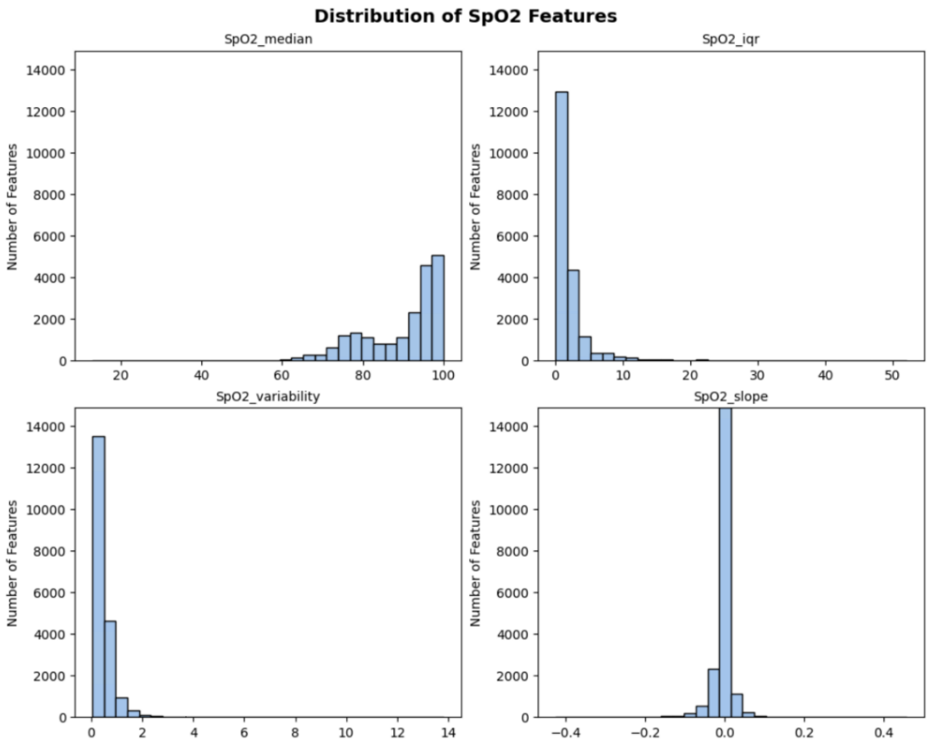

*HR = heart rate, IQR = interquartile range*

**Figure 2.** Distributions of the HR features computed over the 3-minute alarm window, demonstrating their ranges and skewness to inform the choice of feature scaling for model development.



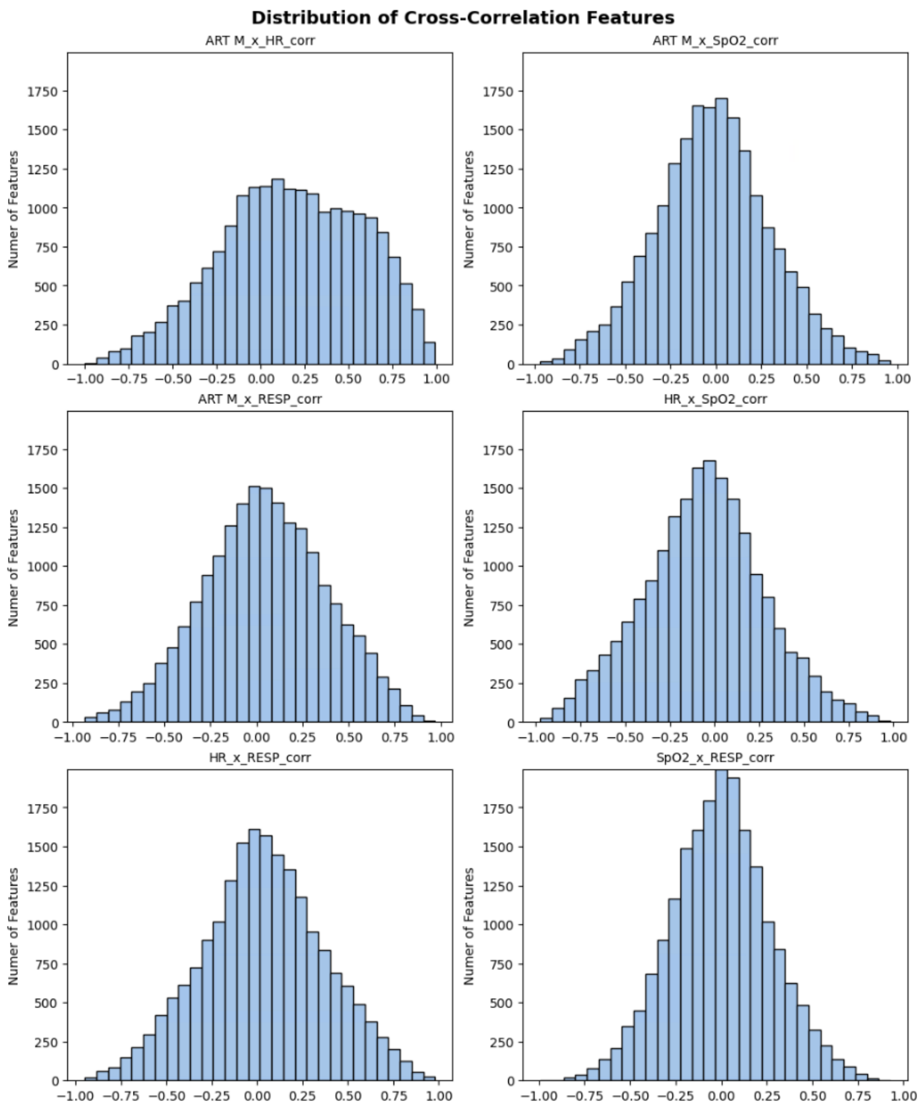

*RESP = respiratory rate, IQR = interquartile range*

**Figure 3.** Distributions of the RESP features computed over the 3-minute alarm window, demonstrating their ranges and skewness to inform the choice of feature scaling for model development.




*$SpO_2$ = oxygen saturation, IQR = interquartile range*

**Figure 4.** Distributions of the $SpO_2$ features computed over the 3-minute alarm window, demonstrating their ranges and skewness to inform the choice of feature scaling for model development.




*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 5.** Distributions of the cross-correlation features computed over the 3-minute alarm window, demonstrating their ranges and skewness to inform the choice of feature scaling for model development.

## Appendix V. Model Evaluation

This appendix presents the performance of the class-weighted and undersampled machine learning models evaluated in this study, reported as medians with interquartile ranges (IQR) across the outer test folds.

### Class Weighting Models

#### Logistic Regression (LR)

| **Alarm Window** *(min)* | **Sensitivity** *Median (Q1-Q3)* | **Specificity** *Median (Q1-Q3)* | **Balanced Accuracy** *Median (Q1-Q3)* | **AUROC** *Median (Q1-Q3)* | **F1-score** *Median (Q1-Q3)* |
|---|---|---|---|---|---|
| 3 | 0.23 (0.14 – 0.48) | 0.76 (0.63 – 0.86) | 0.54 (0.53 – 0.56) | 0.60 (0.56 – 0.62) | 0.20 (0.12 – 0.25) |
| 0.5 | 0.37 (0.17 – 0.51) | 0.69 (0.61 – 0.80) | 0.53 (0.51 – 0.57) | 0.55 (0.50 – 0.58) | 0.21 (0.10 – 0.29) |
| 1 | 0.34 (0.29 – 0.49) | 0.70 (0.63 – 0.77) | 0.54 (0.51 – 0.57) | 0.54 (0.52 – 0.60) | 0.22 (0.14 – 0.29) |
| 5 | 0.40 (0.24 – 0.57) | 0.69 (0.63 – 0.78) | 0.54 (0.54 – 0.59) | 0.58 (0.55 – 0.66) | 0.20 (0.16 – 0.22) |

**Table 1.** Class weighting performance of LR across alarm windows, reported as median with IQR (Q1-Q3).

#### Decision Tree (DT)

| **Alarm Window** *(min)* | **Sensitivity** *Median (Q1-Q3)* | **Specificity** *Median (Q1-Q3)* | **Balanced Accuracy** *Median (Q1-Q3)* | **AUROC** *Median (Q1-Q3)* | **F1-score** *Median (Q1-Q3)* |
|---|---|---|---|---|---|
| 3 | 0.42 (0.33 - 0.55) | 0.60 (0.55 – 0.70) | 0.53 (0.49 – 0.55) | 0.53 (0.52 – 0.56) | 0.19 (0.10 – 0.27) |
| 0.5 | 0.48 (0.40 – 0.60) | 0.51 (0.43 – 0.64) | 0.53 (0.49 – 0.56) | 0.53 (0.50 – 0.58) | 0.22 (0.11 – 0.28) |
| 1 | 0.40 (0.19 – 0.67) | 0.66 (0.45 – 0.78) | 0.52 (0.48 – 0.53) | 0.55 (0.52 – 0.58) | 0.20 (0.13 – 0.27) |
| 5 | 0.36 (0.29 – 0.48) | 0.70 (0.49 – 0.81) | 0.54 (0.52 – 0.57) | 0.55 (0.54 – 0.58) | 0.24 (0.12 – 0.26) |

**Table 2.** Class weighting performance of DT across alarm windows, reported as median with IQR (Q1-Q3).

#### Random Forest (RF)

| **Alarm Window** *(min)* | **Sensitivity** *Median (Q1-Q3)* | **Specificity** *Median (Q1-Q3)* | **Balanced Accuracy** *Median (Q1-Q3)* | **AUROC** *Median (Q1-Q3)* | **F1-score** *Median (Q1-Q3)* |
|---|---|---|---|---|---|
| 3 | 0.37 (0.08 – 0.48) | 0.72 (0.58 – 0.87) | 0.55 (0.51 – 0.58) | 0.58 (0.53 – 0.64) | 0.21 (0.06 – 0.29) |
| 0.5 | 0.21 (0.01 – 0.33) | 0.71 (0.69 – 0.92) | 0.50 (0.47 – 0.55) | 0.55 (0.50 – 0.59) | 0.09 (0.02 – 0.27) |
| 1 | 0.29 (0.05 – 0.37) | 0.77 (0.62 – 0.90) | 0.53 (0.51 – 0.57) | 0.56 (0.52 – 0.61) | 0.16 (0.08 – 0.29) |
| 5 | 0.42 (0.04 – 0.57) | 0.68 (0.58 – 0.81) | 0.57 (0.50 – 0.61) | 0.60 (0.55 – 0.66) | 0.19 (0.03 – 0.28) |

**Table 3.** Class weighting performance of RF across alarm windows, reported as median with IQR (Q1-Q3).

#### Extreme Gradient Boosting (XGBoost)

| **Alarm Window** *(min)* | **Sensitivity** *Median (Q1-Q3)* | **Specificity** *Median (Q1-Q3)* | **Balanced Accuracy** *Median (Q1-Q3)* | **AUROC** *Median (Q1-Q3)* | **F1-score** *Median (Q1-Q3)* |
|---|---|---|---|---|---|
| 3 | 0.05 (0.00 – 0.22) | 0.90 (0.89 – 0.96) | 0.51 (0.48 – 0.54) | 0.54 (0.51 – 0.60) | 0.07 (0.00 – 0.19) |
| 0.5 | 0.08 (0.01 – 0.16) | 0.90 (0.89 – 0.98) | 0.51 (0.50 – 0.52) | 0.55 (0.50 – 0.61) | 0.10 (0.02 – 0.16) |
| 1 | 0.05 (0.002 – 0.20) | 0.94 (0.82 – 0.98) | 0.51 (0.50 – 0.52) | 0.53 (0.51 – 0.57) | 0.07 (0.00 – 0.16) |
| 5 | 0.08 (0.03 – 0.18) | 0.95 (0.83 – 0.95) | 0.51 (0.50 – 0.52) | 0.59 (0.53 – 0.64) | 0.10 (0.05 – 0.13) |

**Table 4.** Class weighting performance of XGBoost across alarm windows, reported as median with IQR (Q1-Q3).

## Undersampling Models

### Logistic Regression (LR)

| **Alarm Window** *(min)* | **Sensitivity** *Median (Q1-Q3)* | **Specificity** *Median (Q1-Q3)* | **Balanced Accuracy** *Median (Q1-Q3)* | **AUROC** *Median (Q1-Q3)* | **F1-score** *Median (Q1-Q3)* |
|---|---|---|---|---|---|
| 3 | 0.23 (0.14 – 0.48) | 0.76 (0.63 – 0.86) | 0.54 (0.53 – 0.56) | 0.60 (0.56 – 0.62) | 0.20 (0.12 – 0.25) |
| 0.5 | 0.13 (0.06 – 0.15) | 0.92 (0.83 – 0.94) | 0.50 (0.49 – 0.54) | 0.54 (0.49 – 0.60) | 0.14 (0.06 – 0.20) |
| 1 | 0.12 (0.09 – 0.14) | 0.91 (0.86 – 0.92) | 0.52 (0.50 – 0.53) | 0.56 (0.53 – 0.60) | 0.15 (0.09 – 0.16) |
| 5 | 0.42 (0.28 – 0.60) | 0.67 (0.63 – 0.73) | 0.55 (0.55 – 0.60) | 0.59 (0.56 – 0.66) | 0.20 (0.15 – 0.24) |

**Table 5.** Undersampling performance of LR across alarm windows, reported as median with IQR (Q1-Q3).

### Decision Tree (DT)

| **Alarm Window** *(min)* | **Sensitivity** *Median (Q1-Q3)* | **Specificity** *Median (Q1-Q3)* | **Balanced Accuracy** *Median (Q1-Q3)* | **AUROC** *Median (Q1-Q3)* | **F1-score** *Median (Q1-Q3)* |
|---|---|---|---|---|---|
| 3 | 0.42 (0.33 – 0.55) | 0.60 (0.55 – 0.70) | 0.53 (0.49 – 0.55) | 0.53 (0.52 – 0.56) | 0.19 (0.10 – 0.27) |
| 0.5 | 0.20 (0.09 – 0.28) | 0.74 (0.66 – 0.82) | 0.49 (0.45 – 0.52) | 0.49 (0.47 – 0.53) | 0.14 (0.06 – 0.21) |
| 1 | 0.19 (0.03 – 0.31) | 0.76 (0.73 – 0.86) | 0.50 (0.49 – 0.52) | 0.53 (0.51 – 0.56) | 0.14 (0.05 – 0.18) |
| 5 | 0.54 (0.36 – 0.75) | 0.66 (0.39 – 0.79) | 0.57 (0.54 – 0.58) | 0.59 (0.57 – 0.63) | 0.23 (0.16 – 0.28) |

**Table 6.** Undersampling performance of DT across alarm windows, reported as median with IQR (Q1-Q3).

### Random Forest (RF)

| **Alarm Window** *(min)* | **Sensitivity** *Median (Q1-Q3)* | **Specificity** *Median (Q1-Q3)* | **Balanced Accuracy** *Median (Q1-Q3)* | **AUROC** *Median (Q1-Q3)* | **F1-score** *Median (Q1-Q3)* |
|---|---|---|---|---|---|
| 3 | 0.37 (0.08 – 0.48) | 0.73 (0.58 – 0.87) | 0.55 (0.51 – 0.58) | 0.58 (0.53 – 0.64) | 0.21 (0.06 – 0.29) |
| 0.5 | 0.14 (0.08 – 0.22) | 0.86 (0.80 – 0.95) | 0.52 (0.50 – 0.53) | 0.56 (0.54 – 0.63) | 0.13 (0.10 – 0.21) |
| 1 | 0.13 (0.08 – 0.28) | 0.90 (0.79 – 0.92) | 0.52 (0.52 – 0.55) | 0.60 (0.54 – 0.62) | 0.14 (0.11 – 0.19) |
| 5 | 0.27 (0.12 – 0.50) | 0.78 (0.61 – 0.86) | 0.54 (0.50 – 0.59) | 0.61 (0.55 – 0.64) | 0.17 (0.10 – 0.27) |

**Table 7.** Undersampling performance of RF across alarm windows, reported as median with IQR (Q1-Q3).

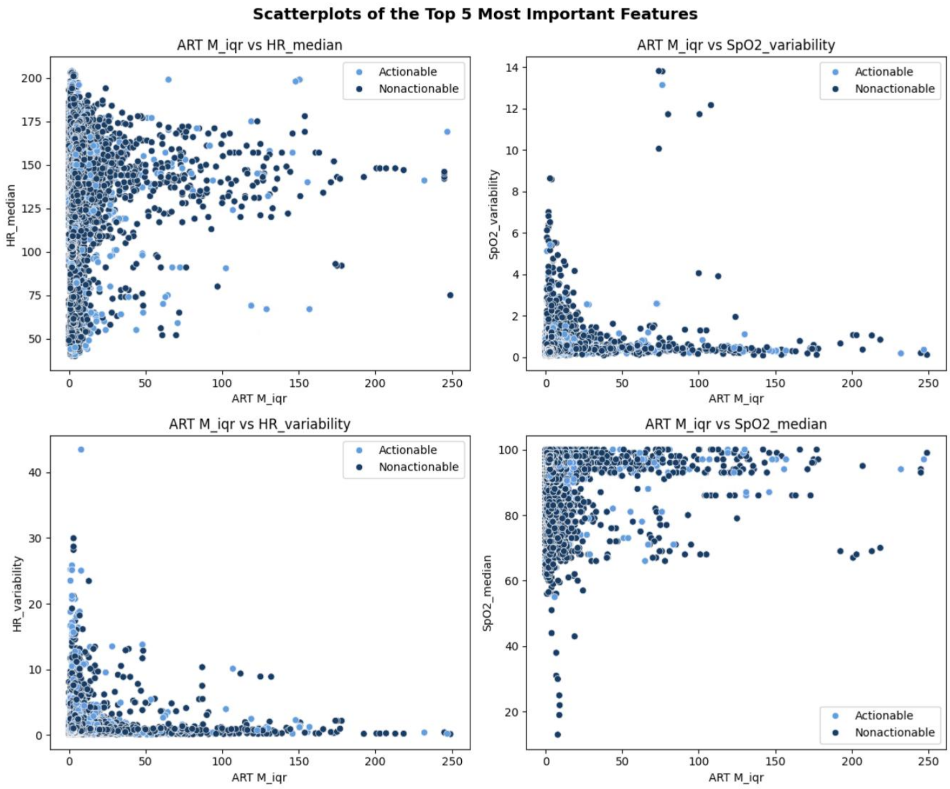
### Extreme Gradient Boosting (XGBoost)

| **Alarm Window** *(min)* | **Sensitivity** *Median (Q1-Q3)* | **Specificity** *Median (Q1-Q3)* | **Balanced Accuracy** *Median (Q1-Q3)* | **AUROC** *Median (Q1-Q3)* | **F1-score** *Median (Q1-Q3)* |
|---|---|---|---|---|---|
| 3 | 0.05 (0.00 – 0.22) | 0.90 (0.89 – 0.96) | 0.51 (0.48 – 0.54) | 0.54 (0.51 – 0.60) | 0.07 (0.00 – 0.19) |
| 0.5 | 0.14 (0.00 – 0.32) | 0.80 (0.77 – 0.90) | 0.51 (0.50 – 0.53) | 0.55 (0.53 – 0.60) | 0.11 (0.00 – 0.21) |
| 1 | 0.17 (0.03 – 0.33) | 0.85 (0.79 – 0.90) | 0.52 (0.50 – 0.54) | 0.56 (0.54 – 0.59) | 0.14 (0.03 – 0.25) |
| 5 | 0.16 (0.10 – 0.25) | 0.86 (0.83 – 0.92) | 0.52 (0.51 – 0.53) | 0.60 (0.56 – 0.62) | 0.14 (0.08 – 0.19) |

**Table 8.** Undersampling performance of XGBoost across alarm windows, reported as median with IQR (Q1-Q3).

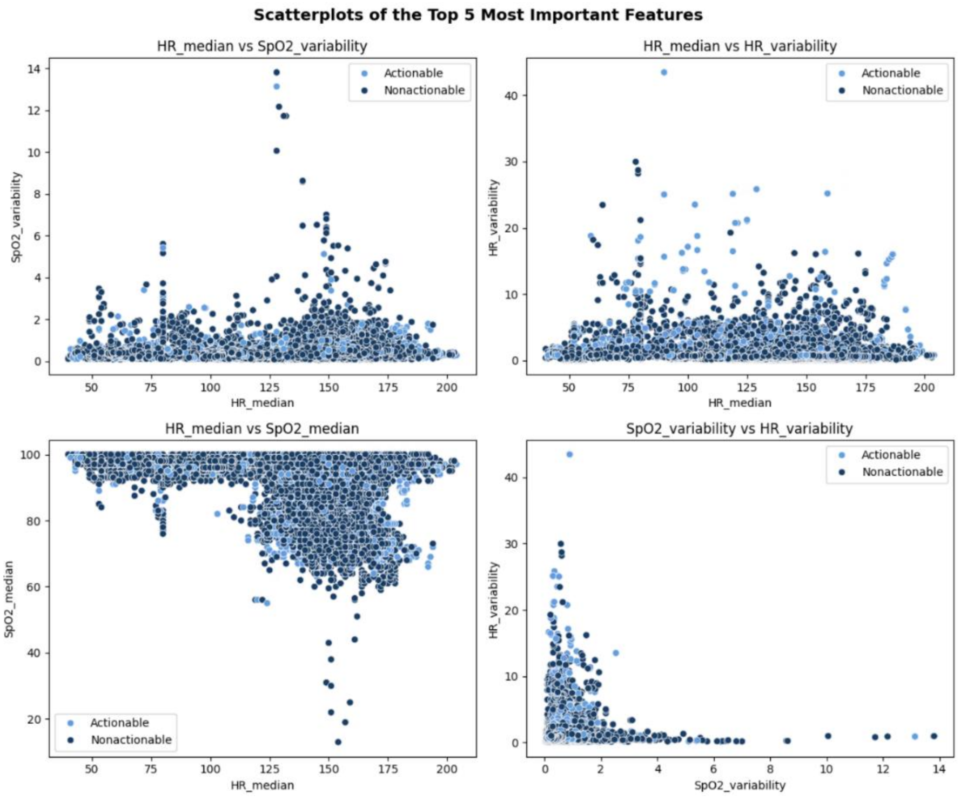
# Appendix VI. Post Hoc Analysis

## Scatterplots



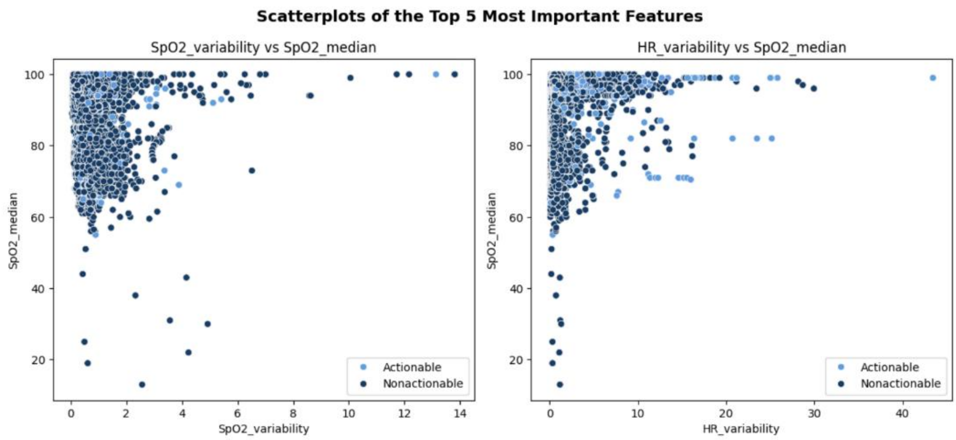

*ART M = mean invasive arterial blood pressure, HR = heart rate, IQR = interquartile range, $SpO_2$ = oxygen saturation*

**Figure 1.** Scatterplots of the five most important features of the DT for the 3-minute alarm window. Each plot shows the relationship between pairs of features (‘ART M IQR’, ‘HR Median’, ‘$SpO_2$ Variability’, ‘HR Variability’ and ‘$SpO_2$ Median’). Alarms classified as actionable are shown in light blue and nonactionable alarms are shown in dark blue. (Figure 1 of 3)




*HR = heart rate, SpO₂ = oxygen saturation*

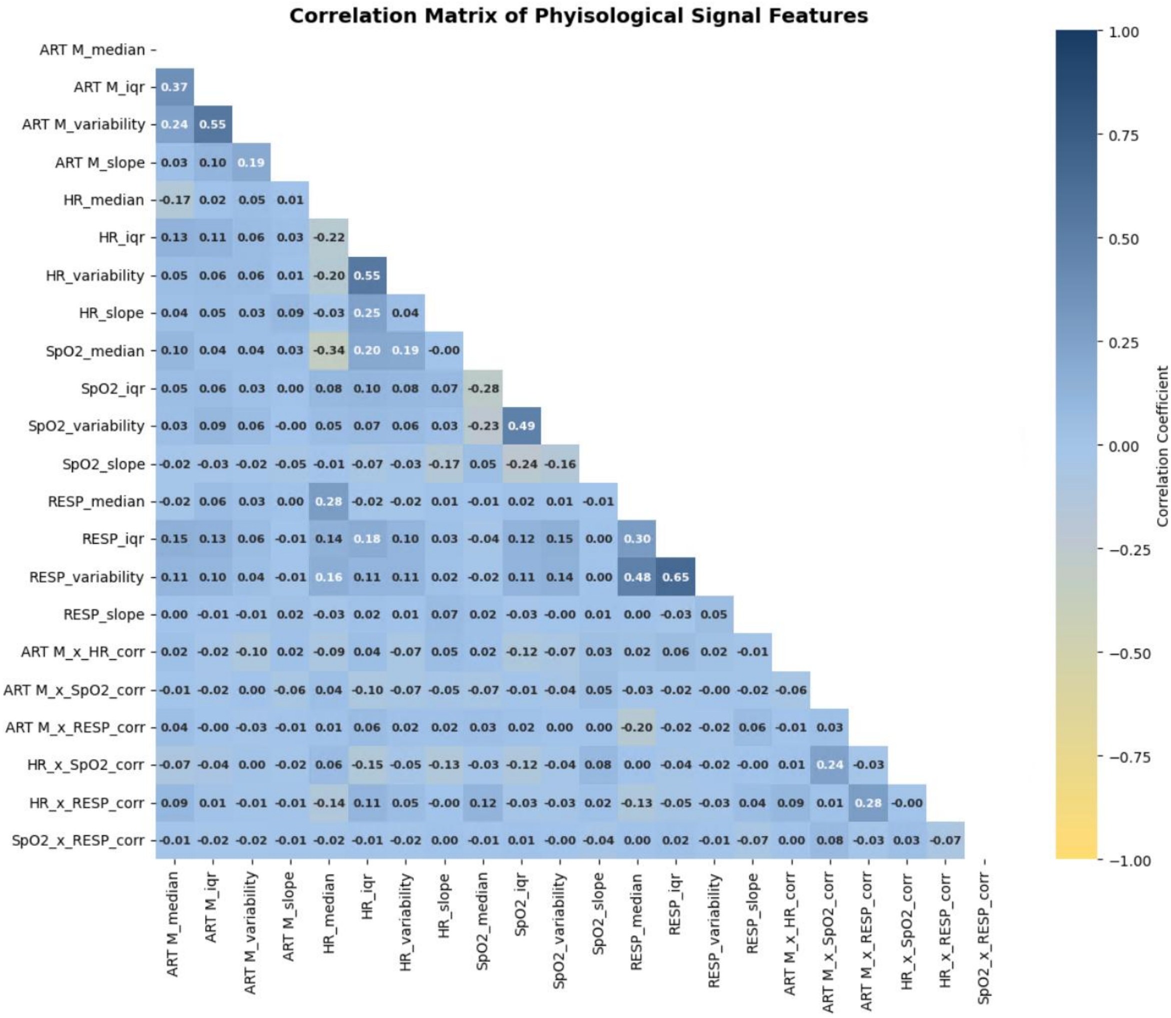
**Figure 2.** Scatterplots of additional feature pairs among the five most important features of the DT for the 3-minute alarm window. Each plot shows the relationship between two of the selected features. Actionable alarms are shown in light blue and nonactionable alarms are shown in dark blue. (Figure 2 of 3).




*HR = heart rate, $SpO_2$ = oxygen saturation*

**Figure 3.** Scatterplots of the remaining feature pairs among the five most important features of the DT for the 3-minute alarm window. Each plot shows the relationship between two of the selected features. Actionable alarms are shown in light blue and nonactionable alarms in dark blue. (Figure 3 of 3).

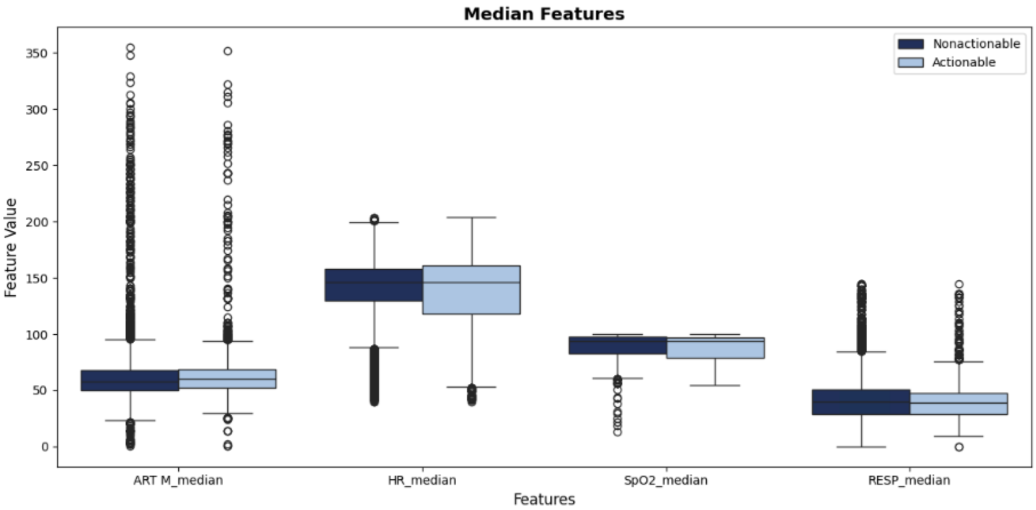
## Correlation Matrix




*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

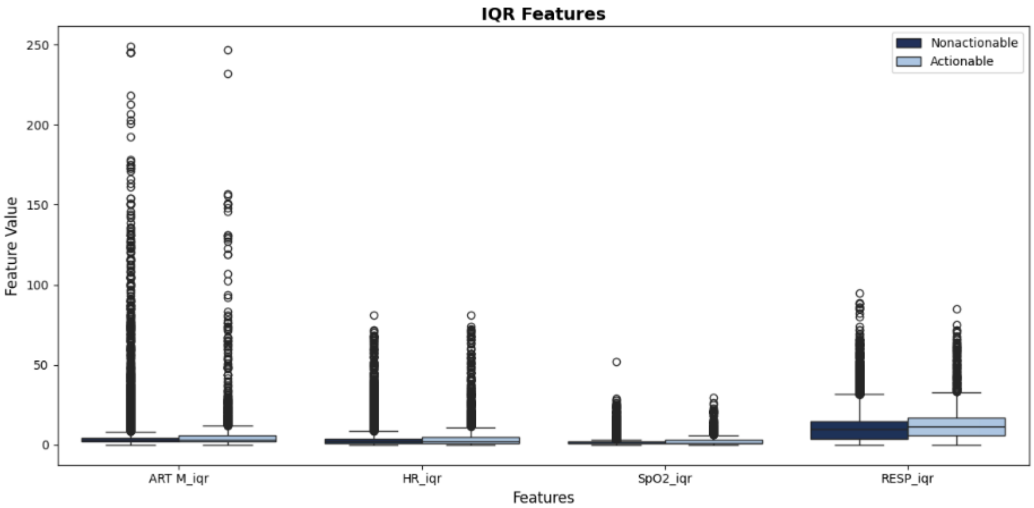
**Figure 4.** Correlation matrix of physiological signal features for the 3-minute alarm window. Pairwise Pearson correlation coefficients are shown for all extracted features, with colour intensity indicating the strength and direction of the correlation.

## Boxplots



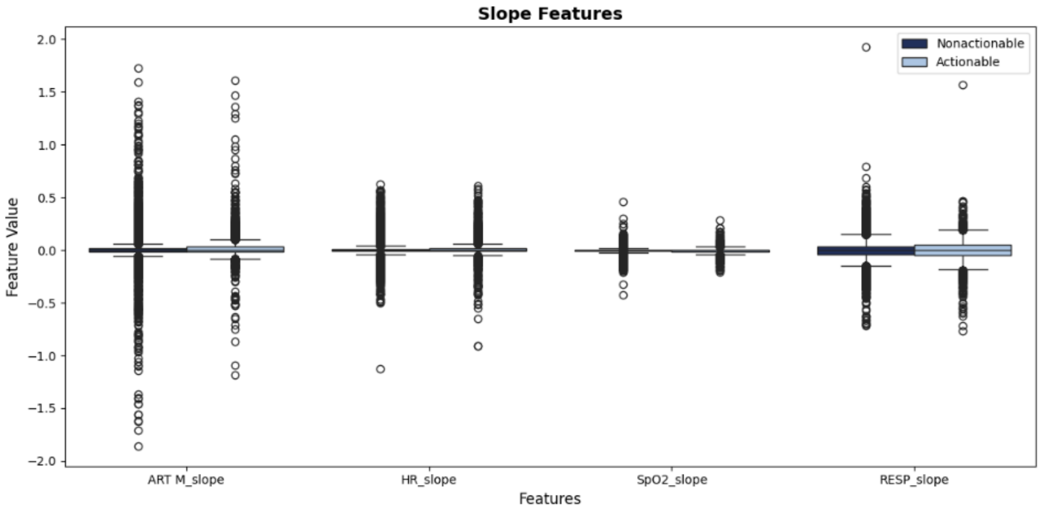

*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 5.** Boxplot distributions of the median values of ART M, HR, $SpO_2$ and RESP features within the 3-minute alarm window.



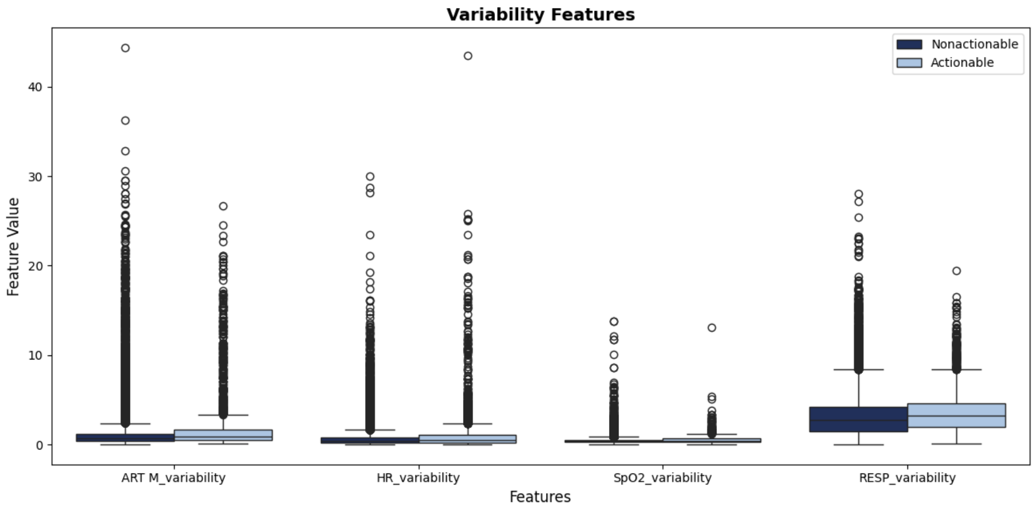

*ART M = mean invasive arterial blood pressure, HR = heart rate, IQR = interquartile range, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 6.** Boxplot distributions of the interquartile range (IQR) values of ART M, HR, $SpO_2$ and RESP features within the 3-minute alarm window.



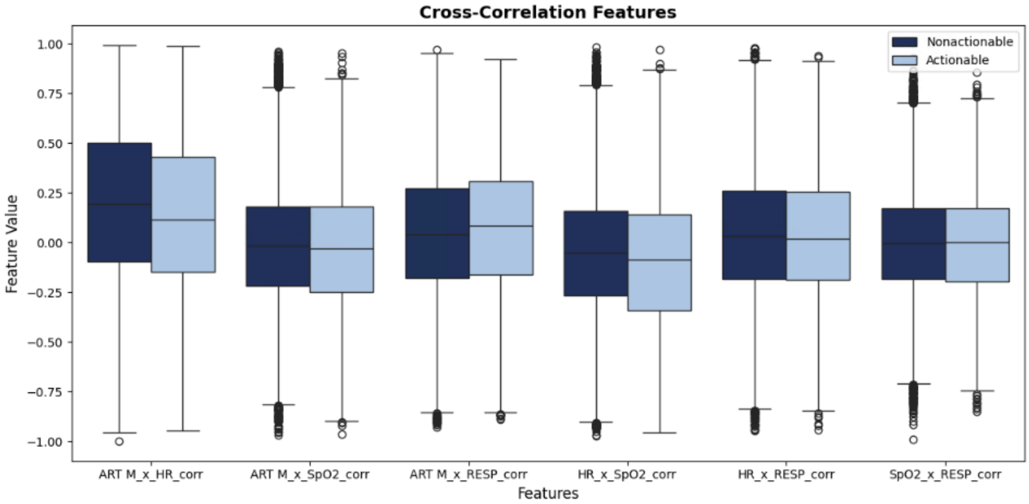

*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 7.** Boxplot distributions of the slope values of ART M, HR, $SpO_2$ and RESP features within the 3-minute alarm window.




*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 8.** Boxplot distributions of the variability values of ART M, HR, $SpO_2$ and RESP features within the 3-minute alarm window.




*ART M = mean invasive arterial blood pressure, HR = heart rate, RESP = respiratory rate, $SpO_2$ = oxygen saturation*

**Figure 9.** Boxplot distributions of cross-correlation values between pairs of ART M, HR, $SpO_2$ and RESP features within the 3-minute alarm window.

# Appendix VII. Stakeholder Input

## Alarm Management Practices and Challenges

### Interview Questions

- What are your perspectives on the use of alarms in the paediatric intensive care unit (PICU)?
- What issues are presently associated with alarm management in the PICU?
- How do these alarms influence your professional practice or personal experience?
- In which situations do you respond to alarms and in which do you decide not to? What actions do you take in each case?

### Findings

Interviews with four nurses identified several themes.

**Alarm thresholds are routinely adjusted at the start of each shift.**

Nurses reported tailoring these settings to the patient's age, underlying pathology and their familiarity with the patient's condition. For example, children with oesophageal atresia often required higher respiratory alarm limits. Narrow thresholds are commonly applied early in a shift to facilitate prompt detection of clinical deterioration and are subsequently widened to reduce clinically irrelevant alarms. Nurses consistently aim to balance the need for early recognition of adverse events with the minimisation of unnecessary noise.

**Repeated exposure to alarm sounds contributes to elevated stress levels and desensitisation, causing nurses to become less responsive and more likely to interpret alarms as non-urgent.**

Nurses described clinical alarms as indispensable, referring to them as their 'eyes and ears' in situations where continuous bedside observation is not feasible. Nonetheless, they reported finishing their shifts feeling cognitively overstimulated and physically fatigued. This fatigue arises not only from the general activity and noise generated by patients and their families and from frequent workflow interruptions, but also from the constant auditory presence of alarms. Moreover, continuous movement between patients is not feasible. Consequently, in certain instances, nurses may delay their response to assess whether the alarm will self-resolve, intervening only if it persists.

**Nurses respond to all high-priority (red) alarms immediately, irrespective of whether these originate from patients under their direct care or from those assigned to colleagues.**

Because the specific source of an alarm is not always immediately identifiable, nursing staff are routinely exposed to alerts generated by multiple patients across the unit. These alarms are typically emitted at a high volume and may even be audible in neighbouring units when doors are open, thereby increasing the overall acoustic burden. Although yellow alarms are equally perceptible to nursing staff, they do not prompt the same immediacy of response as red alarms. It appears that nurses employ an unconscious cognitive filter that enables them to prioritise red alarms, thereby initiating immediate action.

**Alarms are frequently perceived as disruptive during specific clinical situations.**

Nurses emphasised that alarms are particularly disruptive during clinical interventions, as they are already positioned at the bedside and unable to silence them repeatedly because their hands are occupied. A high frequency of nonactionable alarms was also reported, often initiated by patient movement, particularly among those recovering and therefore more physically active. Respiratory alarms were identified as a notable source of unnecessary auditory input, contributing to ambient noise. Nurses expressed a preference for enabling such alarms only when a patient is in respiratory insufficiency. However, even in such cases, the added value of these alarms was considered questionable, as patients in respiratory distress are already closely monitored. Infusion pumps were likewise cited as frequent sources of unnecessary alarms. Nurses suggested that non-critical medications, such as paracetamol, could be limited to a single alert, whereas critical medications should retain the full range of alarm notifications.

**Nurses reported minimal concern regarding duplicate alarms.**
A typical example is when a ventilator alarm is simultaneously displayed on the Dräger monitoring system. In these cases, it is common practice to silence one of the devices' alarms.

**Beyond the impact of alarms on staff, alarms were also described as disruptive to patients and their families.**
Continuous exposure to alarm signals was described as a significant source of stress for parents and has been associated with adverse effects on the patient's recovery. Children were reported to become overstimulated, as they are exposed not only to alarms from their own monitoring equipment but also to those generated by other patients within the unit.

## Perspectives on Alarm Prediction

### Interview Questions

- Is it possible to predict alarms, and if so, on what basis are such predictions made?
- Is it feasible to predict alarms that are specifically actionable?
- Do you observe any recurring patterns in the occurrence of alarms?
- Are you able to distinguish between actionable and nonactionable alarms?
- Do you consider it feasible for an algorithm to identify actionable alarms?
- What is your perspective on the potential implementation of such an algorithm?

### Findings

Insights in this section are based on an interview with two nurses.

**Alarms can often be anticipated through information obtained during handover, observations of patient movement or clinical interventions.**
First, nurses draw on details provided at handover, particularly when multiple incidents, such as episodes of oxygen desaturation, have occurred during the preceding shift. Secondly, alarms may be predicted when a child exhibits restlessness and excessive movement, which increases the likelihood of motion artefacts. Finally, the performance of clinical interventions is recognised as a context in which a substantial number of alarms are likely to be generated.

**The patient's clinical context is of considerable importance in predicting alarms.**
For instance, in the case of a critically ill child, a high probability of adverse events and the consequent need for intervention can often be anticipated. Such assessments are informed not only by physiological signals but also by factors such as the patient's overall appearance, the clinician's prior experience with the patient and an intuitive sense that may accompany clinical evaluation. For example, when a patient is experiencing respiratory insufficiency and blood gas analysis indicates physiological distress, alarms requiring clinical attention can reasonably be expected. Similarly, an elevated HR may be associated with subsequent alarms for reduced $SpO_2$. However, the actionability of such alarms can only be determined by considering the clinical presentation of the patient. A raised heart rate may equally be attributable to a crying child rather than to a pathological process. In such cases, clinicians evaluate behaviour, breathing patterns, skin colour and, where necessary, undertake a structured ABCDE assessment.

**Distinguishing actionable from nonactionable alarms is challenging.**
A profound decrease in, for example, $SpO_2$ is almost always actionable and requires intervention. However, this does not mean that brief desaturation episodes are by definition nonactionable. For instance, if several nonactionable alarms occur within a given timeframe, a response may still be warranted. The frequency of these alarms, the duration of the timeframe in which they occur and the

clinical context in which an intervention is deemed necessary are, however, highly variable and cannot be fully captured by standardised coding rules.

**Limited confidence was expressed in the current potential of a machine learning algorithm designed to transmit only actionable alarms to handheld devices.**
The nurses highlighted the importance of being informed about nonactionable alarms, as these may nevertheless serve as indicators of clinical deterioration. For example, a patient who exhibits a persistently elevated HR with intermittent desaturations that spontaneously resolve may still warrant close observation. Similarly, if four nonactionable alarms of one minute each occur within the span of an hour, this pattern may hold clinical significance and justify intervention. The nurses regarded it as challenging for a model to classify alarms accurately, given the susceptibility of measurements to error and the importance of the patient's clinical context. In addition, concerns were raised about the potential risk of missing important alarms.

**Actionable events, such as episodes of oxygen saturation, can be documented by nurses in HiX.**
One nurse emphasised that clinically important events are sometimes recorded in this way. For example, in one instance a desaturation was accompanied by visible discolouration of the child and the subsequent need for supplemental oxygen. When such an event occurs, the monitoring function allows the clinician to return to the relevant timeframe and document associated parameters, such as HR, $SpO_2$ and RESP. These events are then available for retrospective review within the HiX system. Although not all nurses consistently record them, documentation is considered valuable in cases of clinically significant incidents.

## Strategies for Reducing Alarm Burden

### Interview Question

- What strategies do you recommend for reducing the frequency of alarms?

### Findings

Interviews with four nurses highlighted several proposed strategies.

**Many of the proposed interventions focus on reducing the impact of alarm noise.**
The volume of the alarms was regarded as particularly problematic. One nurse suggested the use of wireless earphones with noise-cancelling functionality, which would enable staff to hear only the alarms of their assigned patient while filtering out environmental alarms and background noise. Another nurse reported already using *Loop Earplugs* to reduce ambient sound during shifts. Suggestions were also made regarding greater reliance on visual alarms, for instance by maintaining an audible alarm at the patient's bedside while displaying a prominent visual signal at the nurses' station. On the NICU, quieter and more acoustically friendly alarms are already in use, which has been reported to create a calmer ward environment. In addition, concern was expressed about conversational noise on the ward, with proposals such as the introduction of a visual indicator (e.g., a large ear symbol that changes colour when the noise level rises).

**It is important to also consider interventions directed towards patients.**
A patient who had undergone prolonged ECMO treatment later reported that the most distressing aspect of the experience was the loud sound of the alarms. Whereas a nurse is able to go home and recover at the end of a shift, patients remain continuously exposed to these loud noises. Addressing this issue is challenging, as it is difficult for patients to wear headphones while lying on their side and children may place earplugs in their mouths. Potential solutions considered included, for example, the use of acoustic screens.

**A need was identified for a foot pedal or push-button to silence alarms during the performance of interventions.**

Such procedures are frequently accompanied by alarms, yet these cannot easily be silenced as both hands are often occupied. Although it is possible to deactivate alarms entirely, this carries the risk that they may not subsequently be reactivated. Similarly, widening alarm thresholds may reduce the number of alarms triggered during interventions, but also entails potential risks. The ability to silence alarms independently was emphasised by nurses, and it was considered essential that any pedal or button be placed in a fixed and safe location to avoid slips or accidents. As nurses usually stand on the right side of the bed, a button attached to the bed frame or a foot pedal in that position was regarded as most practical. The option of a silencing button at the nurses' station was also proposed; however, concern was raised that this might encourage silencing alarms without reviewing the associated waveforms.

**Alarm settings should be adjustable for each infusion pump.**

For pumps administering critical medication, it would be desirable to activate all alarms, whereas for less critical infusions, such as paracetamol, not all alarms would necessarily need to be enabled.

**Nurses expressed a preference for the introduction of audiovisual support via handheld devices at the current stage, in order to familiarise themselves with their use prior to the transition to single-patient rooms.**

Nurses expressed concern that, with the transition to single-patient rooms, they would no longer be able to observe patients immediately and would have reduced control. For this reason, it was considered beneficial to become accustomed to handheld devices in advance. Ideally, handheld devices would be equipped with the functionality to display monitor waveforms and incorporate a camera to provide additional support. This would allow nurses to review the waveforms and assess whether, for example, patient movement is the probable cause of an alarm. It was further proposed that all red alarms be transmitted to all staff members via handheld devices. A buddy system was also suggested, whereby alarms not acknowledged on one device would automatically be forwarded to another nurse.

## Patient Stress and Experience

### Interview Questions

- What are the experiences of patients in the PICU with alarms and how do they reflect upon these in the longer term?
- How do the families of patients experience alarms and how do they perceive them in the longer term?
- What strategies do you consider effective in mitigating the impact of alarms?
- To what extent does prior explanation of the meaning of alarms help to mitigate their impact?
- How do you anticipate single-patient rooms will influence patient and family experiences of alarms?

### Findings

Interview with the psychologist yielded the following insights regarding the impact of alarms on patients and their families.

**Research within the PICU primarily focuses on the impact of stress on parents, as they are continuously exposed to the clinical environment and often feel a heightened sense of responsibility.**

For parents, this experience can be particularly overwhelming, leading them to act with great caution, whereas children are frequently more eager to resume normal activities. Nonetheless, children also demonstrate signs of distress, such as difficulties with sleep, nightmares and challenges in articulating their feelings.

**For parents, the principal source of stress is not the alarms themselves, but rather the severity of their child's illness.**

Other devices, signals and the stress expressed by fellow parents are perceived as less burdensome. Alarms nevertheless contribute to parental strain, as a highly stimulating environment is likely to intensify their emotional burden. Consequently, many parents also experience poorer sleep quality. Additional stressors within the PICU include the need to communicate with multiple healthcare professionals, who may provide differing or even conflicting messages, as well as the inherent difficulty of understanding their child's medical condition.

**Alarms cannot be directly related to patient recovery or prolonged hospital stay.**

A high frequency of alarms often occurs because the patient's condition is deteriorating, which in turn results in a longer stay in the PICU. However, patients are consistently overstimulated by the general noise on the ward and therefore indirectly by alarms as well.

**Auditory stimuli resembling alarms from the PICU frequently evoke recollections of parents' experiences on the unit.**

Families often report that stimuli encountered through television, as well as distinct sounds such as alarm tones or the voices of healthcare professionals, remind them of this period. Comparable responses may also be elicited by particular smells or pieces of music, which families commonly seek to avoid.

**Alarms may also contribute to a perceived sense of safety among parents.**

They provide reassurance that their child is under continuous monitoring and that nursing staff are consistently present on the PICU. When alarms indicate potential clinical changes, parents feel confident that timely intervention can be initiated. This is also related to the transition towards single-patient rooms. Such rooms are often perceived as challenging for parents, as children are no longer continuously visible to nursing staff and parents also value the ease of informal interaction with other families. Nonetheless, single-patient rooms have a positive impact on the child's care, as they reduce background noise and enable more individualised treatment.

**A parental dashboard has been developed to provide contextual information about continuous monitoring and the meaning of alarms.**

For example, a 'beep' may indicate that a measurement has fallen outside a predefined range, which is then displayed on the dashboard together with the corresponding values. The system also clarifies that not every alarm represents a clinical emergency, thereby linking auditory signals with visual information. When multiple alarms occur, these are displayed in real time on the dashboard. Shown on an iPad placed bedside the child's bed, the dashboard presents trends in vital signs, alarm thresholds and, for example, medications being administered.

# Appendix VIII. Literature Review

# Reducing Clinical Alarms to Mitigate Alarm Fatigue in the Paediatric Intensive Care Unit

*C.A. van de Ruit*


## Abstract

**Introduction:** Paediatric intensive care units (PICUs) are high-acuity environments where continuous patient monitoring, facilitated by several medical devices, generates frequent clinical alarms intended to support timely intervention. However, between 87% and 97% of these alarms in PICU settings are nonactionable, contributing to alarm fatigue. This is a condition in which clinicians become desensitised, leading to delayed responses and compromised patient safety. At the PICU of the Erasmus MC Sophia Children's Hospital, this issue has highlighted an urgent need to reduce nonactionable alarms.

**Objectives:** This scoping review investigates strategies to mitigate alarm fatigue by reducing clinical alarms in PICUs. It addresses three objectives: (1) to identify the causes of clinical alarms in PICUs; (2) to evaluate interventions implemented in paediatric hospital settings; and (3) to explore practices from the process industry that may be applicable to PICUs.

**Methods:** A systematic search was conducted across five databases to evaluate current interventions, supplemented by exploratory research to investigate causes of clinical alarms and relevant practices from the process industry.

**Results:** A total of 44 studies were included in the analysis. SpO2 alarms were identified as the predominant source of clinical alarms and perceived as the greatest contributor to alarm fatigue. Key interventions included the adjustment of alarm thresholds, implementation of alarm delays, daily review of alarm parameters, daily electrode replacement with proper skin preparation and the application of machine learning algorithms to improve alarm accuracy. The process industry offers valuable strategies for reducing alarms, including alarm grouping and correlation analysis based on multimodal data.

**Conclusion:** The review emphasises the importance of personalised, data-driven approaches to alarm management. Further research should focus on the development and implementation of a machine learning algorithm that incorporates multimodal vital signs. Incorporating the perspectives of clinicians, patients and families, along with an evaluation of patient safety, will be crucial to ensure effective implementation in PICUs.


*Keywords* Clinical Alarms; Alarm Fatigue; Nonactionable Alarms; Patient Monitoring; Paediatric Intensive Care Unit

## 1. Introduction

Paediatric intensive care units (PICUs) are dynamic and high-pressure clinical environments where continuous patient monitoring and intensive care are essential. Children admitted to these units are critically ill and at risk of acute, life-threatening conditions, necessitating constant observation and (acute) life support and treatment. (1)

In order to facilitate timely clinical assessment and intervention, bedside physiological monitors are employed to continuously track vital parameters such as heart rate (HR), blood pressure (BP) and oxygen saturation ($SpO_2$). (1) These monitors are designed to generate both audible and visual alarms when parameters deviate from predefined thresholds, irrespective of the signal quality or the cause of deviation. (2, 3) Continuous monitoring is widely adopted in critical care settings, as it has been shown to significantly enhance patient safety by ensuring that healthcare professionals are immediately alerted to physiologic changes, enabling prompt and effective response to important deterioration events. (1, 2, 4-8) In addition to physiological monitoring, numerous medical devices commonly utilised in PICUs, such as ventilators, intravenous infusion pumps, feeding pumps and hospital beds, possess their own integrated alarm systems. These devices may signal a variety of problems, including technical malfunction, sensor disconnection and treatment interruption, further increasing the number of alarms in the clinical environment. (2, 9)

In order to minimise the probability of monitors missing indications of deterioration, alarm algorithms and default parameters are frequently configured to maximise sensitivity, often at the expense of specificity. (3) Consequently, this results in a high number of nonactionable alarms being generated by monitors. Nonactionable alarms are defined as both false alarms, which do not reflect the patient's true status, and nuisance alarms, which reflect the true patient status but do not require clinical attention or intervention. (1, 10) False alarms may arise from various factors, including motion artefacts and technical or equipment-related issues. (1) In contrast, nuisance alarms are characterised by deviations that are clinically insignificant. For instance, a patient may experience a desaturation event, falling just below the minimum acceptable pulse oximeter level for a brief period and subsequently return to the established parameters without intervention. (7) In addition, redundant alarms represent a specific subset of nuisance alarms. While these alarms are technically accurate, they appear to duplicate information already communicated by other monitoring devices or alarm sources. For instance, a bedside monitor and a ventilator may both generate alarms in response to the same desaturation event. Despite their clinical validity, these alarms offer no additional information and typically do not necessitate further clinical intervention. (11) Prior research has demonstrated that the proportion of nonactionable alarms has been found to range between 87% and 97% in PICU settings. (12-16)

The responsibility for distinguishing between true, actionable alarms and false or nonactionable ones is typically delegated to clinicians, most often nurses. In the majority of clinical settings, nurses are required to continuously assess whether to respond to alarms from multiple patients or to continue with their current tasks, presuming that the alarms do not necessitate their immediate attention. (7) Excessive noise, including frequent alarms, has been shown to elevate stress levels among staff, potentially impairing their ability to concentrate on tasks and provide optimal care for critically ill patients. (17) Furthermore, frequent exposure to nonactionable alarms has been associated with the development of alarm fatigue, a condition characterised by desensitisation and diminished responsiveness among nurses. (18) This phenomenon is supported by the findings of Bonafide et al. (12), who demonstrated that nurses subjected to a high frequency of nonactionable alarms exhibit slower response times to subsequent alarms. Alarm fatigue has been shown to result in delayed recognition of critical changes in patient status, thereby compromising patient safety and increasing the risk of missed alarms. (12, 13, 17-23)

Alarm fatigue has been a major patient safety concern for the Joint Commission in the United States since 2013, when a report revealed 80 alarm-related deaths between 2009 and 2012. (24, 25) In response, the 2014 Joint Commission National Patient Safety Goal called for hospitals to enhance alarm system safety by developing policies and procedures to mitigate alarm fatigue. (26) Since then, clinical alarm safety has remained a key priority, consistently appearing in the Joint Commission's National Patient Safety Goals and frequently ranking among the "Top 10 Health Technology Hazards," thereby highlighting the ongoing challenges in effectively addressing this issue. (27, 28). In the Netherlands, the Dutch Society for Medical Physics (NVKF) has similarly identified alarm fatigue as a serious alarm-related risk, as outlined in its guideline 'Leidraad medische bewakings- en alarmeringssystemen'. (29) This document highlights the considerable challenge encountered by healthcare providers due to the high prevalence of clinically irrelevant alarms. Alarm fatigue in hospital settings has been shown to share similarities with issues encountered in industrial process control systems. (30) The implementation of strategies employed in these settings to mitigate alarm fatigue may offer valuable insights for hospital environments.

Clinical alarms not only impact healthcare providers but also affect patients and their families. In the PICU, alarm-related noise has been identified as a major source of disruption, contributing to sleep disturbances and increased anxiety among patients and their families. (1) Research has shown that alarm noise is perceived as one of the most disruptive factors by PICU staff and families alike. (31, 32) This disruption can hinder patient recovery, as noise exposure has been linked to various negative physiological responses, including elevated heart rates, increased respiratory rates, and sleep deprivation. (33, 34) Furthermore, continuous monitoring of patients may capture regular physiological fluctuations, often resulting in unnecessary diagnostic workups and interventions, which may contribute to longer hospital stays. (35)

At the PICU of the Erasmus MC Sophia Children's Hospital, there is an urgent need to reduce nonactionable alarms with a view to mitigating alarm fatigue. Pilot data collected between January and June 2023 revealed that the average total number of alarms generated by Dräger monitors across all units per month was 148 557. Of these, high-priority alarms constituted an average of 8.23%, while medium- and low-priority alarms accounted for 66.52% and 25.25%, respectively. (36) High-priority alarms are defined as those necessitating immediate clinical intervention. By contrast, medium- and low-priority alarms comprise the majority and are frequently amenable to reduction through human factors interventions or system-level improvements. Such interventions may include the customisation of alarm limits, or the integration of smart algorithms capable of prioritising or eliminating alarms. (37) The substantial volume of alarms, particularly the predominance of non-urgent alarms, underscores the imperative for strategies that are oriented towards the minimisation of nonactionable alarms.

Kalden et al. (38) indicated that both the patient environment and the methods of alarm transmission contribute to alarm burden. Erasmus MC Sophia Children's Hospital is scheduled to move to a new building, wherein the PICU will transition to single-patient rooms, replacing the multi-patient ward setup. It is important to note that this transition may also compromise nurses' situational awareness in the absence of continuous bedside presence. (38-41) To address this issue, handheld devices will be introduced to support safe alarm management. These devices have the capability to modify alarm pathways, thereby enabling the implementation of effective alarm management interventions aimed at reducing the number of nonactionable alarms.

The objective of this scoping review is, therefore, to investigate strategies to mitigate alarm fatigue by reducing clinical alarms in the PICU. The objective is underpinned by three specific aims:

1. To identify the causes of clinical alarms in the PICU;
2. To evaluate the interventions currently implemented to reduce clinical alarms in paediatric hospitals;
3. To explore and compare practices from the process industry to identify potential lessons that could be applied to the PICU context.

# 2. Methods

### 2.1 Search strategy

In order to achieve the second objective, a systematic literature search was conducted to identify prior interventions employed to mitigate clinical alarms in paediatric hospital settings. In collaboration with the Erasmus MC medical library, search terms were developed for the following databases: Medline ALL, Embase, Web of Science Core Collection, Cochrane Central Register of Controlled Trials and Google Scholar. A detailed list of the search terms is provided in Appendix I. Duplicate studies were removed prior to screening.

### 2.2 Study inclusion

The initial screening of the identified literature was conducted based on titles and abstracts, and following this preliminary assessment, the selected studies underwent a comprehensive full-text review. The screening was conducted using Covidence (42).

To be considered for inclusion in the analysis, studies were required to specifically address interventions aimed at reducing clinical alarms within paediatric hospital settings. Exclusion criteria encompassed studies where interventions for reducing clinical alarms were inadequately described, interventions targeting the reduction of medication alarms, studies exclusively focused on the education of healthcare professionals, approaches solely aimed at reducing ward noise without addressing alarm management and the use of alternative alarm systems, such as handheld devices, as a means of alarm reduction. Additionally, studies that were not available in English or lacked full-text accessibility were excluded.

### 2.3 Supplementary research

Supplementary research was conducted for the first and third objectives in order to investigate the causes of clinical alarms in PICUs, as well as to examine approaches to alarm reduction in the process industry. In contrast to the systematic search strategy previously delineated, this literature search did not adhere to predefined search terms. Instead, an exploratory approach was adopted, whereby relevant studies were identified and reviewed based on their relevance to the objectives of the review.

### 2.4 Data extraction

The included studies were imported into EndNote (43) and data were extracted for analysis. The data extracted included the author, publication year, department, type of patient rooms (single-patient room versus multi-patient ward), described interventions, parameters, validation of interventions and study outcomes. The most commonly employed interventions are presented in the results section.

## 3. Results

### 3.1 Systematic search

The search was conducted on 14 January 2025, resulting in the identification of 1650 records. The detailed results of the search terms for each database are presented in Appendix I. Following the removal of duplicates, a total of 1003 records remained. Of these, 908 were excluded based on title and abstract screening and a further 60 full-text articles were excluded with reasons provided. Consequently, 35 studies were included in the review.

In addition, supplementary research included 12 studies concerning the causes of clinical alarms in PICU and four studies related to alarm reduction strategies in the process industry. However, seven of these studies were found to overlap with the 35 studies that were included through the systematic search, and as a result, they were excluded. The final number of studies included in this scoping review was 44. A summary of the study selection process is presented in Figure 1.

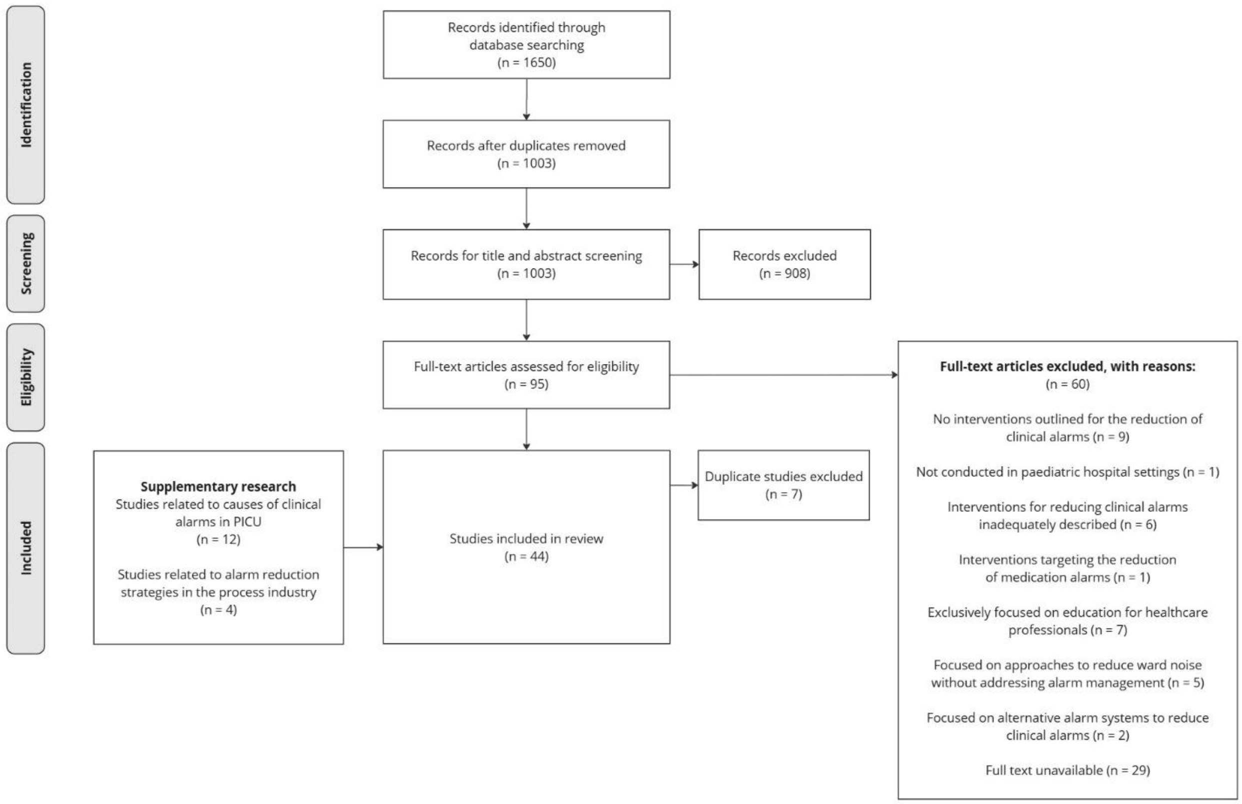



**Figure 1.** *Flow diagram of the study selection process.*

### 3.2 Study characteristics

A review of studies concerning the causes of clinical alarms in PICUs was conducted, and it was found that one study was conducted specifically within PICU settings, five were conducted in NICU settings and six were undertaken across paediatric hospital-wide environments. Comprehensive details concerning the characteristics of the studies evaluating interventions currently implemented to reduce clinical alarms in paediatric hospitals are provided in Appendix II. Of these studies, five were carried out in PICUs, 14 in NICUs and 16 across paediatric hospital-wide settings. Furthermore, seven studies were conducted in units with single-patient rooms, while four were conducted in multi-patient ward settings. The remaining 24 studies did not specify the type of rooms. In addition, the four studies related to alarm reduction strategies within the process industry were all focused on conditions specific to industrial process environments.

### 3.3 Causes of clinical alarms

The majority of studies examining the causes of clinical alarms have been conducted in NICUs, where $SpO_2$ alarms, both high and low, have consistently been identified as the predominant source. (44-46) Surveys conducted in NICU settings further indicated that nonactionable $SpO_2$ alarms were perceived by staff as the greatest contributor to alarm fatigue. (47, 48)

In a study conducted within PICUs, Cvach et al. (49) identified the ten most common causes of monitor alarms. These included low $SpO_2$, poor signal quality, elevated HR, electrocardiogram (ECG) lead disconnection, $SpO_2$ low perfusion, inability to analyse ECG, low HR and irregular HR.

At the paediatric hospital-wide level, Schondelmeyer et al. (50) reported that 33% of total alarms were technical alarms, such as those initiated by artefact or lead failure, while the remaining 67% were related to clinical conditions. Across paediatric hospitals, $SpO_2$ alarms has been identified in numerous studies as the primary contributor to alarm frequency. (13, 51-53). The effectiveness of low $SpO_2$ threshold alarms is influenced by multiple factors, including patient size, skin condition, sensor technology, patient movement and the signal processing algorithm used. Consequently, a significant proportion of low $SpO_2$ alarms do not necessitate clinician intervention. (54) Following $SpO_2$ alarms, the most commonly observed alarm types were those related to elevated HR, elevated RR and failures of ECG and respiratory leads. (50, 52)

### 3.4 Current interventions to reduce clinical alarms in paediatric hospitals

Within the context of the PICU, the reduction of clinical alarms has emerged as a critical concern, given their potential role in contributing to alarm fatigue. This section delineates the various interventions currently employed in paediatric hospitals to decrease the frequency of clinical alarms. The interventions commonly implemented in the reviewed studies include the adjustment of alarm thresholds, the introduction of alarm delays, daily manual procedures and the application of machine learning algorithms.

#### *3.4.1 Alarm thresholds*

The adjustment of alarm thresholds has been identified as a prevalent strategy in the field of paediatric hospital settings, aimed at reducing alarm frequency and enhancing management. (10, 55, 56) The widening of alarm thresholds has been demonstrated to decrease alarm sensitivity to minor or transient deviations, whether due to genuine instability or self-correcting monitoring artefacts. (10, 21, 57, 58) A significant reduction in alarm burden has been demonstrated by lowering the lower $SpO_2$ threshold to 80%-88%. (45, 47, 59)

Beyond widening thresholds, a promising approach identified in the literature is the customisation of alarm settings based on individual patient data, whereby a patient's history and condition are considered when defining alarm parameters. (46, 48) It has been highlighted by numerous studies that reference ranges for paediatric vital signs are often dependent on limited observational data or expert consensus. (60, 61) Bonafide et al. (62) reported that 54% of vital sign measurements in hospitalised children fall outside widely accepted limits, questioning their validity for clinical alarms.

In order to address this issue, age-based and data-driven alarm thresholds have been explored. (49, 63) Johnson et al. (47) developed postmenstrual age-based $SpO_2$ alarm profiles, while other studies have used percentile curves from hospitalised children to define age-appropriate alarm thresholds for HR, RR, and $SpO_2$. (60, 64)

Hravnak et al. (57) and Yang et al. (5) further advocate for the establishment of alarm thresholds based on baseline vital signs, ensuring that alarms are triggered exclusively by clinically significant deviations from an established baseline. Similarly, Herrera et al. (1) recommend adjusting default alarm thresholds to 10% above or below the patient's baseline values.

Schmid et al. (65) finally explored autoregressive models in order to enable real-time, personalised alarm settings by dynamically integrating individual patient data.

#### *3.4.2 Alarm delays*

The implementation of a delay prior to the activation of alarms has been identified as an effective strategy for the reduction of clinical alarms. Alarm thresholds are frequently exceeded during patient care or movement, resulting in the generation of frequent audible alarms. A brief delay has been shown to help minimise recurrent alarms triggered by transient, self-resolving threshold breaches. If the alarm persists beyond this designated period, it is more likely to signify a clinically significant event rather than an artefact. (1, 7, 45, 49, 55, 56, 66)

There are several methods that can be employed for the incorporation of an alarm delay. One such approach involves the implementation of fixed alarm delays. Studies have demonstrated that extending the $SpO_2$ alarm delay to a range of 10 to 60 seconds results in a substantial reduction in the number of false alarms. (45, 47-49, 63, 66, 67) McClure et al. (44) found that combining alarm delays with shorter averaging times further decreases alarm frequency and duration while improving oxygenation data accuracy. Specifically, a 15-second delay with a 2-second averaging time resulted in a 67% reduction in $SpO_2$ alarms.

Nevertheless, the occurrence of simple delays has the potential to result in the failure to detect brief yet critical events. A graduated delay strategy enhances safety and flexibility by ensuring severe deviations trigger alarms more rapidly, while minor, clinically insignificant deviations allow for a longer delay. (57, 65, 68) Exploration of hierarchical time delays and conditional triggers for $SpO_2$ has been undertaken by Yang et al. (5) and Pater et al. (66) In the study by Yang et al. (5), conditional alarm delays varied according to the severity of hypoxaemia, ranging from 30 to 60 seconds for $SpO_2$ levels between 80% and 89%, with immediate notification for $SpO_2$ levels below 60%.

A further refinement of this approach is the SatSeconds™ algorithm, incorporated in the Nellcor™ device manufactured by Medtronic. This algorithm considers both the magnitude and duration of desaturation events when adjusting alarm activation. It has been demonstrated to trigger immediate alarms in cases of severe desaturation episodes, while minor fluctuations result in delayed activation. (1, 7, 69-71)

#### *3.4.3 Daily manual procedures*

In order to achieve an effective reduction in the number of clinical alarms, two key daily manual procedures should be implemented: the structured review of alarm data and routine electrode replacement with proper skin preparation.

<u>Daily review of alarm parameters</u>

Structured daily reviews of alarm data, such as safety huddles, have been identified as effective in reducing alarm burdens for individual patients. (48, 63, 66) Bonafide et al. (72) demonstrated that safety huddles, supported by customised alarm dashboards, reduced alarm frequency among high-alarm paediatric patients. These dashboards highlight prevalent alarm types and settings while incorporating a structured script to guide discussions and document agreed modifications. Dewan et al. (4) also implemented a brief script within safety huddles, integrating an alarm data sheet to focus on key parameters (e.g., RR, HR or $SpO_2$). This strategy contributed to a substantial reduction in alarm activations by modifying alarm parameters based on discussions regarding optimal alarm settings. The visualisation of alarm data has been found to provide valuable insights into alarm frequency and impact, assisting in the comparison of filtering methods such as filtering by alarm duration or type. Smit et al. (73) noted that this visualisation could aid in assessing alarm load per nurse and developing filters to further reduce alarm burdens.

<u>Daily electrode replacement and skin preparation</u>

Daily electrode replacement, in conjunction with proper skin preparation, is essential for minimising alarm frequency. (7, 46, 70) Invalid alarms frequently occur due to inadequate contact between the patient's skin and the electrodes, which can be caused by sensor drying or patient movement. (1) Prolonged electrode use may increase signal impedance and noise due to reduced conductivity, resulting in a higher number of false alarms. (63) When continuous monitoring is required for patients, the correct placement of leads, along with proper skin preparation and the replacement of electrodes every 24 hours, can significantly reduce these alarms. (1, 10, 57, 74-76)

#### *3.4.4 Machine learning algorithms*

Machine learning has demonstrated potential in enhancing the efficacy of clinical alarm systems in paediatric hospital settings. The enhancement involves improving alarm accuracy, reducing false positives and personalising responses to individual patients.

Many studies have adopted a multimodal approach, using data from bedside monitors rather than single-parameter models. Schmid et al. (65) and Sabournia et al. (77) found that incorporating multiple parameters reduces false alarms and can predict patient deterioration. For example, a time-clustering algorithm revealed a correlation between simultaneous cardiovascular and pulmonary alarms. Similarly, Ostojic et al. (78) demonstrated that combining standard physiological monitoring with cerebral oximetry data also reduces false alarms. The employment of decision trees, k-nearest neighbours, naïve Bayes, and support vector machines has yielded optimal specificity, with decision trees demonstrating the highest sensitivity and accuracy. Tsien et al. (79) and Monasterio et al. (80) further confirmed that integrating multiple signals and analysing physiological data, such as $SpO_2$, HR and RR, improves alarm accuracy. Decision trees and support vector machines were employed in these studies, with decision trees selected for their interpretability and their capacity to identify the most informative features from a large set of candidate features.

A number of models have been developed with a view to detecting critical alarms and enhancing their interventions. Cabrera-Quirós et al. (81) and Joshi et al. (82) utilised multimodal vital signs to identify critical events and predict alarm escalation, aiming to enhance alarm response and reduce nurse desensitisation. A further focus of research has been the differentiation of true physiological instability from artefacts through the analysis of vital sign patterns (57, 83), with random forest models showing high accuracy (84).

In the context of personalised approaches, Zhang et al. (85) critiqued the utilisation of generalised algorithms, highlighting their inability to account for the variability amongst individual patients. They proposed using classification trees and neural networks to develop models that adapt to both patient characteristics and the target population. Classification trees were found to be effective in handling fixed attributes, while neural networks were found to be more adept at capturing non-linear patterns, integrating multiple physiological signals to predict adverse events.

### 3.5 Lessons from the process industry in alarm reduction

This section outlines the key lessons derived from industrial alarm management, offering potential applications for improving alarm management in the PICU.

An effective approach involves grouping alarms based on causal relationships within the system. When alarms are interconnected, they should be aggregated and presented as a single alarm issue rather than as multiple individual alarms. (86, 87) For instance, a compressor failure may trigger multiple alarms related to pressure, temperature and flow rate, which would typically be displayed sequentially without indicating their underlying connection. By grouping such alarms, the overall number of individual alarm messages can be significantly reduced while preserving essential information. Schleburg et al. (87) propose an automated method for alarm grouping within process automation systems, demonstrating that this strategy can reduce the number of alarms an operator must manage by 70-80%. Their method integrates historical alarm logs, plant connectivity data and interrelation rules to identify alarms with a common cause, ensuring that they are not removed from the alarm log but instead presented in a structured, hierarchical manner. This hierarchical organisation reduces the number of visible messages at the top level while maintaining information density. Similarly, Rodrigo et al. (30) develop a systematic approach to isolating the causal alarm in the event of an alarm flood, defined as the occurrence of more than ten alarms per ten minutes per operator. By analysing alarm logs, process data and system connectivity, their approach successfully groups related alarms triggered during an alarm flood and identifies the causal alarm.

In addition to alarm grouping, alarm correlation analysis has been proposed as a further strategy for refining alarm management. Rao et al. (88) propose a systematic approach for identifying and analysing alarm correlations by applying pattern mining techniques to historical Alarm and Event logs. This method captures the order of alarm occurrences, enabling the identification of alarm directionality and tracing of abnormality propagation paths. Furthermore, graph visualisation techniques generate correlation networks that facilitate the prioritisation of alarm analysis and reveal process interactions.

## 4. Discussion

### 4.1 Discussion of the findings

This scoping review investigated strategies to mitigate alarm fatigue by reducing clinical alarms in the PICU. The objectives of the review were threefold: firstly, to identify the underlying causes of clinical alarms in the PICU;

secondly, to evaluate existing interventions aimed at reducing alarm frequency in paediatric hospital settings; and thirdly, to explore practices from the process industry to identify potential lessons applicable to the PICU context.

*4.1.1 Causes of clinical alarms*

The findings indicate that $SpO_2$ alarms represent the primary source of clinical alarms in the PICU. Research demonstrates that $SpO_2$ alarms, particularly those indicating low $SpO_2$ levels, account for a substantial proportion of total alarms, many of which are false or nuisance alarms requiring no clinical intervention. Technical issues, including the disconnection of ECG and respiratory leads, have also been identified as contributing factors to the frequency of these alarms. These findings are consistent with pilot data from the PICU at Erasmus MC Sophia Children's Hospital (36), collected between January 2023 and June 2023, which identified $SpO_2$ alarms exceeding thresholds as the most common. Other frequent alarms included HR ECG artefacts, mean invasive blood pressures (ART M) exceeding limits, $SpO_2$ sensor disconnections, and HR surpassing the predefined thresholds. However, ECG and respiratory lead disconnections were less frequent causes.

*4.1.2 Current interventions to reduce clinical alarms in paediatric hospitals*

The most frequently cited interventions for reducing clinical alarms in paediatric hospital settings included the adjustment of alarm limits, the introduction of alarm delays, the application of daily manual procedures, such as the regular review of alarm parameters and optimisation of electrode placement and skin preparation, and the application of machine learning algorithms.

The adjustment of alarm limits is achieved through a number of methodologies. These include widening alarm thresholds, customising alarm parameters based on individual patient data, utilising age-based and data-driven thresholds and employing dynamic adjustment models. It is recommended that alarm thresholds are generated based on the method of using baseline vital signs, because the alarm parameters vary significantly between individuals due to age and the course of the disease. To illustrate this point, alarm thresholds may be set at 10% above or below the patient's baseline values. Consequently, this results in the establishment of appropriate alarm thresholds for the patient, thereby minimising nonactionable alarms.

Furthermore, the implementation of alarm delays prior to alarm activation has been demonstrated to be an effective strategy for reducing nonactionable alarms. Graduated delay strategies are of particular value in that they preserve responsiveness to severe clinical deterioration while filtering out less critical fluctuations.

Another effective approach involves the implementation of daily manual procedures. The utilisation of customised dashboards has been shown to facilitate discussions that are focused on the optimisation of alarm settings and the minimisation of nonactionable alarms. For instance, this could be a topic of discussion during the patient's visit on a daily basis. The dashboard presents data on alarm frequency and identifies the three most common sources, thereby informing decisions regarding potential adjustments to alarm settings or technical interventions, such as improving sensor attachment. In addition, it is recommended that the electrodes be changed on a daily basis, in conjunction with appropriate skin preparation, to enhance the signal quality and minimise the likelihood of nonactionable alarms.

The utilisation of machine learning algorithms, notably those that integrate multimodal vital signs, holds considerable promise for improving alarm management in PICUs. Evidence indicates that machine learning models incorporating multimodal vital signs are preferable to single-parameter models, as they improve the identification of critical events while reducing false alarms.

Numerous studies on alarm management within the medical field have focused on the education of staff members, the reduction of noise levels within the ward and the discontinuation of monitoring when deemed unnecessary. Nevertheless, staff education does not involve technical modifications and only indirectly reduces clinical alarms. While lowering noise levels primarily enhances the overall ward environment, it mitigates the impact rather than directly reducing alarm frequency. Moreover, the discontinuation of monitoring is not a viable option in the PICU, given the critical condition of the patients. Consequently, these interventions were not the focus of this study.

*4.1.3 Lessons from the process industry in alarm reduction*

The findings indicate that the utilisation of multimodal vital signs, as opposed to single-parameter models, is consistent with the principles observed in the process industry, wherein alarm grouping and correlation analysis are conducted based on multiple parameters. In clinical settings, the grouping of alarms has been demonstrated to alleviate alarm overload by aggregating related alarms, thereby preserving essential information and reducing cognitive load. Furthermore, alarm correlation analysis, employing pattern mining on historical data, could assist in prioritising critical alarms and filtering out nonactionable ones. The integration of machine learning algorithms that incorporate multimodal vital signs with alarm grouping or correlation analysis has the potential to enhance their ability to reduce false alarms. As previously mentioned in the introduction, the utilisation of handheld devices in single-patient rooms facilitates the customisation of alarm delivery. Consequently, it becomes feasible to group alarms from multiple vital parameters according to a common cause and subsequently to transmit a consolidated alert to handheld devices.

### 4.2 Relevant research in the Netherlands

In the Netherlands, the reduction of clinical alarms to address alarm fatigue has become a key area of focus.

The guideline ‘Leidraad medische bewakings- en alarmeringssystemen’ provides also recommendations on medical alarm systems. (29) The guideline’s requirements align with the interventions identified in this study. The requirement for alarms to be visually prominent, as outlined in NEN-EN-IEC 60601-1-8/A1, is emphasised and the inclusion of an auditory component is mandatory. It should be possible to set alarm thresholds for different patient profiles, apply filtering and delay functions at the level of departments or profiles and generate department-level alarm reports to optimise workflow. A dashboard should provide retrospective data on alarm frequency, escalation times and alarm types. The software must allow for filtering, grouping and delaying alarms prior to transmission, including selecting which alarms to forward and adjusting prioritisation based on predefined parameters. The utilisation of artificial intelligence for decision support within the alarm system necessitates the establishment of specific policies for maintenance, validation, and testing, in accordance with national AI guidelines for healthcare. These requirements should be considered in future research on implementing strategies to reduce clinical alarms.

The Smart and Silent ICU (SASICU) project (89), an initiative involving institutions such as Erasmus MC and UMC Utrecht, aims to enhance clinical decision support and alarm management. This initiative focuses on enhancing medical device interoperability, directing alarms to appropriate staff, silencing unnecessary alarms and developing AI algorithms to predict Post-Intensive Care Syndrome and monitor critical patient developments.

A Technical Medicine graduate student's thesis (90) contributes to these efforts by analysing the auditory alarm landscape at the Leiden University Medical Center's ICU, with a focus on $SpO_2$ alarms. The study's findings revealed that the majority of these alarms were nonactionable, thereby contributing to an excessive number of alarms. The study proposes a solution in the form of annotation based on clinical context, paving the way for the development of predictive algorithms that can identify and suppress nonactionable alarms. Actionable alarms were found to be associated with significant declines in $SpO_2$, while nonactionable alarms were often transient artefacts. FiO2 increases were the most common response to actionable alarms, though these may also reflect routine care. The study suggests incorporating contextual data, such as electronic medical records and clinician logs, to better distinguish between genuine interventions and routine actions.

Professor Carola van Pul, a Professor of Clinical Physics at Eindhoven University of Technology, has specialised in research related to patient monitoring within NICU and ICU settings, as well as complex patient monitoring and medical alarm systems utilising ICT networks. (91) She contributed to the included study conducted by Cabrera-Quirós et al. (81), which investigated machine learning algorithms for the detection of critical events through the analysis of multimodal vital signs. Additionally, she has been involved in other research exploring machine learning methodologies for analysing multimodal signals and identifying patterns in alarm activation. (92-94) Furthermore, her contributions extend to studies comparing alarm management between multi-patient wards and single-patient rooms, as well as research on workflow optimisation, including the development of protocols for electrode placement, periodic review of alarm parameters and adjustments to delay times for $SpO_2$ alarms. (95, 96) These interventions are consistent with the findings of this study.

At the NICU at Erasmus MC Sophia Children's Hospital, Kalden et al. (38) examined the impact of a modified alarm system in the NICU, which combined handheld devices with filtered and delayed alarms. The system reduced alarms by 84% without affecting the number of critical events, ensuring patient safety was maintained. The study emphasises the prevalence of alarm fatigue, especially in open bay units where nurses are overwhelmed with alarms.

The development of machine learning algorithms based on multimodal vital signs would constitute a valuable contribution to research in the Netherlands, particularly in reducing false alarms and mitigating alarm fatigue in PICU settings.

### 4.3 Limitations

This study aimed to identify interventions that have been implemented to reduce clinical alarms in the PICU. However, several limitations must be acknowledged.

Firstly, only a minority of the included studies were conducted specifically in PICU settings, which limits the generalisability of the findings to this context. Further research in PICU environments is warranted to enhance understanding of interventions that may reduce nonactionable alarms in this setting.

Secondly, this scoping review focused exclusively on interventions implemented within paediatric hospital settings. Consequently, relevant strategies developed and evaluated in adult care environments may have been excluded, despite their potential applicability to paediatric contexts.

Furthermore, the review did not consider how the implementation of interventions may differ between multi-patient wards and single-patient rooms, nor whether the interventions can be integrated in the handheld devices. This should be addressed in future research, particularly in light of the transition to single-patient rooms.

Moreover, the findings are presented without a meta-analysis, as the heterogeneity of outcomes and interventions across the included studies rendered this approach unfeasible.

A further limitation is the absence of a formal quality assessment of the studies, which prevented assessment of their methodological quality. As a result, all studies were treated equally.

Finally, the study did not prioritise patient safety, meaning that no analysis was undertaken to assess the safety of interventions in clinical practice. Further research should incorporate patient safety considerations to enhance the applicability and clinical relevance of the findings.

### 4.4 Recommendations for future research

The findings of the present study suggest that further research at the PICU of Erasmus MC Sophia Children's Hospital should prioritise the development of a machine learning algorithm based on multimodal vital signs, as there is considerable evidence supporting the efficacy of such algorithms in reducing nonactionable alarms. The development and implementation of this algorithm will be the focus of the final thesis project, drawing on relevant studies identified in this scoping review as illustrative examples. It is imperative that the perspectives of medical specialists, nurses, patients and families are incorporated into the development process to ensure that the resulting technological solutions are aligned with their needs. In the event that the implementation of a multimodal algorithm is deemed to be unfeasible, it is recommended that subsequent research explore single-parameter interventions, such as data-driven alarm thresholds and graduated time delays. Following the transition to the new building with single-patient rooms, the potential utilisation of customised dashboards should also be considered.

# 5. Conclusion

In conclusion, alarm fatigue remains a critical issue in PICUs, where the frequent occurrence of nonactionable alarms not only diminishes the responsiveness of healthcare professionals but also comprises patient safety. The high volume of nonactionable alarms contributes significantly to alarm fatigue, leading to desensitisation and delays in responding to actual clinical deterioration. This scoping review has highlighted that $SpO_2$ alarms represent the primary source of clinical alarms in the PICU.

A number of interventions have been shown to be effective in reducing clinical alarms, including the optimisation of data-driven alarm thresholds, the adjustment of graduated time delays, the refinement of daily manual procedures and the application of machine learning algorithms. Moreover, valuable lessons can be drawn from the process industry, where similar challenges of alarm management have been addressed through strategies such as multimodal data integration, alarm grouping, and correlation analysis.

Consequently, further research should focus on the development of a machine learning algorithm to reduce nonactionable alarms based on multimodal vital signs. Incorporating the perspectives of medical specialists, nurses, patients and families will be crucial in the development of this intervention. It is vital to consider the implementation of this intervention in clinical practice, alongside a thorough evaluation of patient safety, to ensure successful reduction of clinical alarms and mitigation of alarm fatigue in the PICU.

# References


1. Herrera H, Wood D. Battling Alarm Fatigue in the Pediatric Intensive Care Unit. Crit Care Nurs Clin North Am. 2023;35(3):347-55.
2. Langga L, Oh J, López D, Blake N, McField E, Hotz J, et al. Prevalence, Proportionality, and Cause of Ventilator Alarms in a Pediatric Intensive Care Setting. Respir Care. 2021;66(4):541-50.
3. Drew BJ, Califf RM, Funk M, Kaufman ES, Krucoff MW, Laks MM, et al. Practice standards for electrocardiographic monitoring in hospital settings: an American Heart Association scientific statement from the Councils on Cardiovascular Nursing, Clinical Cardiology, and Cardiovascular Disease in the Young: endorsed by the International Society of Computerized Electrocardiology and the American Association of Critical-Care Nurses. Circulation. 2004;110(17):2721-46.
4. Dewan M, Wolfe H, Lin R, Ware E. Impact of a safety huddle–based intervention on monitor alarm rates in low-acuity pediatric intensive care unit patients. Journal of hospital …. 2017.
5. Yang JK, Su F, Graber-Naidich A, Hedlin H, Madsen N, DeSousa C, et al. Mitigating Alarm Fatigue and Improving the Bedside Experience by Reducing Nonactionable Alarms. J Pediatr. 2025;276:114278.
6. Moreno RP, Rhodes A, Donchin Y. Patient safety in intensive care medicine: the Declaration of Vienna. Intensive Care Med. 2009;35(10):1667-72.
7. Karnik A, Bonafide CP. A framework for reducing alarm fatigue on pediatric inpatient units. Hosp Pediatr. 2015;5(3):160-3.
8. Welch J, Rayo M, Kanter B, Bagian T, Jacobs K, Shanawani H, et al. Framework for Alarm Management Process Maturity. Biomed Instrum Technol. 2016;50(3):165-79.
9. Poncette AS, Wunderlich MM, Spies C, Heeren P, Vorderwülbecke G, Salgado E, et al. Patient Monitoring Alarms in an Intensive Care Unit: Observational Study With Do-It-Yourself Instructions. J Med Internet Res. 2021;23(5):e26494.
10. Paine CW, Goel VV, Ely E, Stave CD. Systematic review of physiologic monitor alarm characteristics and pragmatic interventions to reduce alarm frequency. Journal of hospital …. 2016.
11. Dewan M, Cipriani L, Boyer J, Stark J, Seger B, Tegtmeyer K. Reducing Redundant Alarms in the Pediatric ICU. Multimodal Technologies and Interaction. 2019;3(1):11.
12. Bonafide CP, Lin R, Zander M, Graham CS, Paine CW, Rock W, et al. Association between exposure to nonactionable physiologic monitor alarms and response time in a children's hospital. J Hosp Med. 2015;10(6):345-51.
13. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. Crit Care Med. 1994;22(6):981-5.
14. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. Crit Care Med. 1997;25(4):614-9.
15. Talley LB, Hooper J, Jacobs B, Guzzetta C, McCarter R, Sill A, et al. Cardiopulmonary monitors and clinically significant events in critically ill children. Biomed Instrum Technol. 2011;Suppl:38-45.
16. Rosman EC, Blaufox AD, Menco A, Trope R, Seiden HS. What are we missing? Arrhythmia detection in the pediatric intensive care unit. J Pediatr. 2013;163(2):511-4.
17. Morrison WE, Haas EC, Shaffner DH, Garrett ES, Fackler JC. Noise, stress, and annoyance in a pediatric intensive care unit. Crit Care Med. 2003;31(1):113-9.
18. Voepel-Lewis T, Parker ML, Burke CN, Hemberg J, Perlin L, Kai S, et al. Pulse oximetry desaturation alarms on a general postoperative adult unit: a prospective observational study of nurse response time. Int J Nurs Stud. 2013;50(10):1351-8.
19. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. AACN Adv Crit Care. 2013;24(4):378-86; quiz 87-8.
20. Wilken M, Hüske-Kraus D, Klausen A, Koch C, Schlauch W, Röhrig R. Alarm Fatigue: Causes and Effects. Stud Health Technol Inform. 2017;243:107-11.
21. Turmell JW, Coke L, Catinella R, Hosford T, Majeski A. Alarm Fatigue: Use of an Evidence-Based Alarm Management Strategy. J Nurs Care Qual. 2017;32(1):47-54.
22. Rasooly IR, Kern-Goldberger AS, Xiao R, Ponnala S, Ruppel H, Luo B, et al. Physiologic Monitor Alarm Burden and Nurses' Subjective Workload in a Children's Hospital. Hosp Pediatr. 2021;11(7):703-10.
23. Lee SJ, Lee YM, Seo EJ, Son YJ. Impact of Hospital Nurses' Perception on Clinical Alarms and Patient Safety Culture on Alarm Management Practice. Int J Environ Res Public Health. 2021;18(8).

24. Medical device alarm safety in hospitals. Sentinel Event Alert. 2013(50):1-3.
25. Jones K. Alarm fatigue a top patient safety hazard. Cmaj. 2014;186(3):178.
26. Commission TJ. Alarm system safety 2013 [Available from: https://www.jointcommission.org/assets/1/18/R3_Report_Issue_5_12_2_13_Final.pdf.
27. Commission TJ. Hospital: 2025 National Patient Safety Goals 2025 [Available from: https://www.jointcommission.org/standards/national-patient-safety-goals/hospital-national-patient-safety-goals/.
28. ECRI. Top 10 Health Technology Hazards Executive Brief 2025 [Available from: https://home.ecri.org/blogs/ecri-thought-leadership-resources.
29. (NVKF) NVvKF. Leidraad medische bewakings- en alarmeringssystemen. 2023.
30. Rodrigo V, Chioua M, Hägglund T, Hollender M. Causal analysis for alarm flood reduction. IFAC-PapersOnLine. 2016;49:723-8.
31. Kaur H, Rohlik GM, Nemergut ME, Tripathi S. Comparison of staff and family perceptions of causes of noise pollution in the Pediatric Intensive Care Unit and suggested intervention strategies. Noise Health. 2016;18(81):78-84.
32. Korniewicz DM, Clark T, David Y. A national online survey on the effectiveness of clinical alarms. Am J Crit Care. 2008;17(1):36-41.
33. Choiniere DB. The effects of hospital noise. Nurs Adm Q. 2010;34(4):327-33.
34. Kudchadkar SR, Aljohani OA, Punjabi NM. Sleep of critically ill children in the pediatric intensive care unit: a systematic review. Sleep Med Rev. 2014;18(2):103-10.
35. Schondelmeyer AC, Dewan ML, Brady PW, Timmons KM, Cable R, Britto MT, et al. Cardiorespiratory and Pulse Oximetry Monitoring in Hospitalized Children: A Delphi Process. Pediatrics. 2020;146(2).
36. Kalden MACM. Dräger Monitor Alarmen - IC Kinderen. Erasmus MC Sophia Children's Hospital; 2023.
37. Bostan I, Ozcan Vieira E, Gommers D, van Egmond R. Annoyance by Alarms in the ICU: A Cognitive Approach to the Role of Interruptions by Patient Monitoring Alarms. Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2022; Torino, Italy2022. p. 133-48.
38. Kalden M, Goos TG, Kalden N, Groenendaal LA, Reiss IKM, Van Bommel J, et al. The Physiological Effects of an Adjusted Alarm Architecture on a Neonatal Intensive Care Unit. Acta Paediatr. 2025.
39. Stevens DC, Helseth CC, Thompson PA, Pottala JV, Khan MA, Munson DP. A Comprehensive Comparison of Open-Bay and Single-Family-Room Neonatal Intensive Care Units at Sanford Children's Hospital. Herd. 2012;5(4):23-39.
40. Almadhoob A. Sound reduction management in the neonatal intensive care unit for preterm or very low birth weight infants. Cochrane Database of Systematic Reviews. 2020(1).
41. Szymczak SE, Shellhaas RA. Impact of NICU design on environmental noise. J Neonatal Nurs. 2014;20(2):77-81.
42. Innovation VH. Covidence systematic review software Melbourne, Australia [Available from: www.covidence.org.
43. The Endnote Team. EndNote. EndNote 20.6 ed. Philadelphia, PA: Clarivate; 2023.
44. McClure C, Jang SY, Fairchild K. Alarms, oxygen saturations, and SpO2 averaging time in the NICU. J Neonatal Perinatal Med. 2016;9(4):357-62.
45. McCauley KE, Schroeder AA, DeBoth TK, Wiebe AM, Bosley CL, Ballweg DD, et al. Reducing Alarm Burden in a Level IV Neonatal Intensive Care Unit. Pediatr Qual Saf. 2021;6(2):e386.
46. Stiglich YF, Dik PHB, Segura MS, Mariani GL. The Alarm Fatigue Challenge in the Neonatal Intensive Care Unit: A "before" and "after" Study. Am J Perinatol. 2024;41(S 01):e2348-e55.
47. Johnson KR, Hagadorn JI, Sink DW. Reducing Alarm Fatigue in Two Neonatal Intensive Care Units through a Quality Improvement Collaboration. Am J Perinatol. 2018;35(13):1311-8.
48. Benincasa CV. Alarm Fatigue Reduction in a Single Patient Room Neonatal Intensive Care Unit. Online Journal of Issues in Nursing. 2024.
49. Cvach M, Kitchens M, Smith K, Harris P. Customizing alarm limits based on specific needs of patients: array.aami.org; 2017.
50. Schondelmeyer AC, Bonafide CP, Goel VV, Blake N, Cvach M, Sucharew H, et al. The frequency of physiologic monitor alarms in a children's hospital. J Hosp Med. 2016;11(11):796-8.
51. Schondelmeyer AC, Brady PW, Goel VV, Cvach M, Blake N, Mangeot C, et al. Physiologic Monitor Alarm Rates at 5 Children's Hospitals. J Hosp Med. 2018;13(6):396-8.

52. Mullen J, Sattari S, Rauch M, Stein F, Roy K, Acorda DE. Utilizing Data and Alarm Champions to Enhance Alarm Management: A Pediatric Quality Improvement Initiative. J Nurs Care Qual. 2024;39(4):369-75.
53. Ruppel H, Makeneni S, Rasooly IR, Ferro DF, Bonafide CP. Pediatric Characteristics Associated With Higher Rates of Monitor Alarms. Biomed Instrum Technol. 2023;57(4):171-9.
54. Nguyen H, Jang S, Ivanov R, Bonafide CP, Weimer J, Lee I. Reducing Pulse Oximetry False Alarms Without Missing Life-Threatening Events. Smart Health (Amst). 2018;9-10:287-96.
55. Cole R, Roderick G, Cheema O. Crying wolf, alarm safety and management in paediatrics: A scoping review. Journal of Advanced …. 2024.
56. Gul G, Intepeler SS, Bektas M. The effect of interventions made in intensive care units to reduce alarms: A systematic review and meta-analysis study. Intensive and Critical Care Nursing. 2023.
57. Hravnak M, Pellathy T, Chen L, Dubrawski A. A call to alarms: current state and future directions in the battle against alarm fatigue. Journal of …. 2018.
58. Lansdowne K, Strauss DG, Scully CG. Retrospective analysis of pulse oximeter alarm settings in an intensive care unit patient population. BMC Nursing. 2016;15(1):36.
59. Berg KJ, Johnson DP, Nyberg G, Claeys C, Ausmus A, Wilkinson E, et al. Reducing the Frequency of Pulse Oximetry Alarms at a Children's Hospital. Pediatrics. 2023;151(5).
60. Goel VV, Poole SF, Longhurst CA, Platchek TS, Pageler NM, Sharek PJ, et al. Safety analysis of proposed data-driven physiologic alarm parameters for hospitalized children. J Hosp Med. 2016;11(12):817-23.
61. Fleming S, Thompson M, Stevens R, Heneghan C, Plüddemann A, Maconochie I, et al. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. Lancet. 2011;377(9770):1011-8.
62. Bonafide CP, Brady PW, Keren R, Conway PH, Marsolo K, Daymont C. Development of heart and respiratory rate percentile curves for hospitalized children. Pediatrics. 2013;131(4):e1150-7.
63. Dandoy CE, Davies SM, Flesch L, Hayward M, Koons C, Coleman K, et al. A team-based approach to reducing cardiac monitor alarms. Pediatrics. 2014;134(6):e1686-94.
64. Jacques SJ, Fauss EK, Sanders JA, Howell MJ, Stein F, Roy K, et al. Patient-centered design of alarm limits in a complex pediatric population. Health Technol. 2017;7(2-3):189-95.
65. Schmid F, Goepfert MS, Reuter DA. Patient monitoring alarms in the ICU and in the operating room: Springer; 2013.
66. Pater CM, Sosa TK, Boyer J, Cable R, Egan M, Knilans TK, et al. Time series evaluation of improvement interventions to reduce alarm notifications in a paediatric hospital. BMJ Qual Saf. 2020;29(9):717-26.
67. Poets CF. Reducing alarms in the NICU. Arch Dis Child Fetal Neonatal Ed. 2018;103(4):F297-F8.
68. Schmid F, Goepfert MS, Franz F, Laule D, Reiter B, Goetz AE, et al. Reduction of clinically irrelevant alarms in patient monitoring by adaptive time delays. J Clin Monit Comput. 2017;31(1):213-9.
69. Brostowicz HM, Rais-Bahrami K. Oxygen saturation monitoring in the neonatal intensive care unit (NICU): Evaluation of a new alarm management. J Neonatal-Perinat Med. 2010;3(3):201-5.
70. Probst PA. Alarm safety in a regional neonatal intensive care unit: search.proquest.com; 2015.
71. Sharkey T. Fact or artifact? Breakthroughs in pulse oximetry alarm management. Neonatal Intensive Care. 2003;16(4):51-3.
72. Bonafide CP, Localio AR, Stemler S. Safety huddle intervention for reducing physiologic monitor alarms: a hybrid effectiveness-implementation cluster randomized trial. Journal of hospital …. 2018.
73. Smit M. Smart Alarm Management System for Patient Monitoring in Intensive Care: studenttheses.uu.nl; 2024.
74. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol. 2012;46(4):268-77.
75. Winters BD, Cvach MM, Bonafide CP, Hu X, Konkani A, O'Connor MF, et al. Technological Distractions (Part 2): A Summary of Approaches to Manage Clinical Alarms With Intent to Reduce Alarm Fatigue. Crit Care Med. 2018;46(1):130-7.
76. Jepsen S. Managing Alarms in Acute Care Across the Life Span: Electrocardiography and Pulse Oximetry. Critical Care Nurse. 2018;38(2).
77. Sabournia P. Analyzing Temporal Patterns of ICU Alarms: A Time-Series Clustering Approach: studenttheses.uu.nl; 2024.
78. Ostojic D, Guglielmini S, Moser V, Fauchere JC, Bucher HU, Bassler D, et al. Reducing False Alarm Rates in Neonatal Intensive Care: A New Machine Learning Approach. Adv Exp Med Biol. 2020;1232:285-90.

79. Tsien CL, Kohane IS, McIntosh N. Multiple signal integration by decision tree induction to detect artifacts in the neonatal intensive care unit. Artificial Intelligence in Medicine. 2000.
80. Monasterio V, Burgess F, Clifford GD. Robust classification of neonatal apnoea-related desaturations. Physiol Meas. 2012;33(9):1503-16.
81. Cabrera-Quirós L, Varisco G, Zhan Z, Long X. Listen to the Real Experts: Detecting Need of Caregiver Response in a NICU using Multimodal Monitoring Signals. … Publication of the 2021 …. 2021.
82. Joshi R, Peng Z, Long X, Feijs L, Andriessen P, Van Pul C. Predictive Monitoring of Critical Cardiorespiratory Alarms in Neonates Under Intensive Care. IEEE J Transl Eng Health Med. 2019;7:2700310.
83. Hravnak M, Chen L, Dubrawski A, Bose E, Clermont G, Pinsky MR. Real alerts and artifact classification in archived multi-signal vital sign monitoring data: implications for mining big data. J Clin Monit Comput. 2016;30(6):875-88.
84. Chen L, Dubrawski A, Wang D, Fiterau M, Guillame-Bert M, Bose E, et al. Using Supervised Machine Learning to Classify Real Alerts and Artifact in Online Multisignal Vital Sign Monitoring Data. Crit Care Med. 2016;44(7):e456-63.
85. Zhang Y. Real-time development of patient-specific alarm algorithms for critical care. Annu Int Conf IEEE Eng Med Biol Soc. 2007;2007:4351-4.
86. Schleburg M, Christiansen L, Fay A, Thornhill NF, editors. Reduction of alerts in automated systems based on a combined analysis of process connectivity and alarm logs2011.
87. Schleburg M, Christiansen L, Thornhill NF, Fay A. A combined analysis of plant connectivity and alarm logs to reduce the number of alerts in an automation system. Journal of Process Control. 2013;23:839-51.
88. Rao HRM, Zhou B, Brown K, Chen T, Shah SL. Alarm correlation analysis with applications to industrial alarm management. Control Engineering Practice. 2024.
89. Smart and Silent ICU 2025 [Available from: https://www.sasicu.eu/.
90. van Kekem MJ. Towards a Quieter ICU: Mapping Alarms and Exploring Opportunities for SpO2 Alarm Reduction: Delft University of Technology; 2025.
91. Technology EUo. Carola van Pul [Available from: https://www.tue.nl/en/research/researchers/carola-van-pul.
92. Cabrera-Quiros L, Kommers D, Wolvers MK, Oosterwijk L, Arents N, van der Sluijs-Bens J, et al. Prediction of Late-Onset Sepsis in Preterm Infants Using Monitoring Signals and Machine Learning. Crit Care Explor. 2021;3(1):e0302.
93. Varisco G, Peng Z, Kommers D, Cottaar EJE, Andriessen P, Long X, et al. Detecting central apneas using multichannel signals in premature infants. Physiol Meas. 2024;45(2).
94. Joshi R, van Pul C, Atallah L, Feijs L, Van Huffel S, Andriessen P. Pattern discovery in critical alarms originating from neonates under intensive care. Physiol Meas. 2016;37(4):564-79.
95. Joshi R, Straaten HV, Mortel HV, Long X, Andriessen P, Pul CV. Does the architectural layout of a NICU affect alarm pressure? A comparative clinical audit of a single-family room and an open bay area NICU using a retrospective study design. BMJ Open. 2018;8(6):e022813.
96. Varisco G, van de Mortel H, Cabrera-Quiros L, Atallah L, Hueske-Kraus D, Long X, et al. Optimisation of clinical workflow and monitor settings safely reduces alarms in the NICU. Acta Paediatr. 2021;110(4):1141-50.

# Appendix I.

To identify literature on methods for reducing alarms in paediatric hospitals, search terms were developed in collaboration with the Erasmus MC Medical Library. This appendix presents the search results and details the search terms used for each database.

**Table 1.** *Results of database searches based on search terms.*

| Database searched | Platform | Years of coverage | Records | Records after duplicates removed |
|---|---|---|---|---|
| Medline ALL | Ovid | 1946 - 2025 | 384 | 381 |
| Embase | Embase.com | 1971 - 2025 | 553 | 284 |
| Web of Science Core Collection* | Web of Knowledge | 1975 - 2025 | 474 | 228 |
| Cochrane Central Register of Controlled Trials** | Wiley | 1992 - 2025 | 39 | 12 |
| Additional Search Engines: Google Scholar*** | | | 200 | 105 |
| **Total** | | | **1650** | **1010** |

*Science Citation Index Expanded (1975-present) ; Social Sciences Citation Index (1975-present) ; Arts & Humanities Citation Index (1975-present) ; Conference Proceedings Citation Index- Science (1990-present) ; Conference Proceedings Citation Index- Social Science & Humanities (1990-present) ; Emerging Sources Citation Index (2005-present)
** Manually deleted abstracts from trial registries
***Google Scholar was searched via "Publish or Perish" to download the results in EndNote.

No other database limits were used than those specified in the search strategies

The following search terms were used to identify relevant studies from the databases:

<u>Medline</u>
(exp "Clinical Alarms"/ OR exp "Alert Fatigue, Health Personnel"/ OR (alarm OR alarms OR ((alert*) ADJ3 (fatigue*))).ab,ti,kf. OR (alert* AND monitor*).ti.) **AND** (exp "Intensive Care Units, Pediatric"/ OR exp *"Infant, Newborn"/ OR (PICU OR NICU OR ((pediatr* OR paediatr* OR child* OR neonate* OR neo-nate* OR newborn* OR new-born*) ADJ4 (intensive*) ADJ4 (care* OR unit* OR ward* OR room* OR department*)) OR ((pediatr* OR paediatr* OR child*) ADJ4 (ICU)) OR ((pediatr* OR paediatr* OR child* OR neonat* OR neo-nat* OR newborn* OR new-born*) ADJ3 (hospital*))).ab,ti,kf.)

<u>Embase</u>
('alarm monitor'/exp OR 'alert fatigue (health care)'/exp OR (alarm OR alarms OR ((alert*) NEAR/3 (fatigue*))):ab,ti,kw OR (alert* AND monitor*):ti) **AND** ('pediatric intensive care unit'/exp OR 'neonatal intensive care unit'/exp OR 'newborn'/de/mj OR (PICU OR NICU OR ((pediatr* OR paediatr* OR child* OR neonate* OR neo-nate* OR newborn* OR new-born*) NEAR/4 (intensive*) NEAR/4 (care* OR unit* OR ward* OR room* OR department*)) OR ((pediatr* OR paediatr* OR child*) NEAR/4 (ICU)) OR ((pediatr* OR paediatr* OR child* OR neonat* OR neo-nat* OR newborn* OR new-born*) NEAR/3 (hospital*))):ab,ti,kw)

<u>Web of Science</u>
(TS=(alarm OR alarms OR ((alert*) NEAR/2 (fatigue*))) OR TI=(alert* AND monitor*)) AND (TS=(PICU OR NICU OR ((pediatr* OR paediatr* OR child* OR neonate* OR neo-nate* OR newborn* OR new-born*) NEAR/4 (intensive*) NEAR/4 (care* OR unit* OR ward* OR room* OR department*)) OR ((pediatr* OR paediatr* OR child*) NEAR/4 (ICU)) OR ((pediatr* OR paediatr* OR child* OR neonat* OR neo-nat* OR newborn* OR new-born*) NEAR/2 (hospital*))))

Cochrane CENTRAL

((alarm OR alarms OR ((alert*) NEAR/3 (fatigue*))):ab,ti,kw OR (alert* AND monitor*):ti) **AND** ((PICU OR NICU OR ((pediatr* OR paediatr* OR child* OR neonate* OR neo NEXT/1 nate* OR newborn* OR new NEXT/1 born*) NEAR/4 (intensive*) NEAR/4 (care* OR unit* OR ward* OR room* OR department*)) OR ((pediatr* OR paediatr* OR child*) NEAR/4 (ICU)) OR ((pediatr* OR paediatr* OR child* OR neonat* OR neo NEXT/1 nat* OR newborn* OR new NEXT/1 born*) NEAR/3 (hospital*))):ab,ti,kw)

Google Scholar

alarm|alarms PICU|NICU|'pediatric|paediatric|neonatal intensive care'|'children|pediatric|paediatric|neonatal hospital'

# Appendix II.

**Table 1.** *Study characteristics*

| Author | Publication year | Department | Type of rooms | Interventions | Parameters | Validation of interventions | Results |
|---|---|---|---|---|---|---|---|
| Benincasa et al. | 2024 | NICU | Single-patient rooms | Extension of alarm delays to 60 seconds, daily review of pulse oximeter prescriptions, pulse oximeter parameters for specified patient populations and education regarding alarm settings | $SpO_2$ | Survey and alarms rates per patient day | More than 40% alarm reduction within four months. There was an increase in nurse satisfaction. |
| Berg et al. | 2023 | Paediatric hospital wide (NICU and PICU excluded) | Single-patient rooms | Modification of default alarm $SpO_2$ limits on monitors to <88% | $SpO_2$ | $SpO_2$ alarm rates per patient day, alarms rates for $SpO_2 \geq 88\%$ per patient day | Relative reduction of $SpO_2$ alarms per patient day was 17.93% and relative reduction for $SpO_2$ alarms ≥ 88% per patient day was 35.8%, both between January 2021 and June 2022. |
| Bonafide et al. | 2018 | Paediatric hospital wide | NA | Structured safety huddle review of alarm data from high-frequency patients, with discussions on reducing alarms | NA | Unit-level alarm rates per patient day, individual patient alarm rates per patient day | Safety huddle-based alarm discussions did not influence unit-level alarm rates due to low intervention dose but were effective in reducing alarms for individual patients. |
| Brinks | 2015 | PICU | NA | Priority levels based on urgency, specific parameter settings, alarm delay, naïve signal filtering, graphical monitor interface design and multi-parametric approaches (clinical decision support systems and machine learning techniques) | NA | NA | NA |
| Brostowicz et al. | 2010 | NICU | NA | SatSeconds™ alarm feature (extended alarm delays) | $SpO_2$ | Percentage of alarm rate reduction using various SatSeconds™ settings | Overall decrease in alarms by 40% with a setting of 50 SatSeconds™. |
| Cabrera-Quirós et al. | 2021 | NICU | Single-patient rooms | Machine learning algorithm to detect urgent moments using multimodal vital signs from bedside patient monitors and caregiver alarm responses | ECG and $SpO_2$ | ROC and AUC curves | General detection of caregiver response with a mean AUC of 0.82. Classifiers perform better in distinguishing alarms requiring no immediate response from those that do when trained solely with stable and genuinely deteriorating samples. |

| Author | Publication year | Department | Type of rooms | Interventions | Parameters | Validation of interventions | Results |
|---|---|---|---|---|---|---|---|
| Cole et al. | 2024 | Paediatric hospital wide | NA | Scoping review of interventions to improve alarm management and reduce alarm fatigue: changing alarm parameters (modifying alarm limits and extending alarm delay), clinician education, communication and planning, technology changes, alarm ordering, standardisation and guidelines | NA | NA | Most studies focused on changing alarm parameters. |
| Cvach et al. | 2017 | ICUs, IMCUs, paediatrics and emergency departments | NA | Alarm customization (alarm parameter limits based on patient need), alarm delays ($SpO_2$: 15-second; ST: 1-minute) and patient profiles (age range or disease conditions) | NA | NA | NA |
| Dandoy et al. | 2014 | Paediatric hospital wide (BMTU) | NA | Cardiac Monitor Care Process (CMCP): age-appropriate monitor parameters, daily electrode replacement, individualized assessment of cardiac monitor parameters, appropriate discontinuation method, customized monitor delays ($SpO_2$ alarm delay from 5 to 10 seconds) and increased threshold settings (increased high RR limit) | Cardiac monitor parameters (e.g., $SpO_2$, RR) | Alarm rates per patient day | CMCP resulted in an 80% decrease in alarm rates per patient day after the full implementation of the intervention process. They achieved a 55% reduction in the number of alarms through human factor-dependent processes, such as changing leads, with a further 25% reduction attained through the customisation of monitor settings. |
| Dewan et al. | 2017 | PICU | Single-patient rooms | Data-driven monitor alarm discussions in safety huddles and monitor parameter adjustments | NA | Priority alarm activation rate per 24 hours before and after the huddle | Reduction of 116 priority alarms (95% CI, 37-194) per 24 hours (P=.004) |
| Goel et al. | 2016 | Paediatric hospital wide (general medical and surgical units) | NA | Modification of alarm limits using data-driven, age-stratified 5$^{th}$ and 95$^{th}$ percentile values | HR and RR | Proportion of out-of-range observations with the intervention versus the current (NIH) reference range | 55.6% reduction in out-of-range measurements |

| Author | Publication year | Department | Type of rooms | Interventions | Parameters | Validation of interventions | Results |
|---|---|---|---|---|---|---|---|
| Gul et al. | 2023 | ICUs (including NICUs and PICUs) | NA | Meta-analysis and systematic review on interventions for alarm management: electrode evaluation/replacement, suitability assessment, customised alarm parameters and thresholds, training, multidisciplinary communication, alarm delays, auxiliary screens alarm notification, volume adjustments, skin preparation, clinical workflow adjustments, identifying primary alarm response responsibility, alarm analysis sharing, disposable ECG leads, recurrent alarm elimination and physical reminders | NA | Heterogeneity analysis, random effects model (effect size indicating the clinical usability of the study results) | The studies were heterogeneous and showed varied distributions. The combined effect size for reducing alarms was weak, with minimal impact in clinical settings. The level of evidence for the effectiveness of interventions in reducing alarm numbers remained low. |
| Herrera et al. | 2023 | PICU | NA | Daily ECG electrode changes, pulse oximeter sensor replacement as needed, age-related parameters (1st and 99th percentile for age), patient-specific parameters (alarm defaults ± 10% of baseline), alarm delays (SatSeconds ™), secondary notification systems and monitor watchers | HR, RR, $SpO_2$, BP and temperature | NA | NA |
| Hravnak et al. | 2018 | ICU | NA | Adequate skin preparation, daily electrode changes, disposable wiring and sensors, customised alarm thresholds and delays, environment/education/organization strategies, secondary notification systems, alarm suppression and artifact discrimination algorithms, an integrated monitoring system displaying a single risk score using neural networking and machine learning algorithms | NA | NA | More accurate and unbiased alarm fatigue metrics need to be developed to assess the impact of interventions. Currently, evaluation is based on comparing alarm rates or alarm fatigue surveys. However, these metrics should also incorporate safety, quality, process of care and human factors outcomes. |
| Jacques et al. | 2017 | Paediatric hospital wide (PCU) | NA | Age-related alarm limits derived from hospitalised population data by characterising percentile curves for parameters | HR, RR and $SpO_2$ | Percentile curves of parameters (current alarm limits versus actual physiological data) and histograms of parameter distributions for the 6–12-month-old population | Comparison of current alarm limits with actual physiological data highlights a mismatch between alarm settings and patient physiology. Age-based recommended alarm limits, derived from healthy cohorts, can result in higher-than-desired alarm loads in the paediatric population. |

| Author | Publication year | Department | Type of rooms | Interventions | Parameters | Validation of interventions | Results |
|---|---|---|---|---|---|---|---|
| Johnson et al. | 2018 | NICU | Multi-patient ward | Quality improvement study: cycles included lowering the low alarm limit (<85%), increasing the low alarm delay (15 seconds), developing postmenstrual age-based alarm profiles and updating bedside visual reminders | $SpO_2$ | Total number of nonactionable $SpO_2$ alarms per patient per hour and number of nonactionable low $SpO_2$ alarms per patient per hour | After the improvement cycles, the mean total nonactionable alarms per patient per hour decreased by 78% from baseline and the mean number of nonactionable low alarms per patient per hour decreased by 80% from baseline. |
| Joshi et al. | 2019 | NICU | Single-patient rooms | Machine learning model using a multiparametric approach to predict whether a yellow alarm will escalate to a red alarm within a short time window | HR, RR and $SpO_2$ | AUROC (performance of alarm classification) | Performance of boosted trees on the test set: AUROC of 0.89 with a sensitivity of 0.33 and specificity fixed at 0.98. Implementing this model could reduce the total number of auditory alarms by nearly 80% while increasing the number of red alarms by 7%. |
| Karnik et al. | 2015 | Paediatric hospital wide | NA | The framework conceptualised various interventions, including monitoring only patients at significant risk of life-threatening events, skin preparation and daily electrode changes, age-based alarm limits for heart and respiratory rates (1st and 99th percentiles), $SpO_2$ alarm limit <80%, allowing nurses to adjust alarm parameters within a ± 10% margin, increasing alarm delays (SatSeconds ™), secondary notification systems and monitor watchers | HR, RR and $SpO_2$ | NA | NA |
| Lilja et al. | 2017 | NICU | NA | Neural network approach for the automatic detection of acoustic alarms | Period duration, frequency components, samples and frames of acoustic alarms | Evaluation of detection performance by frame-level metrics (MR, FAR and EER) and event-level metrics (PB-ERR) | Both generic and class-specific models were proposed. The class-specific input model, which utilises knowledge of alarm frequency components, produced better results than the generic input model. |
| McCauley et al. | 2021 | NICU | Multi-patient ward | Quality improvement study: lowering alarm limits to 88% and 86%, trialling a new pulse oximeter and increasing the low-limit yellow alarm delay to 20 seconds | $SpO_2$ | Alarms per patient hour, staff survey | The improvement cycles reduced yellow $SpO_2$ self-resolving alarms by 64%. There was a reduced need for staff to modify alarm limits following the interventions and parental concern regarding staff responsiveness to alarms decreased. |

| Author | Publication year | Department | Type of rooms | Interventions | Parameters | Validation of interventions | Results |
|---|---|---|---|---|---|---|---|
| McClure et al. | 2016 | NICU | Multi-patient ward | $SpO_2$ averaging times of 2, 8 and 16 seconds and a 15-second alarm delay for $SpO_2$ ranges of 88%-95% and 70%-98% | $SpO_2$ | Events per day per infant, mean seconds per event and seconds per day per infant | Longer averaging times mask the number and severity of aberrant oxygenation events in preterm infants without reducing total alarm time. Incorporating an alarm delay with shorter $SpO_2$ averaging times can reduce alarm frequency and duration, enabling more accurate assessment of oxygenation. Implementing a 15-second alarm delay to 2-second $SpO_2$ averaging in this analysis decreased $SpO_2$ alarms by 67%. |
| Monasterio et al. | 2012 | NICU | NA | Multimodal analysis framework to reduce the false alarm rate in neonatal apnoea monitoring. | $SpO_2$, HR, RR and signal quality of ECG, impedance pneumogram and photoplethy-mographic signals | Classifier performance: sensitivity, specificity and accuracy | Optimal classification performance was achieved with a combination of 13 features, yielding sensitivity, specificity and accuracy of 100% in the training set and sensitivity of 86%, specificity of 91% and accuracy of 90% in the validation set. The most useful feature for false alarm detection was the minimum HR within the 30-second interval before a desaturation. |
| Nguyen et al. | 2018 | Paediatric hospital wide | NA | AdaBoost machine learning classifier with a reject option, specifically tuned to avoid silencing valid alarms while suppressing as many false low $SpO_2$ alarms as possible | $SpO_2$, HR and RR | Classifier performance: specificity and sensitivity | The classifier is able to silence 23.12% of false $SpO_2$ alarms while maintaining clinically significant alarm sensitivity at 99.27% |
| Ostojic et al. | 2020 | NICU | NA | Machine learning algorithm that intelligently analyses data from standard physiological monitoring with cerebral oximetry data. Four algorithms were used to categorise the alarms: Decision tree, 5-nearest neighbours, naïve Bayes and support vector machine | HR, $SpO_2$ and $StO2$ | Classifier performance: specificity, sensitivity and accuracy | All four approaches achieved a specificity of >99%. The decision tree showed the highest sensitivity (87.52%) and accuracy (98.67%). Cerebral oximetry data enhanced classification accuracy. |
| Paine et al. | 2016 | Paediatric hospital wide (one of eight intervention studies included paediatric patients) | NA | Systematic review evaluating the following interventions: widening alarm parameters, alarm delays, reconfiguring alarm acuity, secondary notifications, daily ECG electrode changes or disposable ECG wires, universal monitoring in high-risk populations and timely discontinuation of monitoring in low-risk populations | $SpO_2$, ECG-parameters and BP | NA | Widening alarm parameters, implementing alarm delays and using disposable ECG lead wires and/or changing electrodes daily are the most effective interventions for reducing alarms. |

| Author | Publication year | Department | Type of rooms | Interventions | Parameters | Validation of interventions | Results |
|---|---|---|---|---|---|---|---|
| Pater et al. | 2020 | Paediatric hospital wide (ACCU) | Single-patient rooms | Quality improvement study: technology interventions, alarm delays based on alarm limits, escalation algorithms to reduce initial and secondary notifications, team discussions of patient-specific vital sign parameters and the need for continuous monitoring, turn off in-room alarm volume and lead changes every 24 hours | $SpO_2$, HR and RR | Reduction in initial alarm notifications per monitored bed per day, averaged per month, and nursing satisfaction survey | The number of alarm notifications was reduced by 68% over a 3.5-year period using quality improvement methodology. Alarm notifications decreased successfully, leading to improved nursing satisfaction, with no negative impact on patient safety. |
| Poets et al. | 2018 | NICU | NA | Staff education, modification of alarm limits, alarm delays and averaging times | $SpO_2$ and ventilator parameters | NA | Appropriate parameter settings and the introduction of alarm delays are likely to lead to a significant reduction in alarm rates for both ventilators and patient monitors, particularly pulse oximeters. |
| Probst et al. | 2015 | NICU | NA | Evidence-based practice (EBP) intervention protocols: monitoring parameter EBP protocol and electrode lead and probe changing EBP protocol | $SpO_2$, HR, RR and BP | Monitoring parameter EBP protocol: average number of nuisance alarms per hour per bed; electrode lead and probe changing EBP protocol: average number of false alarms per hour per bed | The monitoring parameter EBP protocol reduced the average nuisance alarms per hour per bed by 85%. The electrode lead and probe changing EBP protocol decreased the average number of false alarms per hour per bed by 57%. The percentage of clinical alarms increased post-intervention, indicating improvement in alarm safety according to the model. |
| Sabournia et al. | 2024 | NICU and PICU | NA | Machine learning approach (time-series clustering algorithm with dynamic time warping) to identify subgroups of ICU patients and examine the relationships between different alarm types | Cardiovascular and pulmonary physiological parameters | Temporal analysis of alarm data | The study identified the simultaneous occurrence of cardiovascular and pulmonary physiological alarms, suggesting a correlation between these two. Additionally, patterns of stable alarms followed by surges provide early warning of patient functional decline, aiding in resource optimisation, prioritisation of interventions and the tailoring of monitoring protocols to individual patient needs. |

| Author | Publication year | Department | Type of rooms | Interventions | Parameters | Validation of interventions | Results |
|---|---|---|---|---|---|---|---|
| Schmid et al. | 2013 | ICU and OR (including paediatric hospital) | NA | Phase specific settings to reduce false alarms (e.g., for specific ICU settings), integrated alarm validation, alarm delays (e.g., 14-second, 19-second or graduated delays), statistical approaches for artifact reduction (autoregressive models, self-adjusting thresholds, statistical process control, median filter) and artificial intelligence (rule-based expert systems, neural networks, fuzzy logic, Bayesian networks) | NA | NA | Many promising approaches using statistical methods and artificial intelligence have been developed to reduce false alarms, yet no obvious changes in false alarms have been observed in clinical practice. |
| Smit et al. | 2024 | PICU | NA | Visual system displaying alarms per bed and alarm load for nurses. Filtering options include duration of alarms (1-60 second delay) and input-based filtering (select when the alarm is silenced based on input). | NA | Percentage of filtered alarms and alarm rate per nurse per hour. | The percentage of filtered alarms increases with the application of a 1-second filter (9.5%) and a 3-second filter (28.9%). The number of technical alarms filtered out (10.1% and 30.4%) exceeds the number of physiological alarms filtered out (8.6% and 26.7%). The filters reduce the average alarm load from 28 alarms per nurse per hour to 25 alarms per nurse per hour with the 1-second filter and to 20 alarms per nurse per hour with the 3-second filter. |
| Stiglich et al. | 2024 | NICU | Multi-patient ward | Alarm Management Program (AMP): correct and individual setting of alarm limits based on each patient's orders in the electronic medical record, proper use of each device (alarm signal automatically stops when the triggering event ceases), role assignment for adjusting alarm limits, sensor checking schedule and brief alarm pauses during patient manipulation | $SpO_2$, BP, HR, tidal volume, peak inspiratory pressure, positive end-expiratory pressure and leak percentage | Proportion of true and nonactionable alarms, response time and variables associated with nonactionable alarms | The proportion of true alarms before and after AMP was 31% versus 57% ($p=0.001$), while the proportion of nonactionable alarms was 69 versus 43% ($p=0.001$). Median response time was significantly reduced (37 seconds versus 12 seconds; $p=0.001$). Neonates with less intensive care needs exhibited a higher proportion of nonactionable alarms and a longer response time. The need for respiratory support was significantly associated with true alarms ($p=0.001$). In the adjusted analysis, response time ($p=0.001$) and respiratory support ($p=0.003$) remained associated with nonactionable alarms. |

| Author | Publication year | Department | Type of rooms | Interventions | Parameters | Validation of interventions | Results |
|---|---|---|---|---|---|---|---|
| Tsien et al. | 2000 | NICU | NA | Detection algorithm for artifact patterns across multiple physiological data signals using decision tree induction | HR, BP, pCO2 and pO2 (collected transcutaneously) | Classifier performance: sensitivity, specificity, PPV, accuracy, and AUROC | HR decision tree model: sensitivity 65.4%, specificity 99.8%, PPV 91.4%, accuracy 98.5%, AUROC 92.8%.<br>BP decision tree model: sensitivity 57.7%, specificity 99.9%, PPV 90.0%, accuracy 98.9%, AUROC 89.4%.<br>pCO2 decision tree model: sensitivity 82.5%, specificity 99.2%, PPV 84.4%, accuracy 98.3%, AUROC 93.3%.<br>pO2 decision tree model:<br>sensitivity 87.5%, specificity and PPV 100%, accuracy 99.8%, AUROC 99.9%. |
| Yang et al. | 2025 | PICU and ACCU | Single-patient rooms | Quality improvement study: hierarchical time delays and conditional alarm triggers for $SpO_2$ and alarm limit modifications for RR and PVCs | $SpO_2$, RR and PVCs | Median alarm rates per monitored patient day and surveys | The median numbers of alarms per monitored patient day decreased by 75% in PICU ($P < .001$) and 82% in the ACCU ($P<.001$) with a sustained effect at the 2-year follow-up. Nursing surveys reported improved capacity to respond to alarms and fewer perceived nonactionable alarms. However, family surveys did not demonstrate improved sleep quality. |
| Zhang | 2007 | PICU | NA | Patient-specific alarm algorithms using machine learning techniques: classification tree learning and neural network learning | All consistently and frequently monitored parameters (i.e., HR, $SpO_2$, RR and BP) | Classifier performance: sensitivity, specificity, PPV and accuracy | Neural networks: sensitivity 0.96, specificity 0.99, PPV 0.79, accuracy 0.99.<br>Classification trees: sensitivity 0.84, specificity 0.98, PPV 0.72, accuracy 0.98.<br><br>The neural network models performed better than the classification trees. |

Abbreviations: ACCU, Acute Care Cardiology Unit; AUC, Area Under the Curve; AUROC, Area Under the Receiver Operating Characteristic Curve; BMTU, Bone Marrow Transplant Unit; BP, Blood Pressure; ECG, Electrocardiogram; HR, Heart Rate; ICU, Intensive Care Unit; IMCU, Intermediate Care Unit; NICU, Neonatal Intensive Care Unit; OR, Operating Room; PCU, Progressive Care Unit; pCO2, Partial Pressure of Carbon Dioxide; PICU, Paediatric Intensive Care Unit; pO2, Partial Pressure of Oxygen; PPV, Positive Predictive Value; PVC, Premature Ventricular Contractions; ROC curve, Receiver Operating Characteristic Curve; RR, Respiratory Rate; $SpO_2$, Peripheral Capillary Oxygen Saturation.