Towards a Quieter ICU: Mapping Alarms and Exploring Opportunities for SpO2 Alarm Reduction

Master Thesis Technical Medicine

Maxime Julia van Kekem

Universiteit Leiden

TU Delft

ERASMUS UNIVERSITEIT ROTTERDAM

Cover image sources in order of appearance:

1. Image source unknown. Used for illustrative purposes under fair use.
2. Swanepoel J. *Alert flat line on heart monitor - 3D rendering on pixel screen* [Internet]. Shutterstock; 2011 [cited 2024 Dec 30]. Available from: https://www.shutterstock.com/nl/image-illustration/flat-line-alert-on-heart-monitor-78054850
3. Stockpro. *Medical monitoring equipment in a modern hospital intensive care unit with blurred background of a healthcare professional* [Internet]. Adobe Stock; 2024 [cited 2024 Dec 30]. Available from: https://stock.adobe.com/nl/search/images?k=icu+monitor&asset_id=873237567

Towards a Quieter ICU: Mapping Alarms and Exploring Opportunities for SpO_2 Alarm Reduction

Maxime Julia van Kekem

Student number : 4648897

January 6, 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 Intensive Care Medicine, Leiden University Medical Center, Leiden
6 May 2024 – 16 January 2025

Supervisor(s):
Dr. D. J. van Westerloo, MD, PhD, LUMC
Dr. L. Kervezee, PhD, LUMC
Drs. F. W. Hiemstra, MSc, LUMC

Thesis committee members:
Dr. D. J. van Westerloo, MD, PhD, LUMC (chair)
Dr. L. Kervezee, PhD, LUMC
Dr. E. Özcan Vieira, PhD, TU Delft

Universiteit Leiden

TU Delft Delft University of Technology

ERASMUS UNIVERSITEIT ROTTERDAM

This page was intentionally left blank.

Preface and Acknowledgments

This thesis marks the culmination of my Master's studies in Technical Medicine at Delft university of Technology, Erasmus University Rotterdam, and Leiden University. The journey has been both challenging and rewarding, and above all, it has been inspiring. Over the past years, I have had the privilege of exploring the intersection of healthcare and technology, an area that has deepened my passion for improving patient care through innovation.

Like most of my peers, my academic journey began with the Bachelor's program in Clinical Technology. Following that, I took a gap year as PR & Marketing manager at Forze Hydrogen Racing, an experience that not only broadened my perspective but also helped me train skills that complemented my technical education. Transitioning into my Master's studies, each phase of this journey has deepened my curiosity and fostered both academic and personal growth. Working on this project has been a defining moment in my studies, allowing me to contribute to the field of ICU alarm reduction while expanding my expertise in bridging technology and medicine.

I am immensely grateful to my supervisors, David van Westerloo, Laura Kervezee, and Floor Hiemstra for their invaluable guidance, expertise, and encouragement throughout this process. Your insights and feedback have challenged me to think critically and approach problems with confidence and creativity. I would also like to thank Elif Özcan Vieira for taking the time for being part of my thesis committee.

To my fellow students who conducted their TM or research internships at the LUMC ICU, thank you for your camaraderie, support, and the many coffee and lunch breaks that provided much-needed moments of reprieve from work.

Finally, I want to express my deepest gratitude to my parents, sister, friends, and roommates. Your unwavering support and encouragement has been a constant source of motivation. The happy memories we've shared over the years have brought me immense joy and inspiration, reminding me why I started on this path in the first place.

As I conclude this chapter and look toward the future, I am excited to apply the lessons learned here to my future professional endeavours, wherever they may lead me.

Maxime van Kekem

Delft, January 2025

Contents

Preface and Acknowledgments 3
Contents 4
Summary 5
List of Abbreviations 5
1. Introduction 6
1.1. Background 7
2. Goals and Objectives 7
3. Methods 8
3.1. Data Acquisition 8
3.2. Data Preparation 8
3.3. Data Analysis 9
3.4. Actionability Annotation 10
3.5. SpO_2 Trends at Time of Alarms 11
4. Results 11
4.1. Data Acquisition 11
4.2. Alarmscape 12
4.3. Actionability Annotation 15
4.4. SpO_2 Trends at Time of Alarms 16
5. Discussion 17
6. Conclusions 19
References 19
Appendix 22

Summary

Introduction: The Intensive Care Unit (ICU) is a highly monitored environment where alarm overload poses risks to patient safety and staff well-being. SpO_2 alarms contribute significantly and are often clinically irrelevant. This contributes to alarm fatigue, delayed responses, and disrupted workflows while excessive noise adversely affects patient recovery and nurse mental health. The aim was to analyse the auditory alarm landscape in the ICU at Leiden University Medical Center (LUMC) with a focus on SpO_2 alarms. Furthermore, SpO_2 alarms were annotated for actionability through the integration of contextual data and SpO_2 trends were evaluated.

Methods: A retrospective analysis was conducted on ICU patients admitted between December 2023 and October 2024. Alarm, oxygen saturation, and patient data were extracted from monitoring systems and patient data management systems. The alarm dataset was explored using descriptive statistics. Audible SpO_2 alarms were annotated for actionability using predefined criteria based on clinical context, signal quality, and response to alarms, including respiratory support therapy escalation and ventilation or oxygen parameter adjustments. SpO_2 trends surrounding alarms were analysed to find patterns between actionable and non-actionable events.

Results: Among 635,717 auditory alarms recorded over 2261 patient-days, 32% were SpO_2 alarms, with 88.7% classified as non-actionable. The median response time for actionable alarms was 8.17 minutes, with most interventions involving FiO_2 increases. Temporal analyses revealed alarm frequency peaks during morning and afternoon shifts. SpO_2 trends at the time of actionable alarms correlated with significant desaturation events, while non-actionable alarms reflected minor, transient changes.

Conclusion: This study analysed the auditory alarmscape in the LUMC ICU, revealing that SpO_2 alarms, while significant contributors to alarm burden, are mostly non-actionable, increasing the alarm load and its associated challenges unnecessarily. By annotating alarms based on clinical context, it has laid groundwork for developing robust predictive algorithms to suppress non-actionable alarms.

List of Abbreviations

ABP = Arterial blood pressure
ASV = Adaptive support ventilation
CI = Confidence interval
CVVH = Continuous Veno-Venous Hemofiltration
DST = Daylight saving time
DWH = Data Warehouse Connect
ECG = Electrocardiogram
ECMO = Extracorporeal membrane oxygenation
ECRI = Emergency Care Research Institute
EHR = Electronic health records
FiO_2 = Fraction of inspired oxygen
IEC = International Electrotechnical Commission
ICU = Intensive care unit
INOP = Inoperative (technical alarm)
IQR = Interquartile range
IRB = Institutional review board
ISO = International Organisation for Standardisation
LOESS = Locally estimated scatterplot smoothing
LOS = Length-of-stay
LUMC = Leiden University Medical Center
PDMS = Patient data management system
PEEP = Positive end-expiratory pressure
PFI = Pulsatile flow indicator
PPV = Positive predictive value
RST = Respiratory support therapy
SpO_2 = Peripheral oxygen saturation
WMO = Wet medisch-wetenschappelijk onderzoek

1. Introduction

The intensive care unit (ICU) is designed for extensive monitoring and treatment for patients with life-threatening conditions. Many medical devices such as bedside monitors, infusion pumps, ventilators, and haemodialysis machines play a vital role in this environment, contributing to a high alarm load in ICUs (1, 2). Technological advancements in healthcare have significantly improved patient care and monitoring capabilities (3), but they have also led to a substantial increase in the number of alarms in ICUs, often exceeding 87 audible alarms per bed per day (4).

Among the routinely monitored vital signs in the ICU is the peripheral arterial oxygen saturation (SpO_2). SpO_2 alarms are a significant contributor to the alarm load in ICUs, since as many as 43% of all ICU alarms are related to SpO_2 levels (5). However, their accuracy can often be limited as they are often false due to motion artifacts and signal quality issues (6). This highlights the significant problem that while many SpO_2 alarms may be triggered, they may not even be clinically relevant.

This large number of alarms in the ICU has serious consequences, first of all for nurses, who are the primary responders to these alerts. Alarm fatigue is a phenomenon in which healthcare providers become desensitised to alarms due to frequent exposure and sensory overload (7-9). This can lead to delayed responses or even missed alarms, resulting in significant risk to patient safety (9). The Emergency Care Research Institute (ECRI) has consistently ranked alarm hazards among the top 10 Health Technology Hazards for over a decade (10, 11).

Furthermore, alarm overload has been linked to significant psychological stress among ICU nurses, with 95% reporting feeling overwhelmed due to alarm burden (12). Nurses, who spend on average 35% of their time responding to alarms (13), often experience increased stress, decreased job satisfaction, and overall declines in mental health (14), all of which affects their ability to respond to alarms effectively.

Secondly, alarm overload disrupts the workflow of patient care, increasing the risk of errors and contributing to noise pollution in the unit (15). With alarms designed with high sensitivity and low specificity, leading to a positive predictive value (PPV) sometimes as low as 27% (16), many alarms are clinically irrelevant. In a paediatric ICU, for instance, only 10% of alarms prompted changes in patient care (8). Low perceived clinical relevance of alarms result in slow response times, missed alarms, and an increased probability of critical errors (8, 17, 18). Research suggests that reducing the total number of false alarms and associated noise levels can improve nurses' perception of alarm relevance, leading to better satisfaction and response times (19).

In addition to that, excessive alarm noise profoundly affects patients. The continuous presence of audible alarms creates an often stressful and disruptive auditory environment (20). Audible alarms disrupt patients' sleep, altering their sleep architecture and reducing their overall sleep quality (21, 22), a key concern for ICU patients whose recovery is already often compromised by poor sleep (23). Both subjective and objective sleep studies underscore their impact: patients frequently report alarms as disruptive and distressing, linking them to frequent awakenings and transitions between sleep stages (24-26). Although alarms are not the sole disruptor of sleep in ICUs, where other environmental factors such as staff conversations and patient care activities also play a role (25, 27, 28), the need to reduce alarm-related noise remains critical. Therefore, addressing noise level reduction and alarm management optimisation is crucial to improving patient experience and outcomes.

Prioritising clinically significant alarms by reducing the number of non-actionable alarms is essential to minimising alarm fatigue and creating a safer, more efficient environment for both patients and healthcare providers. Alarm-driven advancements in alarm management offer promising solutions to address the ICU alarm burden by tailoring systems to prioritise relevant alerts that genuinely require staff intervention. A critical step in this approach is the differentiation between actionable and non-actionable, as well as true and false alarms, to increase the PPV and emphasise clinically relevant alarms (19). For example, a smart alarm delay algorithm for SpO_2 monitoring aims to lower the frequency of alarms by allowing minor deviations from alarm thresholds to self-correct before triggering an alarm, while maintaining patient safety as larger deviations beyond the alarm limit are only allowed for shorter periods of time (19, 29, 30). Other efforts focus on integrating adaptive technologies that adjust alarm thresholds in real-time based on dynamic changes in patient conditions (31, 32).

Future opportunities lie in developing data-driven systems capable of forecasting alarms and predicting whether they will be clinically relevant. Additionally, improving the design of alarm systems to reduce ambiguity and enhance clinical effectiveness remains a critical area of focus (33). Tough no fully integrated data-driven systems for alarm forecasting currently exist (32), their development could revolutionise alarm management by further reducing clinically irrelevant alarms while maintaining high sensitivity to clinically relevant alarms and enhancing clinical outcomes through predictive analytics (33), creating an ICU environment that enhances patient safety, guarantees patients' rest and recovery, and reduces alarm fatigue.

1.1. Background

ICU alarms can be divided into two main categories: clinical and technical alarms (Figure 1). Clinical alarms are related to the patient's physiological condition, while technical alarms are related to the functionality of monitoring equipment. Both types of alarms can either be true or false. True alarms accurately reflect the patient's current condition or the status of the equipment. For example, a true clinical alarm could indicate a dangerously low oxygen saturation level, while a true technical alarms might signal disconnected ventilator tubing. In contrast, false alarms provide misleading information and have no therapeutical consequence (e.g. an alarm indicating lead disconnection, while all electrocardiogram (ECG)-leads are securely attached).

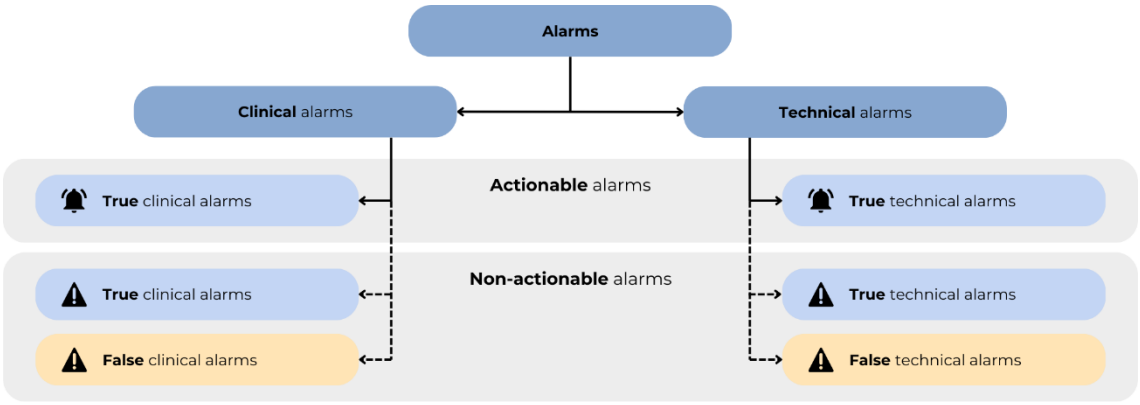


Figure 1: Classification of ICU alarms into clinical and technical categories, further subdivided into actionable and non-actionable types, highlighting true and false alarm distinctions. Subdivision of alarms is based on International Electrotechnical Commission standard 60601-1-8:2006/AMD2:2020 (34).

Another key classification is the distinction between actionable and non-actionable alarms, in which actionable alarms require a response from an operator. According to the International Electrotechnical Commission (IEC) standard IEC 60601-1-8:2006/AMD2:2020, a clinically actionable alarm is defined as an alarm condition for which a response or intervention within a high-priority timeframe from clinical staff is necessary to prevent patient harm that could ensure if ignored (34) (e.g. significant change in heart rate or equipment malfunction). These alarms are generally relevant to healthcare staff and provide meaningful information about a patient's condition or the status of medical equipment.

In contrast, non-actionable alarms lack clinical value for healthcare providers and do not require any clinical intervention (15) (e.g. temporary increase in respiratory rate). Not all true alarms will always be clinically relevant and may therefore be considered non-actionable (35) (e.g. alarms for conditions of which staff is already aware or low battery levels of non-critical equipment). These alarms often act as background noise, diluting the significance of truly important alerts and contributing to alarm overload.

2. Goals and Objectives

The primary goal of this master's thesis is to explore opportunities to reduce the number of non-actionable alarms in the ICU, particularly during the night, to foster a quieter environment, that promotes better sleep quality and reduces stress for patients. A quieter ICU environment not only benefits patient recovery but also enhances staff working conditions by reducing alarm fatigue.

This thesis focuses on understanding the auditory alarm landscape ("alarmscape") at the Leiden University Medical Center's (LUMC) ICU, with a particular emphasis on SpO_2 alarms, and exploring opportunities to reduce the incidence of non-actionable alarms. Specifically, it will focus the annotation of SpO_2 alarms as actionable or non-actionable through integration of data, thereby combining information from various datasets such as vital signs and interventions to create a more holistic view of the patient's condition at the time of an alarm.

By establishing this foundation, the study aims to pave the way for the development of a predictive algorithm capable of distinguishing and suppressing non-actionable SpO_2 alarms, prioritising clinically significant alarms, and ultimately enhancing the overall quality of care. If successful, this approach could be adapted to other vital signs, such as heart rate, blood pressure, or equipment-related technical alarms. In the long term, implementing such a model could significantly reduce the total number of ICU alarms, streamline clinical workflows, enhance patient safety, and create a more supportive environment for patient recovery.

3. Methods

3.1. Data Acquisition

Study cohort

A retrospective study was conducted among patients admitted to the ICU at LUMC between December 1, 2023 and October 1, 2024. To be eligible for inclusion, patients required a minimum ICU admission time of 24 hours. Patients admitted during daylight saving time (DST) transitions were excluded to avoid time data inconsistencies. Additionally, patients who received extracorporeal membrane oxygenation (ECMO) treatment during their admission were excluded. Patients for which not all necessary datasets were available were also excluded from analysis.

Dataset

All patients were monitored using IntelliVue MX750 bedside monitors (Koninklijke Philips N.V., Amsterdam, the Netherlands). All monitoring data was saved to a secure server, the Data Warehouse Connect (DWH) (Koninklijke Philips N.V., Amsterdam, the Netherlands), with a sample frequency of 1 Hz. The extracted datasets included alarm data, oxygen saturation data, and pulsatile flow indicator (PFI) data. Patient demographics, as well as data on respiratory support (oxygen support, (non)-invasive ventilation) for alarm contextualisation, were exported from the patient data management system (PDMS), MetaVision (iMDsoft, Israel).

The protocol of this study and the use of patient data for retrospective analysis was approved by the institutional review board (IRB) (nWMO Commissie Divisie 1) of the Leiden University Medical Center (reference number 2024-020). The IRB concluded that Medical Research Involving Human Subjects Act (Dutch acronym: WMO) does not apply to this study.

3.2. Data Preparation

Data preparation was performed using Spyder (Python Software 3.11.7) (36), supported by the following packages: contextlib, numpy, os, pandas, re, sys, and time.

All relevant treatment and alarm data was grouped by hospital admission to ensure performance efficiency, as patients could be admitted multiple times. Admission and discharge days were trimmed from the data, so only full 24-hour patient days were included, to avoid any bias due to heightened activity during these times. Admission periods that did not have at least one remaining 24-hour period after trimming were excluded.

The alarm dataset from the patient monitors was labelled according to predefined Philips labels, which were first reclassified into broader categories based on the Philips IntelliVue MX750 user manual (37). The Philips monitors announce alarms in two ways: auditory alarms, which are accompanied by visual cues such as lights or on-screen messages, and visual-only alarms. Alarms can additionally be silenced for a period of time, during which auditory alarms become inaudible. Figure 2 provides a visual representation of the alarm classification structure within

the alarms dataset, as exported from the patient monitors. Only auditory alarms were considered for this analysis, thus strictly visual alarms were excluded from the study dataset since they do not directly contribute to alarm noise exposure to patients.

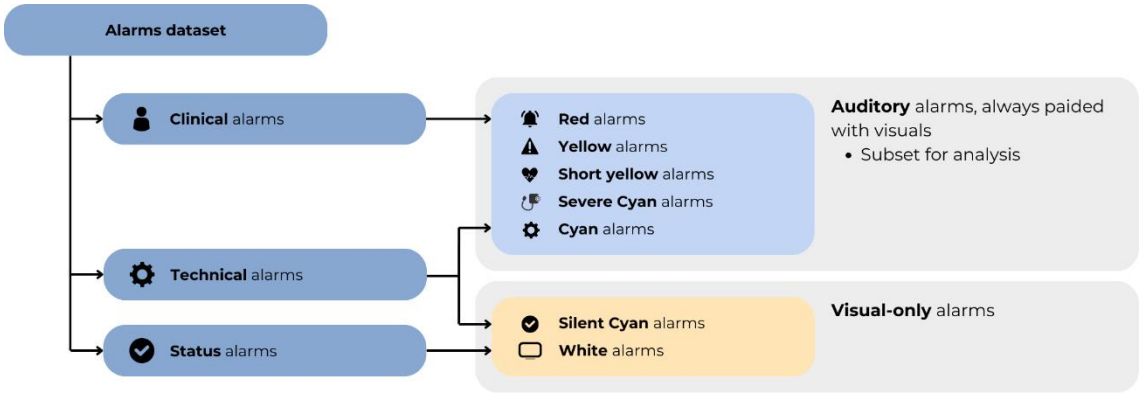


Figure 2: Visualisation of the alarm classification structure within the Philips alarm database, as exported from patient monitors. The chart categorizes alarms into clinical, technical, and status alarms, with further subdivisions indicating the type (e.g., auditory versus visual-only) and severity (e.g., red, yellow, cyan).

Respiratory support therapy (RST) modes (Table 1), along with ventilation and oxygen parameter data, were extracted from the PDMS and combined into a unified respiratory support dataset. Before merging the data, the oxygen parameter data was refinement. Oxygen parameters were recorded in two units of measurement (L/min and % oxygen) depending on the ventilation mode. These were separated into two distinct parameters: Oxygen Flow (0–16 L/min) and fraction of inspired oxygen (FiO_2) (21–100%) (38). Any values falling outside the expected ranges or deemed nonsensical were removed to ensure data accuracy.

3.3. Data Analysis

All data analyses were performed using Spyder (Python 3.11.7) (36), supported by the following packages: contextlib, matplotlib, numpy, os, pandas, scipy, seaborn, statsmodels, sys, and time.

ICU Alarmscape

To map the ICU alarmscape, the monitoring alarm dataset and subset of oxygen saturation alarms were first described through exploratory data analysis and visualisation of the data. Categorical data was presented using pie and bar charts to illustrate the frequency of different alarm types.

The duration of alarms was calculated by subtracting start time from end time of the alarm. Alarms with abnormally long durations (> 30 minutes) were treated as outliers and excluded from analysis. For alarms that were silenced for a maximum period of three minutes at a time, according to the standard settings of the monitors in the LUMC, the duration for which the alarm was silenced was also calculated. Some alarms occur simultaneously, in which case, only the most urgent alarm is audible. Combined with silenced periods during the alarm condition, this means patients are not exposed to all alarms or their entire duration. To determine the effective sound exposure to patients, coinciding alarms were identified, and the effective alarm duration was calculated by taking the earliest start time until the latest end time of overlapping alarms and subtracting the duration of silenced periods. The effective alarm durations for all alarms and the subset of SpO_2 alarms were visualised in a histogram. To illustrate when alarms occur and their distribution throughout the day, the alarm load, defined as the number of alarms per patient-hour, was calculated for intervals representing nursing shift durations (morning: 7:30 to 15:30, afternoon: 15:30 to 23:00, and night: 23:00 to 7:30 the next morning). The number of alarms throughout the day was furthermore visualised using a time-series chart representing the average number of alarms per time-interval of 10 minutes throughout the day.

3.4. Actionability Annotation

Annotation of alarms was performed using the guidelines published by Klopfenstein et al. (39), which are based on the IEC standard (IEC 60601-1-8:2006/AMD2:2020, (34)) definition of clinically actionable alarms and the International Organisation for Standardisation (ISO) classification of lung ventilators and related equipment (ISO 19223:2019, (40)). Alarms were classified as actionable or non-actionable based on the clinical context provided by the available datasets.

First, oxygen saturation alarms were evaluated based on the quality of the signal. The PFI is an index of the SpO_2 measurement quality (<0.3: poor, 0.3-1: acceptable, >1: optimal (37)). A low PFI indicates a poor pulsatile flow, and therefore poor signal quality. If the PFI at the time of an alarm is unacceptable (PFI <0.3), the pulse oximeter reading is not supported, and thus the alarm is labelled as non-actionable. After initial PFI evaluation, annotation criteria were applied.

To determine if an action was taken in response to an alarm, the ventilation modes and the parameters at the time of an alarm were compared to the first instance of a change in settings immediately after the alarm. Possible clinical actions in response to oxygen saturation alarms were defined, along with specific criteria to classify the alarms based on the actions taken. Respiratory support therapies were organised in a hierarchy based on the type and invasiveness of the therapy, taking inspiration from Klopfenstein's guidelines (39). Adaptive ventilation was categorised separately, as it is an automatically adapting ventilation mode that does not require healthcare worker intervention. As such, alarms generated while a patient received Adaptive Support Ventilation (ASV), could not be labelled as actionable or non-actionable by an operator based on changes in settings, and were therefore excluded from the analysis.

Table 1: Respiratory Support Therapies

Support Category	Hierarchy	Therapies Within Category
No support	0	
Oxygen supply	1	
High Flow support	2	HiFlow OptiFlow
Non-invasive ventilation	3	Continuous positive airway pressure (CPAP) Non-invasive ventilation (NIV) Non-invasive ventilation, spontaneous timed mode (NIV-ST)
Invasive ventilation	4	Spontaneous mode (Spont) Dual positive airway pressure (DuoPAP) Pressure-synchronised intermittent mandatory ventilation (PSIMV) Pressure-controlled mandatory ventilation (P-CMV) Spontaneous-controlled mandatory ventilation (S-CMV)
Adaptive ventilation*	5	Adaptive pressure ventilation with synchronised intermittent mandatory ventilation (APVsimv) Adaptive support ventilation (ASV) Intellivent-ASV

The hierarchy is adapted from the guidelines by Klopfenstein et al., simplifying their classification system (39). Unlike Klopfenstein's approach, this analysis focuses solely on the invasiveness of the RST method, irrespective of the airway device uses.

* Adaptive ventilation modes (e.g., ASV) were excluded from analysis as their automated adjustments prevent classification of alarms as actionable or non-actionable by operators.

Actionability criteria were defined as an escalation in RST, as outlined in Table 1, an increase in ventilation parameters (Positive End-Expiratory Pressure (PEEP), Pressure Control, and Pressure Support), or an increase in oxygen parameters (FiO_2 and Oxygen Flow) that occurred within 30 minutes of an alarm. An alarm followed by a decrease or no change in settings was not considered actionable, as this meant no action in response to the alarm was taken. An overview of the labelling process can be found in Figure 3.

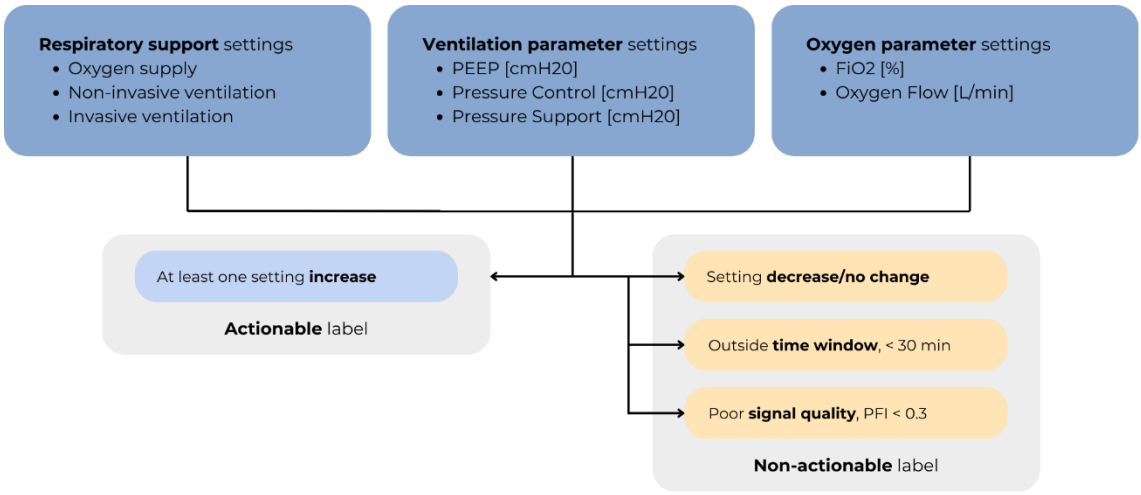


Figure 3: Flowchart of the actionability labelling process for ICU alarm settings. Alarms are classified as actionable if followed by a setting increase within the time window, and as non-actionable for setting decreases/no changes, events outside the time window, or poor signal quality.

3.5. SpO_2 Trends at Time of Alarms

To better understand SpO_2 behaviour around the time of oxygen saturation alarms, SpO_2 trends were analysed over a 30-minute window before and after each alarm event. The objective was to evaluate the pattern of oxygen saturation leading up to and following an alarm, and make a comparison between actionable and non-actionable alarms. Oxygen saturation waveforms were pre-processed to ensure consistency across the dataset. Locally Estimated Scatterplot Smoothing (LOESS) was applied to all alarm windows to remove noise and reveal underlying trends in the data. The average SpO_2 trends were calculated for each patient separately, distinguishing between actionable and non-actionable alarms. Subsequently, averages were calculated across all patients to identify overarching trends for the two alarm groups.

The averages of the actionable and non-actionable alarms were visualised with a 95% confidence interval (CI). The plots were standardised so that the x-axis was normalised to the time of the alarm (t=0), ensuring alignment of the trends relative to the alarm event.

4. Results

4.1. Data Acquisition

Study population

Between December 1, 2023, and October 10, 2024, 2176 patients were admitted to the LUMC ICU, who accounted for 2384 admissions, with a median length-of-stay (LOS) of 0.98 days. After applying the exclusion criteria to the dataset, and trimming the data (removing admission and discharge day from the admission periods), 523 admissions remained in the study dataset. The patient selection process is displayed in Figure 4. Table 2 summarises the study population's characteristics.

The study cohort had a median LOS of 2.92 days, which is significantly longer than the ICU population (0.98

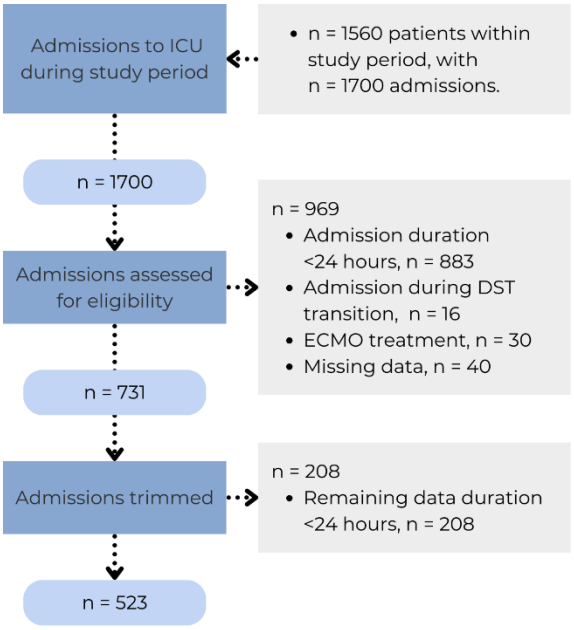


Figure 4: Flow chart depicting the patient admission inclusion process for the study. The chart shows the sequential steps of eligibility assessment, trimming criteria, and final selection for analysis, along with the reasons for exclusion at each stage.

days, p < 0.001). APACHE IV scores were also significantly higher in the study group (63.8 vs. 53.9, p < 0.001), reflecting a sicker and more complex patient cohort. This difference arises from the exclusion of short admission periods, resulting in a patient cohort with more severe conditions and longer stays than the general ICU population.

Table 2: Patient characteristics

	Study population	ICU population	P-value
Number of patients, n	487	1560	
Number of ICU admissions, n	523	1700	
Female, n [%]	172 [35.3%]	539 [34.5%]	0.572[a]
Age, years, mean (± SD)	61.3 (± 14.5)	60.7 (± 15.8)	0.271[b]
APACHE IV, mean (± SD)	63.8 (± 24.0)	53.9 (± 24.2)	< 0.001[b]*
Length-of-stay, median (IQR)	2.92 days (1.95 – 5.92)	0.98 days (0.78 – 2.00)	< 0.001*

[a] Chi-squared test was used for categorical variables.
[b] Student's t-test was used for continuous variables with normally distributed data.
For non-normally distributed continuous data, Wilcoxon rank-sum test was used.
* Indicates statistical significance (p < 0.05).

4.2. Alarmscape

Overview

The study population accounted for 2261 patient-days, capturing a total of 635,717 auditory monitoring alarms (clinical and technical) that were not silenced. Of these, 203,387 (32.0%) were oxygen saturation alarms. Table 3 summarizes the characteristics of all audible alarms and SpO_2 alarms.

A Mann-Whitney U test identified significant differences between all alarms and the SpO_2 alarm subset in terms of duration and effective duration (p < 0.001 for both comparisons). Auditory SpO_2 alarms had a longer average duration (median 12.54 seconds vs. 8.96 seconds). Alarm frequency varied by shift. The highest median alarm rates occurred during morning shifts, with 8.39 for all alarms and 2.06 for SpO_2 alarms, while night shifts recorded the lowest rates (4.91 and 1.37 alarms per patient-hour, respectively).

Table 3: Alarm characteristics

	All Audible Alarms	Audible SpO_2 Alarms	P-value
Number of alarms, n	635,717	203,387	
Number of patients, n	487	484	
Number of admissions, n	523	519	
Alarms per patient-hour, median (IQR)	8.15 (5.41 – 12.49)	2.21 (1.16 – 3.78)	
Morning Shift, median (IQR)	8.39 (4.52 – 15.23)	2.06 (0.77 – 4.51)	
Evening Shift, median (IQR)	6.80 (3.33 – 12.93)	1.60 (0.67 – 4.27)	
Night Shift, median (IQR)	4.91 (2.51 – 9.49)	1.37 (0.46 – 3.09)	
Alarm duration, median (IQR)	8.96 s (5.38 – 21.76)	12.54 s (6.66 – 30.21)	<0.001*
Effective alarm duration, median (IQR)	8.45 s (5.12 – 19.46)	11.264 (6.14 – 26.11)	<0.001*

Mann-Whitney U test was used to compare groups due to non-normally distributed data. Effective alarm duration is the total time patients were exposed to audible alarms with periods when alarms were silenced subtracted.
* Indicates statistical significance (p < 0.05).

Alarm distribution

All 635,717 audible alarms in the study were categorised into three severity levels: Red (1.9%, n = 12,016), Yellow (41.9%, n = 266,193), and technical (INOP) alarms (56.2%, n = 357,508). Figure 5 illustrates the proportional distribution of these categories, along with a detailed breakdown of specific alarm types.

Red alarms, indicating potentially life-threatening changes in vital signs, accounted for only 1.9% of all alarms, with SpO_2 desaturation alarms being the most common (73.4%), followed by apnoea (12.4%) and tachycardia alarms (5.3%). Yellow alarms, which are less critical than Red alarms, made up 41.9% of all alerts and were

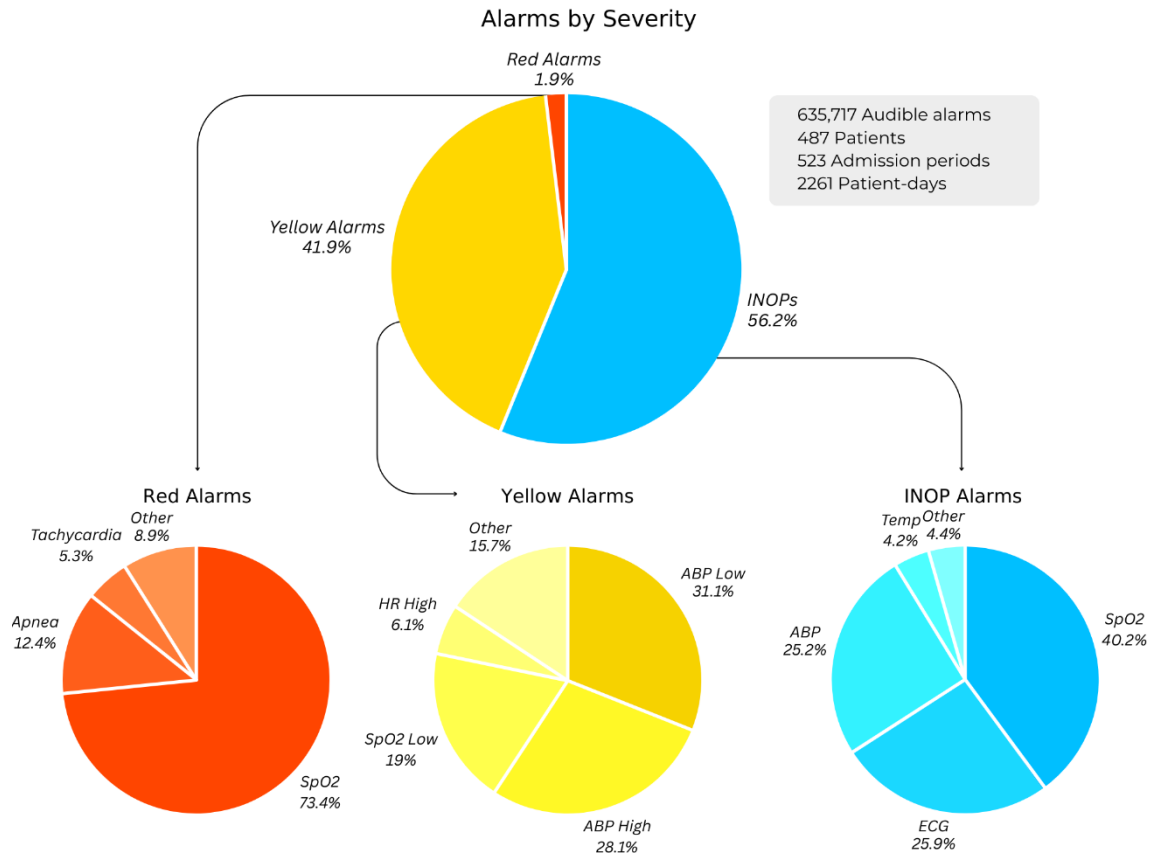


Figure 5: Pie-of-pie chart illustrating the distribution of all audible alarms in the monitoring dataset, categorised by severity (Red, Yellow, INOP), with each severity level further subdivided into specific alarm labels.

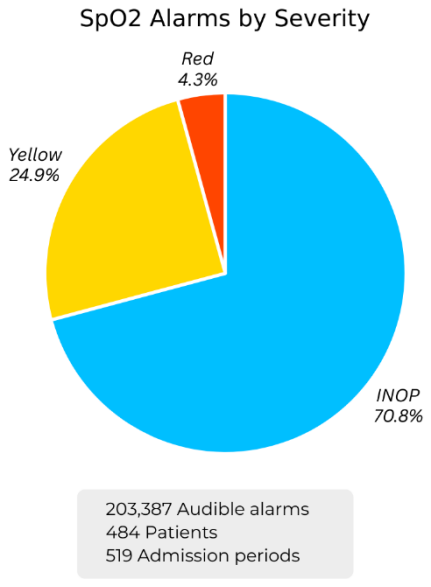


Figure 6: Distribution of audible SpO_2 alarms in the monitoring dataset, categorised by severity (Red, Yellow, INOP).

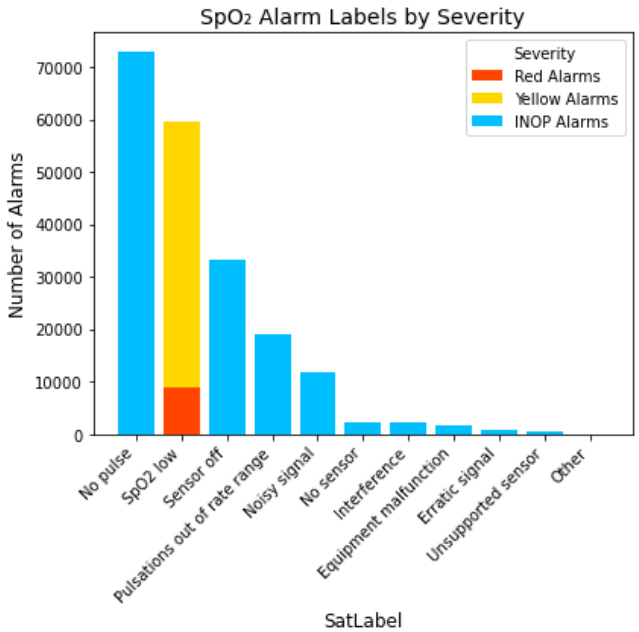


Figure 7: Stacked bar chart illustrating the frequency of different audible SpO_2 alarm labels in the monitoring dataset, categorised by severity (Red, Yellow, INOP).

dominated by high and low arterial blood pressure (ABP) alarms (28.1% and 31.1%, respectively). SpO_2 low alarms contributed 19% to this category. INOP alarms, which signal technical issues, dominated the dataset at 56.2%. Among these, SpO_2-related INOPs were the most prevalent (40.2%), followed by ECG-related INOPs (25.9%), with temperature and other INOPs accounting for 4.4%. The dominance of INOP alarms underscores the need for strategies to address technical issues and alarm reduction strategies, particularly for SpO_2 and ABP-related alarms, to reduce ICU noise and improve alarm management.

SpO_2 alarm characteristics

SpO_2 alarms accounted for 203,387 alarms (32.0% of all audible alarms), among which 4.3% (n = 8817) and 24.9% (n = 50,651) consisted of red and yellow alarms, respectively. Technical alarms contributed to the majority of 70.8% (n = 143,919), as illustrated in Figure 6.

Figure 7 highlights the specific triggers for SpO_2 alarms. The most common issue was "No Pulse," and "SpO_2 Low", for which the red alarms indicated a critically low SpO_2, followed by "Sensor Off," "Pulsations Out of Range," and "Noisy Signal." This emphasises the burden of INOP alarms. Furthermore, the prominence of the clinical SpO_2 alarms highlights the importance of refining alarm thresholds and delays to minimise unnecessary alerts while maintaining patient safety.

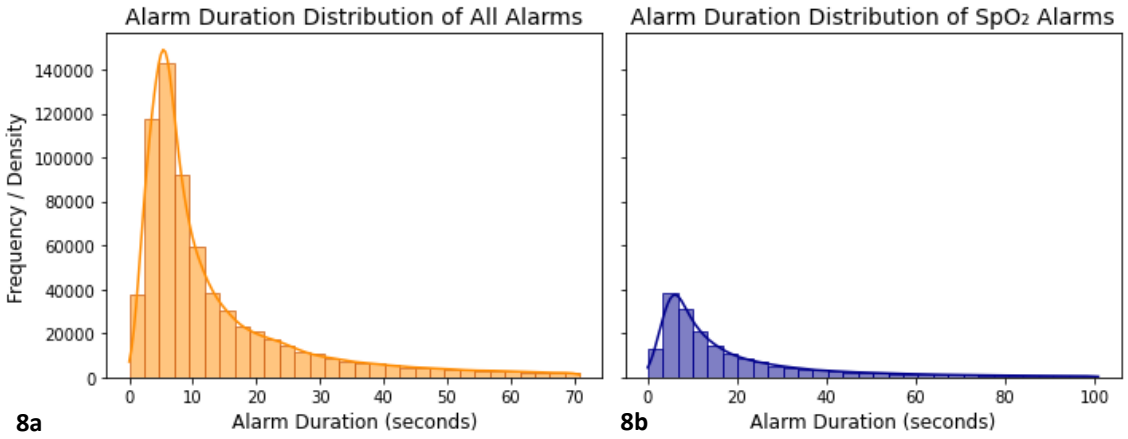


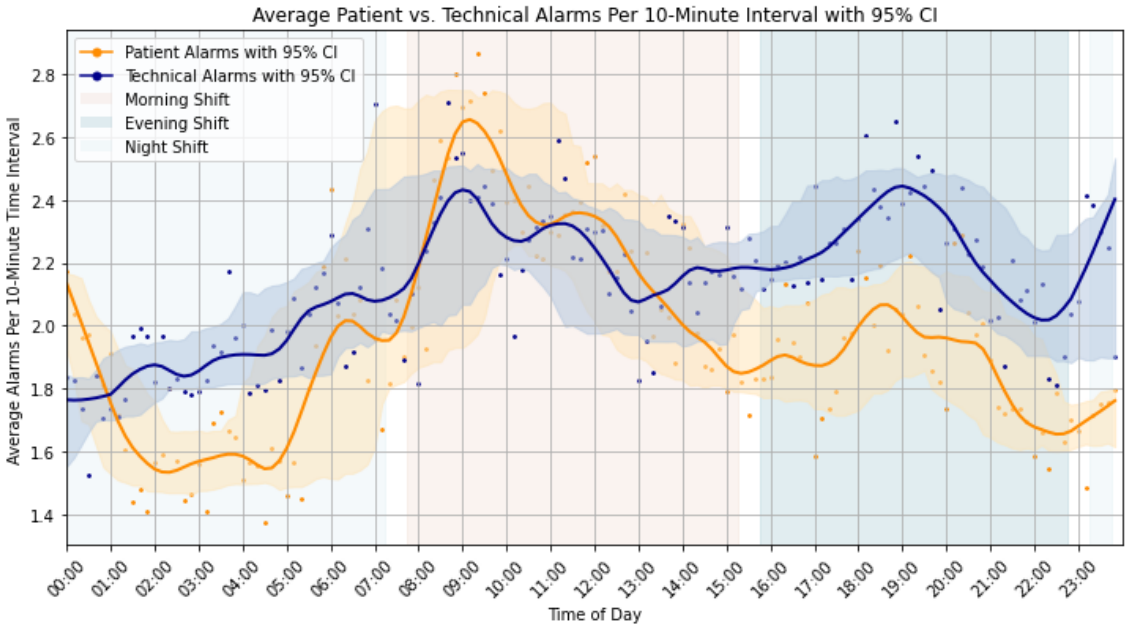
Figure 8: Distribution of alarm durations for all audible alarms (8a) and audible oxygen saturation alarms (8b) in the monitoring dataset. For visibility, the range of these histograms were capped at 1.5x the upper quartile.

Alarm duration and temporal patterns

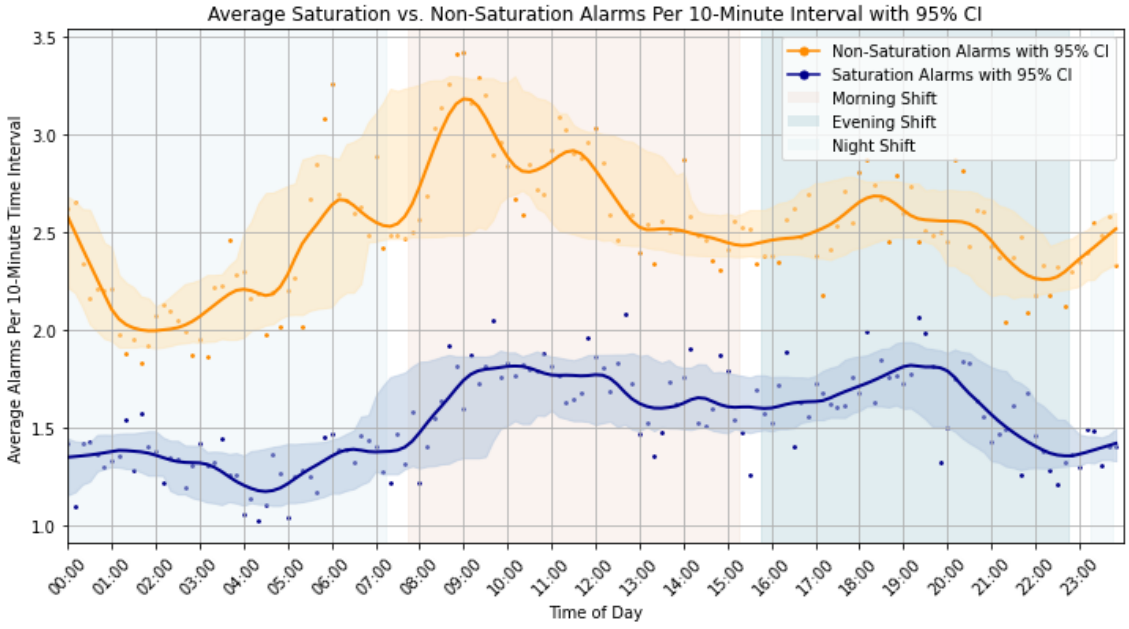
Histograms of alarm duration distributions for all audible alarms and SpO_2 alarms are presented in figure 8. The histograms demonstrated a high prevalence of short-duration alarms, with the majority lasting less than 10 seconds, as indicated by the peak frequency that occurs within this range. SpO_2 alarms, however, showed a slightly longer duration profile, with a noticeable tail extending beyond 30 seconds. This highlights the need for strategies to manage both short-duration alarms, which contribute to ICU noise, and longer-duration alarms, which cause more noise.

Figure 9 illustrates temporal variations in alarm frequency. Clinical and technical alarms (Figure 9a) showed distinct 24-hour patterns, with peaks during the morning and early afternoon, coinciding with scheduled periods of patient care (between 8:00 – 10:00 and 18:00 – 22:00). Technical alarms occurred more frequently than clinical alarms overall. Saturation alarms (Figure 9b) followed a similar 24-hour trend but were consistently less frequent than non-saturation alarms (all audible alarms in dataset excluding saturation alarms). Non-saturation alarms peaked sharply in the morning and remained elevated until the evening shift. Both graphs demonstrate a drop in alarm frequency during the night, yet the alarm load during night shifts is still 4.91 and 1.37 for all alarms and SpO_2 alarms respectively (Table 3), indicating there is still more to gain in alarm reduction during this period to create a more restful environment for patients, especially during the night. These patterns underline the

importance of temporal factors in alarm patterns, in that alarm frequency increases during times of high clinical activity. Alarm frequency peaks in the morning and early afternoon are likely due to artefacts and patient movement.



9a



9b

Figure 9: Distribution of audible alarms throughout the day, comparing clinical alarms and technical alarms (9a), and saturation and non-saturation alarms (9b). The graphs show the mean normalised number of alarms per 10-minute interval across all patients, with LOESS smoothing and 95% CI. Shaded regions represent morning, evening, and night shifts.

4.3. Actionability Annotation

Out of a total of 51,742 clinical oxygen saturation alarms, excluding the alarms that occurred during adaptive ventilation settings, 5841 (11.3%) were classified as actionable according to the proposed annotation algorithm, while the remaining 45,901 (88.7%) were deemed non-actionable. Among the non-actionable alarms, 1119 (2.4% of non-actionable alarms) were associated with low PFI values (<0.3), indicating poor signal quality at the time of the alarm. The remainder of the non-actionable alarms lacked any recorded intervention within the specified time window, suggesting they did not prompt a clinical response.

The vast majority of actionable alarms (97.5%, n = 5695) were classified as actionable based on a single condition. A smaller proportion (2.5%, n = 146) met two or more independent criteria, reinforcing their classification as actionable. The most frequent intervention for actionable alarms was an increase in FiO_2, occurring in 67.7% of cases, underscoring its critical role in managing oxygen saturation. Other responses included escalations in RST mode (18.2%), oxygen flow increase (6.4%), and PEEP increases (4.1%), as shown in figure 10.

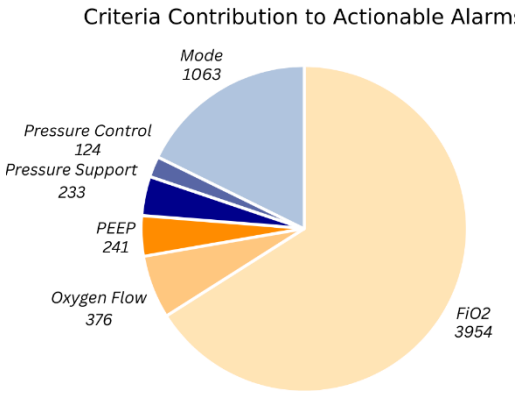


Figure 10: Actionability criteria contributions to determining actionable alarms, with the numbers of alarms that were made actionable labelled.

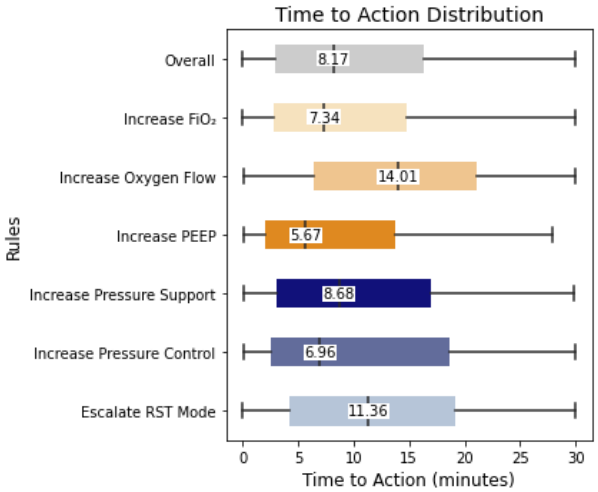


Figure 11: Distribution of time to intervention for actionable alarms, overall and by individual criteria. Times were capped at 30 minutes. The plot shows the frequency and density of intervention times.

The median time to intervention for actionable alarms was 8.17 minutes (IQR: 2.98 – 16.31 minutes), reflecting a relatively prompt response to clinically significant events. Among interventions, PEEP increases had the fastest response time, with a median of 5.67 minutes (IQR: 2.00 – 13.77), followed by Pressure Control increase (6.96 minutes, IQR: 2.51 – 18.59), and FiO_2 increase (7.34 minutes, IQR: 2.76 – 14.78), as detailed in Figure 11.

4.4. SpO_2 Trends at Time of Alarms

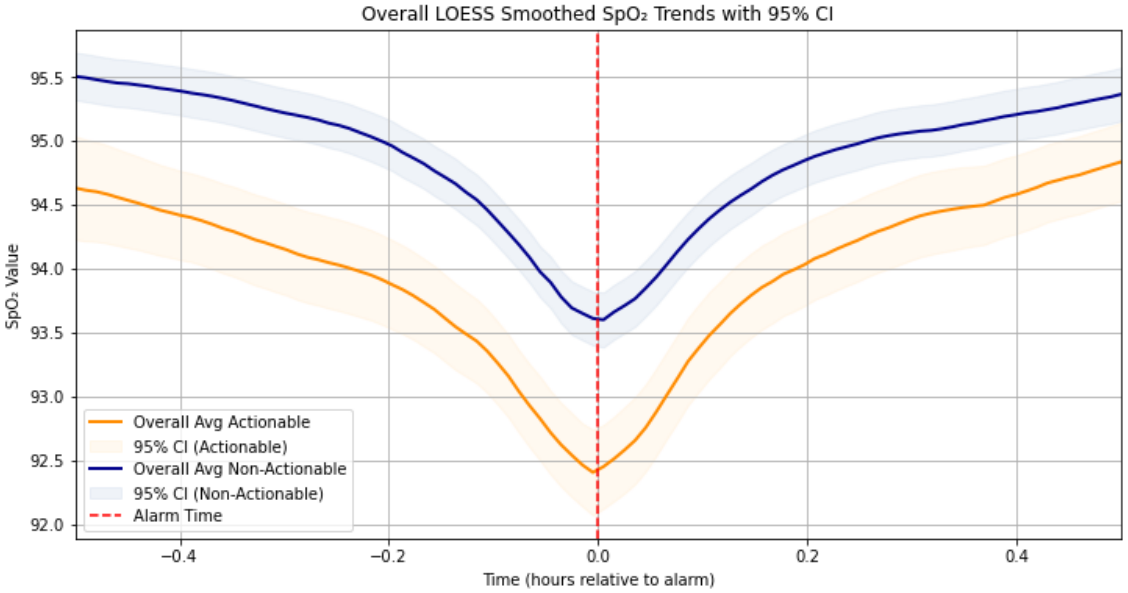


Figure 12: LOESS-smoothed SpO_2 trends with 95% CI relative to audible SpO_2 alarms, highlighting differences between actionable and non-actionable alarms. The plot spans a 1-hour window, spanning 30 minutes before and after the alarm time, standardised at t=0 for consistent comparison of alarm trends.

Figure 12 illustrates the SpO_2 value trends before and after audible SpO_2 alarms, distinguishing between actionable and non-actionable alarms. Notable differences in SpO_2 trajectories emerged between the two categories. For actionable alarms, SpO_2 values show a pronounced decrease leading up to the alarm time, reaching their lowest point at t = 0, followed by a gradual recovery over the subsequent 30 minutes. This pattern reflects clinically significant desaturation events that necessitate intervention. In contrast, non-actionable alarms exhibit a more gradual and shallow decline in SpO_2 values, suggesting they are more often triggered by transient or minor fluctuations that do not indicate sustained or serious issues. These transient changes contribute disproportionately to ICU noise. To further explore patient-specific trends, supplementary Figure A1 provides LOESS-smoothed actionable and non-actionable SpO_2 trajectories for each patient separately.

5. Discussion

This thesis aimed to examine the alarmscape at the LUMC ICU and identify opportunities to reduce non-actionable alarms SpO_2 alarms, to foster a quieter ICU environment to improve patient recovery, satisfaction, and staff working conditions.

The analysis revealed an alarm load of 8.15 alarms per patient-hour, with SpO_2 alarms comprising one-third (203,387 of 635,717) of all audible alarms. Alarm patterns varied over 24-hours, peaking during patient care activities, which likely generated false alarms. Nocturnal alarm rates were notably lower at 4.91 alarms per patient-hour, with SpO_2 alarms accounting for 1.37 alarms per patient-hour. These alarms disrupt patients' sleep and hinder their recovery, highlighting a critical area for intervention. Most alarms lasted less than 10 seconds, but saturation alarms tended to persist slightly longer than non-saturation alarms. Only 11.3% of SpO_2 alarms were actionable, with FiO_2 increases being the most common intervention, occurring in two-thirds of cases. The median response time for actionable alarms was 8.17 minutes, a relatively prompt response but potentially lengthy for urgent critical alarms. These findings underscore the need for strategies to suppress non-actionable alarms without compromising patient safety.

The findings of this study align with previous reports on alarm frequencies. Poncette et al. observed an alarm load of 152.5 alarms per bed per day (6.35 alarms per patient-hour), comparable to the 8.15 alarms per patient-hour reported here. They also noted similar daily patterns, with increased alarm activity during morning and afternoon shifts and reduced activity at night. However, their distribution of alarm types differed significantly, with 79% yellow alarms, 18% red alarms, and only 3% technical alarms. This discrepancy may stem from differences in ICU protocols, patient populations, or alarm settings (41).

In contrast, Li et al. reported distributions more closely matching the findings of this study: 55.0% yellow alarms, 41.4% technical alarms, and 3.6% red alarms. Their study, however, was conducted in a neonatal ICU, which differs substantially from our patient population (42).

The high proportion of non-actionable alarms observed in this study (88.7%) is consistent with previous reports, which indicate 72-99% of alarms are non-actionable (43). Chromik et al. using a similar annotation method, recently found that 9.2% of saturation alarms were actionable, closely aligning with our finding of 11.3% (44). Like this study, they reported that most actionable alarms were classified based on a single criterion, and interventions were typically documented within 15 minutes of alarm onset, similar to our median intervention time of 8.17 minutes.

This study employed predefined annotation criteria to label SpO_2 alarms as actionable or non-actionable, providing a foundation for developing prediction models to improve alarm management. Our insights into SpO_2 suggest that actionable alarms are associated with more pronounced SpO_2 declines, while non-actionable alarms often reflect more transient artefacts. These distinctions warrant further exploration and characterisation to inform alarm suppression algorithms that reduce non-actionable alarms while maintaining specificity and patient safety.

We also evaluated responses to actionable alarms and found that FiO_2 increases were the most frequent response to actionable SpO_2 alarms, likely due to their simplicity and ease of implementation. However, these adjustments may also reflect routine patient care rather than alarm-driven responses. Future alarm management

strategies could incorporate contextual data, such as electronic medical record (EHR) entries, clinician logs, and timestamps of care activities, to distinguish between interventions and genuine alarm-driven actions.

The median response time to actionable alarms was 8.25 minutes, within the IEC's defined 30-minute window. However, this may be insufficient for urgent alarms. Delayed registration in PDMS, potentially due to time required for patient assessment and intervention preparation, may introduce bias in response time evaluations. Further refinement of current annotation criteria is needed to evaluate alarm urgency more accurately.

This study's strengths include its large dataset, which enhances the generalisability of our findings by capturing diverse alarm scenarios, patient conditions, and clinical contexts. The dataset provides a comprehensive view of the alarmscape of our units, offering valuable insights into the frequency, types, and patterns of alarms, particularly in the context of nocturnal disruptions. This allows for the identification of trends, such as differences in SpO_2 patterns, that might not be apparent in smaller studies.

Another key strength is its contribution to the limited research on separating actionable from non-actionable alarms. Few studies have explored this distinction yet (44), but it is a promising avenue in alarm algorithm design. This work adds to the field by providing insights into alarm actionability based on clinical interventions in our dataset. By focusing specifically on SpO2 alarms and integrating contextual data, this study lays the groundwork for improving alarm management strategies, offering a novel perspective on how actionable alarms can be identified and addressed more effectively.

However, limitations include the fact that the current dataset only contained alarms from the Philips bedside monitors, excluding alarms from devices like ventilators, perfusion pumps, and continuous veno-venous hemofiltration (CVVH) machines. This exclusion likely underestimates the total alarm burden of our ICU. Future studies should incorporate data from all ICU alarm sources to present the complete picture. As medical devices are being developed and manufacturers allow for data integration, robust data management systems to unify device-specific information is required. Additionally, the assumption that all actionable alarms prompt an intervention response may not hold as alarms could have been missed or ignored by healthcare providers due to alarm fatigue. Alarms could have been mislabelled as non-actionable when, in reality, they were actionable but did not receive the appropriate intervention. Incorporating real-time feedback mechanisms where staff annotate alarm response and what action was taken could improve labelling accuracy and through that support predictive model development.

Alarm labelling focused exclusively on clinical SpO_2 alarms, as the annotation criteria were defined based on what was possible with the available data, thus excluding technical alarms. However, technical alarms, can provide valuable context. Including these alarms in future analyses could reveal interactions between clinical and technical alarms that may otherwise be overlooked, useful for refining annotation criteria and enhancing alarm management strategies.

Recommendations for alarm management and future research

Tailored strategies are essential for reducing non-actionable alarms. For technical alarms, which are more frequent during patient care, disabling alarms temporarily during routine activities can minimise artefact-related triggers. This is standard practice in proper alarm management, but is not always adhered to. Additional strategies include timely sensor replacement, setting appropriate thresholds and regular reassessment during nursing handovers (19, 45), and staff training on optimal device usage (46). Although adherence to alarm management guidelines was not assessed in this study, emphasising these practices remains critical to reducing alarm burden and improving workflow with the least interventions possible.

For short-term non-actionable alarms, implementing the Philips Smart Alarm Delay algorithm for SpO_2 alarms is a promising solution. This feature delays alarm activation, allowing transient vital sign changes to self-correct and reducing unnecessary alarms. Expanding on this algorithm to other alarm types could further enhance alarm management. Its implementation could also reduce the need for staff to constantly adjust alarm thresholds, minimising workflow disruptions. Before implementation on our units, its effectiveness should be evaluated in our dataset to assess its effectiveness. If validated, it represents a low-effort, high-impact solution to addressing short-duration non-actionable alarms.

For longer-duration non-actionable alarms, examples of management options are optimising thresholds and leveraging insights from actionable versus non-actionable alarm trends, particularly SpO_2 patterns before and after alarms. This is where predictive algorithms can be a promising solution. Such an algorithm could suppress irrelevant alarms before activation, reducing patient disruption and enhancing ICU efficiency.

While this study focused on SpO_2 alarms due to their prevalence and impact in the ICU environment, future research should expand to include alarms for all vital signs. A comprehensive labelling framework for all alarms is essential before predictive algorithm development. Incorporating true versus false classifications alongside actionability would refine alarm differentiation and improve predictive accuracy as well. Additionally, integrating data from clinician logs and nursing reports from the EHR could provide critical context for alarm actionability. This added layer of context will enhance labelling accuracy, identify actionable alarm patterns or conditions associated with actionable alarms, paving the way for a robust predictive model.

6. Conclusions

This thesis examined the auditory alarm landscape in the LUMC ICU, with a focus on SpO_2 alarms, to identify opportunities for reducing non-actionable alarms. The findings revealed that SpO_2 alarms comprise a significant proportion of the overall alarm burden, yet most are non-actionable, contributing to alarm overload, patient disturbances, increased alarm fatigue, and disrupted workflows. By annotating alarms based on clinical context and analysing actionable versus non-actionable trends, this research has laid groundwork for developing robust predictive algorithms capable of suppressing non-actionable alarms while prioritising clinically relevant ones.

The integration of expanded contextual data and implementation of smart alarm delay algorithms offer promising solutions to mitigate the ICU alarm overload, reduce noise, and enhance both patient safety and the working conditions of healthcare providers.

References

1. Eliash C, Lazar I, Nissim N. SEC-CU: the security of intensive care unit medical devices and their ecosystems. IEEE Access. 2020;8:64193-224.
2. Kleber K. ICU Equipment List – A Comprehensive Guide to Critical Care Unit Equipment for Nurses FreshRN2024 [updated 23/05/2024. Available from: https://www.freshrn.com/icu-equipment-list/.
3. Chander S, Kumari R, Sadarat F, Luhana S. The evolution and future of intensive care management in the era of telecritical care and artificial intelligence. Current Problems in Cardiology. 2023;48(10):101805.
4. Sowan AK, Gomez TM, Tarriela AF, Reed CC. Changes in default alarm settings and standard in-service are insufficient to improve alarm fatigue in an intensive care unit: a pilot project. JMIR human factors. 2016;3(1):e5098.
5. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. Critical care medicine. 1997;25(4):614-9.
6. Jubran A. Pulse oximetry. Critical Care. 2015;19(1):272.
7. Cvach M. Monitor alarm fatigue: an integrative review. Biomedical instrumentation & technology. 2012;46(4):268-77.
8. Johnson KR, Hagadorn JI, Sink DW. Alarm safety and alarm fatigue. Clinics in Perinatology. 2017;44(3):713-28.
9. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. AACN advanced critical care. 2013;24(4):378-86.
10. Keller Jr JP. Clinical alarm hazards: a “top ten” health technology safety concern. Journal of electrocardiology. 2012;45(6):588-91.
11. Top 10 health technology hazards for 2020. ECRI Institute; 2019.
12. Lewandowska K, Weisbrot M, Cieloszyk A, Mędrzycka-Dąbrowska W, Krupa S, Ozga D. Impact of alarm fatigue on the work of nurses in an intensive care environment—a systematic review. International journal of environmental research and public health. 2020;17(22):8409.
13. Bitan Y, Meyer J, Shinar D, Zmora E. Nurses’ reactions to alarms in a neonatal intensive care unit. Cognition, Technology & Work. 2004;6:239-46.
14. Nyarko BA, Yin Z, Chai X, Yue L. Nurses’ alarm fatigue, influencing factors, and its relationship with burnout in the critical care units: a cross-sectional study. Australian Critical Care. 2024;37(2):273-80.

15. Ruskin KJ, Hueske-Kraus D. Alarm fatigue: impacts on patient safety. Current Opinion in Anesthesiology. 2015;28(6):685-90.
16. Chambrin M-C, Ravaux P, Calvelo-Aros D, Jaborska A, Chopin C, Boniface B. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. Intensive care medicine. 1999;25:1360-6.
17. Andrade-Méndez B, Arias-Torres D, Gómez-Tovar L. Alarm fatigue in the intensive care unit: Relevance and response time. Enfermería Intensiva (English Ed). 2020;31(3):147-53.
18. Ergezen FD, Kol E. Nurses' responses to monitor alarms in an intensive care unit: An observational study. Intensive and Critical Care Nursing. 2020;59:102845.
19. 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. Critical care medicine. 2018;46(1):130-7.
20. Johansson L, Bergbom I, Lindahl B. Meanings of being critically ill in a sound-intensive ICU patient room - a phenomenological hermeneutical study. Open Nurs J. 2012;6:108-16.
21. Burdick KJ, Callahan CJ. Sleeping soundlessly in the intensive care unit. Multimodal Technologies and Interaction. 2020;4(1):6.
22. Hinton L, Group HER. The intensive care unit was so noisy I couldn't sleep. BMJ. 2016;353.
23. Oztürk L, Pelin Z, Karadeniz D, Kaynak H, Cakar L, Gözükirmizi E. Effects of 48 hours sleep deprivation on human immune profile. Sleep research online: SRO. 1999;2(4):107-11.
24. Elbaz M, Léger D, Sauvet F, Champigneulle B, Rio S, Strauss M, et al. Sound level intensity severely disrupts sleep in ventilated ICU patients throughout a 24-h period: a preliminary 24-h study of sleep stages and associated sound levels. Ann Intensive Care. 2017;7(1):25.
25. Gabor JY, Cooper AB, Crombach SA, Lee B, Kadikar N, Bettger HE, et al. Contribution of the intensive care unit environment to sleep disruption in mechanically ventilated patients and healthy subjects. Am J Respir Crit Care Med. 2003;167(5):708-15.
26. Hilton BA. Quantity and quality of patients' sleep and sleep-disturbing factors in a respiratory intensive care unit. J Adv Nurs. 1976;1(6):453-68.
27. Alsulami G, Rice AM, Kidd L. Prospective repeated assessment of self-reported sleep quality and sleep disruptive factors in the intensive care unit: acceptability of daily assessment of sleep quality. BMJ Open. 2019;9(6):e029957.
28. Ding Q, Redeker NS, Pisani MA, Yaggi HK, Knauert MP. Factors Influencing Patients' Sleep in the Intensive Care Unit: Perceptions of Patients and Clinical Staff. Am J Crit Care. 2017;26(4):278-86.
29. Lansdowne K, Strauss DG, Scully CG. Retrospective analysis of pulse oximeter alarm settings in an intensive care unit patient population. BMC nursing. 2016;15:1-8.
30. IntelliVue MP5 patient monitor: Philips Medical Systems International B.V.; 2021.
31. Antink CH, Leonhardt S, Walter M. Reducing false alarms in the ICU by quantifying self-similarity of multimodal biosignals. Physiological measurement. 2016;37(8):1233.
32. Chromik J, Klopfenstein SAI, Pfitzner B, Sinno Z-C, Arnrich B, Balzer F, et al. Computational approaches to alleviate alarm fatigue in intensive care medicine: A systematic literature review. Frontiers in digital health. 2022;4:843747.
33. Oleksy AJ, Schlesinger JJ. What's all that noise—Improving the hospital soundscape. Journal of clinical monitoring and computing. 2019;33(4):557-62.
34. Commission IE. Medical electrical equipment - Part 1-8: General requirements for basic safety and essential performance - Collateral standard: General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems. IEC-60601-1-8:2006-AMD2:2020. Geneva: IEC; 2020.
35. Schmid F, Goepfert MS, Reuter DA. Patient monitoring alarms in the ICU and in the operating room. Critical care. 2013;17:1-7.
36. van Rossum G, Drake Jr FL. Python tutorial. Python 3.9.13 ed: Centrum voor Wiskunde en Informatica Amsterdam, the Netherlands; 1995.
37. IntelliVue-patiëntmonitor MX750/850: Philips Medical Systems International B.V.; 2019.
38. Fuentes S, Chowdhury YS. Fraction of inspired oxygen. 2020.
39. 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. 2023.
40. Standardization tIOo. Lung ventilators and related equipment — Vocabulary and semantics. ISO 19223:2019: ISO; 2019.

41. Poncette A-S, 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. Journal of Medical Internet Research. 2021;23(5):e26494.
42. Li T, Matsushima M, Timpson W, Young S, Miedema D, Gupta M, et al. Epidemiology of patient monitoring alarms in the neonatal intensive care unit. Journal of Perinatology. 2018;38(8):1030-8.
43. Hravnak M, Pellathy T, Chen L, Dubrawski A, Wertz A, Clermont G, et al. A call to alarms: current state and future directions in the battle against alarm fatigue. Journal of electrocardiology. 2018;51(6):S44-S8.
44. 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.
45. Yeh J, Wilson R, Young L, Pahl L, Whitney S, Dellsperger KC, et al. Team-based intervention to reduce the impact of nonactionable alarms in an adult intensive care unit. Journal of Nursing Care Quality. 2020;35(2):115-22.
46. 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.

Appendix

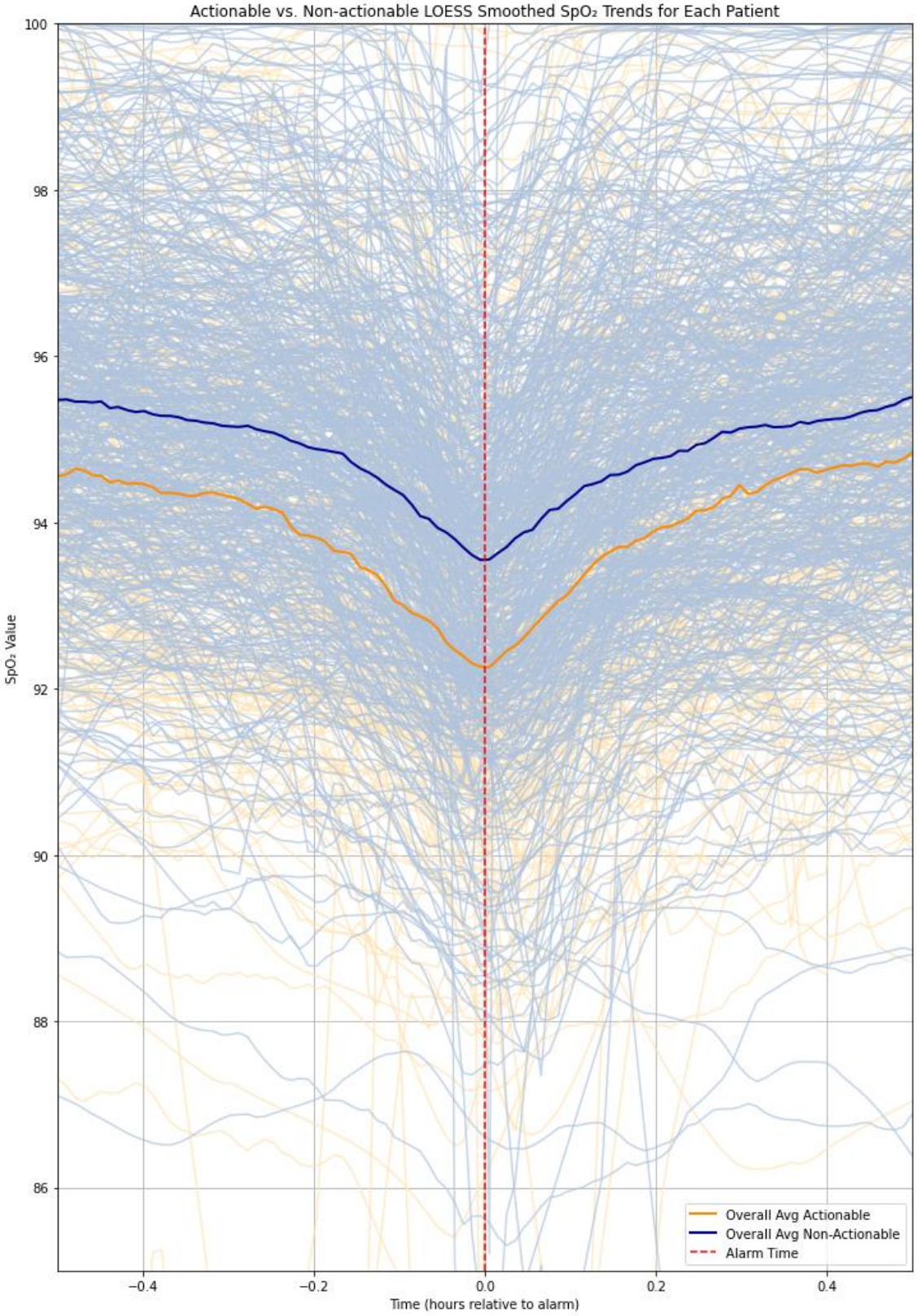


Figure A1: Trends of SpO_2 values relative to audible SpO_2 alarms, showing LOESS-smoothed actionable vs. non-actionable trends averaged for each patient. The plot spans a 1-hour window, spanning 30 minutes before and after the alarm time, standardised at t=0 for consistent comparison of alarm trends.