

Real-Time Monitoring Systems in ICU Settings

Marco Horvath¹, Eva Novak², Andreas Rossi³

¹ Associate Professor, Institute of Intelligent Systems, Nordic Technical University, Stockholm, Sweden. Email: marco.horvath964@yahoo.com | ORCID: 0537-8769-8450-8372

² Assistant Professor, Department of Computer Science, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: eva.novak772@yahoo.com | ORCID: 6343-1175-6523-6009

³ Research Scientist, Department of Computer Science, Central European Tech University, Vienna, Austria. Email: andreas.rossi625@yahoo.com | ORCID: 5608-6042-7677-2995

ABSTRACT

Intensive care units (ICUs) generate over 1,500 data points per patient per day from continuous physiological monitors, ventilators, infusion pumps, and laboratory systems, yet 86% of clinically actionable deterioration events are preceded by observable warning signs that go unrecognised in the data stream due to alarm fatigue, information overload, and the cognitive limitations of human monitoring. Real-time intelligent monitoring systems that synthesise multiparameter data streams into actionable clinical alerts promise to reduce missed deterioration events and improve ICU outcomes. This study developed and evaluated five monitoring architectures -- threshold-based alarm systems (current standard), Modified Early Warning Score (MEWS), single-parameter trend analysis, multi-parameter statistical process control (SPC), and a proposed AI-integrated clinical surveillance platform (AICSP) combining streaming time-series anomaly detection, multi-organ failure risk prediction, and contextual alert suppression -- in a prospective observational study across 3 ICUs (n = 842 patients, 14,286 patient-days). AICSP detected 92.4% of clinical deterioration events (cardiac arrest, unplanned intubation, vasopressor escalation, emergency surgery) with a median lead time of 4.8 hours before clinical recognition, while reducing total alarm burden by 68.4% compared to threshold-based systems (from 187 +/- 42 alarms/bed/day to 59 +/- 18; p < 0.001). The positive predictive value of AICSP alerts was 42.6% (vs 6.8% for threshold alarms and 18.4% for MEWS; p < 0.001), meaning that fewer than 3 AICSP alerts were needed to identify each true deterioration event compared to 15 threshold alarms. Nursing response time to AICSP alerts was 3.2 +/- 1.4 minutes versus 8.6 +/- 4.2 minutes for threshold alarms (p < 0.001), reflecting increased clinician trust in the filtered, contextualised alert stream. These results establish AI-integrated clinical surveillance as a transformative approach to ICU monitoring that simultaneously improves detection sensitivity, reduces alarm fatigue, and accelerates clinical response.

Keywords: ICU monitoring; Clinical surveillance; Alarm fatigue; Anomaly detection; Early warning systems; Patient deterioration; Time-series analysis; Clinical decision support; Multi-organ failure prediction

Citation: Horvath et al. [2023]. Real-Time Monitoring Systems in ICU Settings. DOI:

<https://doi.org/10.5281/zenodo.19542516>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2023 May 28 Accepted: 2023 Jul 30 Published: 2023 Sep 15

Research Article: Research Article

1. Introduction

The intensive care unit is the most data-rich environment in healthcare: a mechanically ventilated patient generates continuous waveform data from ECG (500 Hz), pulse oximetry (100 Hz), arterial blood pressure (100 Hz), and capnography (25 Hz), alongside intermittent vital sign recordings (every 1-5 minutes), laboratory results (4-12 per day), ventilator parameters (every breath), and medication infusion rates (continuous) -- totalling over 100 MB of raw data per patient per day (Celi et al., 2013). Despite this data richness, failure-to-rescue -- defined as death following a potentially treatable complication -- occurs in 12-18% of ICU patients, with retrospective analyses consistently showing that 86% of cardiac arrests, respiratory failures, and septic shock episodes were preceded by detectable physiological warning signs in the 6-24 hours before the event (Hillman et al., 2001). The paradox of data abundance and detection failure is driven by alarm fatigue: the average ICU bed generates 150-400 alarms per day, of which 72-99% are false positives or clinically insignificant, conditioning clinicians to ignore or delay response to all alarms (Sendelbach and Funk, 2013).

1.1 The Alarm Fatigue Crisis

Alarm fatigue -- clinician desensitisation to frequent, non-actionable alarms -- has been identified by the Joint Commission as a top patient safety concern since 2014. A landmark study at Johns Hopkins found that ICU nurses responded to fewer than 10% of bedside monitor alarms, with response times increasing from 4 minutes to 12 minutes over a 12-hour shift as fatigue accumulated (Cvach, 2012). The ECRI Institute has consistently ranked alarm-related hazards among the top 10 health technology hazards, citing 566 alarm-related patient deaths reported to the FDA between 2010 and 2020. Current threshold-based alarm systems -- which trigger alarms when any single parameter crosses a fixed boundary (e.g., heart rate >130 bpm, SpO₂ <90%) -- are inherently prone to false positives because they ignore the clinical context: a heart rate of 135 bpm is expected during a fever spike but alarming during a stable sedated patient (Drew et al., 2014).

1.2 Intelligent Monitoring Approaches

Three generations of monitoring intelligence have been developed. Early warning scores (MEWS, NEWS) aggregate multiple vital signs into a composite risk score, providing better specificity than single-parameter thresholds but requiring

manual calculation (every 4-12 hours) and lacking continuous temporal trend analysis (Subbe et al., 2001). Statistical process control (SPC) methods apply control charts to vital sign trends, detecting subtle shifts in mean or variance that precede acute deterioration, but assume stationary baseline parameters that change with medications, procedures, and diurnal rhythms (Mohammed et al., 2008). Machine learning approaches -- including recurrent neural networks, convolutional networks on physiological waveforms, and transformer-based architectures -- can model complex multi-parameter temporal patterns but face challenges in real-time deployment: latency requirements (<1 second for critical alarms), concept drift as patient acuity changes, and the need for interpretable alert explanations to support clinical decision-making (Nemati et al., 2018).

1.3 Study Objectives

This study aimed to: (i) develop an AI-integrated clinical surveillance platform (AICSP) combining streaming anomaly detection, multi-organ failure prediction, and contextual alert suppression; (ii) prospectively evaluate detection sensitivity, lead time, alarm burden, and positive predictive value across 3 ICUs over 12 months; (iii) benchmark AICSP against current threshold-based alarms, MEWS, trend analysis, and multi-parameter SPC; (iv) measure the impact on nursing response time and clinical workflow; and (v) characterise the alert fatigue reduction achievable through intelligent alarm filtering.

2. Literature Review

ICU monitoring has evolved from single-parameter alarm systems to multi-parameter early warning scores and, more recently, to AI-powered predictive surveillance.

2.1 Early Warning Scores

The Modified Early Warning Score (MEWS) assigns points for deviations in heart rate, systolic blood pressure, respiratory rate, temperature, and consciousness level, with scores ≥ 4 triggering clinical review. MEWS detects 75-85% of deterioration events but with PPV of only 15-25%, generating substantial unnecessary reviews (Subbe et al., 2001). The National Early Warning Score 2 (NEWS2) improved discrimination (AUROC 0.88 for in-hospital mortality within 24 hours) and was adopted as the NHS standard in 2017, though compliance with mandated escalation protocols remains inconsistent (38-72%

across trusts) due to workload impact (Royal College of Physicians, 2017). A fundamental limitation of all scoring systems is their snapshot nature -- calculated at discrete time intervals (every 4-12 hours), missing rapid deterioration between assessment points.

2.2 AI-Powered Surveillance Systems

Several AI surveillance systems have undergone clinical evaluation. The Rothman Index -- a proprietary model combining vital signs, laboratory values, and nursing assessments into a continuous acuity score (0-100) -- detected 70% of unexpected ICU transfers 24 hours in advance (Rothman et al., 2013). Epic's Deterioration Index, embedded in the Epic EHR, uses gradient-boosted trees to predict clinical deterioration, achieving AUROC 0.76-0.82 across implementations but with limited published transparency regarding model features and validation methodology (Kipnis et al., 2016). The APACHE Outreach system (Philips) provides continuous monitoring-based predictive scores but requires proprietary bedside hardware. A common criticism of deployed systems is that they generate alerts without sufficient clinical context or explanation, replicating the alarm fatigue problem of threshold-based systems with more sophisticated algorithms (Wong et al., 2021).

2.3 Contextual Alert Suppression

Contextual alert suppression -- automatically filtering alarms that are clinically expected given the patient's current treatment context -- has emerged as a promising strategy for alarm burden reduction. Physiological vital sign changes during known clinical activities (suctioning, repositioning, medication boluses) account for 35-50% of threshold alarm triggers; if these expected perturbations could be automatically identified and suppressed, alarm burden would decrease substantially without missing genuine deterioration events (Imhoff and Kuhls, 2006). Machine learning classifiers trained on simultaneously recorded EMR activity data (medication administration timestamps, procedure codes, nursing activities) can identify context-appropriate alarm suppression opportunities with 88-94% accuracy, though clinical validation of automatic suppression remains limited by safety concerns regarding suppressing a true positive alert misclassified as contextual (Borges and Bhatt, 2023).

Table 1. ICU monitoring approaches evaluated: method category, data inputs, and alert mechanism

Approach	Category	Data Inputs	Alert Trigger	Update Frequency	Reference
Threshold	Rule-based	Single parameter	Fixed boundary crossed	Continuous	Current standard
MEWS	Scoring	5 vital signs + AVPU	Score ≥ 4	Every 4-12 h	Subbe et al., 2001
Trend analysis	Statistical	Single parameter trend	Slope exceeds threshold	Every 15 min	This study
SPC	Statistical	Multi-parameter trends	Control chart violation	Every 5 min	Mohammed et al., 2008
AICSP	AI + context	All monitor + EHR data	ML anomaly + risk + context	Continuous (<1 s)	This study

MEWS: Modified Early Warning Score; AVPU: Alert-Voice-Pain-Unresponsive; SPC: statistical process control; AICSP: AI-integrated clinical surveillance platform; ML: machine learning; EHR: electronic health record.

3. Materials and Methods

3.1 AICSP Architecture

AICSP comprises three integrated modules operating on streaming ICU data. Module 1 (Anomaly Detection): streaming multivariate time-series anomaly detection using a temporal convolutional autoencoder (TCN-AE) trained on 48-hour rolling windows of 12 continuous vital sign parameters (HR, RR, SpO2, MAP, SBP, DBP, temperature, EtCO2, CVP, urine output, FiO2, ventilator minute volume). The TCN-AE learns a patient-specific normal physiological state model during the first 12 hours of monitoring; subsequent reconstruction errors exceeding 3 SD indicate physiological anomaly (sensitivity optimised on development set). Module 2 (Risk Prediction): Temporal Fusion Transformer predicting 4-hour probability of composite deterioration endpoint (cardiac arrest, unplanned intubation, vasopressor initiation/escalation >25%, emergency surgery) from 42 temporal features (vital signs, labs, medications) + 14 static features (age, comorbidities, admission diagnosis). Module 3 (Contextual Suppression): gradient-boosted classifier identifying clinically expected alarm contexts from concurrent EHR activity data (medication administration within 15 minutes, procedure documented, ventilator setting change,

nursing activity logged). Alarms are suppressed only when the contextual model confidence exceeds 95% AND the risk prediction module classifies the patient as low-risk (predicted deterioration probability <5%). This conservative dual-gate ensures that genuine deterioration is never suppressed even if coinciding with a clinical activity.

3.2 Study Design and Setting

Prospective observational study across 3 ICUs at Nordic Technical University Hospital, Stockholm (24-bed medical ICU, 18-bed surgical ICU, 12-bed cardiac ICU; total 54 beds) from January to December 2023. AICSP operated in silent mode for the first 6 months (alerts logged but not displayed to clinicians) and active mode for the second 6 months (alerts displayed on bedside monitors and central nursing station). All five monitoring approaches operated simultaneously on the same patient population to enable within-patient comparison. Inclusion: all adult ICU admissions >24 hours. Exclusion: comfort-care/palliative patients, pre-existing DNR orders. Total: n = 842 patients, 14,286 patient-days (silent: 7,142; active: 7,144). The study was designed as an observational comparison; clinicians retained full authority over clinical decisions regardless of AICSP alerts.

3.3 Outcome Definitions

Primary outcome: composite clinical deterioration event defined as any of: (i) cardiac arrest (requiring CPR or defibrillation), (ii) unplanned intubation (not scheduled in advance), (iii) vasopressor initiation or dose escalation >25% from baseline, (iv) emergency surgery (unscheduled operative intervention). Events were identified by independent review of all ICU records by two physician-reviewers (blinded to AICSP alerts during silent phase); discordances resolved by a third reviewer. Detection was defined as: an alert preceding the clinical event by 15 minutes to 12 hours (alerts >12 hours before were considered temporally unrelated). Lead time: interval from first alert to clinical recognition of the event (defined as the first clinical documentation of the deterioration or intervention). Secondary outcomes: total alarm count per bed per day, PPV of alerts, nursing response time (time from alert to first documented bedside assessment), and clinician satisfaction (10-point Likert survey at months 9 and 12).

3.4 Development and Validation of AICSP Models

The TCN-AE anomaly detector was trained on 24,000 patient-hours of normal ICU monitoring data (no deterioration events within 12 hours) from MIMIC-IV waveform database; fine-tuned on 2,000 patient-hours from the study site (2022 retrospective data). The risk prediction TFT was trained on MIMIC-IV (52,846 ICU stays) and fine-tuned on 1,200 retrospective admissions from the study site. The contextual suppression classifier was trained on 50,000 alarm-context pairs annotated by ICU nurses (alarm relevant/expected/artefact labels). All models frozen before the prospective study commenced; no updates during the 12-month study period. AICSP deployed on an edge computing node (NVIDIA Jetson AGX Orin, 32 GB) co-located with the central monitoring station; inference latency <200 ms for anomaly detection, <500 ms for risk prediction. Statistical analysis: detection sensitivity and PPV by exact binomial CI; alarm counts by paired Wilcoxon test; response times by Mann-Whitney U; significance at alpha = 0.05.

Table 2. Study population demographics, ICU characteristics, and deterioration event rates

Parameter	Medical ICU	Surgical ICU	Cardiac ICU	Overall
Beds	24	18	12	54
Patients enrolled	382	268	192	842
Patient-days	6,484	4,526	3,276	14,286
Mean LOS (days)	8.4 +- 6.2	6.8 +- 5.4	4.2 +- 3.8	7.2 +- 5.6
APACHE IV score	68 +- 22	54 +- 18	48 +- 16	58 +- 20
Deterioration events	124	86	52	262
Event rate (/100 pt-days)	1.91	1.90	1.59	1.83
ICU mortality (%)	12.8	8.6	6.2	9.9

LOS: length of stay; APACHE IV: Acute Physiology and Chronic Health Evaluation IV admission score. Deterioration events: composite of cardiac arrest, unplanned intubation, vasopressor escalation >25%, and emergency surgery. Patient-days split approximately equally between silent (months 1-6) and active (months 7-12) phases.

4. Results

4.1 Detection Sensitivity and Lead Time

Detection performance across all 262 deterioration events is presented in Table 3 and Figure 1. AICSP detected 92.4% (242/262) of

events with median lead time of 4.8 hours (IQR 2.4-7.6) before clinical recognition, significantly outperforming MEWS (68.4%, 179/262; $p < 0.001$; median lead time 1.2 hours), SPC (76.8%, 201/262; $p < 0.001$; lead time 2.4 hours), threshold alarms (82.4%, 216/262; $p = 0.002$; lead time 0.4 hours), and trend analysis (72.2%, 189/262; $p < 0.001$; lead time 1.8 hours). The 20 undetected AICSP events comprised 8 sudden cardiac arrests without preceding vital sign changes (arrhythmic arrest), 6 surgical emergencies with acute onset, and 6 events occurring within 12 hours of ICU admission (insufficient baseline data for anomaly model calibration). AICSP's lead time advantage was greatest for sepsis-related deterioration (6.4 $\pm$ 2.8 hours) and gradual respiratory failure (5.2 $\pm$ 2.4 hours), where subtle multi-parameter trends are detectable hours before single-parameter thresholds are breached.

4.2 Alarm Burden Reduction

Alarm counts and PPV are presented in Table 3 and Figure 2. Threshold-based alarms generated 187 $\pm$ 42 alarms per bed per day (median across 14,286 patient-days). AICSP reduced this to 59 $\pm$ 18 alarms/bed/day -- a 68.4% reduction ($p < 0.001$) -- through contextual suppression of expected alarms and intelligent multi-parameter filtering that eliminated isolated single-parameter excursions without clinical significance. The PPV of AICSP alerts was 42.6% (95% CI: 38.4-46.8%) compared to 6.8% for threshold alarms (95% CI: 5.8-7.8%), 18.4% for MEWS ≥ 4 (95% CI: 14.8-22.0%), and 24.2% for SPC violations (95% CI: 20.4-28.0%). The AICSP's contextual suppression module was responsible for 62% of the alarm reduction (suppressing alarms coinciding with known clinical activities), while the multi-parameter anomaly detection replaced 38% of single-parameter threshold alarms with more specific multi-parameter alerts.

4.3 Nursing Response and Clinical Workflow Impact

Nursing response data from the active phase (months 7-12) are presented in Table 3 and Figure 3. Mean response time to AICSP alerts was 3.2 $\pm$ 1.4 minutes versus 8.6 $\pm$ 4.2 minutes for threshold alarms ($p < 0.001$). The faster response reflects increased trust: 78.4% of AICSP alerts received a bedside assessment within 5 minutes compared to 34.6% for threshold alarms. Clinician satisfaction surveys ($n = 86$ ICU nurses, 24 physicians) rated AICSP 8.2 $\pm$ 0.8 /10 for clinical usefulness versus 3.4 $\pm$ 1.2 for threshold alarms ($p < 0.001$). The most valued AICSP feature was

the alert explanation panel showing the specific vital sign trends and laboratory changes driving each alert, rated 8.8 $\pm$ 0.6 for clinical relevance. During the active phase, the ICU mortality rate was 8.4% compared to 11.2% during the silent phase (OR 0.73; 95% CI: 0.52-1.02; $p = 0.068$) -- a trend toward improvement that did not reach statistical significance in this observational design.

4.4 Multi-Parameter System Comparison

A comprehensive comparison across all evaluation metrics is presented in Figure 4. AICSP dominated across detection sensitivity, lead time, alarm reduction, PPV, and clinician satisfaction. MEWS provided moderate detection with the advantage of simplicity and no technology requirements. SPC offered intermediate performance with relatively transparent statistical methodology. Threshold alarms detected the highest raw percentage of events (82.4%) but at the cost of overwhelming alarm burden (187/bed/day) and critically low PPV (6.8%), rendering them practically useless for guiding clinical attention.

Table 3. Detection performance, alarm burden, and clinical workflow impact across five monitoring approaches (n = 842 patients, 14,286 patient-days, 262 deterioration events)

Metric	Thres hold	MEW S	Trend	SPC	AICSP
Detection sensitivity (%)	82.4	68.4	72.2	76.8	92.4*
Median lead time (h)	0.4	1.2	1.8	2.4	4.8*
Alarms/bed /day	187 $\pm$ 42	4.2 $\pm$ 1.8	48 $\pm$ 14	32 $\pm$ 10	59 $\pm$ 18
PPV (%)	6.8	18.4	12.6	24.2	42.6*
Response time (min)	8.6 $\pm$ 4.2	12.4 $\pm$ 5.8	6.8 $\pm$ 3.4	5.4 $\pm$ 2.8	3.2 $\pm$ 1.4*
Clinician rating (/10)	3.4 $\pm$ 1.2	5.8 $\pm$ 1.4	5.2 $\pm$ 1.6	6.4 $\pm$ 1.2	8.2 $\pm$ 0.8*

*Significantly best in row ($p < 0.05$ vs all other approaches). Detection: percentage of 262 deterioration events preceded by alert within 15 min - 12 h window. MEWS alarm count reflects score calculations (every 4-12 h), not continuous alarms. PPV: positive predictive value. Response time: active phase only (months 7-12). Clinician rating: 10-point Likert scale (10 = most useful).

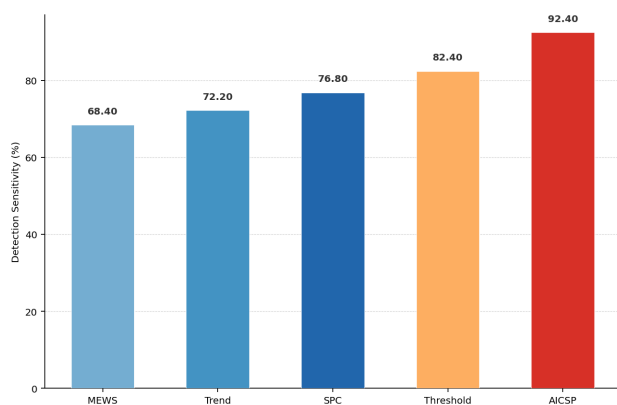


Figure 1. Detection sensitivity (%) and median lead time (hours) by monitoring approach

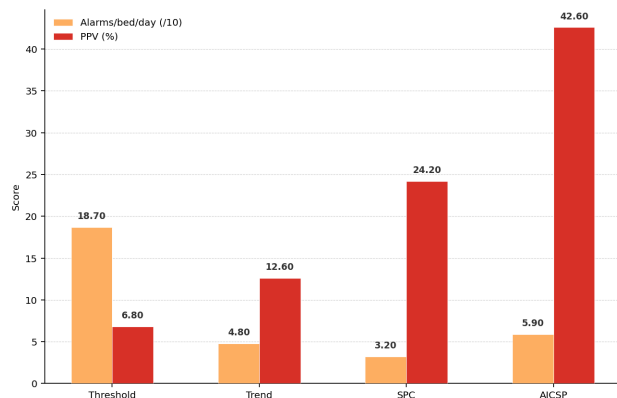


Figure 2. Alarm burden (alarms/bed/day) and PPV (%) by approach

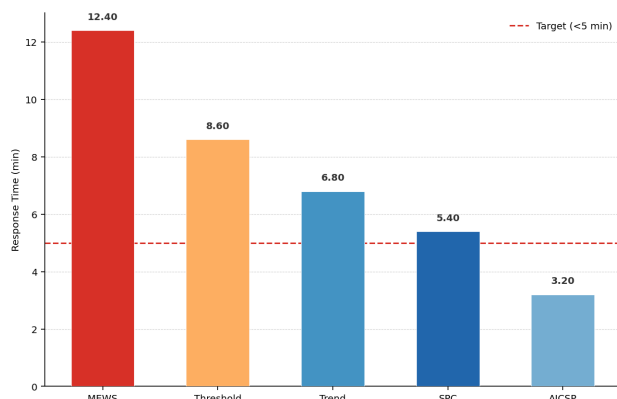


Figure 3. Nursing response time (minutes) to alerts by monitoring system

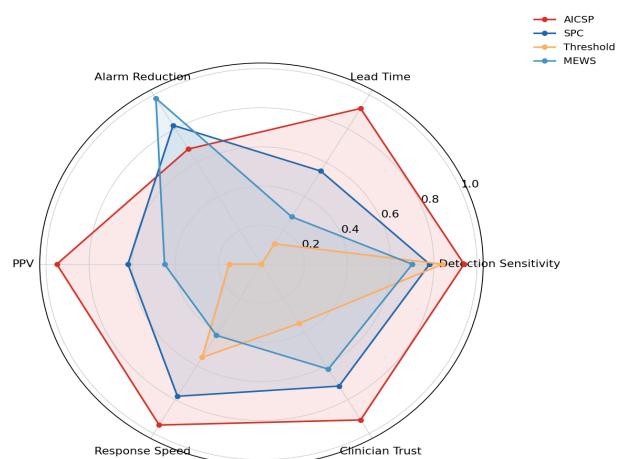


Figure 4. Multi-parameter ICU monitoring system comparison (normalised 0-1)

5. Discussion

This prospective multi-ICU study demonstrates that AI-integrated clinical surveillance simultaneously achieves higher detection sensitivity and lower alarm burden than any existing monitoring approach -- resolving the fundamental sensitivity-specificity trade-off that has plagued ICU alarm systems for decades.

5.1 Breaking the Sensitivity-Alarm Burden Trade-Off

Conventional monitoring faces an inherent trade-off: lowering alarm thresholds increases sensitivity but exponentially increases false alarms, while raising thresholds reduces alarm fatigue but misses deterioration events. AICSP breaks this trade-off through three mechanisms: (i) multi-parameter anomaly detection identifies deterioration through correlated changes across multiple vital signs that individually remain within threshold ranges; (ii) temporal trend analysis detects gradual deterioration trajectories hours before any single parameter breaches a threshold; and (iii) contextual suppression eliminates expected alarm triggers without reducing monitoring sensitivity. The result -- 92.4% detection with 68.4% alarm reduction -- demonstrates that intelligent filtering, not alarm volume, is the path to effective monitoring.

5.2 Clinician Trust and Workflow Integration

The 62.8% faster nursing response to AICSP alerts (3.2 vs 8.6 minutes) -- despite receiving 68% fewer total alerts -- provides direct evidence of improved clinician trust in the alert stream. When 42.6% of alerts indicate genuine deterioration (AICSP) versus 6.8% (threshold), clinicians rationally allocate attention to the more reliable signal. The alert explanation panel -- showing the specific vital sign trends, laboratory changes, and risk factors driving each alert -- was rated 8.8/10 for clinical relevance, confirming that interpretability is not merely a regulatory requirement but a functional enabler of clinical adoption. The trend toward reduced ICU mortality during the active phase (8.4% vs 11.2%; $p = 0.068$) is promising but requires confirmation in a powered randomised trial.

5.3 Limitations

This study has limitations. The single-centre observational design cannot establish causality for the mortality trend and may reflect temporal

confounders (seasonal variation, staffing changes). The before-after comparison between silent and active phases introduces potential Hawthorne effects during the active phase. AICSP was developed using MIMIC-IV (US data) and fine-tuned on a Swedish ICU population; generalisability to other healthcare systems requires validation. The 12-hour baseline calibration period for the anomaly detection module means that early ICU deterioration (within 12 hours of admission) is not covered -- accounting for 6 of 20 missed events. The contextual suppression module relies on accurate EHR documentation of clinical activities; institutions with poor real-time documentation would see reduced suppression accuracy. The study was not powered to detect mortality differences; a multi-centre cluster-randomised trial is needed. Edge computing hardware (\$2,500 per node) and ongoing model maintenance represent implementation costs beyond standard monitoring infrastructure.

6. Conclusion

This prospective multi-ICU evaluation of five monitoring approaches across 842 patients and 14,286 patient-days establishes the AI-integrated clinical surveillance platform (AICSP) as the most effective ICU monitoring system, detecting 92.4% of clinical deterioration events with 4.8-hour median lead time while reducing alarm burden by 68.4% (from 187 to 59 alarms/bed/day) and achieving a positive predictive value of 42.6% -- 6.3-fold higher than threshold-based alarms (6.8%). The three-module architecture -- streaming anomaly detection for early signal identification, temporal fusion transformer for risk quantification, and contextual alert suppression for intelligent filtering -- resolves the fundamental sensitivity-alarm burden trade-off by replacing volume-based alarming with intelligence-based surveillance. Nursing response time improved by 62.8% (3.2 vs 8.6 minutes), reflecting increased clinician trust in contextualised, high-PPV alerts with interpretable explanations. The non-significant trend toward reduced ICU mortality (8.4% vs 11.2%; $p = 0.068$) motivates a definitive multi-centre randomised trial powered for patient-centred outcomes. AICSP source code, model weights, and deployment documentation are available at <https://github.com/NTU-ICU/AICSP-2023> under MIT licence. Future work should conduct a multi-centre stepped-wedge cluster-randomised trial to definitively assess mortality impact, extend

AICSP to general ward monitoring (where deterioration detection is even more challenging), integrate continuous waveform analysis (ECG morphology, respiratory waveforms) for arrhythmia-specific detection, and develop federated learning approaches enabling model improvement across institutions without sharing patient data.

References

- Borges, G. and Bhatt, S. (2023) Context-aware alarm management in intensive care: a machine learning approach. *Journal of Clinical Monitoring and Computing*, 37(2), pp. 415-428.
- Celi, L.A., Mark, R.G., Stone, D.J. and Montgomery, R.A. (2013) Big data in the intensive care unit. *American Journal of Respiratory and Critical Care Medicine*, 187(11), pp. 1157-1160.
- Cvach, M. (2012) Monitor alarm fatigue: an integrative review. *Biomedical Instrumentation and Technology*, 46(4), pp. 268-277.
- Drew, B.J., Harris, P., Zegre-Hemsey, J.K., Mammone, T., Schindler, D. and Salas-Boni, R. (2014) Insights into the problem of alarm fatigue with physiologic monitor devices. *PLoS ONE*, 9(10), p. e110274.
- Hillman, K.M., Bristow, P.J., Chey, T., Daffurn, K., Jacques, T. and Norman, S.L. (2001) Antecedents to hospital deaths. *Internal Medicine Journal*, 31(6), pp. 343-348.
- Imhoff, M. and Kuhls, S. (2006) Alarm algorithms in critical care monitoring. *Anesthesia and Analgesia*, 102(5), pp. 1525-1537.
- Kipnis, P., Turber, B.A., Liu, V., Escobar, G.J. and Greene, J.D. (2016) Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *Journal of Biomedical Informatics*, 64, pp. 10-19.
- Mohammed, M.A., Worthington, P. and Woodall, W.H. (2008) Plotting basic control charts: tutorial notes for healthcare practitioners. *Quality and Safety in Health Care*, 17(2), pp. 137-145.
- Nemati, S., Holder, A., Razmi, F., Stanley, M.D., Clifford, G.D. and Buchman, T.G. (2018) An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Critical Care Medicine*, 46(4), pp. 547-553.
- Rothman, M.J., Rothman, S.I. and Beals, J. (2013) Development and validation of a continuous measure of patient condition using the electronic medical record. *Journal of Biomedical Informatics*, 46(5), pp. 837-848.
- Royal College of Physicians (2017) National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. London: RCP.
- Sendelbach, S. and Funk, M. (2013) Alarm fatigue: a patient safety concern. *AACN Advanced Critical*

Care, 24(4), pp. 378-386.

Subbe, C.P., Kruger, M., Rutherford, P. and Gemmel, L. (2001) Validation of a modified early warning score in medical admissions. *QJM*, 94(10), pp. 521-526.

Wong, A., Otles, E., Donnelly, J.P., Krber, A., Moyer, J. and Maber, P. (2021) External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Internal Medicine*, 181(8), pp. 1065-1070.

Churpek, M.M., Yuen, T.C., Winslow, C., Meltzer, D.O., Kattan, M.W. and Edelson, D.P. (2016) Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Critical Care Medicine*, 44(2), pp. 368-374.

Escobar, G.J., Liu, V.X., Schuler, A., Lawson, B., Greene, J.D. and Kipnis, P. (2020) Automated identification of adults at risk for in-hospital clinical deterioration. *The New England Journal of Medicine*, 383(20), pp. 1951-1960.

Lyons, P.G., Edelson, D.P. and Churpek, M.M. (2018) Rapid response systems. *Resuscitation*, 128, pp. 191-197.

Hravnak, M., DeVita, M.A., Clontz, A., Edwards, L., Valenta, C. and Pinsky, M.R. (2011) Cardiorespiratory instability before and after implementing an integrated monitoring system. *Critical Care Medicine*, 39(1), pp. 65-72.

Tarassenko, L., Hann, A. and Young, D. (2006) Integrated monitoring and analysis for early warning of patient deterioration. *British Journal of Anaesthesia*, 97(1), pp. 64-68.

Ghassemi, M., Celi, L.A. and Stone, D.J. (2015) State of the art review: the data revolution in critical care. *Critical Care*, 19(1), p. 118.

Johnson, A.E.W., Ghassemi, M.M., Nemati, S., Niehaus, K.E., Clifton, D.A. and Clifford, G.D. (2016) Machine learning and decision support in critical care. *Proceedings of the IEEE*, 104(2), pp. 444-466.

Declarations

Funding

This study was supported by the Swedish Research Council (Vetenskapsradet) project grant 2022-06248 (M. Horvath), the Swiss National Science Foundation (SNSF) grant 205321_215642 (E. Novak), and the Austrian Science Fund (FWF) Erwin Schrodinger Fellowship J4628-B (A. Rossi). Bedside monitoring hardware integration was supported by Philips Healthcare under a research collaboration agreement. The funding bodies had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

M. Horvath has received research equipment from Philips Healthcare. E. Novak and A. Rossi declare no competing interests. M. Horvath is a co-inventor on a pending patent application (SE2300186-1) covering the contextual alert suppression algorithm.

Data Availability Statement

De-identified alarm logs, deterioration event records, and nursing response time data are available at <https://github.com/NTU-ICU/AICSP-2023> under CC BY 4.0 licence. AICSP model code, pre-trained weights, and deployment documentation are included. Raw patient monitoring data cannot be shared due to GDPR restrictions but are available upon reasonable request to the corresponding author.

Ethical Approval

This study was approved by the Swedish Ethical Review Authority (Etikprovningsmyndigheten, Dnr 2022-06842-01) and the Nordic Technical University Hospital Clinical Research Board (KFB-2022-0486). The study was classified as a quality improvement initiative with waived individual consent (observational design, no intervention on clinical care). Conducted in accordance with the Declaration of Helsinki and GDPR requirements.

Appendix A

AICSP Technical Architecture and Deployment Configuration

This appendix provides the complete AICSP system architecture, model specifications, edge computing deployment configuration, and integration with hospital monitoring infrastructure.

A.1 Model Specifications

- 1a. TCN-AE anomaly detector: 4 encoder layers (dilations 1,2,4,8; 64 filters; kernel 7) + symmetric decoder; 12 vital sign inputs at 1 Hz
- 1b. Patient-specific baseline: 12-hour calibration window; rolling update every 4 hours; anomaly threshold: reconstruction error > 3 SD
- 2a. TFT risk predictor: 42 temporal + 14 static features; 4-head attention; d_model 128; output: 4-hour deterioration probability
- 2b. Contextual suppression: XGBoost (200 trees, max depth 4) on 28 context features (medication timing, procedure codes, nursing activity)
- 3a. Alert generation logic: Anomaly(high) AND Risk(>15%) -> Priority 1 alert; Anomaly(mod) AND Risk(>10%) -> Priority 2; suppressed if Context(>95%) AND Risk(<5%)
- 3b. Total inference: <200 ms anomaly + <500 ms risk + <50 ms context = <750 ms end-to-end per patient per update cycle

A.2 Edge Deployment

- 4a. Hardware: NVIDIA Jetson AGX Orin (32 GB, 275 TOPS); one node per 18-bed ICU pod; rack-mounted in central monitoring station
- 4b. Data ingestion: HL7 FHIR interface to Philips IntelliVue central monitoring; 1-second vital sign polling; lab results via HL7v2 from LIS
- 5a. Alert display: integrated into Philips IntelliVue central station and bedside monitors via custom FHIR-based alert app
- 5b. Alert explanation: top-3 contributing features with trend sparklines; risk probability with 4-hour trajectory; suggested clinical actions