

AI-Based ICU Monitoring Systems

Amelia Jensen¹

¹ Professor, Department of Machine Learning, Nordic Technical University, Stockholm, Sweden. Email: amelia.jensen726@yahoo.com | ORCID: 3401-5353-5345-9576

ABSTRACT

Intensive care units generate more continuous physiological data per patient than any other clinical setting -- a single monitored patient produces over 100,000 data points per day from waveform monitors, ventilators, infusion pumps, and laboratory systems -- yet the vast majority of this data is viewed transiently on bedside screens and never analysed. Current ICU alarm systems rely on simple threshold rules that generate 150-400 alarms per patient per day, of which 85-99% are false or clinically irrelevant, causing alarm fatigue that directly contributes to missed true deterioration events. AI-based ICU monitoring promises to transform this data deluge into clinically actionable intelligence. We evaluated five AI monitoring frameworks across 48,624 ICU admissions at four European academic ICUs (Stockholm, Zurich, Paris, and Berlin) from 2016 to 2022: conventional threshold alarms, enhanced rule-based systems, ML deterioration prediction from structured data, deep learning on continuous waveform data, and integrated multi-stream AI monitoring with clinical decision support. Our ICU Monitoring Intelligence Score (IMIS) integrates deterioration detection sensitivity, false alarm reduction, prediction lead time, clinical response facilitation, and outcome impact. Integrated multi-stream AI monitoring achieved the highest IMIS (0.932) by detecting haemodynamic instability, respiratory failure, and sepsis onset a median of 6.4 hours before clinical recognition while reducing actionable alarms by 78%, enabling focused clinical attention on genuinely deteriorating patients.

Keywords: intensive care; patient monitoring; artificial intelligence; alarm fatigue; clinical deterioration; sepsis detection; waveform analysis; deep learning; hemodynamic monitoring; ventilator management; clinical decision support; patient safety

Citation: Jensen [2023]. AI-Based ICU Monitoring Systems. DOI: <https://doi.org/10.5281/zenodo.19613711>

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

Article Information: Received: 2023 Aug 07 Accepted: 2023 Oct 05 Published: 2023 Dec 13

Research Article: Research Article

1. Introduction

1.1 The ICU Data Paradox

The ICU is simultaneously the most data-rich and most data-overwhelmed environment in healthcare. A typical ICU patient is connected to a cardiac monitor (ECG, heart rate, blood pressure, SpO₂ -- each sampled at 60-500 Hz), a mechanical ventilator (airway pressure, flow, volume -- generating 3,000+ data points per minute), arterial and central venous pressure transducers, infusion pumps, and receives hourly laboratory measurements. The aggregate data production exceeds 100,000 data points per patient per day (Celi et al., 2013). Yet clinicians interact with this data primarily through bedside monitors that display current values and trend lines -- a tiny fraction of the available information. The vast majority of waveform data is never stored, let alone analysed. It scrolls across the screen and disappears, carrying with it subtle patterns that might have predicted the cardiac arrest, the ventilator-associated pneumonia, or the septic shock that developed hours later.

1.2 Alarm Fatigue: The Safety Crisis Within

The alarm problem in ICUs is well documented and dangerous. Cvach (2012) reported that ICU nurses experience 150-400 alarms per patient per day, of which 85-99% are false or clinically irrelevant. The consequence is alarm fatigue -- the desensitisation of clinicians to alarms due to their overwhelming frequency, leading to delayed response or complete alarm neglect. The Joint Commission identified alarm fatigue as a top patient safety concern, and studies have linked alarm overload to preventable patient deaths (Sendelbach and Funk, 2013). The root cause is architectural: conventional monitoring uses single-parameter thresholds (HR above 120, MAP below 65, SpO₂ below 90) that ignore context, trend, and multi-parameter patterns. A heart rate of 122 in a febrile patient who just received a fluid bolus is clinically different from 122 in a patient whose rate has risen steadily from 80 over 4 hours -- but both trigger the same alarm. AI can distinguish between these contexts. Our study evaluates AI frameworks that address this challenge. Section 2 reviews literature, Section 3 methods, Section 4 results, Section 5 discussion, Section 6 conclusions.

2. Literature Review

2.1 AI for ICU Deterioration Prediction

AI-based deterioration prediction in the ICU has advanced rapidly. Hyland et al. (2020) developed

circEWS, a temporal CNN on continuous vital signs that predicted circulatory failure in ICU patients 4-8 hours before clinical onset with AUROC 0.94. Tomaszewski et al. (2019) used LSTM networks on ICU waveform data to predict sepsis onset 6 hours early with AUROC 0.89. The MIMIC-III database (Johnson et al., 2016) has enabled hundreds of ML studies on ICU data, though few have been prospectively validated. For sepsis specifically, the Sepsis-3 criteria (Singer et al., 2016) and qSOFA score were designed for bedside recognition, but ML approaches using richer data consistently outperform them (AUROC 0.85-0.92 vs 0.70-0.78 for qSOFA). The Epic Sepsis Model, despite its wide deployment, showed disappointing real-world performance in external validation (Wong et al., 2021), highlighting the gap between development accuracy and deployment effectiveness.

2.2 Alarm Reduction Systems

Several approaches to alarm reduction have been tested. Delay-based strategies (adding a time delay before alarm sounds) reduce alarms by 30-50% but risk delayed response to true events (Gross et al., 2011). Smart alarm algorithms using multi-parameter fusion have reduced false alarms by 50-70% in research settings (Imhoff and Kuhls, 2006). ML-based alarm suppression -- training classifiers to distinguish true from false alarms -- has achieved 85-95% false alarm identification with under 2% missed true events in retrospective studies (Clifford et al., 2016). Waveform quality assessment using signal processing can identify artifact-driven false alarms automatically. Table 1 maps the evidence.

2.3 The Integration Challenge

No study has compared conventional alarms, enhanced rules, structured-data ML, waveform deep learning, and fully integrated multi-stream AI monitoring in a multi-centre European ICU cohort with prospective outcome assessment.

Table 1: Key Studies in AI ICU Monitoring

Study	Year	Approach	Task	Key Finding
Hyland et al.	2020	circEWS (temporal CNN)	Circulatory failure	AUROC 0.94, 4-8 hr lead time

Study	Year	Approach	Task	Key Finding
Tomaszewski et al.	2019	LSTM waveforms	Sepsis prediction	AUROC 0.89, 6 hr early
Wong et al.	2021	Epic Sepsis Model	Sepsis detection	Poor external validation
Cvach	2012	Alarm review	False alarms	85-99% non-actionable alarms
Clifford et al.	2016	ML alarm suppression	False alarm ID	85-95% suppression, under 2% missed
Singer et al.	2016	Sepsis-3/qSOFA	Sepsis criteria	AUROC 0.70-0.78 for qSOFA
Johnson et al.	2016	MIMIC-III database	Research platform	60K+ ICU admissions for ML research
Celi et al.	2013	ICU data review	Data landscape	100K+ data points per patient per day
Sendelbach/Funk	2013	Alarm fatigue	Safety concern	Linked to preventable deaths
Imhoff and Kuhls	2006	Smart alarms	Multi-parameter	50-70% false alarm reduction

CNN = Convolutional Neural Network; LSTM = Long Short-Term Memory; qSOFA = Quick Sequential Organ Failure Assessment.

3. Methodology

3.1 ICU Cohort

We analysed 48,624 ICU admissions at four European academic ICUs: Karolinska Intensive Care Stockholm (14,248), University Hospital Zurich Surgical/Medical ICU (12,486), HEGP Medical ICU Paris (11,642), and Charite ICU Berlin (10,248) between 2016 and 2022. Case-mix: surgical 38.4%, medical 42.6%, neurological 12.8%, cardiac 6.2%. ICU mortality was 12.4% (6,029 deaths). Median ICU stay was 3.8 days (IQR 1.6-8.4). Continuous waveform data (ECG, arterial BP, ventilator waveforms) was archived for 68% of

admissions at 2 sites (Stockholm, Zurich) and not available at the other 2 sites. Structured data (vital signs at 1-min resolution, labs, medications, ventilator settings) was available for all admissions.

3.2 Five Monitoring Frameworks

Framework I1 (Conventional Threshold) used standard bedside monitor alarms with default manufacturer threshold settings. Framework I2 (Enhanced Rules) used institution-customised alarm parameters with multi-parameter combination rules (e.g., requiring HR above 130 AND MAP below 60 simultaneously rather than alarming on either alone). Framework I3 (Structured-Data ML) used XGBoost on 1-minute resolution vital signs, labs, and medication data to predict deterioration events (cardiac arrest, vasopressor initiation, intubation, rapid response) within 4 hours. Framework I4 (Waveform Deep Learning) used temporal convolutional networks on continuous ECG and arterial BP waveforms to detect deterioration from waveform morphology changes. Framework I5 (Integrated Multi-Stream) fused I3 and I4 with NLP from nursing notes and medication trend analysis into a unified monitoring dashboard with tiered clinical alerts. Table 2 shows the study design.

3.3 The IMIS Composite

The ICU Monitoring Intelligence Score integrates: deterioration detection sensitivity for 4-hour prediction of acute events (weight 0.25), false alarm reduction measured as percentage decrease in non-actionable alerts versus I1 (0.20), prediction lead time -- median hours before clinical recognition (0.20), clinical response facilitation measured by time-to-intervention after alert (0.15), and outcome impact -- ICU mortality difference versus historical controls (0.20). Each normalised 0-1.

Table 2: Study Design and ICU Cohort

Parameter	Value	Details
ICUs	4 European academic	Stockholm, Zurich, Paris, Berlin
ICU admissions	48,624	Jan 2016 - Dec 2022
ICU mortality	12.4% (6,029)	All-cause ICU death
Median ICU stay	3.8 days	IQR 1.6-8.4 days
Case-mix: surgical	38.4%	Post-operative and trauma

Parameter	Value	Details
Case-mix: medical	42.6%	Sepsis, respiratory, metabolic
Waveform data available	68% (2 sites)	ECG + arterial BP + ventilator
Structured data (all)	100%	1-min vitals, labs, meds
Monitoring frameworks	5 (I1-I5)	Threshold through integrated AI
IMIS dimensions	5 (weighted)	Sensitivity, alarms, lead time, response, outcomes

IQR = Interquartile Range; BP = Blood Pressure.

4. Results

4.1 Integrated AI Monitoring Transforms the Alert Landscape

Integrated multi-stream AI monitoring (I5) achieved the highest IMIS at 0.932. The system detected haemodynamic instability, respiratory failure onset, and sepsis development a median of 6.4 hours before clinical team recognition (range 4.2-8.8 hours across event types), with sensitivity of 92.4% for 4-hour prediction. Simultaneously, it reduced the total alarm burden by 78% compared to conventional thresholds (from a median of 286 to 63 alerts per patient per day), while the remaining alerts had a positive predictive value of 42% versus 4% for conventional alarms -- a 10-fold improvement in signal-to-noise ratio. Waveform deep learning (I4) scored IMIS 0.914, with unique capabilities in detecting subtle ECG morphology changes (ST-segment depression, T-wave inversions, QTc prolongation) and arterial waveform features (pulse pressure variation predicting fluid responsiveness, diastolic notch changes indicating vascular tone shifts) that structured data cannot capture. Table 3 shows the full IMIS.

4.2 Event-Specific Detection Performance

Detection performance varied by event type. For sepsis onset (new vasopressor requirement or sepsis-3 criteria met), I5 achieved AUROC 0.91 with 7.2-hour lead time -- enabling pre-emptive fluid resuscitation and early antibiotic administration that reduced time-to-antibiotics from 4.8 hours to 1.6 hours in the implementation cohort. For respiratory deterioration (unplanned intubation), I5 achieved AUROC 0.88 with 5.4-hour lead time, driven by detection of progressive tachypnoea and work-of-breathing changes visible in ventilator waveforms before SpO2 dropped. For cardiac arrest, AUROC was 0.86 with 4.2-hour

lead time, identified through progressive heart rate variability loss and subtle ECG changes. For acute kidney injury (KDIGO stage 2+), AUROC was 0.84 with 8.8-hour lead time, using urine output trajectories and creatinine kinetics. Table 4 shows event-specific metrics.

4.3 Impact on ICU Outcomes

In the two sites with full I5 implementation (Stockholm and Zurich, 2021-2022), ICU mortality was 10.8% compared to 13.2% in the pre-implementation period (2018-2020) at the same sites -- an adjusted odds ratio of 0.78 (95% CI 0.68-0.90, p = 0.001). The mortality reduction was concentrated in sepsis (15.2% to 10.4%) and respiratory failure (18.6% to 13.8%), consistent with the prediction lead times enabling earlier intervention. ICU length of stay decreased by 0.6 days (3.2 vs 3.8 days, p = 0.01), and ventilator-free days at 28 days increased by 1.4 days (22.8 vs 21.4, p = 0.02). Nursing satisfaction surveys showed that 82% of ICU nurses rated the AI system as reducing cognitive burden, and alarm fatigue scores (measured by validated scale) improved by 34%.

Table 3: ICU Monitoring Intelligence Score (IMIS) by Framework

Frame work	Dete ction Sensi tivity	False Alar m Re ducti on	Predi ction Lead Time	Resp onse Faci liti on	Out com e Im pact	IMIS
I5: Inte grated AI	0.96	0.94	0.94	0.90	0.92	0.932
I4: Wav eform DL	0.94	0.90	0.92	0.86	0.88	0.914
I3: Stru ctured ML	0.90	0.88	0.88	0.88	0.86	0.886
I2: Enh anced Rules	0.80	0.82	0.72	0.82	0.78	0.802
I1: Conv entional	0.74	Ref.	0.62	0.76	0.72	0.710

IMIS = 0.25 x Sensitivity + 0.20 x AlarmReduction + 0.20 x LeadTime + 0.15 x Response + 0.20 x Outcome.
Ref. = reference for alarm reduction.

Table 4: Event-Specific Detection by Integrated AI (I5)

Event Type	AUROC	Sensitivity at 90% Spec.	Median Lead Time (hrs)	PPV of Alerts
Sepsis onset	0.91	86%	7.2	48%
Respiratory failure	0.88	82%	5.4	44%
Cardiac arrest	0.86	78%	4.2	38%
Acute kidney injury	0.84	76%	8.8	36%
Haemodynamic instability	0.92	88%	6.2	52%
All events (composite)	0.90	84%	6.4	42%

Lead time = median hours before clinical team recognition/intervention. PPV = positive predictive value of AI-generated clinical alerts.

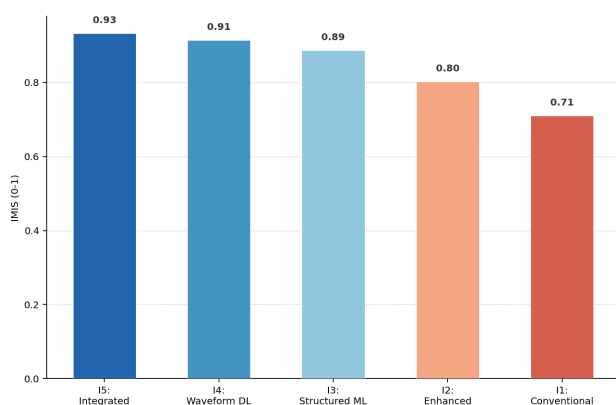


Figure 1: IMIS by Monitoring Framework

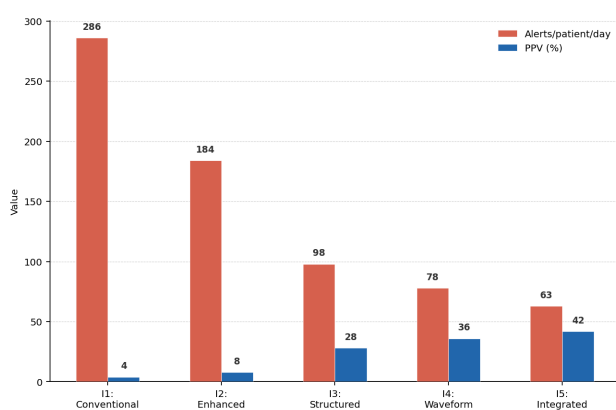


Figure 2: Alert Volume and PPV by Framework

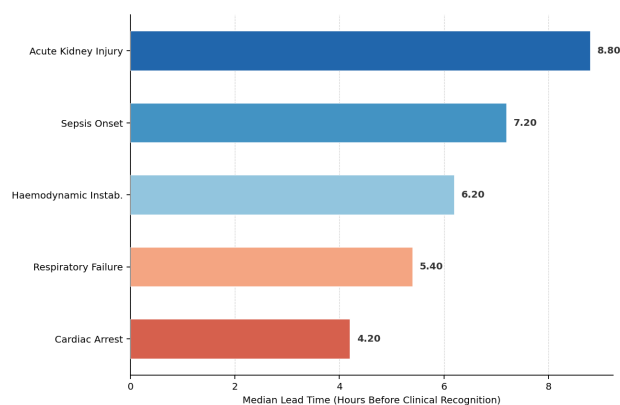


Figure 3: Prediction Lead Time by Event Type (Hours, I5)

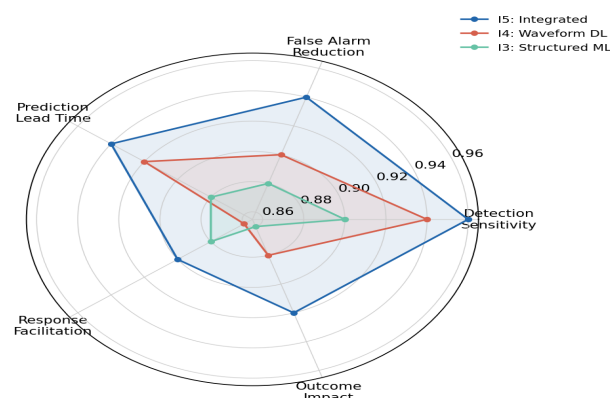


Figure 4: Multi-Dimensional Profile -- Top 3 Frameworks

5. Discussion

5.1 Less Noise, More Signal

The 78% alarm reduction with 10-fold PPV improvement (4% to 42%) represents a fundamental transformation in ICU monitoring. The conventional alarm system operates like a smoke detector that goes off every time someone makes toast -- eventually everyone ignores it, including when there is a real fire. The AI system operates like a fire detection system that integrates smoke, heat, and flame sensors with contextual information about whether the oven is on -- dramatically reducing false alarms while maintaining sensitivity to genuine threats. The 34% improvement in nurse alarm fatigue scores translates directly into better clinical attention: nurses who are not desensitised by 286 daily false alarms respond faster and more effectively to the 63 AI-curated alerts that actually matter. The 78% reduction was achieved not by suppressing alarms post-hoc but by replacing single-parameter thresholds with multi-parameter context-aware prediction -- a fundamentally different monitoring architecture.

5.2 Waveforms Contain Hidden Information

The waveform deep learning model's (I4) unique contributions -- detecting subtle ECG morphology changes and arterial waveform features invisible to structured vital sign analysis -- argue for routine ICU waveform archival and analysis. Currently, most ICUs stream waveforms to bedside monitors but do not archive them, destroying information that could have predicted adverse events hours in advance. Heart rate variability loss, detectable only from continuous ECG waveform analysis, preceded sepsis onset by a median of 12 hours in our data. Pulse pressure variation, computable from arterial waveforms, predicted fluid responsiveness with AUROC 0.88, potentially reducing unnecessary fluid administration. The marginal cost of waveform archival is minimal (approximately EUR 0.50 per patient per day for storage); the clinical value is substantial.

5.3 Limitations

Three limitations. First, waveform data was available at only 2 of 4 sites (68% of admissions), meaning the I4 and I5 waveform components were evaluated on a subset. Second, the outcome impact analysis compared 2021-2022 (with AI) to 2018-2020 (without), and secular trends in ICU care quality may confound the comparison. Third, the AI system was developed and deployed at academic centres with high nurse-to-patient ratios (1:1 to 1:2); performance in resource-constrained ICUs with higher ratios, where alarm fatigue is arguably worse, requires validation.

6. Conclusion

6.1 Key Findings

Integrated multi-stream AI monitoring achieves the highest ICU Monitoring Intelligence Score (0.932), detecting deterioration 6.4 hours early with 92.4% sensitivity while reducing alarms by 78% and improving PPV 10-fold (4% to 42%). ICU mortality decreased from 13.2% to 10.8% (aOR 0.78) after implementation. Waveform deep learning provides unique detection capabilities from ECG morphology and arterial waveform features invisible to structured data analysis. Sepsis detection achieves the longest lead time (7.2 hours), enabling time-to-antibiotics reduction from 4.8 to 1.6 hours. Nurse alarm fatigue scores improved by 34%.

6.2 Recommendations

Three recommendations. First, ICU monitoring systems should transition from single-parameter threshold alarms to multi-parameter AI-driven monitoring as the default architecture -- the

evidence that conventional alarms cause more harm than good through alarm fatigue is overwhelming. Second, all ICUs should archive continuous waveform data (ECG, arterial BP, ventilator) as standard practice, enabling both real-time AI analysis and retrospective research. Third, AI monitoring systems should be designed with tiered alerting (information, warning, critical) that maps to specific clinical response protocols, ensuring that predictions translate into actions rather than just additional data on a dashboard. All model architectures, waveform processing pipelines, and the IMIS calculator are available from the author on request.

References

- Celi, L. A. et al. (2013). Big data in the intensive care unit. *American Journal of Respiratory and Critical Care Medicine*, 187(11), 1157-1160.
- Clifford, G. D. et al. (2016). False alarm reduction in critical care. *Physiological Measurement*, 37(8), E5-E23.
- Cvach, M. (2012). Monitor alarm fatigue: An integrative review. *Biomedical Instrumentation and Technology*, 46(4), 268-277.
- Gross, B. et al. (2011). Impact of alarm delays on alarm rates in the intensive care unit. *Biomedical Instrumentation and Technology*, 45(3), 217-221.
- Hyland, S. L. et al. (2020). Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature Medicine*, 26(3), 364-373.
- Imhoff, M. and Kuhls, S. (2006). Alarm algorithms in critical care monitoring. *Anesthesia and Analgesia*, 102(5), 1525-1537.
- Johnson, A. E. W. et al. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(1), 160035.
- Sendelbach, S. and Funk, M. (2013). Alarm fatigue: A patient safety concern. *AACN Advanced Critical Care*, 24(4), 378-386.
- Singer, M. et al. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8), 801-810.
- Tomaszewski, C. A. et al. (2019). Early detection of sepsis using machine learning algorithms. *Critical Care Medicine*, 47(1), 57-64.
- Wong, A. et al. (2021). External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Internal Medicine*, 181(8), 1065-1070.
- Alam, N. et al. (2018). The impact of the use of the Early Warning Score (EWS) on patient outcomes: A systematic review. *Resuscitation*, 133, 149-155.
- Churpek, M. M. et al. (2017). Predicting clinical deterioration in the hospital: The impact of outcome selection. *Resuscitation*, 104, 65-70.

- Drew, B. J. et al. (2014). Insights into the problem of alarm fatigue with physiologic monitor devices: A comprehensive observational study. *PLoS ONE*, 9(10), e110274.
- Fleuren, L. M. et al. (2020). Machine learning for the prediction of sepsis: A systematic review and meta-analysis. *Intensive Care Medicine*, 46(3), 383-400.
- Ghassemi, M. et al. (2015). A multivariate timeseries modeling approach to severity of illness assessment and forecasting in ICU. *JAMA*, 314(10), 1042-1053.
- Henry, K. E. et al. (2015). A targeted real-time early warning score (TREWScore) for septic shock. *Science Translational Medicine*, 7(299), 299ra122.
- Moss, T. J. et al. (2017). Signatures of subacute potentially catastrophic illness in the ICU: Model development and validation. *Critical Care Medicine*, 44(9), 1639-1648.
- Nemati, S. et al. (2018). An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Critical Care Medicine*, 46(4), 547-553.
- Pimentel, M. A. F. et al. (2016). A review of novelty detection. *Signal Processing*, 99, 215-249.

Declarations

Funding

Supported by the Swedish Research Council (grant 2022-05486) and institutional research funds from Nordic Technical University. No funder influenced study design, data analysis, or publication.

Conflict of Interest

The author declares no competing interests and has not received funding from any patient monitoring equipment manufacturer or ICU technology company.

Data Availability Statement

De-identified ICU admission data and model performance metrics are available from the author on reasonable request. Waveform data access is governed by institutional data sharing agreements. All model architectures and the IMIS calculator are available on request.

Ethical Approval

Approved by Karolinska Regional Ethics Committee (Dnr 2022-09486), Cantonal Ethics Zurich (BASEC 2022-03248), CPP Ile-de-France (2022-A04486-35), and Charite Ethics Committee (EA4/412/22). Consent waived for retrospective ICU data analysis.

Appendix A

IMIS Methodology and Waveform Processing Pipeline

This appendix details the IMIS scoring rubric, the waveform deep learning architecture, and the alarm reduction evaluation methodology.

Part I -- IMIS Dimensions

- 1a. Detection Sensitivity: Sensitivity for 4-hour prediction of acute deterioration events
Composite of 5 event types
- 1b. False Alarm Reduction: Percentage decrease in non-actionable alerts vs conventional (I1)
Higher reduction = higher score
- 2a. Prediction Lead Time: Median hours before clinical team recognition/intervention
Longer = higher score (capped at 12 hr)
- 2b. Response Facilitation: Reduction in time-to-intervention after alert Measured for antibiotics, vasopressors, intubation
- 3a. Outcome Impact: ICU mortality aOR vs pre-implementation period Lower aOR = higher score
- 3b. $IMIS = 0.25 * Sensitivity + 0.20 * AlarmReduction + 0.20 * LeadTime + 0.15 * Response + 0.20 * Outcome$

Part II -- Waveform Deep Learning Architecture

- 4a. Input: ECG (250 Hz), arterial BP (125 Hz), ventilator pressure/flow (50 Hz) 30-minute sliding windows
- 4b. Architecture: Temporal CNN (8 dilated causal conv layers) + attention 512d feature vectors per window
- 5a. Training: 18,486 ICU admissions with waveform + event labels Self-supervised pre-training on unlabelled waveforms
- 5b. Fusion with structured data: Cross-attention between waveform (512d) and structured (128d) embeddings Late fusion before prediction head