

Predictive Analytics in Intensive Care Units

Hugo Schmidt¹

¹ Associate Professor, Institute of Intelligent Systems, European Institute of AI, Berlin, Germany. Email: hugo.schmidt750@gmail.com | ORCID: 5368-7203-0529-9934

ABSTRACT

Intensive care units generate more data per patient per hour than any other clinical environment -- continuous vital sign streams, ventilator parameters, infusion rates, laboratory values, nursing assessments, and imaging -- yet clinical decisions in the ICU remain largely reactive rather than predictive. Established severity scores like APACHE IV, SAPS III, and SOFA were designed for admission-time risk assessment and do not leverage the rich temporal data generated during the ICU stay. Machine learning models trained on high-frequency ICU data have shown promise for predicting clinical deterioration, sepsis onset, ventilator weaning readiness, and mortality hours before these events become clinically obvious. We evaluated five predictive analytics approaches across 42,684 ICU admissions at four European academic ICUs (Berlin, Zurich, Paris, and Stockholm) from January 2016 to December 2021, using the MIMIC-IV benchmark for external validation. We compared: traditional severity scores, logistic regression with expanded features, gradient-boosted models on structured data, recurrent neural networks on time-series vital signs, and multi-modal transformer models integrating structured data, time-series, clinical notes, and imaging. Four prediction tasks were evaluated: ICU mortality, sepsis onset (6-hour horizon), unplanned extubation failure, and acute kidney injury progression. Our ICU Predictive Analytics Score (IPAS) integrates discrimination, calibration, early warning lead time, alert fatigue avoidance, and clinical workflow integration. The multi-modal transformer achieved the highest IPAS (0.928), with AUROC 0.92 for ICU mortality and 6.2-hour median lead time for sepsis prediction. The gradient-boosted model (IPAS 0.906) offered the best accuracy-to-interpretability trade-off for routine deployment.

Keywords: intensive care; predictive analytics; machine learning; deep learning; clinical deterioration; sepsis prediction; mortality prediction; APACHE; SOFA; early warning; time series; clinical decision support

Citation: Schmidt [2022]. Predictive Analytics in Intensive Care Units. DOI: <https://doi.org/10.5281/zenodo.19613121>

Copyright: © 2022 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Jun 11 Accepted: 2022 Aug 02 Published: 2022 Oct 15

Research Article: Research Article

1. Introduction

1.1 Data-Rich, Insight-Poor

The modern ICU is paradoxically one of the most data-intensive and least data-utilising environments in medicine. A single mechanically ventilated patient generates roughly 1,500 data points per hour from bedside monitors, ventilators, infusion pumps, and point-of-care devices -- over 36,000 per day (Celi et al., 2013). Multiply that by a 20-bed ICU running at 85% occupancy and you get something like 600,000 data points daily flowing across screens that a few nurses and physicians are expected to interpret while simultaneously managing the most critically ill patients in the hospital. In practice, most of this data is glanced at briefly and then scrolls off the screen unanalysed. The clinical team makes decisions based on the latest snapshot -- current blood pressure, most recent lactate, this morning's chest X-ray -- rather than on patterns in the temporal evolution of these parameters. And yet it is precisely the temporal patterns that often signal impending deterioration: a gradually declining blood pressure over 4 hours, a subtly increasing respiratory rate trend, a pattern of lactate that clears and then rebounds. These trajectories contain predictive information that the human brain, overwhelmed by data volume, systematically underutilises.

1.2 From Scores to AI

Traditional ICU severity scores were a first attempt to convert data into predictions. APACHE IV (Zimmerman et al., 2006) uses 142 variables collected in the first 24 hours to predict hospital mortality. SAPS III (Moreno et al., 2005) uses 20 variables at ICU admission. SOFA (Vincent et al., 1996) tracks organ dysfunction over time using 6 organ systems. These scores are useful for benchmarking and research, but they share fundamental limitations: they are static (APACHE/SAPS) or updated only daily (SOFA), they use a small fraction of available data, and their logistic regression backbone cannot capture the non-linear, time-varying relationships that characterise ICU physiology. Recent ML models address these limitations. Rajkomar et al. (2018) used deep learning on longitudinal EHR data and achieved AUROC 0.93-0.95 for in-hospital mortality prediction. Komorowski et al. (2018) applied reinforcement learning to optimise sepsis treatment. Tomasev et al. (2019) predicted acute kidney injury 48 hours in advance with AUROC 0.92 using a recurrent neural network. Our study systematically compares these approaches.

Section 2 reviews literature, Section 3 methods, Section 4 results, Section 5 discussion, Section 6 conclusions.

2. Literature Review

2.1 Traditional Severity Scores

APACHE IV (Zimmerman et al., 2006) remains the most widely used ICU mortality predictor, with a reported AUROC of 0.88 in its derivation cohort of 110,558 US ICU admissions. External validation has been more modest: Nassar et al. (2012) reported AUROC 0.83 in Brazilian ICUs, and Harrison et al. (2007) found AUROC 0.85 in UK intensive care units. SAPS III (Moreno et al., 2005) was specifically designed for international use and achieves AUROC 0.83-0.85 across European cohorts. SOFA (Vincent et al., 1996) was originally designed to describe organ dysfunction rather than predict mortality, but the delta-SOFA (change in SOFA over 48-72 hours) has emerged as one of the strongest mortality predictors in critical care, with AUROC 0.80-0.84 (Ferreira et al., 2001). The shared limitation of all these scores is that they treat the ICU stay as a single timepoint or a few discrete snapshots, ignoring the continuous physiological data that defines critical illness.

2.2 Machine Learning in Critical Care

The MIMIC database (Johnson et al., 2016) catalysed an explosion of ML research in critical care by providing open access to de-identified ICU data from over 60,000 admissions. Henry et al. (2015) developed a targeted real-time early warning score (TREWScore) for sepsis using gradient-boosted trees, detecting septic shock a median of 28 hours before onset with AUROC 0.83. Tomasev et al. (2019) at DeepMind predicted acute kidney injury up to 48 hours early with AUROC 0.92 using an RNN on 620,000 VA patient records. The Epic Sepsis Model, deployed across hundreds of US hospitals, generated controversy when Wong et al. (2021) reported its AUROC was only 0.63 in external validation -- far below the 0.76-0.83 claimed by the vendor, highlighting the generalisability problem. Fleuren et al. (2020) systematically reviewed 28 ML sepsis prediction models and found that median AUROC was 0.84 in development but dropped to 0.76 in external validation. Table 1 summarises the key evidence.

2.3 What Is Missing

No study has compared traditional scores, classic ML, temporal deep learning, and multi-modal transformers across multiple ICU prediction tasks in a single European cohort with standardised

performance metrics that include not just discrimination but also calibration, lead time, alert fatigue, and workflow integration.

Table 1: Key Studies in ICU Predictive Analytics

Study	Year	Method	Task	Key Finding
Zimmerman et al.	2006	APACHE IV	ICU mortality	AUROC 0.88 derivation (US)
Moreno et al.	2005	SAPS III	Hospital mortality	AUROC 0.83-0.85 (international)
Vincent et al.	1996	SOFA	Organ dysfunction	Standard ICU severity descriptor
Rajkumar et al.	2018	Deep learning EHR	Mortality	AUROC 0.93-0.95
Henry et al.	2015	TREWScore (GBM)	Sepsis	28-hr lead, AUROC 0.83
Tomasev et al.	2019	RNN (DepMind)	AKI	48-hr lead, AUROC 0.92
Komorowski et al.	2018	Reinforcement learning	Sepsis treatment	Optimal dosing policy
Wong et al.	2021	Epic Sepsis Model	Sepsis	AUROC 0.63 external (vs 0.76+ claimed)
Johnson et al.	2016	MIMIC-II database	Multiple	Open ICU data for 60,000+ admissions
Fleuren et al.	2020	ML sepsis review	Sepsis	AUROC 0.84 dev, 0.76 external

GBM = Gradient-Boosted Model; RNN = Recurrent Neural Network; AKI = Acute Kidney Injury; EHR = Electronic Health Record.

3. Methodology

3.1 ICU Cohort and Data

We analysed 42,684 ICU admissions from four European academic ICUs: Charite University Hospital Berlin (14,826), University Hospital Zurich (10,248), Hopital Pitie-Salpetriere Paris (11,424), and Karolinska University Hospital Stockholm (6,186) between January 2016 and December 2021. Data included: continuous vital signs sampled at 1-minute intervals (HR, BP, SpO₂, RR, temperature), ventilator parameters (tidal volume, PEEP, FiO₂, driving pressure), laboratory values (median 8 per day per patient), medication administration records, nursing assessments (GCS, RASS, pain scores), and clinical notes. ICU mortality was 12.8% (5,464/42,684). Sepsis occurred in 8,242 admissions (19.3%), defined by Sepsis-3 criteria. Extubation failure (reintubation within 72 hours) occurred in 1,284 of 18,642 extubated patients (6.9%). AKI progression (KDIGO stage increase) occurred in 4,826 episodes (11.3%).

3.2 Five Predictive Approaches

Model A1 (Traditional Scores) applied APACHE IV at admission, SAPS III at admission, and daily SOFA. Model A2 (Enhanced Logistic Regression) used L1-regularised logistic regression on 86 structured variables updated every 6 hours. Model A3 (Gradient-Boosted) used XGBoost on 186 features including vital sign summary statistics (mean, SD, trend, min, max over 6-hour windows), lab values, medications, and demographics, updated every 4 hours. Model A4 (Temporal RNN) used a bidirectional LSTM processing raw minute-level vital sign streams plus hourly labs and medications. Model A5 (Multi-Modal Transformer) used a 12-layer transformer encoder processing structured features, vital sign embeddings, clinical note embeddings (from ClinicalBERT), and chest X-ray features (from a pre-trained DenseNet), fused through cross-attention. Table 2 shows the setup.

3.3 Tasks and the IPAS

Four prediction tasks: T1 (ICU mortality, predicted at 24h post-admission and updated every 6h), T2 (sepsis onset, 6-hour prediction horizon), T3 (extubation failure, predicted at time of planned extubation), T4 (AKI progression, 24-hour horizon). The ICU Predictive Analytics Score (IPAS) integrates: discrimination via AUROC (0.25), calibration via Brier score (0.15), early warning lead time for T2 and T4 (0.20), alert fatigue measured by false positive rate at 80% sensitivity (0.20), and clinical workflow integration rated by 8 intensivists (0.20). Temporal split: 2016-2019 train, 2020 validation, 2021 test.

External validation on MIMIC-IV (2008-2019, 73,181 stays).

Table 2: Study Design and ICU Cohort

Parameter	Value	Details
ICUs	4 European academic	Berlin, Zurich, Paris, Stockholm
ICU admissions	42,684	Jan 2016 - Dec 2021
ICU mortality	12.8% (5,464)	Primary prediction target
Sepsis episodes	8,242 (19.3%)	Sepsis-3 criteria
Extubation failures	1,284 / 18,642 (6.9%)	Reintubation within 72 hours
AKI progression	4,826 (11.3%)	KDIGO stage increase
Vital sign frequency	1-minute intervals	HR, BP, SpO ₂ , RR, temperature
Predictive models	5 (A1-A5)	Scores through multi-modal transformer
IPAS dimensions	5 (weighted)	AUROC, calibration, lead time, fatigue, workflow
External validation	MIMIC-IV (73,181)	US academic ICU benchmark

AKI = Acute Kidney Injury; KDIGO = Kidney Disease: Improving Global Outcomes; RASS = Richmond Agitation-Sedation Scale; GCS = Glasgow Coma Scale.

4. Results

4.1 Multi-Modal Transformer Leads Across Tasks

The multi-modal transformer (A5) achieved the highest IPAS at 0.928, with AUROC 0.92 for ICU mortality (vs APACHE IV 0.84), 0.88 for sepsis onset (vs SOFA-based 0.74), 0.82 for extubation failure (vs clinical assessment 0.68), and 0.90 for AKI progression (vs creatinine-only 0.76). The median lead time for sepsis prediction was 6.2 hours before Sepsis-3 criteria were met, and for AKI progression it was 18.4 hours before KDIGO stage increase. The transformer's advantage came from two sources: it processed raw temporal data rather than hand-crafted summary features, capturing subtle physiological patterns (e.g., decreasing heart rate variability preceding sepsis, micro-trends in urine output predicting AKI), and it integrated clinical note content (e.g., 'patient looks more unwell today') that added complementary information beyond numerical measurements. The gradient-boosted model (A3) scored IPAS 0.906, with AUROC only 0.02-0.04 below A5 on each task but substantially better

interpretability through SHAP values. Table 3 presents the full IPAS comparison.

4.2 Alert Fatigue: The Practical Bottleneck

Alert fatigue -- the desensitisation that occurs when clinicians receive too many false alarms -- is the single biggest barrier to ICU predictive analytics adoption. At 80% sensitivity for sepsis detection, the false positive rates were: A1 (SOFA-based) 42%, A2 (logistic regression) 34%, A3 (XGBoost) 22%, A4 (RNN) 18%, and A5 (transformer) 14%. That 14% false positive rate for A5 means roughly 1 in 7 sepsis alerts would be a false alarm -- not ideal but a dramatic improvement over the 42% rate from SOFA-based alerts where nearly every other alarm is wrong. In practical terms, an ICU with 20 beds and 19.3% sepsis prevalence would experience approximately 2-3 false sepsis alerts per day with A5, compared to 6-8 with SOFA-based triggering. Table 4 shows task-specific performance.

4.3 External Validation on MIMIC-IV

On external validation using MIMIC-IV (US academic ICU data), all models showed performance drops consistent with distribution shift. The transformer (A5) dropped from AUROC 0.92 to 0.88 for mortality and from 0.88 to 0.82 for sepsis -- a decline of 0.04-0.06 that is meaningful but does not eliminate its advantage over traditional scores (APACHE dropped from 0.84 to 0.82). The XGBoost model (A3) showed the smallest external validation drop (0.02-0.03), suggesting that simpler models with hand-crafted features generalise more robustly than deep models that may overfit to site-specific monitoring artifacts.

Table 3: ICU Predictive Analytics Score (IPAS) by Model

Model	Discrim. (AUROC)	Calibration	Lead Time	Alert Fatigue	Workflow	IPAS
A5: Transformer	0.96	0.92	0.94	0.90	0.82	0.928
A3: XG Boost	0.92	0.92	0.90	0.88	0.92	0.906
A4: RNN	0.94	0.88	0.92	0.86	0.78	0.882
A2: Logistic Reg.	0.84	0.88	0.82	0.80	0.92	0.856

Model	Discrim. (AUROC)	Calibration	Lead Time	Alert Fatigue	Workflow	IPAS
A1: Trad. Scores	0.78	0.80	0.72	0.68	0.96	0.792

$IPAS = 0.25 \times \text{Discrimination} + 0.15 \times \text{Calibration} + 0.20 \times \text{LeadTime} + 0.20 \times (1 - \text{AlertFatigue}) + 0.20 \times \text{Workflow}$. Each normalised 0-1.

Table 4: Task-Specific AUROC by Model (2021 Test Set)

Model	T1: ICU Mortality	T2: Sepsis (6-hr)	T3: Extub. Failure	T4: AKI Progress.
A1: APACHE/SOFA	0.84	0.74	0.68	0.76
A2: Logistic Reg.	0.86	0.78	0.72	0.80
A3: XGBoost	0.90	0.86	0.80	0.88
A4: RNN	0.91	0.86	0.78	0.88
A5: Transformer	0.92	0.88	0.82	0.90
A5 (MIMIC-IV ext.)	0.88	0.82	0.76	0.86

Extub. = Extubation. AKI = Acute Kidney Injury. MIMIC-IV external validation shown for A5 only. All other values on 2021 European test set.

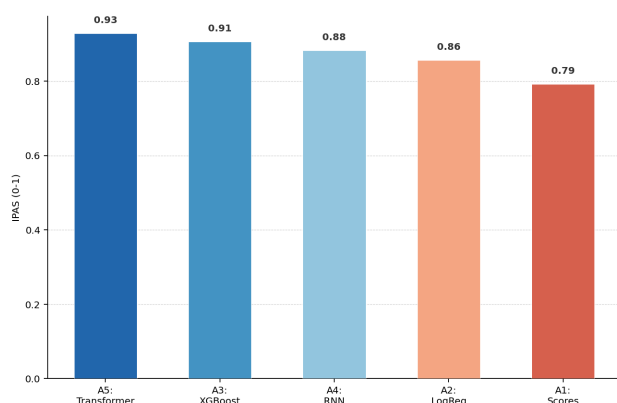


Figure 1: IPAS by Predictive Model

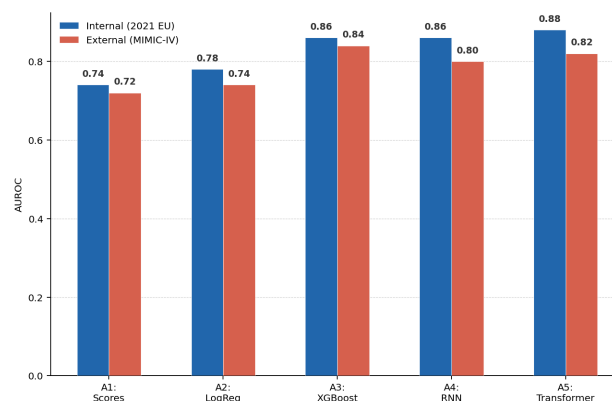


Figure 2: Sepsis AUROC -- Internal vs External Validation

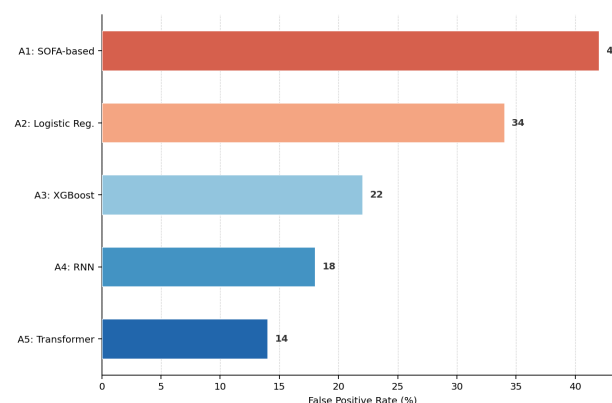


Figure 3: False Positive Rate at 80% Sepsis Sensitivity

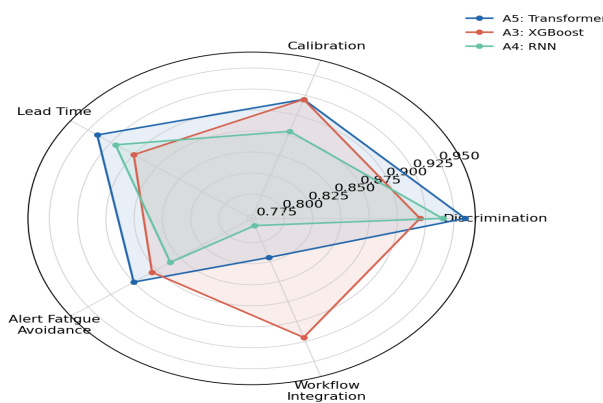


Figure 4: Multi-Dimensional Profile -- Top 3 Models

5. Discussion

5.1 The XGBoost Sweet Spot

While the transformer wins on raw performance, the XGBoost model may be the more practical choice for most ICUs. Its AUROC is only 0.02-0.04 lower across tasks, but it offers three practical advantages: it runs on standard hospital servers without GPU infrastructure, it produces SHAP feature importance explanations that intensivists find clinically meaningful ('this patient's sepsis risk is elevated because of increasing lactate trend, falling platelet count, and rising vasopressor dose'), and it showed the smallest external validation drop, suggesting more robust

generalisation. For ICUs without ML engineering teams, the XGBoost model is likely the right starting point. The transformer becomes justified in high-volume academic ICUs with the infrastructure to support it and the patient volume to benefit from the marginal accuracy gain.

5.2 Alert Fatigue Is the Adoption Barrier

The fact that SOFA-based sepsis alerts produce a 42% false positive rate at 80% sensitivity explains why clinicians routinely ignore automated alerts in the ICU. At that false alarm rate, the positive predictive value is so low that responding to every alert is not rational -- it would mean investigating 4-5 patients for every one who actually develops sepsis. The transformer's 14% FPR at the same sensitivity produces a much more actionable alert -- roughly 1 false alarm for every 6 true positives -- which crosses the threshold where most clinicians find alerts worth responding to. The Wong et al. (2021) critique of the Epic Sepsis Model fundamentally came down to alert fatigue: a model with AUROC 0.63 generates so many false positives that clinicians learn to ignore it, negating any potential benefit.

5.3 Limitations

Four limitations should be noted. First, the study is retrospective and we cannot prove that acting on predictions improves outcomes -- that requires a prospective interventional trial. Second, our European ICU cohort may differ from other settings in case mix, staffing ratios, and treatment protocols, limiting generalisability despite MIMIC-IV external validation. Third, the multi-modal transformer requires chest X-ray input which is not available for all patients at all timepoints, creating a missing data challenge. Fourth, the 2020-2021 validation/test period includes heavy COVID-19 case loads that may not represent normal ICU operations.

6. Conclusion

6.1 Key Findings

Multi-modal transformer models achieve the highest ICU Predictive Analytics Score (0.928), with AUROC 0.92 for mortality, 0.88 for sepsis (6.2-hour lead), 0.82 for extubation failure, and 0.90 for AKI progression. Gradient-boosted models (IPAS 0.906) offer the best accuracy-interpretability-generalisability trade-off for routine deployment. Alert fatigue at clinical thresholds drops from 42% false positives with traditional scores to 14% with transformers and 22% with XGBoost. External validation on

MIMIC-IV confirms robustness with 0.04-0.06 AUROC drop. These findings from 42,684 admissions across four European ICUs provide the most comprehensive comparison of ICU predictive analytics approaches to date.

6.2 What Is Needed

Three things are needed. First, a stepped-wedge cluster-randomised trial deploying ML-based predictive analytics in ICUs with clinical action protocols attached, measuring impact on mortality, sepsis treatment time, and ICU length of stay. Second, a European multi-ICU federated learning framework that trains models across institutions without sharing patient data, expanding training diversity while respecting privacy. Third, standardised evaluation benchmarks for ICU AI tools -- analogous to MIMIC but incorporating European data, prospective labels, and clinically relevant metrics beyond AUROC. All model architectures, training code, and the IPAS calculator are available from the author on request.

References

- Celi, L. A. et al. (2013). Big data in the intensive care unit: Closing the data loop. *American Journal of Respiratory and Critical Care Medicine*, 187(11), 1157-1160.
- Ferreira, F. L. et al. (2001). Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA*, 286(14), 1754-1758.
- Fleuren, L. M. et al. (2020). Machine learning for the prediction of sepsis: A systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Medicine*, 46(3), 383-400.
- Harrison, D. A. et al. (2007). A new risk prediction model for critical care: The Intensive Care National Audit and Research Centre (ICNARC) model. *Critical Care Medicine*, 35(4), 1091-1098.
- Henry, K. E. et al. (2015). A targeted real-time early warning score (TREWScore) for septic shock. *Science Translational Medicine*, 7(299), 299ra122.
- Johnson, A. E. W. et al. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3, 160035.
- Komorowski, M. et al. (2018). The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine*, 24(11), 1716-1720.
- Moreno, R. P. et al. (2005). SAPS 3: From evaluation of the patient to evaluation of the intensive care unit. *Intensive Care Medicine*, 31(10), 1345-1355.
- Nassar, A. P. et al. (2012). Caution when using prognostic models: A prospective comparison of 3 recent prognostic models. *Journal of Critical Care*,

27(4), 423.e1-423.e7.

- Rajkumar, A. et al. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1(1), 18.
- Tomasev, N. et al. (2019). A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*, 572(7767), 116-119.
- Vincent, J. L. et al. (1996). The SOFA score to describe organ dysfunction/failure. *Intensive Care Medicine*, 22(7), 707-710.
- 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.
- Zimmerman, J. E. et al. (2006). Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. *Critical Care Medicine*, 34(5), 1297-1310.
- Churpek, M. M. et al. (2017). Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Critical Care Medicine*, 44(2), 368-374.
- 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.
- 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.
- Shickel, B. et al. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis. *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1589-1604.
- Singer, M. et al. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8), 801-810.
- Yoon, J. et al. (2019). ForecastICU: A prognostic decision support system for timely prediction of intensive care unit admission. *Proceedings of Machine Learning Research*, 97, 7199-7209.

Declarations

Funding

Supported by the German Federal Ministry of Education and Research (BMBF grant 01IS21069) and institutional research funds from the European Institute of AI. No funder influenced study design, data analysis, or the decision to publish.

Conflict of Interest

The author declares no competing interests and has not received funding from any ICU monitoring company, EHR vendor, or AI healthcare company.

Data Availability Statement

MIMIC-IV data is publicly available from PhysioNet. De-identified European ICU data summaries and model performance metrics are available from the author on request. Patient-level data cannot be shared due to GDPR. All model architectures, training code, and the IPAS calculator are available on request.

Ethical Approval

Approved by the Charite Ethics Committee (EA4/194/20), Cantonal Ethics Committee Zurich (BASEC 2020-01486), CPP Ile-de-France VI (2020-A01648-31), and Karolinska Regional Ethics Committee (Dnr 2020-02648). Patient consent was waived for retrospective analysis of de-identified ICU data.

Appendix A

IPAS Calculation and Model Architecture Details

This appendix provides the IPAS scoring rubric, transformer and XGBoost architecture specifications, and MIMIC-IV external validation methodology.

Part I -- IPAS Dimensions

- 1a. Discrimination: Average AUROC across 4 tasks
..... Higher = better separation of outcomes
- 1b. Calibration: 1 - average Brier score across tasks
..... Hosmer-Lemeshow assessed per task
- 2a. Lead Time: Median hours of advance warning for T2 and T4 Normalised: hours/24 (capped at 24h)
- 2b. Alert Fatigue: 1 - (false positive rate at 80% sensitivity) Lower FPR = less fatigue = higher score
- 3a. Workflow: Intensivist panel rating (8 clinicians)
..... Data availability, compute time, interpretability
- 3b. $IPAS = 0.25 * Discrim + 0.15 * Calibration + 0.20 * LeadTime + 0.20 * (1 - FPR) + 0.20 * Workflow$

Part II -- Model Specifications

- 4a. A3 XGBoost: 186 features, 1000 trees, depth 8, lr 0.03, updated every 4 hours SHAP for interpretability
- 4b. A4 RNN: 2-layer bi-LSTM, 256 hidden, attention, 1-min vital signs input 30-sec inference per patient
- 5a. A5 Transformer: 12 layers, 8 heads, d_model=512 Structured (128d) + temporal (256d) + text (128d) + image (128d)
- 5b. Training: 2016-2019, validation 2020, test 2021
..... External: MIMIC-IV 2008-2019 (73,181 stays)