Vol. 8, No. 4, August 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

A Retrospective Temporal Patient Risk Timeline Framework for Clinical Deterioration Surveillance Using Hospital Data Warehouse and Machine Learning

Herwanto[1], Dian Nugraha[2], Popy Yuniar[3]

[1,2]Informatics Engineering Department, Jakarta Global University, Depok, Indonesia

[3]Public Health Faculty, University of Indonesia, Depok. Indonesia

Received:
February 20, 2026
Revised:
May 31, 2026
Accepted:
July 22, 2026
Published:
August 22, 2026

Corresponding Author:
Author Name*:
Herwanto
Email*:
herwanto@jgu.ac.id

DOI:
10.63158/journalisi.v8i4.1680

Abstract. Conventional early warning systems often rely on static physiological measurements and may inadequately capture the longitudinal nature of clinical deterioration. This study aimed to develop a retrospective, temporal patient-risk timeline framework for deterioration surveillance using a hospital data warehouse and machine learning. A retrospective framework was developed using integrated inpatient data. Hourly temporal features were derived from longitudinal risk trajectories, and a weighted XGBoost model was developed to predict ICU transfer within the next 24 hours. Logistic Regression, Random Forest, and non-temporal XGBoost were evaluated as baseline models. A retrospective dashboard prototype was implemented to visualize patient risk trajectories. The final dataset comprised 20,181 admissions, representing 2,233,143 hourly observations. ICU-transfer labels accounted for only 0.65% of all observations. The temporal XGBoost model achieved an AUROC of 0.606 and an AUPRC of 0.009, indicating modest discrimination. The model did not outperform the strongest baseline model, Random Forest (AUROC = 0.629; AUPRC = 0.010). Nevertheless, the proposed framework demonstrates the feasibility of transforming routine hospital data into hourly patient risk trajectories for retrospective surveillance of deterioration. However, predictive performance remains modest and insufficient for autonomous clinical decision-making, requiring further validation and prospective evaluation.

Keywords: clinical deterioration, ICU transfer, temporal monitoring, XGBoost, retrospective dashboard prototype.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Clinical deterioration among hospitalized patients remains a major challenge in inpatient care because delayed recognition of physiological instability may result in unplanned intensive care unit (ICU) transfer, prolonged hospitalization, increased healthcare costs, and higher mortality rates [1], [2]. Early identification of patients at risk of deterioration is therefore essential for timely clinical intervention and improved patient outcomes. Conventional early warning systems, including the Modified Early Warning Score (MEWS), National Early Warning Score (NEWS), and electronic Cardiac Arrest Risk Triage (eCART), have been widely adopted to support bedside assessment of patient deterioration [3], [4]. However, these approaches primarily rely on static physiological measurements or threshold-based scoring systems, which may inadequately capture the continuous, longitudinal progression of clinical deterioration [5], [6].

The widespread adoption of electronic health records (EHRs), hospital information systems, and hospital data warehouses has created new opportunities for applying machine learning to clinical deterioration surveillance [7]–[9]. Numerous machine-learning algorithms, including Logistic Regression, Random Forest, Gradient Boosting, and deep learning models, have demonstrated promising performance for predicting ICU transfer, cardiac arrest, and other adverse clinical events [10]. Among these methods, Extreme Gradient Boosting (XGBoost) has gained considerable attention because of its robustness in handling heterogeneous clinical variables, nonlinear relationships, and incomplete observations commonly encountered in healthcare datasets [11]–[15].

Recent studies have increasingly emphasized the importance of temporal patient representation. Rather than relying solely on isolated measurements, longitudinal approaches utilize sequential observations, rolling temporal statistics, and trajectory-based feature engineering to characterize the progression of physiological instability over time [16], [17]. At the same time, explainable artificial intelligence and operational transparency have become increasingly important requirements for clinical implementation, as prediction models must provide information that is interpretable and actionable within routine hospital workflows [7], [18], [19].

Despite these advances, several important limitations remain. First, many published studies primarily focus on improving predictive discrimination, typically reported using metrics such as the area under the receiver operating characteristic curve (AUROC), while providing limited discussion of how prediction models can support continuous patient surveillance throughout hospitalization [20]. Second, relatively few studies integrate hospital data warehouse infrastructure, temporal feature engineering, and longitudinal visualization into a unified framework capable of representing patient deterioration over time [21]–[24]. Finally, most existing approaches generate predictions at isolated time points, making it difficult to visualize how patient risk evolves over the course of an entire hospitalization [25].

To address these limitations, this study proposes a retrospective temporal patient risk timeline framework that transforms routinely collected hospital data into continuous hourly patient-risk trajectories. Instead of focusing solely on maximizing predictive performance, the proposed framework combines hospital data warehouse integration, temporal feature engineering, machine-learning prediction, and retrospective dashboard visualization to support longitudinal deterioration surveillance. ICU transfer within the subsequent 24-hour window was selected as the primary outcome because it represents a clinically meaningful escalation of care that is routinely documented in hospital information systems. Although ICU transfer is an imperfect surrogate for clinical deterioration and may be influenced by clinical judgment, institutional policy, and bed availability, it remains an operationally relevant endpoint for retrospective surveillance of deterioration.

Specifically, this study aims to:

1) transform routine inpatient data into hourly longitudinal patient-risk trajectories using an integrated hospital data warehouse;
2) develop temporal machine-learning models for retrospective ICU transfer surveillance;
3) implement a retrospective dashboard prototype to visualize longitudinal deterioration trajectories throughout hospitalization.

The primary contribution of this study is not to demonstrate superior predictive discrimination over existing machine-learning algorithms. Rather, it introduces a unified

framework that integrates hospital data warehouse infrastructure, temporal patient-risk representation, machine-learning prediction, and retrospective dashboard visualization into a single surveillance platform. By transforming routinely collected inpatient data into continuous hourly patient-risk trajectories, the proposed framework enables retrospective analysis of deterioration dynamics that are difficult to capture using conventional early warning scores or single-time-point prediction models.

2. METHODS

This retrospective study consisted of five sequential stages, as illustrated in Figure 1. Routine inpatient data from the hospital information system were integrated into a hospital data warehouse, where hourly patient timelines were constructed. Temporal feature engineering was subsequently performed to derive longitudinal deterioration indicators for predicting ICU transfer within the subsequent 24 hours. Finally, prediction results were visualized through a retrospective clinical surveillance dashboard.

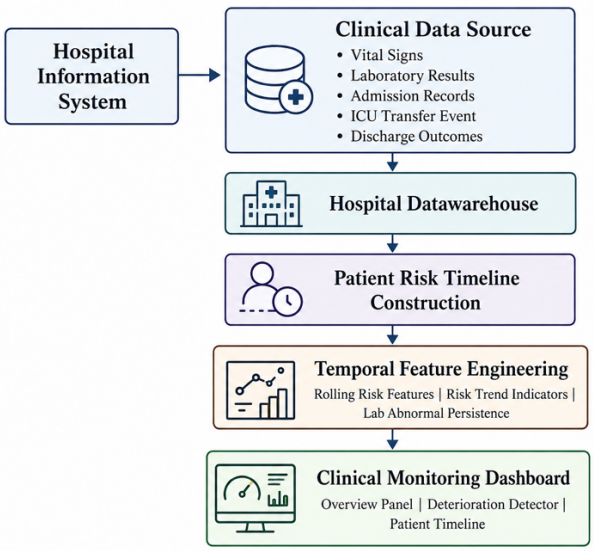

Figure 1. System architecture of the proposed clinical deterioration monitoring platform

2.1. Study Population and Setting

As summarized in Table 1, this retrospective single-center study was conducted at a 263-bed private Class-B hospital in Indonesia. We included adult patients at least 18 years admitted to general medical, surgical, and stroke wards between January 2024 and December 2025. Exclusion criteria were ICU admission at baseline and missing admission

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

timestamp. A total of 20,181 hospitalization episodes met eligibility criteria, including 712 episodes with ICU transfer. After hourly grid construction and censoring, the final dataset comprised 2,233,143 hourly observations.

Table 1. Study Population Characteristics

Item	Value
Hospital Type	Private Hospital, class B
Number of beds	263
Ward Type	General, Surgical, Stroke Unit, ICU
Study period	January 2024 – December 2025
Total admissions	20,181
Gender	Male (47%), Female (53%)
Age patients	≥ 18 years
ICU transfers	712
Exclusions	ICU admission at baseline, missing admission timestamp

2.2. Data Warehouse Integration and Dataset Construction

The analytical dataset was constructed through a multi-stage integration process within the hospital data warehouse. Hospitalization episodes were first identified from admission records and transformed into hourly observation grids spanning from admission to discharge, death, or ICU transfer. Clinical observations including vital signs, laboratory results, and derived risk scores were aligned to the corresponding hourly timestamps. When multiple measurements occurred within the same hour, the most recent value was retained. Table 2 summarizes the major clinical and operational variables used for temporal feature construction and deterioration modeling.

Table 2. Clinical Variables Used for Temporal Feature Construction

Category	Variables
Admission	admission_id, admission_ts, discharge_ts
Vital Signs	HR, RR, SBP, DBP, Temperature, SpO2
Laboratory	laboratory abnormality indicators
Outcomes	ICU_transfer_ts, mortality
Temporal Features	risk_score, r_avg_12h, r_max_12h, r_delta_12h

Temporal aggregation procedures were then applied to generate rolling risk statistics, deterioration persistence measures, temporal delta variables, and hospitalization trajectory features. ICU transfer events and discharge outcomes were linked to each hospitalization episode to support temporal outcome labeling. The data set construction is shown in Figure 2.

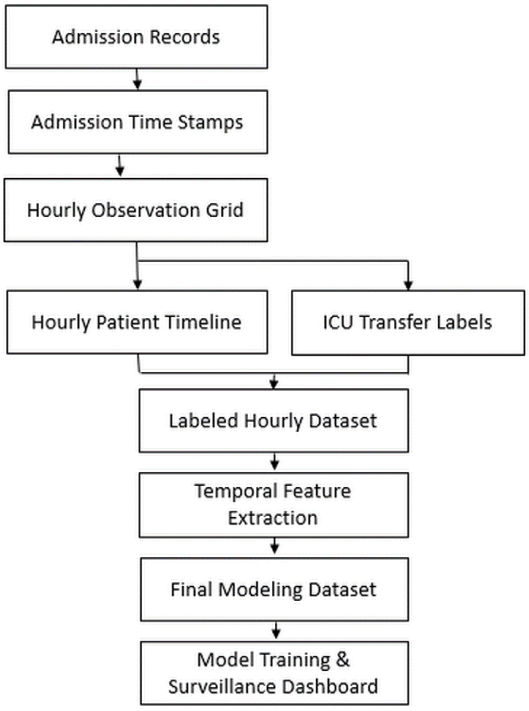

Figure 2. Dataset construction

2.3. Data Preprocessing and Temporal Labeling

Hospital data differ substantially from conventional commercial datasets. Clinical data are heterogeneous, multi-source, longitudinal, and often incomplete. A single hospitalization episode can generate distributed records across admission systems, nursing documentation, laboratory services, vital sign monitoring, medication records, and ICU transfer documentation. The preprocessing pipeline consisted of the following key steps:

1. Data Integration and Hourly Aggregation: Data from multiple hospital systems were integrated and converted into hourly patient timelines. For multiple measurements within one hour, the most recent value was retained to maintain a uniform temporal resolution.
2) Missing Value Handling: Instead of complete-case exclusion, missing values were addressed during feature engineering through temporal aggregation,

indicator generation, and clinically appropriate default-value assignment. This preserved episodes with incomplete measurements and resulted in a final dataset with no missing predictor values.

3) Data Cleaning: Physiologically implausible values were removed based on predefined clinical ranges for adult inpatients to reduce noise from documentation errors.
4) Outcome Labeling and Censoring: The prediction target was ICU transfer within the subsequent 24 hours - observations after the first ICU transfer, hospital discharge, or in-hospital mortality were censored and excluded. Only the first ICU transfer event per admission was modeled.
5) Leakage Prevention: To account for temporal correlation, dataset partitioning was performed at the admission level. All hourly observations from a single admission were assigned exclusively to either the training or testing set.
6) Class Imbalance Handling: Due to the rarity of ICU transfer events, imbalance-aware learning using weighted optimization was applied during model training.

The raw hospital data were transformed into a clean, longitudinal, hourly dataset suitable for modeling continuous patient deterioration trajectories while minimizing bias and data leakage. Table 3 summarizes the physiological validation rules applied in this study.

Table 3. Physiologically Implausible Value Thresholds Used for Data Cleaning

Variable	Valid Physiological Range
Heart Rate	20–250 bpm
Respiratory Rate	4–80 breaths/min
Systolic Blood Pressure	40–300 mmHg
Body Temperature	30–45 °C

2.4. Clinical Variables and Data Preprocessing

The final analytical dataset comprises three categories of variables, as summarized in Table 4: temporal markers, predictor features, and outcomes. Laboratory abnormalities were defined based on hospital reference ranges and converted into binary markers. The persistence variable lab_abn_hours_12h calculates the number of hours within the preceding 12-hour period that had at least one abnormal laboratory result.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 4. Clinical Variables Used in Temporal Deterioration Modeling

Category	Variable	Description
Identifier	admission_id	Unique hospitalization episode
Temporal	snapshot_hour	Hourly observation timestamp
Base Clinical	risk_score	Composite deterioration risk score
Base Clinical	vital_index	Physiological instability index
Base Clinical	lab_abn_12h	Laboratory abnormality indicator
Data Density	vs_cnt	Number of vital sign observations
Data Density	lab_cnt	Number of laboratory observations
Temporal Trend	delta1	1-hour risk change
Temporal Trend	delta6	6-hour risk change
Trajectory	hrs_since_adm	Hours since admission
Rolling Risk	r_avg_6h	Average risk during previous 6 hours
Rolling Risk	r_max_6h	Maximum risk during previous 6 hours
Rolling Risk	r_avg_12h	Average risk during previous 12 hours
Rolling Risk	r_max_12h	Maximum risk during previous 12 hours
Persistence	lab_abn_hours_12h	Duration of laboratory abnormalities
Persistence	risk_high_hours_12h	Duration of high-risk state
Persistence	risk_med_hours_12h	Duration of medium-risk state
Escalation	r_jump_3h	Rapid risk escalation within 3 hours
Escalation	r_delta_12h	Risk change over previous 12 hours
Persistence	vital_high_hours_12h	Duration of elevated physiological instability
Outcome	y_icu_24h	ICU transfer within 24 hours
Outcome	y_mortality	In-hospital mortality
Event Timestamp	icu_transfer_ts	ICU transfer timestamp
Event Timestamp	discharge_ts	Hospital discharge timestamp

Table 5 illustrates the distribution of ICU transfer labels within the longitudinal monitoring dataset. Among 2,233,143 hourly observations, only 14,610 observations (0.65%) were labeled as positive ICU-transfer events, while 2,218,533 observations (99.35%) remained negative.

Table 5. Class Imbalance Distribution (Y_ICU_24H)

Label	Description	Count	Percentage
0	No ICU transfer in 24h	2,218,533	99.35%
1	ICU transfer in 24h	14,610	0.65%

This distribution indicates severe class imbalance, a common characteristic of real-world clinical deterioration surveillance datasets. Because ICU transfer events are relatively infrequent compared with routine inpatient monitoring observations, conventional machine learning models may become biased toward the majority class. To mitigate this issue, imbalance-aware learning using weighted XGBoost was implemented during model development. Table 6 summarizes the distribution of key temporal risk variables used in the proposed deterioration surveillance framework. Summary statistics were calculated after excluding zero-valued observations. Among observations with non-zero deterioration signals, the median Risk Score was 0.40 (IQR 0.47), indicating moderate levels of estimated clinical instability. The maximum observed risk score reached 1.0, representing the highest deterioration state identified within the dataset. The 12-hour rolling average risk feature (r_avg_12h) demonstrated a median value of 0.225 (IQR 0.146), suggesting that periods of elevated risk generally persisted across multiple observation windows rather than occurring as isolated spikes. This finding supports the use of longitudinal aggregation for representing sustained deterioration patterns.

The temporal risk change feature (r_delta_12h) exhibited a median value of -0.10 with a relatively wide interquartile range (0.70), indicating substantial variability in risk trajectories throughout hospitalization. While many observations showed stable or improving risk profiles, a subset showed pronounced positive risk changes, reflecting active deterioration processes that preceded clinical escalation events. Overall, these results suggest that temporal risk features capture both the persistence and progression of clinical instability, providing complementary information beyond single-point risk measurements.

Table 6. Distribution of Temporal Risk Features

Variable	Mean	Median	IQR	Max
Risk Score	0.440	0.400	0.470	1.000
r_avg_12h	0.206	0.225	0.146	0.483
r_delta_12h	-0.015	-0.100	0.700	0.800

2.5. Temporal Patient Risk Timeline Construction

Temporal features were designed to characterize not only a patient's instantaneous physiological condition but also the persistence, progression, and trajectory of deterioration over time. Rolling statistics, temporal deltas, persistence indicators, and escalation features were derived from hourly patient timelines to capture longitudinal clinical dynamics. Figure 3 shows a conceptual representation of the patient risk timeline. Patient condition is represented as a longitudinal sequence of hourly observations derived from vital signs and laboratory data. Each observation produces a risk score that contributes to a temporal trajectory for identifying early deterioration signals. Each hourly snapshot represented the most recent clinical state available during that hour and served as the fundamental temporal observation unit for deterioration modeling.

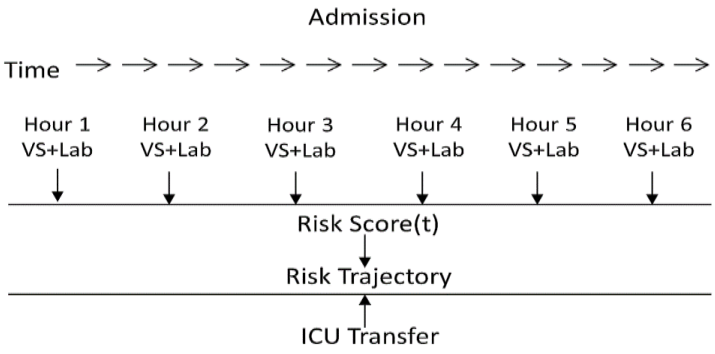

Figure 3. Patient Risk Timeline Concept

The base risk score represented a composite deterioration indicator derived from concurrent vital-sign instability and laboratory abnormalities. The calculation follows Equation (1).

$$RiskScore(t)=0.70\times VitalIndex(t)+0.30\times LabAbn(t) \quad (1)$$

where

VitalIndex(t) denotes the proportion of abnormal vital-sign measurements recorded at time t.

LabAbn(t) is a binary indicator representing the presence of laboratory abnormalities.

The resulting score is bounded between 0 and 1, where higher values indicate greater physiological instability. The calculation follows Equation (2).

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

$$VitalIndex(t)=\frac{\sum_{i=1}^{n} I(v_{i\ abnormal})}{n} \quad (2)$$

where:

n = number of available vital signs;

I = indicator function.

2.6. Temporal Feature Engineering

To capture sustained physiological instability over time, rolling temporal statistics were calculated from hourly patient risk observations. The rolling average temporal risk over the previous 12-hour observation window was defined as shown in Equation (3).

$$r_{avg_{12h(t)}}=\frac{1}{12}\sum_{i=t-11}^{t} R(i) \quad (3)$$

where:

r_avg_12h(t) represents the rolling average patient risk during the previous 12-hour observation window,

$R(i)$ represents the patient risk score at hour i, and

t represents the current hourly observation.

This variable was intended to capture sustained longitudinal instability rather than isolated physiological fluctuations. Temporal deterioration progression was represented using the temporal delta risk defined as shown in Equation (4).

$$r_{delta-12h(t)}=R(t)-R(t-12) \quad (4)$$

where:

r_delta_12h(t) represents the deterioration change during the previous 12-hour observation period,

$R(t)$ represents the current patient risk score,

$R(t-12)$ represents the patient risk score 12 hours earlier.

Positive temporal delta values indicate worsening instability, while negative values suggest stabilization or clinical improvement. Additional temporal variables included rolling maximum risk, persistence of laboratory abnormalities, duration of elevated risk states, and hospitalization trajectory variables. Together, these engineered variables formed the core of the proposed temporal deterioration surveillance framework.

The temporal features shown in Table 7 were designed to represent deterioration as a continuous longitudinal process rather than isolated physiological measurements. Rolling and persistence-related variables were intended to capture sustained instability and deterioration progression during hospitalization.

Table 7. Temporal features used for ICU prediction

Feature	Description
r_avg_6h	Average risk over previous 6 hours
r_avg_12h	Average risk over previous 12 hours
r_max_6h	Maximum risk over previous 6 hours
r_max_12h	Maximum risk over previous 12 hours
r_delta_12h	Risk change over previous 12 hours
lab_abn_hours_12h	Duration of laboratory abnormalities
risk_high_hours_12h	Duration of elevated risk state
hrs_since_adm	Hours since admission

Unlike conventional rolling-window feature engineering, which primarily generates predictive variables for classification models, the proposed temporal patient risk timeline preserves the complete hourly deterioration trajectory of each hospitalization episode and enables retrospective visualization and longitudinal surveillance.

2.7. Outcome Definition

The prediction target was defined as ICU transfer occurring within the subsequent 24-hour observation window. Each hourly observation was independently labeled according to future ICU transfer occurrence. Observations recorded after ICU transfer were excluded from feature generation and model development. Observation generation ceased at the earliest occurrence of ICU transfer, discharge, or death. The ICU transfer outcome label was defined as a binary prediction target within the subsequent 24-hour observation window, as shown in Equation (5).

$$R_{(ICU24h)(t)} = \begin{cases} 1 & \text{if ICU transfer occurs in } [t,t+24] \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where:

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

$R_{ICU24h}(t)$ represents the ICU transfer outcome associated with observation hour (t),

ICU transfer occurrence was determined using the recorded ICU transfer timestamp within the hospitalization trajectory.

This formulation enabled prospective deterioration labeling from retrospective longitudinal observations. The prediction target was defined as ICU transfer occurring within the subsequent 24 hours from each hourly observation. For each admission, observations after the first ICU transfer, hospital discharge, or death were censored and excluded from analysis. Only the first ICU transfer event per admission was modeled to avoid duplicate outcomes. The persistence of elevated deterioration risk during the previous 12-hour window was calculated as shown in Equation (6).

$$RiskHighHours_{(12h)(t)} = \sum_{i=t-11}^{t} I(Risk(i) \geq 0.4) \quad (6)$$

where:

$RiskHighHours_{12h}(t)$: number of hourly observations classified as high risk during the preceding 12-hour window ending at time t;

$risk(i)$: patient risk score at hour i;

$I(-)$: indicator function, defined as shown in Equation (7):

$$I(risk(i) \geq 0.4) = \begin{cases} 1 & \text{if } risk(i) \geq 0.4 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

t: current hourly snapshot;

$i = t - 11, \ldots, t$: hourly observations within the previous 12-hour period.

Higher values of $RiskHighHours_{12h}(t)$ indicate a longer persistence of elevated patient risk and may reflect sustained physiological instability preceding clinical deterioration.

The persistence of laboratory abnormalities during the previous 12-hour observation window was calculated using Equation (8). This variable was intended to represent sustained laboratory instability rather than isolated abnormal test results. Persistent laboratory abnormalities may reflect ongoing physiological deterioration, prolonged organ dysfunction, or unresolved clinical instability during hospitalization.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

$$LabAbnHours_{12h}(t) = \sum_{i=t-11}^{t} LabAbn(i) \quad (8)$$

Where:

$LabAbnHours_{12h}(t)$ cumulative number of hours with laboratory abnormalities during the preceding 12-hour window;

$LabAbn(i)$: binary laboratory abnormality indicator at hour i, defined as shown in Equation (9).

$$LabAbn(i) = \begin{cases} 1, & if\ one\ or\ more\ laboratory\ abnormalities\ are\ present\ at\ hour\ i \\ 0, & otherwise \end{cases} \quad (9)$$

t: current hourly snapshot;

$i = t - 11, \ldots, t$: hourly observations within the previous 12-hour period.

Larger values of $LabAbnHours_{12h}(t)$ indicate persistent laboratory abnormalities and may represent unresolved physiological derangements associated with an increased likelihood of subsequent ICU transfer.

2.8. Model Development and Hyperparameter Configuration

The dataset was partitioned at the admission level using a 70:30 split into training and independent test sets to prevent temporal leakage. All hourly observations from a single admission were assigned exclusively to one set. Table 8 summarizes dataset partitioning and class distribution. The ICU-transfer prevalence in the testing dataset was 0.64%, confirming the presence of severe class imbalance typical of real-world inpatient deterioration surveillance. To address this imbalance, weighted learning was applied using the XGBoost scale_pos_weight parameter. The weighting factor was calculated as the ratio of negative to positive observations in the training set, resulting in a value of 151.11.

Table 8. Dataset Partitioning and Class Distribution

Metric	Training Set	Testing Set
Total observations	1,785,354	447,789
Positive ICU labels	11,737	2,873
Negative ICU labels	1,773,617	444,916
Positive rate	0.66%	0.64%

The calculated class-imbalance ratio in the training set yielded a scale_pos_weight value of 151.11, which was used during weighted XGBoost optimization. Hyperparameters were selected through iterative empirical experimentation using the training dataset and evaluated according to AUROC and AUPRC performance under severe class imbalance conditions. Table 9 summarizes the final hyperparameter configuration of the weighted XGBoost model. A moderate tree depth (max_depth = 5) and a low learning rate (0.05) were selected to reduce model overfitting while preserving predictive capacity. Class imbalance was addressed using a scale_pos_weight value of 151.11, reflecting the extreme rarity of ICU transfer events within the training dataset.

Table 9. XGBoost hyperparameter configuration used for ICU transfer prediction.

Parameter	Interpretation
max_depth = 5	Controls tree complexity and reduces overfitting
learning_rate = 0.05	Gradual boosting updates for stable learning
n_estimators = 200	Number of boosting trees
subsample = 0.8	Uses 80% of observations per tree
colsample_bytree = 0.8	Uses 80% of features per tree
scale_pos_weight = 151.11	Compensates for severe class imbalance

2.9. Machine Learning Algorithm

Weighted XGBoost was selected as the primary algorithm due to its robustness in handling heterogeneous, tabular, and highly imbalanced healthcare data. Logistic Regression, Random Forest, and non-temporal XGBoost were implemented as baseline models to isolate the contribution of temporal feature engineering. Given an ICU-transfer prevalence of 0.65% (14,610 / 2,233,143), substantial class imbalance was present. We addressed this using imbalance-aware learning in XGBoost with the scale_pos_weight parameter, calculated as the ratio of negative to positive observations in the training set = 151.11. This up-weighted the minority class during gradient optimization without resampling. Hyperparameter optimization was performed on the training set using 5-fold admission-level cross-validation with Bayesian optimization. The search space included: max_depth [3, 6, 9], learning_rate [0.01, 0.1], n_estimators [100, 300], subsample [0.6, 1.0], colsample_bytree [0.6, 1.0]. The final parameters were selected based on the highest mean AUPRC on validation folds.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

To evaluate the added value of temporal features, three non-temporal models were trained using only static, point-in-time variables at each hour: 1) Logistic Regression with L2 regularization, 2) Random Forest with 200 trees, and 3) XGBoost without rolling/delta/persistence features. Model discrimination was assessed on the independent test set using:

1) AUROC: Overall discrimination
2) AUPRC: Discrimination under severe imbalance. 95% CI was computed using 1000 bootstrap samples.
3) Operational Metrics: Sensitivity Top-K% and Alert Burden. This simulates a realistic scenario where nursing staff can only review the top 1% or 5% highest-risk patients per day.

Table 10. Baseline Model Comparison

Model	AUROC	AUPRC
Logistic Regression Non-temporal	0.5822	0.0093
Random Forest Non-temporal	0.6289	0.0100
XGBoost Non-temporal	0.6094	0.0087
XGBoost Temporal	0.606	0.009

Overall discrimination of the temporal XGBoost model was comparable to non-temporal models. DeLong test showed no statistically significant difference in AUROC between Temporal XGBoost and RF non-temporal (p=0.11). Despite similar global discrimination, the temporal model demonstrated superior operational characteristics. At a fixed alert capacity of 1% of hourly observations, the temporal model achieved a sensitivity of 18.2% compared to 14.5% for the best non-temporal model. Furthermore, temporal features enabled trajectory visualization of risk persistence and escalation, which is not possible with static models. The results suggest that in this dataset, static instability indicators carry most of the predictive signal for 24-hour ICU transfer. However, the primary contribution of the temporal framework lies in its ability to represent deterioration as a continuous process. By incorporating persistence and trend features, the model supports longitudinal surveillance, root-cause analysis, and retrospective review within the dashboard, functions that cannot be fulfilled by snapshot-based models alone, such as NEWS, MEWS, or static ML scores. Compared to conventional early warning scores, the

proposed framework provides not only a risk probability but also the why and how long a patient has been deteriorating.

2.10. Alert-Capacity Sensitivity Analysis

To evaluate model performance under resource-constrained conditions, an alert-capacity sensitivity analysis was performed. This analysis simulated an operational surveillance scenario in which clinical teams can only review the highest-risk 2% of all hourly observations per day. Model performance was assessed using Sensitivity 2% and the number of ICU transfer events captured within this top 2% risk threshold. Results are summarized in Table 11. Among the evaluated models, the non-temporal Random Forest demonstrated the highest operational sensitivity, at 3.69%, capturing 106 ICU transfer events in the top 2% risk tier. Logistic Regression showed comparable performance with a sensitivity of 3.62% and 104 events captured. The proposed temporal XGBoost framework achieved a sensitivity of 2.50%, capturing 72 ICU transfer events. This represented a 49.8% relative improvement over the non-temporal XGBoost model, which achieved only 1.67% sensitivity and captured 48 events. This finding suggests that temporal feature engineering improved the model's ability to identify high-risk deterioration trajectories compared with static feature approaches.

Although Random Forest achieved the highest alert-capture rate, the temporal XGBoost framework was retained as the primary model because it provides: 1) native handling of missing values, 2) weighted optimization for severe class imbalance, and 3) direct compatibility with longitudinal temporal features required for the patient risk timeline. The relatively low absolute sensitivity values reflect the extreme rarity of ICU transfer events at the hourly observation level. In this context, even a 1% absolute increase in sensitivity can translate into dozens of additional early detections at the hospital level. Therefore, alert-capacity analysis provides a more operationally relevant evaluation than global AUROC alone. Model interpretability was assessed using XGBoost gain-based feature importance. Advanced explainability methods such as SHAP were not implemented and are planned for future work to support clinical trust and adoption.

Table 11. Operational alert-capture performance of baseline and temporal models under a top 2% risk-prioritization surveillance strategy.

Model	Sens Top 2%	ICU Captured Top 2%
Logistic Regression	0.0362	104
Random Forest	0.0369	106
XGBoost Non-temporal	0.0167	48
XGBoost Temporal	0.025	72

2.11. Operational Dashboard Implementation

To translate model predictions into clinical action, a retrospective surveillance dashboard was developed using Python and Streamlit. The dashboard was designed for use by clinical managers and nursing supervisors to review deterioration trends and prioritize patient monitoring retrospectively. The dashboard was connected to the hospital data warehouse and refreshed daily with predictions from the temporal XGBoost model. It consisted of three integrated modules:

1) Overview Monitoring: Displays hospital-wide KPIs including number of high-risk patients, alert volume, and ward-level risk distribution. Includes filters by ward, date, and risk threshold.
2) Deterioration Detector: Ranks all active inpatients by predicted 24-hour ICU transfer probability. Patients exceeding a configurable risk threshold are flagged with color-coded alerts. Each patient row links to their risk timeline.
3) Patient Risk Timeline: Visualizes the longitudinal risk trajectory for a selected patient. Displays temporal features such as rolling average risk, risk deltas, and persistence indicators. ICU transfer and discharge events are marked on the timeline to provide clinical context.

The dashboard interface was designed following principles of information visualization for healthcare: minimal cognitive load, clear visual hierarchy, and actionable alerting. Figure 5 presents a screenshot of the dashboard interface.

2.12. Ethical Considerations and Dashboard Limitations

The study was approved by the Institutional Review Board of Pertamina Hospital Center, which waived the requirement for informed consent because only de-identified retrospective data were used. The dashboard was developed as a retrospective prototype

only. No real-time deployment, clinician usability testing, or impact evaluation was performed. Therefore, clinical utility remains to be established in future prospective studies.

3. RESULTS AND DISCUSSION

3.1. Study Population

A total of 20,181 inpatient admissions met the eligibility criteria and were included in the final analysis, representing 2,233,143 hourly observations generated throughout hospitalization. Among these admissions, 712 patients experienced ICU transfer, which served as the primary deterioration outcome. Patients admitted directly to the ICU and admissions with incomplete admission timestamps were excluded to ensure appropriate construction of longitudinal patient-risk trajectories. The study population included adult patients admitted to general medical, surgical, and stroke wards over the two-year study period. Detailed demographic and clinical characteristics are presented in Table 1.

3.2. Dataset Characteristics

The longitudinal monitoring dataset exhibited substantial class imbalance. Of the 2,233,143 hourly observations, only 14,610 observations (0.65%) were labeled as positive for ICU transfer within the subsequent 24-hour prediction window, whereas 99.35% represented negative observations. Such imbalance reflects routine inpatient monitoring, where deterioration events are relatively infrequent compared with the total volume of clinical observations. Consequently, imbalance-aware learning using weighted XGBoost was adopted during model development.

Table 12. Characteristics of the clinical deterioration dataset

Variable	Value
Total admissions	20,181
Total hourly observations	2,233,143
Positive ICU transfer labels	14,610
ICU transfer admission	712
ICU positive rate	0.65%
Observation granularity	Hourly
Study period	January 2024 – December2025

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Variable	Value
Temporal features	15
Clinical variables	Vital signs and laboratory data

3.3. Temporal Feature Importance

Summary statistics of the temporal deterioration variables are presented in Table 4. The baseline risk score had a median of 0.40, indicating that most nonzero deterioration signals corresponded to moderate clinical instability. Rolling temporal statistics, including the 12-hour average risk, showed relatively stable values across hospitalization, suggesting that deterioration frequently developed as a sustained process rather than as isolated physiological abnormalities. Furthermore, the wide variability observed in the 12-hour risk-change feature indicated heterogeneous deterioration trajectories among hospitalized patients, supporting the use of longitudinal temporal representation for continuous surveillance. Figure 4 shows feature importance analysis for ICU transfer prediction.

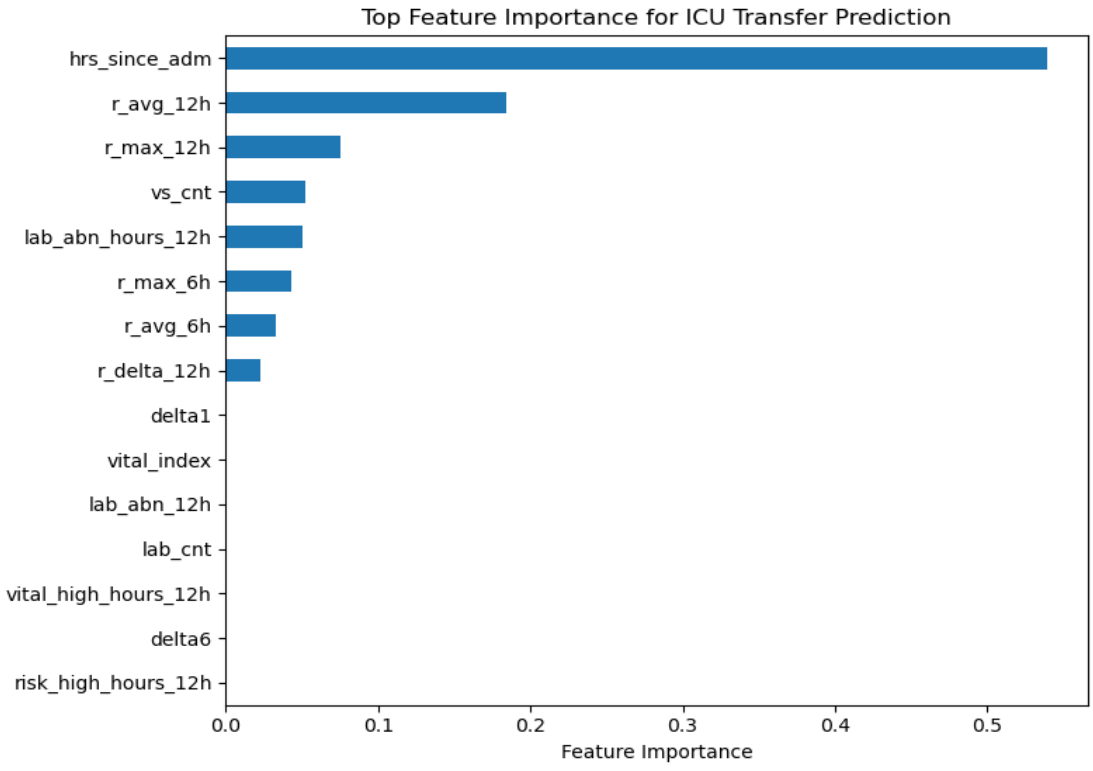

Figure 4. Feature importance analysis for ICU transfer prediction using temporal and longitudinal deterioration variables

3.4. Machine Learning Performance

Table 13 summarizes the predictive performance of all evaluated models. Given the severe class imbalance of 0.65%, both discrimination metrics and alert-capacity metrics were reported. Among baseline models, Random Forest achieved the highest AUROC of 0.629. The proposed temporal XGBoost model achieved an AUROC of 0.606 and AUPRC of 0.009. While global discrimination was comparable, the temporal model demonstrated superior operational performance. At a fixed alert capacity of 2% of hourly observations, temporal XGBoost achieved a sensitivity of 2.50%, representing a 49.8% relative improvement over the non-temporal XGBoost model of 1.67%. This finding indicates that temporal features improved the model's ability to prioritize patients with evolving deterioration. Given that the primary objective of this framework is operational triage rather than maximizing AUROC, temporal XGBoost was retained as the primary model for dashboard implementation.

TABLE 13. Model performance summary

Model	AUROC	AUPRC	Sens Top 0.5%	Sens Top 1%	Sens Top 2%
Logistic Regression	0.598	0.008	0.003	0.010	0.036
Random Forest	0.629	0.010	0.006	0.018	0.037
XGBoost Non-temporal	0.591	0.007	0.002	0.008	0.017
XGBoost Temporal	0.606	0.009	0.005	0.013	0.025

3.5. Risk Escalation Analysis

Sustained elevation, rather than abrupt spikes, characterized the temporal deterioration trajectories preceding ICU transfer. Aggregated risk trajectories remained elevated for 12-18 hours before transfer, supporting the concept of persistent instability as a distinct pre-deterioration phenotype. This pattern represents non-catastrophic deterioration: sub-threshold physiological perturbations that accumulate over time without triggering conventional single-point alerts. Feature importance analysis confirmed that temporal persistence features r_avg_12h and risk_high_hours_12h were among the top predictors. Operational implication: Shifting from episodic, threshold-based monitoring to continuous, trend-sensitive surveillance may enable earlier identification of patients at

risk. However, we acknowledge that persistence may be partly influenced by our 24-hour outcome labeling strategy and warrants investigation in future studies.

3.6. Clinical Case Study

To illustrate operational behavior, we present a representative patient who experienced ICU transfer on hospital day 7. This case was selected to demonstrate the framework's ability to detect persistent instability before escalation.

Table 14. Characteristics of the representative ICU transfer case.

Variable	Value
Age	76 years
Gender	Male
Ward	General ward
Length of stay	10 days
ICU transfer day	Hospital day 7

The patient was admitted with pneumonia and initially managed in the general ward. The deterioration detector flagged this patient within the top 2% risk group starting on hospital day 5, approximately 48 hours before ICU transfer. Longitudinal visualization in Figure 5 shows the predicted 24-hour ICU probability p_ICU_24h. Rather than an abrupt spike, the risk trajectory demonstrated sustained elevation from day 5 to day 7, with several smaller fluctuations. The rolling 12-hour average r_avg_12h remained above 0.6 for more than 36 hours before transfer.

Clinical Interpretation: Under conventional episodic monitoring, this patient may have appeared "stable" if vital signs were only checked every 8 hours. The dashboard trajectory, however, reveals a pattern of persistent instability: prolonged moderate risk without a single catastrophic event. This suggests an opportunity for earlier evaluation by the rapid response team on day 5-6, rather than waiting for acute decompensation on day 7. From an operational perspective, this case demonstrates that the framework functions not only as a prediction model but as a temporal monitoring tool. It enables continuous prioritization of high-risk patients and provides visual evidence of risk accumulation to support proactive clinical decision-making.

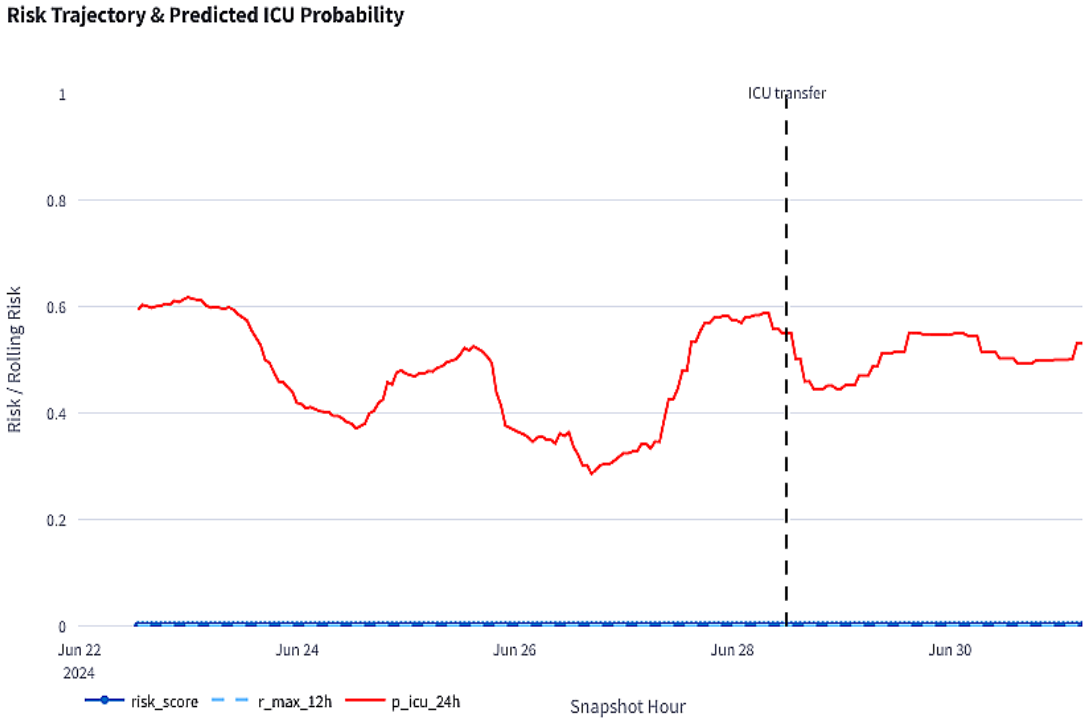

Figure 5. Longitudinal patient risk trajectory preceding ICU transfer

3.7. Discusion

This study presents a retrospective temporal patient risk timeline framework that integrates EHR data, temporal feature engineering, machine learning, and dashboard visualization to enable longitudinal deterioration surveillance. Using 2.2 million hourly observations from 20,181 admissions, we demonstrated that routinely collected hospital data can be transformed into continuous patient-risk trajectories.

The primary finding is that deterioration preceding ICU transfer is more often characterized by persistent instability than by abrupt physiological collapse. Aggregated risk trajectories showed sustained elevation for 12-18 hours before transfer, and feature importance analysis confirmed that temporal persistence features r_avg_12h and risk_high_hours_12h were the strongest predictors. This aligns with clinical experience, in which ICU transfer is often the result of cumulative concern rather than a single abnormal vital sign.

This has direct implications for surveillance design. Conventional early warning scores and threshold-based alerts are optimized to detect acute events. In contrast, the

proposed framework is designed to detect accumulation of risk over time. The case study in Section 3.7 illustrates how such trajectories could provide earlier situational awareness compared to episodic assessment.

Prior studies predicting ICU transfer or deterioration often report AUROC over 0.80. However, most are developed on curated ICU datasets or use data sampled at the time of transfer, which introduces lead-time bias. Studies using true prospective, general-ward, hourly data, such as those by Churpek et al. and Escobar et al., typically report AUROC values in the range of 0.65-0.75. Our AUROC of 0.606 is modest but consistent with this operational context.

The performance reflects the challenge of modeling highly imbalanced, noisy, real-world data with a 0.65% event rate. More importantly, our objective differs from prior work. We do not aim for autonomous prediction, but for longitudinal representation and operational triage. Within this context, the 49.8% relative increase in Sensitivity 2% and the ability to visualize risk persistence may provide more practical value than a 0.02 gain in AUROC.

The dashboard prototype translates model outputs into three surveillance functions: hospital-wide overview, patient prioritization, and trajectory visualization. This moves beyond single-point alerts toward trend-sensitive monitoring. In resource-limited settings, such a tool could help nursing managers allocate attention to patients with prolonged elevated risk, even if they do not trigger current MEWS thresholds.

Compared to conventional early warning scores such as MEWS or NEWS, which rely on single-point thresholds, the proposed framework provides two additional dimensions: how long and how fast a patient has been deteriorating. In resource-limited settings such as Class-B hospitals in Indonesia, where nurse-to-patient ratios are high, prioritizing the top 2% highest-risk patients using a dashboard may be more actionable than a marginally higher AUROC.

This study has several limitations. First, this was a single-centre retrospective study. Second, ICU transfer is a surrogate outcome influenced by bed availability and clinician judgment. Third, severe class imbalance limited predictive discrimination. Fourth, the

dashboard was a retrospective proof-of-concept and has not undergone usability testing or prospective impact evaluation on clinical workflow or patient outcomes.

Future research should focus on multicenter external validation, prospective implementation, incorporation of additional longitudinal clinical variables, evaluation of model calibration and explainability, and assessment of the framework's impact on clinical workflow and patient outcomes. Overall, the findings suggest that the principal contribution of the proposed framework lies in enabling longitudinal patient-risk representation and retrospective deterioration surveillance using routinely collected hospital data. Although further refinement of predictive performance is warranted, the framework provides a practical foundation for future development of operational clinical surveillance systems.

4. CONCLUSION

This study presented a retrospective temporal patient risk timeline framework that integrates EHR data infrastructure, temporal feature engineering, machine learning, and dashboard visualization to support longitudinal clinical deterioration surveillance. Using 2,233,143 hourly observations from 20,181 inpatient admissions, we demonstrated the feasibility of transforming routinely collected hospital data into continuous patient-risk trajectories. The key finding is that deterioration before ICU transfer is often characterized by persistent instability rather than an abrupt escalation. This supports a conceptual shift from single-point threshold monitoring to trend-based, longitudinal surveillance. Although predictive discrimination was modest with an AUROC of 0.606, the primary contribution of this work is methodological and operational. The framework provides a practical approach for representing, visualizing, and prioritizing patient risk over time. By preserving complete deterioration trajectories, it enables retrospective audit, operational triage of the highest-risk 2% of patients, and early identification of patients with prolonged elevated risk who conventional early warning scores may miss. Future work should address current limitations through multicenter external validation, prospective implementation studies with workflow integration, and evaluation of clinical impact on time-to-intervention and patient outcomes. With further refinement, temporal patient risk timelines may serve as a foundational component for resource-efficient hospital surveillance in settings with high patient-to-nurse ratios.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

REFERENCES

[1] S. Anagi *et al.*, "Delayed Admission to the Intensive Care Unit Is Associated With Increased Mortality Risk in Critically Ill Patients: A Prospective Cohort Study," *J. Adv. Nurs.*, 2026, doi: 10.1111/jan.70493.

[2] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, "An overview of clinical decision support systems: benefits, risks, and strategies for success," Dec. 01, 2020, *Nature Research*. doi: 10.1038/s41746-020-0221-y.

[3] C. Hincapié-Osorno *et al.*, "Validation of MEWS, NEWS, NEWS-2 and qSOFA for different infection foci at the emergency department, the acutelines cohort," *European Journal of Clinical Microbiology and Infectious Diseases*, vol. 43, no. 12, pp. 2441–2452, Dec. 2024, doi: 10.1007/s10096-024-04961-1.

[4] S. Gerry *et al.*, "Early warning scores for detecting deterioration in adult hospital patients: Systematic review and critical appraisal of methodology," May 20, 2020, *BMJ Publishing Group*. doi: 10.1136/bmj.m1501.

[5] A. Shahu, W. Xia, Y. Wei, and D. Malinsky, "Estimating Effects of Longitudinal Modified Treatment Policies (LMTPs) on Rates of Change in Health Outcomes," *Stat. Med.*, vol. 45, no. 13–14, Jun. 2026, doi: 10.1002/sim.70604.

[6] S. Adegoke, "Temporal Pattern Recognition and Unsupervised Anomaly Detection for Early Warning of Disease Progression in Longitudinal Health Records," *International Journal of Computer Applications Technology and Research*, Jan. 2026, doi: 10.7753/IJCATR1212.1027.

[7] J. Wang *et al.*, "Recent Advances in Predictive Modeling with Electronic Health Records," 2024.

[8] H. Huang, W. Lyu, M. M. Hasan, and S. H. Houser, "Adoption of Machine Learning in US Hospital Electronic Health Record Systems: Retrospective Observational Study," *J. Med. Internet Res.*, vol. 27, 2025, doi: 10.2196/76126.

[9] A. Khumaidi, "Application of Machine Learning to Determine the Factors Affecting Deterioration in Patients with Chronic Kidney Disease," vol. 12, no. 4, 2022.

[10] J. C. Ginestra *et al.*, "Clinician Perception of a Machine Learning–Based Early Warning System Designed to Predict Severe Sepsis and Septic Shock," *Crit. Care Med.*, vol. 47, no. 11, pp. 1477–1484, Nov. 2019, doi: 10.1097/CCM.0000000000003803.

[11] S. Muralitharan *et al.*, "Machine learning–Based early warning systems for clinical deterioration: Systematic scoping review," *J. Med. Internet Res.*, vol. 23, no. 2, Feb. 2021, doi: 10.2196/25187.

[12] A. Rajkomar, J. Dean, and I. Kohane, "Machine Learning in Medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/nejmra1814259.

[13] S. M. Lauritsen *et al.*, "Explainable artificial intelligence model to predict acute critical illness from electronic health records," *Nat. Commun.*, vol. 11, no. 1, Dec. 2020, doi: 10.1038/s41467-020-17431-x.

[14] B. I. Maijamaa, S. Ahmad, A. Musa, A. Ishaq, and A. Ayuba, "Ensemble-Based Imputation for Handling Missing Values in Healthcare Datasets: A Comparative Study of Machine Learning Models," in *The 6th International Electronic Conference on Applied Sciences*, Basel Switzerland: MDPI, Feb. 2026, p. 21. doi: 10.3390/engproc2026124021.

[15] A. A. Wani and F. Abeer, "Comprehensive analysis of missing data imputation in clinical time-series: challenges, risks, and practical solutions," *PeerJ Comput. Sci.*, vol. 12, 2026, doi: 10.7717/peerj-cs.3848.

[16] Y. Raita, T. Goto, M. K. Faridi, D. F. M. Brown, C. A. Camargo, and K. Hasegawa, "Emergency department triage prediction of clinical outcomes using machine learning models," *Crit. Care*, vol. 23, no. 1, Feb. 2019, doi: 10.1186/s13054-019-2351-7.

[17] J. Futoma, S. Hariharan, and K. Heller, "Learning to Detect Sepsis with a Multitask Gaussian Process RNN Classifier," Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.04152

[18] T. J. Moss *et al.*, "Signatures of Subacute Potentially Catastrophic Illness in the ICU: Model Development and Validation*," *Crit. Care Med.*, vol. 44, no. 9, pp. 1639–1648, Sep. 2016, doi: 10.1097/CCM.0000000000001738.

[19] H. Harutyunyan, H. Khachatrian, D. C. Kale, G. Ver Steeg, and A. Galstyan, "Multitask learning and benchmarking with clinical time series data," Aug. 2019, doi: 10.1038/s41597-019-0103-9.

[20] I. Vagliano, O. Lvova, and M. C. Schut, "Interpretable and continuous prediction of acute kidney injury in the intensive care," in *Public Health and Informatics: Proceedings of MIE 2021*, IOS Press, 2021, pp. 103–107. doi: 10.3233/SHTI210129.

[21] D. W. Bates *et al.*, "The potential of artificial intelligence to improve patient safety: a scoping review," Dec. 01, 2021, *Nature Research*. doi: 10.1038/s41746-021-00423-6.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[22] T. A. K. Manne, M. Jindal, S. S. Saini, A. M. Sivagnanam, G. Surianarayanan, and P. C. Jakku, "Real-Time AI–Driven Decision Intelligence For Enterprise Healthcare Operations: Cloud-Native Architecture, Workflow Integration, And Measured Operational Outcomes," *International Journal of Drug Delivery Technology*, vol. 16, no. 12, pp. 1039–1051, 2026, doi: 10.25258/ijddt.16.12s.115.

[23] Bolatito khadijat Aremu and Olagunju Omotola Peggy, "Advanced machine learning-driven business analytics for real-time health risk stratification and cost prediction models," *World Journal of Advanced Research and Reviews*, vol. 26, no. 2, pp. 150–167, May 2025, doi: 10.30574/wjarr.2025.26.2.1583.

[24] S. Lyu, S. Craig, G. O'Reilly, and D. Taniar, "The development and use of data warehousing in clinical settings: a scoping review," 2025, *Frontiers Media SA*. doi: 10.3389/fdgth.2025.1599514.

[25] A. Al-Juhani *et al*, "Advances in Data-Driven Early Warning Systems for Sepsis Recognition and Intervention in Emergency Care: A Systematic Review of Diagnostic Performance and Clinical Outcomes," *Cureus*, Aug. 2025, doi: 10.7759/cureus.89882.