

ORIGINAL RESEARCH

Open access

Rule-Augmented Artificial Intelligence Framework for Detecting Clinically Significant Abnormal Laboratory Result Patterns in Hospitalized Adults Using Sequential Blood Chemistry Panels, Vital Sign Trends, and Physician Response Times

Wei Chen^{1*}, Li Zhang¹

Abstract

Inpatient laboratory monitoring produces frequent blood chemistry results that must be reviewed in relation to the patient's evolving clinical state. Although many results are statistically abnormal, only a smaller subset require urgent interpretation, escalation, or therapeutic action. Conventional rule-based critical value systems depend heavily on fixed thresholds and may generate non-actionable notifications. Pure machine-learning classifiers may detect complex patterns but can be difficult to explain, audit, or align with institutional clinical policies. This article proposes a rule-augmented artificial intelligence framework for detecting clinically significant abnormal laboratory result patterns in hospitalized adults. The framework uses established clinical logic as a structured skeleton and enriches it with sequential laboratory patterns, vital sign trends, and physician response-time feedback. The framework contains a clinical rule knowledge base, a sequential blood chemistry encoder, a vital sign fusion module, and a significance calibration layer. Together, these components would support interpretable pattern detection while allowing alert thresholds to adapt to observed clinical behavior. The proposed architecture could reduce non-actionable alerts by distinguishing isolated statistical abnormalities from evolving clinical patterns. It would also be expected to support patient-specific baselines and integrate into existing inpatient electronic health record workflows. A rule-augmented AI framework offers a pathway toward safer, smarter, and less disruptive laboratory result surveillance. Its value would depend on careful rule curation, transparent model governance, and prospective evaluation in real clinical settings.

Keywords Clinical decision support, Alert fatigue, Vital signs, Rule-augmented artificial intelligence, Laboratory result interpretation, Inpatient monitoring

*Correspondence:

Wei Chen

wei.chen@outlook.com

¹ Department of Health Information Systems, Faculty of Medicine, School of Public Health, Peking University, Beijing, China

Introduction

Hospitalized adults often undergo repeated blood chemistry testing, producing dense streams of laboratory values that must be interpreted alongside diagnoses, medications, procedures, and changing physiology. Electronic health record environments can make these results readily visible, but visibility alone does not ensure that clinically meaningful abnormality patterns are recognized at the right time. Laboratory decision support initiatives such as AMPEL illustrate the

potential of structured laboratory surveillance, while broader EHR-based prediction systems show that sequential clinical data can support automated risk recognition when embedded carefully in care processes [1, 2]. The clinical burden is therefore not simply the number of abnormal results, but the need to distinguish expected, chronic, or transient abnormalities from patterns requiring timely attention.

Current critical value notification systems commonly rely on fixed institutional thresholds for individual analytes, which can be effective for extreme results but limited when clinical meaning depends on trajectory, context, or patient-specific baseline. Closed-loop critical value notification systems can strengthen accountability, yet they may still inherit the rigidity of threshold-driven alerting when abnormality is treated as equivalent to clinical significance [3]. Alert burden is a persistent concern in clinical decision support, and laboratory alerts may contribute to fatigue when duplicate, low-context, or poorly prioritized notifications interrupt clinicians without improving decision-making [4, 5]. A framework for abnormal laboratory pattern detection should therefore preserve safety-critical rules while reducing low-value interruptions through contextual prioritization.

Temporal patterns in blood chemistry panels and vital signs can contain early signals of deterioration that are not captured by isolated abnormal values. Multianalyte delta checks, recurrent clinical time-series models, and EHR-based deterioration prediction systems suggest that longitudinal changes across laboratory and physiological streams may be more informative than single-point measurements [6-8]. Vital sign trends add physiologic context to biochemical abnormalities, particularly for syndromes such as acute kidney injury, sepsis, circulatory failure, and general inpatient deterioration [9-11]. Physician response time is conceptually important because delayed acknowledgment, rapid action, or repeated dismissal may reveal whether alerts were perceived as clinically urgent, although this signal should be treated as noisy and socially mediated rather than a direct measure of ground truth [3, 4].

This article proposes a rule-augmented AI framework that retains clinical interpretability while learning to identify clinically significant abnormality patterns. The framework is grounded in knowledge-based decision support principles, but it uses machine-learning components to represent sequential blood chemistry trajectories, synchronize vital sign context, and calibrate alert significance from clinician response behavior. Rather than replacing clinical rules with an opaque classifier, the framework treats rules as a transparent scaffold that constrains, explains, and structures learned pattern recognition. The intended contribution is conceptual: an AI systems architecture that could be evaluated for safer inpatient laboratory surveillance without reporting experimental performance claims.

Figure 1 presents the proposed rule-augmented artificial intelligence architecture for transforming sequential laboratory values, vital sign trends, and physician response-time signals into explainable and clinically governed abnormality alerts.

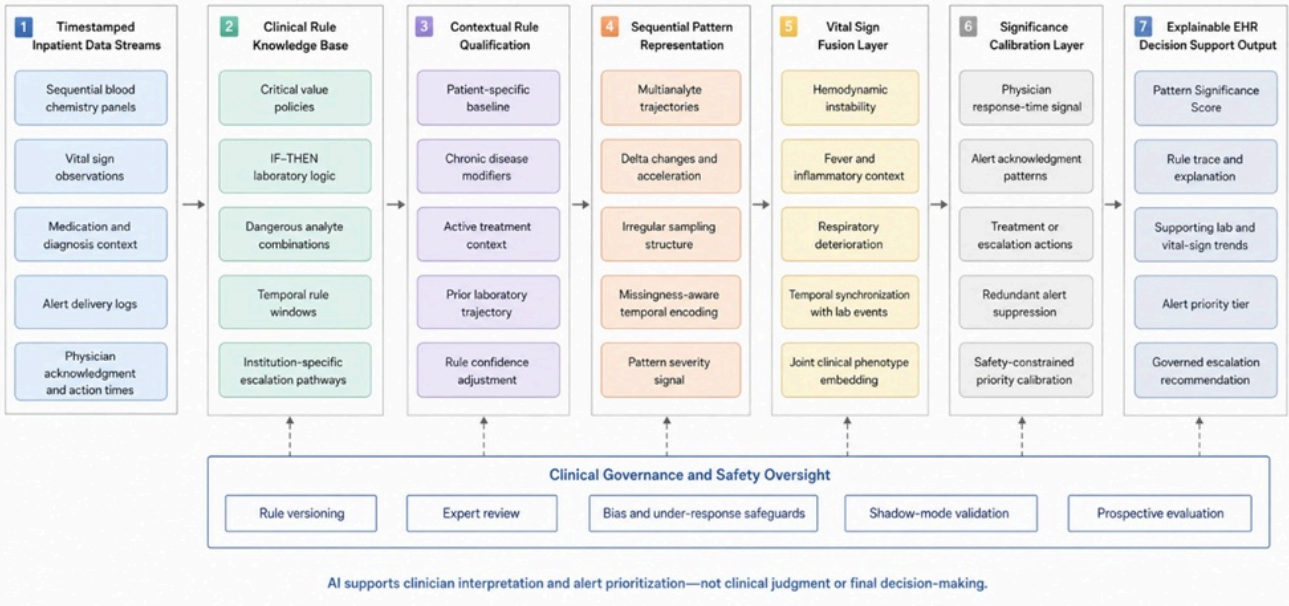


Figure 1. Rule-Augmented AI Framework for Contextual Detection of Clinically Significant Abnormal Laboratory Patterns

Background

Laboratory result interpretation in hospitalized adults

Common inpatient blood chemistry panels include analytes such as electrolytes, renal function markers, liver-associated measurements, glucose, and hematologic indices, each of which may vary because of illness severity, treatment, sampling frequency, chronic disease, or preanalytic factors. A statistically abnormal value may not be clinically significant when it reflects a known baseline, a predictable treatment effect, or a minor excursion outside a reference interval. Conversely, a value within the reference range may be clinically important when it represents a rapid change from the patient’s prior state, especially in the context of evolving deterioration or medication exposure. Laboratory surveillance systems should therefore interpret abnormality as a contextual pattern rather than as a binary property of a single result [1, 6, 9].

Table 1 distinguishes statistical laboratory abnormality from clinically significant abnormality patterns and clarifies how the proposed framework operationalizes this distinction.

Table 1. Analytical Distinction between Statistical Abnormality and Clinically Significant Laboratory Pattern

Interpretive Dimension	Statistical Abnormality	Clinically Significant Pattern	Framework Mechanism That Supports Distinction	Clinical Value Added
Primary unit of interpretation	Single laboratory value outside a reference range	Temporal and contextual episode across multiple data streams	Sequential blood chemistry encoder and rule knowledge base	Reduces over-alerting from isolated low-context abnormalities
Temporal meaning	Often evaluated at one time point	Interpreted through trajectory, rate of change, persistence, and recurrence	Delta change tracking, multianalyte trajectories, irregular sampling representation	Identifies deterioration before extreme thresholds are reached

Patient baseline	Population reference interval dominates interpretation	Patient-specific baseline and recent clinical state shape interpretation	Baseline modeling and contextual rule modifiers	Prevents chronic or expected abnormalities from being over-prioritized
Physiologic context	May be disconnected from bedside status	Gains urgency when paired with vital sign deterioration	Vital sign fusion and synchronization layer	Distinguishes isolated laboratory noise from coherent clinical instability
Rule behavior	Fixed critical value threshold triggers notification	Rule activation is qualified by trajectory, context, and safety constraints	Rule confidence adjustment and contextual modifiers	Preserves safety-critical logic while improving prioritization
Clinician response signal	Usually not incorporated into alert meaning	Used cautiously as a weak behavioral signal of perceived actionability	Physician response-time calibration layer	Helps identify repeatedly non-actionable or high-attention alert patterns
Alerting consequence	High risk of duplicate or low-value alerts	Episode-based prioritization and explanation	Significance calibration and redundant alert suppression	Supports less disruptive and more meaningful clinical decision support
Governance requirement	Threshold review and policy maintenance	Continuous rule, model, and workflow oversight	Expert review, shadow-mode testing, prospective validation	Prevents unsafe adaptation and supports institutional accountability

Rule-based decision support and its limitations

Rule-based clinical decision support has traditionally encoded expert knowledge through computable logic, threshold policies, and IF-THEN statements that can be reviewed by clinicians and aligned with institutional practice. Such systems are attractive because they are transparent and auditable, but their rigidity can become problematic when rules are insensitive to baseline variation, comorbidity, medication context, and temporal change. Modern clinical environments also face alert fatigue when rule-triggered notifications accumulate without sufficient prioritization, duplication control, or contextual explanation [4, 5, 12]. A rule-augmented framework should therefore keep explicit rules for safety and accountability while adding adaptive mechanisms that refine the clinical significance of abnormal patterns.

Sequences and trends in laboratory data

Sequential laboratory interpretation considers not only whether a result is abnormal, but how rapidly it changed, whether multiple related analytes moved together, and whether the pattern is consistent with evolving pathology. Delta checks, cumulative trends, moving averages, and multianalyte temporal representations can help distinguish analytical anomalies, chronic abnormalities, and clinically meaningful trajectories. Neural-network-based multianalyte delta checks and EHR time-series benchmarks suggest that sequential modeling can represent irregular clinical data more flexibly than static threshold rules [6, 8]. For an AIF laboratory framework, these methods would be used conceptually to encode trajectories rather than to assert validated performance in a particular hospital.

Vital sign integration with laboratory data

Vital signs provide continuous or repeated physiological context for interpreting biochemical abnormalities, because laboratory changes often gain urgency when accompanied by tachycardia, hypotension, fever, tachypnea, or other signs

of deterioration. Machine-learning work on inpatient deterioration, sepsis, acute kidney injury, and circulatory failure supports the premise that physiology and laboratory data are complementary streams in clinical prediction [7, 10, 11, 13]. Adding vital sign trends to laboratory interpretation could help distinguish isolated abnormalities from patterns embedded in a broader decompensation phenotype. The framework therefore treats vital signs not as secondary variables, but as synchronized context for determining whether an abnormal laboratory pattern is likely to require attention [9, 14].

Physician response time as a relevance metric

Physician response time to abnormal results or alerts can be conceptualized as a weak behavioral signal of perceived clinical importance. Rapid acknowledgment, order entry, medication adjustment, escalation, or repeat testing may indicate that clinicians considered an abnormality actionable, whereas repeated nonresponse or dismissal may suggest low relevance, notification overload, or competing clinical demands. Studies of closed-loop critical value notification and alert compliance show that clinician response data can be captured and analyzed within clinical decision support workflows, although interpretation requires caution because response behavior is influenced by workload, staffing, interface design, and institutional culture [3, 4]. In the proposed framework, response time would calibrate alert prioritization rather than define clinical truth by itself.

Framework Overview

High-level architecture

The proposed framework ingests timestamped laboratory results, vital sign observations, alert events, acknowledgment logs, and relevant EHR context into a rule-augmented architecture. A clinical rule layer first evaluates explicit policies for critical values, known dangerous combinations, and institutionally defined escalation triggers, while a machine-learning layer scores the severity and trajectory of multivariate abnormality patterns. A feedback loop then uses physician response behavior to calibrate alert thresholds and prioritization logic, with safeguards to prevent reinforcement of unsafe nonresponse patterns. Similar modular thinking is consistent with EHR-based prediction, interoperable clinical decision support, and explainable acute illness modeling, where structured workflows and model outputs must be integrated rather than merely displayed [2, 15, 16].

Core assumptions

The framework assumes access to timestamped blood chemistry panels, vital sign feeds or repeated bedside measurements, medication and diagnosis context, and EHR logs showing alert delivery and acknowledgment. It also assumes that institutional critical value policies can be represented in computable form and periodically reviewed by laboratory medicine specialists, hospitalists, intensivists, informaticians, and patient safety leaders. These assumptions are realistic for many digital hospitals but would vary by EHR configuration, data governance arrangements, and clinical documentation practices. Because clinical decision support effectiveness depends on local workflow integration, the framework should be adapted to institutional conditions rather than treated as a universal plug-in [1, 5, 16].

Design principles

The framework is explainable by construction because the rule layer provides a transparent account of which clinical policies, analyte thresholds, temporal patterns, and contextual modifiers contributed to an alert. It is clinically grounded because machine-learned scores are constrained by expert-authored logic rather than allowed to operate as isolated opaque classifications. It learns from clinician behavior by incorporating response-time feedback as a calibration signal while retaining expert oversight and safety rules for high-risk abnormalities. This design aligns with the broader movement toward interpretable, workflow-aware clinical AI systems that support rather than replace professional judgment [15, 17, 18].

Rule Knowledge Base and Clinical Logic Layer

Encoding clinical guidelines and critical value policies

The rule knowledge base translates institutional laboratory policies into computable IF-THEN logic, such as rules for severe electrolyte disturbances, critical renal markers, marked anemia, or combinations suggesting urgent deterioration. Each rule would specify the analyte, threshold or pattern, temporal window, relevant exclusions, recommended notification level, and explanatory text displayed to clinicians. Unlike a simple threshold table, the knowledge base would allow rules to reference prior values, concurrent laboratory abnormalities, vital sign context, and documented clinical status. This approach extends rule-based laboratory decision support while retaining the auditability required for clinical governance [1, 3, 12].

Rule confidence and contextual modifiers

A rule confidence layer would qualify raw rule activation using contextual modifiers such as baseline chronic kidney disease, active potassium replacement, dialysis status, transfusion, diuretic exposure, or recent operative events. These modifiers would not eliminate safety-critical alerts automatically, but they could adjust priority, explanation, or routing when an abnormal result is expected or already addressed. Logistic and interpretable modeling traditions in clinical prediction suggest that transparent contextual variables can be useful when they are clinically meaningful and reviewed by domain experts [19, 20]. In this framework, context modifies the significance of abnormality while preserving the rule trace needed for accountability.

Interaction with machine-learned scores

The rule layer serves both as a standalone safety mechanism and as a structured input to the learning component. Rule activations, rule confidence values, trend flags, and contextual modifiers could become features in a sequential model that estimates the significance of the evolving pattern. This hybrid arrangement would allow the system to benefit from flexible temporal representation while keeping clinical logic visible, reviewable, and correctable. Such integration is consistent with the need for AI systems that combine EHR-scale pattern recognition with explainability, clinical oversight, and deployment-aware design [5, 17, 18].

Sequential Blood Chemistry Integration

Time-series representation of blood panels

The sequential laboratory encoder would construct per-analyte trajectories for values such as creatinine, potassium, sodium, bicarbonate, hemoglobin, platelet count, bilirubin, and glucose, while preserving measurement time, sampling interval, and missingness pattern. Because hospitalized patients are not sampled at uniform intervals, the representation would need to distinguish true stability from absence of testing and should avoid assuming that missing values are clinically neutral. Clinical time-series modeling work has emphasized the importance of sparse, heterogeneous, and irregular EHR data when representing illness trajectories [8, 17, 21]. The framework would therefore encode both the values and the temporal structure of measurement as part of the abnormality pattern.

Concept drift and patient-specific baselines

Patient-specific baselines are essential because clinically significant change may occur even when a result remains within a population reference interval, while chronic abnormalities may remain non-urgent despite appearing statistically extreme. The framework would model each patient's recent trajectory, historical baseline when available, and expected variation under current treatment conditions. Personalized prediction approaches and broad EHR risk modeling literature support the concept that patient similarity, longitudinal context, and baseline state can improve interpretive relevance [2,

19, 22]. Within this AIF design, the purpose of baseline modeling is not to claim a measured gain, but to ensure that abnormality is judged against the patient's evolving clinical context.

Sequential pattern detection module

A recurrent neural network, transformer, temporal convolutional model, or other time-aware representation could be used to recognize clinically meaningful laboratory trajectories such as rising creatinine, worsening acidosis, falling platelets, increasing bilirubin, or converging multianalyte abnormalities. The model would receive rule activations and contextual modifiers alongside raw trajectories, allowing learned representations to remain connected to clinical logic. Prior clinical AI studies have shown that temporal EHR models can support early recognition of sepsis, acute kidney injury, mortality risk, and broader deterioration, although any new deployment should be evaluated locally and prospectively [8, 10, 13, 23-33]. In the proposed framework, the sequential module would generate a pattern severity signal that supports alert prioritization rather than replacing clinician review.

Vital Sign Trend and Multi-Modal Fusion

Concurrent vital sign analysis

The vital sign module would extract concurrent trends from heart rate, blood pressure, respiratory rate, oxygenation when available, and temperature to provide physiological context for abnormal laboratory patterns. Instead of treating vital signs as isolated observations, the framework would represent direction, persistence, variability, and co-occurrence with laboratory changes. This is important because acute deterioration often emerges through combined physiologic and biochemical signals rather than through a single abnormal value. Early warning score reviews and deterioration prediction studies support the conceptual role of vital sign trajectories in identifying patients whose laboratory abnormalities may require more urgent interpretation [7, 13, 14].

Alignment and synchronization with lab data

Laboratory measurements and vital signs are collected at different frequencies, so the framework would require time-aware alignment rather than simple row-wise merging. A synchronization layer would map laboratory events to recent and subsequent vital sign windows, preserving the temporal distance between biochemical change and physiologic response. This design would allow the model to distinguish, for example, an isolated abnormal potassium value from an abnormal potassium value occurring alongside tachycardia, hypotension, fever, or respiratory instability. Work on sparse heterogeneous clinical time series and ICU trajectory modeling supports the need to represent irregular sampling explicitly when combining EHR data streams [8, 11, 21].

Joint embedding of lab and vital-sign phenotypes

The multi-modal fusion layer would produce a unified representation of the patient's evolving state by combining sequential blood chemistry trajectories, vital sign trends, rule activations, and contextual modifiers. This joint embedding would allow the framework to identify phenotypes such as biochemical deterioration without physiologic instability, physiologic instability preceding laboratory confirmation, or combined deterioration across both streams. Such representations could help prioritize abnormal patterns that are clinically coherent and suppress isolated alerts that lack supportive context. Prior EHR deep learning and explainable acute illness modeling studies suggest that joint representations can capture complex patient trajectories, but the resulting outputs should remain interpretable and clinically reviewable [2, 15, 17, 18].

Physician Response Time Feedback and Significance Calibration

Capturing clinician attention from acknowledgment logs

The physician response-time module would use alert acknowledgment logs, order timing, repeat testing, escalation documentation, and related clinician actions as weak signals of attention to abnormal laboratory patterns. These signals would not be treated as perfect labels because delayed response can reflect workload, handoffs, alert fatigue, or unclear responsibility rather than low clinical importance. However, when aggregated and interpreted cautiously, response behavior could help distinguish alerts that repeatedly prompt action from alerts that are routinely ignored or deferred. Closed-loop critical value systems and alert compliance modeling show that response data can be captured within decision support workflows and used to understand notification effectiveness [3, 4].

Learning to prioritise alerts

The significance calibrator would use response-time feedback to adjust alert priority over time while preserving hard safety rules for immediately dangerous abnormalities. A reinforcement-like feedback loop could down-weight recurrently non-actionable patterns and elevate patterns associated with rapid acknowledgment, escalation, treatment, or repeat testing. This calibration should be constrained by expert review to prevent unsafe learning from under-response, especially in settings where staffing or workflow barriers delay action. Clinical decision support research emphasizes that adaptive systems should reduce alert fatigue without weakening safety-critical notification pathways [4, 5, 24].

Clinical Decision Support and Alerting Integration

Seamless EHR integration

The framework would be embedded within the inpatient EHR results viewer rather than implemented as a separate dashboard requiring clinicians to leave their normal workflow. For each abnormal pattern, the interface could display a Pattern Significance Score accompanied by the rule triggers, relevant trends, supporting vital sign context, and recent clinician actions. Alerts would be generated only when the calibrated significance signal exceeds a clinically governed threshold or when a non-negotiable safety rule is activated. Interoperable clinical decision support initiatives and laboratory CDS implementations highlight the importance of fitting decision support into existing systems, governance structures, and clinician review practices [1, 5, 16].

Reducing alert fatigue and improving actionable notification

Alert fatigue reduction would be pursued through suppression of redundant notifications, grouping of related abnormalities into coherent episodes, and explanation of why a pattern is being escalated. Instead of notifying separately for every abnormal value, the framework would summarize clinically connected changes such as worsening renal function with hyperkalemia and hypotension. The alert explanation would show both rule-based justification and temporal evidence, allowing clinicians to judge whether the recommendation is credible. Prior work on duplicate laboratory alerts, critical value notification, and clinical decision support risks supports the need for prioritization strategies that reduce interruption while preserving timely recognition of meaningful abnormalities [3, 4, 12, 34].

Evaluation Strategy

Retrospective annotation of clinically significant abnormalities

A retrospective evaluation would begin by constructing a reference standard for clinically significant abnormality episodes using expert chart review, documented escalation, urgent interventions, repeat confirmatory testing, and clinician response patterns. The annotation process would distinguish isolated statistical abnormalities from episodes that plausibly required clinical attention. Reviewers would examine laboratory trajectories, vital sign context, medication changes, and notes surrounding abnormal results to determine whether the framework's conceptual alerts correspond to

clinically meaningful events. Similar EHR-based modeling studies underscore the importance of careful outcome definition and clinical context when developing AI systems for deterioration, sepsis, kidney injury, and mortality risk [7, 10, 19, 25-27, 35].

Model performance metrics

Evaluation should focus on clinically meaningful episodes rather than individual abnormal values, because the target of the framework is pattern significance rather than laboratory abnormality detection alone. Appropriate metrics could include event-based sensitivity, precision for clinically significant episodes, timeliness relative to standard notification, proportion of redundant alerts suppressed, and clinician review burden. These metrics should be interpreted in shadow mode before live deployment, without claiming clinical benefit until prospective evaluation demonstrates safe workflow integration. Prior machine-learning work in acute illness prediction, sepsis recognition, and EHR benchmarking provides methodological context for evaluating temporal clinical models, while also reinforcing the need for local validation [8, 13, 23, 25, 27, 29-32].

Prospective shadow-mode validation

Prospective shadow-mode validation would run the framework silently alongside standard critical value notification without changing clinician behavior. During this phase, investigators could compare framework-generated alerts with existing notifications, clinician acknowledgment, treatment actions, escalation events, and expert review determinations. This approach would allow assessment of whether the framework identifies coherent abnormality episodes, suppresses low-value patterns, and provides explanations that clinicians judge useful before any live alerting begins. Shadow-mode evaluation is especially important for adaptive clinical AI because models that appear plausible retrospectively may behave differently when exposed to real-time data latency, missingness, workflow constraints, and changing clinical practice [5, 11, 15, 24, 28].

Table 2 maps each major framework component to its analytical role, implementation risk, and governance requirement.

Table 2. Functional Roles, Risks, and Governance Requirements of the Proposed Rule-Augmented AI Components

Framework Component	Functional Role	Main Analytical Contribution	Key Risk if Poorly Designed	Required Governance Safeguard
Clinical rule knowledge base	Encodes institutional laboratory policies and critical value logic	Provides transparent, auditable clinical structure	Rigid or outdated rules may misclassify evolving clinical significance	Versioned rule review by laboratory medicine, hospital medicine, and informatics teams
Contextual rule qualification layer	Adjusts rule interpretation using baseline, treatment, and clinical context	Separates expected abnormalities from potentially urgent deviations	Unsafe suppression of important alerts in complex patients	Hard safety rules and clinician-reviewable explanation traces
Sequential blood chemistry encoder	Represents longitudinal analyte trajectories and multianalyte changes	Captures deterioration patterns missed by single thresholds	Misinterpretation of sparse or irregular laboratory sampling	Missingness-aware modeling and local validation
Vital sign fusion module	Adds physiologic context to biochemical abnormalities	Identifies coherent clinical phenotypes across laboratory and bedside data	False urgency from noisy or poorly synchronized vital signs	Time-window validation and explicit display of supporting trends

Physician response-time calibrator	Uses acknowledgment and action timing as weak relevance signals	Learns which alert patterns tend to prompt clinical action	Reinforcement of unsafe under-response or workflow inequities	Safety-constrained calibration and expert audit of down-weighted alerts
EHR alerting interface	Presents prioritized alerts with rule trace, trends, and explanation	Converts model output into interpretable clinical decision support	Alert fatigue, poor usability, or workflow disruption	Shadow-mode testing, clinician usability review, and prospective evaluation
Evaluation layer	Assesses episode-level performance and workflow impact	Aligns validation with clinical significance rather than raw abnormality detection	Overclaiming benefit from retrospective plausibility alone	Prospective shadow-mode validation before live deployment

Limitations

Rules require expert maintenance

The rule knowledge base would require ongoing expert maintenance because laboratory policies, clinical guidelines, institutional workflows, and medication practices change over time. Rules developed at one hospital may not transfer directly to another because critical value thresholds, escalation pathways, staffing models, and EHR configurations vary. The framework could therefore become unsafe or ineffective if rules are not versioned, audited, and periodically reviewed by laboratory medicine and clinical informatics teams. Experience with clinical decision support implementation indicates that governance, maintenance, and local adaptation are central to sustainable use.

Data quality and availability

The framework would depend on the quality, completeness, and timeliness of laboratory results, vital signs, alert logs, and clinician response documentation. Missing vital signs, irregular laboratory draws, delayed result posting, undocumented verbal communication, and inconsistent acknowledgment behavior could bias the learned significance calibrator. Patient-specific baseline modeling could also be limited when historical data are unavailable or when care is fragmented across institutions. These limitations are consistent with broader challenges in EHR-based AI, including sparse data, heterogeneous measurement practices, and the risk that models learn documentation behavior rather than clinical state.

Conclusion

The proposed rule-augmented AI framework integrates sequential blood chemistry panels, vital sign trends, and physician response-time feedback to detect clinically significant abnormal laboratory result patterns in hospitalized adults. It treats abnormality as a temporal and contextual phenomenon rather than a single-threshold event. By combining explicit rules with adaptive pattern recognition, the framework offers a clinically grounded path for more meaningful inpatient laboratory surveillance.

A key strength of the framework is its emphasis on interpretability. The rule layer would allow clinicians to see which critical value policies, trajectory patterns, contextual modifiers, and physiologic signals contributed to an alert. The adaptive calibration layer could help prioritize actionable notifications while reducing repeated interruptions from patterns that are unlikely to require immediate attention.

Important challenges remain before such a framework could be considered ready for routine clinical use. The rule base would need continuous expert maintenance, the data pipeline would need safeguards for missingness and latency, and response-time feedback would need careful governance to avoid reinforcing unsafe workflow patterns. Prospective clinical trials and pragmatic implementation studies would be necessary to evaluate safety, usability, and clinical impact.

Future work should focus on collaborative rule curation across laboratory medicine, hospital medicine, critical care, nursing, and clinical informatics. Institutions should evaluate rule-augmented AI in shadow mode before live alerting and should include clinicians in the design of explanations, thresholds, and escalation pathways. With careful governance, this approach could support a safer and less disruptive model of laboratory result monitoring.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 21 Jul 2020

Revised: 23 Aug 2020

Accepted: 15 Sep 2020

Published online: 25 February 2021

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Costa MB, Wernsdorfer M, Kehrer A, Voigt M, Cundius C, Federbusch M, et al. The clinical decision support system AMPEL for laboratory diagnostics: implementation and technical evaluation. *JMIR Med Inform.* 2021;9(6):e20407.
<https://doi.org/10.2196/20407>.
- Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med.* 2018;1(1):18.
<https://doi.org/10.1038/s41746-018-0029-1>.

Li R, Wang T, Gong L, Dong J, Xiao N, Guo M, et al. Enhance the effectiveness of clinical laboratory critical values initiative notification by implementing a closed-loop system: a five-year retrospective observational study. *J Clin Lab Anal.* 2020;34(2):e23038.
<https://doi.org/10.1002/jcla.23038>.

Baron JM, Huang R, McEvoy D, Dighe AS. Use of machine learning to predict clinical decision support compliance, reduce alert burden, and evaluate duplicate laboratory test ordering alerts. *JAMIA Open.* 2021;4(1):ooab006.

Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med.* 2020;3(1):17.
<https://doi.org/10.1038/s41746-020-0221-y>.

Jackson CR, Cervinski MA. Development and characterization of neural network-based multianalyte delta checks. *J Lab Precis Med.* 2020;5:12.
<https://doi.org/10.21037/jlpm.2020.03.03>.

Escobar GJ, Liu VX, Schuler A, Lawson B, Greene JD, Kipnis P, et al. Automated identification of adults at risk for in-hospital clinical deterioration. *N Engl J Med.* 2020;383(20):1951-60.
<https://doi.org/10.1056/NEJMsa2001090>.

Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. *Sci Data.* 2019;6(1):96.
<https://doi.org/10.1038/s41597-019-0103-9>.

Ueno R, Xu L, Uegami W, Matsui H, Okui J, Kusunoki Y, et al. Value of laboratory results in addition to vital signs in a machine learning algorithm to predict in-hospital cardiac arrest: a single-center retrospective cohort study. *PLoS One.* 2020;15(7):e0235835.
<https://doi.org/10.1371/journal.pone.0235835>.

Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature.* 2019;572(7767):116-9.
<https://doi.org/10.1038/s41586-019-1390-1>.

Hyland SL, Faltys M, Hüser M, Lyu X, Gumbsch T, Esteban C, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nat Med.* 2020;26(3):364-373.
<https://doi.org/10.1038/s41591-020-0789-4>.

Beeler PE, Bates DW, Hug BL. Clinical decision support systems. *Swiss Med Wkly.* 2014;144:w14073.
<https://doi.org/10.4414/smww.2014.14073>.

Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG, et al. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med.* 2018;46(4):547-53.
<https://doi.org/10.1097/CCM.0000000000002936>.

Fu LH, Schwartz J, Moy A, Knaplund C, Kang MJ, Schnock KO, et al. Development and validation of early warning score system: a systematic literature review. *J Biomed Inform.* 2020;105:103410.
<https://doi.org/10.1016/j.jbi.2020.103410>.

Lauritsen SM, Kristensen M, Olsen MV, Larsen MS, Lauritsen KM, Jørgensen MJ, et al. Explainable artificial intelligence model to predict acute critical illness from electronic health records. *Nat Commun.* 2020;11(1):3852.
<https://doi.org/10.1038/s41467-020-17431-x>.

Kawamoto K, Kukhareva PV, Weir C, Flynn MC, Nanjo CJ, Liu M, et al. Establishing a multidisciplinary initiative for interoperable electronic health record innovations at an academic medical center. *JAMIA Open.* 2021;4(3):ooab041.

Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform.* 2018;22(5):1589-604.
<https://doi.org/10.1109/JBHI.2017.2767063>.

Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *J Am Med Inform Assoc.* 2018;25(10):1419-28.

Beaulieu-Jones BK, Lavage DR, Snyder JW, Moore JH, Pendergrass SA, Bauer CR, et al. Characterizing and managing missing structured data in electronic health records: data analysis. *JMIR Med Inform.* 2018;6(1):e11.
<https://doi.org/10.2196/medinform.8960>.

Lynam AL, Dennis JM, Owen KR, Oram RA, Jones AG, Shields BM, et al. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. *Diagn Progn Res.* 2020;4(1):6.
<https://doi.org/10.1186/s41512-020-00075-2>.

Bennett CE, Wright RS, Jentzer J, Gajic O, Murphree DH, Murphy JG, et al. Severity of illness assessment with application of the APACHE IV predicted mortality and outcome trends analysis in an academic cardiac intensive care unit. *J Crit Care.* 2019;50:242-6.
<https://doi.org/10.1016/j.jcrc.2018.12.019>.

de Munter L, Polinder S, Lansink KW, Cnossen MC, Steyerberg EW, de Jongh MA, et al. Mortality prediction models in the general trauma population: a systematic review. *Injury.* 2017;48(2):221-9.
<https://doi.org/10.1016/j.injury.2016.11.009>.

Kwon JM, Lee Y, Lee Y, Lee S, Park H, Park J, et al. Validation of deep-learning-based triage and acuity score using a large national dataset. *PLoS One.* 2018;13(10):e0205836.
<https://doi.org/10.1371/journal.pone.0205836>.

Bedoya AD, Clement ME, Phelan M, Steorts RC, O'Brien C, Goldstein BA, et al. Minimal impact of implemented early warning score and best practice alert for patient deterioration. *Crit Care Med.* 2019;47(1):49-55.
<https://doi.org/10.1097/CCM.0000000000003499>.

Giannini HM, Ginestra JC, Chivers C, Draugelis M, Hanish A, Schweickert WD, et al. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. *Crit Care Med.* 2019;47(11):1485-92.
<https://doi.org/10.1097/CCM.0000000000003891>.

Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC, et al. Prediction of sepsis patients using machine learning approach: a meta-analysis. *Comput Methods Programs Biomed.* 2019;170:1-9.
<https://doi.org/10.1016/j.cmpb.2018.12.027>.

Fleuren LM, Klausch TL, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med.* 2020;46(3):383-400.
<https://doi.org/10.1007/s00134-019-05872-y>.

Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med.* 2018;24(11):1716-20.
<https://doi.org/10.1038/s41591-018-0213-5>.

Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early prediction of sepsis in the ICU using machine learning: a systematic review. *Front Med (Lausanne).* 2021;8:607952.
<https://doi.org/10.3389/fmed.2021.607952>.

Masino AJ, Harris MC, Forsyth D, Ostapenko S, Srinivasan L, Bonafide CP, et al. Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data. *PLoS One.*

2019;14(2):e0212665.

<https://doi.org/10.1371/journal.pone.0212665>.

Futoma J, Hariharan S, Heller K. Learning to detect sepsis with a multitask Gaussian process RNN classifier. In: Proceedings of the 34th International Conference on Machine Learning. PMLR; 2017. p.1174-82.

Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. *Comput Biol Med.* 2017;89:248-55.

<https://doi.org/10.1016/j.compbimed.2017.08.015>.

Huff K, Rose RS, Engle WA. Late preterm infants: morbidities, mortality, and management recommendations. *Pediatr Clin North Am.* 2019;66(2):387-402.

<https://doi.org/10.1016/j.pcl.2018.12.008>.

Wasylewicz AT, Scheepers-Hoeks AM. Clinical decision support systems. In: Ristevski B, Chen M, editors. *Fundamentals of clinical data science*. Cham: Springer; 2019. p.153-69.

https://doi.org/10.1007/978-3-319-99713-1_10.

Zhang Z, Bokhari F, Guo Y, Goyal H. Prolonged length of stay in the emergency department and increased risk of hospital mortality in patients with sepsis requiring ICU admission. *Emerg Med J.* 2019;36(2):82-7.

<https://doi.org/10.1136/emermed-2017-207332>.