

ORIGINAL RESEARCH

Open access

Explainable Gradient Boosting Framework for Identifying Key Drivers of Delayed Emergency Department Discharge Using Diagnostic Order Completion Times, Bed Availability, Consultant Response Delays, and Patient Acuity Scores

Maria Hernandez^{1*}, Carlos Vega¹

Abstract

Emergency department discharge delays are a critical operational bottleneck shaped by diagnostic completion intervals, inpatient bed scarcity, consultant responsiveness, and patient acuity. Understanding these drivers in real time is essential for improving flow and reducing avoidable crowding. Current ED analytics often describe aggregate delays after they occur. Clinicians and operational managers therefore lack transparent patient-level tools that indicate which factor is most responsible for a specific delayed discharge episode. This article proposes an explainable gradient-boosting framework for identifying key contributors to delayed ED discharge. The framework focuses on diagnostic order completion times, bed availability, consultant response delays, and patient acuity scores. The proposed framework uses a gradient-boosted tree ensemble trained on historical ED visit data and paired with SHAP-based post-hoc explanations. It is designed conceptually for real-time use with live electronic health record, bed-board, order, and consultation data. Conceptually, the framework would generate both a delay-risk score and an interpretable decomposition of that risk. These explanations could support targeted actions such as expediting a pending diagnostic test, escalating bed-management review, or re-contacting a delayed consultant service. The framework would shift ED discharge management from reactive reporting toward proactive operational decision support. Explainability is positioned as the foundation for clinician trust, workflow alignment, and accountable deployment.

Keywords Explainable artificial intelligence, SHAP, Emergency department, Gradient boosting, Discharge delay, Bed availability

*Correspondence:

Maria Hernandez
maria.hernandez@gmail.com

¹ Department of Health Informatics and Smart Systems, Faculty of Medicine, University of Valencia, Valencia, Spain

Introduction

Emergency department crowding and delayed discharge remain central challenges for acute-care systems because they can degrade patient safety, prolong exposure to crowded environments, and intensify staff workload. Operational delay is not only a throughput concern but also a clinical governance issue, because patients awaiting discharge, transfer, or admission continue to occupy scarce ED capacity while new arrivals compete for attention. Prior work on ED crowding, boarding, and operational delay has emphasized that downstream hospital congestion can translate into upstream

emergency care pressure [1-4]. In this context, delayed ED discharge should be treated as a system-level phenomenon rather than a narrow documentation or transport problem.

The operational drivers of delayed discharge are usually intertwined: diagnostic order completion determines medical clearance, bed availability determines downstream movement, consultant responsiveness affects disposition certainty, and acuity modifies how rapidly pathways can safely proceed. Laboratory and imaging turnaround times may delay readiness for discharge or admission, while bed occupancy and boarding pressures may hold patients in the ED after a disposition decision has been reached [5, 6]. Consultant decision time can further extend the interval between clinical readiness and operational departure, especially when specialty assessment is required before a final plan is accepted [2, 3, 7]. Patient acuity scores also shape delay trajectories because high-acuity patients may require more complex diagnostic sequencing, monitoring, or interdepartmental coordination [8, 9].

Machine learning has been proposed as a way to anticipate ED length of stay, admission, and discharge outcomes, but operational uptake is limited when predictions are not interpretable to frontline users. Recurrent and tree-based models can capture nonlinear patterns in clinical and operational data, yet opaque risk estimates alone may not tell a charge nurse, bed manager, or physician what should be done next [10-12]. Broader healthcare AI discussions have similarly warned that predictive performance without transparent reasoning can undermine trust, accountability, and clinical adoption [13-15]. For ED operations, the relevant question is therefore not only whether delay can be predicted, but whether the prediction can be decomposed into actionable causes.

This article proposes an explainable gradient-boosting framework that would estimate the risk of delayed ED discharge while attributing that risk to operational drivers visible in real time. Gradient-boosted decision trees are attractive for this setting because they can represent nonlinear interactions among timestamps, census variables, order states, and acuity indicators while remaining compatible with post-hoc explanation methods. SHAP, LIME, partial dependence, and related interpretation tools could translate model behavior into ranked driver explanations for individual patients and departmental dashboards. The central thesis is that an XAI framework should move ED analytics beyond retrospective averages toward patient-level explanations that support timely, targeted interventions.

Background

ED discharge process and bottlenecks

The ED discharge process extends from clinical disposition to physical departure, and the interval can be prolonged by incomplete documentation, pending medication dispensing, transport delays, family coordination, bed assignment, and final consultant approval. These bottlenecks often emerge after the clinical decision has been made, making them less visible in traditional length-of-stay summaries that treat the ED visit as a single continuous interval [1, 3]. Operational analytics should therefore distinguish diagnostic readiness, disposition certainty, and physical exit, because each may be delayed by a different mechanism [16, 17]. An explainable framework would be expected to make these distinctions explicit rather than compressing them into a single retrospective throughput metric.

Diagnostic order completion times

Diagnostic order completion times are central to discharge readiness because laboratory, imaging, and cardiology results often determine whether clinicians can safely close an ED episode. Longer order-to-result intervals can extend medical clearance time, especially when CT, X-ray, serial laboratory testing, or repeat diagnostics are required before final disposition [5, 6]. Machine learning models for ED length of stay and discharge outcomes would be expected to treat pending diagnostics, elapsed time since order placement, and modality-specific turnaround as dynamic predictors rather than static descriptors [11, 18]. In an explainable framework, diagnostic delay explanations could reveal whether a patient's predicted discharge delay is mainly driven by an unresolved test rather than by bed pressure or consultant response time.

Bed availability and access block

Bed availability and access block are among the most persistent system-level causes of ED boarding and discharge delay because ED flow depends on inpatient capacity beyond the direct control of ED staff. High inpatient occupancy can prevent admitted patients from leaving the ED, indirectly constraining treatment spaces for new arrivals and slowing downstream disposition for others [2, 7]. Studies of operational delay and ED crowding have therefore treated bed scarcity as a structural driver rather than a patient-specific inconvenience [1, 3]. In the proposed framework, bed-board variables would be expected to explain delay risk when downstream capacity, rather than clinical workup, dominates the patient's pathway.

Consultant response delays and specialty decision time

Consultant response delays can lengthen ED stays when specialty review, admission acceptance, or service-specific recommendations are required before discharge or transfer can proceed. These delays may arise from variable specialty workload, competing inpatient responsibilities, shift timing, or informal communication channels that are not consistently captured in structured records [3, 17]. Predictive models that ignore consultation timestamps may misattribute delay to patient complexity or diagnostic burden when the dominant operational factor is actually delayed decision ownership [8, 18]. An explainable framework should therefore represent consult request time, first response, recommendation completion, and admission decision time as separate operational signals where data systems allow.

Explainable gradient boosting in healthcare operations

Gradient-boosted tree models such as XGBoost, LightGBM, and CatBoost are well suited to ED operations because they can incorporate heterogeneous predictors, missingness patterns, nonlinear thresholds, and interactions among operational variables. Their compatibility with SHAP and related explanation methods makes them especially useful for separating global patterns from local patient-level drivers [19, 20]. Interpretable machine learning research has emphasized that explanations should support meaningful clinical and operational reasoning rather than merely decorate black-box predictions [15, 21, 22]. In ED discharge management, this means that feature importance should be translated into workflow-relevant explanations for charge nurses, physicians, consultants, and bed managers.

Framework Overview

High-level architecture

The proposed architecture would ingest data from electronic health records, admission-discharge-transfer feeds, diagnostic order systems, consultant logs, and bed-board platforms. These streams would be transformed into time-stamped features describing order status, elapsed diagnostic time, real-time bed availability, consultation progress, and triage acuity before being passed into a gradient-boosted tree model [8, 11, 12]. A SHAP-based explanation layer would then translate each prediction into a ranked contribution profile for the patient and a departmental dashboard view [19, 20]. This design would allow the same model output to support both bedside prioritization and system-level flow coordination.

Figure 1 presents the proposed explainable gradient-boosting architecture for translating real-time emergency department operational data into delay-risk predictions, SHAP-based driver explanations, and stakeholder-specific escalation pathways.

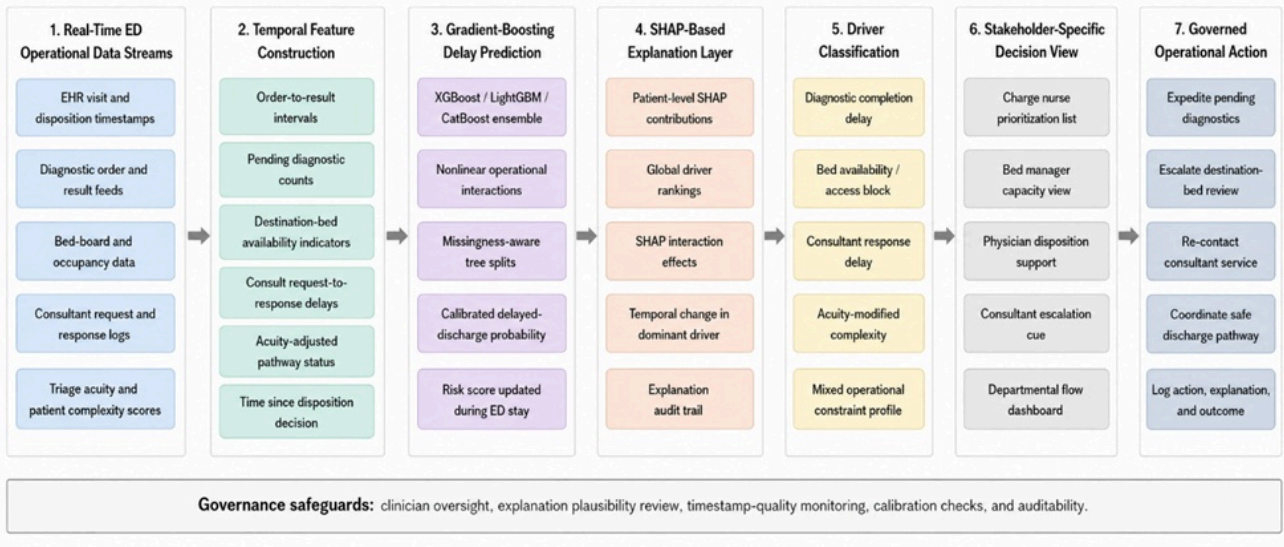


Figure 1. Explainable Gradient-Boosting Framework for Identifying Operational Drivers of Delayed Emergency Department Discharge

Core inputs and outputs

The core inputs would include diagnostic order timestamps, pending result indicators, bed census and occupancy signals, consultant request-to-response intervals, and patient acuity scores. The primary output would be a conceptual delay probability score accompanied by a ranked list of contributing factors that identifies whether the dominant driver is diagnostic completion, bed availability, consultation delay, acuity, or an interaction among them [9, 18, 23]. Rather than functioning as a generic black-box alert, the framework would be expected to show why a given patient is at risk and which operational pathway may require attention [13, 15]. This design aligns prediction with actionability, which is essential for clinical decision-support acceptance.

Design principles

The framework is organized around explainability by design, real-time computability, workflow alignment, and auditability. SHAP values would serve not only as interpretive outputs but also as an audit trail showing how model reasoning changes as new diagnostic, bed, or consult information arrives [19, 20]. Real-time updating should be aligned with ED operational rhythms, because charge nurses and bed managers make repeated prioritization decisions throughout a shift rather than relying on end-of-day reports [16, 17]. The system should therefore be evaluated not only for predictive behavior but also for whether explanations are plausible, timely, and usable in emergency care.

Data Sources and Delay Target Definition

Defining delayed discharge

Delayed discharge should be defined using operational timestamps that distinguish disposition decision time from actual ED departure. A framework could define delay thresholds according to acuity level, disposition pathway, and local policy, while avoiding claims that one universal threshold fits every ED context [1, 3, 8]. The target should reflect a clinically meaningful excess interval after the patient is ready to leave, transfer, or be admitted, rather than total ED length of stay alone [11, 18]. This distinction is important because a long diagnostic workup and a post-disposition bed wait may require entirely different interventions.

Extracting diagnostic order completion times

Diagnostic order features would be extracted from order-entry and result-reporting systems using intervals such as order-to-collection, collection-to-result, imaging order-to-scan, and scan-to-report completion. These features could also include pending order counts, elapsed time since order placement, modality type, and whether a critical diagnostic result remains unavailable [5, 6]. Prior work on ED prediction and operational analytics suggests that such time-stamped process variables are more actionable than broad visit-level summaries because they identify where the care pathway is currently constrained [11, 16]. In an explainable model, these variables would allow a delay prediction to be linked to specific laboratory or imaging bottlenecks.

Bed availability and consultant response metrics

Bed availability metrics would be drawn from real-time bed-board systems and could include unit-level occupancy, blocked beds, pending discharges, and availability of clinically appropriate destination beds. Consultant response metrics would be derived from consultation orders, paging logs, callback documentation, and recorded decision times, although informal communication may remain difficult to capture [2, 7, 17]. Combining these variables with acuity indicators would help the model distinguish a patient who is delayed because the hospital lacks capacity from one delayed because a specialty decision remains incomplete [8, 9]. These data sources should be integrated cautiously because timestamp accuracy and workflow documentation practices may vary across services.

Table 1 defines the operational driver taxonomy that links real-time ED variables to explainable delay mechanisms, stakeholder interpretation, and feasible escalation actions.

Table 1. Operational Driver Taxonomy for Explainable Delayed ED Discharge Prediction

Driver domain	Core variables	Mechanism of delayed discharge	Expected SHAP interpretation	Primary stakeholder	Actionable response
Diagnostic completion delay	Order-to-collection time; collection-to-result time; imaging order-to-report time; pending diagnostic count	Disposition cannot proceed because medical clearance remains incomplete	Positive contribution when unresolved or prolonged diagnostic intervals elevate delay risk	Charge nurse; physician; diagnostic coordination team	Expedite laboratory, imaging, or reporting workflow
Bed availability / access block	Destination-unit occupancy; blocked beds; pending inpatient discharges; appropriate-bed availability	Patient remains in ED despite disposition because downstream capacity is unavailable	Positive contribution when bed scarcity dominates risk after clinical readiness	Bed manager; hospital flow coordinator	Escalate bed review, placement prioritization, or command-center coordination
Consultant response delay	Consult request time; first response time; recommendation completion time; admission acceptance time	Disposition remains uncertain because specialty decision ownership is delayed	Positive contribution when specialty-response intervals exceed expected pathway timing	ED physician; consultant service; charge nurse	Re-contact service, clarify decision responsibility, or escalate delayed consult

Acuity-modified pathway complexity	Triage acuity score; monitoring needs; complexity indicators; high-risk pathway flags	Higher acuity requires more cautious sequencing, monitoring, or coordination	Contextual contribution that may amplify diagnostic, consult, or bed-related delay	ED physician; charge nurse	Review whether delay reflects appropriate clinical caution or modifiable operational blockage
Mixed operational constraint profile	Interactions among diagnostics, bed status, consult delay, and acuity	Multiple constraints jointly prolong discharge beyond any single bottleneck	SHAP interaction values identify combined driver effects	ED leadership; flow team	Coordinate multi-domain response rather than single-task escalation
Documentation or timestamp uncertainty	Missing disposition time; informal consult communication; delayed documentation; inconsistent event logging	Apparent delay may reflect data-quality gaps rather than true workflow delay	Unstable or implausible explanations indicate need for audit rather than direct action	Quality/safety team; informatics team	Review timestamp validity, documentation workflows, and model-input reliability

Explainable Gradient Boosting Architecture

Model choice and training regimen

The proposed model would use a gradient-boosted tree ensemble such as XGBoost or LightGBM because these methods can accommodate nonlinear relationships among diagnostic intervals, bed availability, consultation response, and acuity. Monotonic constraints could be considered where operational logic is strong, such as expecting longer unresolved diagnostic intervals or higher bed occupancy to increase delay risk rather than reduce it [12, 23]. The training objective would be framed conceptually as calibrated binary delay classification, with probability outputs intended for prioritization rather than deterministic decision-making [13, 18]. Model development should emphasize transparency, clinical review, and workflow validity rather than presenting isolated performance numbers.

Post-hoc explanation layer with SHAP

The explanation layer would use TreeSHAP to compute patient-level feature contributions and global driver summaries from the trained gradient-boosting model. These explanations could identify whether a given delay prediction is primarily influenced by pending CT completion, high inpatient occupancy, delayed consultant response, or high acuity interacting with other constraints [19, 20]. LIME, partial dependence, and individual conditional expectation could complement SHAP by testing whether local explanations remain plausible under small feature perturbations [21, 22]. Counterfactual-style what-if analysis should be framed cautiously as operational guidance rather than proof that changing one variable alone would guarantee discharge acceleration.

Handling temporal dynamics and censoring

Delayed discharge risk is time-dependent, so the framework would construct repeated snapshots of each active ED visit as new orders, results, consultant responses, and bed-board updates arrive. This approach would allow explanations to evolve over time, showing whether early diagnostic uncertainty gives way to later access block or consultant delay as the dominant driver [3, 5, 11]. Patients still in the ED at prediction time require careful handling because their eventual

departure time is not yet observed, and naïve labeling could introduce censoring bias [10, 14]. A framework-oriented implementation should therefore specify exclusion rules, delayed-label updating, or survival-oriented adaptations without reporting unsupported experimental outcomes.

Key Driver Identification Mechanism

Global driver rankings

Global driver rankings would use SHAP summary plots and aggregated contribution measures to reveal which operational variables consistently shape delayed ED discharge risk across the department. Bed occupancy, CT completion time, pending laboratory results, consultant response intervals, and acuity scores could be compared across shifts, weekdays, and service lines to identify recurring bottleneck patterns [2, 5-7]. These rankings should be interpreted as model-derived operational signals rather than causal proof, because crowding, diagnostics, and consultation processes often influence one another [1, 3]. Used appropriately, global explanations could help leaders distinguish chronic capacity constraints from time-limited diagnostic or specialty-response problems.

Patient-level local explanations

Patient-level local explanations would decompose a single patient's delay prediction into additive contributions from current operational conditions. A SHAP waterfall-style explanation could show whether pending imaging, a delayed consultant reply, limited destination-bed availability, or high acuity is most responsible for the patient's elevated delay risk [9, 19, 20]. This local interpretability is essential because two patients with similar delay scores may require different interventions depending on the dominant driver [21, 22]. The explanation should therefore be phrased in operational language that connects model reasoning to feasible ED action.

Interaction effects between drivers

Interaction analysis would examine whether one driver becomes more influential under specific operational contexts. For example, consultant delay may matter more when inpatient bed occupancy is high, while diagnostic turnaround may dominate earlier in the visit before disposition is clear [2, 3, 5]. SHAP interaction values, partial dependence, and individual conditional expectation curves could help visualize such nonlinear relationships without reducing them to oversimplified one-factor explanations [19, 20, 23]. These interaction displays should be reviewed with clinicians and operational leaders to ensure that apparent patterns are plausible within local workflow.

Temporal trends in driver influence

Temporal explanation tracking would show how the dominant source of predicted delay changes as the patient's ED stay progresses. Early in an episode, diagnostic uncertainty and pending order completion would be expected to carry greater explanatory weight, whereas later snapshots may increasingly reflect bed availability, transport readiness, or consultant decision delay [5, 6, 11]. Repeated explanations could therefore support escalation timing by showing when a delay is shifting from clinical workup to operational blockage [16, 17]. This temporal view would be especially valuable for distinguishing unavoidable clinical observation from modifiable throughput failure.

Interpretability for Emergency Department Stakeholders

Explanations for the charge nurse

For the charge nurse, explanations should provide a concise patient-level summary of the leading delay drivers and whether they are modifiable within the current shift. A driver list could distinguish actionable issues such as pending laboratory results or an unanswered specialty page from less immediately modifiable factors such as destination-bed

scarcity [3, 5, 17]. This format would align model output with existing prioritization work, allowing the charge nurse to identify which patients may benefit from escalation, coordination, or reassignment of attention [8, 16]. The explanation must remain clear enough to support rapid operational judgment rather than adding cognitive burden.

Explanations for the bed manager

For the bed manager, the framework would aggregate patient-level explanations into a department-wide view of how many active delays are primarily linked to bed unavailability. Such a view could separate delays driven by inpatient occupancy from those driven by diagnostics or consultation, supporting more targeted bed-clearing and placement decisions [2, 7]. Bed-focused explanations would be expected to be most useful when they identify destination-unit constraints rather than merely reporting that the ED is crowded [1, 3]. This aggregation could also support communication between ED leadership, inpatient units, and hospital command centers.

Counterfactual reasoning and actionable suggestions

Counterfactual reasoning would translate explanations into cautious what-if statements about how delay risk might change if an operational barrier were resolved. For example, the system could indicate that completing a pending CT, obtaining a consultant recommendation, or releasing an appropriate inpatient bed would be expected to reduce the model's estimated delay risk [6, 19, 20]. These statements should not be presented as guaranteed causal effects, because operational systems contain dependencies that the model may not fully capture [13, 15]. Instead, counterfactual outputs should function as structured prompts for human review and prioritization.

Building trust through transparency and auditability

Trust would depend on transparent logging of predictions, explanations, user actions, and subsequent outcomes. Regular review could compare model explanations with clinician judgment, identify implausible driver attributions, and reveal whether explanations remain stable across patient groups, shifts, and operational states [13, 21, 22]. Auditability is especially important in emergency care because poorly explained alerts can increase fatigue, reduce confidence, and encourage workarounds [15, 16]. The framework should therefore support feedback loops in which ED staff can challenge explanations and contribute to refinement.

Clinical Integration and Real-Time Decision Support

ED dashboard embedding

The framework would be embedded within the existing ED information system rather than introduced as a separate destination that disrupts workflow. Risk scores and driver lists could update at regular operational intervals using live order, consult, acuity, and bed-board data [11, 16, 17]. Dashboard design should emphasize explanation first, so users see not only that a patient is at risk of delayed discharge but also why the model has reached that assessment [19, 20]. This integration would be expected to support situational awareness while preserving clinician control over decisions.

Tiered interventions based on driver profiles

Tiered interventions would route escalation according to the dominant explanation profile. When diagnostic delay dominates, the system could prompt review by laboratory or radiology coordination; when bed delay dominates, it could support escalation to bed management; when consultant delay dominates, it could trigger review of specialty communication pathways [2, 3, 5, 6]. Acuity information should moderate these suggestions so that intervention recommendations remain clinically appropriate rather than purely throughput driven [8, 9]. The goal is not to automate discharge decisions but to make operational barriers visible early enough for staff to intervene.

Evaluation Strategy

Predictive performance

Predictive evaluation should assess discrimination, calibration, and reliability across operational contexts without treating accuracy as the sole measure of success. Metrics such as AUROC, precision-recall behavior, and calibration curves could be examined conceptually across morning, evening, overnight, weekday, and weekend periods [11, 16, 23]. Because ED operations are dynamic, evaluation should also consider whether predictions remain meaningful as diagnostic, consultation, and bed-board states update during the visit [10, 14]. Performance assessment should be paired with workflow review to ensure that the model supports action rather than simply forecasting unavoidable delay.

Explanation quality metrics

Explanation quality should be evaluated through clinician-rated plausibility, actionability, clarity, and consistency. ED physicians, charge nurses, bed managers, and consultants could review whether model explanations align with observed workflow constraints and whether the ranked drivers suggest reasonable next steps [15, 21, 22]. Fidelity checks should examine whether local explanations reflect the underlying model behavior rather than producing persuasive but misleading narratives [19, 20]. Evaluation should also consider whether explanations reduce or worsen alert fatigue, because interpretability must improve usability rather than adding another layer of noise.

Table 2 provides a deployment-oriented evaluation matrix that separates predictive performance from explanation fidelity, clinical plausibility, actionability, and governance readiness.

Table 2. Evaluation Matrix for Prediction Quality, Explanation Utility, and Safe ED Deployment

Evaluation layer	Key question	Recommended assessment criteria	Failure mode detected	Governance implication
Predictive discrimination	Can the model identify patients at elevated risk of delayed discharge?	AUROC; precision-recall behavior; subgroup performance across shifts and dispositions	High average performance but weak detection during crowded or overnight periods	Require stratified validation before deployment
Calibration	Are predicted delay probabilities reliable enough for prioritization?	Calibration plots; calibration slope/intercept; observed-to-expected delay rates	Overconfident risk estimates that may trigger unnecessary escalation	Use recalibration and threshold review
Temporal stability	Do predictions remain meaningful as ED state changes?	Snapshot-based performance; time-dependent risk trajectories; update consistency	Risk scores fluctuate without plausible operational reason	Review feature timing and update intervals
Local explanation fidelity	Do SHAP explanations accurately reflect model behavior for individual patients?	Perturbation tests; explanation consistency; comparison with model-output changes	Persuasive but misleading driver narratives	Restrict or redesign explanation display
Clinical plausibility	Do frontline users judge explanations as operationally believable?	Structured review by physicians, charge nurses, bed managers, and consultants	Model attributes delay to implausible or non-actionable features	Add domain review and feature governance

Actionability	Does the explanation identify a feasible next operational step?	Proportion of explanations linked to diagnostic, bed, consult, or coordination actions	Explanations describe risk but do not support workflow decisions	Redesign outputs around stakeholder tasks
Equity and subgroup reliability	Are predictions and explanations consistent across patient groups and acuity strata?	Subgroup calibration; explanation distribution review; error-pattern audits	Certain groups receive systematically different or less plausible explanations	Require fairness monitoring and local policy review
Prospective workflow impact	Does explanation-guided use improve coordination without increasing burden?	Silent-mode testing; staged rollout; user workload assessment; delay-process metrics	Alert fatigue, workarounds, or no measurable operational benefit	Continue human oversight and staged deployment only

Prospective impact assessment

Prospective assessment should begin with silent-mode deployment, allowing the framework to generate predictions and explanations without influencing workflow. This phase could compare explanation patterns with retrospective operational baselines and identify whether alerts would have been timely, plausible, and actionable [3, 16, 17]. A later staged or randomized rollout should evaluate whether explanation-guided interventions appear to improve discharge coordination, while avoiding unsupported claims before implementation evidence exists [13, 15]. The strongest evaluation would include multiple EDs so that site-specific bed policies, documentation practices, and consultation workflows can be tested.

Limitations

Data fragmentation and timestamp accuracy

The framework depends on accurate timestamps, but ED operational data are often fragmented across EHR, bed-board, order-entry, paging, and documentation systems. Disposition time may be recorded after the clinical decision, consultant response may occur through informal communication, and diagnostic status may not fully capture real-world workflow delays. Missing or inconsistent timestamps could bias both predictions and explanations, particularly if undocumented work is systematically associated with specific services or patient groups. These limitations mean that explanation outputs should be reviewed as operational decision support rather than definitive accounts of causality.

Generalizability across EDs and patient populations

The framework may require local adaptation because discharge processes, inpatient bed policies, consultation norms, pediatric pathways, rural transfer patterns, and staffing models differ across EDs. A driver that appears dominant in one hospital may be less relevant in another, especially when bed management, diagnostic capacity, or specialty coverage is organized differently. Explainability may expose these site-specific idiosyncrasies, which is valuable for local improvement but limits direct portability. Multi-site evaluation and transparent recalibration would therefore be needed before broader deployment.

Conclusion

An explainable gradient-boosting framework for delayed ED discharge would combine real-time prediction with patient-level attribution of operational drivers. By focusing on diagnostic order completion times, bed availability, consultant

response delays, and acuity scores, the framework would make delay risk more interpretable to frontline and managerial stakeholders.

The main strength of this approach is its ability to connect a risk estimate with an explanation that can guide action. Rather than presenting delay as a generic probability, the system would identify whether a patient's risk is most related to pending diagnostics, bed scarcity, specialty response, or acuity-modified pathway complexity.

Important challenges remain in data completeness, timestamp validity, real-time integration, and user trust. Explanations must be understandable, clinically plausible, and operationally useful, and they must be monitored to ensure that they change behavior in constructive rather than burdensome ways.

Future work should include prospective, multi-site demonstration studies and the development of ED-specific explainability benchmarks. Such studies should examine not only prediction quality but also whether explanation-guided interventions improve coordination, reduce avoidable delay, and support safer emergency care operations.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 07 Nov 2021

Revised: 15 Dec 2021

Accepted: 04 Feb 2022

Published online: 25 February 2022

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

Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: a systematic review of causes, consequences and solutions. *PLoS One*. 2018;13(8):e0203316.
<https://doi.org/10.1371/journal.pone.0203316>.

Kenny JF, Chang BC, Hemmert KC. Factors affecting emergency department crowding. *Emerg Med Clin North Am*. 2020;38(3):573-87.
<https://doi.org/10.1016/j.emc.2020.04.002>.

Joseph JW, White BA. Emergency department operations: an overview. *Emerg Med Clin North Am*. 2020;38(3):549-62.
<https://doi.org/10.1016/j.emc.2020.04.001>.

Mueller B, Kinoshita T, Peebles A, Graber MA, Lee S. Artificial intelligence and machine learning in emergency medicine: a narrative review. *Acute Med Surg*. 2022;9(1):e740.
<https://doi.org/10.1002/ams2.740>.

Bourchany A, Thauvin-Robinet C, Lehalle D, Bruel AL, Masurel-Paulet A, Philippe C, et al. Reducing diagnostic turnaround times of exome sequencing for families requiring timely diagnoses. *Eur J Med Genet*. 2017;60(11):595-604.
<https://doi.org/10.1016/j.ejmg.2017.07.006>.

Jaffe TA, Goldstein JN, Yun BJ, Etherton MR, Leslie-Mazwi TM, Feske SK, et al. Impact of emergency department crowding on delays in acute stroke care. *West J Emerg Med*. 2020;21(4):892-901.

Boudi Z, Lauque D, Alsabri M, Östlundh L, Oneyji C, Khalemsky A, et al. Association between boarding in the emergency department and in-hospital mortality: A systematic review. *PLoS One*. 2020;15(4):e0231253.
<https://doi.org/10.1371/journal.pone.0231253>.

De Hond AA, Raven W, Schinkelshoek L, Gaakeer MI, Ter Avest E, van den Berg E, et al. Machine learning for developing a prediction model of hospital admission of emergency department patients: hype or hope? *Int J Med Inform*. 2021;152:104496.
<https://doi.org/10.1016/j.ijmedinf.2021.104496>.

Alanazi A. Using machine learning for healthcare challenges and opportunities. *Inform Med Unlocked*. 2022;30:100924.
<https://doi.org/10.1016/j.imu.2022.100924>.

Yu Y, Si X, Hu C, Zhang J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput*. 2019;31(7):1235-70.
https://doi.org/10.1162/neco_a_01199.

Hunter-Zinck HS, Peck JS, Strout TD, Gaehde SA. Predicting emergency department orders with multilabel machine learning techniques and simulating effects on length of stay. *J Am Med Inform Assoc*. 2019;26(12):1427-36.

Zeng X. Length of stay prediction model of indoor patients based on light gradient boosting machine. *Comput Intell Neurosci*. 2022;2022:9517029.

Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. 2019;25(9):1337-40.
<https://doi.org/10.1038/s41591-019-0548-6>.

Cheung DS, Grubenhoff JA. Machine learning in clinical medicine still finding its way. *JAMA Netw Open*. 2019;2(1):e186926.
<https://doi.org/10.1001/jamanetworkopen.2018.6926>.

Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1(5):206-15.
<https://doi.org/10.1038/s42256-019-0048-x>.

Pianykh OS, Guitron S, Parke D, Zhang C, Pandharipande P, Langlotz CP, et al. Improving healthcare operations management with machine learning. *Nat Mach Intell.* 2020;2(5):266-73.
<https://doi.org/10.1038/s42256-020-0176-9>.

Musen MA, Middleton B, Greenes RA. Clinical decision-support systems. In: *Biomedical informatics: computer applications in health care and biomedicine*. Cham: Springer; 2021. p.795-840.
https://doi.org/10.1007/978-3-319-99713-1_32.

Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLoS One.* 2018;13(7):e0201016.
<https://doi.org/10.1371/journal.pone.0201016>.

Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30:4765-74.

Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat Mach Intell.* 2020;2(1):56-67.
<https://doi.org/10.1038/s42256-019-0138-9>.

Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Ahmad MA, Eckert C, Teredesai A. Interpretable machine learning in healthcare. In: *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2018:559-60.
<https://doi.org/10.1145/3240508.3243650>.

Habehh H, Gohel S. Machine learning in healthcare. *Curr Genomics.* 2021;22(4):291-300.
<https://doi.org/10.2174/1389202922666210122211612>.