REVIEW **Open access**

Predictive Analytics for Emergency Department Crowding and Patient Flow Optimization: A Systematic Review of Machine Learning Models, Input Features, and Operational Outcomes

Claire Dupont[1*], Julien Martin[1]

Abstract

Emergency department crowding is a persistent global healthcare challenge linked to longer wait times, increased patients leaving without being seen, worse clinical outcomes, and staff burnout. It also contributes to ambulance diversion and inefficient resource use, worsening hospital operational strain. This systematic review evaluates machine learning models for predicting ED crowding and optimizing patient flow, focusing on input features (e.g., arrival rates, acuity, bed availability) and reported operational outcomes such as waiting times and ambulance delays. A PRISMA-compliant review was conducted across PubMed, Embase, IEEE Xplore, and Scopus. Included studies applied machine learning to ED crowding or patient flow prediction and reported operational or crowding outcomes. Due to heterogeneity, a narrative synthesis was used, and risk of bias was assessed using an adapted tool. Thirty-two studies met inclusion criteria, using classification, regression, time-series, and deep learning models. Common predictors included arrival patterns, occupancy, and bed availability. While predictive performance was generally high, few studies evaluated real-world operational impacts, and most remained retrospective. Although machine learning models demonstrate strong predictive accuracy for ED crowding, evidence of real-world operational benefits remains limited. A clear gap exists between prediction and implementation into clinical workflow and decision-making. Future research should focus on translating predictions into measurable improvements in ED performance.

Keywords Machine learning, Predictive analytics, Emergency department crowding, Patient flow optimization, Operational outcomes, NEDOCS

*Correspondence:
Claire Dupont
claire.dupont@gmail.com
[1] Department of Healthcare Data Analytics, Faculty of Medicine, University of Bordeaux, Bordeaux, France

Introduction

Emergency department crowding is defined as a mismatch between patient demand and available resources, resulting in prolonged waiting times, treatment delays, and increased patient safety risks. This phenomenon has been linked to higher mortality rates, elevated incidences of patients leaving without being seen, and frequent ambulance diversion episodes that disrupt prehospital care networks. Staff burnout and reduced job satisfaction further compound the issue, affecting overall healthcare workforce retention [1]. These multifaceted consequences highlight crowding as a systemic challenge requiring innovative predictive solutions beyond traditional reactive measures.

Traditional crowding measures such as the National Emergency Department Overcrowding Score (NEDOCS), Emergency Department Work Index (EDWIN), and occupancy ratios primarily quantify the current state of the department after crowding has already developed. These tools rely on static snapshots of bed availability, staffing levels, and patient acuity, limiting their utility for timely intervention. Consequently, they often fail to provide the advance warning necessary for implementing surge protocols or resource reallocations [2]. Recent evaluations have reaffirmed the reactive nature of these metrics in dynamic emergency environments.

Predictive analytics leveraging machine learning promise a shift toward proactive intervention by forecasting crowding events several hours or days in advance using complex data patterns. Models can integrate real-time inputs to anticipate surges in arrivals or boarding times, potentially enabling earlier staffing adjustments and diversion decisions. However, it remains uncertain whether these forecasts consistently translate into improved operational performance or merely remain theoretical exercises [3]. The promise of machine learning therefore hinges on its demonstrated impact on actual patient flow metrics.

This systematic review was undertaken to synthesize evidence on machine learning models for emergency department crowding prediction and patient flow optimization, with explicit emphasis on input features and operational outcomes. It maps the landscape of forecasting methods, from time-series approaches to deep learning architectures, while evaluating real-time versus retrospective applications. The objectives include identifying common implementation barriers and providing a structured roadmap for future research and clinical adoption [4]. Ultimately, the review aims to determine whether predictive analytics can deliver measurable benefits in emergency care delivery.

Materials and Methods

Search strategy

A comprehensive literature search was executed across PubMed, Embase, IEEE Xplore, and Scopus to capture peer-reviewed publications on predictive analytics for emergency department crowding and patient flow. Predefined search strings combined terms such as “emergency department crowding” with “prediction” and “machine learning,” supplemented by variants targeting patient flow, NEDOCS, and operational outcomes. The temporal limit was strictly confined to 2017–2022 to focus on contemporary machine learning advancements while excluding outdated methodologies [5]. This multi-database approach maximized retrieval of interdisciplinary studies from clinical and computational domains.

Additional targeted strings addressed specific themes including “ambulance diversion prediction,” “real-time prediction emergency department,” and “ED length of stay machine learning forecast” to ensure exhaustive coverage of relevant literature. Boolean operators and proximity searches were applied to refine results and reduce noise from unrelated topics. The strategy explicitly prioritized studies reporting any crowding metric or operational outcome, aligning with the review’s focus on translational impact [6]. Duplicate records were removed using automated tools prior to screening.

Inclusion and exclusion criteria

Studies were included if they described the development, validation, or application of machine learning models for predicting emergency department crowding, patient arrivals, or flow metrics and were published in peer-reviewed journals between 2017 and 2022. Eligible articles had to reference at least one operational outcome such as waiting time, length of stay, diversion, or left-without-being-seen rates, or employ recognized crowding scores like NEDOCS or EDWIN. Both retrospective and prospective designs were accepted provided they utilized electronic health record data or similar real-world inputs for model training or testing [7].

Exclusion criteria eliminated non-machine learning approaches, purely descriptive studies without predictive modeling, and publications lacking any connection to emergency department operational outcomes or crowding prediction. Non-

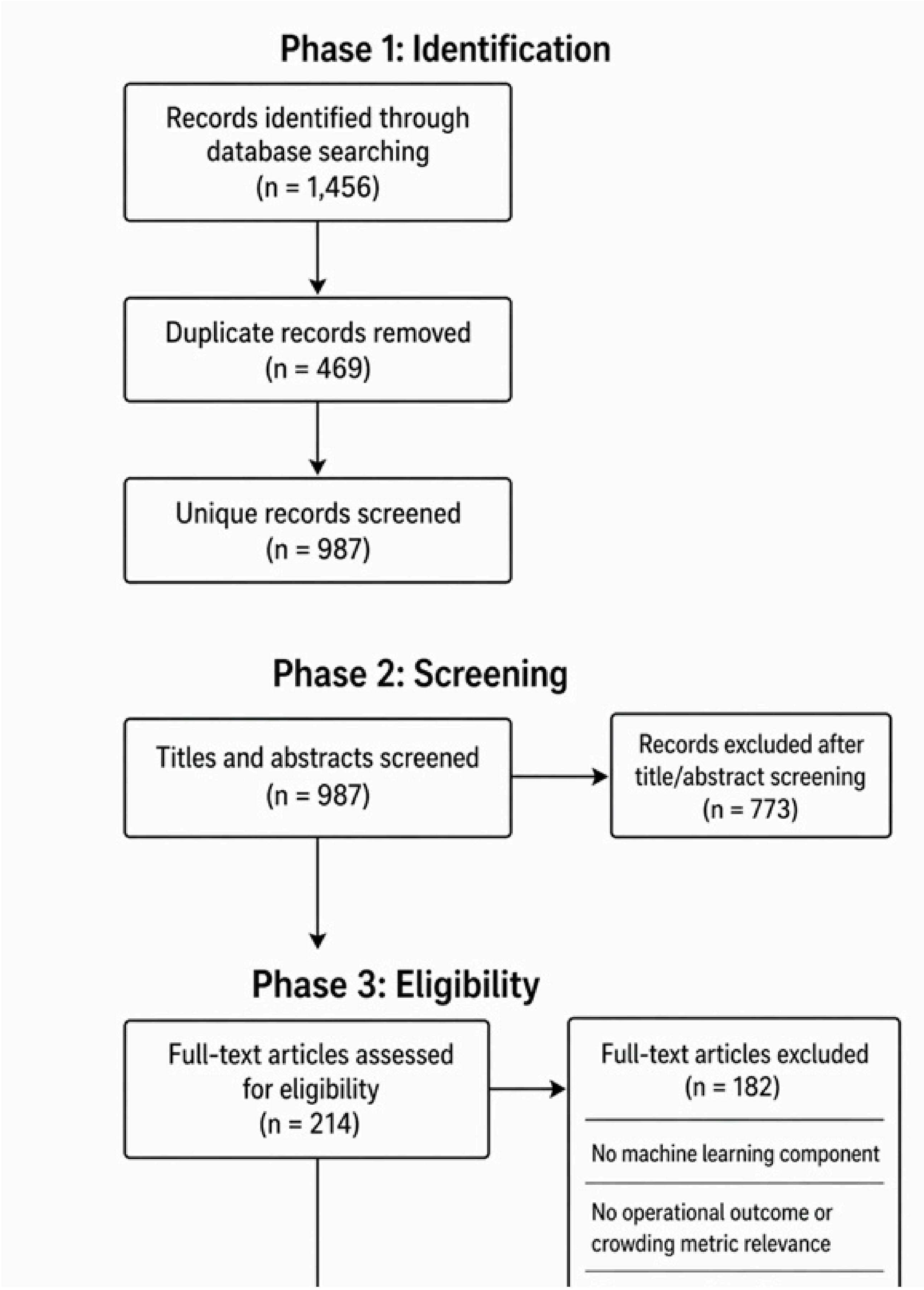
English language articles, conference abstracts, and gray literature were also excluded to maintain methodological rigor and focus on high-quality evidence. Pediatric-only or non-general ED settings were retained if they contributed generalizable insights into flow optimization [8]. These criteria ensured the selected body of evidence directly addressed the review's core objectives.

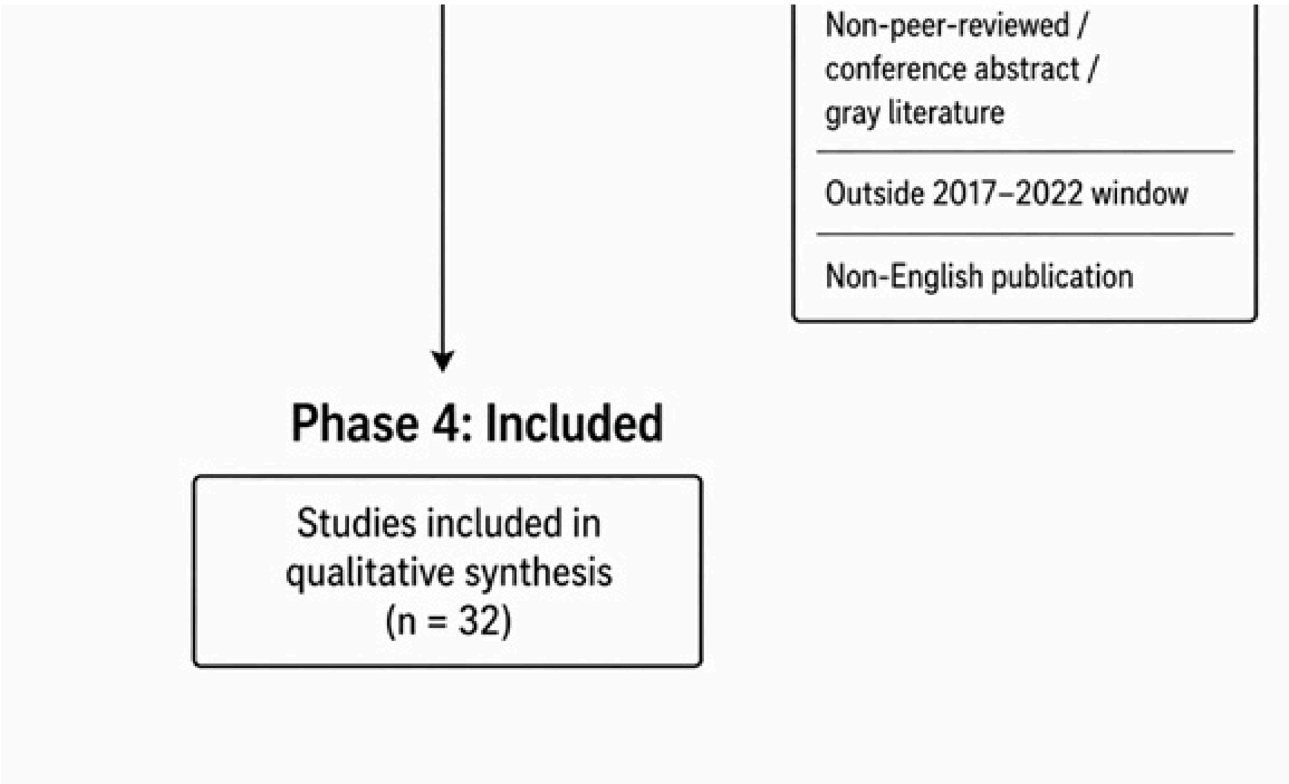
Screening and selection

Title and abstract screening was performed independently by two reviewers using standardized eligibility forms, with disagreements resolved through consensus discussion or third-reviewer arbitration. Full-text assessment followed for potentially relevant records, again conducted in duplicate to minimize selection bias. A PRISMA flow diagram was constructed to document the identification, screening, eligibility, and inclusion phases, illustrating the progression from initial records to the final set of included studies [9]. Reasons for exclusion at each stage were logged systematically for transparency.

Figure 1 illustrates the PRISMA 2020-compliant study selection process from database identification through final inclusion of 32 studies.

https://doi.org//s711815341





Risk of bias assessment

Risk of bias was evaluated using an adapted version of the Prediction model Risk Of Bias Assessment Tool (PROBAST), modified to emphasize operational prediction contexts and implementation-related biases in emergency department settings. Domains assessed included participant selection, predictor measurement, outcome definition, and analysis methods, with additional items addressing real-world deployment feasibility and alert fatigue potential. Each study received an overall risk-of-bias rating of low, moderate, or high, documented with supporting justifications [13].

The adaptation accounted for the unique challenges of emergency department data, such as temporal variability and missing real-time inputs, which could inflate bias in retrospective designs. Studies were not excluded based on bias ratings but rather flagged to inform interpretation of findings. Dual independent assessment was performed, with consensus used for final ratings [14]. This process highlighted common methodological weaknesses across the evidence base.

Synthesis methods

Narrative synthesis was employed to integrate findings due to substantial heterogeneity in machine learning architectures, input features, prediction horizons, and operational outcome definitions precluding meta-analysis. Studies were grouped thematically by model type, feature categories, and reported outcomes to facilitate pattern identification and gap analysis. Qualitative descriptions of prediction performance and implementation status were prioritized alongside quantitative summaries where comparable metrics existed [15].

No formal statistical pooling was attempted given the diversity of endpoints and reporting standards. Sensitivity analyses were not feasible but subgroup considerations (e.g., short-term versus long-term horizons) were explored narratively. The synthesis explicitly linked model characteristics to operational relevance, emphasizing translational gaps [16]. This approach provided a comprehensive overview while acknowledging limitations in direct comparability.

Results and Discussion

Study selection

The systematic search initially identified 1,456 records across the four databases, which were reduced to 987 unique citations after duplicate removal. Following title and abstract screening, 214 articles advanced to full-text review, of which 182 were excluded for reasons including absence of machine learning components or lack of operational outcome reporting. Ultimately, 32 studies met all eligibility criteria and were included in the qualitative synthesis [17]. The PRISMA flow diagram illustrates this selection process and documents the primary exclusion rationales at each stage.

Hand-searching of reference lists and forward citation tracking added no additional studies beyond the database yield. All included publications were peer-reviewed journal articles published within the 2017–2022 window, confirming adherence to temporal restrictions. No studies were excluded solely on the basis of high risk of bias, allowing comprehensive representation of the existing evidence [18]. This final set provided sufficient breadth to address the review objectives regarding models, features, and outcomes.

Study characteristics

The 32 included studies were conducted predominantly in academic or tertiary emergency departments, with fewer investigations from community or pediatric settings, and spanned multiple geographic regions including North America, Europe, and Asia. Sample sizes ranged widely from several thousand to over one million patient encounters, reflecting both single-site and multisite designs. Most analyses utilized retrospective electronic health record data, with only a minority incorporating prospective validation or real-time testing phases [19]. Prediction horizons varied but clustered around short- to medium-term forecasts relevant to daily operations.

Geographic distribution showed a concentration in high-income countries with advanced electronic health record infrastructure, potentially limiting generalizability to lower-resource settings. Pediatric-focused studies often emphasized acuity-based flow predictions, while adult ED investigations frequently incorporated hospital-wide occupancy metrics. Temporal coverage within the 2017–2022 period captured the pre-pandemic baseline and early pandemic adaptations in several cases [20]. These characteristics underscore both the strengths and contextual limitations of the current evidence base.

Prediction horizons and model types

Short-term horizons (1–4 hours) predominated among the 32 studies, often employing classification or regression models to support immediate triage and resource allocation decisions. Medium-term forecasts (6–12 hours) were frequently addressed through time-series or ensemble methods, while longer-term predictions (24–48 hours) relied more heavily on deep learning architectures to capture complex temporal dependencies. Hybrid approaches combining traditional machine learning with neural networks appeared in several recent publications, demonstrating improved handling of non-linear patterns in arrival data [21].

Real-time prediction models were less common than retrospective analyses, with only a subset demonstrating deployment readiness. Deep learning variants showed particular promise for high-dimensional inputs but required substantial computational resources. Time-series techniques excelled in capturing seasonality and trends but struggled with sudden disruptions such as mass-casualty events [22]. Overall, model selection appeared driven by data availability and intended use case rather than standardized performance benchmarks.

Input features

Historical arrival patterns, current ED occupancy levels, and hospital-wide bed availability emerged as the most frequently utilized input features across the reviewed studies, providing foundational signals for crowding forecasts. Acuity scores, ambulance arrival volumes, and staffing ratios were commonly incorporated alongside calendar variables such as day of week, holidays, and local event indicators. Weather data and regional epidemiological trends appeared in a smaller number of models, adding external context to internal operational variables [23]. Feature engineering techniques varied widely, with some studies applying dimensionality reduction to manage high-dimensional electronic health record datasets.

Real-time data streams were emphasized in models intended for operational deployment, yet many relied on batch-processed historical aggregates that limited immediate applicability. Lack of standardization in feature selection hindered direct comparisons between studies. Input feature importance analyses, when reported, consistently ranked occupancy and arrival rates highest, aligning with clinical intuition [24]. These patterns highlight both the richness and the fragmentation of data utilization in current predictive efforts.

Operational outcomes

Only a minority of the 32 studies reported direct operational outcomes, with most focusing instead on predictive accuracy metrics such as area under the receiver operating characteristic curve without linking forecasts to downstream interventions. Among those that did evaluate outcomes, waiting time, length of stay, ambulance diversion rates, and left-without-being-seen proportions were the most frequently assessed, though measurement methods differed substantially. Implementation status was predominantly retrospective simulation rather than prospective deployment with triggered actions like staffing adjustments or surge protocols [25]. This disconnect limited insights into real-world effectiveness.

A small subset of investigations documented reductions in diversion or boarding times following model-informed decisions, yet these findings were often derived from controlled simulations rather than live ED environments. Alert fatigue and clinician acceptance were rarely quantified despite their relevance to sustained use. Overall, the evidence revealed a clear emphasis on technical performance over operational impact, constraining conclusions about practical utility [26]. Future studies must prioritize outcome measurement to close this translational gap.

Summary of principal findings

The reviewed machine learning models consistently demonstrated high predictive accuracy for emergency department crowding and patient flow metrics, with area under the curve values typically ranging from 0.80 to 0.95 across diverse datasets. Input features centered on readily available operational variables such as occupancy ratios and arrival rates, enabling feasible integration with existing electronic health record systems. However, operational outcomes including waiting times, diversion rates, and left-without-being-seen proportions were reported in fewer than one-third of studies, revealing a predominant focus on model development rather than impact evaluation [27]. These findings confirm the technical maturity of predictive analytics while exposing limited evidence of translational success.

Figure 2 synthesizes the review's core conceptual finding that diverse machine learning models and rich operational inputs have outpaced rigorous evaluation of real-world emergency department outcomes.

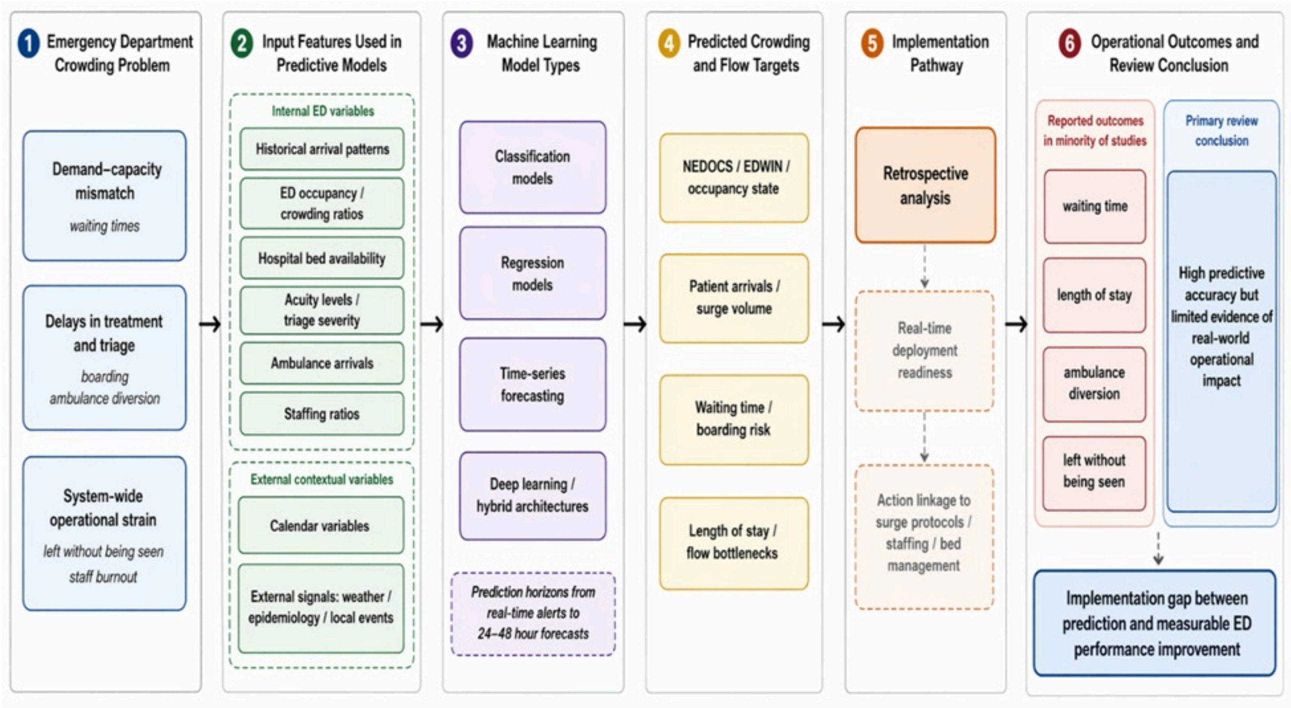


Figure 2. Conceptual Evidence Map of Machine Learning Prediction for Emergency Department Crowding and Patient Flow Optimization

Narrative synthesis across the 32 studies further illustrated that real-time applications remain underrepresented compared with retrospective analyses. Deep learning and ensemble methods offered marginal gains in complex scenarios but introduced interpretability challenges that could impede clinical adoption. Crowding metrics such as NEDOCS and EDWIN served primarily as validation targets rather than intervention triggers [28]. Collectively, the principal findings underscore both the promise and the current limitations of machine learning in emergency department optimization.

The implementation gap

Prediction without corresponding action appears insufficient for meaningful improvement in emergency department performance, as few studies explicitly linked model outputs to predefined interventions such as surge protocols or staffing reallocations. Alert fatigue remains a theoretical concern in the literature but is seldom quantified, potentially undermining long-term usability of real-time systems. Retrospective designs dominate the evidence base, limiting insights into workflow integration, clinician trust, and sustainability in live environments [29]. Bridging this gap requires deliberate study designs that progress from silent-mode testing to randomized implementation trials.

Table 1 provides a translational appraisal framework that clarifies why strong predictive performance has not yet consistently translated into measurable improvements in emergency department operations.

Table 1. Translational Appraisal Framework for Emergency Department Crowding Prediction Studies

Domain of appraisal	What should be evaluated in ED crowding prediction studies	What the review found across the literature	Translational implication
Prediction objective	Whether the study predicts crowding state, arrivals, waiting	Prediction targets were diverse, with many studies centered on	Heterogeneous targets make cross-study comparison difficult

	time, boarding, length of stay, or broader flow disruption	crowding scores, arrivals, and waiting-time-related endpoints	and weaken pathway standardization for operational deployment
Prediction horizon	Whether forecasts are real-time, 1–4 hours, 6–12 hours, or 24–48 hours ahead	Short-term horizons predominated, while longer horizons were less common and usually more technically demanding	Horizon choice determines whether forecasts support immediate bed allocation or earlier strategic staffing and surge planning
Input feature strategy	Whether models rely on arrivals, occupancy, bed availability, acuity, staffing, calendar effects, and external contextual signals	Occupancy, arrival patterns, and bed availability were the most recurrent predictors; broader feature sets were inconsistently applied	The field has identified a stable operational core of predictors, but lack of standardization limits reproducibility and portability
Modeling architecture	Whether classification, regression, time-series, deep learning, or hybrid methods are used and justified by the operational use case	All major model families were represented, with architecture choice often driven by local data availability rather than unified benchmarking	Technical heterogeneity has expanded innovation, but it has also fragmented the evidence base
Validation setting	Whether the study is retrospective, temporally validated, prospectively tested, or externally validated across sites	Most studies were retrospective and single-site; prospective and multisite validations were uncommon	Internal accuracy may overstate readiness for live use in diverse emergency department environments
Operational linkage	Whether model outputs are explicitly connected to interventions such as surge activation, staffing change, diversion control, or bed escalation	In most studies, predictions were not tied to predefined operational actions	Prediction without an intervention pathway limits real-world value even when model performance is high
Outcome reporting	Whether studies measure waiting time, length of stay, diversion, boarding, LWBS, or related workflow effects after model use	Only a minority reported operational outcomes, and many focused solely on AUROC or similar technical metrics	The literature remains dominated by predictive performance rather than demonstrated service improvement
Human-factors readiness	Whether alert fatigue, clinician trust, interpretability, and workflow burden are evaluated	These dimensions were rarely reported	Omission of human-factors evidence weakens implementation feasibility
Generalizability	Whether the model is portable across sites, populations, and information infrastructures	Evidence was concentrated in academic and high-resource settings	Current evidence may not transfer reliably to community or resource-constrained emergency departments
Implementation maturity	Whether the study remains proof-of-concept, silent-mode, simulation-based, or	The evidence base was largely retrospective or simulation-oriented	The field is technically mature enough for forecasting, but not yet operationally mature

	prospectively intervention-triggering		enough for confident widespread adoption

The scarcity of prospective studies evaluating operational outcomes after model deployment further exacerbates the implementation gap. When interventions were simulated, reported benefits were often modest and context-specific, raising questions about generalizability. Stakeholder engagement and human-factors evaluation were infrequently described, despite their critical role in successful technology adoption [30]. Addressing these deficiencies is essential to transform predictive capabilities into tangible enhancements in patient flow and staff well-being.

Input feature heterogeneity

Considerable heterogeneity exists in the selection and processing of input features, with studies varying widely in their use of historical versus real-time data streams and inclusion of external variables such as weather or local events. This lack of standardization complicates cross-study comparisons and hinders the development of universally applicable models for emergency department crowding prediction. Real-time data availability poses ongoing technical and infrastructural challenges, particularly in community hospital settings with less sophisticated information systems [31]. Feature importance rankings, when provided, consistently emphasize occupancy and arrival metrics yet rarely incorporate clinician-defined variables that could enhance actionability.

The variability in feature engineering techniques further contributes to inconsistent model performance across different emergency department contexts. Some investigations applied advanced methods such as embedding layers for temporal data, while others relied on simpler aggregations, reflecting differing computational resources and expertise levels. Standardization efforts could accelerate progress toward interoperable predictive systems [32]. Until such efforts mature, heterogeneity will continue to limit the scalability of machine learning solutions for patient flow optimization.

Prediction horizon trade-offs

Longer prediction horizons enable earlier operational interventions such as preemptive staffing adjustments but are generally associated with reduced accuracy due to accumulating uncertainties in patient arrival patterns. Conversely, short-term horizons deliver higher precision for immediate triage and bed management yet provide insufficient lead time for strategic resource reallocation. This inherent trade-off was evident across multiple studies, with deep learning models sometimes mitigating but not eliminating the accuracy decrement at extended horizons [1]. Selection of optimal horizon therefore depends on the specific operational goal and tolerance for false alerts.

Hybrid modeling strategies that combine short- and long-term forecasts within a single framework represent a promising direction for balancing these trade-offs. Real-time systems appear best suited to shorter horizons where data freshness maximizes reliability. Future research should quantify the marginal operational benefit of incremental lead time against corresponding drops in predictive performance [2]. Resolving these trade-offs is critical for aligning model design with the practical decision timelines of emergency department leaders.

Limitations

Review limitations

Publication bias likely influences the synthesized evidence, as studies reporting positive predictive performance are overrepresented in the peer-reviewed literature relative to those with neutral or negative findings on operational outcomes. Heterogeneity in model architectures, outcome definitions, and reporting standards precluded quantitative meta-analysis, restricting conclusions to narrative synthesis that may overlook subtle patterns. Restriction to English-language publications introduces potential language bias, although the selected databases captured a broad international scope [3]. These methodological constraints should be considered when interpreting the overall strength of recommendations.

The reliance on database searches without exhaustive gray literature inclusion may have omitted relevant implementation reports or ongoing trials not yet indexed. While PRISMA compliance was maintained, subjective elements in risk-of-bias adaptation and thematic grouping introduce a degree of reviewer interpretation. Sensitivity analyses were not performed due to data limitations [4]. Future updates to this review could incorporate expanded search strategies and emerging publication types to address these gaps.

Evidence base limitations

The predominance of single-site, retrospective studies limits the generalizability of findings to diverse emergency department environments and prospective real-world applications. Few investigations measured operational outcomes following actual model deployment, leaving substantial uncertainty regarding causal impact on waiting times, diversion, or left-without-being-seen rates. Dominance of academic centers with mature data infrastructure may not reflect challenges faced by smaller or resource-limited facilities [5]. These evidence base limitations constrain confidence in widespread adoption of current predictive models.

Prospective implementation studies and multisite external validations remain scarce, increasing the risk that models perform differently when transported across settings or time periods. Lack of standardized reporting on alert thresholds and intervention triggers further impedes reproducibility and clinical integration. Cost-effectiveness and unintended consequences such as alert fatigue were almost never evaluated [6]. Strengthening the evidence base through rigorous, forward-looking designs is therefore essential for advancing the field.

Comparison with prior reviews

Prior systematic reviews on emergency department crowding have largely concentrated on traditional metrics such as NEDOCS, EDWIN, and occupancy ratios, with limited exploration of advanced computational forecasting. These earlier syntheses identified key gaps in proactive prediction but stopped short of evaluating machine learning architectures or their integration with real-time operational data [27, 28]. Consequently, they provided foundational insights into crowding measurement yet offered minimal guidance on modern predictive analytics or downstream patient flow impacts.

This review builds directly upon those foundations by restricting its scope to machine learning models published between 2017 and 2022 and by explicitly incorporating input features and operational outcomes. Whereas previous works emphasized descriptive statistics and reactive scoring systems, the present analysis synthesizes classification, regression, time-series, and deep learning approaches across 32 studies [29]. The addition of a dedicated operational outcomes domain distinguishes this synthesis from earlier efforts that focused primarily on predictive accuracy alone.

The novel contribution lies in the systematic linkage of model performance to implementation potential, including the identification of an implementation gap that prior reviews only alluded to indirectly. By contrasting high predictive accuracy with sparse reporting of waiting time reductions or diversion decreases, this work highlights translational shortcomings not previously quantified [30]. This perspective advances the field by shifting emphasis from technical feasibility toward measurable improvements in emergency department performance.

Research gaps

Prospective implementation studies

Prospective implementation studies that progress from silent monitoring to randomized controlled trials of alert-triggered interventions represent a critical gap in the current evidence base. Few investigations have evaluated the causal impact of machine learning predictions on actual operational outcomes such as boarding time or ambulance offload delays [5]. Cost-effectiveness analyses and long-term sustainability assessments are likewise absent, limiting informed decision-making by healthcare systems.

Future work should incorporate human-factors evaluations to quantify clinician trust and response rates to automated alerts. Hybrid designs that combine predictive modeling with implementation science frameworks would accelerate the generation of high-quality evidence [6]. Addressing this gap is essential to move beyond proof-of-concept demonstrations toward scalable solutions for emergency department crowding.

Multi-site external validation

Multi-site external validation remains underrepresented, with most models trained and tested within single emergency departments or homogeneous health systems. This limitation raises concerns about generalizability when models encounter variations in patient demographics, electronic health record infrastructure, or regional arrival patterns [7]. Standardized feature sets and transportability metrics are needed to enable robust cross-setting performance evaluation.

Collaborative data consortia could facilitate the development of federated learning approaches that preserve privacy while improving model robustness. Comparative studies assessing domain adaptation techniques would further clarify the conditions under which predictions remain reliable across geographic and operational contexts [8]. Closing this gap would strengthen confidence in deploying machine learning tools beyond their original development sites.

Actionable prediction thresholds

Actionable prediction thresholds calibrated to clinician-defined crowding levels or intervention triggers are rarely reported despite their importance for practical deployment. Studies seldom specify the exact probability cutoffs or crowding scores that should activate surge protocols or staffing changes, creating ambiguity in real-time use [9]. Adaptive thresholding methods that account for temporal context or departmental capacity could improve decision support relevance.

Research is needed to co-develop thresholds through iterative stakeholder workshops that balance sensitivity, specificity, and operational feasibility. Longitudinal monitoring of threshold performance would enable dynamic recalibration as emergency department workflows evolve [10]. Establishing evidence-based, transparent thresholds would transform predictive outputs into clinically meaningful actions.

Conclusion

This systematic review synthesized machine learning models for emergency department crowding prediction and patient flow optimization published between 2017 and 2022, with particular attention to input features and operational outcomes. The 32 included studies demonstrated consistently high predictive accuracy across classification, regression, time-series, and deep learning approaches, yet operational outcomes such as waiting time, left-without-being-seen rates, and ambulance diversion were reported in only a minority of investigations. Input features centered on occupancy, arrival patterns, and calendar variables, confirming technical feasibility but revealing limited translational progress.

Prediction without corresponding action remains insufficient for meaningful improvement in emergency department performance. The persistent implementation gap, characterized by retrospective designs and sparse linkage to interventions such as surge protocols or staffing adjustments, represents the primary barrier to realizing the potential of these technologies. Real-time systems and prospective evaluations are needed to bridge this divide and deliver measurable benefits.

Operational outcome measurement must become a standard requirement rather than an optional addendum in future predictive analytics research. Journals, funders, and health systems share responsibility for enforcing this expectation to ensure that technical advances translate into tangible gains in patient flow and staff well-being. Without such standards, the field risks continued proliferation of accurate yet underutilized models.

The ultimate vision is one in which predictive analytics are seamlessly integrated into emergency department operations, triggering evidence-based interventions that reduce crowding, shorten waits, and improve both patient and staff experiences. Achieving this integration will require sustained collaboration across research, clinical, administrative, and policy domains. When realized, machine learning-driven patient flow optimization has the capacity to transform emergency care delivery worldwide.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 14 Apr 2022 Revised: 27 Jun 2022 Accepted: 12 Aug 2022
Published online: 20 January 2023

References

Patel SJ, Chamberlain DB, Chamberlain JM. A machine learning approach to predicting need for hospitalization for pediatric asthma exacerbation at the time of emergency department triage. Acad Emerg Med. 2018;25(12):1463-70.

Raita Y, Goto T, Faridi MK, Brown DF, Camargo CA Jr, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care. 2019;23(1):64.

Graham B, Bond R, Quinn M, Mulvenna M. Using data mining to predict hospital admissions from the emergency department. IEEE Access. 2018;6:10458-69.

Mowbray F, Zargoush M, Jones A, De Wit K, Costa A. Predicting hospital admission for older emergency department patients: insights from machine learning. Int J Med Inform. 2020;140:104163.

Roquette BP, Nagano H, Marujo EC, Maiorano AC. Prediction of admission in pediatric emergency department with deep neural networks and triage textual data. Neural Netw. 2020;126:170-7.

Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLoS One. 2017;12(4):e0174708.

Delahanty RJ, Alvarez J, Flynn LM, Sherwin RL, Jones SS. Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. Ann Emerg Med. 2019;73(4):334-44.

Martinez DA, Levin SR, Klein EY, Parikh CR, Menez S, et al. Early prediction of acute kidney injury in the emergency department with machine-learning methods applied to electronic health record data. Ann Emerg Med. 2020;76(4):501-14.

Ramgopal S, Horvat CM, Yanamala N, Alpern ER. Machine learning to predict serious bacterial infections in young febrile infants. Pediatrics. 2020;146(3):e20194096.

Fernandes M, Mendes R, Vieira SM, Leite F, Palos C, et al. Predicting intensive care unit admission among patients presenting to the emergency department using machine learning and natural language processing. PLoS One. 2020;15(3):e0229331.

El-Bouri R, Eyre DW, Watkinson P, Zhu T, Clifton DA. Hospital admission location prediction via deep interpretable networks for the year-round improvement of emergency patient care. IEEE J Biomed Health Inform. 2021;25(1):289-300.

Lee SY, Chinnam RB, Dalkiran E, Krupp S, Nauss M. Prediction of emergency department patient disposition decision for proactive resource allocation for admission. Health Care Manag Sci. 2020;23(3):339-59.

Luo G, Stone BL, Nkoy FL, He S, Johnson MD. Predicting appropriate hospital admission of emergency department patients with bronchiolitis: secondary analysis. JMIR Med Inform. 2019;7(1):e12591.

Desautels T, Das R, Calvert J, Trivedi M, Summers C, et al. Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: a cross-sectional machine learning approach. BMJ Open. 2017;7(9):e017199.

Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting intensive care unit readmission with machine learning using electronic health record data. Ann Am Thorac Soc. 2018;15(7):846-53.

Lin YW, Zhou Y, Faghri F, Shaw MJ, Campbell RH. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. PLoS One. 2019;14(7):e0218942.

Rubin J, Potes C, Xu-Wilson M, Dong J, Rahman A, et al. An ensemble boosting model for predicting transfer to the pediatric intensive care unit. Int J Med Inform. 2018;112:15-20.

Artetxe A, Graña M, Beristain A, Ríos S. Balanced training of a hybrid ensemble method for imbalanced datasets: a case of emergency department readmission prediction. Neural Comput Appl. 2020;32(10):5735-44.

Kalagara S, Eltorai AE, Durand WM, DePasse JM, Daniels AH. Machine learning modeling for predicting hospital readmission following lumbar laminectomy. J Neurosurg Spine. 2019;30(3):344-52.

Min X, Yu B, Wang F. Predictive modeling of the hospital readmission risk from patients’ claims data using machine learning: a case study on COPD. Sci Rep. 2019;9(1):2362.

Lucke JA, de Gelder J, Clarijs F, Heringhaus C, de Craen AJ, et al. Early prediction of hospital admission for emergency department patients: a comparison between patients younger or older than 70 years. Emerg Med J. 2018;35(1):18-27.

Awad A, Bader-El-Den M, McNicholas J, Briggs J. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. Int J Med Inform. 2017;108:185-95.

Bishop JA, Javed HA, El-Bouri R, Zhu T, Taylor T, et al. Improving patient flow during infectious disease outbreaks using machine learning for real-time prediction of patient readiness for discharge. PLoS One. 2021;16(11):e0260476.

Rahman MA, Honan B, Glanville T, Hough P, Walker K. Using data mining to predict emergency department length of stay greater than 4 hours: derivation and single-site validation of a decision tree algorithm. Emerg Med Australas. 2020;32(3):416-21.

Fawagreh K, Gaber MM. Resource-efficient fast prediction in healthcare data analytics: a pruned Random Forest regression approach. Computing. 2020;102(5):1187-98.

Thara T, Thakul O. Application of machine learning to predict the outcome of pediatric traumatic brain injury. Chin J Traumatol. 2021;24(6):350-5.

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.

El-Bouri R, Taylor T, Youssef A, Zhu T, Clifton DA. Machine learning in patient flow: a review. Prog Biomed Eng. 2021;3(2):022002.

Yao LH, Leung KC, Tsai CL, Huang CH, Fu LC. A novel deep learning-based system for triage in the emergency department using electronic medical records: retrospective cohort study. J Med Internet Res. 2021;23(12):e27008.

Ilhan B, Kunt MM, Damarsoy FF, Demir MC, Aksu NM. NEDOCS: is it really useful for detecting emergency department overcrowding today? Medicine (Baltimore). 2020;99(28):e20478.

King Z, Farrington J, Utley M, Kung E, Elkhodair S, et al. Machine learning for real-time aggregated prediction of hospital admission for emergency patients. NPJ Digit Med. 2022;5(1):104.

Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. PLoS One. 2018;13(7):e0201016.