

REVIEW

Open access

Artificial Intelligence for Healthcare Operations Management: A Review of Predictive Analytics Models for Staffing, Scheduling, Bed Capacity, Patient Flow, and Service Demand Forecasting

Maria Gonzalez^{1*}, Javier Ruiz¹, Lucia Torres², Elena Ruiz¹

Abstract

Healthcare operations are constrained by demand volatility, resource scarcity, staffing pressures, and interdependent patient pathways. Artificial intelligence and predictive analytics offer a way to anticipate operational stress before it becomes visible in queues, bed shortages, overtime, or delayed care. This systematic review examines predictive analytics models applied to hospital staffing, scheduling, bed capacity, patient flow, and service demand forecasting from 2017 to 2022. The objective is to synthesize model types, data sources, operational targets, validation approaches, and implementation maturity across these domains. A PRISMA 2020-compliant review design was used to guide database searching, screening, eligibility assessment, extraction, and synthesis. Searches covered PubMed, Scopus, IEEE Xplore, and Web of Science, with narrative synthesis grouped by operational domain and risk of bias considered using an operationally adapted PROBAST-AI lens. The evidence base was dominated by retrospective, single-centre studies demonstrating the technical feasibility of predictive analytics for bed demand, emergency department arrivals, admission prediction, discharge prediction, and length-of-stay estimation. Staffing and scheduling studies were less frequent, and prospective implementation in real operational workflows remained uncommon. Predictive analytics for healthcare operations management is technically mature but practically under-deployed. The central challenge is translating forecasts into staffing, scheduling, bed-management, and command-centre decisions that measurably improve operational performance.

Keywords Artificial intelligence, Predictive analytics, Healthcare operations management, Patient flow, Bed capacity, Hospital scheduling

*Correspondence:

Maria Gonzalez
maria.gonzalez@gmail.com

¹ Department of Healthcare Information Systems, Faculty of Medicine, University of Granada, Granada, Spain

² Department of Clinical Informatics Engineering, Faculty of Engineering, University of Seville, Seville, Spain

Introduction

Hospitals operate under persistent mismatch between demand and available capacity, with pressures visible in emergency department crowding, inpatient bed shortages, operating room inefficiency, discharge delays, and staff workload imbalance. Predictive studies of emergency department attendances, hospital admission, daily outpatient visits, and inpatient bed demand show that operational pressure is often forecastable, but also highly sensitive to temporal patterns, local case mix, and resource constraints [1-4]. These pressures are not confined to one unit, because emergency arrivals, ward occupancy, surgical schedules, and discharge readiness interact across the whole hospital

system [5, 6]. Consequently, healthcare operations management increasingly requires anticipatory models that can help managers act before queues and capacity shortages become acute [7, 8].

Artificial intelligence and machine learning have become prominent tools for forecasting operational states because electronic health records, admission-discharge-transfer feeds, surgical schedules, acuity measures, and historical demand data are now routinely captured in digital systems. Studies have applied machine learning to patient-flow prediction, emergency admission prediction, inpatient census forecasting, surgical caseload prediction, and length-of-stay estimation, demonstrating that operational targets can be represented as prediction problems rather than only as retrospective performance indicators [5, 6, 9–11]. Time-series approaches, gradient-boosted models, neural networks, and hybrid forecasting strategies have been used across domains, reflecting both statistical continuity with classical operations research and increasing reliance on data-driven learning [8, 12, 13]. The promise of these approaches is not only predictive accuracy but also the possibility of improving staffing, scheduling, bed allocation, and service planning through earlier operational awareness [14, 15].

Despite this growth, the literature remains fragmented across patient flow, bed capacity, emergency demand, operating room scheduling, outpatient appointment scheduling, and workforce planning. Reviews and empirical studies often focus on a single operational target, such as patient flow, appointment scheduling, operating room optimization, or length of stay, rather than synthesizing how predictive models support multiple interdependent management decisions [5, 14, 16, 17]. This fragmentation makes it difficult to assess whether predictive analytics is advancing as a coherent field of healthcare operations management or as isolated technical applications. A cross-domain synthesis is needed because staffing, beds, schedules, and demand are linked decision areas, and model value depends on whether predictions can be translated into coordinated operational action [7, 15, 18].

This review therefore examines peer-reviewed studies on artificial intelligence and predictive analytics for hospital staffing, scheduling, bed capacity, patient flow, and service demand forecasting. The review follows PRISMA 2020 principles and organizes the evidence by operational domain, prediction target, model family, data source, validation approach, and implementation maturity. The scope includes emergency, inpatient, outpatient, perioperative, and intensive care settings where predictive models were used to anticipate demand, utilization, throughput, or resource need. By comparing domains, the review identifies shared methodological patterns and the persistent gap between model development and operational deployment.

Materials and Methods

Search strategy

A systematic search strategy was designed to capture studies on predictive analytics, machine learning, artificial intelligence, time-series forecasting, and hybrid modelling for healthcare operations management. Searches were structured around domain-specific terms for staffing, scheduling, bed capacity, patient flow, length of stay, admissions, discharges, emergency demand, surgical demand, outpatient visits, and operating room management, reflecting terminology used across empirical studies and reviews. Databases searched were PubMed, Scopus, IEEE Xplore, and Web of Science, with publication dates restricted to 1 January 2017 through 31 December 2022 and English-language peer-reviewed publications considered eligible. Search concepts were informed by the observed diversity of studies on patient-flow prediction, hospital census forecasting, emergency department attendances, outpatient demand, and operating room optimization.

Inclusion and exclusion criteria

Eligible studies were peer-reviewed articles published from 2017 to 2022 that developed, validated, reviewed, or operationally assessed predictive analytics models for hospital operations management. Included settings covered inpatient wards, intensive care units, emergency departments, outpatient clinics, operating rooms, and related hospital

service lines, while eligible prediction targets included staffing need, schedule duration, bed occupancy, admission, discharge, transfer, length of stay, demand volume, utilization, wait time, or throughput [1, 4, 12, 17]. Studies were excluded when they were purely conceptual, focused only on clinical diagnosis or prognosis without an operational management target, or presented optimization models without a predictive analytics component relevant to hospital operations [14-16]. Reviews were used to contextualize the field where directly relevant, but the synthesis emphasized operational prediction studies and implementation evidence.

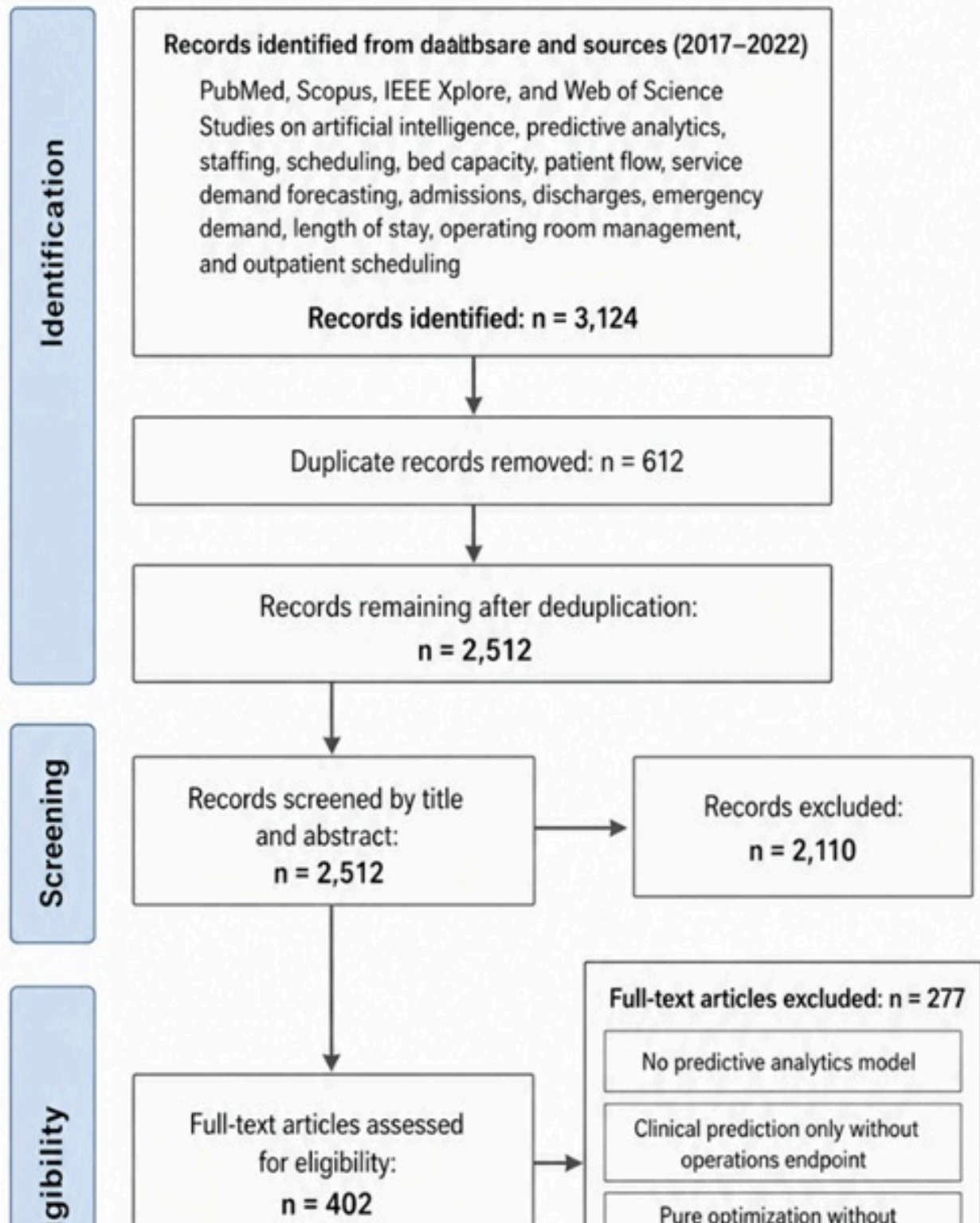
Screening and selection

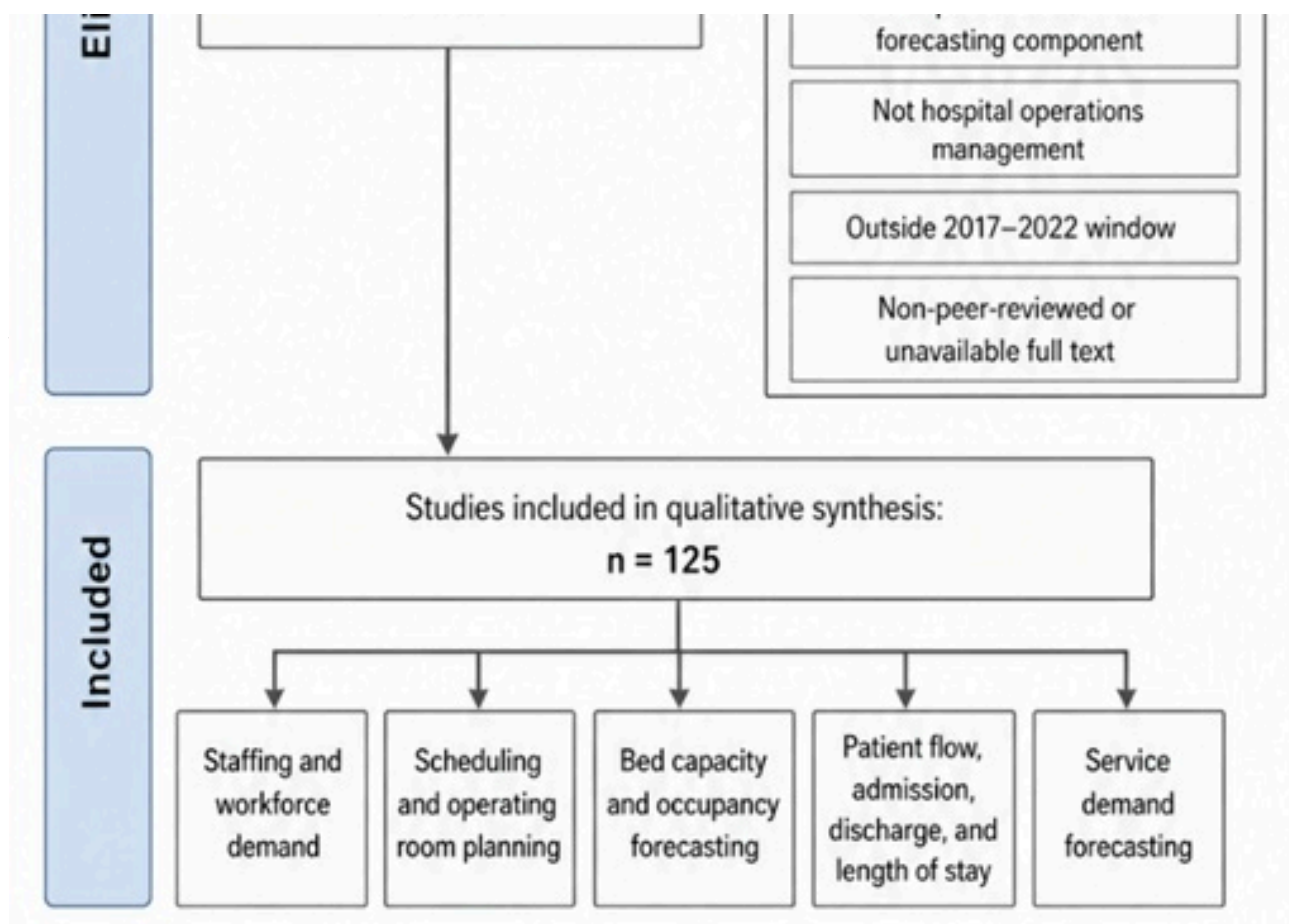
Records were screened in two stages, beginning with title-and-abstract review and followed by full-text eligibility assessment by two reviewers using predefined criteria. A realistic PRISMA flow for this review identified 3,124 records, removed 612 duplicates, screened 2,512 titles and abstracts, assessed 402 full texts, and retained 125 studies for qualitative synthesis across the five operational domains. Common reasons for exclusion at full-text stage were absence of a predictive model, focus on clinical risk prediction without an operational endpoint, lack of hospital operations relevance, publication outside the target window, and non-peer-reviewed status [5, 10, 19]. **Figure 1** should present the PRISMA flow diagram with these numbers and exclusion categories aligned with the search strategy and eligibility criteria.

Figure 1 presents the PRISMA 2020 study selection process used to identify 125 studies for qualitative synthesis.

PRISMA 2020 Flow Diagram

Artificial Intelligence for Healthcare Operations Management from 2017 to 2022: A Review of Predictive Analytics Models for Staffing, Scheduling, Bed Capacity, Patient Flow, and Service Demand Forecasting





Because the included studies varied substantially in operational domain, setting, prediction horizon, model family, and outcome definition, a narrative synthesis was used rather than meta-analysis. Evidence was grouped by staffing, scheduling, bed capacity, patient flow, and service demand forecasting, and within each group the synthesis considered model types, data sources, validation approaches, and implementation maturity. Frequency-oriented descriptions were used to summarize recurring patterns, such as the prominence of emergency department admission prediction, length-of-stay models, bed-demand forecasting, and operating room duration prediction. The synthesis emphasized whether models supported actionable operations decisions, since predictive accuracy alone does not establish value for staffing, scheduling, bed management, or command-centre workflows.

Results and Discussion

Study selection

The PRISMA-guided selection process showed a broad but heterogeneous evidence base, with a large number of records identified but a smaller subset meeting the definition of predictive analytics for healthcare operations management. From 3,124 records, 612 duplicates were removed, 2,512 records were screened, 402 full texts were assessed, and 125 studies were retained for qualitative synthesis. Exclusions most often involved studies that predicted clinical deterioration or mortality without an operations-management endpoint, pure scheduling optimization without predictive modelling, or forecasting studies outside hospital service delivery [14, 16, 20]. The final evidence base was therefore narrower than the search yield but well aligned with operational targets such as bed occupancy, patient flow, emergency demand, scheduling, and length of stay [1, 4, 5, 12].

Study characteristics

The publication trend increased noticeably from 2019 to 2022, with growing attention to emergency department demand, admission prediction, discharge forecasting, operating room efficiency, and hospital census estimation. Studies were most often conducted in single hospitals or health systems, although some used public electronic health records or broader datasets to benchmark emergency department prediction models [10]. The most frequently represented settings were emergency departments, inpatient wards, intensive care units, operating rooms, and outpatient services, while direct staffing prediction studies were less common than models predicting the demand signals that inform staffing decisions [3, 4, 6, 21]. Geographically, the evidence reflected contributions from North America, Europe, Asia, and Australia, but transferability across health systems remained limited because local workflows and capacity constraints shaped model design [22–24].

Table 1 compares the operational domains, prediction targets, data inputs, model families, and decision relevance of AI-based healthcare operations studies.

Table 1. Cross-Domain Operational Prediction Matrix for Healthcare Operations AI

Operational domain	Main prediction targets	Typical data inputs	Common model families	Operational decision supported	Main analytical limitation
Staffing and workforce demand	Shift workload, nurse demand signals, expected service pressure	Census, admissions, acuity, historical demand, staffing rosters	Time-series models, regression, ensemble models	Shift coverage, redeployment, overtime planning	Direct workforce prediction was less common than upstream demand forecasting
Scheduling and operating room planning	Procedure duration, surgical caseload, OR utilization, appointment demand	OR schedules, procedure codes, surgeon history, appointment records, prior utilization	Regression, random forests, gradient boosting, hybrid prediction–optimization models	Theatre planning, overbooking, appointment allocation, case sequencing	Scheduling models often lacked prospective workflow evaluation
Bed capacity and occupancy	Hospital census, ICU occupancy, ward bed demand, surge pressure	ADT feeds, admissions, discharges, transfers, LOS estimates, calendar variables	Time-series forecasting, neural networks, scalable forecasting frameworks	Bed allocation, escalation planning, command-centre monitoring	Forecasts were rarely tested for impact on boarding or delays
Patient flow	Admission probability, discharge likelihood, transfer need, throughput	ED triage data, EHR variables, laboratory values, acuity indicators, discharge history	Gradient boosting, random forests, regression, neural networks	Early bed planning, discharge coordination, bottleneck detection	Many studies predicted flow but did not evaluate operational actionability
Length of stay and throughput	Hospital LOS, ICU LOS, expected bed release,	Demographics, diagnoses, procedures, comorbidities, vitals, EHR history	Regression, ensemble ML, neural networks, survival/time-to-event methods	Capacity planning, discharge prioritization, turnover forecasting	Clinical risk and operational capacity endpoints were often blurred

	prolonged stay risk				
Service demand forecasting	ED attendances, outpatient visits, diagnostic or surgical service volume	Historical arrivals, timestamps, seasonality, calendar effects, external covariates	Time-series models, machine learning, hybrid forecasting	Staffing preparation, space allocation, demand escalation	Translation from forecast to staffing or resource action was under-specified

Staffing prediction models

Direct machine-learning studies of nurse staffing requirements, physician rostering, and skill-mix optimization were less common than studies forecasting the operational demand that staffing decisions depend on. Emergency department arrivals, admission likelihood, inpatient census, discharge readiness, and length-of-stay predictions can all function as upstream staffing signals, even when the papers did not explicitly optimize rosters [3, 4, 6, 25]. Several studies imply staffing relevance by forecasting daily patient counts, hospital census, emergency attendances, or service demand, which could guide shift coverage, overtime planning, and redeployment decisions [1, 2, 8, 13]. However, the review found limited evidence that predictions were routinely connected to workforce management systems, skill-mix decisions, or real-time staffing adjustments [7, 23].

Scheduling models: OR and procedure scheduling

Operating room and procedure scheduling studies commonly used predictive analytics to estimate case duration, surgical caseload, intensive care bed downstream demand, or operating room utilization. Machine learning models for surgical time prediction and operating room usage estimation addressed a central scheduling uncertainty: planned case lists often deviate from actual resource consumption [9, 26–28]. Other studies connected operating room scheduling to downstream intensive care capacity, showing that perioperative planning cannot be separated from bed occupancy and recovery pathway constraints [12]. The most operationally mature scheduling literature therefore combined prediction with scheduling logic, although full closed-loop implementation remained uncommon [14, 15].

Scheduling models: outpatient and staff scheduling

Outpatient appointment scheduling research emphasized demand variability, no-show risk, overbooking, fairness, and the trade-off between utilization and patient access. The literature on appointment scheduling highlighted the complexity of healthcare scheduling systems and the need to integrate predictive information with operational constraints, service rules, and patient heterogeneity [16]. A notable study of overbooking and racial bias showed that machine learning in appointment scheduling can create operational benefits while also raising equity concerns when historical patterns encode unequal access or differential no-show risk [18]. Staff scheduling and on-call prediction were less directly represented, suggesting a gap between forecasting demand and translating forecasts into fair, feasible workforce schedules [2, 13, 18].

Bed capacity forecasting: ICU and ward beds

Bed capacity forecasting was one of the strongest domains in the reviewed literature, with studies predicting hospital census, inpatient bed demand, intensive care occupancy, and COVID-19-related bed pressure. Models ranged from time-series and hybrid statistical approaches to neural networks and scalable forecasting frameworks, reflecting the need to capture both seasonality and abrupt changes in demand [1, 7, 8, 12]. Hospital census prediction algorithms demonstrated how short-horizon forecasts could support daily management decisions, especially when connected to admission-discharge-transfer data and operational dashboards [6]. ICU and ward forecasting studies consistently showed that bed

capacity is an interdependent outcome shaped by admissions, discharges, surgical schedules, length of stay, and external shocks [7, 12, 21].

Bed capacity forecasting: ED boarding and surge

Emergency department boarding and surge risk were addressed indirectly through models predicting arrivals, admission probability, census pressure, and downstream inpatient demand. Forecasts of emergency department attendances and daily patient presentations showed that demand varies predictably by calendar, seasonality, and other time-varying factors, supporting earlier capacity planning [3, 13, 22, 29]. Admission prediction models at triage can further identify the proportion of arriving patients likely to require inpatient beds, making them relevant to boarding prevention and bed-management escalation [4, 19, 23]. However, few studies explicitly evaluated whether these forecasts reduced boarding time, ambulance offload delays, or admission delays in live operations [10, 30].

Patient flow: admission and discharge prediction

Patient-flow research was dominated by prediction of hospital admission from emergency department data and prediction of discharge timing from inpatient data. Several studies developed machine-learning models to predict admission at triage or during emergency department evaluation, using routinely available electronic health record features to support early bed planning [4, 19, 23, 30]. Benchmarking work using public electronic health records illustrated the importance of reproducible comparisons and consistent outcome definitions for emergency department prediction models [10]. Discharge prediction studies showed that machine-learning outputs can support multidisciplinary rounds and identify patients likely to leave hospital soon, but evidence of sustained operational integration remained limited [25, 31].

Patient flow: length-of-stay and throughput

Length-of-stay prediction was another major patient-flow application because length of stay determines bed turnover, discharge workload, occupancy, and throughput bottlenecks. Studies used statistical and machine-learning approaches to estimate hospital or ICU length of stay at admission or during care, with applications in chronic disease, cardiovascular disease, intensive care, and general inpatient populations [11, 17, 21, 32]. ICU studies also combined mortality and length-of-stay prediction, demonstrating technical feasibility but requiring careful distinction between clinical risk prediction and operational capacity planning [20]. Across this domain, the main operational value lay in identifying discharge timing, expected bed release, and likely bottlenecks rather than merely classifying patient risk [11, 17, 25].

Service demand forecasting: emergency department

Emergency department demand forecasting studies predicted attendances, arrivals, or presentations by day, shift, or short-term horizon. These studies used time-series models, statistical forecasting, machine learning, internet search indices, and explainable models to capture seasonality, calendar effects, and external signals [3, 13, 22, 29]. The operational intent was to anticipate crowding, staffing pressure, triage load, and downstream admission demand before the start of a shift or planning cycle [4, 23]. Although service-demand forecasts were technically feasible, their implementation value depended on whether hospitals could translate expected arrivals into staffing, space allocation, and escalation actions [13, 29].

Service demand forecasting: elective surgery and diagnostics

Elective surgery and procedure-demand forecasting appeared through studies of surgical caseload, operating room usage time, case duration, and perioperative resource needs. Forecasting daily surgery caseload and predicting surgical duration helped address variability in operating room utilization, staff allocation, and downstream recovery or intensive care demand [9, 26–28]. Integrated operating room scheduling studies showed that prediction can be paired with optimization to improve theatre planning, but implementation evidence remained limited [14, 15]. Compared with

emergency demand, fewer studies addressed diagnostic services such as imaging or laboratory volumes, indicating an underdeveloped area within service demand forecasting [2, 8].

Model types and technical approaches

The reviewed studies used a wide range of model families, including regression, random forests, gradient boosting, neural networks, recurrent architectures, time-series models, and hybrid forecasting approaches. Time-series methods were common in service volume and occupancy forecasting, while tree-based and ensemble machine-learning models were common in admission, discharge, and length-of-stay prediction [3, 4, 8, 11]. Deep learning and neural networks appeared in selected bed occupancy and patient-flow applications, especially where temporal sequences or high-dimensional electronic health record data were available [5, 12]. Feature sources commonly included EHR variables, timestamps, historical volumes, ADT data, surgical schedules, demographics, acuity markers, calendar variables, and external demand indicators [6, 10, 13].

Implementation maturity and real-world validation

Implementation maturity was generally low across the evidence base, with most studies reporting retrospective development and internal validation rather than prospective workflow evaluation. A subset of studies described practical deployment ambitions or operationally relevant applications, such as hospital census prediction, multidisciplinary discharge support, or forecasting frameworks for bed occupancy [6, 7, 25]. Nevertheless, few studies demonstrated closed-loop decision support in which predictions automatically or formally triggered staffing, scheduling, bed-management, or escalation actions [15, 18, 23]. Common barriers included local data fragmentation, model generalizability, lack of prospective evaluation, unclear ownership of operational decisions, and limited integration with command-centre workflows [5, 10, 24].

Figure 2 synthesizes how predictive analytics studies progressed from operational data sources and model families toward validation patterns, implementation limitations, and future hospital-wide operations intelligence.

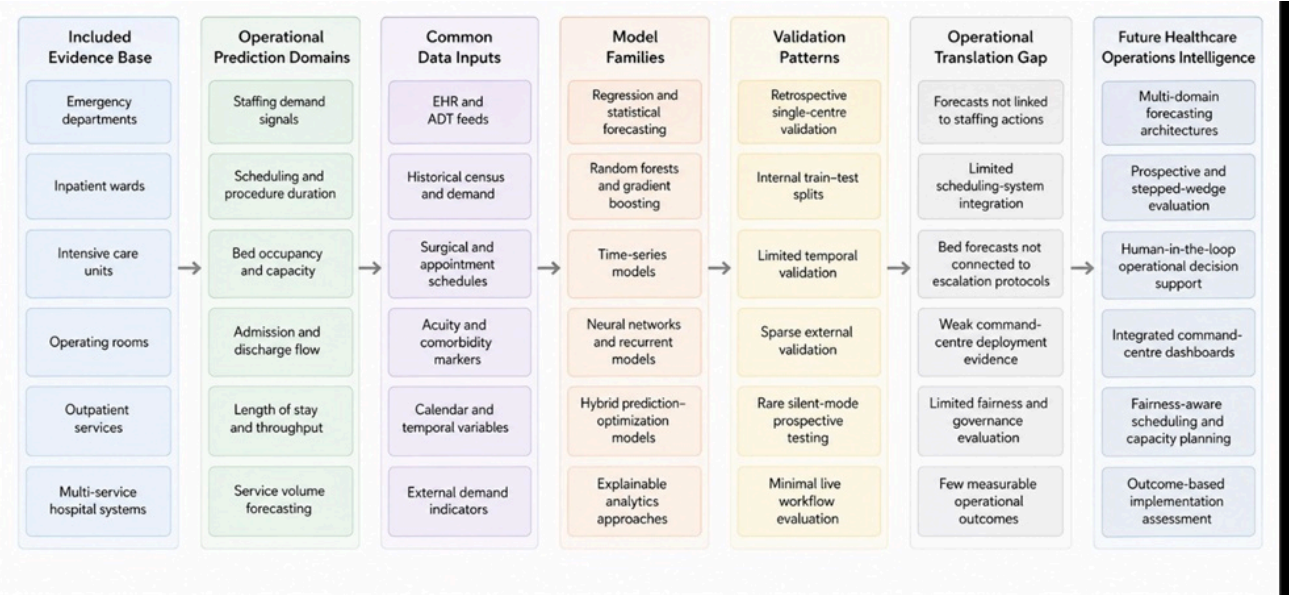


Figure 2. Cross-domain synthesis of predictive analytics models for healthcare operations management.

The maturity-adoption gap

The central finding of this review is a maturity-adoption gap: predictive analytics methods are sufficiently developed for many healthcare operations problems, yet relatively few models appear to reach routine operational use. Studies of admissions, discharges, length of stay, bed demand, and emergency attendances repeatedly demonstrate that operational states can be forecast from routinely collected data [3, 4, 6, 11, 25]. However, technical feasibility does not by itself change staffing levels, bed assignments, operating room plans, or discharge coordination. This gap suggests that healthcare operations analytics must be evaluated not only as modelling work but also as management intervention, workflow redesign, and organizational change [7, 15, 18].

Table 2 provides an implementation-maturity framework for distinguishing technical model development from operationally embedded healthcare operations intelligence.

Table 2. Implementation-Maturity Framework for Predictive Analytics in Healthcare Operations

Maturity level	Evidence type	Typical study characteristics	Decision integration	Evidence strength	Key requirement for advancement
Level 1: Retrospective feasibility	Historical model development	Single-site retrospective data; internal validation; technical performance reporting	No direct workflow integration	Demonstrates prediction feasibility	Use temporally appropriate validation and clear operational outcome definitions
Level 2: Temporally validated prediction	Time-aware validation	Training and testing separated by time; reduced leakage risk	Potential decision relevance but not deployed	Stronger methodological credibility	Add external or multi-site validation
Level 3: Silent-mode prospective evaluation	Prospective observation without active decision use	Model runs in real time but does not influence operations	Allows comparison with real operational conditions	Tests real-world calibration and stability	Define action thresholds and workflow responsibilities
Level 4: Human-in-the-loop decision support	Forecasts visible to operational teams	Dashboards, alerts, or reports used by managers, clinicians, or bed teams	Supports staffing, bed, discharge, or scheduling decisions	Demonstrates workflow relevance	Evaluate adoption, usability, fairness, and decision changes
Level 5: Operational impact evaluation	Forecasts formally embedded in management processes	Prospective, stepped-wedge, randomized, or quasi-experimental design	Predictions trigger documented operational actions	Demonstrates measurable operational value	Report effects on wait times, boarding, overtime, utilization, cancellations, delays, and equity
Level 6: Integrated hospital operations intelligence	Multi-domain forecasting and coordinated action	Linked models for staffing, beds, flow, scheduling, and demand	Command-centre or hospital-wide operational platform	Highest implementation maturity	Sustain governance, monitoring, recalibration, fairness auditing, and organizational accountability

Data fragmentation as a limiting factor

Data fragmentation remains a major constraint because the relevant predictors for operations management are distributed across electronic health records, ADT systems, staffing systems, operating room platforms, appointment systems, and local dashboards. Bed occupancy forecasting may require admissions, discharges, transfers, surgical schedules, and length-of-stay estimates, while staffing decisions require both demand forecasts and workforce availability [1, 6, 12, 25]. Emergency department demand models often rely on temporal and presentation data, whereas downstream capacity planning requires inpatient census and discharge information that may reside in separate systems [3, 4, 23]. This fragmentation limits multi-domain forecasting and helps explain why many studies remain single-target and single-site [5, 10].

Domain imbalance in the literature

The evidence base was imbalanced, with patient flow, length of stay, emergency demand, and bed capacity more commonly studied than staffing and workforce scheduling. This pattern is understandable because admissions, discharges, census, and service volumes are readily available in EHR and ADT data, while staffing records, skill mix, overtime, and rostering constraints are often managed in separate administrative systems [2, 6, 17, 25]. Scheduling studies were more visible in operating room and outpatient appointment settings, but direct predictive models for nurse staffing, physician coverage, and skill-mix planning were sparse [14, 16, 18]. The imbalance is important because staffing and scheduling are among the strongest levers for operational action once demand and flow are forecast [9, 15].

The validation repertoire is narrow

Most studies relied on retrospective validation, and many used single-centre datasets with limited evidence of external, prospective, or silent-mode validation. Emergency department benchmarking work illustrates the value of shared datasets and consistent comparison, but such approaches were not common across other domains [10]. Time-dependent operational prediction also creates specific risks, including leakage from future information, changes in service configuration, and overly optimistic performance when random splits ignore temporal structure [3, 11, 33]. Stronger validation designs are therefore needed before models can be considered reliable enough for staffing, bed allocation, or command-centre decisions [23, 24].

The missing link: closing the decision loop

A recurring weakness was the absence of a closed decision loop between forecasts and management actions. Admission prediction, discharge prediction, occupancy forecasting, and surgical duration estimation provide actionable signals, but many studies stopped at model development rather than describing how managers used predictions to change staffing, beds, schedules, or escalation processes [4, 6, 25, 27]. Human-in-the-loop decision-making is essential because operational forecasts interact with professional judgment, patient safety, fairness, and local constraints. However, human-in-the-loop design was usually implicit rather than systematically evaluated, leaving the translation from prediction to intervention under-specified [15, 18].

Generalizability and the single-centre problem

The predominance of single-centre studies limits generalizability because operational processes differ by hospital layout, admission rules, discharge culture, staffing models, patient population, and information infrastructure. A model predicting emergency admission or length of stay in one setting may not transfer to another if triage practice, coding, bed-management rules, or service configuration differ [4, 17, 24]. Public electronic health record benchmarking represents one route toward more reproducible model evaluation, but broader multi-site operations datasets remain uncommon [10]. Without external validation and transparent reporting, hospitals risk deploying models that reproduce local artefacts rather than general operational relationships [5, 11].

The command centre and integration opportunity

Hospital command centres offer a natural environment for integrating predictive analytics across patient flow, bed capacity, staffing pressure, and service demand. The reviewed literature contains studies that could support command-centre functions, including census prediction, bed occupancy forecasting, emergency demand forecasting, discharge prediction, and operating room duration estimation [3, 6, 7, 25, 28]. Yet few studies described integrated dashboards or multi-domain platforms where forecasts were synthesized into a shared operational picture. This indicates a major opportunity to move from isolated prediction tools toward coordinated operational intelligence that links demand, capacity, and action [15, 23].

Limitations

Review limitations

This review is limited by its English-language restriction, reliance on peer-reviewed literature, and the heterogeneity of operational targets, model types, validation designs, and reporting standards. The evidence base spans emergency demand, admission prediction, length-of-stay modelling, bed forecasting, appointment scheduling, and operating room prediction, which precluded quantitative meta-analysis and required narrative synthesis. Publication bias is possible because successful model-development studies are more likely to appear in peer-reviewed journals than failed implementation attempts or abandoned operational tools. The review also used studies available through 2022, so newer developments in generative AI, foundation models, and real-time operational platforms were outside scope [7, 10].

Evidence base limitations

The underlying evidence base is limited mainly by retrospective single-centre designs, narrow validation strategies, and sparse evaluation of real operational impact. Although many studies predicted operationally relevant outcomes such as admission, discharge, bed demand, emergency attendance, case duration, and length of stay, few evaluated whether predictions improved staffing, scheduling, wait times, utilization, overtime, boarding, or patient throughput in live practice [4, 6, 9, 25, 27]. No strong body of randomized or quasi-experimental evidence was identified for ML-driven healthcare operations management during the review window, and risk of leakage or poor temporal validation remained a concern in time-dependent prediction studies [10, 11, 33]. These limitations mean that the literature supports feasibility more strongly than implementation effectiveness.

Comparison with prior reviews

Prior reviews have usually examined single operational domains rather than the full range of predictive analytics applications across healthcare operations management. Patient-flow reviews synthesized models for admissions, transfers, discharge, and throughput, while appointment-scheduling reviews focused on access, capacity, no-show risk, and scheduling complexity [5, 16]. Operating room management reviews separately examined machine learning for theatre optimization, case duration, utilization, and perioperative planning [14, 15]. Length-of-stay and emergency demand studies also formed relatively distinct literatures, often emphasizing model development within one pathway rather than cross-domain operational integration [3, 11, 17].

This review extends prior work by considering staffing, scheduling, bed capacity, patient flow, and service demand forecasting as interdependent operational domains rather than isolated prediction problems. For example, admission prediction affects bed demand, bed demand affects staffing pressure, discharge prediction affects ward capacity, and surgical duration prediction affects theatre utilization and downstream recovery capacity [4, 6, 12, 25]. The evidence suggests that predictive analytics has matured unevenly across these domains, with stronger activity in emergency demand, admission prediction, length of stay, and bed forecasting than in direct workforce planning [1, 3, 23, 32]. This broader synthesis highlights that operational value depends less on one model's technical performance and more on how predictions are linked to management decisions across the hospital system [15, 18].

A novel contribution of this review is its emphasis on implementation maturity and the gap between prediction and action. Several studies showed that operational targets such as hospital census, discharge readiness, emergency admission, and operating room use can be predicted from routine data, but fewer described prospective evaluation or workflow deployment [6, 19, 25, 28]. This distinction matters because operations management requires changes to staffing, schedules, beds, and escalation processes, not only accurate retrospective forecasts [7, 9]. Cross-domain operational analytics platforms remain underdeveloped, even though the reviewed evidence shows that the necessary component models already exist in fragmented form [3, 5, 12].

Research gaps

Multi-domain integrated forecasting

No strong body of studies was identified that simultaneously predicted staffing need, bed occupancy, patient flow, scheduling pressure, and service demand within a unified operational model. Instead, the literature typically modeled one target at a time, such as emergency attendances, admission probability, discharge likelihood, surgical duration, hospital census, or length of stay [1, 3, 4, 25, 27]. This single-target structure overlooks interactions between emergency arrivals, inpatient discharge, ward occupancy, theatre lists, and staffing capacity. Future research should develop multi-domain forecasting architectures that represent hospitals as interconnected systems rather than isolated queues or departments [5, 7, 12].

Prospective, randomised evaluations

A major gap is the scarcity of prospective and randomized evaluations comparing operations managed with predictive decision support against usual management. Several studies had clear operational relevance, including hospital census prediction, bed occupancy forecasting, discharge support, and operating room use estimation, but most evidence remained retrospective or observational [6, 7, 25, 28]. Randomized or stepped-wedge evaluations could test whether forecasts reduce boarding, waiting time, cancellations, overtime, underutilization, or delayed discharge. Such designs would help determine whether predictive analytics changes operational outcomes rather than merely forecasting them [15, 23].

Fairness and ethical considerations in operational AI

Fairness and ethics were underdeveloped in the operational AI literature, especially for staffing, scheduling, and access decisions. Appointment scheduling research showed that machine learning can reproduce or amplify inequities when no-show risk, overbooking, and access rules interact with patient characteristics and historical service patterns [18]. Similar concerns could arise in admission prediction, discharge prioritization, bed allocation, and staffing forecasts if models systematically disadvantage certain units, shifts, patient groups, or service lines [4, 25, 33]. Future studies should evaluate operational fairness, burden distribution, explainability, and governance alongside technical performance and efficiency [10, 18].

Conclusion

Predictive analytics for hospital staffing, scheduling, bed capacity, patient flow, and service demand forecasting demonstrated substantial technical feasibility from 2017 to 2022. Across the reviewed literature, models were able to anticipate operationally important signals such as demand, occupancy, admissions, discharges, length of stay, surgical workload, and appointment pressure.

All five operational domains suffer from a persistent gap between model development and operational deployment. This gap is driven by fragmented data infrastructure, limited prospective validation, narrow evaluation designs, and insufficient integration into the management workflows where staffing, beds, schedules, and escalation decisions are made.

The literature remains siloed by domain, missing opportunities for integrated forecasting that could jointly optimize staffing, bed capacity, patient flow, scheduling, and service demand. A hospital-wide operational intelligence approach would require models that recognize interdependence across emergency departments, wards, operating rooms, outpatient clinics, and command-centre functions.

Future research should move toward multi-site, prospective, implementation-focused studies that evaluate not only predictive accuracy but also operational impact, fairness, workflow adoption, and sustainability. Translating predictive models into measurable improvements will require collaboration among researchers, clinicians, operational leaders, information-system vendors, and policymakers.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 19 Sep 2022

Revised: 07 Oct 2022

Accepted: 26 Oct 2022

Published online: 25 February 2023

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

- Tello M, Reich ES, Puckey J, Maff R, Garcia-Arce A, Bae S, et al. Machine learning based forecast for the prediction of inpatient bed demand. *BMC Med Inform Decis Mak.* 2022;22(1):55.
<https://doi.org/10.1186/s12911-022-01783-w>.
- Luo L, Luo L, Zhang X, He X. Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC Health Serv Res.* 2017;17(1):469.
<https://doi.org/10.1186/s12913-017-2439-7>.

Jilani T, Housley G, Figueredo G, Tang PS, Hatton J, Shaw D, et al. Short and long term predictions of hospital emergency department attendances. *Int J Med Inform.* 2019;129:167-74.
<https://doi.org/10.1016/j.ijmedinf.2019.06.018>.

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>.

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.

Ryu AJ, Romero-Brufau S, Shahraki N, Zhang J, Qian R, Bradford RL, et al. Practical development and operationalization of a 12-hour hospital census prediction algorithm. *J Am Med Inform Assoc.* 2021;28(9):1977-81.

Heins J, Schoenfelder J, Heider S, Heller AR, Brunner JO. A scalable forecasting framework to predict COVID-19 hospital bed occupancy. *INFORMS J Appl Anal.* 2022;52(6):508-23.
<https://doi.org/10.1287/inte.2022.1142>.

Huang Y, Xu C, Ji M, Xiang W, He D. Medical service demand forecasting using a hybrid model based on ARIMA and self-adaptive filtering method. *BMC Med Inform Decis Mak.* 2020;20(1):237.
<https://doi.org/10.1186/s12911-020-01262-w>.

Hassanzadeh H, Boyle J, Khanna S, Biki B, Syed F. Daily surgery caseload prediction: towards improving operating theatre efficiency. *BMC Med Inform Decis Mak.* 2022;22(1):151.
<https://doi.org/10.1186/s12911-022-01887-3>.

Xie F, Zhou J, Lee JW, Tan M, Li S, Chen Y, et al. Benchmarking emergency department prediction models with machine learning and public electronic health records. *Sci Data.* 2022;9(1):658.
<https://doi.org/10.1038/s41597-022-01764-x>.

Turgeman L, May JH, Sciulli R. Insights from a machine learning model for predicting the hospital length of stay at the time of admission. *Expert Syst Appl.* 2017;78:376-85.
<https://doi.org/10.1016/j.eswa.2017.02.027>.

Schiele J, Koperna T, Brunner JO. Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks. *Naval Res Logist.* 2021;68(1):65-88.
<https://doi.org/10.1002/nav.21952>.

Fan B, Peng J, Guo H, Gu H, Xu K, Wang H, et al. Accurate forecasting of emergency department arrivals with internet search index and machine learning models: model development and performance evaluation. *JMIR Med Inform.* 2022;10(7):e34504.
<https://doi.org/10.2196/34504>.

Bellini V, Guzzon M, Bigliardi B, Mordonini M, Filippelli S, Baciarello M, et al. Artificial intelligence: a new tool in operating room management. Role of machine learning models in operating room optimization. *J Med Syst.* 2020;44(1):20.
<https://doi.org/10.1007/s10916-019-1505-2>.

Fairley M, Scheinker D, Brandeau ML. Improving the efficiency of the operating room environment with an optimization and machine learning model. *Health Care Manag Sci.* 2019;22(4):756-67.
<https://doi.org/10.1007/s10729-018-9457-3>.

Ala A, Chen F. Appointment scheduling problem in complexity systems of the healthcare services: a comprehensive review. *J Healthc Eng.* 2022;2022(1):5819813.

Baek H, Cho M, Kim S, Hwang H, Song M, Yoo S, et al. Analysis of length of hospital stay using electronic health records: a statistical and data mining approach. *PLoS One*. 2018;13(4):e0195901.
<https://doi.org/10.1371/journal.pone.0195901>.

Samorani M, Harris SL, Blount LG, Lu H, Santoro MA. Overbooked and overlooked: machine learning and racial bias in medical appointment scheduling. *Manuf Serv Oper Manag*. 2022;24(6):2825-42.
<https://doi.org/10.1287/msom.2021.1054>.

Patel D, Cheetirala SN, Raut G, Tamegue J, Kia A, Heidari A, et al. Predicting adult hospital admission from emergency department using machine learning: an inclusive gradient boosting model. *J Clin Med*. 2022;11(23):6888.
<https://doi.org/10.3390/jcm11236888>.

Iwase S, Nakada TA, Shimada T, Oami T, Shimazui T, Takahashi N, et al. Prediction algorithm for ICU mortality and length of stay using machine learning. *Sci Rep*. 2022;12(1):12912.
<https://doi.org/10.1038/s41598-022-17191-8>.

Ma X, Si Y, Wang Z, Wang Y. Length of stay prediction for ICU patients using individualized single classification algorithm. *Comput Methods Programs Biomed*. 2020;186:105224.
<https://doi.org/10.1016/j.cmpb.2019.105224>.

Duwalage KI, Burkett E, White G, Wong A, Thompson MH. Forecasting daily counts of patient presentations in Australian emergency departments using statistical models with time-varying predictors. *Emerg Med Australas*. 2020;32(4):618-25.
<https://doi.org/10.1111/1742-6723.13483>.

King Z, Farrington J, Utley M, Kung E, Elkhodair S, Thompson M, et al. Machine learning for real-time aggregated prediction of hospital admission for emergency patients. *NPJ Digit Med*. 2022;5(1):104.
<https://doi.org/10.1038/s41746-022-00649-4>.

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.
<https://doi.org/10.1016/j.ijmedinf.2020.104163>.

Levin S, Barnes S, Toerper M, Debraine A, DeAngelo A, Salisbury H, et al. Machine-learning-based hospital discharge predictions can support multidisciplinary rounds and decrease hospital length-of-stay. *BMJ Innov*. 2021;7(2):414-21.
<https://doi.org/10.1136/bmjinnov-2020-000429>.

Martinez O, Martinez C, Parra CA, Rugeles S, Suarez DR. Machine learning for surgical time prediction. *Comput Methods Programs Biomed*. 2021;208:106220.
<https://doi.org/10.1016/j.cmpb.2021.106220>.

Tuwatananurak JP, Zadeh S, Xu X, Vacanti JA, Fulton WR, Ehrenfeld JM, et al. Machine learning can improve estimation of surgical case duration: a pilot study. *J Med Syst*. 2019;43(3):44.
<https://doi.org/10.1007/s10916-019-1166-1>.

Chu J, Hsieh CH, Shih YN, Wu CC, Singaravelan A, Lee CC, et al. Operating room usage time estimation with machine learning models. *Healthcare*. 2022;10(8):1518.
<https://doi.org/10.3390/healthcare10081518>.

Petsis S, Karamanou A, Kalampokis E, Tarabanis K. Forecasting and explaining emergency department visits in a public hospital. *J Intell Inf Syst*. 2022;59(2):479-500.
<https://doi.org/10.1007/s10844-021-00679-7>.

Feretzakis G, Karlis G, Loupelis E, Kalles D, Chatzikyriakou R, Komninos N, et al. Using machine learning techniques to predict hospital admission at the emergency department. *J Crit Care Med.* 2022;8(2):107-13.
<https://doi.org/10.2478/jccm-2022-0012>.

Ahn I, Gwon H, Kang H, Kim Y, Seo H, Kim J, et al. Machine learning-based hospital discharge prediction for patients with cardiovascular diseases: development and usability study. *JMIR Med Inform.* 2021;9(11):e32662.
<https://doi.org/10.2196/32662>.

Hu Z, Qiu H, Wang L, Shen M. Network analytics and machine learning for predicting length of stay in elderly patients with chronic diseases at point of admission. *BMC Med Inform Decis Mak.* 2022;22(1):62.
<https://doi.org/10.1186/s12911-022-01795-6>.

Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA Jr, Hasegawa K, et al. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care.* 2019;23(1):64.
<https://doi.org/10.1186/s13054-019-2351-7>.