

REVIEW

Open access

Machine Learning for Hospital Length-of-Stay Prediction: A Systematic Review of Electronic Health Record Features, Model Architectures, Validation Methods, and Operational Implementation Outcomes

Hiroshi Nakamura^{1*}, Yuta Kato¹

Abstract

Hospital length-of-stay is a central operational metric for inpatient capacity planning, discharge coordination, and resource allocation. Accurate prediction remains difficult because patient trajectories are heterogeneous, nonlinear, and shaped by evolving clinical events during admission. Traditional statistical models often have limited flexibility for high-dimensional and sequential electronic health record data. Across the literature, there is no settled consensus regarding optimal model architecture, feature representation, validation design, or clinical implementation strategy. This systematic review synthesizes machine learning approaches for hospital length-of-stay prediction published from 2017 to 2022. It focuses on EHR feature types, model architectures, validation methods, interpretability strategies, and reported operational outcomes. A structured review of peer-reviewed literature was conducted using targeted search strings related to machine learning, deep learning, electronic health records, discharge prediction, and hospital length-of-stay. The review included studies across emergency, inpatient, surgical, pediatric, cardiovascular, and intensive care settings. The literature suggests that gradient boosting, random forest, ensemble learning, and recurrent neural networks are common approaches for LOS prediction. However, external validation remains uncommon, prediction horizons vary widely, and operational implementation outcomes are reported less consistently than model development results. Future research should prioritize external validation, prospective implementation studies, standardized outcome definitions, and transparent reporting of workflow barriers. Shared benchmarking datasets and multi-center validation consortia would strengthen comparability across LOS prediction studies.

Keywords Machine learning, Electronic health records, Predictive modeling, Clinical informatics, Hospital length-of-stay, Discharge planning

*Correspondence:

Hiroshi Nakamura

hiroshi.nakamura@outlook.com

¹ Department of Smart Healthcare Informatics, Faculty of Medicine, Nagoya University, Nagoya, Japan

Introduction

Hospital length-of-stay prediction is important because inpatient duration influences bed occupancy, discharge planning, patient flow, staffing, and institutional costs. Over-prediction may reserve capacity unnecessarily, whereas under-prediction can contribute to discharge bottlenecks and disrupted patient flow. Systematic and methodological reviews have therefore framed LOS prediction as both a clinical informatics problem and an operational management problem [1, 2]. Studies across general medicine, cardiovascular care, intensive care, and surgery have reported that prediction models are increasingly used to support planning rather than to replace clinical judgment [3-6].

Earlier approaches to LOS prediction often relied on regression models, administrative risk adjustment, or severity scores, but these strategies may not adequately represent individualized clinical trajectories. In critical care, studies have compared machine learning models with severity-based approaches for outcomes such as mortality and prolonged ICU stay, while highlighting the limits of static admission information [7-9]. Traditional models remain useful as interpretable baselines, but the literature suggests that they can struggle when LOS depends on nonlinear combinations of comorbidity, laboratory trends, interventions, and discharge constraints [4, 10, 11]. This motivates interest in flexible models that can incorporate broader EHR-derived feature spaces.

Machine learning approaches offer a way to represent nonlinear associations, interactions among clinical variables, and sequential patterns in EHR data. Investigations using neural networks, recurrent architectures, hidden Markov models, and mixed-data deep learning have explored how temporal trajectories may inform LOS prediction during hospitalization [12-15]. Tree-based models and ensemble methods have also been widely investigated because they can handle heterogeneous structured variables and provide feature-importance summaries [16-19]. However, the field remains methodologically heterogeneous, and evidence is fragmented across specialties, data sources, and prediction targets.

This review examines studies on machine learning for hospital LOS prediction using structured and unstructured EHR data. It synthesizes feature types, preprocessing methods, model architectures, validation approaches, interpretability techniques, and operational implementation outcomes. The scope includes ICU, emergency, medical, surgical, pediatric, cardiovascular, respiratory, oncology, and orthopedic settings because these domains reveal different challenges in LOS definition and deployment. The review is critical and conceptual, avoiding pooled performance claims because heterogeneity in outcomes and validation designs limits direct comparison.

Background

Clinical and operational importance of LOS prediction

LOS prediction is clinically important because expected discharge timing affects care coordination, bed allocation, and downstream patient movement across emergency departments, wards, and intensive care units. The literature links LOS modeling to discharge planning, multidisciplinary rounds, and resource allocation, especially when predictions are intended to support operational decisions rather than merely describe risk [4, 20]. In specialty settings such as cardiology, orthopedics, pediatrics, and ICU care, studies have reported that LOS forecasts may help anticipate care complexity and post-acute needs [4, 6, 21, 22]. Nevertheless, operational usefulness depends on whether predictions are timely, understandable, and integrated into clinical workflows.

Traditional approaches and their limitations

Traditional LOS prediction approaches include logistic regression, linear regression, survival models, and severity scores, which remain important comparators in the literature. Reviews and modeling studies suggest that these methods provide useful baselines but may inadequately capture nonlinear, time-varying, and institution-specific determinants of discharge readiness [1, 2, 10]. In ICU-focused work, severity scoring systems and admission-level predictors have been used to benchmark machine learning models, yet such comparisons are complicated by differences in patient mix and outcome definitions [18, 19, 22]. The literature therefore treats conventional models less as obsolete methods and more as necessary reference points for evaluating added complexity.

EHR data modalities for LOS prediction

EHR-based LOS prediction studies draw on structured data such as demographics, diagnoses, procedures, laboratory values, vital signs, admission source, medications, and service-line information. Some investigations incorporate unstructured data, including clinical notes, where narrative descriptions may contain information about functional status, complications, discharge barriers, or clinician expectations not captured in coded variables [23]. Mixed-data and deep

learning studies have explored ways to combine structured time series with text-derived or high-dimensional clinical representations [13, 14]. However, variability in documentation practices and coding systems makes feature harmonization a persistent challenge across institutions.

Prediction horizons: static vs dynamic

LOS models vary in whether they predict expected duration at admission or update predictions as hospitalization unfolds. Admission-time models may be operationally attractive because they support early bed management, but they may miss complications, response to treatment, and evolving discharge barriers [4, 16, 24]. Dynamic models, including repeated prediction and hidden Markov approaches, attempt to revise LOS estimates as new clinical information becomes available [14, 15]. The literature suggests that this distinction is critical because a model designed for admission triage may not answer the same operational question as a model designed for daily discharge planning.

Outcome definitions: regression vs classification

LOS has been modeled as a continuous outcome, a count outcome, a skewed regression target, or a categorical classification endpoint such as prolonged stay. Two-stage modeling approaches have been proposed because LOS distributions are often highly skewed, with a small subset of patients experiencing very long admissions [10]. Other studies define prolonged ICU stay, extended orthopedic admission, or specialty-specific thresholds, making direct comparison difficult [6, 8, 22, 25]. The literature therefore indicates that outcome definition is not merely a technical choice but a determinant of interpretability, clinical actionability, and benchmark comparability.

Table 1 presents a conceptual typology of LOS prediction targets, showing how outcome definition determines the operational meaning, timing, and methodological risks of model evaluation.

Table 1. Conceptual Typology of LOS Prediction Targets and Their Operational Meaning

Prediction target	Typical modeling formulation	Main operational question answered	Appropriate timing	Key methodological risk	Review-level interpretation
Total LOS at admission	Regression or count prediction	How long is this patient likely to remain hospitalized?	At admission or early hospitalization	Misses evolving complications and discharge barriers	Useful for early capacity planning but limited for daily discharge management
Prolonged LOS	Binary or multiclass classification	Is this patient at risk of exceeding a clinically or operationally meaningful threshold?	Admission, transfer, or specialty entry point	Thresholds vary across settings and reduce comparability	Actionable for risk stratification but difficult to compare across studies
Remaining LOS	Dynamic regression	How many days remain before discharge from the current clinical state?	Daily or repeated prediction	Requires careful handling of time-dependent data and censoring	More aligned with discharge planning than static total LOS models
Specialty-specific LOS	Regression or classification	What duration is expected for this	Service-line admission or	May not generalize outside the specialty	Clinically interpretable but

	within defined service lines	specialty pathway?	procedure date	cohort	transportability is limited
ICU or high-acuity LOS	Regression, prolonged-stay classification, or competing-risk framing	How long will high-acuity resource use continue?	ICU entry or repeated ICU landmarks	Death, transfer, and escalation complicate interpretation	Requires clearer distinction between discharge readiness and survival-related endpoints
Operational discharge-readiness proxy	Classification or ranking	Which patients may need early coordination to prevent delay?	Morning rounds or discharge-planning windows	May conflate clinical readiness with institutional workflow constraints	Strong operational potential but requires prospective workflow evaluation

Feature types and preprocessing

Figure 1 summarizes the review-level synthesis across study settings, EHR feature domains, model architectures, prediction targets, validation patterns, interpretability strategies, and operational implementation gaps.

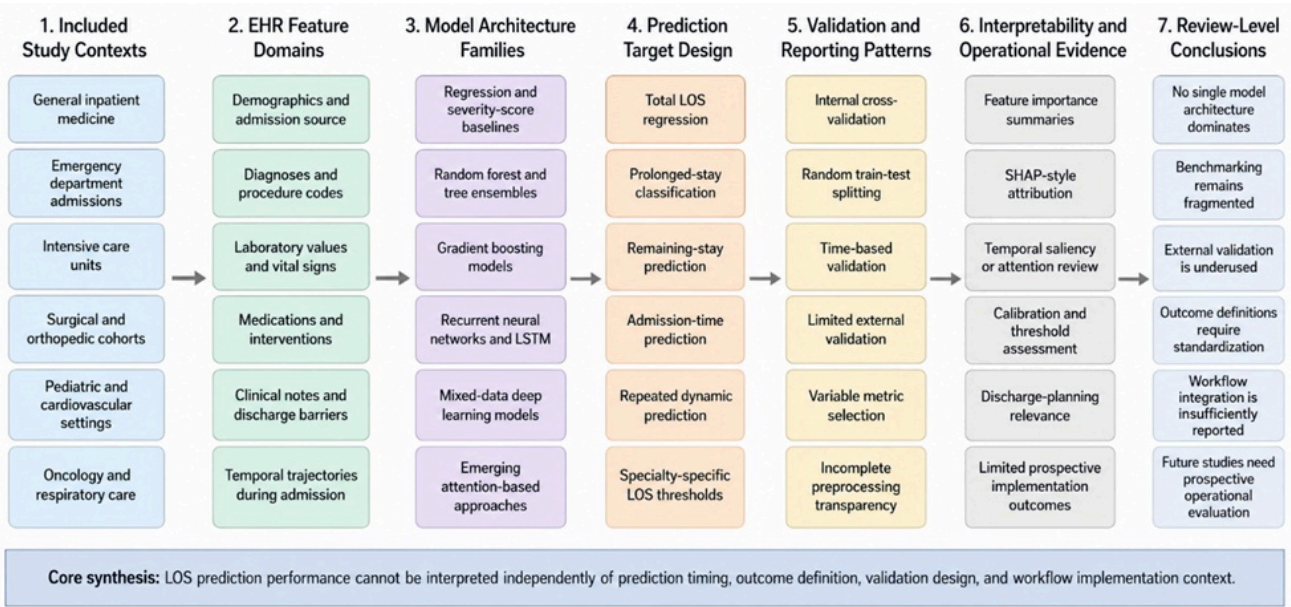


Figure 1. Results Synthesis Framework for Machine Learning-Based Hospital Length-of-Stay Prediction Studies.

Structured features: demographics, labs, vitals

Structured EHR features are the most common inputs in LOS prediction studies because they are readily extractable and can be aligned with admission, ward, or ICU workflows. Investigations have used demographic characteristics, admission route, diagnosis groups, laboratory results, vital signs, procedure codes, and general admission features to represent baseline risk and evolving acuity [11, 24, 26, 27]. Cardiovascular and ICU studies have also used structured indicators to characterize disease severity, treatment burden, and physiologic instability [4, 22, 28, 29]. The literature suggests that the practical advantage of structured features is balanced by the risk that important discharge constraints, social factors, and clinician assessments remain unmeasured.

Unstructured features: clinical notes and discharge summaries

Unstructured clinical text can capture symptoms, clinical reasoning, functional limitations, anticipated discharge barriers, and care coordination issues that may not be present in structured fields. Studies using unstructured data have reported that clinical notes may enrich LOS prediction by representing nuanced information embedded in narrative documentation [23]. Deep learning and mixed-data approaches have also explored medical-record representations that integrate textual or high-dimensional information with structured variables [13, 14]. However, note availability, timing, documentation style, and institutional templates create challenges for reproducibility and external validation.

Handling missing data and variable-length sequences

Missingness is intrinsic to EHR data because measurements are ordered according to clinical need rather than research protocol. LOS studies using ICU vitals, laboratory trends, and repeated prediction frameworks must address irregular observation, sparse measurements, and variable-length hospital courses [7, 14, 24]. Common strategies in the literature include imputation, feature aggregation, time-window summaries, masking, padding, and sequence modeling, although reporting is often insufficient to assess how preprocessing choices influence predictions [9, 12, 13]. This variability limits comparability because two studies may use similar model families while representing missing and temporal information in substantially different ways.

Model Architectures for LOS Prediction

Gradient boosting and tree-based methods

Gradient boosting, random forest, and related ensemble methods are prominent in LOS prediction because they accommodate nonlinear interactions among heterogeneous structured features. Studies have applied XGBoost-like approaches, random forest models, and ensemble learning to hospital, ICU, cardiovascular, and specialty-specific LOS outcomes [16–19]. These approaches are often treated as strong baselines because they can handle mixed feature types and support post hoc attribution through feature importance or SHAP-style explanations. Nevertheless, the literature suggests that their apparent practicality depends on careful validation, calibration, and workflow alignment rather than model class alone.

Recurrent neural networks and LSTM

Recurrent neural networks and LSTM-style models are used when LOS prediction depends on sequences of vital signs, laboratory values, or events observed over time. Neural-network studies using MIMIC-style or clinical record data have explored whether sequential representations can reflect evolving patient status more naturally than static admission snapshots [12, 13]. Repeated-prediction and mixed-data deep learning work further suggests that temporal updating may be useful when the clinical question concerns expected remaining stay rather than initial total stay [14]. However, recurrent models can be difficult to interpret, and their value depends on transparent reporting of time windows, sequence construction, and missing-data handling.

Transformer and attention-based models

Transformer and attention-based approaches are conceptually attractive for LOS prediction because they can represent long-range dependencies, irregular clinical histories, and relationships among heterogeneous events. Within the 2017–2022 LOS literature, attention mechanisms appear more often as an emerging direction or as part of broader deep learning discussions than as a standardized benchmark across institutions [1, 2, 13]. The promise of attention lies in linking model behavior to specific timepoints, notes, or clinical variables, but attention weights should not automatically be interpreted as causal explanations. The literature therefore supports cautious use of attention-based architectures, especially when external validation and clinical interpretability remain limited.

Temporal Modeling and Dynamic Prediction

Static vs dynamic prediction windows

Static LOS prediction models estimate expected stay from information available at admission, while dynamic models revise predictions after new measurements, interventions, or clinical notes become available. Admission-level models have been investigated in cardiovascular, asthma, orthopedic, and ICU settings because early estimates can support bed planning and triage workflows [4, 11, 25, 27]. Dynamic prediction studies, including repeated deep learning and hidden Markov approaches, suggest that updating predictions may better match discharge planning processes during hospitalization [14, 15]. The choice between static and dynamic prediction should therefore be guided by the operational use case rather than by model sophistication alone.

Handling irregularly sampled time series

Temporal EHR data are irregular because sicker patients receive more frequent measurements, while stable patients may have sparse observations. LOS studies using ICU vital signs, cardiovascular ICU records, and general admission features show that preprocessing decisions can shape how models interpret physiologic trajectories [7, 24, 28, 29]. Neural and sequential approaches may use aggregation windows, padding, masking, or learned representations, whereas tree-based models often summarize temporal information into engineered features [12, 13, 18]. The literature suggests that transparent reporting of these design choices is essential because irregular sampling may encode both clinical severity and care-process variation.

Landmark analysis and competing risks

Dynamic LOS prediction is complicated by the fact that discharge is not the only possible endpoint during hospitalization. In ICU and high-acuity cohorts, death, transfer, readmission risk, and treatment escalation may alter the meaning of remaining LOS, making simple discharge-time prediction clinically ambiguous [7-9]. Landmark-style approaches and competing-risk framing are therefore conceptually relevant, even when not consistently implemented in LOS machine learning studies. The literature indicates that future models should more clearly distinguish prediction of discharge timing from prediction of prolonged hospitalization among patients still eligible for discharge.

Validation Strategies and Performance Benchmarking

Internal validation: time-based splitting

Internal validation strategies vary widely across LOS prediction studies, with cross-validation, random train-test splitting, and time-based splitting all appearing in the literature. For hospital prediction tasks, time-based splitting is especially important because future patients, coding changes, workflow shifts, and evolving treatment practices can otherwise leak indirectly into model evaluation [1, 2]. Studies using repeated prediction, ICU time series, and specialty-specific cohorts demonstrate that validation design must reflect the intended deployment environment, not merely the convenience of retrospective data partitioning [7, 14, 24]. The literature suggests that internal validation should be interpreted cautiously unless temporal separation, preprocessing pipelines, and outcome definitions are clearly reported.

External validation across institutions or populations

External validation remains one of the most important limitations in LOS machine learning research. Many investigations are developed within single hospitals, single health systems, or specialty cohorts, which limits claims about transportability to different case mixes, coding systems, bed-management practices, and discharge policies [4-6, 27]. Pediatric, cardiovascular, orthopedic, oncology, and ICU studies illustrate that LOS determinants may differ substantially by specialty and institution, even when similar modeling approaches are used [19, 21, 22, 25]. The literature therefore supports external validation as a prerequisite for generalizability claims rather than an optional methodological extension.

Performance metrics and their trade-offs

LOS prediction studies use different metrics depending on whether LOS is framed as regression, classification, ranking, or prolonged-stay detection. Regression studies commonly report error-based metrics, while classification studies often report discrimination-oriented measures, but these metrics answer different operational questions and should not be treated as interchangeable [1, 10, 16]. The literature suggests that metric selection should align with the intended use case, such as discharge planning, ICU capacity forecasting, prolonged-stay screening, or service-line resource allocation [3, 8, 17]. Because thresholds, LOS distributions, and patient populations differ across studies, this review does not pool metrics or report aggregate performance claims.

Interpretability and Clinical Explainability

Feature attribution methods for tree-based models

Interpretability is central to clinical adoption because LOS predictions may influence discharge planning, staffing expectations, or escalation of care coordination. Tree-based models are often favored in applied studies because they can provide feature-importance summaries, support attribution methods, and offer clinicians a more transparent view of influential predictors than many deep learning architectures [16, 18, 19]. Studies in cardiovascular, ICU, and specialty populations suggest that feature attribution may help identify clinically plausible contributors such as age, admission source, physiologic instability, diagnosis category, and treatment burden [4, 24, 29]. However, attribution should be interpreted as model explanation rather than causal evidence, particularly when predictors reflect care processes as well as patient severity.

Attention weights and saliency maps for deep learning

Deep learning models raise different explainability challenges because their internal representations may be difficult to map onto clinically meaningful events. Recurrent and mixed-data deep learning studies suggest that temporal models can represent evolving trajectories, but their predictions may require additional explanation through saliency, attention, or timepoint-specific attribution methods [12-14]. Attention-based representations are promising because they may highlight specific clinical notes, laboratory changes, or time windows, yet attention weights alone do not guarantee faithful explanation of model reasoning [1, 2]. The literature therefore supports combining deep temporal models with careful clinical review, transparent preprocessing, and post hoc interpretability checks.

Operational Outcomes and Implementation Barriers

Reported operational improvements

Operational outcomes are less consistently reported than retrospective model-development results. Some studies have linked LOS or discharge prediction models to multidisciplinary rounds, discharge coordination, bed management, and patient-flow planning, suggesting that predictive tools may support operational decision-making when embedded into clinical routines [3, 20]. Specialty studies in surgery, cardiovascular care, pediatrics, and ICU settings further indicate that LOS forecasts may be useful for anticipating bed needs or identifying patients who require early discharge planning [6, 21, 22]. However, the literature generally reports fewer prospective implementation outcomes than algorithmic evaluations, limiting conclusions about real-world effect.

Implementation barriers: workflow, governance, and generalizability

Implementation barriers include integration into EHR workflows, clinician trust, governance of model updates, data quality monitoring, and uncertainty about accountability when predictions conflict with clinical judgment. Single-center development studies may fit local documentation habits and discharge policies, but such models can degrade when

transferred to hospitals with different workflows, population structures, or resource constraints [2, 5, 10]. Explainable machine learning work highlights the need for transparent outputs, yet interpretability alone does not resolve operational barriers such as alert fatigue, unclear ownership, and lack of prospective evaluation [19]. The literature suggests that implementation should be treated as a socio-technical process rather than a final step after retrospective model development.

Gaps and Heterogeneity Across Studies

Lack of standardised benchmark datasets

A major gap in the LOS prediction literature is the absence of a widely adopted benchmark dataset and shared prediction task. Public ICU datasets and institutional EHR repositories have enabled methodological development, but studies differ in inclusion criteria, temporal windows, preprocessing rules, discharge definitions, and specialty focus [7, 12, 13]. Systematic reviews have therefore emphasized that apparent differences between model families may reflect dataset construction rather than intrinsic model superiority [1, 2]. Without standardized benchmarks, it remains difficult to determine whether gradient boosting, recurrent neural networks, or hybrid architectures are preferable for specific LOS use cases.

Table 2 provides a methodological maturity matrix for interpreting LOS prediction studies beyond reported performance metrics.

Table 2. Methodological Maturity Matrix for Machine Learning LOS Prediction Studies

Evidence dimension	Low maturity	Moderate maturity	High maturity	Why this matters for the review
Feature representation	Static administrative or admission-only variables	Structured EHR variables with limited temporal summaries	Multimodal EHR features with transparent temporal and missing-data handling	Determines whether the model captures evolving hospitalization trajectories
Model architecture	Single conventional model without justification	Comparison of tree-based, regression, and neural approaches	Architecture selected according to prediction horizon, data type, and operational use case	Prevents unsupported claims that one model family is universally superior
Validation design	Random split only	Cross-validation or internal temporal split	External validation across hospitals, specialties, or time periods	Separates retrospective model fit from real-world transportability
Outcome definition	Ambiguous LOS endpoint	Defined LOS threshold or regression target	Clearly justified target linked to discharge planning, bed management, or resource allocation	Makes performance interpretable and comparable
Interpretability	No explanation or global feature list only	Feature importance or post hoc attribution	Clinically reviewed explanations with calibration and uncertainty reporting	Supports safe use without implying causal interpretation

Implementation evidence	Retrospective performance only	Simulated workflow or decision-support discussion	Prospective evaluation of clinician use, workflow effect, calibration drift, and governance	Determines whether prediction improves operations rather than only model metrics
Reporting transparency	Limited preprocessing and cohort detail	Basic cohort, metric, and feature reporting	Reproducible pipeline description with missingness, timing, exclusions, and deployment context	Enables synthesis across heterogeneous studies

Inconsistent outcome definitions and prediction horizons

Outcome definitions vary across studies, with some predicting total LOS in days, others predicting prolonged stay, and others estimating remaining stay during hospitalization. Prediction horizons also differ, ranging from admission-time prediction to repeated daily prediction and ICU-specific forecasting [8, 9, 14, 26]. Studies using two-stage modeling, hidden Markov approaches, and specialty thresholds illustrate that LOS is not a single homogeneous endpoint but a family of related operational targets [10, 15, 22]. This heterogeneity limits cross-study synthesis and reinforces the need for clearer reporting of when predictions are made, which patients are included, and how discharge outcomes are defined.

Underreporting of negative findings and failed deployments

The LOS prediction literature appears weighted toward model-development studies that report feasible or promising retrospective results, while failed deployments, negative evaluations, and workflow misalignment are less visible. This imbalance makes it difficult to assess how often models fail because of data drift, clinician nonuse, poor calibration, or lack of operational integration [1, 3, 20]. Studies emphasizing explainability, repeated prediction, and operational discharge support suggest that implementation challenges are central to clinical value, yet these challenges are not consistently measured across the field [14, 19]. The literature would benefit from more transparent reporting of unsuccessful implementation attempts, model maintenance burdens, and unintended workflow consequences.

Limitations of the Review

Publication and language bias

This review is limited by its focus on peer-reviewed literature from 2017 to 2022 and by reliance on studies available in the indexed academic record. Excluding grey literature, internal hospital evaluations, commercial deployments, and non-English publications may underrepresent operationally important work that was not published as a conventional journal article or conference paper [1, 2]. Publication bias may also favor studies with positive model-development findings over studies reporting weak transportability, poor workflow fit, or failed implementation [3, 19]. As a result, the synthesis should be interpreted as a review of the published research landscape rather than a complete inventory of all deployed LOS prediction systems.

Heterogeneity precluding meta-analysis

Meta-analysis was not appropriate because the included studies differed substantially in populations, settings, predictors, model architectures, validation strategies, outcome definitions, and reporting practices. ICU studies, cardiovascular cohorts, orthopedic populations, pediatric surgical settings, asthma admissions, and general medicine cohorts represent different clinical and operational prediction problems [6, 21, 22, 27, 28]. In addition, regression and classification outcomes are not directly comparable, and studies use different temporal anchors such as admission, daily updates, or

ICU entry [8, 10, 14]. For these reasons, this review provides a conceptual and critical synthesis rather than pooled estimates or aggregate performance rankings.

Conclusion

Machine learning for hospital length-of-stay prediction expanded substantially from 2017 to 2022, with gradient boosting, ensemble learning, random forests, neural networks, and recurrent architectures appearing across diverse clinical settings. The literature suggests that these approaches can represent complex EHR-derived predictors, but their clinical usefulness depends on validation rigor, interpretability, workflow fit, and alignment with operational decisions. External validation and prospective implementation outcomes remain less common than retrospective model-development studies.

A central gap is the lack of a standardized benchmarking corpus or shared prediction tasks for LOS modeling. Without common datasets, prediction horizons, feature definitions, and reporting standards, comparisons between model architectures remain uncertain. Future work should distinguish clearly between predicting total LOS at admission, predicting prolonged stay, and dynamically estimating remaining stay during hospitalization.

Prospective implementation studies are needed to determine whether LOS prediction tools improve discharge planning, bed allocation, staffing, and financial performance in real clinical environments. Such studies should measure not only predictive accuracy but also usability, calibration over time, clinician response, governance processes, and unintended consequences. Operational value cannot be inferred from retrospective discrimination or error metrics alone.

The field would benefit from reporting guidelines tailored to LOS prediction studies and from multi-center validation consortia that test models across institutions, specialties, and EHR systems. Standardized descriptions of features, missing-data handling, prediction timing, outcome definitions, and implementation context would make future evidence more comparable. Stronger methodological transparency will be essential for moving LOS prediction from retrospective modeling toward safe and useful clinical decision support.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 26 Jul 2022

Revised: 18 Sep 2022

Accepted: 20 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

- Stone K, Zwiggelaar R, Jones P, Mac Parthaláin N. A systematic review of the prediction of hospital length of stay: towards a unified framework. *PLOS Digit Health*. 2022;1(4):e0000017.
<https://doi.org/10.1371/journal.pdig.0000017>.
- Lequertier V, Wang T, Fondrevelle J, Augusto V, Duclos A. Hospital length of stay prediction methods: a systematic review. *Med Care*. 2021;59(10):929-38.
<https://doi.org/10.1097/MLR.0000000000001623>.
- Levin S, Barnes S, Toerper M, Debraine A, DeAngelo A, Hamrock E, et al. Machine-learning-based hospital discharge predictions can support multidisciplinary rounds and decrease hospital length-of-stay. *BMJ Innov*. 2021;7(2):388-95.
<https://doi.org/10.1136/bmjinnov-2019-000428>.
- Daghistani TA, Elshawi R, Sakr S, Ahmed AM, Al-Thwayee A, Al-Mallah MH. Predictors of in-hospital length of stay among cardiac patients: a machine learning approach. *Int J Cardiol*. 2019;288:140-7.
<https://doi.org/10.1016/j.ijcard.2019.01.046>.
- Bacchi S, Tan Y, Oakden-Rayner L, Jannes J, Kleinig T, Koblar S, et al. Machine learning in the prediction of medical inpatient length of stay. *Intern Med J*. 2022;52(2):176-85.
<https://doi.org/10.1111/imj.15235>.
- Sridhar S, Whitaker B, Mouat-Hunter A, McCrory B. Predicting length of stay using machine learning for total joint replacements performed at a rural community hospital. *PLoS One*. 2022;17(11):e0277479.
<https://doi.org/10.1371/journal.pone.0277479>.
- 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/s41598-022-17191-6>.
- Wu J, Lin Y, Li P, Hu Y, Zhang L, Kong G. Predicting prolonged length of ICU stay through machine learning. *Diagnostics*. 2021;11(12):2242.
<https://doi.org/10.3390/diagnostics11122242>.
- 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>.
- Xu Z, Zhao C, Scales CD Jr, Henao R, Goldstein BA. Predicting in-hospital length of stay: a two-stage modeling approach to account for highly skewed data. *BMC Med Inform Decis Mak*. 2022;22(1):110.
<https://doi.org/10.1186/s12911-022-01833-y>.
- Abd-Elrazek MA, Eltahawi AA, Abd Elaziz MH, Abd-Elwhab MN. Predicting length of stay in hospitals intensive care unit using general admission features. *Ain Shams Eng J*. 2021;12(4):3691-702.
<https://doi.org/10.1016/j.asej.2021.04.007>.

Gentimis T, Ala'J A, Durante A, Cook K, Steele R. Predicting hospital length of stay using neural networks on MIMIC III data. In: 2017 IEEE 15th Intl Conf Dependable Autonomic Secure Comput, 15th Intl Conf Pervasive Intell Comput, 3rd Intl Conf Big Data Intell Comput Cyber Sci Technol Congr. Piscataway (NJ): IEEE; 2017. p. 1194-201.
<https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2017.203>.

Zebin T, Rezvy S, Chausalett TJ. A deep learning approach for length of stay prediction in clinical settings from medical records. In: 2019 IEEE Conf Comput Intell Bioinform Comput Biol. Piscataway (NJ): IEEE; 2019. p. 1-5.
<https://doi.org/10.1109/CIBCB.2019.8791463>.

Bacchi S, Gluck S, Tan Y, Chim I, Cheng J, Gilbert T, et al. Mixed-data deep learning in repeated predictions of general medicine length of stay: a derivation study. *Intern Emerg Med.* 2021;16(6):1613-7.
<https://doi.org/10.1007/s11739-020-02629-w>.

Sotoodeh M, Ho JC. Improving length of stay prediction using a hidden Markov model. *AMIA Jt Summits Transl Sci Proc.* 2019;2019:425-34.

Chen Y. [Retracted] Prediction and analysis of length of stay based on nonlinear weighted XGBoost algorithm in hospital. *J Healthc Eng.* 2021;2021:4714898.

Zheng L, Wang J, Sheriff A, Chen X. Hospital length of stay prediction with ensemble methods in machine learning. In: 2021 Int Conf Cyber-Physical Soc Intell. Piscataway (NJ): IEEE; 2021. p. 1-5.
<https://doi.org/10.1109/ICCSI52924.2021.9670330>.

Kuang NL, Leung CH. Performance effectiveness of multimedia information search using the epsilon-greedy algorithm. In: 2019 18th IEEE Int Conf Mach Learn Appl. Piscataway (NJ): IEEE; 2019. p. 929-36.
<https://doi.org/10.1109/ICMLA.2019.00156>.

Alsinglawi B, Alshari O, Alorjani M, Mubin O, Alnajjar F, Novoa M, et al. An explainable machine learning framework for lung cancer hospital length of stay prediction. *Sci Rep.* 2022;12(1):607.
<https://doi.org/10.1038/s41598-021-04609-1>.

Ahn I, Gwon H, Kang H, Kim Y, Seo H, Choi H, 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>.

Corren YB, Merrill J, Wilkinson R, Cannon C, Bickel J, Reis BY. Predicting surgical department occupancy and patient length of stay in a paediatric hospital setting using machine learning: a pilot study. *BMJ Health Care Inform.* 2022;29(1):e100498.
<https://doi.org/10.1136/bmjhci-2021-100498>.

Chen Q, Zhang B, Yang J, Mo X, Zhang L, Li M, et al. Predicting intensive care unit length of stay after acute type A aortic dissection surgery using machine learning. *Front Cardiovasc Med.* 2021;8:675431.
<https://doi.org/10.3389/fcvm.2021.675431>.

Chrusciel J, Girardon F, Roquette L, Laplanche D, Duclos A, Sanchez S. The prediction of hospital length of stay using unstructured data. *BMC Med Inform Decis Mak.* 2021;21(1):351.
<https://doi.org/10.1186/s12911-021-01722-z>.

Alghatani K, Ammar N, Rezgui A, Shaban-Nejad A. Predicting intensive care unit length of stay and mortality using patient vital signs: machine learning model development and validation. *JMIR Med Inform.* 2021;9(5):e21347.
<https://doi.org/10.2196/21347>.

Han C, Liu J, Wu Y, Chong Y, Chai X, Weng X. To predict the length of hospital stay after total knee arthroplasty in an orthopedic center in China: the use of machine learning algorithms. *Front Surg.* 2021;8:606038.
<https://doi.org/10.3389/fsurg.2021.606038>.

Alabbad DA, Almuhaideb AM, Alsunaidi SJ, Alqudaihi KS, Alamoudi FA, Alhobaishi MK, et al. Machine learning model for predicting the length of stay in the intensive care unit for COVID-19 patients in the eastern province of Saudi Arabia. *Inform Med Unlocked*. 2022;30:100937.

<https://doi.org/10.1016/j.imu.2022.100937>.

Jayamini WK, Mirza F, Naeem MA, Chan AH. State of asthma-related hospital admissions in New Zealand and predicting length of stay using machine learning. *Appl Sci*. 2022;12(19):9890.

<https://doi.org/10.3390/app12199890>.

Alsinglawi B, Alnajjar F, Mubin O, Novoa M, Alorjani M, Karajeh O, et al. Predicting length of stay for cardiovascular hospitalizations in the intensive care unit: machine learning approach. In: 2020 42nd Annu Int Conf IEEE Eng Med Biol Soc. Piscataway (NJ): IEEE; 2020. p. 5442-5.

<https://doi.org/10.1109/EMBC44109.2020.9176654>.

Karboub K, Tabaa M. A machine learning based discharge prediction of cardiovascular diseases patients in intensive care units. *Healthcare (Basel)*. 2022;10(6):966.

<https://doi.org/10.3390/healthcare10060966>.