Supporting Hospital Operational Decision-Making Through Early Prediction of Length of Stay

Matilde Sousa[1], João Lopes[1(✉)], Tiago Guimarães[1], Agostinho Barbosa[2], and Manuel Filipe Santos[1]

[1] ALGORITMI/LASI Research Centre, University of Minho, Braga, Portugal
jlopes@dsi.uminho.pt
[2] Unidade Local de Saúde Tâmega e Sousa (ULSTS), Penafiel, Portugal

Abstract. Early prediction of hospital Length of Stay (LOS) is essential for effective bed management, discharge planning, and hospital operational coordination. While Machine Learning (ML) has been widely applied to LOS prediction, limited evidence exists on how model performance varies across different hospital operational contexts when predictions rely only on information available at admission or transfer. This study compares classical ensemble models and neural network approaches for early LOS prediction in two operational scenarios: emergency admissions and internal hospital transfers. A retrospective analysis was conducted using anonymized administrative and temporal data from Unidade Local de Saúde Tâmega e Sousa (ULSTS). Results show that gradient boosting models achieved strong ordinal performance in large emergency datasets, whereas neural networks provided more balanced predictions in smaller, highly imbalanced transfer data. These findings highlight the importance of context-aware model selection when developing predictive decision-support tools for hospital operations.

Keywords: Healthcare · Machine Learning · Length of Stay

1 Introduction

Efficient inpatient management represents a major operational challenge for modern healthcare systems worldwide. The growing demand for hospital care, driven by ageing populations, increasing prevalence of chronic diseases, and rising patient complexity, combined with inherently limited physical and human resources, contributes to persistent hospital overcrowding and negatively affects key performance indicators such as waiting times, bed availability, staff workload, and overall operational efficiency [1–5]. These pressures place considerable strain on healthcare institutions, undermining their capacity to deliver timely, high-quality care and highlighting the urgent need for more effective approaches to inpatient resource planning and management [1,3].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16761, pp. 631–643, 2027.
https://doi.org/10.1007/978-3-032-30527-5_41

Length of Stay (LOS) is widely recognised as a central indicator for hospital resource management, directly influencing bed allocation, discharge planning, patient flow coordination, and overall capacity management. Prolonged or unpredictable hospitalizations generate cascading inefficiencies throughout the hospital system, from delayed admissions in emergency departments to suboptimal utilisation of surgical and intensive care capacity [6,7]. Accurate LOS prediction at the time of admission allows hospital managers and clinicians to anticipate bed requirements, optimise patient pathways and workflows, support earlier discharge planning, and reduce costs associated with unnecessarily prolonged hospitalizations, all while maintaining the quality and safety of care delivered to patients [6–8]. Beyond its administrative relevance, reliable LOS estimation also carries direct clinical implications, enabling more proactive and personalised care coordination throughout the inpatient episode [8].

In this context, data-driven approaches have gained substantial traction as a means of improving predictive accuracy and supporting clinical decision-making. Recent studies have increasingly applied Machine Learning (ML) techniques to LOS prediction, demonstrating that predictive models are capable of capturing complex, non-linear interactions among demographic, clinical, and operational variables that traditional statistical methods often fail to adequately represent [6–11,17]. Algorithms such as gradient boosting, random forests, and deep learning architectures have shown promising results across a range of clinical settings, suggesting the potential for ML-based tools to become integrated components of hospital management systems [9–11]. Nevertheless, several methodological challenges persist and limit the broader adoption of these approaches in routine clinical practice. These include the appropriate selection and engineering of predictive features, the inherent heterogeneity and incompleteness of clinical data collected in real-world settings, the risk of overfitting to specific institutional contexts, and the limited generalisability of trained models across different patient populations, care pathways, or clinical environments [6,7,18]. Addressing these challenges is essential to developing LOS prediction systems that are not only statistically robust but also practically deployable and equitable across diverse healthcare contexts.

1.1 Operational Pressure and the Strategic Role of Length of Stay

LOS is a key operational indicator in hospital management, as it directly affects bed occupancy, discharge planning, staffing needs, and overall healthcare costs [6,7]. Variations in LOS can lead to capacity constraints that extend across multiple medical specialties, disrupting patient flow and limiting the system's ability to accommodate new admissions [1,19]. For this reason, a reliable estimate plays a key role in hospital capacity planning and day-to-day operational decisions [2,3,19]. Its relevance becomes even more critical in high-uncertainty settings, such as Emergency Departments. In these environments, demand variability and limited information at the time of admission make timely decision-making particularly challenging. Overcrowding in emergency departments is consistently associated with longer wait times and reduced operational efficiency [4,11,20].

This metric also has a direct relationship with the hospital's internal dynamics, particularly in the context of interdepartmental transfers, where it influences both workload distribution and bed turnover. However, despite its clear operational importance, the application of predictive models in transfer scenarios remains relatively limited in the existing literature.

1.2 Predictive Modeling of LOS: From Statistical Approaches to Deep Learning

Predictive modelling has become an important research direction for estimating hospital LOS, driven by the increasing availability of electronic health records and administrative Healthcare datasets. Early studies relied mainly on traditional statistical methods, such as linear regression, logistic regression, and survival analysis, which provided interpretable insights into hospitalization duration but often assumed linear relationships between variables [9,18].

With the growth of Healthcare data, ML techniques have increasingly been adopted to capture nonlinear relationships and complex interactions among demographic, clinical, and operational variables. Tree-based ensemble models, including Random Forest and Gradient Boosting, have shown strong predictive performance in structured datasets and are widely used in LOS prediction tasks [7,12,21].

More recently, deep learning approaches have been explored for modeling high-dimensional clinical data, demonstrating promising results in predicting outcomes such as LOS, mortality, and readmission [10,13,14]. Neural architectures designed for tabular Healthcare data have further expanded the applicability of deep learning methods in structured datasets [11,15].

Despite these advances, systematic reviews indicate that no single modelling approach consistently outperforms others across all Healthcare contexts, as performance often depends on dataset size, eature availability, class imbalance, and the clinical environment in which predictions are applied [6,7,22].

1.3 Study Objective and Contributions

Despite the growing use of ML techniques for predicting hospital LOS, several methodological and operational limitations remain. Many models rely on clinical variables that become available only during hospitalization, such as laboratory tests or treatment information. Although these variables may improve predictive performance, they limit the practical applicability of models intended to support early operational decision-making at the time of admission or transfer [6,7,18].

Additionally, there is no consensus on the most appropriate formulation of the LOS prediction problem. Some studies treat LOS as a regression task, while others adopt binary classification to identify prolonged hospitalizations, approaches that may not reflect the operational need to interpret LOS through clinically meaningful time intervals for capacity planning [9,21].

Unlike many previous approaches that rely on variables collected during hospitalization, the proposed models use information realistically available for operational planning [6,7]. The study evaluates LOS prediction across two distinct hospital contexts: emergency admissions and internal patient transfers. These scenarios differ in dataset scale, feature availability, and class imbalance, allowing analysis of model behaviour under heterogeneous operational conditions. Classical ensemble ML models and neural network architectures are compared using consistent validation procedures and imbalance-aware evaluation metrics. The main contributions are:

1. Context-aware formulation of the LOS prediction problem aligned with hospital operational workflows, using only information available at the time of admission or transfer;
2. Systematic comparison between classical ensemble and deep learning models across two operationally distinct hospital contexts with different data regimes;
3. Analysis of how dataset scale, class imbalance, and structural heterogeneity influence predictive performance and model generalisation;
4. Discussion of the trade-off between predictive optimality and operational maintainability as an explicit criterion for model selection in hospital decision-support contexts.

This paper is organised as follows. Chapter 2 describes the research methodology, including the study design, ethical considerations, problem formulation, data understanding and preprocessing procedures, and the Machine Learning models developed for each operational context. Chapter 3 presents the comparative results across the emergency admission and internal hospital transfer scenarios. Chapter 4 discusses the findings in relation to the existing literature and reflects on the practical implications for hospital decision support. Finally, a conclusion for the paper, emphasizing directions for future research.

2 Methods

2.1 Ethical Considerations

This study used exclusively anonymous data provided by ULSTS and did not involve any sensitive personal information. As a result, review by an ethics committee was not required, in accordance with the General Data Protection Regulation (GDPR).

2.2 Study Design

The study followed a Design Science Research (DSR) methodology [24], operationalized through the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework [25].

DSR provided the framework for the research, guiding the development and evaluation of artifacts designed to address a specific organizational problem [24].

Within this paradigm, the research problem was identified, and the resulting predictive models were treated as artifacts whose value is ultimately assessed by their practical utility.

CRISP-DM supplemented this methodological framework by providing a structured process model for data mining activities [25]. The framework organized the analytical workflow into six sequential phases:

1. Business Understanding, in which the organisational objectives and predictive requirements were translated into concrete data mining goals;
2. Data Understanding, encompassing the initial collection, exploration, and quality assessment of the available clinical and operational data;
3. Data Preparation, involving the transformation, cleaning, and feature engineering steps necessary to construct analysis-ready datasets;
4. Modelling, in which a range of Machine Learning algorithms were selected, configured, and trained;
5. Evaluation, where model performance was assessed against both technical metrics and the broader research objectives;
6. and Deployment, addressing the practical considerations for integrating the resulting models into hospital decision-support workflows.;

The iterative nature of CRISP-DM allowed the findings from later phases to serve as a foundation and refine the decisions made in earlier phases, ensuring that the final results were methodologically sound and aligned with the operational realities of the clinical setting.

2.3 Problem Statement

Efficient management of hospital resources requires forecasting patient flow and length of stay. LOS is a critical operational indicator that influences bed occupancy, discharge planning, staffing, and the overall performance of the system. An inaccurate estimate of LOS can lead to overcrowding, delays in admissions, and inefficient allocation of resources.

Although hospital information systems collect large volumes of administrative and operational data, planning decisions are often based on historical averages or subjective clinical judgments. Furthermore, many predictive studies use variables that are not available at the time of operational decision-making, limiting their real-world applicability. This study addresses two operational challenges:

1. **Early LOS prediction in emergency admissions (Problem A):** Emergency departments operate under high demand variability and limited downstream capacity. Admission decisions must often be made with limited information, yet accurate LOS estimates are essential for proactive bed management and atient flow coordination.
2. **LOS prediction in internal hospital transfers (Problem B):** Transfers between hospital units affect workload distribution and bed turnover. Unanticipated prolonged stays may generate congestion across departments, yet

predictive tools supporting LOS estimation at the time of transfer remain scarce.

2.4 Data Understanding and Preprocessing

Data for both use cases were extracted from the hospital information systems of ULSTS, covering emergency admission episodes and internal hospital transfer events. In both contexts, only administrative, demographic, and temporal variables available at the time of admission or transfer decision were included, ensuring real-world operational applicability and preventing information leakage from variables that arise only during hospitalization.

The emergency admissions dataset comprised historical inpatient episodes with predictors including patient demographic attributes (e.g., age, sex), admission timestamps, day-of-week and seasonal indicators, and operational descriptors such as the care location within the emergency department. LOS was defined as the elapsed time between admission and discharge and transformed into five ordinal categories ($\leq$3 days, 4–7 days, 8–14 days, 15–30 days, and >30 days), framing the prediction task as an ordinal multiclass classification problem aligned with clinically and operationally meaningful duration thresholds. Episodes were further segmented by care location, as each unit reflects a distinct clinical workflow and patient case-mix that may differentially influence LOS dynamics.

The internal transfer dataset comprised intra-hospital movement episodes between clinical specialties and physical units. Available predictors included demographic attributes, origin and destination units, transfer timestamps, and calendar-derived temporal features. LOS in this context was discretized into operationally relevant multiclass intervals, with short stays (1–3 days) aggregated into a single category to reduce class sparsity, while longer duration intervals were preserved as distinct classes to support meaningful operational differentiation.

Data preprocessing involved systematic cleaning, treatment of missing values, categorical variable encoding, and feature engineering of temporal variables to derive calendar-based indicators (e.g., day of week, month, admission period). To prevent post hoc bias, only predictors available at the time of the operational decision were retained in the modelling pipeline. Both datasets exhibited asymmetric LOS distributions characterised by a predominance of short hospitalizations and a smaller proportion of prolonged stays, a structural imbalance that informed the selection of evaluation metrics and the design of training strategies. During the exploratory phase, multiple LOS interval aggregation schemes were assessed; the final class definitions were selected based on predictive performance, class distribution stability, and alignment with operational planning requirements.

2.5 Proposed Machine Learning Models

ML Approach for Problem A (Emergency Admissions). In this context, LOS prediction was formulated as an ordinal multiclass classification task. Given

the relatively large dataset, both classical ensemble models and deep learning architectures were evaluated. Classical approaches included tree-based ensemble methods, which are well suited for structured Healthcare datasets due to their ability to capture nonlinear relationships and handle mixed feature types. In parallel, neural networks were implemented to assess whether learned embeddings and nonlinear feature representations could improve predictive performance.

ML Approach for Problem B (Internal Hospital Transfers). For internal transfers, LOS prediction was formulated as a multiclass classification task. The smaller dataset and stronger class imbalance required careful attention to model stability and generalization. Baseline classical models included Random Forest and Histogram-based Gradient Boosting, selected for their robustness in structured datasets. To mitigate class imbalance, class weighting and sampling strategies were applied during training.

Hyperparameters were optimized through cross-validation. Neural models incorporated embedding layers for categorical variables and fully connected layers to capture nonlinear feature interactions. Regularization techniques and early stopping were used to reduce overfitting.

3 Results

3.1 Comparative Results Across Operational Contexts

Although both problems involve predicting LOS, they differ substantially in class distribution and operational constraints. These differences directly influence the model's behavior, generalization ability, sensitivity to class imbalance, and the criteria used for final model selection. In the context of the emergency department, episodes were grouped by care location, representing distinct operational units within the emergency department. Each location corresponds to a specific clinical workflow, a set of patients, and a configuration of factors that can influence the dynamics of LOS. The ML models were trained independently for each location to capture specific patterns and relationships within the data. The performance of the Random Forest (RF) model across the different locations is summarized in Table 1.

Table 1. Random Forest Performance by Emergency Context

Location	Accuracy	F1-macro	AUC	QWK
1	0.832	0.830	0.971	0.875
3	0.915	0.867	0.987	0.919
4	0.651	0.658	0.891	0.758
9	0.838	0.820	0.969	0.853

The model performed well in locations 1, 3, and 9, with QWK values above 0.85, indicating agreement between the predicted and actual LOS classes. However, performance was notably lower at location 4, where both Precision and QWK decreased significantly. This decline is consistent with the reduced number of available training observations at that location. HistGradientBoosting (HGB), shown in Table 2, demonstrated consistent improvements over Random Forest at most locations.

Table 2. HGB Performance by Location in Emergency Context

Location	Accuracy	F1-macro	AUC	QWK
1	0.863	0.863	0.981	0.906
3	0.937	0.896	0.987	0.933
4	0.651	0.658	0.891	0.758
9	0.877	0.862	0.976	0.882

HGB achieved strong predictive performance at locations 1 and 3, with QWK values exceeding 0.90, indicating high agreement among the LOS categories. This confirms the effectiveness of the methods for structured data in high-volume contexts. At location 4, results stabilized due to data scarcity and a more pronounced class imbalance, which limited the model's ability to learn robust decision boundaries.

Error analysis revealed that most misclassifications occurred between adjacent LOS categories. Performance was also stronger for extreme LOS categories compared to intermediate ranges, reflecting the difficulty of distinguishing hospitalizations of average duration.

Although site-specific models achieved high performance in some units, maintaining multiple models introduced scalability and maintenance challenges. To address this limitation, a global ordinal neural network was trained using all emergency episodes collectively. The validation results for this model are presented in Table 3.

Table 3. Global Neural Network

Metric	Value
F1-macro	0.365
QWK	0.453
Within ± 1 class	$\sim$0.65

Although the macro-F1 and QWK scores were lower than those achieved by the best classical models per location, approximately 65% of the predictions fell

within ±1 LOS of the true class. The neural model preserved the ordinal structure and incorporated an auxiliary regression task to improve training stability. It is important to note that it provided several structural advantages, including a single global model applicable across all locations, embedding-based categorical representations, and improved scalability. For these reasons, the neural model was selected as the final solution for the emergency context.

The transfer setting presented a different scenario, with a smaller dataset and severe class imbalance. The classic ensemble baselines exhibited severe overfitting, and to improve generalization, a robust pipeline incorporating temporal validation, correlation filtering, and ensemble techniques was introduced. The best-performing classic model (Bagging) is presented in Table 4.

Table 4. Advanced Temporal Pipeline (Bagging)

Model	Holdout Accuracy	Macro F1	Weighted F1	Holdout Kappa
Bagging	0.888	0.220	0.870	0.711

The holdout Kappa result indicates a significant predictive structure, beyond random agreement. The F1-macro performance is below the desired threshold, reflecting a persistent difficulty in recognizing minority LOS categories due to class imbalance. To improve performance, a neural network was trained using early stopping and weighted sampling. The final validation results are presented in Table 5.

Table 5. Neural Network

Metric	Value
Accuracy	0.855
Balanced Accuracy	0.855
Macro F1	0.849
Macro Precision	0.862
Macro Recall	0.855
Kappa	0.850

The global ordinal neural network achieved balanced and stable predictive performance, with Macro F1 $\approx$ 0.85 and QWK $\approx$ 0.85. Low variability across validation folds indicated strong robustness and generalization. The neural approach provided consistent performance across LOS categories and mitigated class imbalance by improving discrimination in minority and intermediate classes while preserving ordinal consistency. This resulted in more homogeneous predictions across operational units, supporting a scalable and unified deployment.

In the emergency setting, the larger dataset enabled strong performance of classical per-location models, with QWK values above 0.90 in some units. However, maintaining multiple models introduced fragmentation and scalability challenges. In contrast, the transfer context was characterised by a smaller dataset and severe class imbalance. Classical models showed poor generalisation, whereas the neural model achieved balanced macro-level performance and stronger inter-class agreement.

4 Discussion

The results align with the broader literature on ML-based LOS prediction [6,17]. The strong performance of gradient boosting in the emergency context is consistent with evidence that tree-based ensembles effectively capture nonlinear interactions in structured administrative data [9]. The superiority of the neural model in the transfer context corroborates findings that deep architectures offer advantages under class imbalance and limited sample sizes [11]. A notable divergence, however, is that the global neural model did not surpass the best per-location classical models in the emergency setting, suggesting that gradient boosting remains highly competitive when sufficient structured data are available. The near-zero Cohen's Kappa of classical baselines in the transfer context, despite moderate accuracy, further highlights that accuracy is a misleading metric in severely imbalanced multiclass settings. Macro F1 and Kappa proved essential for detecting model collapse in minority classes.

From a practical standpoint, these results reinforce the idea that model selection in healthcare data analysis must balance predictive performance with operational feasibility. The global neural model in the emergency department setting, despite exhibiting slightly lower peak metrics, reduces implementation complexity by providing a single artifact for all care settings (a consideration that is often just as important as marginal gains in metrics in real-world hospital settings).

These findings are specific to ULSTS and should not be generalised without further validation. Differences in patient case-mix, specialty configurations, coding conventions, and data quality across institutions may substantially affect model performance. External validation and context-specific recalibration would be required before deployment in other settings.

5 Conclusion

This study evaluated ML approaches for early LOS prediction across two operationally distinct contexts at ULSTS demonstrating that no single modelling approach dominates across settings. The effectiveness of each approach depends on the type of data: Gradient Boosting performed well on large, structured datasets, while neural networks proved to be more robust with small sample sizes and severe class imbalances. Model selection was therefore guided not only by predictive metrics but also by operational considerations related to scalability and implementation feasibility.

The results are limited to the scope of a single institution. Implementing these models as operational decision-support tools would require several steps that were not addressed in this study. From an integration perspective, the predictive results should be incorporated into existing hospital workflows, with bed managers, clinical coordinators, and admissions staff as the primary stakeholders. Explanation mechanisms, such as summaries of the contribution of features or risk alerts for predicted prolonged stays, would be essential to bolster clinicians' confidence and feasibility. From an operational perspective, implementation would require automated data flows linked to the hospital information system, mechanisms for detecting data anomalies, and periodic retraining of the models to account for changes in the composition of the patient population or in care pathways. Finally, evaluation studies (e.g., bed turnover rates, unplanned prolonged stays, and discharge preparation times) would be needed to determine whether predictive accuracy translates into measurable operational benefits. Additionally, validation across multiple sites would be necessary to assess the model's transferability between institutions with different organizational structures and patient populations.

Acknowledgments. This work has been supported by FCT–Fundação para a Ciência e Tecnologia within the R&D Unit Project Scope UID/00319/2025 - Centro ALGORITMI (ALGORITMI/UM).

References

1. Bagust A, Place M, Posnett, J.W.: Dynamics of bed use in accommodating emergency admissions: stochastic simulation model. BMJ. **319**, 155 (1999). https://doi.org/10.1136/bmj.319.7203.155
2. Hulshof, P., Kortbeek, N., Boucherie, R., et al.: Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. Health Syst. **1**, 129–175 (2012). https://doi.org/10.1057/hs.2012.18
3. Litvak, E., Bisognano, M.: More patients, less payment: increasing hospital efficiency in the aftermath of health reform. Health Affairs. **30**(1), 76–80 (2011). https://doi.org/10.1377/hlthaff.2010.1114
4. Hoot, N.R., Aronsky, D.: Systematic review of emergency department crowding: causes, effects, and solutions. Ann Emerg Med. **52**(2), 126–136 (2008). https://doi.org/10.1016/j.annemergmed.2008.03.014
5. Morley, C., Unwin, M., Peterson, G.M., Stankovich, J., Kinsman, L.: Emergency department crowding: a systematic review of causes, consequences and solutions. PLoS One. **13**(8), e0203316 (2018). https://doi.org/10.1371/journal.pone.0203316
6. Gokhale, S., Taylor, D., Gill, J., et al.: Hospital length of stay prediction tools for all hospital admissions and general medicine populations: systematic review and meta-analysis. Front Med (Lausanne). **10**, 1192969 (2023). https://doi.org/10.3389/fmed.2023.1192969
7. Lequertier, V., Wang, T., Fondrevelle, J., Augusto, V., Duclos, A.: Hospital length of stay prediction methods: a systematic review. Med Care. **59**(10), 929–938 (2021). https://doi.org/10.1097/MLR.0000000000001596

8. Beam, A.L., Kohane, I.S.: Big data and machine learning in health care. JAMA. **319**(13), 1317–1318 (2018). https://doi.org/10.1001/jama.2017.18391
9. Baek, H., Cho, M., Kim, S., Hwang, H., Song, M., Yoo, S.: Analysis of length of hospital stay using electronic health records: a statistical and data mining approach. PLoS One. **13**(4), e0195901 (2018). https://doi.org/10.1371/journal.pone.0195901
10. Rajkomar, A., Oren, E., Chen, K., et al.: Scalable and accurate deep learning with electronic health records. NPJ Digit. Med. **1**, 18 (2018). https://doi.org/10.1038/s41746-018-0029-1
11. Chen, J., Qi, T.D., Vu, J., Wen, Y.: A deep learning approach for inpatient length of stay and mortality prediction. J. Biomed. Inform. **147**, 104526 (2023). https://doi.org/10.1016/j.jbi.2023.104526
12. Breiman, L.: Random Forests. Machine Learning. **45**, 5–32 (2001). https://doi.org/10.1023/A:1010933404324
13. Shickel, B., Tighe, P.J., Bihorac, A., Rashidi, P.: Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record Analysis. IEEE J. Biomed. Health Inform. **22**(5), 1589–1604 (2018). https://doi.org/10.1109/JBHI.2017.2767063
14. Harutyunyan, H., Khachatrian, H., Kale, D.C., Ver Steeg, G., Galstyan, A.: Multitask learning and benchmarking with clinical time series data. Sci. Data. **6**(1), 96 (2019). https://doi.org/10.1038/s41597-019-0103-9
15. Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., Kasneci, G.: Deep Neural Networks and Tabular Data: A Survey. IEEE Trans. Neural Netw. Learn. Syst. **35**(6), 7499–7519 (2024). https://doi.org/10.1109/TNNLS.2022.3229161
16. Goldstein, B.A., Navar, A.M., Pencina, M.J., Ioannidis, J.P.: Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc. **24**(1), 198–208 (2017). https://doi.org/10.1093/jamia/ocw042
17. Farimani, R.M., Karim, H., Atashi, A., et al.: Models to predict length of stay in the emergency department: a systematic literature review and appraisal. BMC Emergency Med. **24**(1), 54 (2024). https://doi.org/10.1186/s12873-024-00965-4
18. Futoma, J., Simons, M., Panch, T., Doshi-Velez, F., Celi, L.A.: The myth of generalisability in clinical research and machine learning in health care. Lancet Digit. Health. **2**(9), e489–e492 (2020). https://doi.org/10.1016/S2589-7500(20)30186-2
19. Gualandi, R., Masella, C., Tartaglini, D.: Improving hospital patient flow: a systematic review. Bus. Process Manage. J. **26**(6), 1541–1575 (2020). https://doi.org/10.1108/BPMJ-10-2017-0265
20. Obermeyer, Z., Emanuel, E.J.: Predicting the future — Big Data, Machine Learning, and Clinical Medicine. New England J. Med. **375**, 1216–1219 (2016). https://doi.org/10.1056/NEJMp1606181
21. Esteva, A., Robicquet, A., Ramsundar, B., et al.: A guide to deep learning in healthcare. Nat. Med. **25**, 24–29 (2019). https://doi.org/10.1038/s41591-018-0316-z
22. Sendak, M.P., D'Arcy, J., Kashyap, S., et al.: A path for translation of machine learning products into healthcare delivery. EMJ Innov. (2020). https://doi.org/10.33590/emjinnov/19-00172
23. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manage. Inf. Syst. **24**(3), 45–77 (2007). https://doi.org/10.2753/MIS0742-1222240302

24. Design Science Methodology for Information Systems and Software Engineering. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-43839-8_20
25. Chapman, P., Clinton, J., Kerber, R., et al.: CRISP-DM 1.0: Step-by-step data mining guide. CRISP-DM Consortium (2000)