Intelligent Human Systems Integration (IHSI): Vol. XXX, 2025
doi: 10.54941/ahfeXXXX

Machine Learning-Based Prediction of Emergency Department Prolonged Length of Stay: A Case Study from Italy

Paolo Perliti Scorzoni, Anita Giovanetti, Federico Bolelli, and Costantino Grana

Università degli Studi di Modena e Reggio Emilia - Dipartimento di Ingegneria “Enzo Ferrari”, Modena, Italy

ABSTRACT

Overcrowding in Emergency Departments (EDs) is a pressing concern driven by high patient demand and limited resources. Prolonged Length of Stay (pLOS), a major contributor to this congestion, may lead to adverse outcomes, including patients leaving without being seen, suboptimal clinical care, increased mortality rates, provider burnout, and escalating healthcare costs.
This study investigates the application of various Machine Learning (ML) algorithms to predict both LOS and pLOS. A retrospective analysis examined 32,967 accesses at a northern Italian hospital’s ED between 2022 and 2024. Twelve classification algorithms were evaluated in forecasting pLOS, using clinically relevant thresholds. Two data variants were employed for model comparison: one containing only structured data (e.g., demographics and clinical information), while a second one also including features extracted from free-text nursing notes. To enhance the accuracy of LOS prediction, novel queue-based variables capturing the real-time state of the ED were incorporated as additional dynamic predictors. Compared to single-algorithm models, ensemble models demonstrated superior robustness in forecasting both ED-LOS and ED-pLOS. These findings highlight the potential for integrating ML into EDs practices as auxiliary tools to provide valuable insights into patient flow. By identifying patients at high risk of pLOS, healthcare professionals can proactively implement strategies to expedite care, optimize resource allocation, and ultimately improve patient outcomes and ED efficiency, promoting a more effective and sustainable public healthcare delivery.

Keywords: ED-LOS, pLOS, Emergency Department, Overcrowding, Machine Learning, Public Healthcare, Sustainability

INTRODUCTION

Emergency Departments (EDs) represent an essential pillar within the public health infrastructure, functioning as the primary point of access for emergency and urgent medical care on a continuous basis (24/7). Staffed by highly qualified physicians, EDs provide comprehensive evaluation and treatment for a wide range of illnesses and injuries.

The inherent unpredictability of emergencies necessitates a flexible approach within these departments, as activities cannot be predetermined. Patients may self-present or arrive by emergency vehicle, requiring a diverse array of treatments to address their individual needs. As a result, ED overcapacity poses a major global challenge to healthcare systems, primarily due to the widening gap between the ever-increasing demand for emergency services and the limited resources available. This imbalance is further compounded by an ageing population and rigorous cost-containment policies. The effects of ED overcrowding are multiple, ranging from extended wait times and delayed treatment to ambulance rerouting and higher expenditures for hospitals. Additionally, overcrowding is associated with an escalating incidence of patients leaving the ED without physician consultation due to dissatisfaction with prolonged wait times (Phillips et al. 2017).

Addressing these challenges requires a comprehensive approach, combining effective triage systems and strategic resource allocation. Over the past two decades, EDs have been implementing initiatives like specialized Fast-track models (Manno et al. 2015), and Short-Stay Observation (SSO) units (Daly et al. 2003; Baugh et al. 2011). Although these interventions demonstrated benefits, there remains substantial potential to further optimize ED resource management through the integration of accurate Length of Stay (LOS) prediction tools at the triage stage, an innovation that is currently lacking in many healthcare systems.

ED-LOS, defined as the time elapsed from patient's arrival in ED to departure (encompassing discharge, transfer to another healthcare setting, or admission to a medical ward), is influenced by numerous variables. These include hospital organization, staffing levels, triage procedures, the need for consultations or advanced imaging (Yoon et al. 2003), bed availability (Liew et al. 2003), and patient characteristics like age, sex, and comorbidities (Sarıyer et al. 2020).

Prolonged LOS (pLOS) in EDs is not merely symptomatic of overcrowding but also a contributing factor, leading to deterioration of clinical conditions for waiting patients (Kenny et al. 2020), higher mortality rates (Savioli et al. 2022), and severe provider burnout (Gaeta et al. 2019; Phillips et al. 2022).

The development of predictive models to forecast ED-LOS upon triage completion represents an emerging area of research. Application of analytics and Machine Learning (ML) techniques holds significant promise for early identification of patients at elevated risk for pLOS, enabling targeted interventions to prevent complications, streamline care processes, and ensure timely transitions between care settings. Such a proactive approach not only helps alleviate the adverse effects of overcrowding but also safeguards the long-term effectiveness and sustainability of EDs within the public health network.

The present study aimed to develop ML-based models for predicting ED-pLOS in general patient populations. Multiple algorithms were compared to identify the most effective models, examining the impact of different LOS thresholds. Additionally, key features influencing predictions were analyzed.

RELATED WORKS

Recent applications of ML models for predicting LOS in healthcare settings have evidenced their potential in dealing with the inherent complexities of Emergency Departments, where conventional statistical techniques often falter due to their sensitivity to outliers and reliance on linearity assumptions.

Rahman et al. (2020) utilized a decision tree to identify patients at risk of prolonged ED-LOS (defined as LOS exceeding 4 hours) in an Australian public hospital, achieving an accuracy of 85%. Similarly, Etu et al. (2022) employed gradient boosting to predict ED-pLOS for COVID-19 patients in a Michigan hospital, also attaining 85% accuracy. D'Etienne et al. (2021) developed a two-step predictive model for early detection of extended ED-LOS (>16 hours), demonstrating 67.8% accuracy. Walsh et al. (2004) applied Artificial Neural Network (ANN) to forecast ED disposition for infants with bronchiolitis, considering patient demographics and medical tests, reporting an accuracy of 81%. Ricciardi et al. (2024) explored multiple ML algorithms to predict prolonged ED-LOS (> 3 hours) in a university hospital in Salerno, Italy, with Random Forest performing with the highest accuracy (74.9%).

The importance of feature selection in predicting LOS has been recognized in other works. Azari et al. (2015) proposed a framework that integrates data mining and class imbalance learning to estimate prolonged LOS (>14 hours). Naemi et al. (2021) emphasized the critical role of handling missing values and data skewness in LOS prediction, for both regression and classification tasks, yielding an AUC-ROC between 66% and 82%, respectively.

Despite extensive efforts in developing predictive models for prolonged length of stay in EDs, a key challenge remains the lack of a standardized definition for pLOS, with thresholds varying considerably across studies (3-24 hours). This inconsistency hinders the comparability of research findings, making it difficult to draw definitive conclusions about the effectiveness of different prediction models. Our study aims to address this gap by applying ML models to predict pLOS across multiple thresholds, thereby assessing the impact of choosing the most appropriate cut-offs. Notably, our approach leverages readily available data obtained after nursing assessment, such as demographics, triage codes, ICD-9 codes, and ED status metrics.

MATERIALS AND METHODS

This retrospective study was conducted in collaboration with Ospedale di Sassuolo, a secondary-level hospital located in Emilia-Romagna, Italy. The facility comprises 19 clinical units, whose Emergency and Admission Department records an average of 40,000 annual visits.

We examined a dataset of 32,967 encounters, representing 24,829 unique patients discharged between November 2022 and January 2024. All visitors had a minimum documented length of stay in the ED of one hour. Data included information on patient demographics, clinical features, and access details.

To mitigate potential biases and ensure data quality and integrity, we excluded patients deceased upon arrival, visitors who left without being seen (0.27% of the dataset), and sixteen outliers with stays exceeding the 99.95th percentile of the LOS distribution (corresponding to 54 hours).

The dataset was enriched with historical information from prior encounters and fundamental ED metrics. These encompassed patient-specific factors, such as the number of visits within the previous month (particularly those requiring SSO or involving critical conditions) and their associated LOS statistics. ED factors included current patient load (categorized by triage code), admission trends over the preceding 24 hours, and the average LOS for all patients admitted within the last 12 hours (stratified by triage code). Triage codes followed the Italian five-tier numerical priority system, ranging from 1 (most critical) to 5 (non-urgent), with a corresponding color-coded scheme (1=Red, 2=Orange, 3=Light Blue, 4=Green, 5=White). This system provides maximum waiting times indications for care pathway initiation, which vary depending on the priority code, with immediate access granted for emergencies and up to 240 minutes for non-urgent cases. Notably, these waiting times represent only a portion of the overall length of stay.

Datasets and Models Development

Two distinct datasets were employed for this study. Dataset A, the primary dataset, consisted exclusively of structured data, including patient demographic, clinical details, and admission-related information. Dataset B, which extended Dataset A, incorporated supplementary features derived from free-text triage notes documented by nursing personnel. To extract potentially valuable information from triage notes, a text preprocessing pipeline was implemented. This process involved removing common stop words and non-specialized terms, tokenizing the text into manageable units, and generating embeddings using a pre-trained Italian BERT base model. Principal Component Analysis (PCA) was applied to the embeddings in Dataset B to reduce dimensionality to 100 components. A data preprocessing pipeline was employed to normalize all numerical variables using a StandardScaler and to one-hot encode categorical features. This ensured all features were on a comparable scale for model training.

A series of twelve classification algorithms was utilized to forecast ED-LOS as a multi-class target, categorized into intervals of 1-4 hours, 5-8 hours, and greater than 8 hours. The same classifiers were subsequently used to predict pLOS, using four distinct thresholds: 4 hours, 6 hours, 8 hours, and 12 hours.

To enhance predictive accuracy and mitigate biases associated with individual models, ensemble methods such as Voting and Stacking were also implemented.

Both datasets A and B were randomly partitioned, allocating 80% (n=26,373) to the training set and the remaining 20% (n=6,594) to a holdout validation set. Model evaluation employed a five-fold cross-validation strategy to identify top-performing classifiers, followed by systematic hyperparameter tuning using a grid search approach. Classification performance was assessed using a comprehensive suite of metrics including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUROC). Final model performance was evaluated on the independent test set.

Classification Models

Adhering to international guidelines for patients presenting to the Emergency Department, the Italian Ministry of Health recommends establishing a clinical diagnosis within a maximum of four hours. Additionally, for patients requiring hospitalization, the total length of stay in the ED should not exceed eight hours from the time of initial evaluation. These directives apply even in cases of complex clinical presentations and aim to facilitate effective management of the diagnostic-therapeutic pathway.

In alignment with these benchmarks, our study stratified ED admissions into three distinct LOS categories: less than 5 hours (26,392 admissions), between 5 and 8 hours (5,205 admissions), and greater than 8 hours (1,370 admissions).

Binary Classification Models

To further classify ED stays as either "short" or "prolonged", we also investigated four distinct thresholds for pLOS: 4, 6, 8, and 12 hours.

The significant disproportion in the dataset between the majority class (patients with "short" length of stay) and the minority class (patients with prolonged LOS) posed a potential challenge. In such scenarios, models may overfit the majority class and struggle to accurately identify instances in the minority class, leading to misleading accuracy metrics. To address this issue, both ADASYN and SMOTE techniques were evaluated to balance the class distribution by synthesizing new data points for the minority class. However, no significant improvement was observed following oversampling, likely because ensemble methods, which combine predictions from multiple models, can effectively mitigate the impact of class imbalance.

To maintain methodological coherence, pLOS prediction leveraged the same battery of classifiers and evaluation metrics employed for the multi-class LOS task.

RESULTS

Classification Models

Among the classifiers employed to forecast LOS on Dataset A, StackingClassifier exhibited superior performance (accuracy 88.60%, F1-score 87.86%, AUROC 94.16%). The LGBMClassifier, XGBClassifier, and CatBoostClassifier trailed closely behind. As for Dataset B, StackingClassifier again emerged as the top performer (accuracy 87.64%, F1-score 86.77%, AUROC 93.51%).

Table 1. Results of tuned models on datasets A and B (multi-class classification task).

	Dataset A				**Dataset B**			
Model	**Acc.**	**F1**	**A.ROC**	**A.PRC**	**Acc.**	**F1**	**A.ROC**	**A.PRC**
Stacking	**0,886**	**0,879**	**0,942**	**0,820**	**0,876**	**0,868**	**0,935**	**0,802**
LGBM	0,880	0,871	0,940	0,811	0,872	0,862	0,935	0,798
XGB	0,878	0,868	0,939	0,810	0,874	0,864	0,934	0,800
CatBoost	0,875	0,866	0,935	0,805	0,876	0,867	0,933	0,797
Voting Soft	0,877	0,865	0,938	0,812	0,873	0,860	0,933	0,802
GB	0,874	0,864	0,932	0,786	0,873	0,862	0,930	0,784
MLP	0,860	0,858	0,911	0,750	0,858	0,855	0,912	0,745
RF	0,867	0,851	0,929	0,789	0,849	0,823	0,909	0,745
LR	0,866	0,849	0,911	0,733	0,865	0,847	0,909	0,731
DT	0,863	0,848	0,910	0,727	0,864	0,849	0,909	0,725
KNN	0,844	0,813	0,864	0,676	0,834	0,796	0,844	0,643
AdaBoost	0,842	0,812	0,909	0,695	0,842	0,812	0,909	0,695

Legend: LightGBM (LGBM), XGBoost (XGB), Gradient Boost (GB), Multi-layer Perceptron (MLP), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbors (KNN).

While accuracy is a less reliable metric for imbalanced datasets, as in the current case, the Area Under the Precision-Recall Curve (AUPRC) effectively captures the trade-off between precision and recall, a critical consideration when modeling infrequent yet impactful events like extended stays.

Notably, the inclusion of embedded representations derived from free-text notes did not result in performance improvement. This unexpected outcome may be attributed to the weaknesses of BERT in capturing the nuances of medical language specific to ED triage notes. These limitations could manifest as misinterpretations of abbreviations, medical jargon, and local conventions within the notes, hindering the model's ability to fully capture the semantic meaning.

The feature importance analysis conducted using the XGBoost classifier on Dataset A identified several key factors influencing ED-LOS. ED occupancy was a significant contributor to extended LOS. Treatment modality was also influential: general treatment, commonly involving more extensive procedures, contrasted with Fast-track processing (designed for less severe cases), which generally correlated with shorter stays. Temporal factors, including admission hour and day of the week, were found to affect LOS, potentially due to variations in staffing levels, diagnostic service availability, and overall workload within the ED. Patient age was a critical determinant, with pediatric cases frequently receiving prioritized treatment, which impacted LOS. Specific diagnoses and conditions, such as orthopedic issues, gastrointestinal complaints, and lower back pain, required extensive evaluation, resulting in longer stays. Cardiac-related conditions, like angina, also necessitated thorough assessment, contributing to prolonged LOS. Triage levels were another important factor, with higher severity categories (e.g., orange, red) typically entailing more comprehensive care, thus increasing LOS. Insurance status, particularly coverage by the Italian National Institute for Insurance against Accidents at Work (INAIL), correlated with specific injury types and treatment protocols, leading to extended stays. Patients suspected of COVID-19 infection also experienced longer LOS due to additional testing and isolation measures.

Importantly, upon applying the same analysis to Dataset B, PCA components – derived from embedded triage text notes – started to emerge as significant predictors. While PCA is a powerful technique for addressing the curse of dimensionality, it inherently sacrifices model interpretability by projecting the original feature space onto a lower-dimensional subspace. Consequently, the linear combinations of variables often lack a direct correspondence to the underlying, human-interpretable features.

Binary Classification Models

When assessing prediction of pLOS using a four-hour cutoff on Dataset A (Table 2), the XGBClassifier exhibited superior performance, achieving the highest AUPRC, F1-Score, and AUROC. Given the moderately imbalanced dataset (20% positive class), the XGBClassifier proved optimal.

Table 2. Results of top-performing models on Dataset A (4h cutoff).

Model	Acc.	Precision	Recall	F1	AUROC	AUPRC
XGB	**0,906**	**0,826**	**0,670**	**0,740**	**0,938**	**0,843**
CatBoost	0,906	0,834	0,659	0,736	0,938	0,842
Stacking	0,906	0,834	0,659	0,736	0,938	0,842
LGBM	0,904	0,821	0,660	0,732	0,938	0,841
Voting Soft	0,904	0,833	0,646	0,728	0,937	0,836

As the pLOS cutoff extended to six hours, class imbalance intensified (8.3% positive class). The LGBMClassifier achieved the highest F1-Score and closely approached the top performance in AUPRC and AUROC (Table 3).

Table 3. Results of top-performing models on Dataset A (6h cutoff).

Model	Acc.	Precision	Recall	F1	AUROC	AUPRC
LGBM	**0,958**	**0,860**	**0,585**	**0,696**	**0,963**	**0,808**
XGB	0,957	0,868	0,565	0,684	0,964	0,812
CatBoost	0,955	0,848	0,563	0,677	0,961	0,806
Stacking	0,955	0,848	0,563	0,677	0,961	0,806
Voting Soft	0,955	0,870	0,539	0,666	0,963	0,807

For the eight-hour cutoff, characterized by severe class imbalance (4.2% positive class), the XGBClassifier again excelled, yielding the best F1-score and AUROC, while maintaining a competitive AUPRC (Table 4).

Table 4. Results of top-performing models on Dataset A (8h cutoff).

Model	Acc.	Precision	Recall	F1	AUROC	AUPRC
XGB	**0,980**	**0,892**	**0,602**	**0,719**	**0,978**	**0,808**
LGBM	0,980	0,887	0,602	0,717	0,977	0,811
CatBoost	0,978	0,883	0,551	0,679	0,975	0,804
Stacking	0,978	0,883	0,551	0,679	0,975	0,804
GB	0,978	0,898	0,544	0,677	0,969	0,784

The twelve-hour cutoff scenario resulted in an extremely imbalanced dataset, with only 2.3% of instances belonging to the positive class. Despite all top models achieving high accuracy (0.991), this metric was deemed unreliable for model selection. AUPRC was prioritized as the most critical indicator, reflecting sensitivity to minority class prediction. XGBClassifier once again demonstrated superior overall performance (Table 5).

Table 5. Results of top-performing models on Dataset A (12h cutoff).

Model	Acc.	Precision	Recall	F1	AUROC	AUPRC
XGB	**0,991**	**0,896**	**0,687**	**0,777**	**0,991**	**0,857**
LGBM	0,991	0,925	0,660	0,770	0,991	0,862
Voting Soft	0,991	0,951	0,647	0,770	0,990	0,848
MLP	0,990	0,815	0,707	0,757	0,985	0,819
CatBoost	0,990	0,914	0,640	0,753	0,987	0,832

The XGBClassifier consistently demonstrated robust performance across all pLOS thresholds (4, 6, 8, and 12 hours). While other models, such as LGBMClassifier and VotingClassifierSoft, performed well in specific scenarios, the robustness of XGBClassifier, particularly in handling highly imbalanced datasets, is noteworthy and suggests an optimal balance between identifying patients at risk of prolonged stays (high recall) and reducing unnecessary interventions for patients not truly at risk (high precision).

In addition, calibration curves were generated for the XGBClassifier at each cutoff to assess the alignment between predicted probabilities and observed outcomes.

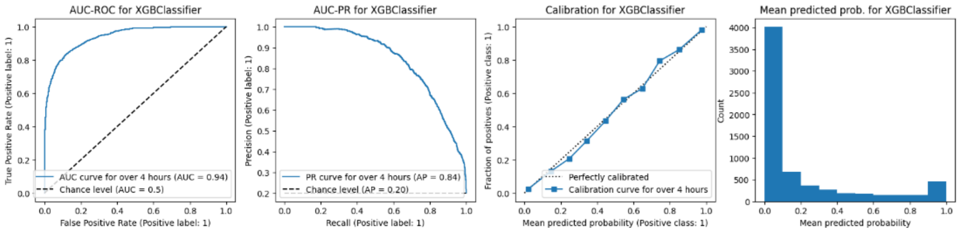

Figure 1: AUROC, AUPRC, and calibration plot for XGBClassifier on Dataset A (4h cutoff).

Given that an ideal model would exhibit a diagonal calibration curve (reflecting perfect concordance between predicted probabilities and observed event frequencies), XGBClassifier proved well-calibrated for the four-hour cutoff in Dataset A, with plot points closely clustering around the ideal diagonal line (Figure 1). However, as the threshold increased, the calibration curve worsened. Overestimation of mid to high-range probabilities became evident at the eight-hour cutoff, escalating to severe miscalibration at the twelve-hour cutoff. These findings indicate that, in the examined context, the XGBClassifier is most reliable for short-term (4 and 6 hours) than longer-term resource planning.

Consistent with findings from the multi-class classification, incorporating embeddings from triage notes (Dataset B) did not improve results.

DISCUSSION

This study employed a comprehensive machine learning approach to explore factors influencing ED-LOS and develop predictive models. A comparative analysis of twelve algorithms revealed the superiority of ensemble models in accurately forecasting LOS across three distinct categories (1-4 hours, 5-8 hours, >8 hours) and predicting the likelihood of prolonged stays using four predefined thresholds (4, 6, 8, and 12 hours). By aggregating multiple base learners, ensembles effectively mitigate bias and improve overall predictive capability (Valentini et al. 2002). Specifically, tree-based ensemble models provide a transparent decision-making process, thereby facilitating adoption in clinical practice. While nearly all ensemble models exhibited high accuracy for the investigated ED, the XGBoost classifier demonstrated superior consistency and a well-balanced trade-off between precision and recall.

The inclusion of queue-based indicators (both global and triage code-specific), representing real-time and historical ED status, significantly bolstered the models' predictive capacity, underscoring the importance of considering operational contexts in LOS forecasting. Unexpectedly, embeddings derived from free-text triage notes had minimal influence on model performance within the studied ED, contrasting our previous findings on hospital admissions (Perliti et al. 2024). This discrepancy may be partly attributed to the inclusion of ICD-9 codes as features in the ED models, which likely capture relevant diagnostic information, thus diminishing the added value of textual embeddings. Our prior hospital LOS models did not include diagnosis codes, rendering free-text notes more impactful.

Although this study focused on a specific ED, the underlying methodology exhibits potential for broader applicability across diverse ED settings, requiring minimal site-specific adjustments. Notably, while Italian guidelines recommend static triage code-based thresholds – primarily focused on initial nursing assessment and representing only a limited portion of the ED experience – our model’s ability to forecast the entire LOS offers a more comprehensive and adaptive strategy for ED management.

This research acknowledges certain limitations. The retrospective nature of the analysis may introduce selection bias, and the limited feature set, despite favoring readily available data, excludes potentially valuable information such as vital signs and laboratory test results. While incorporating these additional data points could refine predictions, it might limit the model’s transferability, presenting a trade-off that requires careful consideration. Additionally, the monocentric nature of the study, relying on data from a single ED, restricts external validation. However, it is worth noting that ED-LOS is also influenced by operational workflows, regional governance, and available resources that may vary across facilities. Consequently, comprehensive generalizability may not be entirely feasible in this context.

CONCLUSION

According to the National Agency for Regional Health Services (AGENAS), Italian Emergency Departments recorded a total of 18.27 million visits in 2023, representing a 6% increase compared to 2022. As patient volumes rise, staff efficiency diminishes, resulting in extended waiting times. Moreover, overcrowding in EDs, often exacerbated by organizational inefficiencies and personnel shortages, contributes to patients leaving without receiving medical attention and to elevated patient mortality and morbidity rates.

Predictive analytics, particularly when integrated into the triage process, offer a viable solution to alleviate these challenges. Early identification and monitoring of patients at risk for prolonged LOS allows ED administrators to initiate timely interventions, such as adjusting staffing levels and refining patient triage and assessment protocols. For example, anticipating which patients are likely to benefit from SSO units may enhance operational planning.

In the contemporary landscape of constrained public funding, optimizing ED operations, preserving assets, and mitigating costs are paramount. The proposed data-driven framework has the potential to improve resource management and promote proactive, targeted actions, particularly for patients with severe conditions. Furthermore, the inference model can be readily embedded in many EDs, leveraging their unique data patterns to implement tailored strategies to reduce prolonged stays, thereby contributing to the sustainability and resilience of public healthcare systems over generations.

REFERENCES

Azari, A., Janeja, V. P., & Levin, S. (2015). Imbalanced learning to predict long stay Emergency Department patients. In 2015 IEEE international conference on bioinformatics and biomedicine (BIBM) (pp. 807-814). IEEE.

Baugh, C. W., Venkatesh, A. K., & Bohan, J. S. (2011). Emergency department observation units: a clinical and financial benefit for hospitals. Health care management review, 36(1), 28-37.

d'Etienne, J. P., Zhou, Y., Kan, C., Shaikh, S., ... & Wang, H. (2021). Two-step predictive model for early detection of emergency department patients with prolonged stay and its management implications. The American Journal of Emergency Medicine, 40, 148-158.

Daly, S., Campbell, D. A., & Cameron, P. A. (2003). Short-stay units and observation medicine: a systematic review. Medical Journal of Australia, 178(11), 559-563.

Etu, E. E., Monplaisir, L., Arslanturk, S., Masoud, S., Aguwa, C., Markevych, I., & Miller, J. (2022). Prediction of length of stay in the emergency department for COVID-19 patients: A machine learning approach. IEEE Access, 10, 42243-42251.

Gaeta, S., Edwards, T., Bourenane, S., Gonzalez, C. E., McFarland, K., & Milling, D. (2019). Emergency department surge and overcrowding: An interdisciplinary solution for an institutional issue.

Kenny, J. F., Chang, B. C., & Hemmert, K. C. (2020). Factors affecting emergency department crowding. Emergency Medicine Clinics, 38(3), 573-587.

Liew, D., Liew, D., & Kennedy, M. P. (2003). Emergency department length of stay independently predicts excess inpatient length of stay. Medical Journal of Australia, 179(10), 524-526.

Manno, E., Pesce, M., Stralla, U., Festa, F., Geninatti, S., Balzarro, M. F., ... & Gelain, B. (2015). Specialized fast track: A sustainable model to improve emergency department patient flow. J. Hosp. Adm, 4, 40.

Naemi, A., Schmidt, T., Mansourvar, M., Ebrahimi, A., & Wiil, U. K. (2021). Quantifying the impact of addressing data challenges in prediction of length of stay. BMC Medical Informatics and Decision Making, 21, 1-13.

Perliti, P., Giovanetti, A., Bolelli, F., & Grana, C. (2024). Sustainable Use of Resources in Hospitals: A Machine Learning-Based Approach to Predict Prolonged Length of Stay at the Time of Admission. In Proceedings of 12th IHIET International Conference.

Phillips, J. L., Jackson, B. E., Fagan, E. L., Arze, S. E., Major, B., Zenarosa, N. R., & Wang, H. (2017). Overcrowding and its association with patient outcomes in a median-low volume emergency department. Journal of clinical medicine research, 9(11), 911.

Phillips, K., Knowlton, M., & Riseden, J. (2022). Emergency department nursing burnout and resilience. Advanced emergency nursing journal, 44(1), 54-62.

Rahman, M.A., Honan, B., Glanville, T., Hough, P. and Walker, K. (2020), Using data mining to predict emergency department length of stay greater than 4 hours. Emergency Medicine Australasia, 32: 416-421.

Ricciardi, C., Marino, M. R., Trunfio, T. A., Majolo, M., Romano, M., Amato, F., & Improta, G. (2024). Evaluation of different machine learning algorithms for predicting the length of stay in the emergency departments: a single-centre study. Frontiers in Digital Health, 5, 1323849.

Sarıyer, G., Ataman, M. G., & Kızıloğlu, İ. (2020). Factors affecting length of stay in the emergency department: a research from an operational viewpoınt. International Journal of Healthcare Management, 13(sup1), 173-182.

Savioli, G., Ceresa, I. F., Gri, N., Bavestrello Piccini, G., Longhitano, Y., Zanza, C., ... & Bressan, M. A. (2022). Emergency department overcrowding: understanding the factors to find corresponding solutions. Journal of personalized medicine, 12(2), 279.

Valentini, G., & Masulli, F. (2002). Ensembles of learning machines. In Neural Nets: 13th Italian Workshop on Neural Nets, WIRN VIETRI 2002 Vietri sul Mare, Italy, May 30–June 1, 2002, Revised Papers 13 (pp. 3-20). Springer Berlin Heidelberg.

Walsh, P., Cunningham, P., Rothenberg, S. J., O'Doherty, S., Hoey, H., & Healy, R. (2004). An artificial neural network ensemble to predict disposition and length of stay in children presenting with bronchiolitis. European Journal of Emergency Medicine, 11(5), 259-264.

Yoon, P., Steiner, I., & Reinhardt, G. (2003). Analysis of factors influencing length of stay in the emergency department. Canadian Journal of Emergency Medicine, 5(3), 155-161.