Human Interaction and Emerging Technologies (IHIET 2024), Vol. 157, 2024, 603–613
https://doi.org/10.54941/ahfe1005520

Sustainable Use of Resources in Hospitals: A Machine Learning-Based Approach to Predict Prolonged Length of Stay at the Time of Admission

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

Introduction: Length of Stay (LOS) and Prolonged Length of Stay (pLOS) are critical indicators of hospital efficiency. Reducing pLOS is crucial for patient safety, autonomy, and bed allocation. This study investigates different machine learning (ML) models to predict LOS and pLOS.
Methods: We analyzed a dataset of patients discharged from a northern Italian hospital between 2022 and 2023 as a retrospective study. We compared sixteen regression algorithms and twelve classification methods for predicting LOS as either a continuous or multi-class variable (1–3 days, 4–10 days, >10 days). We also evaluated pLOS prediction using the same models, having pLOS defined as any hospitalization with LOS longer than 8 days. We further analyzed all models using two versions of the same dataset: one containing only structured data (e.g. demographics and clinical information), whereas the second one also containing features extracted from free-text diagnosis.
Results: Our results indicate that ensemble models achieved the highest prediction accuracy for both LOS and pLOS, outperforming traditional single-algorithm models, particularly when using both structured and unstructured data extracted from diagnoses.
Discussion: The integration of ML, particularly ensemble models, can significantly improve LOS prediction and identify patients at increased risk of pLOS. This information can guide healthcare professionals and bed managers in making informed decisions to enhance patient care and optimize resource allocation.

Keywords: LOS, pLOS, Machine learning, Hospital admissions, Public healthcare, Sustainability

INTRODUCTION

Over the past twenty years, the number of beds in Italian hospitals has decreased by 30 percent (OECD, 2021) while bed occupancy has notably increased, causing congestion and lengthening patient stays. With 11.6 beds per 100,000 inhabitants in 2021, Italy still maintains a provision below the Organisation for Economic Cooperation and Development average of 16.9 (OECD, 2023). Moreover, a significant escalation in national health expenditure, from approximately 80 billion euros in 2002 to 129 billion euros in 2022 (Italian Ministry of Economy and Finance, 2023), with 20% attributed to inpatient expenses, has placed hospitals under considerable pressure to implement effective cost containment strategies. Several trends, including the rise of patient volumes and changing working practices (as the COVID-19 pandemic forced to do), contribute to the complexity of managing patient flow. Furthermore, an aging population, higher co-morbidity rates, and delayed discharges (especially into nursing and residential care) add to the challenges.

Current bed modeling techniques, often based on midnight occupancy census, lack the information needed for effective space management. At the same time, the advent of AI and Machine Learning (ML) may assist bed managers in their daily routine, offering valuable tools.

BACKGROUND

The Length of Stay (LOS), represented as the interval time between admission and discharge (i.e., total bed-days occupied by a patient), plays a fundamental role in evaluating the quality of care services.

In the literature, it has been demonstrated that LOS does have correlations to disease severity, readmission rates, and mortality (Marfil-Garza et al., 2018). Reducing LOS in public healthcare systems (i.e. enabling early discharge and fast turnover) benefits both patients – preventing complications, limiting the risk of adverse events (Ackroyd-Stolarz et al., 2011) such as falls, thrombosis, drug reactions and hospital-acquired infections, improving patient autonomy (Hauck and Zhao, 2011) – and hospitals – optimizing treatment plans and resource utilization (such as bed allocation) (Molloy et al., 2017), up to maintaining control over the growth of waiting lists. A prolonged Length of Stay (pLOS), in fact, is associated with cognitive impairment, functional limitations, and elevated burdens of comorbidity (Bo et al., 2016), and it generally leads to cancellations of elective operations, additional use of resources and thus increased medical costs (especially in ICU units); pLOS could also adversely affect the admission process for critically ill patients and hinder timely access to treatment. Moreover, a small percentage of patients with pLOS could consume a large proportion (up to 50%) of the limited resources available (Evans et al., 2018).

With an aging population, the anticipated increase of LOS underscores the urgency of responsive healthcare planning to ensure safety, satisfaction, and accessibility. This is particularly evident in Italy, where demographic shifts like rising life expectancy (reaching 80.6 years for men and 84.8 for women in 2022) (ISTAT, 2022) contribute to the prevalence of chronic and degenerative diseases. Identifying inpatients with extended stays – often referred to as 'bed-blockers' – necessitate early recognition for appropriate treatment planning, including potential admission to long-stay wards. Hence, prolonged LOS serves as a key metric, being associated with escalated costs and reduced capacity.

The present study aimed to extensively evaluate ML-based models to predict both LOS and pLOS for general patients. A comprehensive comparison among these methods was performed to assess their respective capabilities. Additionally, an investigation was conducted to identify the most relevant features for predicting LOS and pLOS.

RELATED WORKS

Over the last two decades, various prediction models employing statistical techniques have been developed to investigate LOS and the influence of covariates such as age, gender, diagnosis, illness severity, type of admission, and hospital characteristics. Nevertheless, the use of Machine Learning and Deep Learning (DL) has recently gained attention in health service research as a viable alternative to those established methodologies (Gholipour et al., 2015; Barnes et al., 2016). There is a quite heterogeneous literature exploring LOS patterns (Stone et al., 2022), often focusing on broader patient cohorts (Mekhaldi et al., 2020), with particular emphasis on limited age intervals (Thompson et al., 2018; Hesselink et al., 2019), explicit discipline areas (Chen et al., 2023) and medical specialties (e.g., cardiology, Daghistani et al., 2019), surgical procedures (Chuang et al., 2018) and cancer surgeries (Jo et al., 2021), and specific medical units.

Unfortunately, only a minority of these reviews take Italian public health into consideration. In a study by Zeleke et al. (2023) six classification algorithms were developed to study the prediction of pLOS for 12,858 inpatients admitted through the emergency department (ED) of an Italian hospital ("Sant'Orsola-Malpighi University Hospital", Bologna). The pLOS threshold was defined as any stay longer than the average LOS (6 days). The authors also developed eight regression models for LOS prediction. The Gradient Boosting classifier best predicted pLOS (accuracy 75%, AUC 75.4%). Ridge and XGBoost regressors best predicted LOS, with an overall prediction error between 6 and 7 days.

Trunfio et al. (2022) analyzed LOS for 2,515 patients undergoing hip-replacement surgery at the "San Giovanni di Dio e Ruggi d'Aragona" University Hospital of Salerno. Several regression and classification algorithms were implemented in order to predict the total length of stay. The results from the regression analysis showed that the best model was Multiple Linear Regression (R^2 0.616). As for classification analysis, LOS was divided into 3 classes (LOS <= 6 days, 6 days < LOS <= 12 days, LOS > 12 days). In terms of overall accuracy, Random Forest achieved a value of 71.76%.

Olivato et al. (2022) developed a machine learning-driven system to forecast pLOS for COVID-19 patients. Their model, trained on demographic information and laboratory test values from over 6,000 patients admitted to the Spedali Civili di Brescia, in northern Italy, achieved a ROC-AUC of 0.76.

In another study conducted at A.O.R.N. "Antonio Cardarelli", Naples (Italy) (D'Onofrio et al. 2023), the Electronic Medical Records (EMRs) of 989 patients who underwent mastectomy surgery were used to predict LOS as a binary outcome. Random Forest showed the best accuracy (77.79%).

Age and the presence of comorbidities such as hypertension, diabetes, obesity, and tumor stage were considered factors affecting the LOS.

Di Matteo et al. (2023) implemented a custom neural network to forecast the length of stay (as a binary outcome) for patients undergoing hip or knee arthroplasty in "Humanitas Research" Hospital, Milan (Italy). The model, trained on a dataset of 1,517 patients, leveraged a combination of clinical and textual data. They achieved an AUC of 0.789 in predicting "Short LOS" (<= 6 days) over "Long LOS" (> 7 days).

Altogether, these findings suggest that the ML approach could assist hospital systems in forecasting and addressing bed capacity requirements. However, most of these studies focus only on a specific department or rely on data not readily available at admission (such as lab results).

Our research employs a variety of supervised ML algorithms to predict the length of stay for patients in general inpatient settings, examining LOS as a continuous, multi-class, and dichotomous variable using data extracted from a medico-administrative platform. This investigation intentionally encompasses all medical-surgical departments with the aim of developing robust and adaptable models for effective generalization. This decision is supported by the observation that when a department reaches full capacity, patients are relocated to alternate wards with available beds (often regardless of their primary service). Consequently, evaluating all medical units collectively provides greater consistency. Additionally, our analysis solely considers data available at the time of admission. While this approach may potentially overlook a substantial portion of clinical data, it ensures good performance from the very beginning of hospitalization and can be implemented across diverse hospital settings.

MATERIALS AND METHODS

The present study examines the departments of a general hospital located in Emilia-Romagna, Italy. The facility is organized by intensity care and structured around 19 clinical units. We analyzed a dataset of 12,471 hospitalizations from 10,145 unique patients discharged between February 2022 and November 2023, with a length of stay of at least 24 hours. Data were extracted from an EBMS (Electronic Bed Management System) which included information on patient demographics, admission type, clinical features, and hospitalization features.

Patients undergoing Day Surgery or Day Hospital procedures were excluded from the analysis due to their predetermined LOS. This exclusion aimed to mitigate potential biases in the model's performance arising from an overrepresentation of cases with a fixed LOS of one day. To ensure the quality and integrity of the data, we also excluded patients deceased during hospitalization, inpatients with LOS values exceeding the 99.95th percentile of the LOS distribution (outliers), and maternity/infancy wards due to their distinct clinical characteristics and potential data collection biases.

Models Development

The study employed two versions of the dataset: the first one (A) containing only structured data (demographics, clinical information, admission details) and the second one (B) incorporating features extracted from unstructured free-text diagnoses documented by practitioners.

Initially, fourteen regression algorithms were employed for predicting LOS as a continuous value. Performance was evaluated using mean absolute error (MAE), root mean squared error (RMSE), R-squared (R^2), and adjusted R-squared scores. Additionally, ten classification methods were utilized to predict LOS as a multi-class target (1–3 days, 4–10 days, >10 days). Evaluation metrics included accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUROC). The same ten classification models were also used to predict pLOS, defined as any hospitalization longer than 8 days, corresponding to the 75th percentile.

To extract features from diagnoses, a text cleaning process was implemented, removing stop words and irrelevant or non-domain-specific terms. A pre-trained BERT model was then employed to tokenize the text and generate embeddings. Moreover, principal component analysis (PCA) was applied to embeddings from dataset B to reduce the dimensionality of the data to 100 components, thereby decreasing computational complexity.

Each dataset was randomly partitioned into a training set comprising 9,976 admissions (80%) and a holdout/validation set encompassing 2,495 admissions (20%). A five-fold cross-validation approach was employed for each task to compare algorithms and identify the top performers. Hyperparameter tuning was then performed for each model. Finally, performance was assessed on the test set.

Additionally, a Voting ensemble and a Stacking ensemble were included in the final evaluation. These methods aggregate predictions from multiple base models (using voting and stacking aggregation respectively) to improve overall accuracy and reduce bias, yielding a superior ensemble model.

RESULTS

Regression Models

In predicting LOS as a continuous variable (Table 1), the ensemble Stacking-Regressor demonstrated the highest accuracy for dataset A (MAE 2.81, R^2 score 0.635), followed by VotingRegressor (MAE 2.82, R^2 score 0.634) and XGB-Regressor (MAE 2.84, R^2 score 0.632). When considering datasets containing unstructured data (B), CatBoost outperformed other models in predicting LOS (MAE 2.73, R^2 score 0.649), followed closely by Voting-Regressor (MAE 2.72, R^2 score 0.647) and XGBRegressor (MAE 2.78, R^2 score 0.639).

Notably, incorporating embedded representations derived from free-text diagnoses led to a modest yet noticeable performance improvement across various models. This enhancement can be attributed, in part, to the ability of embeddings to capture the semantic meaning of diagnoses, which can be challenging to represent using traditional structured features. Additionally, embeddings offer the advantage of modeling the relationships between different diagnoses, which becomes particularly important in the presence of comorbidities. It is noteworthy that including the admitting diagnosis does not introduce bias or confound the study endpoint (e.g., data leakage), given its inherent availability at the time of hospitalization.

Table 1. Results of tuned models on datasets A and B (regression task).

	Dataset A				Dataset B			
Model	MAE	RMSE	R^2	Ad. R^2	MAE	RMSE	R^2	Ad. R^2
Stacking Regressor	**2.806**	**4.614**	**0.635**	**0.633**	2.705	4.622	0.633	0.632
Voting Regressor	2.824	4.617	0.634	0.633	2.722	4.537	0.647	0.645
XGB Regressor	2.844	4.634	0.632	0.630	2.776	4.585	0.639	0.638
CatBoost Regressor	2.831	4.639	0.631	0.629	**2.726**	**4.520**	**0.649**	**0.648**
Linear Regression	2.976	4.692	0.622	0.621	2.911	4.622	0.633	0.632
Ridge	2.963	4.706	0.620	0.618	2.862	4.620	0.634	0.632
GB Regressor	2.946	4.723	0.617	0.616	2.923	4.739	0.615	0.613
LGBM Regressor	2.903	4.770	0.609	0.608	2.756	4.616	0.634	0.633
Elastic-net	2.995	4.776	0.609	0.607	2.887	4.679	0.624	0.623
SVR	2.883	4.784	0.607	0.606	2.792	4.694	0.622	0.620
Lasso	3.051	4.867	0.594	0.592	3.004	4.811	0.603	0.601
RF Regressor	2.968	4.903	0.588	0.586	2.843	4.787	0.607	0.605
KNN Regressor	2.989	5.012	0.569	0.567	2.868	4.957	0.578	0.577
AdaBoost Regress.	3.601	5.275	0.522	0.521	3.576	5.285	0.521	0.519
MLP Regressor	3.298	5.587	0.464	0.462	3.090	5.350	0.509	0.507
DT Regressor	3.882	6.617	0.249	0.246	3.757	6.604	0.252	0.249

Classification Models

Among the ensemble-based models (Table 2), VotingClassifierSoft exhibited superior performance in predicting LOS for both datasets A (accuracy 73.55%, AUROC 87.94%) and B (accuracy 76.27%, AUROC 89.60%). As observed in the regression analysis, the employment of embeddings extracted from diagnoses evidences a modest augmentation in performance.

Table 2. Results of tuned models on datasets A and B (multi-class classification task).

	Dataset A				Dataset B			
Model	Acc.	F1	ROC	PRC	Acc.	F1	ROC	PRC
Voting Soft	**0.735**	**0.732**	**0.879**	**0.777**	**0.763**	**0.760**	**0.896**	**0.794**
Stacking	0.732	0.730	0.879	0.774	0.763	0.761	0.893	0.789
CatBoost	0.733	0.729	0.878	0.776	0.758	0.754	0.894	0.788
XGB	0.728	0.724	0.875	0.775	0.762	0.760	0.892	0.789
RF	0.730	0.728	0.873	0.767	0.750	0.745	0.886	0.778
GB	0.729	0.726	0.872	0.764	0.760	0.757	0.890	0.783
LGBM	0.724	0.722	0.870	0.760	0.748	0.745	0.887	0.778
Log. Regress.	0.730	0.727	0.864	0.750	0.740	0.737	0.883	0.773
KNN	0.702	0.697	0.847	0.726	0.718	0.710	0.862	0.733
MLP	0.678	0.679	0.827	0.707	0.719	0.717	0.852	0.729
AdaBoost	0.710	0.707	0.787	0.651	0.738	0.737	0.817	0.677
DT	0.654	0.652	0.724	0.519	0.655	0.655	0.720	0.513

CatBoost was also employed to determine the relative significance of features (Figure 1). Analyzing dataset A, the most salient features in descending order of importance were: the overall average length of stay for same-service hospitalizations within the previous 30 days (*ba_specialty_prev_hosp_avg_los*); the number of recent transfers across wards (*ba_recent_transfers_count*), possibly indicative of increasing medical complexity (i.e. patient cases that require multiple departments or specialized care); the surgery hospitalization area (*b_hosp_area_G*); the number of bed movements during hospitalization (*ba_movements_count*), including those within the same ward; low-intensity care level (*ba_intenscare_L*), supporting the idea that stays associated with critical conditions require only the time necessary to stabilize the patient's condition; age in range 80–89 years (*p_age_range_80-89*), suggesting that inpatients in this age group may have an increased risk of prolonged stays, potentially due to age-related vulnerabilities or comorbidities; the average LOS for same-patient hospitalizations in the prior year (*ba_patient_prev_hosp_avg_los*); and the need for a multidimensional geriatric assessment (*cs_is_uvm_req*), typically associated with elderly or frail individuals.

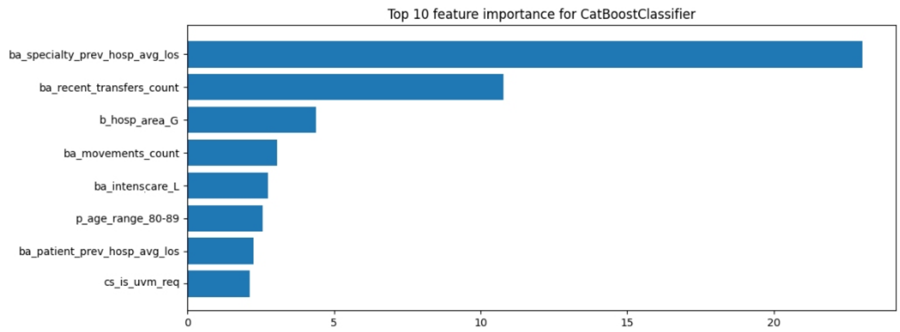

Figure 1: Feature importance for CatBoostClassifier on dataset A (multi-class classification task).

Binary Classification Models

Similar to the multi-class classification task (Table 3), Area Under the Receiver Operating Characteristic Curve (AUROC) and Area Under the Precision-Recall Curve (AUPRC) were employed to assess the performance of the models. Ranging from 0 to 1, AUROC effectively captures the trade-off between true and false positives across all possible thresholds. Conversely, AUPRC prioritizes the identification of positive samples, making it particularly advantageous for imbalanced datasets.

LogisticRegression provided the most accurate predictions for prolonged length of stay in the A dataset (accuracy 86.61%, AUROC 90.54%), followed by VotingClassifierSoft and CatBoostClassifier. On the other hand, for dataset B, VotingClassifierSoft (accuracy 86.53%, AUROC 91.67%), CatBoostClassifier (accuracy 86.69%, AUROC 91.24%), and StackingClassifier demonstrated superior predictive capabilities for pLOS.

Table 3. Results of tuned models on datasets A and B (binary classification task).

	Dataset A				Dataset B			
Model	Acc.	F1	ROC	PRC	Acc.	F1	ROC	PRC
Log. Regress.	**0.866**	**0.646**	**0.905**	**0.794**	0.866	0.654	0.911	0.798
Voting Soft	0.863	0.636	0.905	0.789	**0.865**	**0.645**	**0.917**	**0.808**
CatBoost	0.867	0.654	0.903	0.785	0.867	0.654	0.912	0.799
Stacking	0.867	0.654	0.903	0.785	0.867	0.654	0.912	0.799
GB	0.861	0.636	0.901	0.782	0.863	0.650	0.908	0.793
LGBM	0.861	0.642	0.898	0.777	0.862	0.642	0.907	0.791
AdaBoost	0.862	0.642	0.896	0.778	0.853	0.634	0.888	0.769
RF	0.863	0.635	0.893	0.777	0.857	0.584	0.904	0.788
XGB	0.842	0.606	0.884	0.751	0.862	0.655	0.901	0.785
KNN	0.848	0.545	0.884	0.743	0.862	0.516	0.874	0.726
MLP	0.827	0.618	0.867	0.733	0.862	0.625	0.874	0.745
DT	0.799	0.569	0.719	0.423	0.862	0.558	0.713	0.412

DISCUSSION

A key strength of this study lies in its comparative analysis of different data sources. The integration of text extracted from diagnoses proved instrumental in capturing subtle nuances not represented in structured data alone. This is particularly relevant to leveraging the expertise of medical staff and the outcomes derived from physical examinations. Our findings align with prior investigations (Zhang et al., 2020; Jiang et al., 2023), which have demonstrated that text-derived features may enhance predictive performance.

This study also provides further evidence of the effectiveness of machine learning ensembles in improving LOS prediction accuracy. By aggregating the outputs of multiple base learners, these ensemble models effectively reduce inherent biases and significantly improve overall predictive power. Moreover, their decision tree structure facilitates the conversion of complex models into transparent decision rules, fostering understanding among practitioners. This enhances model acceptance and facilitates integration into healthcare workflows, which is pivotal for the adoption of ML-driven systems in medical practice.

Compared to previous works focused on specific patient groups, such as hip- or knee-replacement patients or heart failure patients, our model is not limited to narrow clinical subsets and encompasses a heterogeneous patient spectrum, expanding its applicability beyond specific diagnoses or conditions. Furthermore, our approach avoids the limitations of single-ward-based studies (e.g. ICU or general medicine) that may be constrained by specific internal dynamics influencing patient outcomes. While this methodology poses a greater challenge, it also yields broader generalizability, increasing the potential for real-world implementation. Most importantly, our models leverage readily available data from institutional Electronic Health Records (EHR), collected within the first 24 hours of hospitalization, allowing bed managers to utilize them as decision-making aids promptly upon patient admission.

Despite its strengths, our study also exhibits certain limitations that should be acknowledged. Firstly, it is a retrospective study employing historical data. While it utilizes minimal historical information, such as the number of previous hospitalizations and the average LOS for same-specialty past stays, this may potentially introduce biases. Secondly, the exclusion of vital signs and laboratory test results limits the comprehensiveness of the predictor variables. While it is desirable to have input variables as general as possible, avoiding specific test or laboratory results, these vital data points could potentially enhance the prediction models' accuracy and reduce the risk of false positives (overestimation of LOS) and false negatives (underestimation of LOS).

Eventually, the monocentric nature of the study, relying on data from a single hospital, hinders external validation of the prediction models. While this may restrict their applicability to other healthcare settings, it should be recognized that LOS is also influenced by factors that may vary across facilities: striving for absolute generalizability may not be feasible or optimal in this context.

CONCLUSION

Our findings underscore the potential of ensemble-based prediction models to assist medical personnel in identifying patients who are at risk of prolonged LOS across different services and wards. By recognizing these inpatients, healthcare providers can proactively implement targeted interventions, such as close monitoring and early discharge planning, to reduce the likelihood of lengthy stays. This preemptive approach can contribute to smoother patient flow, increased bed availability, a lower rate of rescheduled interventions, improved patient satisfaction, and, ultimately, reduced overall healthcare expenditures. Furthermore, the use of readily available data from EHR in conjunction with algorithms that do not necessitate resource-intensive procedures or specialized hardware encourages the potential integration of this methodology in other settings and workflows (e.g. embedding the inference model within EBMSs as a second-opinion tool, to support both medical staff and healthcare management in their daily tasks).

These results emphasize the priority for hospitals to adapt and innovate in order to meet the demands of an aging population with chronic disorders, while containing costs and optimizing resources to improve the sustainability of public healthcare systems.

REFERENCES

Ackroyd-Stolarz, S., Guernsey, J. R., Mackinnon, N. J., & Kovacs, G. (2011). The association between a prolonged stay in the emergency department and adverse events in older patients admitted to hospital. BMJ Quality & Safety.

Barnes, S., Hamrock, E., Toerper, M., Siddiqui, S., & Levin, S. (2016). Real-time prediction of inpatient length of stay for discharge prioritization. Journal of the American Medical Informatics Association, 23(e1), e2–e10.

Bo, M., Fonte, G., Pivaro, F., Bonetto, M., Comi, C., Giorgis, V., Marchese, L., Isaia, G., Maggiani, G., Furno, E., Falcone, Y., and Isaia, G. C. (2016). Prevalence of and factors associated with prolonged length of stay in older hospitalized medical patients. Geriatrics & gerontology international, 16(3), 314–321.

Chen, R., Zhang, S., Li, J. et al. (2023). A study on predicting the length of hospital stay for Chinese patients with ischemic stroke based on the XGBoost algorithm. BMC Medical Informatics and Decision Making, 23(1), 1–10.

Chuang, M. T., Hu, Y. H., & Lo, C. L. (2018). Predicting the prolonged length of stay of general surgery patients: a supervised learning approach. International Transactions in Operational Research, 25(1), 75–90.

Daghistani, T. A., Elshawi, R., Sakr, S., Ahmed, A. M., Al-Thwayee, A., & Al-Mallah, M. H. (2019). Predictors of in-hospital length of stay among cardiac patients: a machine learning approach. International journal of cardiology, 288, 140–147.

Di Matteo, V., Tommasini, T., Morandini, P., Savevski, V., Grappiolo, G., Loppini, M. Machine Learning Prediction Model to Predict Length of Stay in Patients undergoing Hip or Knee Arthroplasties: Results from a High Volume Single Center Multivariate Analysis. Preprints 2023, 2023110915.

D'Onofrio G., D'Amore A., Onofaro F., Caputi E., Napoli A., Triassi M., Marino M. R. Prediction of Hospital Length Stay for Patients Undergoing Mastectomy. Stud Health Technol Inform. 2023 Jun 29;305: 261–264.

Evans, J., Kobewka, D., Thavorn, K., D'Egidio, G., Rosenberg, E., & Kyeremanteng, K. (2018). The impact of reducing intensive care unit length of stay on hospital costs: evidence from a tertiary care hospital in Canada. Canadian Journal of Anesthesia, 65(6), 627–635.

Gholipour, C., Rahim, F., Fakhree, A., & Ziapour, B. (2015). Using an artificial neural networks (ANNs) model for prediction of intensive care unit (ICU) outcome and length of stay at hospital in traumatic patients. Journal of clinical and diagnostic research: JCDR, 9(4), OC19.

Hauck, K., & Zhao, X. (2011). How dangerous is a day in hospital? A model of adverse events and length of stay for medical inpatients. Medical care, 1068–1075.

Health at a Glance 2023 OECD Indicators https://www.oecd-ilibrary.org/citeas/10.1787/7a7afb35-en, last accessed 2024/01/08

Hesselink, G., Van Den Bogaert, M., Akkermans, R. P., & Schoon, Y. (2019). Risk factors for prolonged length of stay of older patients in an academic emergency department: A retrospective cohort study. Emergency medicine international, 2019.

ISTAT (Istituto Nazionale di Statistica). Annual Italian Statistics 2022 (2022). http://dati.istat.it/, last accessed 2024/01/08.

Italian Ministry of Economy and Finance (2023). Rapporto n. 10: Il Monitoraggio della Spesa Sanitaria. https://www.rgs.mef.gov.it/_Documenti/VERSIONE-I/Attivit-V-i/Spesa-soci/Attivit-monitoraggio-RGS/2023/IMDSS-RS2023.pdf

Jiang, L. Y., Liu, X. C., Nejatian, N. P. et al. (2023). Health system-scale language models are all-purpose prediction engines. Nature, 1–6.

Jo, Y. Y., Han, J., Park, H. W., Jung, H., Lee, J. D., Jung, J., Cha, H. S., Sohn, D. K., Hwangbo, Y. (2021). Prediction of prolonged length of hospital stay after cancer surgery using machine learning on electronic health records: Retrospective cross-sectional study. JMIR medical informatics, 9(2), e23147.

Marfil-Garza, B. A., Belaunzarán-Zamudio, P. F., Gulias-Herrero, A., Zuñiga, A. C., Caro-Vega, Y., Kershenobich-Stalnikowitz, D., & Sifuentes-Osornio, J. (2018). Risk factors associated with prolonged hospital length-of-stay: 18-year retrospective study of hospitalizations in a tertiary healthcare center in Mexico. PloS one, 13(11).

Mekhaldi, R. N., Caulier, P., Chaabane, S., Chraibi, A., & Piechowiak, S. (2020, April). Using machine learning models to predict the length of stay in a hospital setting. In World conference on information systems and technologies (pp. 202–211). Cham: Springer International Publishing.

Molloy, I. B., Martin, B. I., Moschetti, W. E., & Jevsevar, D. S. (2017). Effects of the length of stay on the cost of total knee and total hip arthroplasty from 2002 to 2013. The Journal of bone and joint surgery. American volume, 99(5), 402.

OECD Indicators on Healthcare Resources: Hospital beds by sector between 2001 and 2021 https://stats.oecd.org/

Olivato, M., Rossetti, N., Gerevini, A. E., Chiari, M., Putelli, L., & Serina, I. (2022). Machine Learning Models for Predicting Short-Long Length of Stay of COVID-19 Patients. Procedia Computer Science, 207, 1232–1241.

Stone, K., Zwiggelaar, R., Jones, P., & Mac Parthaláin, N. (2022). A systematic review of the prediction of hospital length of stay: Towards a unified framework. PLOS Digital Health, 1(4), e0000017.

Thompson, B., Elish, K. O., & Steele, R. (2018, December). Machine learning-based prediction of prolonged length of stay in newborns. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1454–1459). IEEE.

Trunfio TA, Borrelli A, Improta G. Is It Possible to Predict the Length of Stay of Patients Undergoing Hip-Replacement Surgery? International Journal of Environmental Research and Public Health. 2022; 19(10):6219.

Zeleke AJ, Palumbo P, Tubertini P, Miglio R and Chiari L (2023) Machine learning-based prediction of hospital prolonged length of stay admission at emergency department: A Gradient Boosting algorithm analysis. Front. Artif. Intell. 6:1179226.

Zhang, D., Yin, C., Zeng, J., Yuan, X., & Zhang, P. (2020). Combining structured and unstructured data for predictive models: a deep learning approach. BMC medical informatics and decision making, 20(1), 1–11.