



HAL
open science

Early Length of Stay Prediction at Admission in Short-Stay Hospitals

Mohamed Gharbi, Christine Verdier, Maria Di Mascolo, Jean-Marc Babouchkine

► **To cite this version:**

Mohamed Gharbi, Christine Verdier, Maria Di Mascolo, Jean-Marc Babouchkine. Early Length of Stay Prediction at Admission in Short-Stay Hospitals. *Research Challenges in Information Science*, May 2025, Sevilla, Spain. pp.52-66, <10.1007/978-3-031-92471-2_4>. <hal-05210216>

HAL Id: hal-05210216

<https://hal.science/hal-05210216v1>

Submitted on 19 Mar 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Early Length of Stay Prediction at Admission in Short-Stay Hospitals.

Mohamed Gharbi^{1,2,3}[0009-0004-8628-4506], Christine Verdier²[0000-0001-9861-5617], Maria Di Mascolo³[0000-0002-8784-9685], and Jean-Marc Babouchkine¹

¹ Calystene, France

`mohamed.gharbi1@univ-grenoble-alpes.fr`

² Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

`christine.verdier@univ-grenoble-alpes.fr`

³ Univ. Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, 38000 Grenoble, France

`maria.di-mascolo@grenoble-inp.fr`

Abstract. Predicting hospital length of stay (LOS) is crucial for optimizing resource allocation and improving patient care. Accurate estimation of a patient’s discharge date enhances operational efficiency and reduces stress on hospital staff. While many studies have addressed this challenge using classification and regression models that consider data from admission to discharge, this research focuses on predicting LOS using only the features available at admission, specifically demographic details and primary diagnosis data, in a case study of patients in surgical wards dealing with musculoskeletal and connective tissue conditions. Advanced machine learning techniques, including boosting algorithms (LightGBM, XGBoost, CatBoost), Random Forest, and Feedforward Neural Networks (FNN), were employed to achieve this. LightGBM, demonstrated the best performance in this context, achieving a precision R^2 score of 0.77 and a mean squared error (MSE) of 1.07. These measures indicate the model’s ability to predict LOS with a high degree of accuracy while minimizing prediction errors.

These results highlight the potential of machine learning, particularly LightGBM, to optimize hospital operations and enhance patient outcomes. Ultimately, the findings underscore the value of leveraging admission-based data for early and accurate LOS predictions, providing a foundation for more efficient resource planning in healthcare settings.

Keywords: Machine learning · Electronic Medical Records · Length Of Stay Prediction · Regression

1 Introduction

The length of stay (LOS) is widely recognized as a key metric for assessing hospital performance and operational efficiency, as emphasized by the World Health Organization (WHO) [1]. Defined as the duration of a patient’s hospitalization for a single episode of care, LOS provides critical insights into the effectiveness

of healthcare delivery. The LOS can vary significantly across different types of hospitalization. For instance, short-stay hospitalizations, which include medical, surgical, and obstetric care, typically have a mean LOS of 6 days. These stays are characterized by unpredictability, as factors such as the patient’s state, diagnosis, comorbidity, and individual characteristics can all influence the length of hospitalization. Efficient management of LOS is crucial across managerial, medical, and financial domains, as it directly impacts resource allocation and patient outcomes. Conversely, delays in discharge or prolonged stays can disrupt workflows, strain resources, and negatively affect the experiences of both patients and staff [2]. Therefore, there is a need for accurate prediction of LOS in short-stay hospitalization to address these issues.

Predicting the length of stay (LOS) can significantly benefit hospitals by enabling more informed medical decision-making and efficient planning. Accurate LOS predictions allow hospitals to provide families with an estimated discharge date, which helps manage expectations. Furthermore, such predictions facilitate the advanced mobilization of resources, including ensuring an adequate number of nursing staff to address workload distribution challenges. Research has demonstrated the value of early LOS prediction in improving hospital operations. For instance, studies such as [3] have shown that predicting LOS can help reduce costs, enhance healthcare delivery, and enable the timely mobilization of critical resources, such as ICU beds.

Many studies have been conducted to predict LOS in short-stay hospitals, leveraging various types of information, such as demographic, clinical, and administrative data. However, most of these studies rely on data available up to the point of discharge to predict LOS. In contrast, our study aims to predict LOS using only the data available at admission.

In this study, we utilized Electronic Medical Records (EMRs), which are digital systems used by hospitals to track patients’ medical history, treatments, and demographic information. Data extracted from EMRs were employed to predict the length of stay (LOS) in short-stay hospitalizations. To achieve this, we implemented several machine learning algorithms, including ensemble models such as LightGBM, Random Forest (RF), CatBoost, and XGBoost, which are well-suited for capturing complex patterns and relationships between different types of variables while effectively handling noise [4]. Additionally, we tested Feedforward Neural Networks (FNNs) for their robustness in modeling intricate patterns and extracting insights from various variables.

This paper is organized as follows: Section 2 reviews the related work, highlighting existing approaches and methodologies for predicting the length of stay (LOS) in hospitals. Section 3 introduces the case study that serves as the basis for our research, describing the hospital setting, and the characteristics of the dataset used. Section 4 outlines our proposed approach for LOS prediction, detailing the methods and techniques employed. This section is further divided into two subsections: Section 4.1, which explains the data processing steps, including cleaning, transformation, and feature selection, and Section 4.2, which describes the machine learning models and algorithms implemented to predict

LOS. Section 4.3 presents the evaluation metrics and results, comparing the performance of different models and discussing their implications. Finally, the Conclusion summarizes the main findings, discusses the limitations of our work, and proposes directions for future research.

2 Related work

Many studies have highlighted the role of artificial intelligence (AI) in the health-care domain [5], particularly in resource management, to address unpredictable situations such as hospital overcrowding, surgery cancellations or delays, and workload management. These challenges can exacerbate difficulties in hospital operations. One solution that has been explored is the prediction of hospital length of stay (LOS) [6].

LOS prediction has been applied across diverse clinical contexts, including lung cancer, neonatal intensive care units (NICUs), ankle arthroplasty, bed census forecasting, and patients undergoing coronary artery bypass graft (CABG) surgery. A variety of machine learning algorithms have been employed, ranging from classification methods and ensemble models to neural networks (NN), with a primary focus on classification tasks. The most commonly used models were k-nearest neighbors (KNN), Random forest (RF), Decision Tree, Support Vector Machine (SVM), Linear Regression (LR), XGBoost, and neural networks (NNs). For instance, a study reported that the random forest classification algorithm achieved the best performance with an accuracy of 97% for binary length-of-stay (LOS) prediction in NICUs. In this study, the length of stay was categorized as a short stay (less than 3 days) or a long stay (more than 3 days) [7].

A recent review conducted in 2024 [6], compared various articles presenting machine learning methods for predicting hospital length of stay (LOS) using both classification and regression approaches. The review found that neural networks achieved superior performance in LOS prediction, while traditional machine learning methods also demonstrated good results, though their effectiveness was highly dependent on dataset characteristics. Classification was the most commonly used approach, with LOS often categorized into intervals.

Some studies have also focused on predicting LOS at admission, leveraging historical data that includes patient characteristics and treatment details up to discharge. However, relatively few studies have explored LOS prediction using only admission data [8]. One of the articles that showcased this is [9], where the authors predicted the LOS of readmitted patients at admission using demographic information and data of the first admission in the classification approach.

In another study [10], the authors applied machine learning models for predicting continuous LOS, using several features at admission for programmed patients admission such as age, primary and secondary diagnosis, comorbidity, and laboratory data. Table 1, summarizes previous works, highlighting the hospital's ward/field, algorithms, used approaches, the number of features considered (N.F), and the specific scenarios for prediction.

Table 1. Overview of articles in predicting Length of stay in hospitals.

Abbreviations: *Ar* – Article, *N.F* – Number of Features, *Perf* - Performance, *ICU* – Intensive Care Unit, *LGB* – Light Gradient Boosting, *UMLS* – Unified Medical Language System, *Cls.* - Classification, *Reg.* - regression. **Performance metrics:** *Acc* - Accuracy, R^2 and *MAE* - Mean Absolute Error

Ar	Ward/ Field	Scenario	Algorithm	Approach	N.F	Perf
[7]	Neonatal	Discharge	RF	Cls. (Acc)	11	97%
[9]	ICU	Admission	RF	Cls. (Acc)	11	71%
[10]	ICU	Admission	LR and RF	Reg. (R^2)	13	36%
[11]	ICU	Discharge	RF	Cls. (Acc)	70	96%
[12]	Clinical ward	Discharge	RF and UMLS	Cls. (Acc)	10	75%
[13]	hospital	Admission	LGB and RF	Cls. (Acc)	9	81%
[14]	ICU	Admission	Ridge and Lasso	Reg. (MAE)	12	0.96
[15]	cardio-thoracic	Preoperative	Bayesian	Reg. (MAE)	12	1.16
[16]	ICU	Admission	Deep attention	Cls. (Acc)	8	81%

Based on the analysis in Table 1, most studies employed machine learning algorithms for classification purposes. Specifically, studies such as [7,9,11] focused on binary classification for predicting the length of stay (LOS) at admission and discharge. In contrast, other works, including [12,13,16], addressed LOS prediction as a multi-classification problem by categorizing LOS into three distinct classes.

A smaller subset of studies [10,14,15], explored LOS as a regression problem. These studies employed a variety of approaches, from traditional machine learning to advanced deep learning techniques, with most focusing on data available up to the point of patient discharge. An exception was observed in [15], where the dataset included patient characteristics, the first laboratory results, the number of diagnoses, and comprehensive demographic details.

In conclusion, most studies on Length of Stay (LOS) prediction in hospitals have relied on historical data that captures patients’ states, characteristics, and treatments, aiming to predict LOS at both admission and discharge. Exceptions include studies such as [13,15], where models incorporated admission data alongside prescriptions specific to the patient’s diagnosis. These studies highlight the potential of leveraging targeted data for more precise LOS predictions. A wide range of data sources were used in these studies, with Electronic Medical Records (EMRs) serving as the primary source. Open-source datasets, like MIMIC [17], were extensively utilized in studies such as [9,15], underscoring their significance in LOS prediction research.

While most existing research has focused on the entire patient’s hospitalization process, from admission to discharge, some studies, such as [15], primarily rely on admission data, supplemented by the patient’s associated diagnosis and initial laboratory results. However, many of these approaches incorporate features that are available only after admission, such as follow-up laboratory tests

or a comprehensive medical history, limiting their applicability in early decision-making.

In contrast, our study aims to predict the length of stay (LOS) using only the data available at the time of patient admission, including the date of birth, gender, principal diagnosis, and the motivation for admission, particularly for planned admissions. By relying solely on admission-specific data, our model offers practical insights that enable hospitals to optimize resource management, including forecasting nurse workload, before additional clinical information becomes available.

Our work fills a critical gap in current research by demonstrating that LOS can be accurately predicted without relying on post-admission data, which primarily provides retrospective insights. Instead, our approach focuses on prospective prediction, enabling real-time decision-making. This is particularly valuable for short-stay patients, where early planning is essential for efficient hospital operations.

3 Case Study

Electronic Medical Records (EMR) in hospitals contain a vast amount of valuable information that can be leveraged for hospital management and workload prediction. While there are already numerous research efforts in the field of AI in healthcare utilizing existing data, EMRs remain a rich source of information to address the needs of healthcare providers further and improve analysis.

For this study, we utilized structured data from the EMRs maintained by Futura Smart Design [18], a system developed by Calystene. The dataset encompasses patient demographics, clinical diagnosis, medical procedures, and recorded treatments throughout hospitalization.

Upon analyzing the EMR, we identified a column labeled 'provisional date of discharge,' which represents an estimated date provided by staff to approximate when a patient might be discharged. However, this column is frequently incomplete or missing. To address this gap, we aim to predict the length of stay (LOS) in hospitals at the time of patient admission using the data available at that point. Accurately predicting LOS is not only crucial for optimizing hospital occupancy management but also serves as a key component for estimating nursing workload. Since nursing workload is influenced by both the LOS and the intensity of treatments administered, LOS prediction provides the first step for future workload forecasting models.

Our research focuses specifically on predicting LOS per ward. For this study, we targeted the Surgical department by focusing on one family of diseases: Diseases of the musculoskeletal system and connective tissue as a case study, where patients stay and treatment plans often vary significantly. By understanding the LOS trends, we can begin to establish a framework for connecting LOS with nursing workload, an aspect we plan to explore in future work.

The following section outlines our proposed approach for predicting Length of Stay.

4 Proposed approach

Prior to this work, we proposed a structured approach that generalized the process of nursing workload prediction [19], based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology [20]. This methodology consists of six steps in a data mining project: business understanding, data understanding, data processing, modeling, evaluation, and deployment.

In this study, we focus on two key components: data processing and modeling for predicting the length of stay at the admission stage, which forms part of the primary goal of forecasting nursing workload in short-stay hospitalizations.

4.1 Data processing

This part of the process is composed of several steps: data collection, feature selection, and data cleaning.

Data collection As previously mentioned, we extracted data from the EMR system of the company, which includes five years of short-stay hospitalization records from 2008 to 2013. The dataset comprises a total of 8152 rows. These records contain information spanning from patient admission to discharge. However, since our objective focuses on using data available at the time of admission, we retained only features such as date of birth (DOB), gender, primary diagnosis (PD), associated diagnoses (AD), and their corresponding labels (PDL and ADL), as well as weight, reason for admission (RA), provisional date of entry (PDE), real date of entry (RDE), provisional date of discharge (PDD), and actual date of discharge (ADD).

In hospitals, diagnoses are coded using the ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th Revision) system, which classifies diseases and health conditions into 22 main categories and their subgroups [21].

As discussed in Section 3, the provisional date of discharge is frequently unavailable. To support our objective of predicting the length of stay (LOS), we included the real date of discharge in the dataset for training and evaluation purposes.

Dataset features selection At the patient admission stage, the available data is limited, and its availability varies depending on the type of hospitalization (short, medium, and long) and the ward. In our case, we used data from the surgical ward in short-term hospitalization. After careful discussion with the company’s experts, we made an initial selection of features, as described in section 4.1.

Subsequently, we made a second selection based on the data available in our dataset. We found that the variables weight, reason for admission, and secondary diagnosis were not reported for more than 85% of the cases. Additionally, the

primary diagnosis labeling showed redundancy in the data. As a result, we retained only the variables age, gender, primary diagnosis codes, and length of stay.

Table 2, provides insight into the percentage of the missing values (MV) in each feature and its description.

Table 2. Features Description and Percentage of Missing Values (MV)

Feature	Description	MV (%)
RA	The primary reason for the patient’s hospitalization	100
PED	The planned admission date for the patient, relevant for scheduled admissions	99.98
AD	Additional medical conditions diagnosed alongside the primary diagnosis	98.64
ADL	Label for the associated diagnoses	98.64
Weight	The weight of the patient (in kilograms)	86.3
DOB	The patient’s date of birth	0
RDE	The actual date of the patient’s hospital admission	0
Age	The patient’s age (calculated from the DOB)	0
Gender ¹	The biological sex of the patient	0
PD	The primary diagnosis assigned to the patient upon admission	0
PDL	Label of the primary diagnosis	0
RDD	The actual date of the patient’s hospital discharge	0
LOS	The total length of the patient’s stay in the hospital (calculated from admission to discharge)	0

¹ In our dataset, gender refers to the biological sex of the patient, which was either male or female.

Data cleaning Working with raw data can be challenging and may impact model performance. To address this, we applied several preprocessing rules specific to the surgery ward. These included retaining only records with a length of hospitalization greater than 1 day and less than 15 days, as the mean length of stay (LOS) in short-stay hospitalizations is approximately six days [22].

Errors detection Several quality control measures were implemented to ensure the integrity of the data: the validation of admission dates; any admission date that was before the date of birth or after the date of discharge was removed, as such records were deemed invalid; removal of invalid diagnoses. Finally, incomplete patient records were excluded, specifically those missing essential information such as date of birth, date of admission, date of discharge, or gender. These fields were considered mandatory, and their absence was attributed to issues in the Electronic Medical Records (EMR) system or incompleteness.

Diagnosis feature processing To analyze the distribution, total number of diagnoses, and hierarchical levels used in the diagnosis, we processed the ICD-10 codes as follows:

- Aggregation of Low-Frequency Categories: Diagnoses with a frequency below 0.3% and above 0.1% were aggregated into their next hierarchical level to ensure adequate representation in the dataset. For instance, if "A047" had a frequency below 0.2%, it was assigned to the broader category "A04."
- Exclusion of Rare Diagnoses: Diagnoses that appeared less than 0.1 in the dataset were excluded from the analysis to focus on more prevalent cases. In addition, diagnoses that didn't follow the ICD-10 codification were deleted.
- Standardization of Code Length: Diagnoses were standardized to include up to four characters (e.g., "A047" instead of "A0471"), maintaining a meaningful detail of the diagnosis.

The results of our data preprocessing revealed that we have a dataset comprising 7880 patients with verified and reliable information, focusing on 42 unique diagnoses. The age of patients ranges from 2 to 96 years. Figure 1 illustrates the interaction of three key features—diagnosis, age, and gender—with the length of stay.

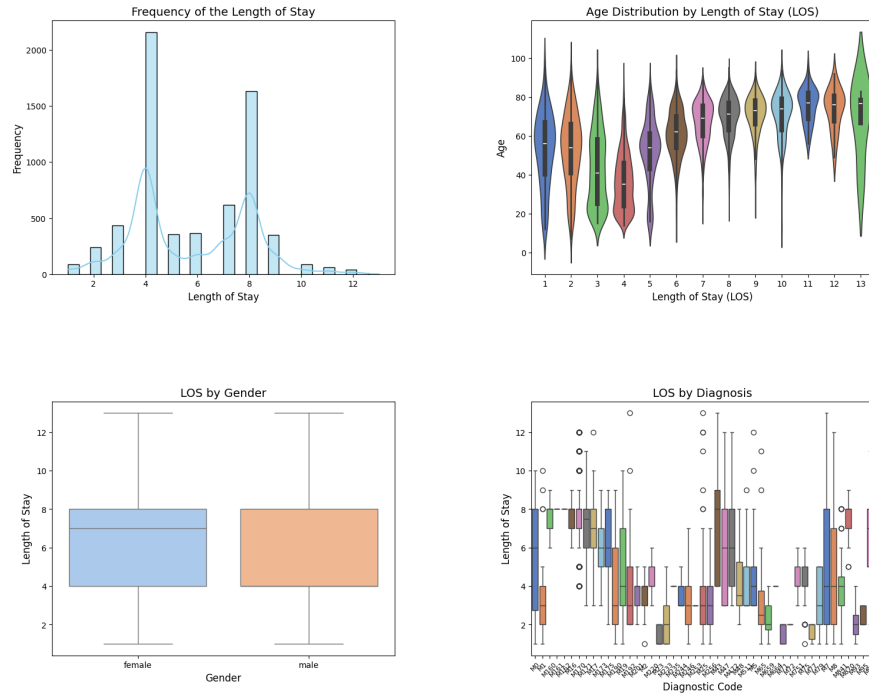


Fig. 1. Interaction of features with length of stay.

Our analysis indicates that gender does not significantly influence the length of stay, as no observable variability was noted based on this feature. In contrast, both diagnosis and age demonstrate a notable influence on the length of stay. For instance, the variability in length of stay across different diagnoses and age highlights the complexity and severity of the conditions being treated.

Outliers detection and processing Outliers were identified in both the diagnosis and age features. These outliers represent valid but extreme cases. They can often be identified by other features such as the associated diagnosis and dependency, which were not documented in the EMR for short-stay hospitalizations. Additionally, these cases are influenced by other factors such as specific treatments and medical practices, which are beyond the scope of the current analysis. To ensure robustness and improve the reliability of our findings, we proceeded to remove these outliers. This was achieved by calculating the first (Q1) and third (Q3) quartile for the relevant features and excluding values outside the bounds defined as (Lower Bound, Upper Bound)

This preprocessing ensured that our analysis focused on the most reliable and representative data, minimizing the potential for skewness while maintaining the integrity of the dataset.

4.2 Modeling

In Section 2, we reviewed various machine learning algorithms commonly employed in similar analyses, ranging from classification and regression techniques to neural networks. Since our task involves predicting a continuous variable, we selected four regression models: three gradient boosting models—XGBoost, CatBoost (which efficiently handles categorical features by reducing preprocessing steps [23]), and LightGBM (a gradient boosting model optimized for speed and efficiency)—as well as an ensemble learning model, Random Forest. We also applied a Feedforward Neural Network model (FNN), a widely used model where the information follows a unique direction.

These models were chosen for their established effectiveness in regression tasks and their capability to capture complex relationships in the data. Given that our dataset includes both categorical and continuous features, we applied label encoding to the categorical features, such as ICD-10 diagnosis codes and gender, for all models except CatBoost.

Model Hyperparameter tuning Hyperparameters are parameters that must be initialized before training and play a crucial role in optimizing model performance, which ultimately leads to better results [24,25]. In our study, we adopted the grid search approach [24]. This method evaluates all possible combinations of hyperparameters to identify the optimal configuration for achieving the best performance. Table 3 summarizes the hyperparameters used for each model. Additionally, in the case of the FeedForward Neural Network, we used a four-layer architecture consisting of an input layer, two hidden layers, and an output

layer. The input and hidden layers utilized the Rectified Linear Unit (ReLU) activation function to capture non-linear patterns, while the output layer adopted a linear activation function appropriate for predicting a continuous target variable.

Table 3. Optimal Hyperparameters of our models.

Model	Hyperparameters	Best Values
XGBoost	learning_rate, max_depth, n_estimators, subsample	0.01, 5, 500, 0.8
RF	min_sample_leaf, max_depth, n_estimators, min_samples_split	4, 10, 500, 10
CatBoost	depth, iterations, l2_leaf_reg, learning_rate	4, 500, 5, 0.1
LightGBM	learning_rate, n_estimators, num_leaves, subsample	0.01, 500, 31, 0.7
FNN	number of layers, activation functions, number of units	4, {ReLU, linear}, {64, 32, 1}

As shown in Table 3, four key hyperparameters were selected for each model and tested to identify the optimal configuration. These parameters were utilized during the training and validation phases, which will be further discussed in section 4.3

4.3 Evaluation and results

The models we employed followed a conventional process, with the dataset split into training and test subsets, accounting for 80% and 20% of the data, respectively.

To evaluate the performance of our models, we relied on a set of well-established metrics commonly used in regression tasks. These metrics—Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score—allowed us to comprehensively assess the models’ performance. Specifically, MAE measured the average magnitude of errors regardless of direction (positive or negative), RMSE quantified the margin between predictions and actual outcomes, and R^2 Score indicated the proportion of variance explained by the model, providing a measure of prediction precision [6]. Table 4, represents all the results of our models.

Based on the results presented in Table 4, the precision of the models is consistent, with minimal differences between the R^2 scores for the training and test datasets. This indicates no significant overfitting, as the models maintain similar performance levels on unseen data.

The LightGBM model yielded the best performance from our experiments, closely followed by XGBoost. Despite being specifically designed to handle categorical data effectively, CatBoost was outperformed by LightGBM. This can

Table 4. Performance metrics for predicting Length of Stay.

Model	Train			Test		
	MAE	RMSE	R ²	MAE	RMSE	R ²
XGBoost	0.609	1.008	0.79	0.627	1.080	0.76
Random Forest	0.585	0.974	0.81	0.623	1.090	0.76
CatBoost	0.614	1.006	0.80	0.633	1.083	0.76
LightGBM	0.60	1.07	0.80	0.60	1.07	0.77
FeedForward NN	0.6	1.00	0.8	0.7	1.4	0.73

be attributed to LightGBM’s robust hyperparameter tuning, which effectively mitigates overfitting, its efficient processing of categorical features, and its architecture, which allows it to explore information deeply and comprehensively. However, the FeedForward Neural Network (FNN) was the least effective model, exhibiting the highest test RMSE among all models. However, the small difference between the train and test scores suggests that the model was not overfitting. This can be credited to early stopping during training, which prevented the model from excessively optimizing the training data.

Going further we explored the best model results to explain the importance and the influence of the variables using Shapley Additive exPlanations (SHAP) [26].

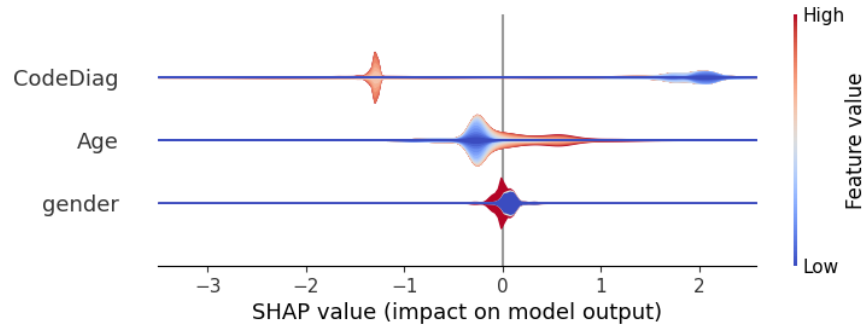
**Fig. 2.** Distribution of SHAP values across the features.

Figure 2, illustrates the importance and impact of each feature on the model’s results. It shows that the diagnoses are the most influential factor, with two distinct groups: one group influences the SHAP positively, explained by a longer length of stay, and the other group influences the SHAP negatively, associated with a shorter length of stay. Next in importance is age, where older patients tend to have higher SHAP values. Finally, gender does not appear to significantly influence the model’s predictions.

Additionally, as shown in figure 3, there is a non-linear relationship between age and LOS, which the model has effectively captured. We can also confirm the

impact of these features and observe how age influences the diagnoses. Specifically, for patients aged 15 to 40, the SHAP values remain stable across a mix of diagnoses. However, for patients over 40, the SHAP values increase with age, following a positive trend. Finally, the interaction with diagnoses further supports this finding.

In conclusion, the interaction between age and diagnoses influences the SHAP values, with changes in age affecting the length of stay for different diagnoses. As age increases, it impacts the length of stay.

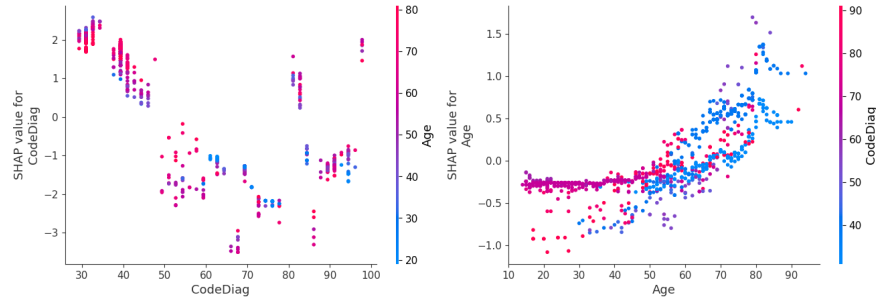


Fig. 3. SHAP Value Analysis for Diagnosis and Age.

5 Conclusion

In this study, we utilized machine learning algorithms to predict patients’ surgical ward Length of Stay (LOS) at the time of admission, providing valuable insights for short-stay hospitalization. We analyzed electronic Medical records (EMR) comprising data from 7880 surgical ward patients after preprocessing. Our models included XGBoost, CatBoost, LightGBM, Random Forest, and a Feedforward Neural Network, achieving comparable performance with R^2 scores ranging from 73% to 77%, with LightGBM delivering the best performance of 77% R^2 and 0.6 MAE.

These results demonstrate that our approach outperforms related studies in LOS prediction at admission despite relying on a more limited feature set. Unlike other studies that incorporate both demographic and comprehensive clinical data available up to discharge, our analysis leveraged only demographic attributes (age and gender), as well as the principal diagnosis at admission.

Despite these promising results, certain limitations must be acknowledged. First, the model’s reliance on limited data at admission may exclude valuable temporal patterns and longitudinal information in post-admission data. Additionally, the absence of key admission-time details such as secondary diagnoses, comorbidities, and the reason for admission, which are typically recorded at admission but unavailable in our dataset may have restricted the model’s ability

to capture important predictive factors. Furthermore, the generalizability of the findings is constrained by the dataset’s single-center origin, which might limit applicability in diverse clinical settings. Expanding the dataset to include multiple hospital wards and a broader range of medical conditions would improve its robustness and applicability. Moreover, leveraging currently underutilized admission-time features —previously unusable due to missing values— could enhance predictive accuracy.

Looking ahead, several avenues for future research could strengthen and extend our work. In addition to dataset expansion, integrating electronic health records (EHRs) and real-time patient monitoring data could further enrich the predictive model by capturing evolving patient conditions.

We also aim to develop near real-time LOS and treatment predictions, potentially utilizing machine learning models such as recurrent neural networks (RNNs) or transformer-based architectures to enable dynamic adjustments in nurses’ workload forecasts. By incorporating a more comprehensive set of features and leveraging temporal data more effectively, future models could better capture variations throughout a patient’s hospital stay, ultimately improving resource allocation and operational efficiency in healthcare settings.

References

1. World Health Organization: (2003). How can hospital performance be measured and monitored?. In *How can hospital performance be measured and monitored?* (pp. 17-17).
2. Tipton, K., Leas, B.F., Mull, N.K., et al.: *Interventions To Decrease Hospital Length of Stay* [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2021 Sep. (Technical Brief, No. 40.) Introduction. Available from: <https://ncbi.nlm.nih.gov/books/NBK574438/>
3. Hunter, A., Johnson, L., Coustasse, A.: Reduction of intensive care unit length of stay: the case of early mobilization. *The Health Care Manager*, 33(2), pp. 128–135 (2014).
4. Wang, Z., Chen, X., Wu, Y. et al.: A robust and interpretable ensemble machine learning model for predicting healthcare insurance fraud. *Sci Rep* 15, 218 (2025). <https://doi.org/10.1038/s41598-024-82062-x>
5. Nadella, G., Satish, S., Meduri, K., Meduri, S.: A Systematic Literature Review of Advancements, Challenges and Future Directions of AI and ML in Healthcare. *Int. J. Mach. Learn. Sustain. Dev.* 5, pp. 115–130 (2023).
6. Almeida, G., Brito Correia, F., Borges, A.R., Bernardino, J.: Hospital Length-of-Stay Prediction Using Machine Learning Algorithms—A Literature Review. *Appl. Sci.* 14, 10523 (2024). <https://doi.org/10.3390/app142210523>
7. Erdogan Yildirim, A., Canayaz, M.: Machine learning-based prediction of length of stay (LoS) in the neonatal intensive care unit using ensemble methods. *Neural Comput. & Applic.* 36, pp. 14433–14448 (2024). <https://doi.org/10.1007/s00521-024-09831-7>
8. Deschepper, M., De Smedt, C., Colpaert, K.: A literature-based approach to predict continuous hospital length of stay in adult acute care patients using admission

- variables: A single university center experience. *Int. J. Med. Inform.* 193, (2025). <https://doi.org/10.1016/j.ijmedinf.2024.105678>
9. Zhang, M., Kuo, T.-T.: Early prediction of long hospital stay for intensive care units readmission patients using medication information. *Comput. Biol. Med.* 174, (2024). <https://doi.org/10.1016/j.compbiomed.2024.108451>
 10. Peres, I.T., Hamacher, S., Oliveira, F.L.C., Bozza, F.A., Salluh, J.I.F.: Data-driven methodology to predict the ICU length of stay: A multicentre study of 99,492 admissions in 109 Brazilian units. *Anaesth. Crit. Care & Pain Med.* 41, (2022). <https://doi.org/10.1016/j.accpm.2022.101142>
 11. Alsinglawi, B., Alshari, O., Alorjani, M., et al.: An explainable machine learning framework for lung cancer hospital length of stay prediction. *Sci. Rep.* 12, 607 (2022). <https://doi.org/10.1038/s41598-021-04608-7>
 12. Chrusciel, J., Girardon, F., Roquette, L., et al.: The prediction of hospital length of stay using unstructured data. *BMC Med. Inform. Decis. Mak.* 21, 351 (2021). <https://doi.org/10.1186/s12911-021-01722-4>
 13. Lee, H., Kim, S., Moon, H., et al.: Hospital Length of Stay Prediction for Planned Admissions Using Observational Medical Outcomes Partnership Common Data Model: Retrospective Study. *J. Med. Internet Res.* 26, e59260 (2024). <https://www.jmir.org/2024/1/e59260>, <https://doi.org/10.2196/59260>
 14. Grampurohit, S., Sunkad, S.: Hospital Length of Stay Prediction Using Regression Models. In: 2020 IEEE International Conference for Innovation in Technology (INOCON), pp. 1–5. IEEE Press, Bengaluru (2020). <https://doi.org/10.1109/INOCON50539.2020.9298294>.
 15. Abdurrab, I., Mahmood, T., Sheikh, S., et al.: Predicting the Length of Stay of Cardiac Patients Based on Pre-Operative Variables—Bayesian Models vs. Machine Learning Models. *Healthcare (Basel)*. 12(2), 249 (2024). <https://doi.org/10.3390/healthcare12020249>
 16. Harerimana, G., Kim, J.W., Jang, B.: A deep attention model to forecast the Length of Stay and the in-hospital mortality right on admission from ICD codes and demographic data. *J. Biomed. Inform.* (2021). <https://doi.org/10.1016/j.jbi.2021.103778>.
 17. Johnson, A., Pollard, T., Mark, R.: MIMIC-III Clinical Database Demo (version 1.4). *PhysioNet*. (2019). <https://doi.org/10.13026/C2HM2Q>.
 18. Futura Smart Design, <https://www.calystene.com/futura-smart-design-solutions-informatiques-pour-hopitaux-et-cliniques/>, last accessed 2024/12/19
 19. Gharbi, M., Verdier, C., Di Mascolo, M.: Nurses' workload prediction in hospitals—a Machine Learning-based approach. In: *INFORSID*, (2024).
 20. Mirza, B., Li, X., Lauwers, K., et al.: A clinical site workload prediction model with machine learning lifecycle. *Healthc. Anal.* 3, 100159 (2023).
 21. ICD-10, <https://icd.who.int/browse10/2019/en>, last accessed 2025/01/02
 22. Mean LOS, https://www.atih.sante.fr/sites/default/files/public/content/4729/synthese_aah_2022_mco.pdf, Last accessed 2025/01/10
 23. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A.: CatBoost: unbiased boosting with categorical features. In: *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 6639–6649. Montreal (2018).
 24. Ogunsanya, M., Isichei, J., Desai, S.: Grid search hyperparameter tuning in additive manufacturing processes. *Manuf. Lett.* 35, Suppl., pp. 1031–1042 (2023). <https://doi.org/10.1016/j.mfglet.2023.08.056>.

25. Belete, D., Manjaiah, D.H.: Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *Int. J. Comput. Appl.* 44, pp. 1–12 (2021). <https://doi.org/10.1080/1206212X.2021.1974663>.
26. Lundberg, S.: A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874 (2017).