ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Pediatric patient hospital length of stay prediction: A comparative analysis of Bayesian inference and machine learning approaches

Sarmad Zafar[1]*, Tariq Mahmood[1], Zahra Hoodbhoy[2], and Babar Hasan[3]

[1]Big Data Analytics Laboratory, Department of Computer Science, School of Mathematics and Computer Science, Institute of Business Administration Karachi, Karachi, Sindh, Pakistan

[2]Department of Paediatrics and Child Health, Medical College, Aga Khan University, Karachi, Sindh, Pakistan

[3]Division of Cardiothoracic Sciences, Division of Cardio-thoracic Sciences, Sindh Institute of Urology and Transplantation, Karachi, Sindh, Pakistan

Abstract

Predicting patient length of stay (LoS) is crucial for optimizing resource allocation and enhancing healthcare efficiency. However, achieving accurate LoS predictions remains a challenging and complex task. This study presents a non-disease-specific predictive model that integrates machine learning (ML) methods and Bayesian inference techniques to accurately predict hospital LoS using static patient admission data. While traditional statistical regression techniques have been widely used for LoS prediction within hospital settings, this research investigates the capabilities of ML and Bayesian inference algorithms in this context. By leveraging Bayesian inference techniques, our model captures complex relationships within the data and quantifies uncertainty, offering a more nuanced understanding of the outcomes. This methodological approach offers a more comprehensive and probabilistically grounded framework for LoS prediction, allowing more informed decision-making in resource allocation and patient management. Among the evaluated models, extreme boosting and support vector machine regressor models demonstrated the highest efficiency, achieving mean squared logarithmic error (MSLE) values of 0.23 and 0.24, respectively. The Bayesian model also showed competitive performance with an MSLE of 0.25. While it did not outperform other models in terms of error metrics, the Bayesian model's ability to provide additional uncertainty output enhances its utility, offering valuable supplementary information for informed decision-making. This research highlights the potential of ML and Bayesian inference in predicting patient LoS, emphasizing their significance in effective resource allocation and patient care management within the healthcare sector.

Keywords: Length of stay; Machine learning; Predictive model; Bayesian inference; Natural language processing

***Corresponding author:**
Sarmad Zafar
(s.zafar@khi.iba.edu.pk)

Citation: Zafar S, Mahmood T, Hoodbhoy Z, Hasan B. Pediatric patient hospital length of stay prediction: A comparative analysis of Bayesian inference and machine learning approaches. *Artif Intell Health*. 2026;3(1):77-87. doi: 10.36922/AIH025160030

Received: April 14, 2025

Revised: June 26, 2025

Accepted: July 7, 2025

Published online: July 22, 2025

1. Introduction

Predictive modeling has become increasingly prevalent in various domains for forecasting future outcomes and facilitating informed decision-making. In healthcare, predictive modeling, especially modern machine learning (ML) methods, is widely used to forecast individual patient outcomes using large patient datasets.[1] These techniques have been used to predict numerous critical features such as mortality, readmission probabilities, recommended treatments, and length of stay (LoS).[2-4] Patient LoS is an important metric in healthcare operations, contributing significantly to the efficient flow of patients through hospital systems. Its strategic importance lies in its role in optimizing resource allocation, bed management, specialist scheduling, billing estimation, discharge planning, and overall enhancement of patient satisfaction and operational efficacy.[5,6]

Accurate LoS prediction is an effective tool for addressing challenges in resource planning, capacity management, and staffing. Reliable forecasts of patient discharge dates enable better scheduling of elective admissions and improve hospital bed occupancy management. In addition, accurate LoS predictions can enhance hospital workflow, improve patients' safety, reduce healthcare costs, and optimize resource utilization.[7] LoS is often used as a proxy indicator when direct measurement of certain outcomes is not feasible; for example, it can be used as a proxy for hospital mortality.[8] Moreover, LoS is used to assess illness severity and estimate healthcare resource utilization.[9]

LoS is a complex metric influenced by multiple factors, such as the type of illness, geographic location and season, and individual characteristics including demographics, age, medical complications, and treatment complexity.[10] Conventionally, statistical techniques have dominated LoS predictive modeling.[11,12] These algorithms assume specific relationships between variables and outcomes and often treat variables as independent. Given the complexity of LoS determinants, these assumptions may not hold, leading to limited predictive performance.[13] Hence, there is a growing emphasis on leveraging ML algorithms to improve prediction accuracy. Figure 1 highlights the methods used in the literature for calculating and predicting LoS.[14]

ML algorithms can process and integrate numerous features, capturing complex and non-linear relationships between them to make accurate predictions. Various ML techniques have been used for LoS prediction, including linear regression, multilayer perceptron, random forest, bagging, boosting, and support vector machine.[15-18] While these methods vary in performance, depending on the specific context and data, they offer a flexible framework that often outperforms traditional approaches in capturing the underlying patterns associated with hospital LoS.

Despite their successes, classical ML approaches sometimes fail to deliver satisfactory results or may be unsuitable due to specific circumstances, available data, desired outcomes, and personal preferences. Consequently, Bayesian inference is used as an alternative to frequentist approaches because it allows the incorporation of prior knowledge during model training, helping to mitigate limitations posed by small or imperfect datasets.[19] In addition, Bayesian methods provide insights into the model's confidence in each prediction by quantifying uncertainty – distinct from the confidence intervals in frequentist methods – and by detecting areas with insufficient training data.[20]

Bayesian methods, unlike classical ML methods, leverage probability distributions for model parameters, enabling uncertainty estimation. This capability allows Bayesian models to quantify uncertainties associated with predicted outputs, offering a more nuanced understanding of the reliability of predictions.[21] In healthcare, if Bayesian predictive models demonstrate comparable performance to frequentist models, they are often preferred due to their ability to provide uncertainty estimates.[22]

The existing body of literature predominantly relies on conventional regression methods or frequently incorporates ML approaches such as random forest and support vector machines for LoS prediction. In this research, we aim to bridge this gap by introducing Bayesian inference-based techniques. The primary objective is to enhance predictive accuracy through improved point estimates while simultaneously strengthening the reliability of uncertainty quantification. In addition, this study compares the performance of Bayesian and ML approaches to evaluate their relative effectiveness in LoS prediction.

2. Data and methods

2.1. Study setting and data variables

This study utilized data sourced from Aga Khan University Hospital, Karachi, a distinguished not-for-profit tertiary healthcare facility. Aga Khan University Hospital serves as a leading hospital in Pakistan, catering to a highly diverse patient population presenting with a wide range of health conditions from across the country and abroad. The study focused on pediatric patients aged less than 18 years admitted as inpatients between January 01, 2015, and November 30, 2019. Records pertaining to inpatients admitted for elective procedures, surgeries, or diagnostic examinations were omitted from the analysis. The dataset consisted of records from approximately 22,106 pediatric patients, encompassing a range of attributes including demographic and health-related parameters, clinical admission information, such as the initial level of care classification (general, intensive, special, or isolation care), and other relevant administrative data. Socio-demographic attributes include age, gender, address, and financial class.

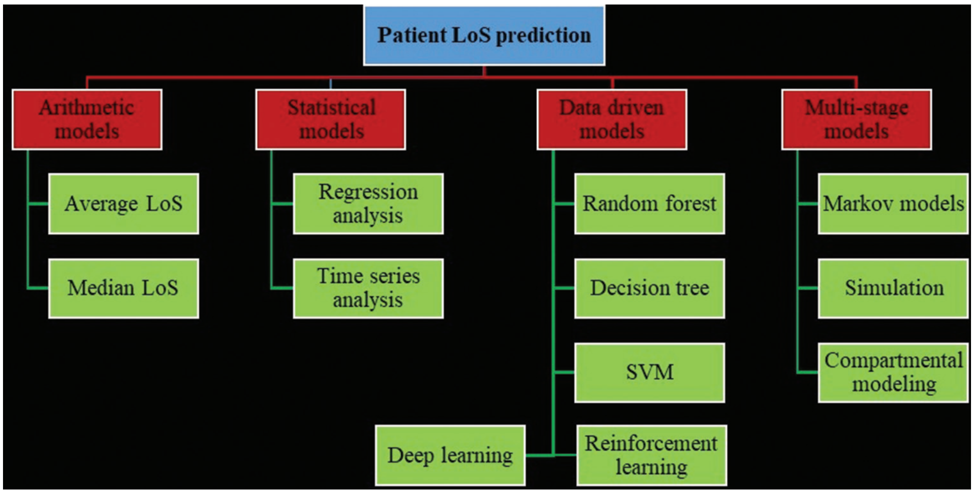

Figure 1. Length of stay prediction methods
Abbreviations: LoS: Length of stay; SVM: Support vector machine.

LoS in hospitals tends to increase when patients have comorbidities, which are additional medical conditions existing alongside the primary ailment.[23,24] In our study, we used the visit reason feature to create a comorbidity indicator. Each patient was assigned an integer value: "1" if they had no comorbidities or a single disease, and higher values (e.g., ≥2) if they had multiple comorbidities. Table 1 presents an overview of all variables utilized in this study, including their brief descriptions and data types. The main outcome variable was the LoS for admitted patients, measured as a continuous numeric variable with a wide distribution of durations. Notably, within the pediatric patient dataset examined, approximately 95% of cases exhibited LoS within 15 days. Instances of LoS surpassing 20 days were deemed outliers due to their infrequency and were consequently eliminated from the dataset.

Table 1. Feature description

Field name	Description	Data type
Patient ID	Unique numeric identifier assigned to each patient	Integer
Address	District-level location of patient residence	String (open-text entry)
Street	Detailed residential location, including street and neighborhood	
Financial class	Classification of financial coverage or funding source	Categorical (5 categories)
Admission date	Calendar date on which the patient was admitted	Date (DD/MM/YYYY)
Discharge date	Calendar date on which the patient was discharged	Date (DD/MM/YYYY)
Visit reason	The initial reason for seeking care, as assessed by the emergency room consultant	String (open-text entry)
Length of stay	Total duration of hospitalization in days	Integer
Admission source	Indicates if admission was emergency-based or scheduled	Categorical (2 categories)
Date of birth	Patient calendar date of birth, used to derive patient age	Date (DD/MM/YYYY)
Admission care type	Accommodation type assigned upon admission (e.g., ward and private room)	Categorical (3 levels)
Admission care level	Classification of care intensity (e.g., general and intensive)	Categorical (4 levels)
Gender	Biological sex of the patient	Categorical (Male/Female)

2.2. Evaluation measures

In this study, we evaluated the predictive model using internal validation, dividing the dataset into training and testing subsets. We randomly partitioned patient data, allocating three-fourths for training and one-fourth for testing. Model performance was primarily assessed using the mean squared log error (MSLE), defined in Equation I.

$$MSLE(y,\hat{y})=\frac{1}{n}\sum_{i=1}^{n}\left(log\left(1+y_i\right)-log\left(1+\hat{y}_i\right)\right)^2 \quad \text{(I)}$$

The MSLE metric is considered a primary evaluation metric due to its ability to penalize errors proportionally to the size of predictions, making it suitable for our applications. For example, if the actual LoS is two days but predicted as 3 days, and another case has an LoS of 15 days but is predicted as 16.5 days, MSLE accounts for the relative significance of errors. Thus, a 1-day error in the shorter stay is treated as more significant than the same absolute error in the longer stay. For a comprehensive evaluation, we also reported other metrics such as mean absolute error (MAE), mean square error, and root mean square error (RMSE). In this study, Python (version 3.12.4) served as the primary tool for data analysis, feature engineering, and model estimation, chosen for its support for ML, Bayesian inference, and natural language processing (NLP)-related tasks.

2.3. Pre-processing and feature engineering

In healthcare, clinical histories often consist of unstructured text, serving as a primary source of information regarding patients' diagnoses, medications, symptoms, and other clinical factors.[25] In our study, we utilized the named entity recognition (NER) technique to process two unstructured features: visit reason and address. The visit reason comprises textual statements of potential diagnoses recorded by attending physicians in the Emergency Room, while the address feature contains detailed textual descriptions of patients' locations.

To convert unstructured features into usable data for ML, we employed a systematic pre-processing approach followed by NER. Initially, lemmatization was applied to standardize word forms and enhance consistency across the dataset. Subsequently, part-of-speech tagging was applied to filter nouns, adjectives, and prepositions, which are indicative of named entities in clinical texts. Further refining steps included spelling error correction using specialized dictionaries and the removal of redundant phrases to streamline the dataset. In addition, medical abbreviations were expanded to ensure coherence. For the address feature, NER involved error correction in street, district, and area names, followed by the application of a location dictionary specific to Pakistan districts, particularly compiled to ensure accuracy. Overall, our NER methodology integrated linguistic analysis techniques with domain-specific knowledge to accurately extract named entities from clinical text, facilitating the creation of usable features for subsequent ML applications. In our study, we utilized the visit reason feature to derive a comorbidity indicator, assigning integer values to quantify the complexity of their medical conditions.

2.4. ML

This study evaluated and compared a range of ML models to identify the most suitable model for our problem, taking into account that variations in dataset characteristics and outcome variables can significantly influence model performance. We tested six standard regression models, encompassing seven different algorithms commonly used in the field[26-28] (Table 2). Patient data were randomly split into training (75%) and testing (25%) sets, and each algorithm was assessed using three metrics: MSLE, MAE, and RMSE.[26] Optimal hyperparameter configurations for each model were determined using an exhaustive grid search over a predefined list of hyperparameters (Table S1). For each hyperparameter combination, the testing error was recorded, and the lowest error was reported in the results.

Table 2. Machine learning models brief description

Classification model	Description
Linear regression	A foundational algorithm used for modeling the relationship between input features and a continuous target variable. Fitting a linear equation to the training data enables predictions of the target variable based on new input data, making it a fundamental technique for regression tasks in predictive modeling. Algorithm: multivariate linear regression
Decision tree	A decision tree structures decisions as branches and outcomes as leaves, where internal nodes correspond to feature-based queries. It recursively partitions the data space to predict output values based on input feature thresholds. Algorithm: decision tree regressor
Bagging	Bagging, or bootstrap aggregation, builds multiple independent models on varied subsets of data. The aggregation of these models, typically decision trees, enhances generalization and minimizes overfitting. Algorithm: random forest regressor
Boosting	Boosting constructs a strong predictor by sequentially training weak learners, each compensating for the errors of its predecessors. Greater weight is given to previously mispredicted instances to refine subsequent models. Algorithm: extreme gradient boosting, Adaptive boosting
Nearest neighbor	The Nearest neighbor approach operates by identifying a set number of closest data points to a given point and predicting its value based on the average or weighted average of these neighboring points. Algorithm: K-nearest neighbor regression
Support vectors	This method aims to find the optimal separating hyperplane that maximizes the margin between data points and the decision boundary. Kernel functions can be applied to capture non-linear relationships in higher-dimensional spaces. Algorithm: support vector regression

2.4. Bayesian modeling

This study also utilized a Bayesian framework to predict the LoS, integrating both prior knowledge and observed data. We started by specifying a probabilistic model for the LoS, assuming it follows a distribution conditioned on parameters. Let L represent the LoS, and θ denotes the parameters governing its distribution. The likelihood function captures the probability of observing a specific LoS given the model parameters. Incorporating prior knowledge, we assigned a prior distribution $P(\theta)$ over the parameters. Then, using Bayes' theorem, the posterior distribution of the parameters given the observed data D is obtained, as shown in Equation II:

$$p(\theta \mid D) = \frac{p(D \mid \theta) \times p(\theta)}{p(D)} \quad \text{(II)}$$

where P(D|θ) is the likelihood function, P(θ) is the prior distribution, and P(D) is the marginal likelihood. To obtain the posterior distribution of the LoS P(L|D), we integrated all possible parameter values, as shown in Equation III:

$$p(L \mid D) = \int \left[p(L \mid \theta, D) \times p(\theta \mid D) \right] d\theta \quad \text{(III)}$$

This integral represents the predictive distribution of the LoS given the observed data. However, obtaining the exact form of the posterior distribution can be computationally challenging, especially for complex models. To address this, we employed the Markov Chain Monte Carlo method, a powerful class of algorithms, for efficient sampling from the posterior distribution.

We assumed non-informative densities as prior distributions for the coefficients, and tested multiple weakly informative priors for the coefficients, including chi-squared, Gamma, bounded normal, and Poisson distributions, with different parameter values. Using a wide prior expanded the search space, increasing the possibility of finding an optimal solution. Given that LoS follows a skewed distribution, the Gamma prior combined with a Poisson likelihood yielded the best results. Other skewed distributions, like Gamma and chi-squared, were also tested.[29] Bayesian multiple linear and non-linear models were trained using different distributions, parameters, and numbers of features to evaluate their predictive performance.

2.4.1. Linear model (single feature)

First, we tested the Bayesian linear regression model with different non-informative prior distributions with one regressor. The first step of the Bayesian prediction is to choose/compute the prior probability distribution and likelihood distribution. Our first model is simple, with a conjugate normal prior with a mean of five and a standard deviation of 10 (Figure S1). The model is specified as below in Equations IV-VII.

$$\alpha, \beta = N(5,10) \quad \text{(IV)}$$

$$\sigma = HC(4) \quad \text{(V)}$$

$$\mu_i = \alpha + \beta x_i \quad \text{(VI)}$$

$$L|\mu_i, \sigma \sim N(\mu_i, \sigma) \quad \text{(VII)}$$

α and β are the prior distributions for the intercept and variable coefficient, respectively. The standard deviation of the likelihood function was modeled using the Half-Cauchy distribution throughout this study.

2.4.2. Linear model (multiple features)

Similarly, the regression scenario was generalized in several ways using multiple variable regression settings, where the mean of a continuous response was written as a linear function of several predictor variables. Similar to a simple linear regression model, a multiple linear regression model assumes an observation-specific mean μ_i for the i^{th} response variable Yi (Equation VIII).

$$L|\mu_i, \sigma \sim N(\mu_i, \sigma), i = 1 \ldots .n \quad \text{(VIII)}$$

In addition, it assumes that the mean of Yi is μ_i, a linear function of all predictors. With four predictors, it is written as in Equations IX-XI.

$$\mu_i = \alpha + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} \quad \text{(IX)}$$

$$\alpha, \beta_r = N(5,10) \quad \text{(X)}$$

$$\sigma = HC(4) \quad \text{(XI)}$$

where $x_{i,1}$ is a predictor, like age and location, for observation i, and β = (β1, β2,⋯, βr) is a vector of unknown regression parameters (coefficients), shared among all observations. Equations X and XI are the prior distribution and standard deviation distribution, respectively.

2.4.3. Non-linear model

While linear models assume a straight line relationship between the input and output data, a non-linear approach is more suitable when variables exhibit a curved relationship. There are many ways to model a curved relationship between two variables, including using higher-order polynomial terms, such as x squared, x cubed, or basic geometry or trigonometry functions, such as exponential or cosine functions. Similar to the linear model, in a non-linear model, an additional term with squares and cubes of the predictor variable was added. This only altered Equation XII while keeping the others unchanged.

$$\mu_i = \alpha + \beta_1 x_{i,1} + \beta_{11} x_{i,1}{}^2 + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} \quad \text{(XII)}$$

3. Results

Table 3 summarizes the general findings from the patient data. The LoS exhibits a skewed distribution, with a mean of 3.82 days and a median of 2 days. Approximately 95% of patients have a LoS of <15 days, and over 82% of patients have a LoS of <5 days. The average age of patients admitted to the hospital is 4.2 years, which also displays a skewed distribution, with 75% of patients being under seven years old. Seasonal patterns are evident in the data, as observed

Table 3. Statistical findings

Variable	Values (n=22,106)
Length of stay (LoS), days (median [IQR])	2 (4)
LoS <5 days	82% (18,114)
LoS <10 days	92% (20,406)
LoS <15 days	95% (21,056)
Sex	
Male	60% (13,283)
Female	40% (8823)
Clinician prediction	
True prediction	64.7% (14,304)
LoS (median [IQR])	1.6 [1.0, 2.4]
False prediction	35.3% (78,02)
LoS (median [IQR])	3.9 [2.6, 7.0]
Age, years (mean, standard deviation)	4.2 (5.0)
Hospital admission (months)	
January	8.5% (1878), mean: 3.8
February	7.7% (1702), mean: 4.2
March	7.6% (1676), mean: 4.1
April	7.5% (1665), mean: 4.1
May	7.9% (1747), mean: 3.9
June	5.9% (1316), mean: 4.5
July	8.5% (1898), mean: 3.8
August	9.1% (2002), mean: 3.6
September	9.5% (2103), mean: 3.8
October	9.2% (2027), mean: 3.7
November	8.1% (1799), mean: 3.4
December	9.0% (1900), mean: 3.2

Note: Data are presented as percentages (numbers), unless specified otherwise.

from the monthly distribution of patient admissions. In addition to these general findings, we examined the independent associations between various factors and LoS. We included seven features, and their importance was evaluated using a random forest model, as shown in Figure 2. The results indicated that age, location, and type of admission care have a more significant impact on LoS compared to other features.

The gradient boosting regressor demonstrated the best performance in terms of MSLE, closely followed by extreme gradient boosting. The support vector regressor achieved the lowest MAE, with values below 1. Notably, there is a significant disparity between the mean and median absolute error, attributable to the skewed distribution of the LoS (Table 4).

Our investigation into predicting LoS using Bayesian models yielded several noteworthy findings. Initially, we

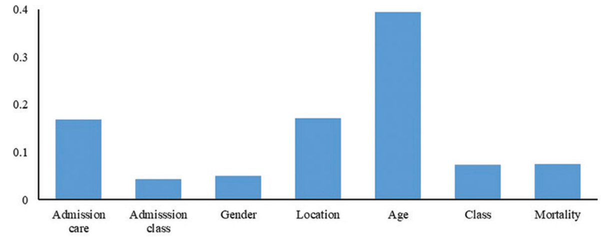

Figure 2. Significance of the random forest-based feature

examined the performance of various Bayesian models, each focused on a single predictor based on the results of a random forest feature importance analysis. Surprisingly, despite age being the most significant predictor in ML, we did not achieve satisfactory results when using it as the sole predictor. However, a linear model incorporating admission care as the predictor, with normal priors and a Poisson likelihood, emerged as the most effective combination (Table 5). Expanding our analysis, we integrated multiple predictors to construct Bayesian models. Utilizing the top five predictors identified by the random forest feature selector – age, admission care, location, gender, and mortality – we computed various combinations to predict LoS. Notably, a linear model incorporating age, admission care, location, and gender produced the most promising results (Table 5). While age stood out as the most significant predictor, our findings revealed that admission care level outperformed age when assuming a linear relationship with LoS in the Bayesian framework. This discrepancy prompted us to explore non-linear models by introducing squared, cubed, and exponential transformations of age. Subsequent analysis demonstrated that incorporating non-linearity in age prediction led to improved results compared to linear models (Table 5).

The optimal configuration among Bayesian models included four predictors: age, admission care, location, and gender, utilizing Gamma prior distributions and Poisson likelihood functions. This Bayesian model yielded results comparable to those of ML algorithms, achieving an MSLE of 0.25. Furthermore, the Bayesian approach offered robust performance across various error measures, underscoring its utility in healthcare decision-making. The incorporation of uncertainty information through predictive distributions enhanced the applicability of results within decision-theoretic frameworks, thereby augmenting the power and generalization of our findings.

4. Discussion

4.1. Key findings

The landscape of healthcare analytics is ever-evolving, largely driven by the exponential growth of patient data

Table 4. Predictive performance of machine learning models

Models	Mean squared log error	Root mean standard error	Mean absolute error	Median absolute error
Linear regression	0.27	2.39	1.68	1.31
Decision tree	0.24	2.22	1.53	1.25
Random forest	0.27	2.35	1.61	1.08
K-nearest neighbor	0.27	2.38	1.62	1.00
Support vector machine	0.24	2.38	1.45	0.96
Extreme gradient boosting	0.23	2.16	1.47	1.06
Adaptive boosting	0.29	2.28	1.72	1.74

Table 5. Predictive performance of Bayesian models

Model	Predictors	Distribution prior/likelihood	Mean squared log error	Root mean standard error	Mean absolute error
Linear regression					
Linear model-0 (with outliers)	0	Normal/Normal	0.42	4.39	2.47
Linear model-1	0	Normal/Normal	0.3	2.52	1.76
Linear model-2	0	Gamma/Normal	0.3	2.52	1.75
Linear model-3	0	Gamma/Gamma	0.3	2.52	1.75
Linear model-4	0	Chi-squared/Chi-squared	0.3	2.52	1.76
Linear model-5	0	Chi-squared/Poisson	0.3	2.51	1.74
Linear model-6	2	Normal/Poisson	0.3	2.48	1.73
Linear model-7	1	Normal/Poisson	0.25	2.21	1.56
Linear model-8	3	Normal/Poisson	0.3	2.49	1.75
Linear model-9	2	Normal/Chi-squared	0.29	2.45	1.73
Linear model-10	1	Normal/Chi-squared	0.27	2.35	1.65
Multivariate linear regression (MLR)					
MLR model-1	0, 2	Normal/Normal	0.44	4.74	2.55
MLR model-2	0, 2	Chi-squared/Chi-squared	0.4	4.76	2.36
MLR model-3	1, 2	Normal/Poisson	0.26	2.27	1.58
MLR model-4	0, 1	Normal/Poisson	0.26	2.24	1.57
MLR model-5	0, 1, 2	Chi-squared/Poisson	0.24	2.27	1.57
MLR model-6	0,1, 2	Gamma, normal, normal/Poisson	0.25	2.32	1.59
MLR model-7	0, 1, 2	Chi-squared, normal, normal/Poisson	0.25	2.28	1.57
MLR model-8	0, 1, 2	Chi-squared, normal, normal/Gamma			
MLR model-9	0, 1, 2, 4	Normal/Normal	0.25	2.22	1.55
MLR model-10	0, 1, 2, 4	Chi-squared/Chi-squared	0.26	2.23	1.6
MLR model-11	0, 1, 2, 4	Chi-squared/Poisson	0.25	2.21	1.55
MLR model-12	0, 1, 2, 4	Gamma/Poisson	0.25	2.20	1.54
MLR model-13	0, 1, 2, 4	Gamma/Poisson	0.25	2.24	1.56
MLR model-14	0, 1, 2, 3, 4	Chi-squared/Poisson	0.26	2.28	1.59
Non-linear regression (NLR)					
NLR model-1	0, 02	Chi-squared, Gamma/Normal	0.3	2.4	1.71
NLR model-2	$0, 0^2, 1, 2$	Chi-squared/Normal	0.26	2.3	1.59
NLR model-3	$0, 0^2, 1, 2$	Chi-squared, Gamma, Gamma/Poisson	0.25	2.26	1.56

(Cont'd...)

Table 5. *(Continued)*

Model	Predictors	Distribution prior/likelihood	Mean squared log error	Root mean standard error	Mean absolute error
NLR model-4	0, 0^2, 1, 2	Chi-squared, normal, Gamma/Chi-squared	0.26	2.27	1.61
NLR model-5	0, 0^3, 1, 2	Chi-squared, Gamma, Gamma/Poisson	0.26	2.26	1.58
NLR model-5	0, 0^2, 1, 1^2, 2	Chi-squared/Normal	0.28	2.66	1.73

Note: Predictor coding: 0: Age; 1: Admission care; 2: Location; 3: Mortality; 4: Gender.

and advancements in predictive methodologies. This study aimed to predict hospital LoS, a critical factor in healthcare resource management and patient care planning. We discovered that both Bayesian and ML models have good predictive power. Particularly, boosting models achieved the lowest MSLE, corroborating existing literature that highlights their superior regression performance. Conversely, support vector regression proved to be the most robust and widely applicable model, given its consistent performance across multiple evaluation metrics. While Bayesian models did not outperform ML algorithms in terms of error measures, they provided a useful measure of uncertainty. This is particularly helpful for medical professionals as the model can identify predictions that have a higher level of uncertainty, thus increasing effectiveness and reliability when making decisions.

4.2. Comparison with similar research

Our research adds to the growing body of literature on the prediction of hospital stay, offering a broader scope by utilizing a large pediatric dataset rather than focusing on a specific disease subgroup. The superior performance of ensemble methods like extreme gradient boosting aligns with other studies, highlighting its ability to model complex and non-linear relationships in healthcare data. The use of Bayesian modeling aims to fill a gap in existing literature by exploring its potential to complement traditional ML models with the added feature of uncertainty estimation. In addition, the integration of NER for feature engineering is a significant methodological contribution, improving the models' predictability while integrating recent advancements in NLP within healthcare. The comparative analysis in Table 6 reveals comparable performance scores in predicting LoS. Our approach stands out for its broader applicability and superior performance across various regression models and metrics.

4.3. Strengths and limitations

This study is distinguished by its comprehensive evaluation of predictive models, incorporating both traditional ML techniques and Bayesian inference methods. The use of Bayesian models for LoS prediction offers the added advantage of uncertainty quantification. By processing unstructured clinical text, NER strengthens the predictive power of our models, contributing to the broader field of NLP within healthcare. However, the study is limited by constraints inherent in the dataset, such as the lack of detailed patient medical conditions, relying instead on initial physician diagnoses. The absence of variables such as temperature, blood pressure, or other physiological indicators restricts the depth of analysis and predictive accuracy.

4.4. Implications and future actions

The findings of this study include several practical implications for hospital administrators, healthcare policymakers, and clinical practitioners aiming to improve operational efficiency and patient care. The predictive capabilities of both ML and Bayesian models have the potential for integrating such tools into operational hospital management systems. Accurate prediction of LoS can significantly enhance bed occupancy planning, reduce bottlenecks in patient flow, and optimize resource distribution across departments, particularly in resource-constrained settings like Pakistan.

Our study also paves the way for future research, particularly in developing hybrid models that combine the strengths of ML and Bayesian approaches. Moreover, the study highlights the value of transforming unstructured clinical text into structured features through NLP. Hospitals should prioritize digitization and structured documentation of clinical records to facilitate future predictive analytics applications. Building robust data infrastructures and standardizing terminology used in reason for visit and diagnostic notes would enhance model accuracy.

On a broader policy level, healthcare authorities should consider investing in model development and validation for other key outcomes, including readmissions and mortality. The integration of Bayesian frameworks, which allow uncertainty quantification, also presents opportunities for embedding probabilistic reasoning into clinical decision-support systems, offering a cautious and more reliable approach to automation in healthcare.

Finally, future work should aim to expand data collection to include real-time physiological metrics, laboratory

Table 6. Summary of results of prior studies

Study	Method used	Results	Data size
Turgeman *et al.*[18]	Regression tree (Cubist) model	MAE: 1.0, R^2: 0.79	20,321
Liu *et al.*[24]	Linear regression	MSE: 0.029, R^2: 0.146	155,474
Lee *et al.*[30]	Multivariable regression model	MAE: 7.6, RMSE: 11	22,824
Muhlestein *et al.*[31]	Gradient boosted trees, SVM, others	RMSLE: 0.631	41,222
Medeiros *et al.*[32]	Regression techniques, ML techniques	RMSE: 2.5 – 4.26	23,551
Alsinglawi *et al.*[33]	Gradient boosting, random forest, DNN	MAE: 2.0, R^2 0.81,	61,532
Zolbanin *et al.*[34]	DNN	MAE: 1.239, RMSE: 2.063 R^2: 0.613	86,338
Fang *et al.*[35]	Bayesian neural network	MAE: 1.955, R^2: 0.098	200,000
Muhlestein *et al.*[31]	ML algorithms	RMLSE: 0.661	41,222
Abdurrab *et al.*[36]	Bayesian regression versus ML regressors	RMSE 3.36, MAE 1.98	5,636
Lequertier *et al.*[37]	Embeddings+FFNN (deep learning)	Accuracy 0.944	515,199
Hu *et al.*[38]	Random forest, XGBoost, SVM, DNN (ML meta-analysis)	RMSE~5.8 – 7.0 R ~0.10 – 0.38	10,700,000
Rocheteau *et al.*[39]	TPC (deep time-series CNN)	MAE 1.55 – 2.28	MIMIC-IV database

Abbreviations: CNN: Convolutional neural network; DNN: Deep neural network; FFNN: Feed forward neural network; MAE: Mean absolute error; ML: Machine learning; MSE: Mean squared error; RMSE: Root mean squared error; RMSLE: Root mean squared log error; SVM: Support vector machine; TPC: Temporal pointwise convolutional; XGBoost: Extreme gradient boosting.

results, and longitudinal comorbidity information. These additional layers of data would likely improve predictive accuracy.

5. Conclusion

This study underscores the potential of both ML and Bayesian models in predicting hospital LoS, with each approach offering unique strengths. ML models, particularly boosting algorithms, have an excellent predictive accuracy, while Bayesian models offer valuable insights into prediction uncertainty. Integrating these models into healthcare systems could significantly improve resource management and patient care. Future research should explore hybrid approaches and incorporate more detailed patient data to further enhance predictive capabilities.

Acknowledgments

We extend our sincere gratitude to the management of Aga Khan University Hospital for providing the data to conduct this research. We are also thankful for their valuable subject expertise, which greatly contributed to the success of our study.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Tariq Mahmood, Babar Hasan, Zahra Hoodbhoy
Formal analysis: Sarmad Zafar, Tariq Mahmood, Zahra Hoodbhoy
Investigation: Sarmad Zafar, Tariq Mahmood, Zahra Hoodbhoy
Methodology: Sarmad Zafar, Tariq Mahmood, Babar Hasan
Writing–original draft: Sarmad Zafar, Tariq Mahmood
Writing–review & editing: All authors

Ethics approval and consent to participate

This study is retrospective in nature, utilizing historical data from patients admitted to the hospital, with all procedures conducted as part of routine care following established clinical practices. The data provided by Aga Khan University Hospital is anonymized to protect patient privacy, aligning with ethical guidelines and ensuring confidentiality. The hospital has obtained consent from all patients to use their data, without disclosing any identifying information, for research purposes.

Consent for publication

The Aga Khan University Hospital has obtained consent from all patients to use their data, without disclosing any identifying information, for research purposes, including the publication of results.

Availability of data

The data used for this research was provided by Aga Khan University Hospital under a confidentiality agreement. Data will be shared upon request to the corresponding author, subject to the permission from Aga Khan University Hospital and the signing of a Non-Disclosure Agreement (NDA) if required.

References

1. Chen M, Hao Y, Hwang K, Wang L, Wang L. Disease prediction by machine learning over big data from healthcare communities. *IEEE Access*. 2017;5:8869-8879.

   doi: 10.1109/access.2017.2694446

2. Jain R, Singh M, Rao AR, Garg R. Predicting hospital length of stay using machine learning on a large open health dataset. *BMC Health Serv Res*. 2024;24(1):860.

   doi: 10.1186/s12913-024-11238-y

3. Bopche R, Gustad LT, Afset JE, Ehrnström B, Damås JK, Nytrø Ø. In-hospital mortality, readmission, and prolonged length of stay risk prediction leveraging historical electronic patient records. *JAMIA Open*. 2024;7(3):ooae074.

   doi: 10.1093/jamiaopen/ooae074

4. Kothinti RR. Deep learning in healthcare: Transforming disease diagnosis, personalized treatment, and clinical decision-making through AI-driven innovations. *World J Adv Res Rev*. 2024;24(2):2841-2856.

   doi: 10.30574/wjarr.2024.24.2.3435

5. Van Houdenhoven M, Nguyen DT, Eijkemans MJ, *et al*. Optimizing intensive care capacity using individual length-of-stay prediction models. *Crit Care*. 2007;11(2):R42.

   doi: 10.1186/cc5730

6. Barnes SL, Hamrock E, Toerper MF, Siddiqui S, Levin SR. Real-time prediction of inpatient length of stay for discharge prioritization. *J Am Med Inform Assoc*. 2016;23(e1):e2-e10.

   doi: 10.1093/jamia/ocv106

7. Mahyoub MA, Dougherty K, Yadav RR, Berio-Dorta R, Shukla A. Development and validation of a machine learning model integrated with the clinical workflow for inpatient discharge date prediction. *Front Digit Health*. 2024;6:1455446.

   doi: 10.3389/fdgth.2024.1455446

8. Wessman T, Ärnlöv J, Carlsson AC, *et al*. The association between length of stay in the emergency department and short-term mortality. *Intern Emerg Med*. 2021;17(1):233-240.

   doi: 10.1007/s11739-021-02783-z

9. Arabi Y, Venkatesh S, Haddad S, Al Shimemeri A, Al Malik S. A prospective study of prolonged stay in the intensive care unit: Predictors and impact on resource utilization. *Int J Qual Health Care*. 2002;14(5):403-410.

   doi: 10.1093/intqhc/14.5.403

10. Gruenberg DA, Shelton W, Rose SL, Rutter AE, Socaris S, McGee G. Factors influencing length of stay in the intensive care unit. *Am J Crit Care*. 2006;15(5):502-509.

    doi: 10.4037/ajcc2006.15.5.502

11. Gustafson DH. Length of stay: Prediction and explanation. *Health Serv Res*. 1968;3(1):12-34.

12. Baek H, Cho M, Kim S, Hwang H, Song M, Yoo S. Analysis of length of hospital stay using electronic health records: A statistical and data mining approach. *PLoS One*. 2018;13(4):e0195901.

    doi: 10.1371/journal.pone.0195901

13. Awad A, Bader-El-Den M, McNicholas J. Patient length of stay and mortality prediction: A survey. *Health Serv Manag Res*. 2017;30(2):105-120.

    doi: 10.1177/0951484817696212

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

    doi: 10.1371/journal.pdig.0000017

15. Morton A, Marzban E, Giannoulis G, Patel A, Aparasu R, Kakadiaris IA. A Comparison of Supervised Machine Learning Techniques for Predicting Short-Term In-Hospital Length of Stay among Diabetic Patients. In: *2014 13th International Conference on Machine Learning and Applications*. IEEE; 2014:428-431.

    doi: 10.1109/icmla.2014.76

16. Ma F, Yu L, Ye L, Yao DD, Zhuang W. Length-of-stay prediction for pediatric patients with respiratory diseases using decision tree methods. *IEEE J Biomed Health Inform*. 2020;24(9):2651-2662.

    doi: 10.1109/jbhi.2020.2973285

17. Goh KH, Wang L, Yeow AYK, *et al*. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat Commun*. 2021;12(1):711.

    doi: 10.1038/s41467-021-20910-4

18. Turgeman L, May JH, Sciulli R. Insights from a machine learning model for predicting the hospital length of stay (LOS) at the time of admission. *Exp Syst Appl*. 2017;78:376-385.

    doi: 10.1016/j.eswa.2017.02.023

19. Fortuin V. Priors in bayesian deep learning: A review. *Int Statis Rev*. 2022;90:563-591.

    doi: 10.1111/insr.12502

20. Austin PC, Naylor CD, Tu JV. A comparison of a bayesian vs. A frequentist method for profiling hospital performance.

J Eval Clin Pract. 2001;7(1):35-45.

doi: 10.1046/j.1365-2753.2001.00261.x

21. Abdullah AA, Hassan MM, Mustafa YT. A review on bayesian deep learning in healthcare: Applications and challenges. *IEEE Access*. 2022;10:36538-36562.

    doi: 10.1109/access.2022.3163384

22. Ruhe D, Cinà G, Tonutti M, de Bruin D, Elbers P. Bayesian Modelling in Practice: Using Uncertainty to Improve Trustworthiness in Medical Applications. *arXiv*. Preprint posted online 2019.

    doi: 10.48550/arXiv.1906.08619

23. Yang Y, Yang KS, Hsann YM, Lim V, Ong BC. The effect of comorbidity and age on hospital mortality and length of stay in patients with sepsis. *J Crit Care*. 2010;25(3):398-405.

    doi: 10.1016/j.jcrc.2009.09.001

24. Liu V, Kipnis P, Gould MK, Escobar GJ. Length of stay predictions: Improvements through the use of automated laboratory and comorbidity variables. *Med Care*. 2010;48(8):739-744.

    doi: 10.1097/mlr.0b013e3181e359f3

25. Durango MC, Torres-Silva EA, Orozco-Duque A. Named entity recognition in electronic health records: A methodological review. *Healthc Inform Res*. 2023;29(4):286-300.

    doi: 10.4258/hir.2023.29.4.286

26. Mitchell T. *Machine Learning*; 1997. Available from: https://www.cs.cmu.edu/~tom/files/machinelearningtommitchell.pdf [Last accessed on 2025 Jul 21].

27. Bishop CM. *Pattern Recognition and Machine Learning*. Berlin: Springer; 2006.

28. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res*. 2014;15:3133-3181.

29. Williford E, Haley V, McNutt LA, Lazariu V. Dealing with highly skewed hospital length of stay distributions: The use of Gamma mixture models to study delivery hospitalizations. *PLoS One*. 2020;15(4):e0231825.

    doi: 10.1371/journal.pone.0231825

30. Lee H, Bennett MV, Schulman J, Gould JB, Profit J. Estimating length of stay by patient type in the neonatal intensive care unit. *Am J Perinatol*. 2016;33(8):751-757.

    doi: 10.1055/s-0036-1572433

31. Muhlestein WE, Akagi DS, Davies JM, Chambless LB. Predicting inpatient length of stay after brain tumor surgery: Developing machine learning ensembles to improve predictive performance. *Neurosurgery*. 2018;85(3):384-393.

    doi: 10.1093/neuros/nyy343

32. Medeiros NB, Fogliatto FS, Rocha MK, Tortorella GL. Forecasting the length-of-stay of pediatric patients in hospitals: A scoping review. *BMC Health Serv Res*. 2021;21(1):938.

    doi: 10.1186/s12913-021-06912-4

33. Alsinglawi B, Alnajjar F, Mubin O, et al. Predicting Length of Stay for Cardiovascular Hospitalizations in the Intensive Care Unit: Machine Learning Approach. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE; 2020:5442-5445.

    doi: 10.1109/embc44109.2020.9175889

34. Zolbanin HM, Davazdahemami B, Delen D, Zadeh AH. Data analytics for the sustainable use of resources in hospitals: Predicting the length of stay for patients with chronic diseases. *Inform Manag.* 2022;59(5):103282.

    doi: 10.1016/j.im.2020.103282

35. Fang J, Zhu J, Zhang X. Prediction of length of stay on the intensive care unit based on bayesian neural network. *J Phys Conf Ser*. 2020;1631(1):012089.

    doi: 10.1088/1742-6596/1631/1/012089

36. 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)*. 2024;12(2):249.

    doi: 10.3390/healthcare12020249

37. Lequertier V, Wang T, Fondrevelle J, Augusto V, Polazzi S, Duclos A. Length of stay prediction with standardized hospital data from acute and emergency care using a deep neural network. *Med Care*. 2024;62(4):225-234.

    doi: 10.1097/mlr.0000000000001975

38. Hu Z, Qiu H, Wang L, Shen M. Network analytics and machine learning for predicting length of stay in elderly patients with chronic diseases at point of admission. *BMC Med Inform Decis Mak*. 2022;22(1):62.

    doi: 10.1186/s12911-022-01802-z

39. Rocheteau E, Liò P, Hyland S. Temporal pointwise convolutional networks for length of stay prediction in the intensive care unit. In: *Proceedings of the Conference on Health, Inference, and Learning.* Association for Computing Machinery; 2021:58-68.

    doi: 10.1145/3450439.3451860