Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Construction of a Machine Learning-Based Model for Evaluating Virtual Bed Load in Comprehensive Hospitals

Yanling Wu
Guangzhou Twelfth People's Hospital
Kai Zhao
First Affiliated Hospital of Guangzhou Medical University
Jinkun Huang
First Affiliated Hospital of Guangzhou Medical University
Weiqing Huang
First Affiliated Hospital of Guangzhou Medical University
Yaotian Li
First Affiliated Hospital of Guangzhou Medical University
Hui Pan
First Affiliated Hospital of Guangzhou Medical University
Jianxin Yu
656820501@qq.com
First Affiliated Hospital of Guangzhou Medical University

Research Article

Keywords: Pre-admission, Virtual beds, Bed load assessment, Machine learning

Posted Date: May 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9534904/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Objective

This study examines optimization strategies for managing virtual bed load within the pre-admission model and develops a machine learning-based evaluation model to improve bed utilization efficiency and increase patient admission capacity.

Methods

We analyzed data from 15,988 pre-admission patients in surgical departments of a large tertiary hospital in 2024. Factors influencing bed efficiency, including patient discharge volume, bed count, and bed occupancy rate, were examined to construct evaluation indicators. By integrating static indicators, dynamic indicators, and derived indicators, we developed a nonlinear regression model using the XGBoost algorithm to establish a dynamic bed load threshold model. We used time-series cross-validation to evaluate the model's robustness and generalization capabilities. We employed SHAP value analysis to interpret the model.

Results

Statistical analysis revealed that 94.2% of patients completed the transition to formal admission within 10 days, although departments such as Thoracic Surgery 9 experienced longer average processing times. Logistic regression analysis identified bed occupancy rate, Case Mix Index, bed turnover rate, and bed count as independent factors influencing prolonged processing time ($P < 0.001$). The XGBoost prediction model achieved an average RMSE of 4.55 and an average R^2 of 0.91 across various time periods, demonstrating strong predictive accuracy and stability. This prediction model functions as an early warning mechanism, facilitating real-time monitoring of bed load peaks and providing a scientific basis for bed resource allocation.

Conclusion

The XGBoost-based pre-admission bed load evaluation model developed in this study effectively predicts bed load peaks, offering valuable decision support for hospital administrators to optimize bed resource allocation. These findings are crucial for enhancing medical service efficiency and improving health resource utilization.

1 Introduction

In 2017, the General Office of the State Council of China released the "Guiding Opinions on Establishing a Modern Hospital Management System". This policy document emphasized the necessity of prioritizing public health and preserving the public welfare character of public hospitals, while advocating for an organic unity between social benefits and operational efficiency. In 2021, the National Health Commission and the National Administration of Traditional Chinese Medicine jointly released the "Action Plan for Promoting High-Quality Development of Public Hospitals (2021–2025)". This document outlined initiatives aimed at enhancing patient experience through the establishment and improvement of ten medical service systems, including appointment-based diagnosis and treatment, to increase the accessibility and convenience of medical services. In recent years, several Chinese hospitals have implemented pre-admission management strategies for elective surgery patients to reduce admission waiting times and enhance operational efficiency[1]. This approach involves allocating "pre-admission" bed numbers within each surgical department, while physical beds are not available during the preoperative examination period. After completing preoperative assessments, patients are evaluated by attending physicians and undergo pre-anesthesia consultations with anesthesiologists prior to formal admission to the ward. Statistical analyses have demonstrated that the pre-admission model significantly decreases bed waiting times, shortens

the average length of stay, and reduces hospitalization costs[2]. This model has gained widespread acceptance among hospital administrators, department heads, and medical insurance authorities.

With the widespread implementation of this model, operational challenges have emerged. Because "pre-admission" beds operate as virtual beds, physicians in outpatient clinics often lack awareness of virtual bed saturation and may issue pre-admission orders based solely on patients' medical needs. Consequently, backlogs have occurred in certain departments, where patients who have completed preoperative examinations face delays in formal admission. In certain cases, waiting times have exceeded the standards established by medical insurance management.

This study addresses the paradox of virtual bed saturation within the pre-admission management model. We investigated the factors affecting pre-admission bed utilization efficiency and developed a resource allocation model based on key variables, including annual patient discharges, physical bed counts, bed utilization rates, bed turnover rates, and the Case Mix Index (CMI)[3]. Although previous research has addressed bed resource allocation, applying machine learning techniques to dynamically assess virtual pre-admission bed load and thereby resolving the conflict between clinical admissions and bed saturation—remains underexplored[4, 5]. Current management practices rely on experiential judgments and lack data-driven, proactive early warning systems, which constitutes the primary issue addressed in this study[6]. To the best of our knowledge, this represents the first comprehensive investigation into factors influencing virtual bed allocation within the context of widespread pre-admission management implementation. These findings will assist hospital administrators in optimizing resource allocation, streamlining patient care processes, enhancing service efficiency, and improving patient satisfaction.

2 Data Sources and Methods

2.1 Data Sources

This study used inpatient data from pre-admission patients in surgical departments of a large tertiary hospital in Guangzhou, collected in 2024. The hospital implemented the pre-admission model in selected departments in 2019 and subsequently expanded it throughout the institution. Considering the temporal distribution of pre-admission patients before formal admission, this study focused exclusively on surgical department patients. Analysis of the hospital's Health Information System (HIS) revealed that 30540 patients used the pre-admission model in 2024. Inclusion criteria, Inpatients in surgical departments, and patients registered in the pre-admission system. Exclusion criteria, patients with a time from pre-admission registration to formal admission of less than 1 day, or patients who did not complete formal admission within 20 days. The final sample comprised 15988 patients from 32 surgical wards, as shown in Fig. 1.

2.2 Evaluation Indicators

According to the resource-demand matching theory[7], bed utilization and turnover rates are direct indicators of bed resource strain. The Case Mix Index (CMI) and annual patient discharges reflect patient resource demands and departmental service volume. When resource supply fails to meet demand—particularly under high load conditions, patient admission waiting times are likely to increase. This scenario constitutes the foundational hypothesis for model construction. Based on this hypothesis, the primary factor contributing to excessive pre-admission waiting times is likely insufficient bed turnover efficiency, which hinders physical bed availability. The national performance evaluation system for public hospitals identifies key bed efficiency indicators, including patient discharges, bed count, and bed utilization rate. This study used the 2024 Case Mix Index (CMI) for each specialty as the variable for evaluating cross-specialty resource consumption[8].

2.3 Statistical Methods

To analyze pre-admission admission thresholds, we employed a multi-stage modeling approach. First, data preprocessing was conducted. The Shapiro-Wilk test was used to assess the normality of continuous variables, specifically the Case Mix

Index (CMI) and bed utilization rate. Meanwhile, categorical variables, including department type and the duration from pre-admission to formal admission, were defined and categorized accordingly. Second, derivative indicator calculations were performed.

$$Workload\ Index = \mathrm{CMI} \times \left\lfloor \frac{Annual\ Discharges}{365} \right\rfloor$$

1

.....

$$Bed\ Turnover\ Density = \frac{Annual\ Turnover\ frequency}{bed\ Count}$$

2

.....

Subsequently, we constructed the predictive model by employing the XGBoost algorithm to develop a nonlinear regression model. This algorithm integrates predictions from multiple weak learners, typically decision trees[9, 10]. In each iteration, a new decision tree is created to address the residual errors of the previous tree, while successive iterations progressively correct residual errors. The weak learners are combined using a weighted summation approach to form the final model[9]. Furthermore, the model's hyperparameters were optimized using Bayesian methods[10]. Finally, we defined a safety margin function to calculate the final threshold as follows:

$$Margin\,(u,d) = \begin{cases} 0.65 - 0.1 \times \min\left(\frac{d}{1000}, 1\right) & u > 0.9 \\ 0.75 & 0.8 < u < 0.9 \\ 0.85 + 0.05 \times \min\left(\frac{d}{2000}, 1\right) & others \end{cases}$$

3

.....

u represents the real-time bed utilization rate, *d* denotes the number of patient discharges over the past 30 days, min() function takes the minimum value. For instance, min(*d*/1000,1) caps the ratio at 1 when d exceeds 1000. The segmented thresholds (0.65, 0.75, 0.85) are empirical constants derived from historical data distributions and clinical management experience. These thresholds provide varying safety buffer margins under different load conditions.

Threshold = max (ŷ × $\mathrm{Margin}\,(\mathrm{u}, \mathrm{d})$ × Department Calibration Coefficient, 0.3 × Bed Count) ..... (4)

In this equation, ŷ represents the model-predicted value, and the Department Calibration Coefficient is a regression coefficient calculated from each department's historical overdue admission rate. Furthermore, the term 0.3 × Bed Count denotes a predetermined minimum threshold, which ensures that a limited number of patients can still be admitted even when predicted values are extremely low.

2.4 Model Training Process and Validation

We used pre-admission patient data from 2024 (n = 15988) to assess the model's predictive performance and robustness. Within this dataset, we employed time series cross-validation (*TSCV*) for model training and hyperparameter optimization to capture time-varying patterns while preventing overfitting[11].

During hyperparameter tuning cross-validation, model training performance was evaluated using multiple metrics. We used Root Mean Square Error (*RMSE*) to quantify deviations between model predictions and actual values, with smaller values indicating superior performance. Mean Absolute Error (*MAE*) was also calculated to evaluate prediction errors. Additionally,

R-squared (R^2) was used to represent the proportion of variability explained by the model, with values approaching 1 indicating better model fit. To improve model interpretability and understand feature contributions to predictions, we employed SHAP (*Shapley Additive exPlanations*) value analysis during model evaluation[12]. SHAP values were computed for all samples in the test set. Using these values, we created global feature importance plots, feature dependence plots, and SHAP force plots to clarify the model's decision-making rationale from global and local perspectives. Descriptive statistics were performed using *SPSS* (version 22). We developed the model using the "xgboost" package in R software (version 4.3.1) and computed SHAP values using the "shap" package. Model visualization was created using the "shinydashboard" package[13]. Statistical significance was established at $P < 0.05$.

3 Results

3.1 Time to Formal Admission

Guangzhou's medical insurance regulations require formal admission within 10 days following registration. During the pre-admission period, laboratory tests and examinations are uniformly covered under inpatient expenses. Statistical analysis revealed that 94.2% of patients in surgical departments achieved formal admission within 10 days in 2024. Bed turnover efficiency was notably high in the departments studied, with 68.6% of patients admitted within 1 to 7 days. Additionally, 15.7% of patients were admitted between 7 and 10 days. We further analyzed which surgical departments had higher patient concentrations within this timeframe. The detailed distribution is shown in Table 1.

Table 1
Distribution of Time from Pre-admission to Formal Admission

Time to Formal Admission (days)	Number of Patients, n	Percentage, %
≥ 1, < 4 days	7495	46.9
≥ 4, < 7 days	5066	31.7
≥ 7, ≤ 10*	2503	15.7
> 10, < 20	924	5.8

Time distribution across specific departments is detailed in Table 2. Eleven departments or wards, including Breast Surgery, Hepatobiliary Surgery, Ophthalmology, and Thoracic Surgery, had higher proportions of borderline cases. In these departments, the percentage of patients admitted between 7 and 10 days exceeded 10%. Notably, Thoracic Surgery Ward 2 and Thoracic Surgery Ward 9 had particularly extended waiting times, with 20.1% and 22.7% of patients experiencing delayed admissions, respectively.

Table 2
Departments Characterized by Excessively Prolonged Waiting Times in Pre-Hospitalization Admission Procedures

group	BRS (*n*, %)	ENT (*n*, %)	GIS (*n*, %)	HPB2(*n*, %)	OD (*n*, %)	TS1(*n*, %)	TS2(*n*, %)	TS3(*n*, %)	TS5(*n*, %)	TS8(*n*, %)	TS9(*n*, %)	TS10(*n*, %)
≥ 1,<4 days	258, 53.5	274, 55.6	256, 61.0	100, 57.1	19, 16.2	403, 40.2	421, 26.9	165, 37.7	676, 57.0	288, 55.0	590, 28.9	142, 26.2
≥ 4,<7 days	111, 23.0	135, 24.9	120, 28.6	38, 21.7	62, 53.0	329, 32.8	529, 33.8	158, 36.1	317, 26.7	167, 31.9	543, 26.6	249, 45.9
≥ 7, ≤ 10 days	91, 18.9	57, 11.6	33, 7.9	35, 20.0	32, 27.4	210, 20.9	449, 28.7	104, 23.7	162, 13.6	57, 10.9	481, 23.5	119, 22.0
> 10, < 20 days	22, 4.6	27, 5.5	11, 2.6	2, 1.1	4, 3.4	61, 6.1	165, 10.5	11, 2.5	32, 2.7	12, 2.3	430, 21.0	32, 5.9

*BRS, Breast Surgery; ENT, Otorhinolaryngology - Head and Neck Surgery; GIS, Gastrointestinal Surgery; HPB, Hepatobiliary Surgery; OD, Ophthalmology; TS, Thoracic Surgery

Table 3
Clinical department service efficiency indicators

Department	Bed count	Bed Turnover rate	Bed occupancy rate	CMI	Average Admission Conversion Time, days	*Sd.*
Otorhinolaryngology - Head and Neck Surgery	44	48.64	0.7483	0.98	4.04	0.14
Gynecology	46	40.91	0.6124	0.89	2.85	0.09
Hepatobiliary Surgery1	43	43.33	0.9654	2.2	2.86	0.17
Hepatobiliary Surgery2	47	36.91	0.9023	1.86	3.82	0.20
Hepatobiliary Surgery3	44	44.18	0.9161	1.42	3.79	0.14
Urology1	45	54.47	0.8868	1.24	2.81	0.09
Urology2	29	65.49	0.9135	1.21	2.99	0.08
Urology3	47	55.79	1.0159	1.39	4.75	0.10
Urology4	18	63.75	0.983	1.28	2.73	0.13
Urology5	25	33.99	0.3691	1.54	3.99	0.14
Female Pelvic Medicine and Reconstructive Surgery	21	33.33	0.5803	1.1	3.15	0.14
Breast Surgery	33	84.37	0.9215	0.85	4.23	0.15
Joint Surgery	56	27.26	0.7134	1.76	3.85	0.15
Spine Surgery	44	22.86	0.6785	1.78	4.57	0.36
Department of General Surgery	11	31.33	0.6493	1.28	3.16	0.17
Gastrointestinal Surgery	46	50.57	0.8006	1.27	3.40	0.13
Cardiac Surgery	69	14.72	0.6236	4.86	5.88	2.30
Vascular Surgery	46	54.37	0.7716	1.8	3.00	0.09
Thoracic Surgery1	40	47.78	0.745	2.78	4.87	0.10
Thoracic Surgery2	70	54.3	0.7578	2.17	5.95	0.09
Thoracic Surgery5	48	64.75	0.8715	2.25	3.80	0.08
Thoracic Surgery6	45	86.51	1.003	1.62	3.65	0.07
Thoracic Surgery8	45	40.24	0.6284	2.41	3.79	0.11
Thoracic Surgery9	48	101.1	1.0766	2.14	6.58	0.09
Thoracic Surgery10	49	35.9	0.9546	3.92	5.39	0.12
Plastic Surgery	8	65.63	0.613	0.79	2.46	0.12

3.2 Clinical Department Service Efficiency

3.3 Analysis of Factors Contributing to Prolonged Time to Formal Admission

We conducted a retrospective comparative study to examine the impact of extended time to formal admission on service efficiency across 11 departments. Patients were divided into two groups based on admission delay: exceeding 10 days and within 10 days. We used multivariate logistic regression analysis to identify key influencing factors. As detailed in Table 4, bed utilization rate, *CMI*, bed turnover rate, and bed count independently influenced prolonged admission time ($P < 0.001$). Notably, a higher CMI was associated with an odds ratio of 2.704 (95% confidence interval: 2.306–3.171)[14]

Table 4
Multivariate Logistic Regression Analysis of Factors Associated with Prolonged Time to Formal Admission

Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Bed occupancy rate	-3.476	.613	32.199	1	< 0.001	.031	.009	.103
CMI	.995	.081	150.115	1	< 0.001	2.704	2.306	3.171
Bed Turnover rate	.054	.004	181.607	1	< 0.001	1.056	1.047	1.064
Bed count	.029	.003	103.590	1	< 0.001	1.029	1.024	1.035
Constant	-6.069	.356	290.976	1	< 0.001	.002	-	-

As elaborated in the Introduction, pre-admission beds function as virtual beds, thereby complicating clinicians' ability to assess whether physical bed demand has reached saturation during patient admissions. To address this challenge, we developed a dynamic threshold model for real-time monitoring of pre-admission patient numbers. Specifically, when the patient count reaches the established threshold, an alert is triggered to ensure that bed waiting times do not exceed 10 days. The dynamic threshold model for pre-admission patient load is constructed based on the XGBoost prediction framework. At the data layer, we established three indicator categories: (1) Static indicators, which include department bed count, annual average CMI, and bed turnover rate; (2) Dynamic indicators, comprising real-time bed utilization rate (U_t) and 30-day patient discharges (D_{30}). The dynamic threshold model is formulated as follows:

$$\hat{y} = f(CMI, U_t, D_{30}, Turnover_Rate, Bed_Count, Workload_Index) \quad (5)$$

Based on model calculations, we established the dynamic safety margin using the safety margin function Margin (U_t, D_{30}), as outlined in Eq. 3. We then computed the threshold (T). This model incorporates a three-tier early warning mechanism for pre-admission bed load: the red level signifies that T is less than the current pre-admission patient count, the yellow level indicates a tolerance band of T ± 10%, and the green level indicates that T exceeds the current count by more than 10%.

We visualized the dynamic threshold model for pre-admission patient load using a Shiny application in R. The Shiny dashboard comprises five modules: Department Threshold Analysis, Real-time Warning Monitoring, Model Performance, Intervention Effect Simulation and Data Overview.

(1) Department Threshold Analysis Module

By inputting dynamic indicators for a specific ward, such as current bed utilization rate and patient discharges over the past 30 days, the model automatically computes the pre-admission bed threshold for that ward. This calculation uses the department's static indicators, including CMI, turnover rate, and workload index, with results displayed on the dashboard. For instance, in Thoracic Surgery Ward 9, when the bed utilization rate is 95% and there have been 150 patient discharges in the past 30 days, the recommended pre-admission bed threshold is 29 with a minimum protection threshold of 14, as illustrated in Fig. 2.

(2) Real-time Warning Monitoring Module

The current number of pre-admission patients in a specific ward is compared to the designated threshold (T) for that ward, with load statuses represented by color-coded blocks. As shown in Fig. 3, when the number of pre-admission patients is 28 and the admission threshold is 29, the system indicates normal status. Additionally, the dashboard bottom displays the current number of pre-admission patients and warning status for each ward.

(3) Model Evaluation

We used time series cross-validation (*TSCV*) to evaluate the model's robustness and generalization capability. Data was divided chronologically into three folds, with the model trained on earlier folds and tested on subsequent ones.

Cross-validation results were assessed using Root Mean Square Error (*RMSE*) to quantify deviations between model predictions and actual values, Mean Absolute Error (*MAE*) to evaluate prediction errors, and R-squared (R^2) to indicate the proportion of variability explained by the model, with values approaching 1 indicating better model fit. The model achieved an average RMSE of 4.55 across various time periods and an average R^2 of 0.91, demonstrating good predictive accuracy and stability. Performance variation among folds was minimal, with an RMSE standard deviation of 0.12, indicating that the model is not sensitive to temporal changes and demonstrates robust generalization capability. This performance level is sufficiently reliable for practical applications in pre-admission bed management, as shown in Table 5 and Fig. 4 of the Shiny dashboard.

Table 5
Evaluation of the robustness and generalization ability of machine learning models

Fold	RMSE	MAE	R_squared
1.00	4.84	3.51	0.90
2.00	4.33	3.35	0.92
3.00	4.49	3.25	0.92

(4) Model Interpretability Analysis Based on SHAP Values

XGBoost generally provides superior accuracy to linear models in predictive tasks, but this advantage comes at the cost of interpretability, which is a hallmark of linear models. Consequently, XGBoost is frequently regarded as a black-box model. In 2017, Lundberg and Lee introduced SHAP values, a versatile technique for interpreting models including classification and regression frameworks[15]. This method has been especially advantageous for elucidating previously opaque black-box models, such as XGBoost and neural networks[16].

Figure 6 SHAP Dependence Plot

Figure 7 SHAP Force Plo

Figure 5 SHAP Summary Plot

We used SHAP values to interpret the developed XGBoost model. As shown in Fig. 5, global importance analysis based on SHAP values indicates that workload index, physical bed count, and annual patient discharges are the three primary factors affecting pre-admission bed load prediction, evidenced by their mean absolute SHAP values, which are significantly higher than those of other features. This supports our logistic regression analysis findings and further elucidates their prominent role in nonlinear models. As shown in Fig. 6, workload index shows a distinct positive correlation with SHAP values. When workload index falls within the 0–10 range, its SHAP value increases and plateaus after reaching a positive value. This trend indicates that a high workload index is the most significant factor driving the model's high load predictions. Figure 7 presents SHAP force plot analysis for Thoracic Surgery Ward 9, which recorded the highest load prediction (base prediction

value f(x) = 70.56). The analysis reveals that higher physical bed count (+ 9.47), exceedingly high annual patient discharges (+ 7.97), and elevated workload index (+ 2.48) collectively elevated its prediction value above the red warning threshold. This finding intuitively highlights the management challenges faced by this department due to multiple overlapping pressures.

4 Discussion

4.1 Contradiction Between Pre-admission Model Promotion and Time to Formal Admission

In 2021, Zhao et al.[17] summarized the implementation of the pre-admission model at a tertiary general hospital in Guangzhou. Their findings indicated that this model effectively shortened patient hospital stays, reduced medical costs, and enhanced service efficiency. Their systematic overview of the pre-admission model addressed patient care processes, centralized renovation of pre-admission examination and testing facilities, information system optimization, medical insurance policy improvements, and establishment of inclusion and exit criteria for pre-admission patients. In recent years, the pre-admission model has been adopted by numerous hospitals nationwide, receiving positive recognition from patients, hospital administrators, and medical insurance managers. However, studies have noted that because "pre-admission" beds are essentially virtual, physicians often cannot perceive bed saturation when consulting patients in outpatient settings, resulting in pre-admission patient accumulation. Research data indicates that in 2024, 5.8% of patients failed to complete formal admission within the designated timeframe, meaning preoperative examination expenses for these patients were not reimbursable by the medical insurance fund. This situation inadvertently increases patient economic burden and negatively impacts their satisfaction with medical services. Guided by a patient-centered service philosophy, hospital administrators must prioritize addressing pre-admission bed load saturation and develop appropriate measures to resolve this conflict.

4.2 Preliminary Construction of a Prediction Model for Pre-admission Bed Load Threshold

Increasing numbers of studies utilize interpretable machine learning models to elucidate prediction mechanisms. XGBoost, an efficient gradient boosting algorithm, is frequently employed to develop disease diagnosis and prognosis models. For instance, Wang et al.[18] applied this algorithm to create an early gastric cancer diagnosis model, wherein deep learning algorithms enhanced diagnostic accuracy. Similarly, a study constructing a ventilator-associated pneumonia risk prediction model for elderly ICU patients employed algorithms including Extreme Gradient Boosting (XGBoost) and Random Forest (RF). Their findings indicated that the XGBoost-based risk prediction model demonstrated superior predictive performance, achieving an AUC of 0.99[19]. By integrating these predictive factors and derived indicators, we used an XGBoost prediction model to forecast dynamic thresholds for pre-admission virtual bed allocation. The model achieved an average RMSE of 4.55 across various time periods and an average R^2 of 0.91, demonstrating robust predictive accuracy and stability. Minimal performance variation between folds suggests that the model is not sensitive to temporal changes and demonstrates strong generalization capability.

4.3 Countermeasures for Handling Peak Pre-admission Bed Load

As previously noted, departments admitting complex and severe cases (high CMI) are more susceptible to virtual bed congestion. Consequently, implementing targeted precision management strategies for high CMI departments is essential, including early warning pre-positioning and resource pre-matching. For these departments, warning thresholds should be established more conservatively. A pre-admission green channel for high CMI patients should be created, accompanied by early multidisciplinary team (MDT) assessments, and hospital-wide bed resources should be prioritized when bed utilization rates approach critical thresholds. Resource pools should be dynamically reconstructed based on data. In the long term, hospitals must continuously evaluate and adjust human and equipment resource pools for intensive care and complex surgeries according to CMI load prediction data, aligning resource supply with case complexity to alleviate system-level bottlenecks.

Robust bed load prediction models can significantly assist hospital administrators and department heads in monitoring and allocating bed resources. Medical management departments should prioritize promoting a one-bed-within-discipline model, transitioning to a one-bed-throughout-hospital approach. Initially, the focus should be on advancing the one-bed management model. For disciplines with multiple wards, promoting one bed within the discipline is essential. In wards with low load thresholds, other wards within the same discipline should provide physical beds and personnel to expedite pre-admission patient admission. Furthermore, based on prediction model evaluation results, it is necessary to reassess physical bed allocation rationality in various wards and increase bed numbers in those with low thresholds. Lastly, the information department should integrate the prediction model into the hospital's Health Information System (*HIS*) to enable real-time reminders, assisting physicians in making informed patient admission decisions.

4.4 Study Limitations and Future Prospects

This study examined CMI influence on virtual bed load, demonstrating scientific rigor. Additionally, it represents the first investigation into virtual bed load, reflecting a forward-looking perspective. However, this study has notable limitations. First, the model was developed using data from a single tertiary class-A general hospital, which may limit absolute threshold value applicability to other healthcare institutions. However, the model's methodology—specifically, the framework for constructing dynamic warning models based on universal indicators such as CMI and bed utilization rate—demonstrates generalizability. When implemented in other hospitals, local data calibration will be necessary. Future research should aggregate multi-center data to assess covariate impacts such as hospital scale, type, and region on the model, aiming to establish national standards with broader adaptability. Furthermore, this study focused on demonstrating XGBoost algorithm applicability for this issue. Although its performance is commendable, systematic comparison with emerging algorithms (such as LightGBM, CatBoost, and deep neural networks) or ensemble learning/stacking models was not performed, leaving exploration of potential algorithm combinations with superior performance or greater efficiency for future investigation.

5 Conclusion

In conclusion, this study analyzed factors influencing delayed pre-admission patient conversion by employing XGBoost to develop a virtual bed load threshold model and assess its robustness. By integrating the SHAP interpretability framework, the study not only created a high-accuracy prediction model but also elucidated the nonlinear influence patterns of key factors—workload index, CMI, and physical bed count—on load prediction. The evaluation produced an efficient and precise prediction tool capable of early detection of virtual bed load peaks. Its warning function holds significant value for medical insurance compliance by alerting hospital administrators to implement preventive measures and mitigate conflicts. This prediction tool demonstrates innovation and is suitable for promotion in public hospitals.

Declarations

Acknowledgments

The authors acknowledge the clinicians who took part in this analysis.

Funding*1. This work was supported by the 2024 Guangdong Provincial Universities characteristic innovation project (NO. 2024WTSCX072)*

2.Guangzhou Medical University's 2024 Research Capacity Enhancement Program NO. 2024SRP241

Clinical trial number: *not applicable*

Human Ethics and Consent to Participate declarations: *not applicable*

Competing interests: *None*

References

1. B B C A, B E D, C J B B. Operating room planning and scheduling: A literature review - ScienceDirect[J]. European Journal of Operational Research, 2010,201(3):921–932.
2. Hui P, Jianxin Y, Zhina R, et al. Practice of pre-hospitalization in operation department of a hospital in Guangzhou[J]. Chinese Journal of Hospital Administration, 2021,37(3):229–231.
3. Litvak E, Bisognano M. More Patients, Less Payment: Increasing Hospital Efficiency In The Aftermath Of Health Reform[J]. Health Affairs, 2011,30(1):76–80.
4. Mcmanus M L, Long M C, Cooper A, et al. Queuing theory accurately models the need for critical care resources.[J]. Anesthesiology, 2004,100(5):1271–1276.
5. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning for electronic health records[J]. npj Digital Medicine, 2018,1(1):18.
6. Reddy S, Fox J, Purohit M P. Artificial intelligence-enabled healthcare delivery[J]. Journal of the Royal Society of Medicine, 2019,112(1):22–28.
7. Zhao Z, Yang K. Suitability matching evaluation of the low-carbon regional integrated energy system based on the 'resource-demand' balance[J]. Sustainable energy technologies and assessments, 2024(Dec.):72.
8. Hughes J S, Averill R F, Goldfield N I, et al. Identifying potentially preventable complications using a present on admission indicator[J]. Health care financing review, 2006,27(3):63–82.
9. Chen T, Guestrin C. XGBoost[Z]. 2016.
10. Bergstra J, Yamins D, Cox D. Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms[Z]. 2013.
11. Bergmeir C, Benítez J M. On the use of cross-validation for time series predictor evaluation[J]. Information Sciences, 2012,191:192–213.
12. Nohara Y, Matsumoto K, Soejima H, et al. Explanation of machine learning models using shapley additive explanation and application for real data in hospital[J]. Computer Methods and Programs in Biomedicine, 2022,214:106584.
13. Ghelfi A, Isobe S. Hayai-Annotation: A functional gene prediction tool that integrates orthologs and gene ontology for network analysis in plant species[J]. Computational and Structural Biotechnology Journal, 2025,27:117–126.
14. Clement J P V R. Executive summary of the Federal Agency for Healthcare Research and Quality (AHRQ) H-CUP project[J]. Journal of Ambulatory Care Management, 2009,4(32):307–317.
15. Lundberg S M, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees[J]. Nature Machine Intelligence, 2020,2(1):56–67.
16. Luo X, Li B, Zhu R, et al. Development and validation of an interpretable machine learning model for predicting in-hospital mortality for ischemic stroke patients in ICU[J]. International Journal of Medical Informatics, 2025,198:105874.
17. Min T, Shuang L, Kai Z, et al. Practice and Exploration the Effect of Pre-Hospitalization in a Class A Tertiary Hospital under the Con cept of Value[J]. Chinese Hospital Management, 2022,42(11):56–58.
18. Shuyu W, Hongyang D, Chenhao Z, et al. Establishment and evaluation of gastric cancer prediction models based on deep learning and machine learning[J]. Journal of QingDao University(Medical Sciences), 2025,61(1):54–58.
19. Wensi G, Yirui H, Xiaoqing W, et al. Establishment of risk prediction model for ventilator-associated pneumonia in elderly ICU patients based on machine learning and its performance[J]. Chinese Journal of Nosocomiology, 2025(2):290–296.

Figures

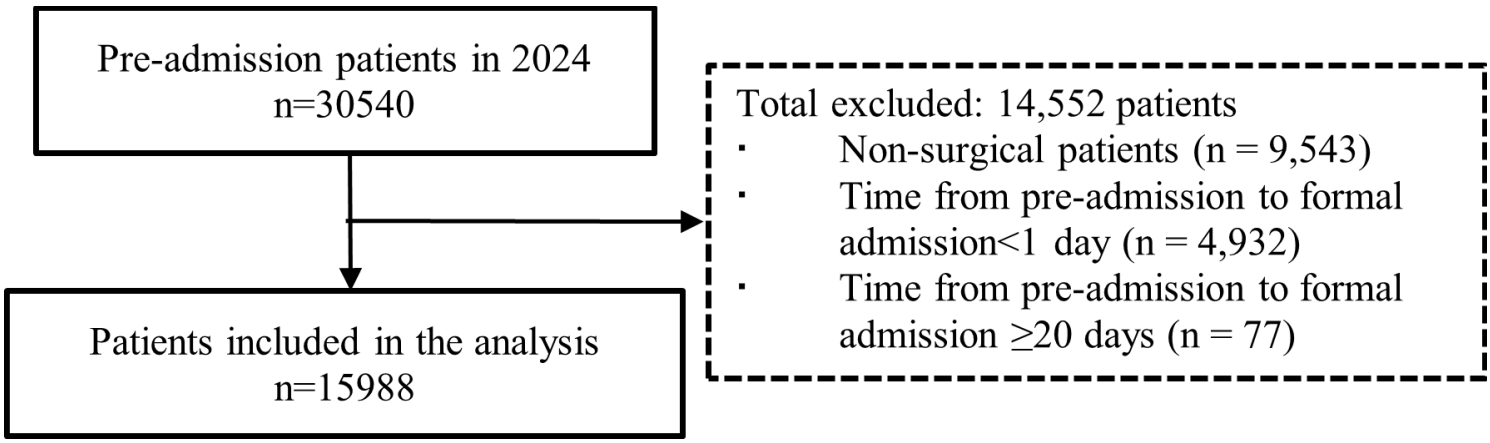


Figure 1

Flow chart of patient selection



Figure 2

Department Threshold Analysis Panel

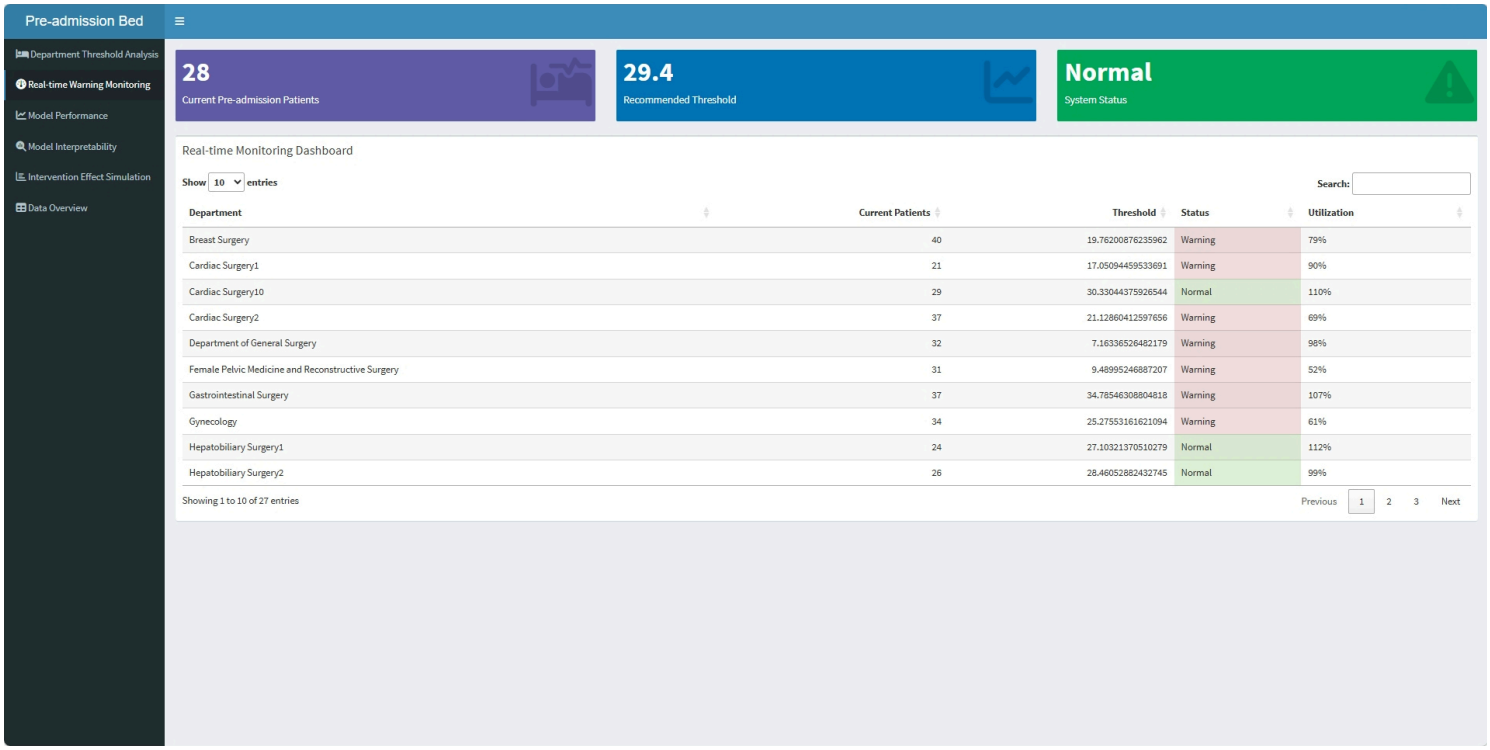


Figure 3

Real-time early warning monitoring panel

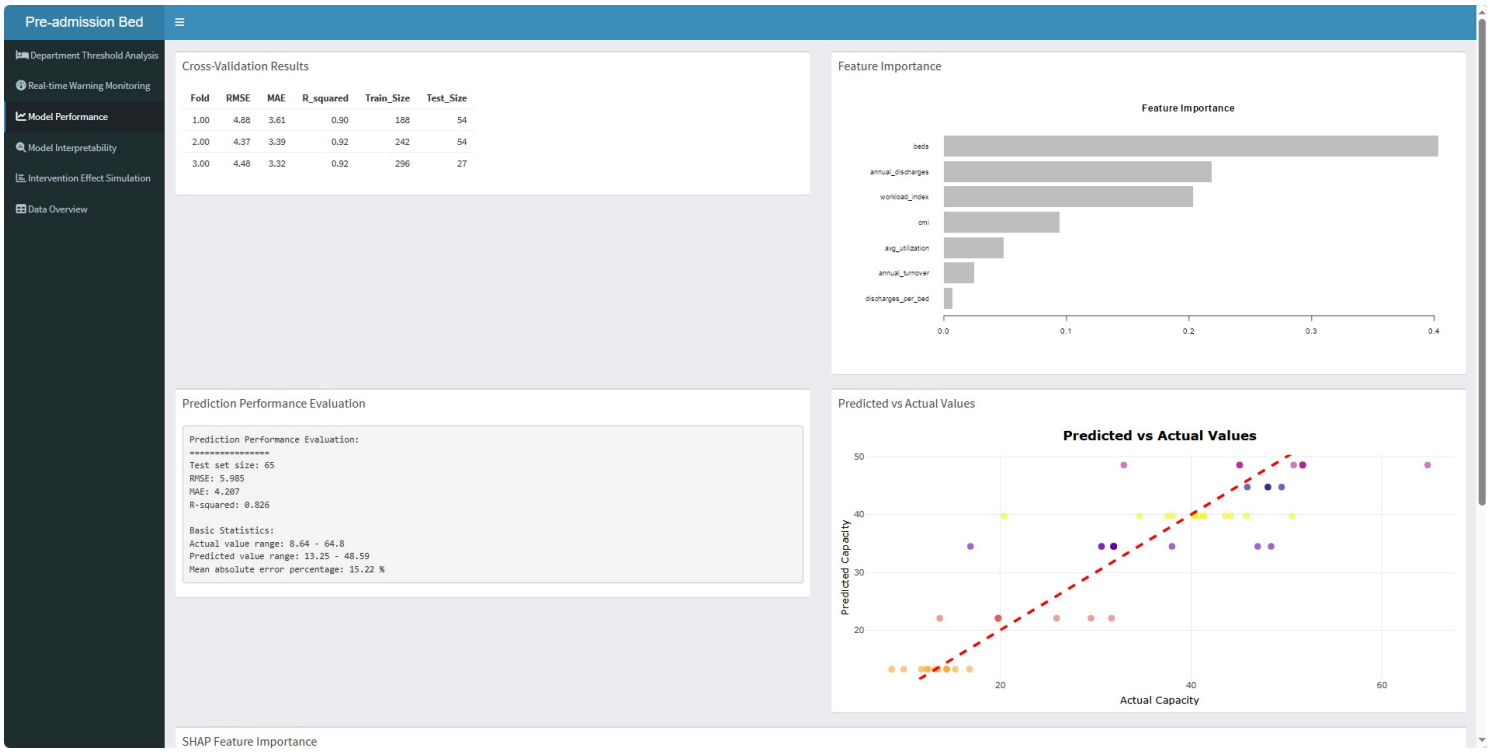


Figure 4

Performance Evaluation Model Panel

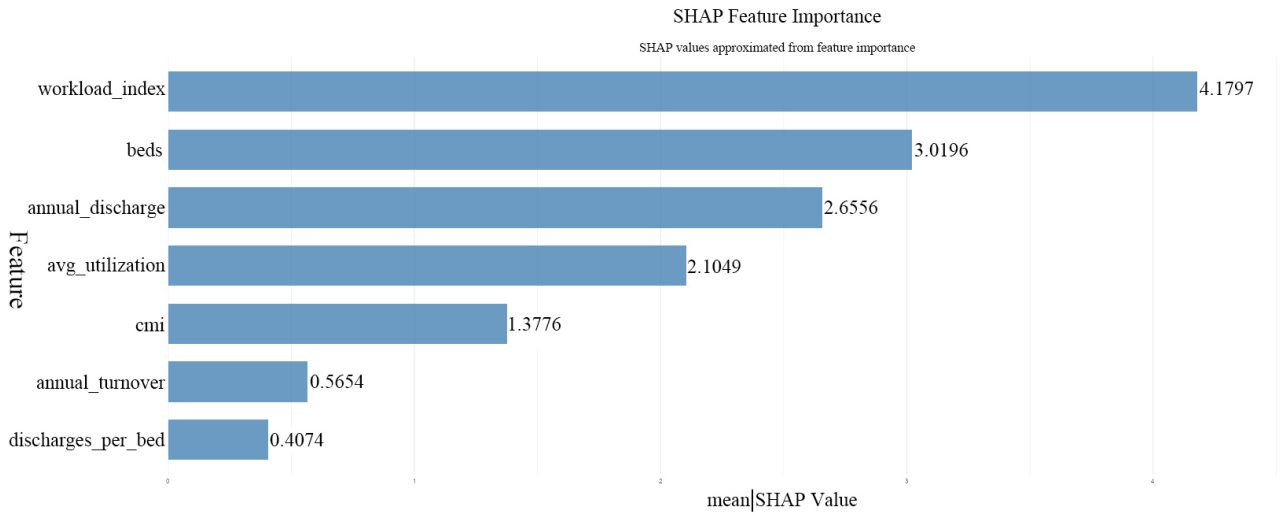


Figure 5

SHAP Summary Plot

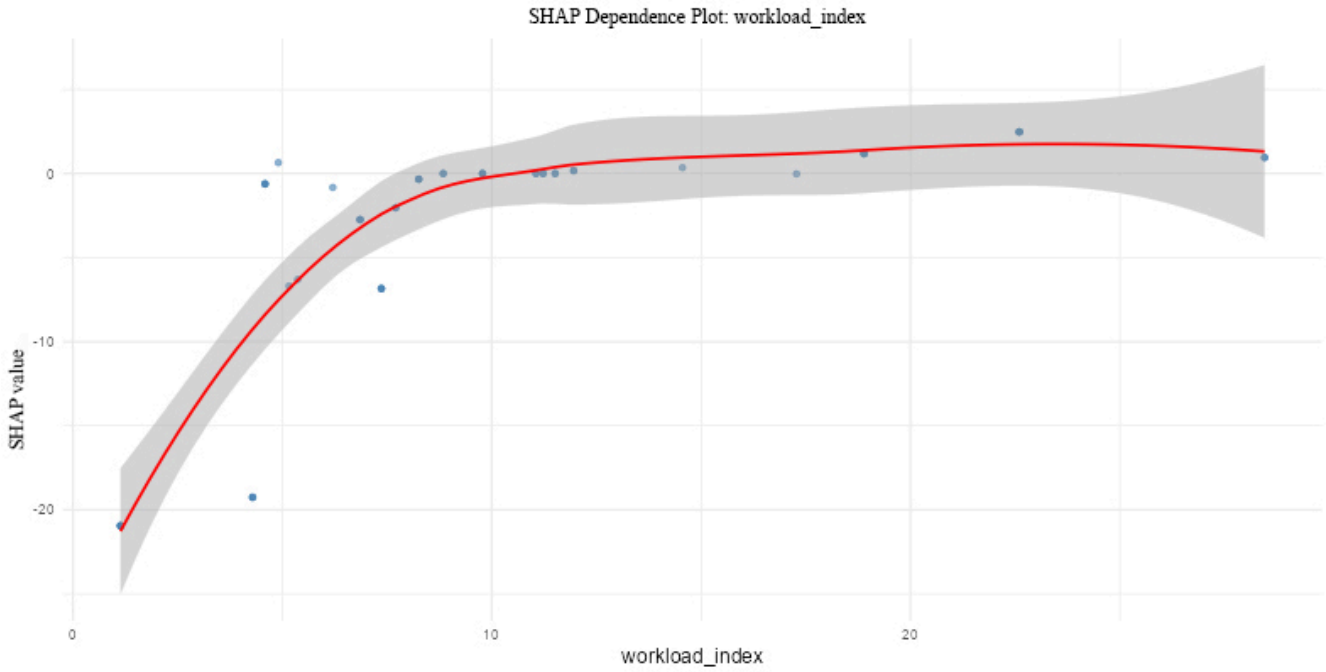


Figure 6

SHAP Dependence Plot

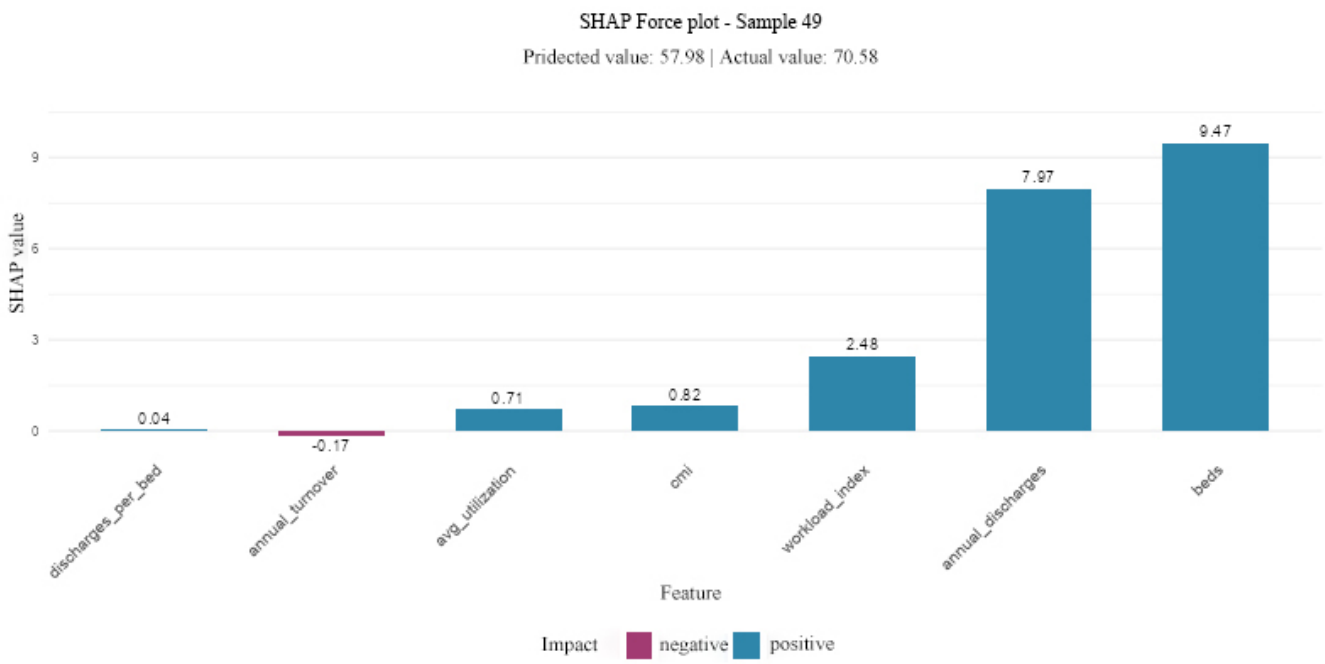


Figure 7

SHAP Force Plo

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Rcode.r