Journal of Advances in Information Technology, Vol. 17, No. 6, 2026

AI-Driven Forecasting of Operating Room Quality Indicators Using LSTM Neural Networks: A Data-Driven Framework for Smart Surgical Management

Pei-Ju Wang [1,2] and Wen-Shin Hsu [1,3,*]

[1] Department of Medical Information, Chung Shan Medical University, Taichung, Taiwan
[2] Operating Room, Changhua Christian Hospital, Changhua, Taiwan
[3] Informatics Office Technology, Chung Shan Medical University Hospital, Taichung, Taiwan
Email: peijuabby@hotmail.com (P.-J.W.); wshsu@csmu.edu.tw (W.-S.H.)
*Corresponding author

***Abstract*—Operating Room (OR) efficiency directly influences patient safety, resource allocation, and overall hospital performance. Despite the routine monitoring of operating room quality indicators, most analyses remain retrospective and provide limited support for forward-looking management. This study develops a forecasting framework based on Long Short-Term Memory (LSTM) neural networks and integrates the prediction results into a Business Intelligence (BI) visualization environment. Monthly specialty-level operational data from a medical center in Taiwan covering 2021 to 2024 were analyzed, including seven key operating room quality indicators. Different historical input windows ranging from 3 to 24 months were examined to evaluate their effects on predictive performance. To assess whether deep learning was necessary for this task, model performance was compared with traditional statistical approaches, including Moving Average, automated Autoregressive Integrated Moving Average (auto-ARIMA), and Naïve forecasting models, using a strictly chronological training and testing split. Indicators characterized by short-term operational variability, such as cancellation and scheduling delay rates, were more accurately predicted using shorter historical windows, whereas structurally stable metrics including utilization and occupancy rates showed improved performance with longer input sequences. While conventional statistical models performed competitively for some stable indicators, LSTM provided more consistent performance across heterogeneous metrics. The forecasting outputs were incorporated into an interactive dashboard to facilitate real-time monitoring and threshold-based management. Although the analysis was conducted using data from a single medical center, the results demonstrate the feasibility of integrating predictive analytics into routine surgical quality management.**

Keywords—business intelligence, artificial intelligence, Long Short-Term Memory (LSTM), Operating Room (OR) quality indicators, time-series forecasting, predictive analytics

I. INTRODUCTION

Operating Rooms (ORs) are among the most resource-intensive units in hospitals and have a direct impact on patient safety, clinical quality, and institutional efficiency. Performance monitoring in the OR environment commonly relies on key quality indicators, including cancellation rates, scheduling delays, first-case punctuality, emergency surgery participation, utilization, and occupancy rates. These indicators are closely linked to workflow coordination, staffing allocation, and resource management. Extensive research has examined operating room scheduling and surgical duration prediction, demonstrating that improved forecasting can enhance utilization efficiency, reduce costs, and improve patient outcomes [1–3]. Systematic reviews further emphasize the complexity of OR planning under uncertainty, where variability in case mix, staffing, and emergency demand complicates operational optimization [4, 5]. The financial implications of inefficient OR time usage have also been well documented [6].

Despite these advances, most existing studies focus on isolated prediction tasks, such as surgical duration estimation, or compare forecasting algorithms for single operational metrics. Relatively limited attention has been given to modeling multiple heterogeneous OR quality indicators within a unified analytical framework. In addition, predictive models are often evaluated in isolation, without being embedded into decision-support systems that can facilitate real-time operational management.

This study addresses these gaps by developing an integrated forecasting framework that combines Long Short-Term Memory (LSTM) neural networks with Business Intelligence visualization. Rather than proposing a new neural architecture, the objective is to establish a consistent and scalable modeling approach capable of

---

Manuscript received January 20, 2026; revised February 20, 2026; accepted March 19, 2026; published June 10, 2026.

doi: 10.12720/jait.17.6.1064-1073

handling diverse OR quality indicators while supporting practical deployment within a clinical management environment. By evaluating multiple historical input windows and benchmarking against conventional statistical forecasting methods, the study examines both methodological suitability and operational applicability [7–11].

LSTM networks were selected due to their established capability in modeling temporal dependencies in sequential data [12–16]. The framework is designed to capture both short-term operational fluctuations and longer-term structural trends across OR indicators. Predictive outputs are subsequently integrated into an interactive dashboard environment to support monitoring, threshold-based alerts, and data-informed decision-making in perioperative management.

II. METHODS

This study analyzed monthly Operating Room (OR) quality indicators collected from a tertiary medical center in Taiwan between January 2021 and December 2024. All indicators were calculated at the specialty level and aggregated on a monthly basis, resulting in multiple parallel time series for each metric.

Seven key indicators were examined: (1) surgery cancellation rate, (2) scheduled surgery delay rate, (3) emergency surgery attendance rate by attending surgeons, (4) on-time completion rate of the first scheduled case, (5) delay rate of patient entry into the preparation room, (6) OR utilization rate during regular hours, and (7) OR occupancy rate.

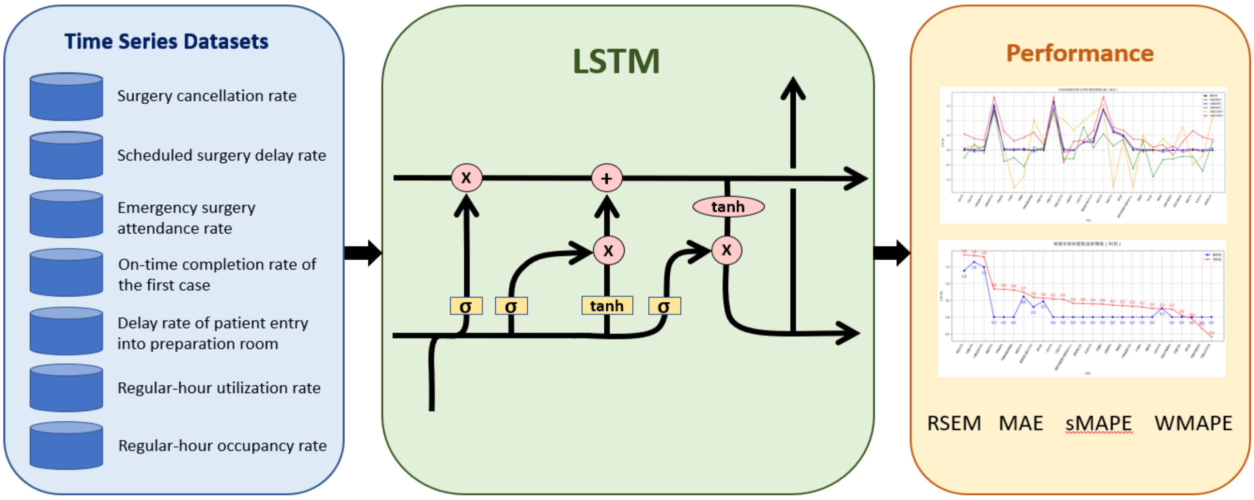

Fig. 1. Architecture of the LSTM model.

Time-series forecasting was conducted to predict one-month-ahead values for each indicator. The overall modeling workflow is illustrated in Fig. 1.

A. *Data Processing*

The dataset consisted of specialty-level monthly time series spanning 48 consecutive months (January 2021 to December 2024). Each operating room quality indicator was calculated at the specialty level and aggregated on a monthly basis, resulting in multiple parallel time series per indicator.

Specifically, the surgery cancellation rate and scheduled surgery delay rate each comprised 1,248 observations across 26 specialties over 48 months. The emergency surgery attendance rate included 960 observations from 20 specialties. The on-time completion rate of the first case and the delay rate of patient entry into the preparation room each contained 1,008 observations from 21 specialties. Regular-hour utilization rate and occupancy rate included 1,152 observations from 24 specialties.

To ensure temporal continuity, each specialty series was aligned to a complete 48-month index. Missing months were addressed using forward and backward filling to preserve local temporal consistency. Remaining missing numerical values were replaced with specialty-specific monthly means derived from historical observations.

Because all indicators were expressed as bounded rates between 0 and 1, MinMax scaling was applied to the training data, and the same transformation parameters were subsequently applied to the test data to prevent information leakage.

B. *Baseline Models*

To assess whether deep learning provided measurable advantages for monthly OR indicator forecasting, three commonly used statistical benchmarks were implemented: Moving Average (MA), automated Autoregressive Integrated Moving Average (auto-ARIMA), and a Naïve persistence model.

All baseline models were trained and evaluated using the same chronological data split as the LSTM models. One-step-ahead forecasting performance was primarily evaluated using Root Mean Squared Error (RMSE), with Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) reported for additional reference.

Traditional statistical forecasting approaches such as ARIMA are grounded in the Box–Jenkins framework [17], and automated variants remain standard baselines in time-series research [18]. In addition, scalable forecasting frameworks designed for real-world applications have also been investigated [19].

C. LSTM Model Architecture

A single-layer Long Short-Term Memory (LSTM) neural network was developed for indicator-level forecasting. Input tensors were structured as (batch size, time steps, features), where feature vectors included specialty-level case counts, indicator values, calendar month information, and one-hot encoded specialty identifiers.

The model architecture consisted of an LSTM layer with 64 hidden units followed by a fully connected output layer:

Input → LSTM (64, activation = 'tanh') → Dense(1)

Multiple historical input window lengths (3, 6, 9, 12, and 24 months) were evaluated to examine the influence of temporal depth on forecasting accuracy.

The hyperparameters of the LSTM model were selected based on commonly adopted configurations in time-series forecasting studies. A single-layer LSTM with 64 hidden units was chosen to balance model capacity and the risk of overfitting given the relatively limited dataset size (48 monthly observations per specialty). Preliminary exploratory experiments indicated that deeper architectures or larger hidden dimensions did not consistently improve predictive stability. Therefore, fixed hyperparameters were adopted across experiments to ensure comparability among different time-window settings and indicators.

The model was trained using Mean Squared Error as the loss function and optimized using the Adam optimizer. Random seeds were fixed (SEED = 42), and deterministic operations were enabled to enhance reproducibility.

Although other deep learning architectures such as Gated Recurrent Units (GRU) and Temporal Convolutional Networks (TCN) have been proposed for time-series forecasting, the present study focuses on LSTM due to its stronger capability in modeling long-term temporal dependencies through its explicit memory cell structure. Compared with GRU, LSTM provides more flexible control over information flow via separate input, forget, and output gates, which can be advantageous for capturing both short-term fluctuations and longer-term patterns in monthly operating room quality indicators.

Given the relatively limited dataset size and the objective of establishing a stable and interpretable forecasting framework for multiple operating room indicators, introducing additional deep learning architectures could increase model complexity without necessarily improving practical applicability. Therefore, LSTM was selected as a representative recurrent architecture for this study. Future research may further evaluate alternative architectures such as GRU or TCN when larger multi-center datasets become available.

D. Model Evaluation

The dataset consisted of specialty-level monthly time series spanning 48 consecutive months (January 2021 to December 2024). Each operating room quality indicator was calculated at the specialty level and aggregated on a monthly basis, resulting in multiple parallel time series per indicator.

Because several OR quality indicators in this study are expressed as rates bounded between 0 and 1 and exhibit relatively small month-to-month variation, absolute error values are naturally constrained. Percentage-based metrics, particularly Symmetric Mean Absolute Percentage Error (sMAPE), may appear inflated when actual values approach zero. Therefore, both absolute and relative error measures are reported to provide balanced interpretation of predictive performance.

To prevent data leakage, all models were evaluated using strictly chronological data splits. The training set included only observations preceding the test period, and no future information was incorporated during feature construction, scaling, or model fitting. One-step-ahead forecasting was performed sequentially without overlapping future targets.

A single chronological hold-out strategy was adopted, reserving the final 8 months as the test set while using all preceding observations for training. This design reflects real-world deployment conditions in which only historical information is available at prediction time. Hyperparameters were fixed prior to evaluation and were not tuned on the test data.

Although rolling-origin evaluation or time-series cross-validation could provide additional robustness, the relatively short time horizon of 48 monthly observations per specialty limits the number of feasible rolling windows without substantially reducing training data. Furthermore, given the limited test length (n = 8 months), formal statistical significance testing across models was not conducted, as small sample sizes may yield unstable inferential results.

E. Model Depth Comparison and Performance Analysis

To examine whether increasing model depth improves predictive performance, two LSTM architectures were compared using the surgery cancellation rate as a representative indicator. This metric was selected because it exhibited moderate variability and demonstrated sensitivity to different historical input windows in preliminary experiments. Both models were trained using identical datasets, feature sets, and hyperparameter configurations (SEED = 42), differing only in architectural depth.

1) Single-layer model

Architecture: Input → LSTM(64) → Dense(1)

Table I presents the forecasting performance of the single-layer LSTM model across different input window lengths.

TABLE I. PERFORMANCE OF THE SINGLE-LAYER LSTM MODEL

Step	RMSE	MAE	sMAPE	WMAPE
3	0.0037	0.0034	154.24%	142.19%
6	0.0004	0.0003	148.22%	14.70%
9	0.0041	0.0035	164.68%	146.18%
12	0.0075	0.0057	166.32%	242.21%
24	0.0007	0.0005	148.46%	22.97%

The 6-month and 24-month configurations achieved the lowest RMSE and MAE values, suggesting that both short- to mid-term and longer historical contexts contain informative temporal patterns for cancellation rate forecasting. Absolute error values remained small across configurations, consistent with the bounded nature of the indicator.

2) *Three-layer model*

Architecture: Input → LSTM(64, return_sequences=True) → LSTM(32, return_sequences=True) → LSTM(16) → Dense(1).

Table II summarizes the performance of the three-layer stacked LSTM model under the same experimental settings.

TABLE II. PERFORMANCE OF THE THREE-LAYER LSTM MODEL

Step	RMSE	MAE	sMAPE	WMAPE
3	0.0096	0.0078	200.00%	329.74%
6	0.0034	0.0021	162.18%	87.36%
9	0.0055	0.0044	179.57%	183.99%
12	0.0073	0.0060	180.99%	251.65%
24	0.0045	0.0035	168.20%	147.31%

Across most time windows, the deeper architecture did not improve predictive accuracy. In several configurations, RMSE and MAE values were higher than those of the single-layer model, particularly for 3-, 9-, and 12-month inputs.

3) *Model comparison*

A direct comparison between the single-layer and three-layer architectures indicates that increased depth did not produce consistent improvements across input window configurations. In particular, the three-layer model yielded higher RMSE values for the 3-, 9-, and 12-month windows, and only marginal differences were observed for the 6- and 24-month settings. Similar patterns were observed for MAE, suggesting that deeper stacking did not enhance predictive stability under the present data structure.

The absence of performance gains may be attributed to the relatively short temporal horizon (48 monthly observations per specialty) and the bounded nature of the target variable. Under such conditions, additional model parameters introduced by stacked LSTM layers increase model complexity without proportionate informational gain, potentially leading to less stable optimization.

Given these findings, the single-layer LSTM architecture was selected for subsequent analyses. This configuration demonstrated comparable or superior accuracy while maintaining structural simplicity, which is advantageous for reproducibility and deployment within a clinical decision-support environment.

III. RESULTS AND DISCUSSION

A. *Overview of Model Training*

Forecasting performance varied across indicators according to their temporal characteristics. Metrics exhibiting short-term operational variability, such as cancellation and scheduling delay rates, showed improved accuracy when modeled using shorter historical windows. In contrast, structurally stable indicators, including utilization and occupancy rates, benefited from longer input sequences.

Across all seven indicators, absolute error values remained small, reflecting the bounded nature of the rate-based metrics. However, relative percentage-based measures such as sMAPE displayed higher variability, particularly for indicators with low average values, such as cancellation rate. Therefore, both absolute and relative error metrics were considered when interpreting predictive performance.

The following subsections present detailed results for each indicator, highlighting differences in optimal time-window selection and forecasting behavior.

B. *Benchmark Comparison with Baseline Models*

To evaluate whether deep learning provides practical advantages for monthly OR quality forecasting, LSTM was compared with three standard statistical benchmarks: Moving Average (MA), auto-ARIMA, and a Naïve persistence model. All models were evaluated using identical chronological train-test splits with one-step-ahead forecasting over an 8-month test period. RMSE was used as the primary comparison metric.

Results indicate that statistical baselines remained highly competitive for structurally stable indicators. MA6 achieved the lowest RMSE for occupancy rate, delay rate, and attending surgeon participation rate, while auto-ARIMA yielded the lowest RMSE for overrun rate. These findings suggest that for relatively smooth and stable time series, conventional univariate methods are sufficient.

In contrast, LSTM achieved the lowest RMSE for utilization rate (0.0416), cancellation rate (0.0009), and first-case on-time rate (0.0038). Although the magnitude of improvement over statistical baselines was modest, LSTM demonstrated consistent performance across indicators characterized by greater short-term variability.

Importantly, the objective of this study is not solely to identify the best-performing model for each individual metric, but to establish a unified forecasting framework capable of handling heterogeneous OR indicators within a single modeling architecture. In this context, LSTM offers a scalable and flexible structure that performs competitively across diverse operational metrics.

It should also be noted that all indicators are expressed as bounded rates between 0 and 1. Consequently, absolute error values are numerically small, and percentage-based measures such as MAPE may appear elevated for indicators with low average values (e.g., cancellation rate). Therefore, both absolute and relative error metrics should be interpreted in relation to the underlying scale and variability of each indicator.

Table III summarizes the detailed benchmark results across all seven OR quality indicators. Best Model is Selected by RMSE (Primary).

TABLE III. ONE-STEP-AHEAD BENCHMARK COMPARISON ACROSS SEVEN OR QUALITY INDICATORS (TEST LENGTH = 8 MONTHS)

Indicator	**Best model (by RMSE)**	**MAE**	**RMSE**	**MAPE (%)**	**n_test**
Occupancy rate	MA6_1step	0.0645	0.0723	6.442	8
Utilization rate	**LSTM_9_ 1step**	0.0332	**0.0416**	4.2067	8
Cancellation rate	**LSTM_9_ 1step**	0.0008	**0.0009**	60.986	8
Delay rate	MA6_1step	0.0015	0.0018	21.66	8
First-case on-time rate	**LSTM_9_ 1step**	0.0036	**0.0038**	0.3649	8
Attending surgeon participation rate	MA6_1step	0.0199	0.0234	3.7799	8
Overrun rate	ARIMA_ auto_1step	0.0052	0.0067	7.5973	8

C. Prediction Results by Quality Indicator

1) Surgery cancellation rate

For the cancellation rate, the 6-month input window yielded the lowest absolute error (RMSE = 0.0004, MAE = 0.0003, WMAPE = 14.70%), as shown in Fig. 2. Although sMAPE values appear elevated (148.22%), this is attributable to the low average magnitude of the cancellation rate, where small absolute deviations translate into large percentage-based fluctuations.

The results suggest that moderate-length historical windows capture short-term temporal patterns more effectively than longer sequences. While the 24-month configuration also produced low absolute errors, extending the input window did not provide additional gains in stability.

2) Delay rate of scheduled surgeries

Similarly, the 6-month configuration achieved the lowest RMSE (0.0038) and MAE (0.0022), with a WMAPE of 2.54% (Fig. 3). Performance declined slightly when the historical window was extended to 24 months, indicating that shorter to mid-range temporal contexts are sufficient to capture recent operational variation in delay rates.

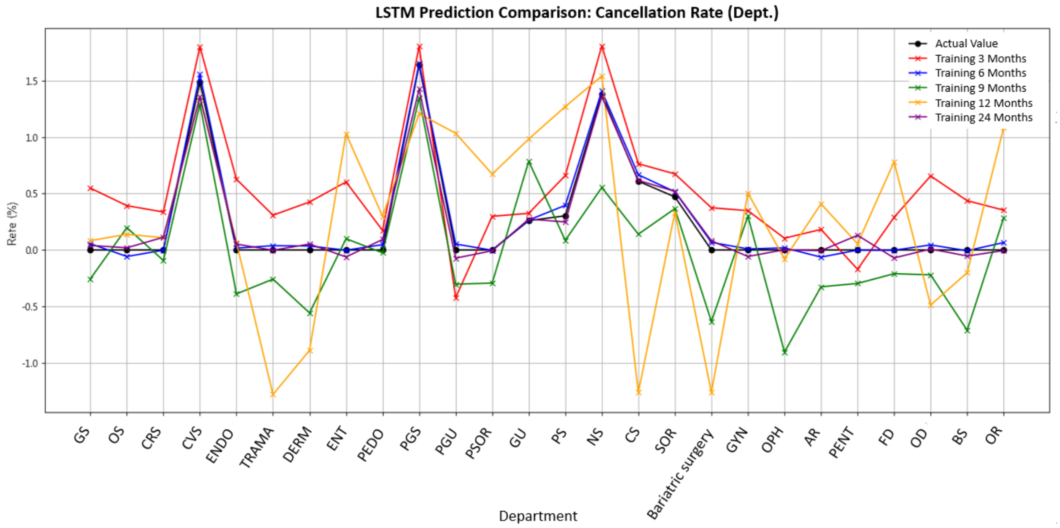

Fig. 2. Comparison of LSTM predictive performance across departments using different training time windows for cancellation rate forecasting.

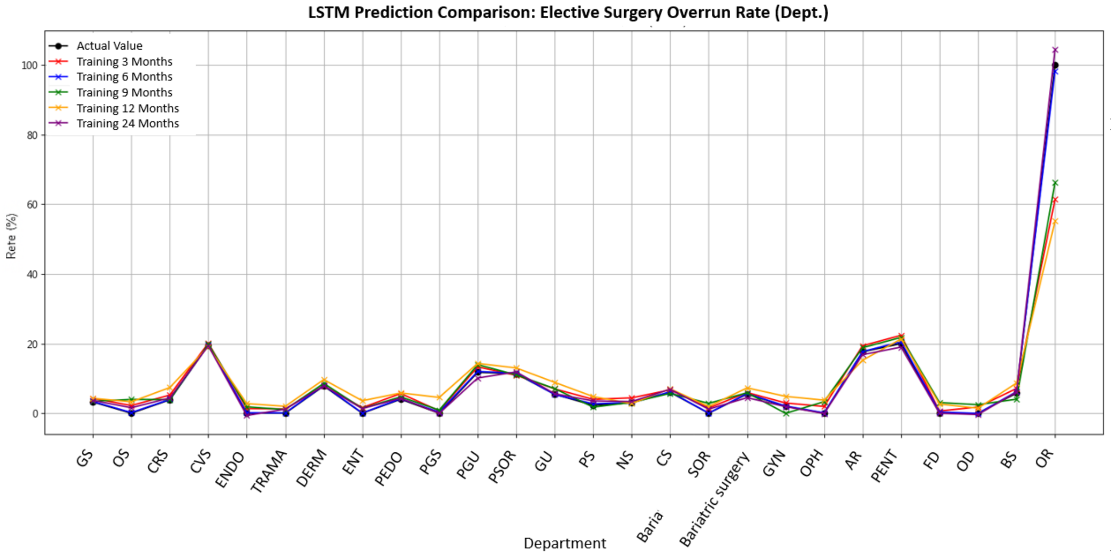

Fig. 3. LSTM prediction comparison for elective surgery overrun rates across departments using different training window lengths.

3) *Emergency surgeries performed by attending surgeons*

For the emergency surgery attendance rate, the 24-month input window achieved the lowest error (RMSE = 0.0013, MAE = 0.0013, WMAPE = 0.15%), as shown in Fig. 4. The small variation across window lengths suggests a relatively stable temporal pattern, where longer historical context improves smoothing and trend consistency.

4) *On-time completion rate of the first case*

The 3-month configuration yielded the lowest RMSE (0.0006) and MAE (0.0004), with minimal relative error (WMAPE = 0.04%), as shown in Fig. 5. These findings indicate that short-term historical information is sufficient to model fluctuations in first-case punctuality, whereas longer sequences did not improve predictive accuracy.

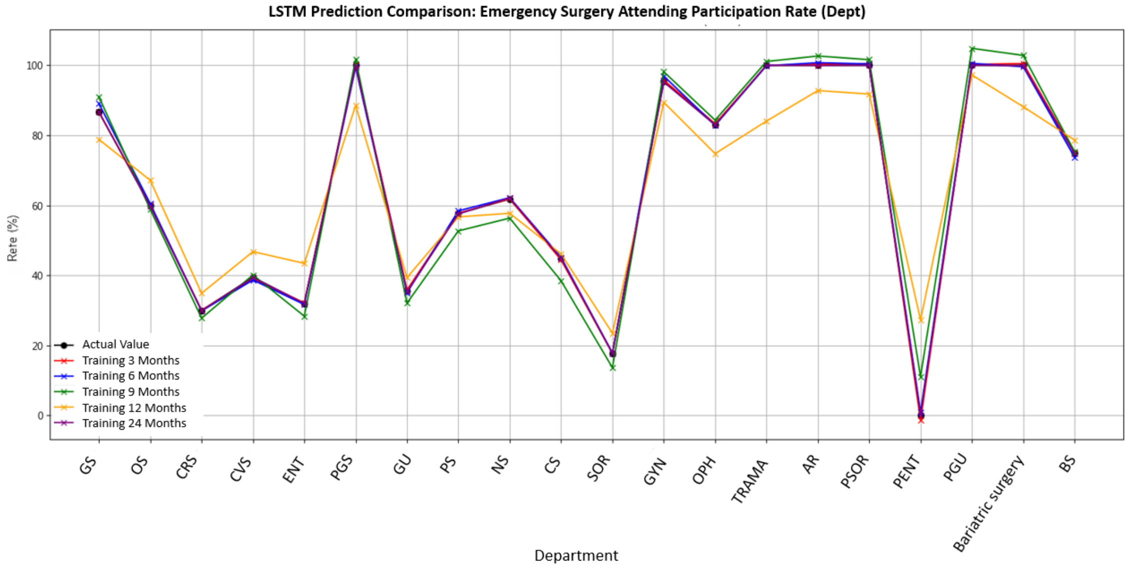

Fig. 4. LSTM prediction comparison for emergency surgery attending participation rates across departments.

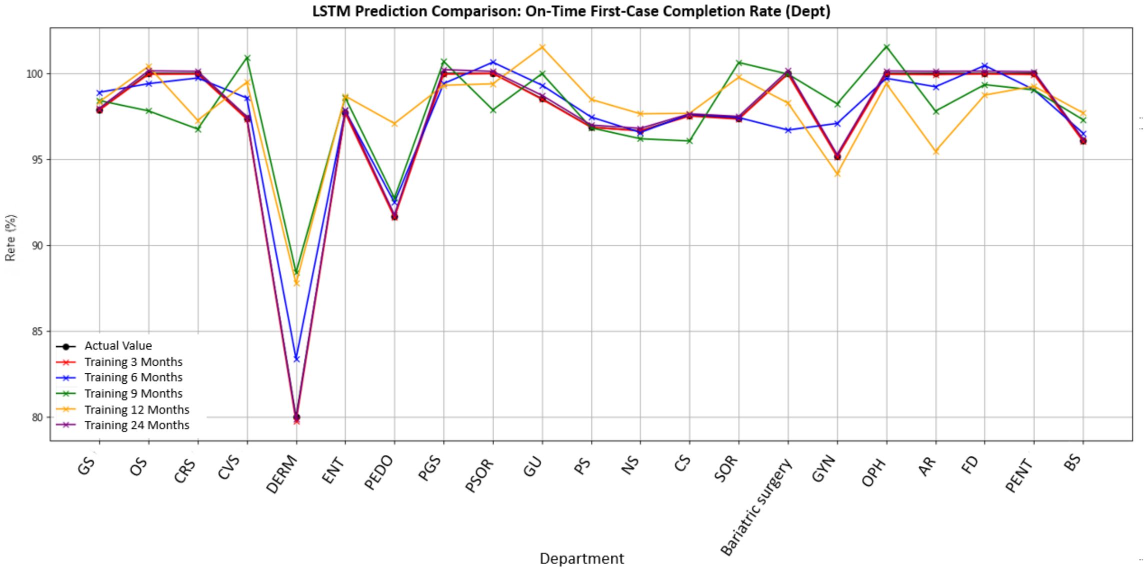

Fig. 5. LSTM prediction comparison for on-time first-case completion rates across departments.

5) *Delay rate of patient entry into preparation room*

The 24-month window produced the lowest error (RMSE = 0.0006, MAE = 0.0005, WMAPE = 13.14%), with limited variation across configurations (Fig. 6). This pattern suggests relatively steady temporal dynamics over longer horizons.

6) *OR utilization rate during regular hours*

For utilization rate, the 24-month input sequence achieved the lowest RMSE (0.0015) and MAE (0.0013), with WMAPE below 1% (Fig. 7). The consistently low error across configurations reflects strong structural regularity and cyclic behavior in utilization patterns.

7) *OR occupancy rate*

Similarly, the 24-month configuration demonstrated the lowest RMSE (0.0037) and MAE (0.0031), with WMAPE of 0.41% (Fig. 8). Shorter windows produced slightly higher deviations, suggesting that occupancy trends are better captured through longer temporal context.

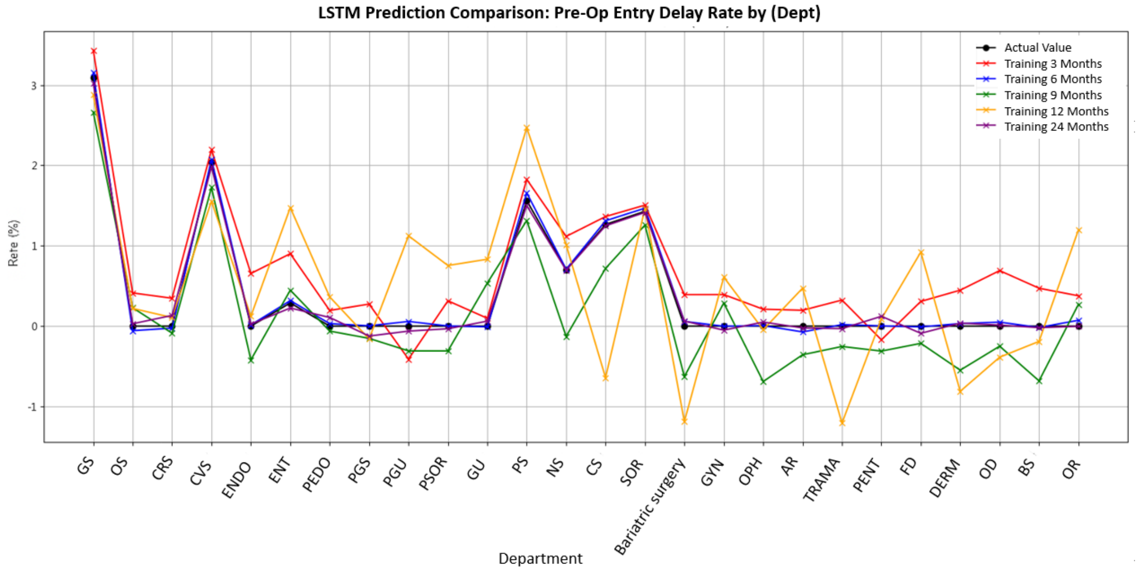

Fig. 6. LSTM prediction comparison for pre-operative entry delay rates across departments.

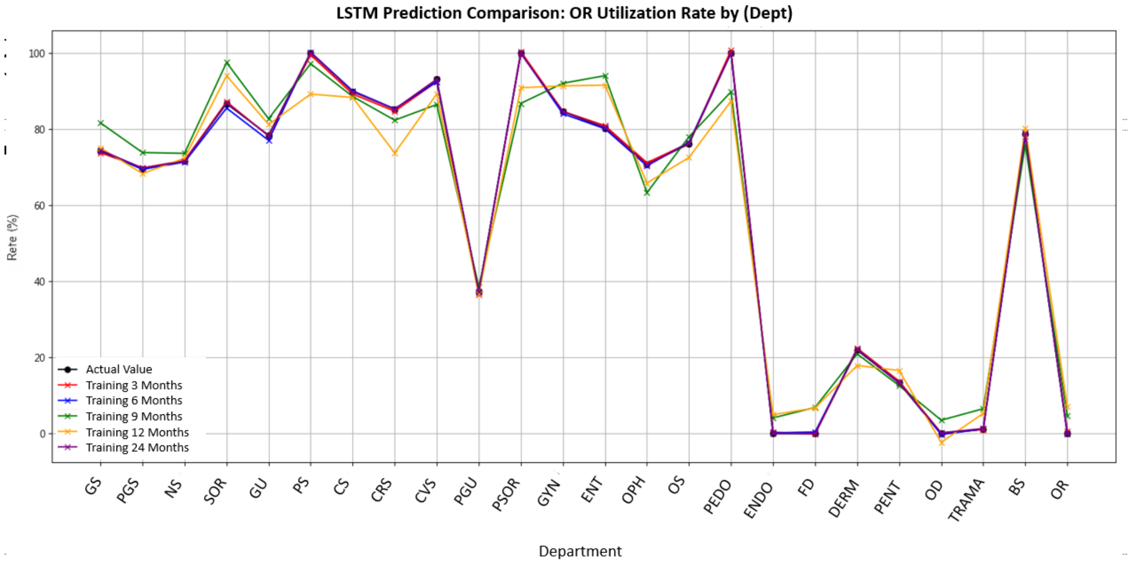

Fig. 7. LSTM prediction comparison for operating room utilization rates across departments.

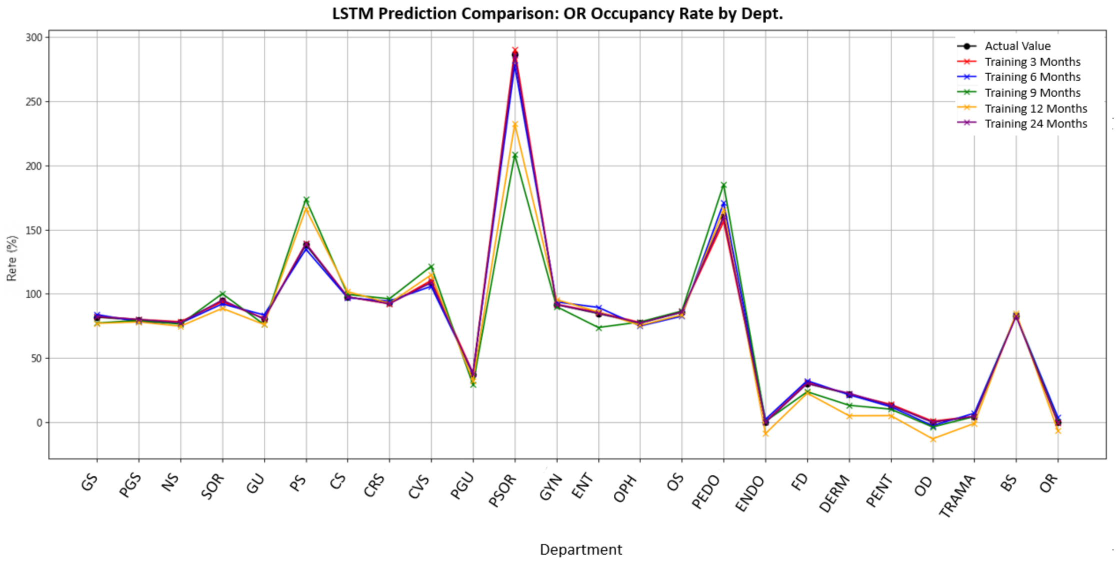

Fig. 8. LSTM prediction comparison for operating room occupancy rates across departments.

D. Comparative Summary and Interpretation

Table IV summarizes the optimal input window and corresponding error metrics for each quality indicator. A consistent pattern emerges in which indicators characterized by higher short-term variability, such as cancellation rate, scheduling delay rate, and first-case punctuality, achieved lower errors when modeled using shorter historical windows (3–6 months). In contrast, structurally stable indicators, including emergency surgeon participation, OR utilization, occupancy rate, and preparation room delay, demonstrated improved performance with longer input sequences (24 months).

TABLE IV. COMPARATIVE SUMMARY AND INTERPRETATION

Quality Indicator	Optimal Time Window	RMSE	MAE	WMAPE	Predictive Characteristics
Surgery cancellation rate	6 months	0.0004	0.0003	14.70%	Influenced by short-term scheduling and patient factors
Delay rate of scheduled surgeries	6 months	0.0038	0.0022	2.54%	Sensitive to short-term fluctuations
Emergency surgeries by attending surgeons	24 months	0.0013	0.0013	0.15%	Highly stable with clear long-term trends
On-time completion of first case	3 months	0.0006	0.0004	0.04%	Affected by short-term staffing and turnover efficiency
Delay of patient entry into preparation room	24 months	0.0006	0.0005	13.14%	Displays periodic variation
OR utilization rate	24 months	0.0015	0.0013	0.22%	Strong regularity and seasonality
OR occupancy rate	24 months	0.0037	0.0031	0.41%	Long-term stability in space utilization

These findings suggest that different OR quality indicators exhibit distinct temporal dynamics. Shorter windows appear sufficient for capturing recent fluctuations, whereas longer windows provide greater stability for indicators with sustained trends or seasonal structure.

While statistical baselines performed competitively for several stable indicators, the LSTM-based framework demonstrated consistent and reliable performance across heterogeneous metrics under a unified modeling structure. Rather than optimizing separate models for each indicator, the present approach emphasizes scalability and structural consistency within an operational deployment context.

In practical OR management environments, maintaining a single modeling architecture simplifies implementation, monitoring, and long-term system governance. From this perspective, the selection of a unified LSTM framework reflects a balance between predictive performance and operational feasibility.

E. Integration with Business Intelligence (BI) System

Within the proposed framework, the single-layer LSTM model was adopted as the primary forecasting component due to its stable performance across heterogeneous indicators and its suitability for deployment within a unified analytical architecture. Statistical baseline models were retained for benchmarking purposes, whereas the LSTM configuration was selected for operational integration.

To facilitate practical application, forecasting outputs were incorporated into a Business Intelligence (BI) visualization environment through automated data pipelines linking historical OR records and predicted values. The resulting dashboards present both observed trends and next-period forecasts for each quality indicator.

As illustrated in Fig. 9, department-level scheduling delay rates are displayed alongside corresponding one-step-ahead predictions, with predefined benchmark thresholds (e.g., 10%) indicated for reference. This visualization framework enables administrators to compare historical patterns with forecasted values and to identify departments that may require further review. Additional dashboard views (Figs. 10 and 11) provide trend summaries and cross-department comparisons.

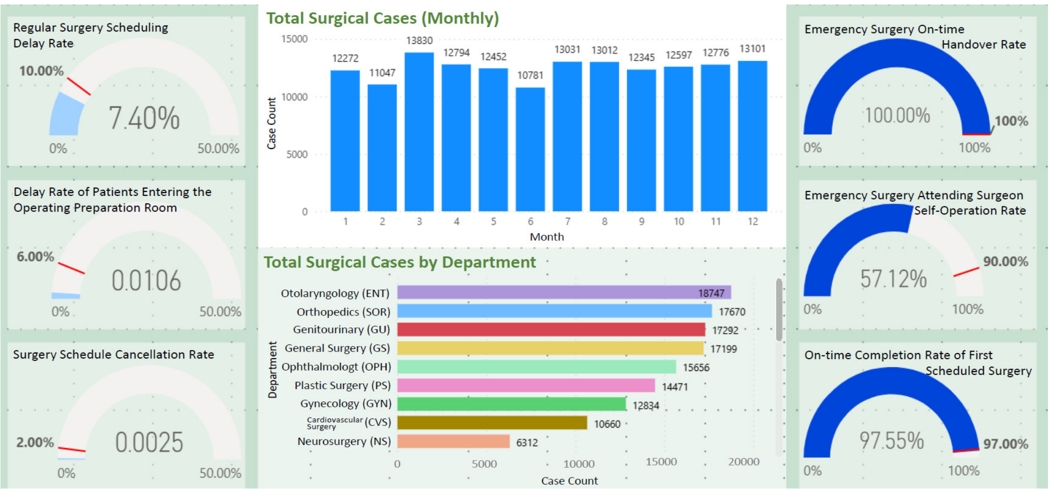

Fig. 9. Dashboard of key surgical performance indicators.

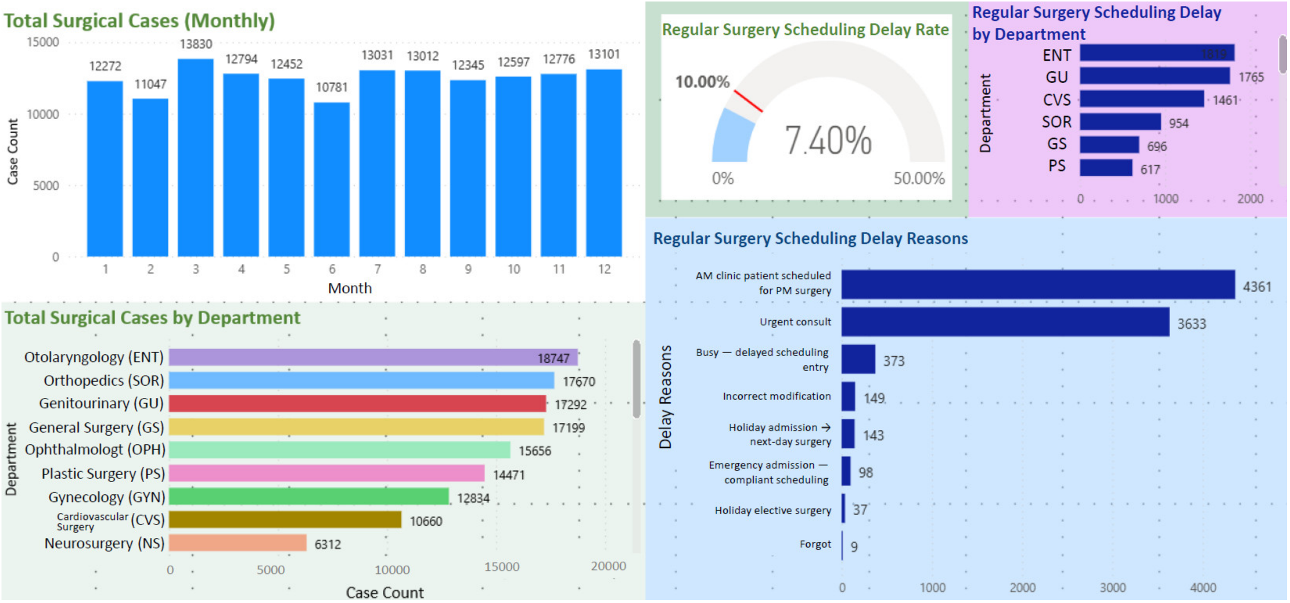

Fig. 10. Dashboard of regular surgery scheduling delays and contributing factors.

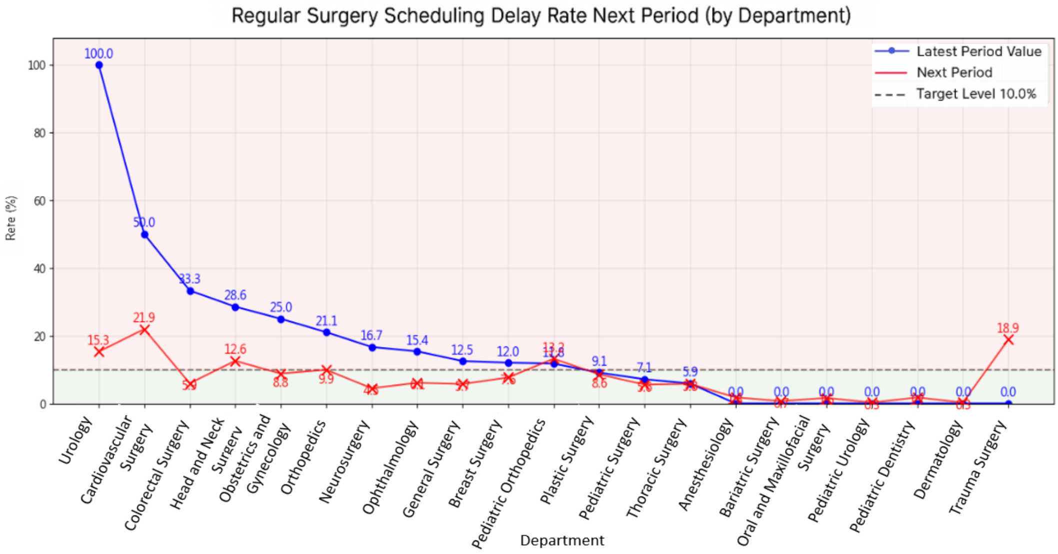

Fig. 11. Next-period regular surgery scheduling delay rates by department, showing historical values, LSTM predictions, and benchmark thresholds.

The integration of forecasting outputs into a BI platform demonstrates the feasibility of embedding predictive analytics within routine OR management workflows. While the current implementation serves as a proof-of-concept rather than a formally evaluated intervention, it illustrates how time-series forecasting can be operationalized to support structured monitoring and decision awareness in perioperative management.

IV. CONCLUSION AND RECOMMENDATIONS

This study examined the feasibility of applying time-series forecasting techniques to Operating Room (OR) quality management using monthly specialty-level data from 2021 to 2024. By evaluating multiple historical input windows and benchmarking against conventional statistical models, the analysis explored how different OR indicators respond to varying temporal contexts. The goal of this study was to evaluate methodological suitability under real-world data constraints.

The findings indicate that OR quality indicators exhibit distinct temporal characteristics. Metrics characterized by short-term variability, such as cancellation and scheduling delay rates, were better modeled using shorter historical windows, whereas structurally stable indicators, including utilization and occupancy rates, benefited from longer input sequences. These results highlight the importance of aligning model configuration with the temporal properties of each operational metric.

Rather than focusing solely on identifying a universally superior forecasting model, this study emphasizes the development of a unified and scalable analytical framework capable of handling heterogeneous OR indicators within a consistent architecture. Under the present data conditions, the selected LSTM configuration demonstrated stable and competitive performance across diverse metrics while supporting practical system integration.

The incorporation of forecasting outputs into a Business Intelligence (BI) environment further illustrates the feasibility of embedding predictive analytics within routine OR monitoring processes. Although the current implementation represents a proof-of-concept, it demonstrates how time-series forecasting can be operationalized to enhance structured monitoring and forward-looking decision awareness in perioperative management.

Several limitations should be acknowledged. First, the analysis was conducted using data from a single medical center, which may limit generalizability; validation across multiple institutions would strengthen external validity. Second, although comparisons were made with standard statistical baselines, additional evaluation against other deep learning architectures could provide further methodological insight. Third, missing data were addressed using deterministic imputation strategies, and future studies should examine alternative approaches through formal sensitivity analyses. Finally, the BI integration was not assessed through formal usability or impact studies, and its effect on operational outcomes remains to be systematically evaluated.

Future research may extend this framework by incorporating additional contextual variables, exploring multi-horizon forecasting strategies, and assessing system robustness under abrupt operational changes. Through continued refinement and validation, predictive analytics may contribute to more informed and adaptive OR quality management.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was reviewed and approved by the Institutional Review Board of a medical center in Changhua, Taiwan (IRB No. 250219). All data used in this study were de-identified prior to analysis, and the requirement for informed consent was waived by the IRB due to the use of secondary, anonymized operational data.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

P.-J. Wang conducted data collection, preprocessing, model development, and experimental analysis, and was responsible for drafting the manuscript. W.-S. Hsu conceptualized the study, supervised the research process, and critically revised the manuscript for important intellectual content. All authors have read and approved the final version of the manuscript.

ACKNOWLEDGMENT

The author expresses sincere gratitude to Medical Director Ming Chen for his guidance and continuous support throughout the course of this study. The author also extends heartfelt appreciation to Supervisor Chia-Lan Chen for her valuable advice and encouragement, which greatly contributed to the completion of this work.

REFERENCES

[1] A. A. Gumbs, I. Frigerio, G. Spolverato, R. Croner, A. Illanes, E. Chouillard, and E. Elyan, "Artificial intelligence surgery: How do we get to autonomous actions in surgery?" *Sensors*, vol. 21, no. 16, 5526, 2021.

[2] C. Spence, O. A. Shah, A. Cebula, K. Tucker, D. Sochart, D. Kader, and V. Asopa, "Machine learning models to predict surgical case duration compared to current industry standards: scoping review," *BJS Open*, vol. 7, no. 6, 2023.

[3] M. Al-Amin, R. Baldacci, and V. Kayvanfar, "A comprehensive review on operating room scheduling and optimization," *Operational Research*, vol. 25, no. 1, pp. 3–25, 2025.

[4] B. Cardoen, E. Demeulemeester, and J. Beliën, "Operating room planning and scheduling: A literature review," *European Journal of Operational Research*, vol. 201, no. 3, pp. 921–932, 2010.

[5] F. Dexter and R. D. Traub, "How to schedule elective surgical cases into specific operating rooms," *Anesthesiology*, vol. 96, no. 4, pp. 933–942, 2002.

[6] A. Macario, "What does one minute of operating room time cost?" *Journal of Clinical Anesthesia*, vol. 22, no. 4, pp. 233–236, 2010.

[7] H. J. Jang and S. M. Lee, "Influence of surgery preparation time on patient outcomes," *Journal of Nursing Management*, vol. 2024, no. 1, 6753210, 2024.

[8] A. Gonzalez-Garcia, S. Pérez-González, C. Benavides, A. Pinto-Carral, E. Quiroga-Sánchez, and P. Marqués-Sánchez, "Impact of artificial intelligence-based technology on nurse management: A systematic review," *Journal of Nursing Management*, vol. 2024, no. 1, 3537964, 2024.

[9] X. Li, Q. Zong, and M. Cheng, "The impact of medical explainable artificial intelligence on nurses' innovation behaviour: A structural equation modelling approach," *Journal of Nursing Management*, vol. 2024, no. 1, 8885760, 2024.

[10] O. Yakusheva, M. J. Bouvier, and C. O. Hagopian, "How artificial intelligence is altering the nursing workforce," *Nursing Outlook*, vol. 73, no. 1, 102300, 2025.

[11] M. K. K. Rony, D. M. Alrazeeni, F. Akter, L. Nesa, D. C. Das, M. J. Uddin *et al*., "The role of artificial intelligence in enhancing nurses' work-life balance," *Journal of Medicine, Surgery, and Public Health*, vol. 3, 100135, 2024.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in *Proc. 15th Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, Singapore, 2014, pp. 338–342.

[14] Y. Lu and F. M. Salem, "Simplified gating in Long Short-Term Memory (LSTM) recurrent neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Boston, MA, USA, Aug. 2017, pp. 1601–1604.

[15] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.

[16] A. Yunita, M. I. Pratama, M. Z. Almuzakki, H. Ramadhan, E. A. P. Akhir, A. B. F. Mansur, and A. H. Basori, "Performance analysis of neural network architectures for time series forecasting: A comparative study of RNN, LSTM, GRU, and hybrid models," *MethodsX*, vol. 15, 103462, 2025.

[17] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.

[18] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for R," *Journal of Statistical Software*, vol. 27, no. 3, pp. 1–22, 2008.

[19] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.