BMC Surgery

RESEARCH Open Access

Assessing artificial intelligence ability in predicting hospitalization duration for pleural empyema patients managed with uniportal video-assisted thoracoscopic surgery: a retrospective observational study

Issa Alnajjar[1], Baraa Alshakarnah[2], Tasneem AbuShaikha[1], Tareq Jarrar[1*], Abed Al-Raheem Ozrail[1] and Yousef Abu Asbeh[2*]

Abstract

Background This retrospective observational research evaluates the potential applicability of artificial intelligence models to predict the length of hospital stay for patients with pleural empyema who underwent uniportal video-assisted thoracoscopic surgery.

Methods Data from 56 patients were analyzed using two artificial intelligence models. A Random Forest Regressor, the initial model, was trained using clinical data unique to each patient. Weighted factors from evidence-based research were incorporated into the second model, which was created using a prediction approach informed by the literature.

Results The two models tested showed poor prediction accuracy. The first one had a mean absolute error of 4.56 days and a negative R^2 value. The literature-informed model performed similarly, with a mean absolute error of 4.53 days and an R^2 below zero.

Conclusions While artificial intelligence holds promise in supporting clinical decision-making, this study demonstrates the challenges of predicting length of stay in pleural empyema patients due to significant clinical variability and the current limitations of AI-based models. Future research should focus on integrating larger, multi-center datasets and more advanced machine learning approaches to enhance predictive accuracy.

Keywords Artificial Intelligence, Machine Learning, Length of Stay, Pleural Empyema, Uniportal Video-Assisted Thoracoscopic Surgery, Predictive Modeling, Random Forest Regressor, Retrospective Study

Background

Lower respiratory tract infections, including pneumonia, pose significant concerns as they rank among the leading causes of mortality and can lead to serious postoperative complications [1]. Parapneumonic effusion, caused by bacteria, viruses, fungi, mycoplasmas, or tuberculosis, affects 2–12% of children with pneumonia, increasing to 28% for hospitalized children [2, 3]. Such effusions, which

*Correspondence:
Tareq Jarrar
tareq.jarrar2@students.alquds.edu
Yousef Abu Asbeh
Dryousefabuasbeh@gmail.com
[1] Faculty of Medicine, Al-Quds University, Jerusalem, Palestine
[2] Faculty of Health Science, Al-Quds University, Jerusalem, Palestine

are usually sterile at the onset, can progress to empyema as a result of bacterial invasion and fibrin deposition [4]. Pleural empyema is a rare disease, affecting 5–10% of parapneumonic effusion cases. It progresses through simple, complicated, and empyema, with advanced treatments aimed at mitigating conditions like trap lung. Recent treatments include ultrasound-assisted drainage, antibiotics, and fibrinolytic agents [5].

Video-assisted thoracoscopic surgery (VATS) effectively treats various thoracic pathologies with minimal invasiveness and operative trauma, but contraindications include severe adhesions, large tumors, or single lung ventilation inability [6]. In patients with primary empyema, VATS has been shown to be more effective than open decortication, with less postoperative pain, fewer complications, and shorter hospital stays, all while achieving comparable resolution rates [7]. The guidelines put forth by the European Association for Cardio-Thoracic Surgery suggest the use of VATS in the management of stage II-III empyema, except for those stage III who presented with symptoms for more than 5 weeks [7, 8]. A study comparing VATS and thoracotomy for chronic stage III empyema found comparable mortality, recurrence, and complication rates, with VATS offering shorter postoperative stay [9]. Recent meta-analyses support VATS as the preferred surgical method due to its lower morbidity, fewer septic complications, and shorter operative time compared to open surgery [10].

The length of hospital stay (LOS) is crucial for assessing hospital performance and patient recovery. Prolonged LOS can lead to negative outcomes like increased healthcare costs, increased risk of infections, and bed shortages. Predicting LOS after pulmonary surgery is challenging due to patient responses, complications, surgical techniques, and postoperative recovery rates. Surgeon-specific practices also complicate LOS prediction. Research aims to improve perioperative care and optimize hospital capacity by examining LOS determinants [11]. There are various patient-related factors that influence LOS. Old age is a significant factor linked to increased LOS due to compromised physiological reserves and increased risk of complications [11–16]. Gender is another variable that has been found to affect the LOS, but the studies on the subject are inconclusive [11, 16]. The nutritional state is crucial as preoperative malnourishment, like low albumin levels, can hinder recovery [11]. Pulmonary function tests like forced expiratory volume 1 and diffusion capacity predict LOS, with lower values indicating longer stays [12, 14–16]. Smoking status does not significantly impact LOS, but active smoking hinders healing and recovery, thereby prolonging LOS [13]. Multiple comorbidities lead to a more complicated postoperative course that impacts LOS significantly [12, 17, 18]. Moreover, the choice of surgical technique significantly impacts LOS, with minimally invasive approaches like VATS reducing LOS due to less surgical injury and quicker recovery [16, 17]. Nevertheless, the development of postoperative complications, especially prolonged air leaks and infections, is still the most important reason for a longer LOS [12, 17, 18]. The complexity and duration of surgical procedures significantly influence patient stays in the hospital [14, 16]. Poor pain relief may not only impair healing rates but also increase the risk of complications and finally result in an increased LOS [12, 13]. In addition, several postoperative parameters have been shown to impact LOS. For example, the duration of chest tube insertion, in particular, is important with regard to the LOS [13, 17]. Blood loss in operative management also influences recovery; drastic loss of blood results in increased recovery times [13, 16]. Postoperative practices like removing the endotracheal tube, initiating physiotherapy, and enduring short periods of starvation reduced the duration of in-hospital treatment [13, 17]. On the other hand, an extended period of monitoring in the Post-Anesthesia Care Unit (PACU) postoperatively refers to a more complicated recovery, which often leads to a longer LOS in the hospital [13].

Artificial intelligence (AI) is revolutionizing healthcare by improving prediction, diagnosis, and management through advanced machine learning and big data analysis. AI models, like random forests and gradient-boosting decision trees, accurately predict diabetes occurrence and identify individuals for early intervention [19]. AI systems and motion sensors enable cost-effective, simple, and non-invasive dementia detection in large dormitory-based mobile screening, crucial for handling sensitive information [20]. For hip fractures, AI-assisted diagnostic tools are sensitive and specific, but their postoperative outcome prediction parameters are similar to conventional methods, aiding in risk management and health resource allocation [21]. According to ventilator-associated pneumonia, AI models increase predictive power due to better integration of clinical, demographic, and laboratory information and thus prompt patients to take measures earlier than expected [22]. As a result, Artificial intelligence enhances prediction quality, operational efficiency, reduces diagnosis time, and allows for personalized patient care in medical, surgical, and radiological practices.

This study aims to explore the potential of AI models in predicting post-operative LOS for pleural empyema, using patient data and literature-based instructions. The aim is to improve perioperative patient care, healthcare resource allocation, and ultimately enhance thoracic surgery patient outcomes.

Methodology

Study design and population

The retrospective observational design was employed to evaluate the ability of AI and Machine Learning (ML) models to predict hospital stay duration in patients diagnosed with thoracic empyema of benign etiologies who underwent uniportal VATS through two experiments on different mathematical models. This study was conducted in Al-Ahli Hospital in Hebron-Palestine on a total of 56 participants who were selected through convenience sampling based on the following inclusion criteria: Patients who underwent VATS for empyema management, operated on by the same surgeon, and their complete preoperative clinical, laboratory, and imaging data are available in the hospital archive. On the other hand, patients who were discharged against medical advice, transferred to other centers to complete management, had malignant empyema, died in the hospital, and patients who stayed postoperatively longer than 30 days or less than 3 days were all excluded.

Data source and variables

This study utilized data collected from the hospital's special software system. The variables included age, sex, smoking status, presence of fever, recurrence of empyema, past medical history, past surgical history, result of pleural culture, computerized tomography stage according to the American Thoracic Society stages of empyema which includes the exudative, fibrinopurulent, and organizing phases [23], white blood cells, hemoglobin, platelets, C-reactive protein, serum sodium, serum chloride, serum creatinine, blood urea nitrogen, and post-operative LOS. The following independent variables were presented as numeric values according to a grading system as follows: smoking status (unknown, non-smoker, smoker), presence of fever (fever presented upon admission, the patient had no fever upon admission), recurrence of empyema (the patient had at least one previous episode of empyema, the patient didn't have previous episodes of empyema), past medical history (free, the patient had a localized/limited disease, the patient has at least one non-respiratory systemic disease, the patient has a positive past medical history that includes at least a respiratory disease), past surgical history (free, the patient underwent a non-chest surgery, the patient underwent a previous chest surgery), the result of pleural culture (null, negative, positive), CT stage (Stage 1: no loculations appeared on CT but may show minimal pleural thickening on imaging, Stage 2: Septations visible on CT with possible enhancement of pleural layers in Contrast CT and may show pleural thickness or consolidation in adjacent lung tissue, Stage 3: thickened pleura with thick wall cavities and calcifications or fibrosis in chronic cases.

Study sample

Fifty-six patients' data were collected. The following variables were available for all patients: age, gender, past medical history, past surgical history, fever presence, recurrence of empyema, laboratory tests, and CT grades. Patients were stratified by CT grades as follows: 13 patients in Stage 1, 23 patients in Stage 2, and 20 patients in Stage 3. However, the following variables were missed for some patients: smoking status (unknown for 13 patients) and pleural culture result (unknown for 18 patients). In addition, one patient was deleted from the data collection in experiment #2 after considering it an outlier.

Written Commands for Experiment #1

Cleaning, arranging, and grading -for some variables- were performed on the collected data, the patients' data was then collected in an Excel file. We created our own GPT model named "Predictor" using the indirect programming model by providing the following instructions: Two groups of patients will be provided to the model. The first group, the "training group," will consist of 28 patients with all data provided, including the date of VATS performing and the discharge date. This group aims to enable our model to do its data analysis to discover the significant relationships between the variables and the post-operative LOS, and then use them in building its algorithmic equations that will be employed in predicting postoperative LOS for a second group. In the second group, the "test group" (which contains another 28 patients), data will be provided to the model without the discharge date.

Written Commands for Experiment #2

Data was prepared, cleaned, graded, and arranged in an Excel file. We assessed the significance of each element in predicting hospital stay duration based on prior research and extant literature to furnish this information to ChatGPT [24–26]. We provided ChatGPT o1 with the following command (in short): "Use the patients'data, including lab values, imaging, and demographic variables. Analyze the data based on numeric encoding for the categorical variables. Use variable significance percentages when predicting post-operative LOS: CT Grade (20%), Pleural Culture Result (15%), Recurrent Empyema (12%), CRP Levels (10%), Fever Presence (8%), WBC Count (8%), Smoking (7%), Platelets (6%), Past Medical History (6%), Hemoglobin (5%), Sex (3%). Consider the following averages for hospital stay [26, 27]: Mean post-operative

LOS (all patients): 9.00 ± 5.59 days, adults (<70 years): 6.44 ± 2.35 days, elderly (≥70 years): 12.29 ± 9.70 days. Avoid extreme or unrealistic predictions. Ignore missing or unavailable data. Be proficient in handling Excel (.xlsx) files for data analysis". Following that, we commanded ChatGPT o1 to create a predictive algorithm model based on the above criteria. Afterward, we copied the model algorithm and equations and provided them to ChatGPT 4o in addition to the data Excel file, and commanded him to apply the model equations in predicting the post-operative LOS for each patient alone. Figure 1 shows the score of each variable in affecting the length of hospital stay.

Statistical analysis

The accuracy of ChatGPT in predicting the post-operative LOS for our patients depends on the following metrics:

1. Mean Absolute Error (MAE): explains the average absolute difference between the predicted and actual post-operative LOS.
2. Root Mean Square Error (RMSE): The square root of the average squared errors, indicating larger deviations.
3. Mean Absolute Percentage Error (MAPE): The average percentage error compared to actual post-operative LOS.
4. R-squared (R^2) and Explained Variance Score (EVS): To assess the proportion of variance the model explains and its overall explanatory power.
5. Median Absolute Error (MedAE): To evaluate the typical extent of errors, less influenced by outliers.
6. Accuracy Percentage (AP): To give a comprehensive idea about the model's predicting abilities.

Mathematical Formulation for Experiment #1

ChatGPT 4o developed a Random Forest Regressor model and trained it (according to the 28 patients of the training group) to predict the post-operative LOS for empyema thoracic patients undergoing uniportal VATS surgery. The model begins its work by collecting relevant information about each patient. These features form a set of input variables called X. For each patient, the model uses this data to predict post-operative LOS. The first mathematical process occurs when the model creates decision trees. Each tree acts as a set of if-then rules. For example, a tree might split patients based on whether their white blood cell count exceeds a certain

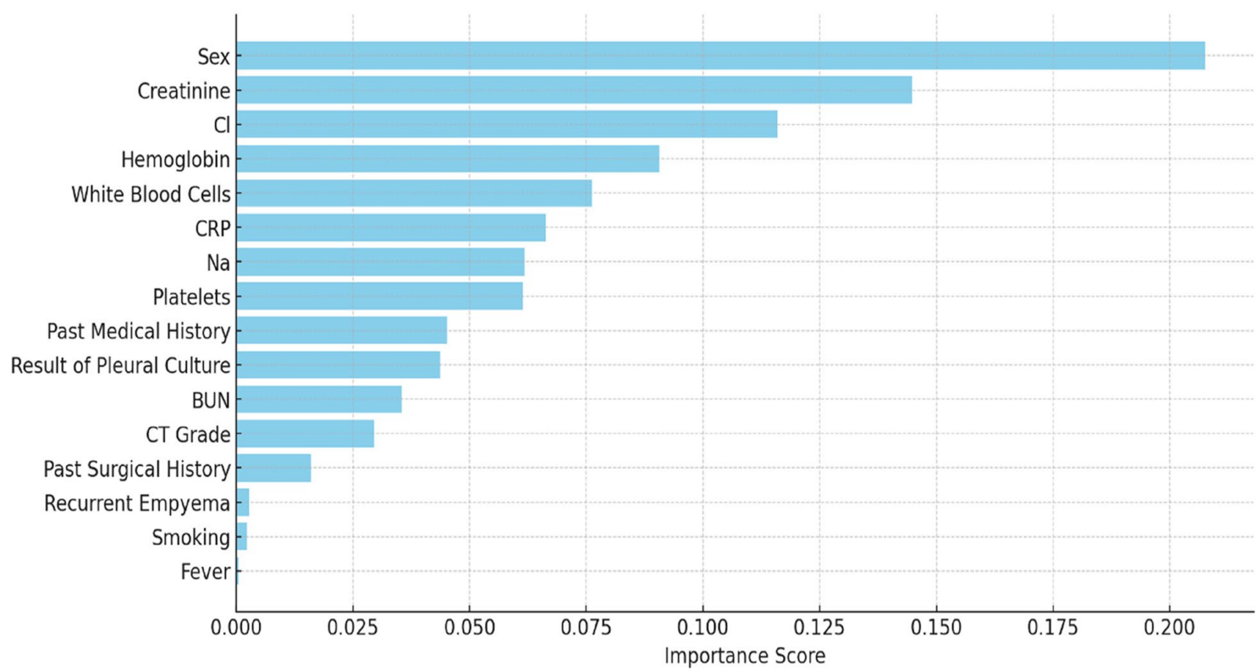

Fig. 1 Feature Importance in Tax Model. This bar chart illustrates the relative importance of various features used in the Tax model to predict postoperative length of stay (LOS) for pleural empyema patients."Sex"has the highest influence, followed by"Creatinine,""Chloride,"and other clinical and demographic factors. These importance scores were calculated using the reduction in prediction error attributable to each feature, highlighting the most impactful variables in the model's decision-making process

threshold or if creatinine is too high. Each split decision follows a condition like:

If: White Blood Cells > 10,000 and Creatinine > 1.2
Then: Predict longer hospitalization (e.g., 12 days).
Else: Predict shorter hospitalization (e.g., 5 days).

These rules are created by analyzing the training data and learning how features relate to the actual LOS. Each tree makes its prediction independently.

After making predictions, the model evaluates how accurate these predictions are using the Mean Squared Error (MSE) equation:

$$MSE = \frac{1}{m}\sum_{i=1}^{M}(y_i - \hat{y}_i)^2$$

Where:

- m: is the number of patients (samples) processed by the tree.
- y_i: is the actual post-operative LOS for patient i.
- $\hat{y}_i$: is the predicted value made by the decision tree.

The model calculates this error to improve future predictions. Lower MSE means the tree is making better predictions.

Once all trees in the Random Forest have made their predictions, the"Tax"model combines them by averaging the outputs using this equation:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} f_i(X)$$

Where:

- $\hat{y}$: is the final predicted post-operative LOS.
- N: is the number of decision trees.
- $f_i(X)$: is the prediction from tree i based on the input data X

This averaging reduces individual errors from each tree, creating a more accurate and stable prediction.

After making the final prediction, the"Tax"model analyzes how much each feature contributed to reducing prediction errors. This is calculated using the Feature Importance Equation:

$$I_k = \sum_{t=1}^{T} \Delta MSE_t$$

Where:

- I_k: is the importance score for feature k.
- T : is the total number of splits where feature k was used.
- ΔMSE : is the reduction in MSE after using feature k at split t

The features that reduce prediction errors the most are considered more important. For example, if"Creatinine"frequently helps the model make better predictions, its importance score increases. The following is a detailed chart showing each variable's weight (importance score).

Mathematical formulation for experiment #2

The core formula for the predicted LOS is:

$$PredictedLOS = BaselineLOS \times [1 + \Sigma(weight_i \times Score_i)]$$

Baseline LOS is 6.44 days if the patient is <70 years old and 12.29 days if ≥70 years old.

The weight is the fractional weight of the item variable, and the score is the score assigned to the item variable based on the patient's clinical values; both are defined in Table 1.

This table lists the features used in the model with their assigned weights and scoring criteria. Each feature, such as"CT Grade"or"Pleural Culture,"has specific score ranges based on patient data, contributing to the overall LOS prediction

Ethical considerations

This study procedure was approved by the research ethics committee at Al-Quds University (Ref#:484/REC/2025), in (Jan 22, 2025). The necessity for individual informed consent was waived due to the study's retrospective nature. All patient data were anonymized and securely stored in accordance with institutional data protection requirements.

Results

Results for experiment #1

Talking about the performance of the model in predicting post-operative LOS for patients with pleural empyema undergoing uniportal VATS, several findings were observed:

As seen in Fig. 2, this box plot compares the actual and predicted postoperative LOS for patients. The actual LOS shows a wider distribution, with some patients staying significantly longer. On the other hand, the predicted LOS is more clustered, showing the model's tendency to provide consistent but less varied predictions. While the model predicts the median LOS well, it has a narrower range of predictions.

Table 1 Variable Weights and Scoring System

Variable	Weight_i	Category/Range	Score
CT Grade	0.20	Stage 1	0.3
		Stage 2	0.6
		Stage 3	1.0
Pleural Culture	0.15	Unknown	0.5
		No Growth	0.0
		Positive	1.0
Recurrent Empyema	0.12	No	0.0
		Yes	1.0
CRP (mg/L)	0.10	<50	0.3
		50–100	0.6
		>100	1.0
Fever	0.08	No	0.0
		Yes	1.0
WBC (x10^9/L)	0.08	≤11	0.3
		>11–20	0.6
		>20	1.0
Smoking	0.07	Unknown	0.5
		Nonsmoker	0.0
		Smoker	1.0
Platelets (x10^9/L)	0.06	≤150	0.3
		150–400	0.0
		>400–800	0.6
		>800	1.0
Past Medical History	0.06	Free	0.0
		Local Non-Resp. Disease	0.3
		Systemic Non-Resp. Disease	0.6
		Chronic Respiratory	1.0
Hemoglobin (g/dL)	0.05	<10	1.0
		10–<12	0.6
		12–<14	0.3
		≥14	0.0
Sex	0.03	Male	0.3
		Female	0.0

Figure 3 shows how individual decision trees in the model predict the post-operative LOS for a single patient. Most predictions are located between 7 and 10 days, reflecting strong agreement among the trees, while a few outliers predict significantly shorter or longer stays (e.g., 5 or 20 days). The dashed line represents the model's final prediction, calculated as the average of all tree predictions.

The model showed suboptimal performance across multiple statistical metrics. The average prediction error (Mean Absolute Error, MAE) was 4.56 days, showing that, on average, the predicted LOS differed from the actual LOS by over four days. The Root Mean Square Error (RMSE) was 6.16 days, emphasizing the effects of large errors on the model's performance. The Mean Absolute Percentage Error (MAPE) showed that the predictions deviated from the actual LOS by more than half of the true value on average. This refers to a prediction accuracy of only 46.21%. The Median Absolute Error (MedAE) demonstrated that typical prediction errors were between 3 and 4 days, excluding outliers. Additionally, the R-squared (R^2) showed that the model performed worse than a simple mean-based prediction. The Explained Variance Score (EVS) was also negative indicating that the model failed to capture any meaningful variance in the target variable. All statistical metrics for experiment #1are listed in Table 2.

Results for experiment #2

Upon reviewing the model's efficacy in predicting post-operative LOS for patients with pleural empyema undergoing uniportal VATS, we came up with the following results:

The box plot in Fig. 4 shows how actual and predicted the LOS values differ. Each box covers the middle half of the data, with a line marking the median. The "whiskers" stretch out to points within one and a half times the box range, and anything beyond that counts as an outlier. The predicted LOS values are more tightly clustered, suggesting the model's results are less scattered. While both sets have outliers, the actual LOS values spread out more, hinting that real-world conditions produce a wider variety of outcomes than the model alone can reflect.

The following bar plot in Fig. 5 shows the absolute errors—calculated as the difference between each patient's predicted and actual LOS—with each bar corresponding to a single patient and its height representing the error magnitude in days. While many patients show relatively moderate gaps, a few stand out as outliers with notably greater errors. The uneven spread of these errors suggests that the model's accuracy is inconsistent across different people.

We discovered the model showed poor performance across multiple statistical metrics as seen in Table 3. On average, the model's predictions didn't get the actual LOS by more than four days. The RMSE reached 6.02 days. The MAPE indicates that the forecasts deviated by around fifty percent of the actual LOS value. Converting this error into an approximate accuracy measure suggests that the model's predictions were only about 49.47% accurate. Additionally, the median absolute error (MedAE) indicated that errors ranged from 3 to 4 days without the inclusion of outliers.

Both R-squared (R^2) and explained variance were negative, indicating that the model performed below the baseline method that assumes each patient's LOS equals the average. The model was unable to identify

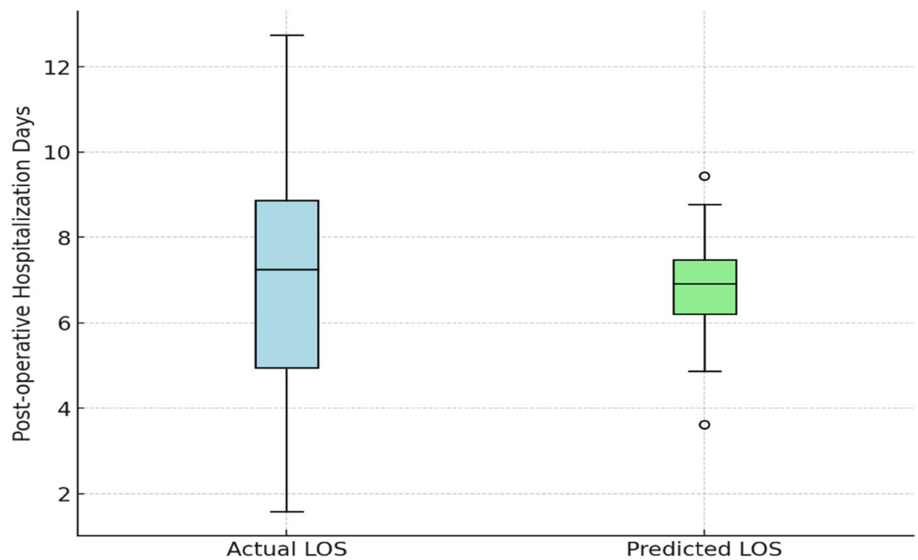

Fig. 2 Box Plot of Actual vs. Predicted LOS. This box plot compares the distribution of actual postoperative LOS values with those predicted by the model. The actual LOS demonstrates greater variability, with some extreme values, whereas the predicted LOS is more tightly clustered around the median. The narrower range of predictions reflects the model's limitations in capturing the full spectrum of clinical variability

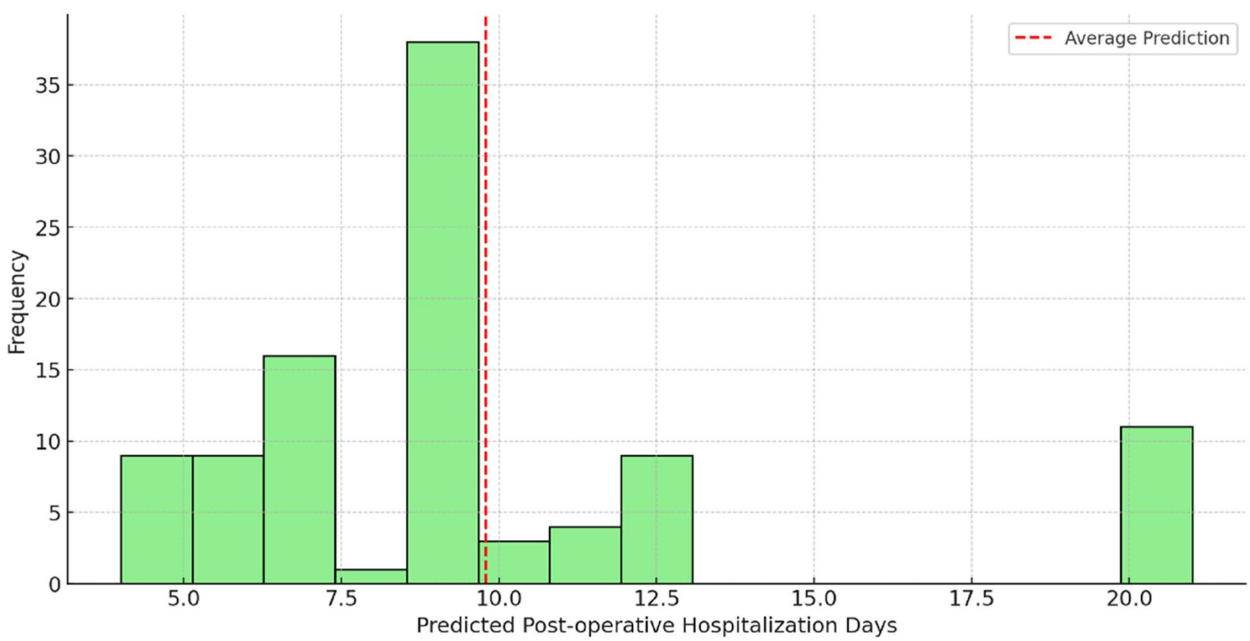

Fig. 3 Tree Predictions Distribution for Sample 1. This histogram visualizes the distribution of LOS predictions made by individual decision trees in the Tax model for a single patient. The majority of predictions fall within a narrow range of 7 to 10 days, while a few trees predict significantly shorter or longer stays. The red dashed line indicates the average prediction, representing the model's final output

any significant correlations between the inputs'weights (as instructed from the literature) and the LOS. Despite the median absolute accuracy being relatively improved, the overall accuracy metrics were adversely affected by substantial outliers.

This table shows the performance results for the literature-informed model. Metrics such as MAE, RMSE, and R-squared are included, demonstrating the model's prediction accuracy and comparison to Experiment #1

Table 2 Statistical metrics for experiment #1

Metric	Value
MAE (days)	4.56
MRSE (days)	6.16
MAPE (%)	53.79%
RR	−0.09
EVS	−0.09
MedAE (days)	3.29
AP (%)	46.21%

This table summarizes the performance of the Random Forest model in predicting LOS. Metrics include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2), indicating the model's accuracy and limitations

Discussion

Discussion for experiment #1

The Random Forest Regressor model displayed significant error rates, expressed by a large Mean Absolute Error (MAE) of 4.56 days and a considerable RMSE of 6.16 days. These findings indicate that the model's predictions fail to correspond with actual clinical outcomes. Similar difficulties in achieving accurate postoperative LOS predictions using ML have been highlighted, underscoring the difficulty of this issue [28].

Both R^2 and Explained Variance Score (EVS) were negative, demonstrating that the model could not exceed basic baseline predictions. This result shows the absence of important relationships identified by the model. Other research has also found difficulty in finding relevant patterns from patient data to predict LOS, pointing out that complex clinical variables limit the prediction effectiveness of basic ML models [29].

The predictions were narrower than the actual LOS values, suggesting insufficient responsiveness to the variability associated with patient scenarios. Comparable studies illustrate that patient heterogeneity, including comorbidities and varying postoperative courses, impairs reliable LOS prediction [30]. This limited variance in model output fails to correspond to the actual world variability.

The model identified a focus on certain features but could not translate these weights into better accuracy. Other experiments have demonstrated that basic feature weighting without solid clinical understanding rarely improves surgical LOS prediction [31]. This highlights the necessity of better clinical variable selection and enhanced modeling methodologies.

Discussion for experiment #2

Despite using literature-based instructions and weighted variables, the model could not reach adequate predicting accuracy, with an MAE over 4 days and an RMSE over 6 days. According to Rajkomar et al, adding written instructions from the literature does not enhance the model if the underlying data complexity is not addressed [31].

Assigning specified significance percentages to each feature did not improve performance measurements, including negative R^2 and EVS. Studies have demonstrated that static, literature-driven weight allocations

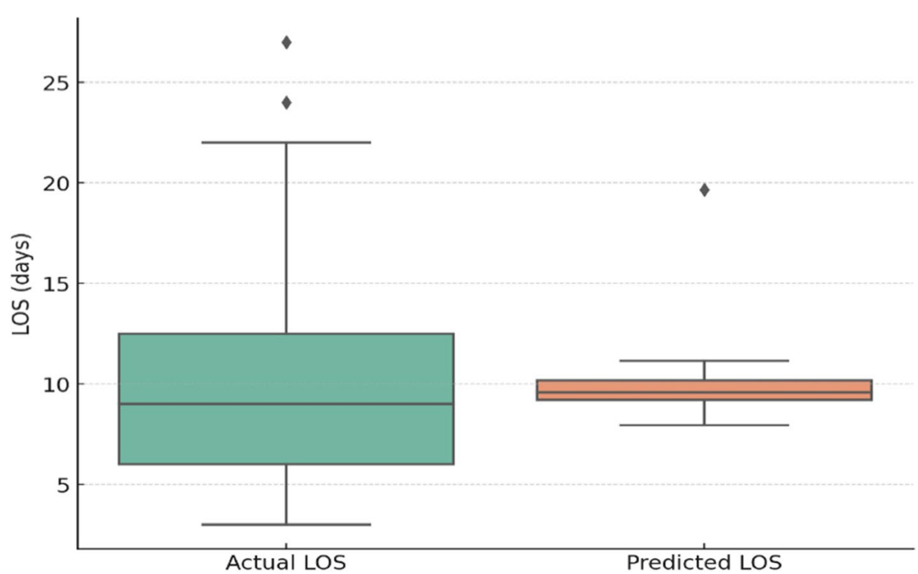

Fig. 4 Box Plot: Predicted vs. Actual LOS. This box plot provides a detailed comparison of actual and predicted LOS distributions. The actual LOS data shows a broader spread, including several outliers, while the predicted LOS is more condensed. The median values differ slightly, highlighting the model's tendency to underestimate extreme cases

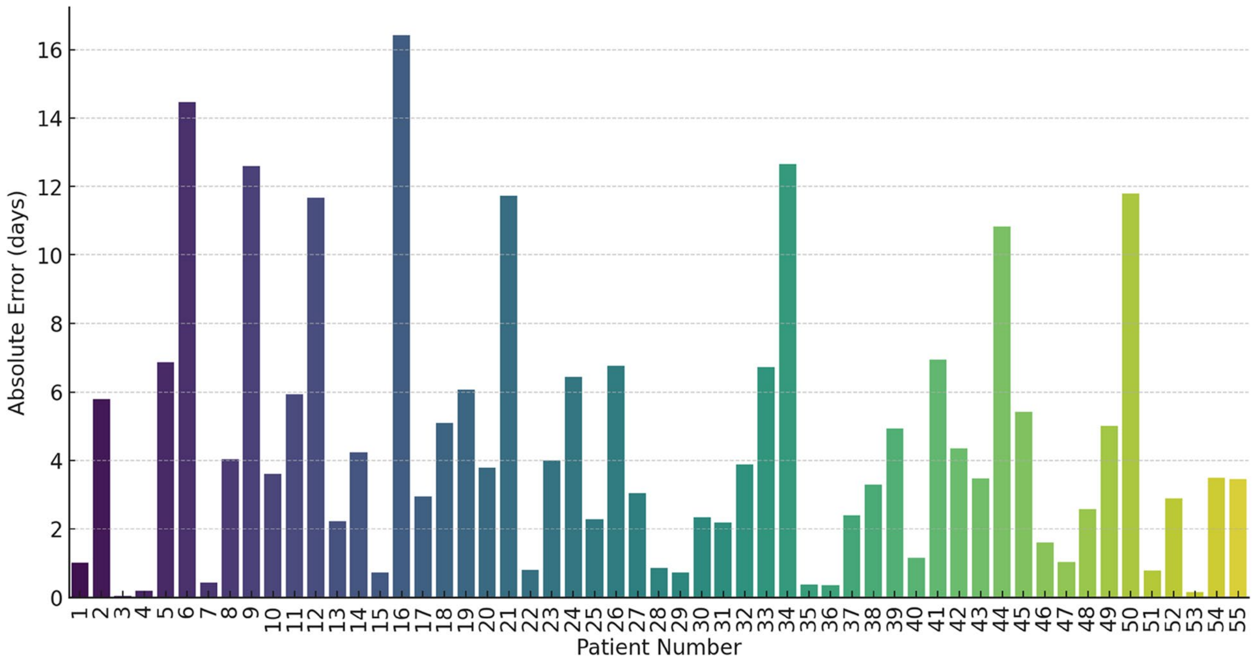

Fig. 5 Absolute Errors by Patient. This bar chart represents the absolute error for each patient, calculated as the difference between actual and predicted LOS. While most errors are moderate, a few outliers exhibit significantly larger deviations, underscoring inconsistencies in the model's accuracy across different cases

Table 3 Statistical Metrics for Experiment #2

Metric	Value
MAE (days)	4.53
MRSE (days)	6.02
MAPE (%)	50.53
RR	−0.10
EVS	−0.10
MedAE (days)	3.50
AP (%)	49.47

generally do not capture individual patient variability, making it difficult to raise LOS estimates [28].

The bar plot of absolute errors revealed differing levels of accuracy among patients. Such inconsistent error patterns have been observed in previous LOS prediction attempts, where unique patient characteristics are particularly difficult to forecast [30]. This discrepancy highlights the necessity for more precise, patient-specific modeling methods.

The model's inability to capture the large distribution of actual LOS lowers its practical utility. Research on LOS prediction frequently points out that models must accurately capture the complete result range to inform decision-making and resource allocation [31]. Without this feature, the model's usefulness in a clinical situation remains seriously limited. AI in thoracic surgery has the potential to improve significantly by integrating larger, diverse datasets and promoting collaboration across multiple centers. These steps can help improve predictions and make AI tools more reliable for clinical use. AI can help plan resources better by predicting longer hospital stays.

Regarding the study limitations, it is worth mentioning that the sample size was small and collected from a single hospital, which restricts the ability to apply the findings to other populations. Some variables, including smoking status and pleural culture results, were missing, means that the data wasn't fully complete. Variables like albumin levels were only available for a few patients, which prevented the inclusion of additional variables in the study. Additionally, the extreme outliers, who had a hospital stay extending to two months or less than one day, were excluded, further reducing the sample size.

A further limitation is the omission of critical AI-related characteristics associated with high-risk patient identification and the prediction of postoperative complications. Clinical indicators, such as complication staging were excluded from the model and search parameters. This may diminish the model's efficacy in identifying early problems and constrains its practical clinical use.

Conclusion

This study evaluates the potential applicability of AI and ML models in predicting the LOS for patients with pleural empyema treated with uniportal VATS. Both the Random Forest Regressor and the literature-informed model demonstrated poor predictive accuracy, with Mean Absolute Errors exceeding four days and negative R-squared values. These results indicate that the current AI approaches and variable weighting strategies can be still inadequate for accurately predicting postoperative LOS in this patient group. Consequently, our research questions and hypotheses were not supported, underscoring the complexity of LOS prediction in clinical settings. However, we believe this study shows a promising future for incorporating AI and ML in the field of medical-care disciplines. Future studies should employ larger collections acquired from more than one hospital, assure inclusive data collection, and apply advanced modeling approaches.

Abbreviations

LOS	Length of Hospital Stay
VATS	Video-Assisted Thoracoscopic Surgery
PACU	Post-Anesthesia Care Unit
AI	Artificial Intelligence
ML	Machine Learning
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
R^2	R-squared
EVS	Explained Variance Score
MedAE	Median Absolute Error
AP	Accuracy Percentage
MSE	Mean Squared Error

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12893-025-02959-w.

Supplementary Material 1.

Acknowledgments

We would like to thank the thoracic surgery team at Al-Ahli hostital - Hebron/Palestine for their efforts in the treatment journey and providing information.

Authors'contributions

I.A., B.A., T.A., and A.O.: collect and prepare data. All authors participate in main manuscript writing. Y.A. and T.J.: review & edit the main manuscript. All authors read and approved the final manuscript.

Funding

This case has not received any financial support.

Data availability

The data that support the findings of this study are available from Al-Ahli hostital - Hebron/Palestine but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Al-Ahli hostital adminstration.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the principles of the Declaration of Helsinki. The study protocol was approved by the research ethics committee at Al-Quds University (Ref#:484/REC/2025). No names or personal identification numbers were included. All authors promise to use this data strictly for research reasons (see Additional file 1).

Consent for publication

Not applicable

Competing interests

The authors declare no competing interests.

Received: 1 February 2025 Accepted: 9 May 2025
Published online: 19 May 2025

References

1. Grott M, Eichhorn M, Eichhorn F, Schmidt W, Kreuter M, Winter H. Thoracic surgery in the non-intubated spontaneously breathing patient. Respiratory Research. 2022 Dec 27;23(1).
2. Byington CL, Spencer LY, Johnson TA, Pavia AT, Allen D, Mason EO, et al. An Epidemiological Investigation of a Sustained High Rate of Pediatric Parapneumonic Empyema: Risk Factors and Microbiological Associations. Clinical Infectious Diseases. 2002;34(4):434–40.
3. Bradley JS, Byington CL, Shah SS, Alverson B, Carter ER, Harrison C, et al. The Management of Community-Acquired Pneumonia in Infants and Children Older Than 3 Months of Age: Clinical Practice Guidelines by the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America. Clinical Infectious Diseases. 2011;53(7):e25-76.
4. Tuomanen EI, Austrian R, Masure HR. Pathogenesis of Pneumococcal Infection. Epstein FH, editor. New England Journal of Medicine. 1995 May 11;332(19):1280–4.
5. Christensen T, Bendixen M, Søren Helbo Skaarup, Jens-Ulrik Stæhr Jensen, René Horsleben Petersen, Christensen M, et al. Intrapleural fibrinolysis and DNase versus video-assisted thoracic surgery (VATS) for the treatment of pleural empyema (FIVERVATS): protocol for a randomised, controlled trial – surgery as first-line treatment. BMJ Open. 2022 Mar 1;12(3):e054236–6.
6. Manolache V, Motas N, Davidescu M, Bluoss C, Rus O, Tanase B, et al. Minimally Invasive Thoracic Surgery - Video Assisted Thoracic Surgery: Technique and Indications. Chirurgia. 2022;117(1):101.
7. Chen KC, Chen HY, Lin JW, Tseng YT, Kuo SW, Huang PM, et al. Acute thoracic empyema: Clinical characteristics and outcome analysis of video-assisted thoracoscopic surgery. Journal of the Formosan Medical Association. 2014;113(4):210–8.
8. Reichert M, Pösentrup B, Hecker A, Schneck E, Pons-Kühnemann J, Augustin F, et al. Thoracotomy versus video-assisted thoracoscopic surgery (VATS) in stage III empyema—an analysis of 217 consecutive patients. Surgical Endoscopy. 2017;32(6):2664–75.
9. Stüben BO, Plitzko GA, Sauerbeck J, Busch P, Melling N, Reeh M, et al. Minimally invasive intrathoracic negative-pressure therapy and flexible thoracoscopy (FlexVATS) for patients with pleural empyema. Scientific Reports [Internet]. 2023 Jul 5 [cited 2023 Jul 15];13(1):10869.
10. Siti Nursyafiqah Sulaiman, Zalina Abu Zaid, Mohd N, A'zim A. Length of Hospital Stay and Its Associated Factors among Surgical Patients in Hospital Serdang. Malaysian Journal of Medicine and Health Sciences/Malaysian journal of medicine and health sciences. 2023 Sep 11;19(5):51–8.
11. Gagarine A, Urschel JD, Miller JD, Bennett WF, Young JEM. Preoperative and intraoperative factors predictive of length of hospital stay after pulmonary lobectomy. Annals of thoracic and cardiovascular surgery : official journal of the Association of Thoracic and Cardiovascular Surgeons of Asia [Internet]. 2003;9(4):222–5.
12. Tat Bang H, Thanh Vy T, Tap NV. Length of Postoperative Hospital Stay and Related Factors After Lobectomy for Lung Cancer: A Pre-enhanced

Recovery After Surgery (ERAS) Single Center Assessment. Cureus. 2024 Feb 22;

13. Giambrone GP, Smith MR, Wu X, Gaber-Baylis LK, Bhat AU, Ramin Zabih, et al. Variability in length of stay after uncomplicated pulmonary lobectomy: is length of stay a quality metric or a patient metric? European Journal of Cardio-Thoracic Surgery. 2016 Apr 1;49(4):e65–71.
14. Pompili C, McLennan Battleday F, Chia WL, Chaudhuri N, Kefaloyannis E, Milton R, et al. Poor preoperative quality of life predicts prolonged hospital stay after VATS lobectomy for lung cancer. European Journal of Cardio-Thoracic Surgery. 2020;59(1):116–21.
15. Maniscalco P, Tamburini N, Fabbri N, Quarantotto F, Rizzardi G, Amore D, et al. Factors Associated with Early Discharge after Thoracoscopic Lobectomy: Results from the Italian VATS Group Registry. Journal of Clinical Medicine. 2022;11(24):7356–6.
16. Clovis, Schmidt A, Elaine Aparecida Felix, Bianchi F, Fernanda Lamego Guerra, Cristiano Feijó Andrade. Risk factors for postoperative pulmonary complications and prolonged hospital stay in pulmonary resection patients: a retrospective study. 2021 Feb 3;71(4):333–8.
17. Zhang G, Liu X, Hu Y, Luo Q, Ruan L, Xie H, et al. Development and comparison of machine-learning models for predicting prolonged postoperative length of stay in lung cancer patients following video-assisted thoracoscopic surgery. Asia-Pacific Journal of Oncology Nursing [Internet]. 2024;11(6): 100493.
18. Nomura A, Noguchi M, Kometani M, Furukawa K, Yoneda T. Artificial Intelligence in Current Diabetes Management and Prediction. Current Diabetes Reports. 2021 Dec;21(12).
19. Puterman-Salzman L, Katz J, Bergman H, Grad R, Vladimir Khanassov, Gore G, et al. Artificial Intelligence for Detection of Dementia Using Motion Data: A Scoping Review. Dementia and geriatric cognitive disorders extra. 2023 Sep 13;13(1):28–38.
20. Lex JR, Di Michele J, Koucheki R, Pincus D, Whyne C, Ravi B. Artificial Intelligence for Hip Fracture Detection and Outcome Prediction. JAMA Network Open [Internet]. 2023;6(3): e233391.
21. Zhang J, Yang P, Zeng L, Li S, Zhou J. Ventilator Associated Pneumonia Prediction Models Based on Artificial Intelligence: Scoping Review (Preprint). JMIR Medical Informatics. 2024;11(12):e57026-6.
22. Towe CW, Carr SR, Donahue JM, Burrows W, Perry Y, Kim S, et al. Morbidity and 30-day mortality after decortication for parapneumonic empyema and pleural effusion among patients in the Society of Thoracic Surgeons' General Thoracic Surgery Database. The Journal of Thoracic and Cardiovascular Surgery. 2019;157(3):1288-1297.e4.
23. Light, R. W. (2013). Pleural diseases (6th ed.). Lippincott Williams & Wilkins.
24. Yang D, Hong Q, Zhao C, Mu J. Postoperative Patient-Reported Outcomes after Uniportal Video-Assisted Thoracoscopic Surgery Using the Perioperative Symptom Assessment for Lung Surgery Scale. Current Oncology. 2022;29(10):7645–54.
25. Luciani C, Scacchi A, Vaschetti R, Giancarlo Di Marzo, Fatica I, Cappuccio M, et al. The uniportal VATS in the treatment of stage II pleural empyema: a safe and effective approach for adults and elderly patients—a single-center experience and literature review. World Journal of Emergency Surgery. 2022 Aug 29;17(1).
26. Yan Y, Huang Q, Han H, Zhang Y, Chen H. Uniportal versus multiportal video-assisted thoracoscopic anatomical resection for NSCLC: a meta-analysis. J Cardiothoracic Surg. 2020;15(1).
27. Kong G, Lin K, Hu Y. Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. BMC Medical Informatics and Decision Making. 2020 Oct 2;20(1).
28. Xu Z, Zhao C, Scales CD, Henao R, Goldstein BA. Predicting in-hospital length of stay: a two-stage modeling approach to account for highly skewed data. BMC Med Informatics Decision Making. 2022;22(1).
29. Gardner A, Ruch A. Not All Stridor Is Croup. Pediatric Emergency Care. 2018 Sep;1.
30. Zhong Z, Yuan X, Liu S, Yang Y, Liu F. Machine learning prediction models for prognosis of critically ill patients after open-heart surgery. Sci Rep. 2021;11(1).
31. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. New Engl J Med. 2019;380(14):1347–58.

Publisher's Note