Original Study

Annals of Surgery Open

OPEN

Development of Multiservice Machine Learning Models to Predict Postsurgical Length of Stay and Discharge Disposition at the Time of Case Posting

Hamed Zaribafzadeh, MS,* T. Clark Howell, MD, MSHS,* Wendy L. Webster, MA, MBA,†
Christopher J. Vail, PA-C, MMCi,* Allan D. Kirk, MD, PhD,* Peter J. Allen, MD,* Ricardo Henao, PhD,‡ and
Daniel M. Buckland, MD, PhD§‖

Objective: Develop machine learning (ML) models to predict postsurgical length of stay (LOS) and discharge disposition (DD) for multiple services with only the data available at the time of case posting.

Background: Surgeries are scheduled largely based on operating room resource availability with little attention to downstream resource availability such as inpatient bed availability and the care needs after hospitalization. Predicting postsurgical LOS and DD at the time of case posting could support resource allocation and earlier discharge planning.

Methods: This retrospective study included 63,574 adult patients undergoing elective inpatient surgery at a large academic health system. We used surgical case data available at the time of case posting and created gradient-boosting decision tree classification models to predict LOS as short (≤1 day), medium (2–4 days), and prolonged stays (≥5 days) and DD as home versus nonhome.

Results: The LOS model achieved an area under the receiver operating characteristic curve (AUC) of 0.81. Adding relative value unit and historical LOS through the similarity cascade increased the accuracy of short and prolonged LOS prediction by 9.0% and 3.9% to 72.9% and 74%, respectively, compared with a model without these features ($P = 0.001$). The DD model had an AUC of 0.88 for home versus nonhome prediction.

Conclusions: We developed ML models to predict, at the time of case posting, the postsurgical LOS and DD for adult elective inpatient cases across multiple services. These models could support case scheduling, resource allocation, optimal bed utilization, earlier discharge planning, and preventing case cancelation due to bed unavailability.

Keywords: machine learning, length of stay, discharge disposition, surgical case posting

INTRODUCTION

Length of stay (LOS) is an important metric for assessing hospital operational capacity. LOS impacts patient flow, outcomes, hospital resource utilization, and surgical capacity.[1–9] Higher LOS increases the cost of care and is a surrogate for high utilization, low efficiency, and increased morbidity, mortality, and readmission risk.[3,6,7,10–14] Intertwined with LOS is discharge disposition (DD), as nonhome DD is associated with longer LOS. Therefore, predicting postoperative LOS and DD at the time of case posting should allow for higher and more consistent patient throughput, thus increasing access to care including surgical care.[6,12,15–17]

Predicting LOS is complex; surgical patients undergoing the same procedure may have variable LOS due to complex individual, social, financial, clinical, and operational factors.[2,5,15] Because of this, human-predicted LOS is inconsistent subject to institutional and individual biases. Few studies have attempted to make LOS predictions with the limited data available at the time of case posting[2,18] when the full clinical background and demographics may not be consistently available for predictive modeling. Similarly, predicting DD is complex and highly dependent on regional and institutional norms and the same extraclinical factors as LOS.

Yet, LOS and DD can be predicted from the data available at the time of case posting. Stonko et al[10] showed that characteristics and decisions from earlier in the hospital encounter are more associated with prolonged LOS than later hospitalization characteristics, indicating that patients with prolonged LOS could be identified as soon as their cases are posted. Similarly, it has been also shown that early hospital encounter characteristics such as patient age, admission type, and hospital type are significantly associated with prolonged LOS.[2] Moreover, identifying patients who will not require prolonged LOS could help allocate the proper attention and resources to patients who may need to stay longer in the hospital.[1,2,19,20] Identifying patients with prolonged LOS and nonhome DD may also facilitate targeted interventions such as improved resource allocation, discharge planning, social support, specialty care, and alternative long-term care facility coordination.[2,6,10,19,21] More accurate and consistent prediction of LOS facilitates better management of bed utilization and patient flow, coordinating bed utilization with surgical volume, and adjusting staff and surgical schedules. Patient-facing,

From the *Department of Surgery, Duke University, Durham, NC; †Perioperative Services, Duke University Health System, Durham, NC; ‡Department of Biostatistics and Bioinformatics, Duke University, Durham, NC; §Department of Emergency Medicine, Duke University, Durham, NC; and ‖Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC.

This work received operational funding from the Department of Surgery, Duke University Medical Center.

Disclosure: The authors declare that they have no conflict of interest.

SDC Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Reprints: Daniel M. Buckland, MD, PhD, Department of Emergency Medicine, Duke University, 2301 Erwin Road, DUMC Box 3096, Durham, NC 27710. Email: dan.buckland@duke.edu.

Annals of Surgery Open (2025) 1:e547

Received: 30 September 2024; Accepted 1 January 2025

Published online 31 January 2025

DOI: 10.1097/AS9.0000000000000547

improved LOS prediction can preallocate appropriate resources and coordinate in-hospital and posthospital patient care to reduce LOS, wait time, and hospitalization costs, thus avoiding overutilization and underutilization for the system.[1,7,8,18,20,22,23]

LOS and DD are central operational factors in the care of surgical patients, and their predictions have previously been shown to be feasible. We hypothesized that LOS and DD can be reliably predicted from the data available at the time of case posting. We have previously reported the development and implementation of a surgical case length prediction model across our institution.[24] This work adapts the methodology of our previously implemented solutions to develop machine learning models to predict, at the time of case posting, whether a patient will (1) have a short, medium, or long LOS and (2) be discharged home or to a nonhome facility after surgery.

METHODS

Source of Data

This study was determined to be exempt by the Duke University Institutional Review Board (protocol number: Pro00112866). Duke University Health System (DUHS) consists of 3 hospitals with 1622 licensed inpatient beds and admitted 66,998 patients in 2021. Elective adult inpatient cardiothoracic, general surgery, gynecology, neurosurgery, obstetrics, orthopedics, otolaryngology, plastic surgery, podiatry, and urology cases across DUHS hospitals from January 2019 to April 2022 were identified. Outpatient or ambulatory cases were excluded as such cases rarely require inpatient beds. Cases with missing timestamps or negative LOS and patients with multiple cases during the same hospital encounter were excluded from the cohort. All remaining missing data were treated by XGBoost sparsity-aware split finding. For the LOS model, 57,912 cases from 2019 to 2021 and 5662 cases in 2022 were selected for training and testing sets, respectively, to consider temporal variations and practical application of the model (Supplemental Figure 1, see http://links.lww.com/AOSO/A455). For the DD model, we also excluded cases with missing, court, or law enforcement disposition labels and expired patients or those who left against medical advice; therefore, 53,203 cases from 2019 to 2021 and 5142 cases in 2022 were selected for training and testing sets, respectively (Supplemental Figure 1, see http://links.lww.com/AOSO/A455).

Outcomes

We explored our electronic health record (EHR) system for an estimated LOS and identified the expected discharge date (EDD) that is generated by case managers and care teams. EDD is a human-predicted LOS that is assigned to a case before or after admission and regularly updated during patient hospitalization. We extracted the first posted EDD, hereafter referred to as EDD, to evaluate the timing of the EDD availability with respect to the day of surgery. Our outcomes were postoperative LOS and DD. LOS was defined as the number of days a patient stayed in the hospital after surgery for inpatient postoperative care, regardless of the hours of surgery and discharge. The LOS distribution is often highly skewed, and cases with long LOS are not as numerous as cases with shorter LOS, which makes predicting long stays extremely challenging. Moreover, the operational impact of patients with long LOS in terms of the number of bed days is quite considerable. Therefore, after discussions with our case management leadership, the following 3 classes were considered for LOS. Prolonged LOS was defined as ≥5 days, which is over the 75th percentile of LOS distribution and includes complications and other issues leading to longer stays.[15,19,25] In fact, ≤5 days is often considered an accepted ideal LOS for major surgeries.[26] Shorter LOS was classified as 2 operationally meaningful periods as 2 to 4 days of stay (medium LOS) and discharge on the same day or 1 day after the procedure (short LOS). These groupings are preferable and more practical than predicted probabilities for operational decision-making[27,28] and can even stratify patients according to their LOS period.[14] DD was classified as home and nonhome destinations. The home destination included home or self-care. Nonhome destinations included other facilities such as skilled nursing facilities, home health services, rehab facilities, hospices, other acute hospitals or healthcare institutions, and long-term acute care (Supplemental Table 1, see http://links.lww.com/AOSO/A455).

Predictors

To ensure the operational utility of our model, we selected data available at the time of case posting. The selected case posting data include age, gender, patient class, service, number of panels, number of current procedural terminology (CPT) codes, primary anesthesia type, location, laterality, EHR system-generated median case length, and surgeon-estimated case length (Supplemental Table 1, see http://links.lww.com/AOSO/A455). We previously showed that CPT conversion to a single work relative value unit (RVU) captures case complexity and reduces sparsity.[29] Moreover, it has been shown that higher work RVU is associated with increased LOS in posterior spine fusion surgical cases.[30] Therefore, we converted all CPTs to total work RVU and incorporated that into our models. We used a similarity cascade[24] to implicitly capture the surgeon's influence on LOS, find historical cases with similar characteristics including primary surgeon, primary CPT, number of CPTs, and RVU range, calculate historical mean and SD of LOS, and incorporate them into the feature set.

LOS Model Development and Evaluation

Among different classification models, the gradient-boosted tree model has been one of the top-performing models that provide balanced bias and variance, is robust to overfitting, and is appropriate for multiclassification prediction.[2,3,23,31] We created 3 gradient-boosted decision tree multiclassification models with cross-entropy loss function using XGBoost[32] (version 1.2.1) in Python (version 3.7.6) to predict case LOS class. All hyperparameter optimization was performed by 5-fold cross-validation on the training set. The first model included all the above predictors except the RVU and historical mean and SD of LOS derived from the similarity cascade. The second model included all the first model's features and the total work RVU. In the third model, we used a similar feature set as the second model and added the historical mean and SD of LOS calculated by the similarity cascade approach. To consider the effect of imbalance among the 3 (short, medium, and long) LOS classes, we applied sample weights when fitting the models. We defined underprediction and overprediction errors as predicted LOS periods less or more than the actual LOS period, respectively, and compared models using one-way ANOVA followed by the Tukey post-hoc test. Accuracy, sensitivity (or recall), specificity, precision (or positive predictive value), balanced accuracy (BA), geometric mean of classwise sensitivity (G-mean), index of BA, unweighted Cohen kappa score, F1 score, and the area under the receiver operating characteristic curve (AUC) were all reported for each model. The 95% confidence intervals (CIs) of these metrics were calculated by bootstrapping over the test set. Fernandez and Vatcheva[33] highlighted the limitations of the Poisson regression model for overdispersed data such as LOS and suggested that the negative binomial (NB) regression model provides a better fit in data with both overdispersion and zero inflation. The LOS model errors were further compared with Poisson and NB regression model errors.

DD Model Development and Evaluation

We similarly created a gradient-boosted decision tree binary classification model using XGBoost[32] (version 1.2.1) in Python (version 3.7.6) to predict whether a patient will be discharged home or to a nonhome facility. Five-fold cross-validation on the training set was performed to optimize the hyperparameters. This model utilized the same feature set as the third LOS model described earlier. Similar performance metrics, geometric mean of sensitivity and specificity (G-mean), receiver operating characteristic, precision-recall (PR) curves, AUC, and the area under the PR curve or average precision, were reported to evaluate the model performance. Bootstrapping over the test set was used to calculate the 95% CI of these metrics.

RESULTS

LOS Distribution and Availability of EDD

The LOS distribution from January 2019 to April 2022 is highly right-skewed, as shown in Figure 1A. Prolonged LOS (ie, the 75th percentile) was calculated at 5 days. Patients with home and nonhome DDs exhibit very different LOS distributions (Fig. 1B). The median LOS for patients who were discharged home and to nonhome facilities were 2 and 6 days, respectively. We quantified the timing of the EDD availability with respect to the day of surgery to show if EDD was generated before, after, or on the day of surgery. The EDD was available only for 17.8% of cases at least 1 day before the day of surgery, which comprised 8.4%, 14.5%, and 33.8% of cases with short, medium, and prolonged LOSs, respectively (Fig. 2A). Interestingly, 95.5% of these cases were patients who were admitted with an assigned case manager, and their surgical case was posted during their hospital encounter. Only the remaining 4.5% of these cases and 0.8% of the whole cohort were patients with an available EDD before their admission to the hospital. Regarding DD, only 13.0% of home and 36.5% of nonhome discharges were assigned the EDD before the day of surgery. Cases discharged to nonhome facilities comprised 11.0% of the entire cohort. Figures 2B and 2C show EDD availability for cases with nonhome and home discharge destinations, respectively. Only 6.2% of home and 1.9% of nonhome discharge cases were patients with EDD before the day of surgery who represented 0.7% and 0.08% of the entire cohort, respectively.

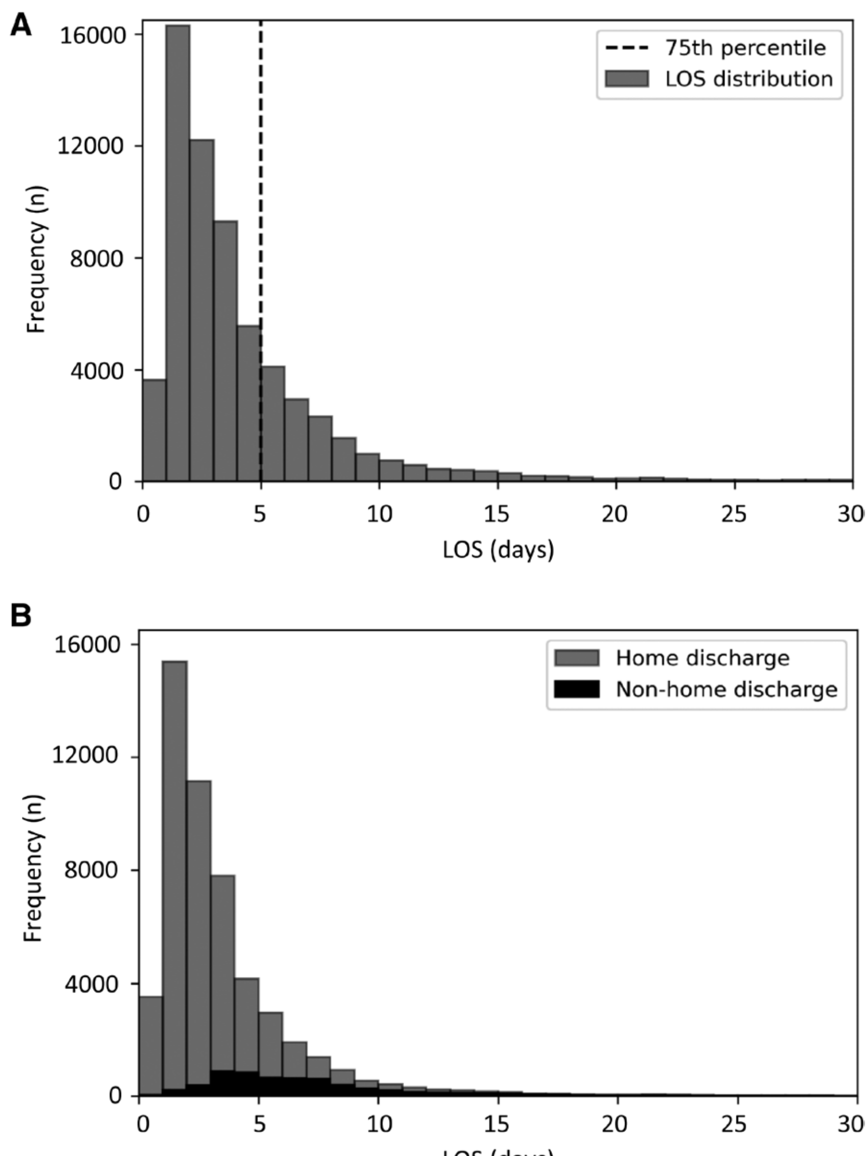

FIGURE 1. LOS distribution. A, LOS distribution of the entire cohort. The dashed line represents the 75th percentile. B, LOS distribution for cases with home and nonhome discharge destinations, illustrated by gray and black lines, respectively.

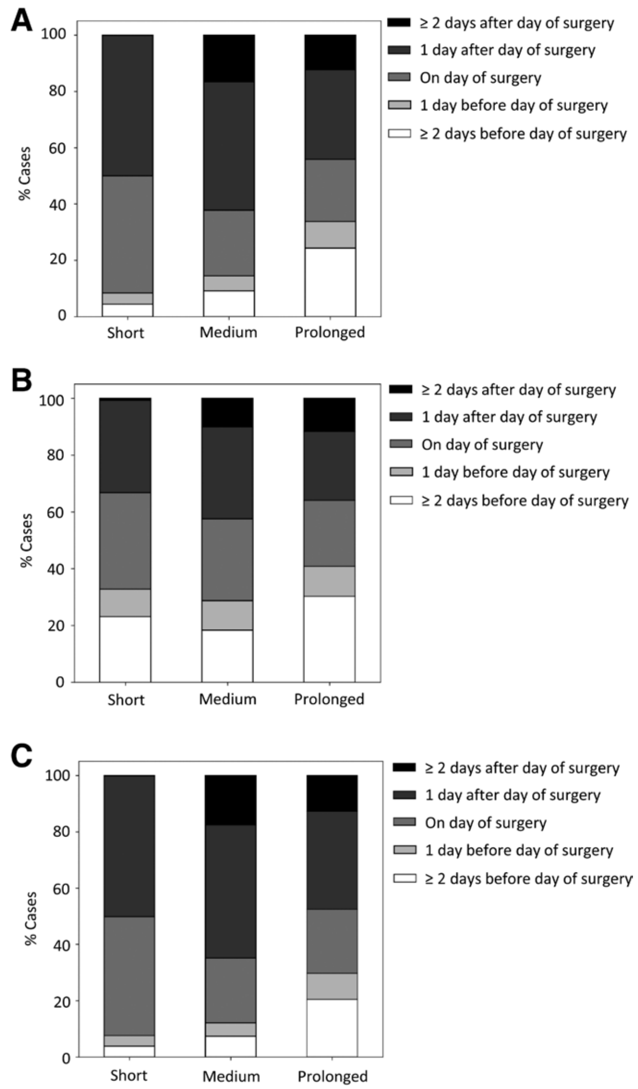

FIGURE 2. EDD availability from the day of surgery. A, The entire cohort. B, Cases with nonhome discharge disposition. C, Cases with home discharge disposition.

LOS Model Performance

We created 3 multiclassification LOS models and compared their accuracy in predicting the LOS classes (Fig. 3). The overall percentage of cases with predicted LOS class as same as the actual LOS group was improved by 4.2% (P = 0.001) and increased to 64.9% (95% CI, 63.2%–66.5%) after adding work RVU, as shown in Figure 3. The percentage of accurately predicted LOS for short cases was increased to 71.8% (95% CI, 68.5%–75.2%), a 7.9% improvement (P = 0.001), by the addition of RVU and resulted in 70.6% (95% CI, 67.2%–73.7%; P = 0.001) accuracy for prolonged stays. Introducing the historical LOS mean and SD of past similar cases calculated by the similarity cascade resulted in 64.6% (95% CI, 62.6%–66.4%) overall accuracy, which was approximately the same as the second model with RVU (P = 0.34). However, adding LOS mean and SD increased the accuracy of short and prolonged LOS prediction by 1.1% (P = 0.001) and 3.4% (P = 0.001) to 72.9% (95% CI, 70.1%–76.2%) and 74.0% (95% CI, 71.2%–76.9%), respectively, compared with the model with RVU. While the third model outperformed the other 2 models in short and prolonged LOS prediction accuracy, all 3 models showed average performance for the medium LOS, and the second model resulted in 56.3% (95% CI, 53.9%–59.6%) accuracy, which was higher than the other 2 models (P = 0.001). We compared the performance of these 3 models without applying the sample weight and observed approximately the same overall accuracy, decreased short and prolonged performance, and improved medium LOS accuracy (Supplemental Figure 2, see http://links.lww.com/AOSO/A455). Other performance metrics are reported in Table 1, showing improvement in all the metrics, except the index of BA by the addition of the RVU. Models 2 and 3 performed similarly in all the metrics with a 1% difference in specificity, precision, weighted average AUC, and F1. We selected model 3 because of higher accuracy in short and prolonged LOSs (Fig. 3) and operational utility and calculated percentage of underprediction and overprediction errors. Figure 4A shows 16.3% underprediction and 19.1% overprediction errors. To further evaluate the percentage of short LOS underprediction and prolonged LOS overprediction, the errors were broken into 2 other false predicted classes. Figure 4B shows that of 27.1% overpredicted short LOS cases, 16.4% were predicted with medium and 10.6% with prolonged LOS. Similarly, 15.4% of underpredicted prolonged cases were assigned a medium and 10.4% a short LOS. The EDD and LOS model accuracies for

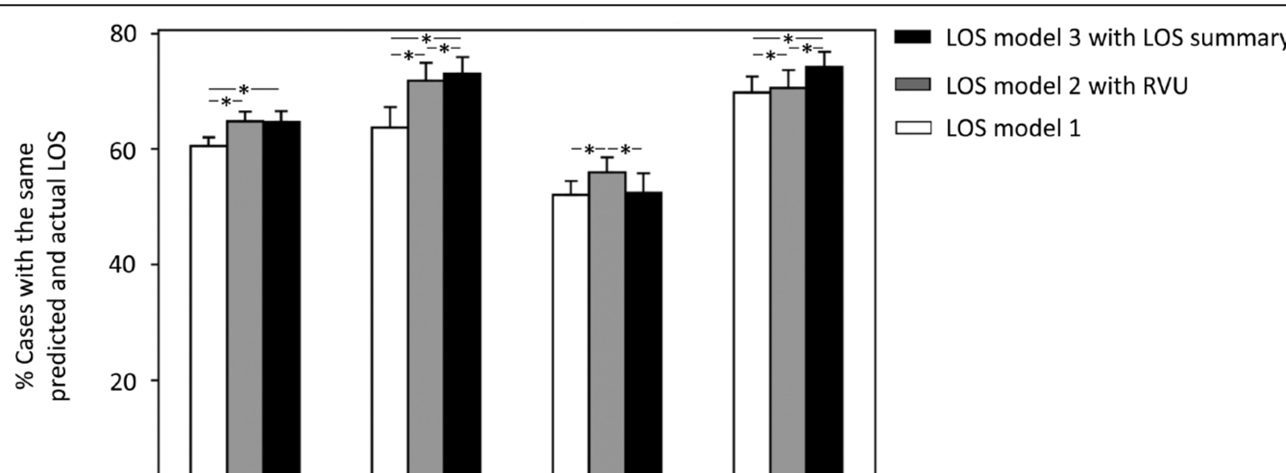

FIGURE 3. LOS models' comparison showing the percentage of cases with accurate LOS prediction in all classes and each individual LOS class. LOS summary includes historical LOS mean and SD calculated via similarity cascade. Error bars represent 95% CIs. *Statistically significant difference at α of 0.05 by the Tukey test.

TABLE 1.

Model Performance Metrics With 95% CIs

Metric	LOS Model 1	LOS Model 2 (With RVU)	LOS Model 3 (With LOS Summary)	DD Model
Accuracy	0.60 (0.59–0.62)	0.65 (0.63–0.66)	0.65 (0.63–0.67)	0.91 (0.90–0.92)
Sensitivity	0.60 (0.59–0.62)	0.65 (0.63–0.66)	0.65 (0.63–0.67)	0.91 (0.90–0.92)
Specificity	0.80 (0.80–0.81)	0.82 (0.81–0.83)	0.83 (0.82–0.84)	0.41 (0.37–0.45)
Precision	0.61 (0.60–0.63)	0.65 (0.64–0.67)	0.66 (0.64–0.68)	0.90 (0.89–0.91)
AUC	0.79 (0.78–0.80)	0.82 (0.81–0.83)	0.81 (0.80–0.83)	0.88 (0.85–0.90)
F1	0.60 (0.59–0.62)	0.65 (0.63–0.66)	0.64 (0.62–0.66)	0.90 (0.88–0.91)
BA	0.62 (0.60–0.63)	0.66 (0.65–0.68)	0.66 (0.65–0.68)	0.66 (0.64–0.69)
G-mean	0.70 (0.68–0.71)	0.73 (0.72–0.74)	0.73 (0.71–0.74)	0.58 (0.53–0.62)
IBA	0.48 (0.46–0.49)	0.52 (0.51–0.54)	0.52 (0.50–0.55)	0.35 (0.30–0.40)
Kappa	0.41 (0.39–0.43)	0.47 (0.45–0.50)	0.47 (0.45–0.50)	0.42 (0.36–0.47)

LOS summary includes historical LOS mean and SD calculated via similarity cascade.
G-mean indicates geometric mean of classwise sensitivity for length of stay model or geometric mean of sensitivity and specificity for discharge disposition model; IBA, index of balanced accuracy.

cases that had their EDD posted at least 1 day before the day of surgery were 37.9% and 60.2%, respectively. We also evaluated Poisson and NB regression models and observed lower overall accuracy compared with our LOS model (Supplemental Figure 3, see http://links.lww.com/AOSO/A455).

DD Model Performance

A binary classification model was created with the same feature set as the third LOS model to predict home versus nonhome discharge destinations. Figures 5A and 5B illustrate the receiver operating characteristics and PR curves. This model achieved an AUC of 0.88 (95% CI, 0.85–0.90) and an average precision of 0.56 (95% CI, 0.50–0.61). Other additional performance metrics are shown in Table 1.

DISCUSSION

Most prior studies were service- or specialty-specific, which limits their generalizability indicating a need for a universal framework to use more routinely collected data to predict LOS.[2,14,15,22,34–37] We report one of the first models, which predicts postsurgical LOS for adult elective inpatient cases across multiple services using only the data available at the time of case posting. Although surgeons can provide their estimated postoperative LOS during surgical case posting as the preliminary EDD, this feature has not yet been fully integrated into our systems planning for bed capacity management. This is partly due to current resource limitations, as well as potential inconsistencies in the initial estimations. Since these estimations are usually entered by various nonclinical users during case posting, their

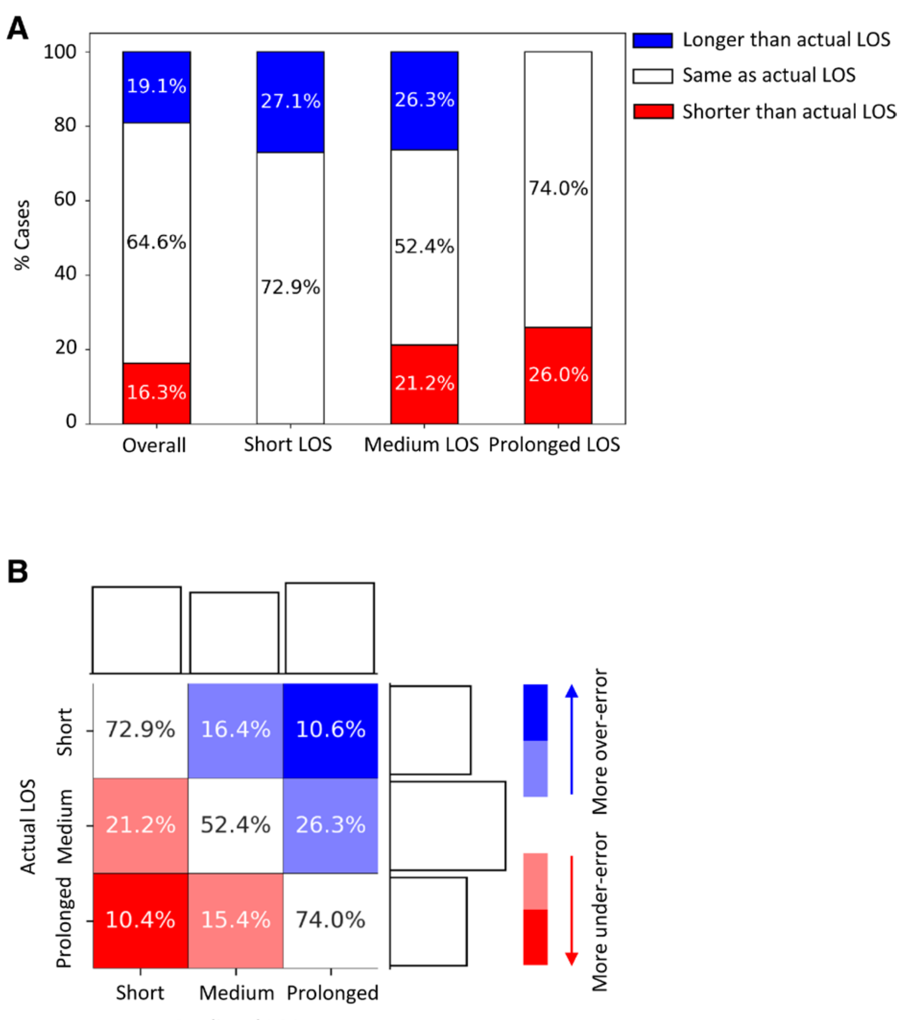

FIGURE 4. Breakdown of underprediction and overprediction errors. A, Under and over error in LOS prediction in all cases, as well as short, medium, and prolonged LOSs shown in red and blue, respectively. B, Heatmap and distribution of actual and predicted LOS classes and underprediction and overprediction errors. The numbers in each block represent the percentage of actual LOS classes that were predicted as short, medium, and prolonged LOSs. The bar charts on the *x* and *y* axes are the distribution of predicted and actual LOS classes, respectively.

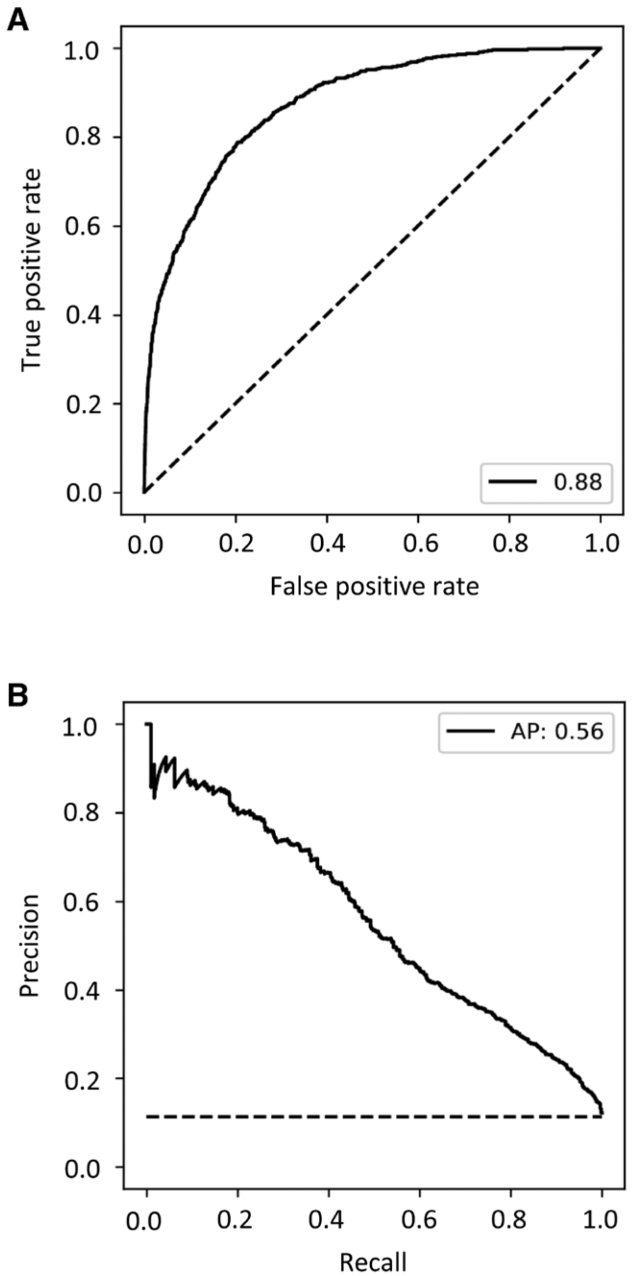

FIGURE 5. Discharge disposition model performance. A, Receiver operating characteristic curve (AUC, 0.88 [95% CI, 0.85–0.90]). B, PR curve with the dashed line represents the prevalence of cases discharged to nonhome facilities at 11.3% (average precision [AP], 0.56 [95% CI, 0.50–0.61]).

overall accuracy can also be quite low. We showed that EDD was less accurate than our LOS model and not available before admission for the majority of the patients; however, consistently available case posting data are sufficient to predict LOS as 3 operationally relevant categories: short, medium, and prolonged stays. We found that converting CPTs to work RVU and adding the historical mean and SD of LOS calculated by the similarity cascade improved our LOS prediction. Predicting LOS is critically important to a surgical system. Indeed, identifying patients with prolonged LOS is one of the top priorities in our institution as they account for a large number of bed days in our hospitals. Identifying short LOS can facilitate surgical case scheduling. For example, one may want to avoid scheduling such cases on Fridays so that those patients could be discharged during the weekdays without weekend-related delays. Furthermore, patients who were discharged to nonhome facilities historically stayed longer as they required specific coordination and planning because the uncertainty of nonhome discharge coordination further prolongs LOS. Existing research highlights the impact of nonclinical factors such as nonhome DD on discharge delays and extended LOS.[38,39] Hwabejire et al[40] found that nonclinical factors including rehabilitation facility placement, hospital operational delays, and payer authorization caused 80% of discharge delays in their trauma patients. More communication and coordination among the care teams have been shown to be significantly associated with lower LOS.[14] Thus, preadmission LOS and DD predictions as preliminary EDD and DD allow case managers to proactively engage with relevant healthcare providers, potentially streamlining discharge planning and improving patient flow.

Therefore, we also created a model to predict home versus nonhome discharge destinations using the same approach as the LOS model to provide case managers and clinicians with a prediction that a patient may require coordination for a nonhome discharge.

We have previously reported a production framework for the surgical case length prediction at case posting time,[24] and we are working to leverage the same pipeline and deploy and silently evaluate our LOS and DD models. Our general approach facilitates incorporating posted data and integrating the predicted LOS and DD into our EHR system for earlier decision-making and discharge planning. We also believe that LOS and DD predictions better inform the case managers and hospital administrators when there are risks of overutilization. Similar to diagnostic-related groups, which helped to quantify and standardize hospital resource utilization and reduce average LOS,[14,41] categories such as short, medium, and prolonged LOSs could also be used to quantify and standardize resource utilization.

There are several limitations to our study. First, LOS may also be affected by other factors unrelated to the procedure such as comorbidities and socioeconomic status and support as disease- and nondisease-related factors, respectively. Often studies attempting to predict LOS and DD utilize patient zip code as a proxy for several socioeconomic factors. We assessed the impact of zip code on the predictive performance of our models and found that zip code did not improve the model's performance (results not shown). Future models may also include intraoperative and postoperative clinical features to provide updated LOS prediction throughout the patient's stay and at critical points along the continuum of care. A deeper study is required to identify and quantify the risk factors and reasons for longer LOS such as comorbidities, socioeconomic factors, pain, physical therapy, and placement[1,11] that are associated with prolonged LOS and may help to change practice and standard procedures and reduce LOS. Gabriel et al[1] highlighted the inherent challenge of ideally predicting the exact LOS given the limitations of preadmission data and compared Poisson regression to a dichotomized approach. They suggested that with models such as Poisson regression, achieving highly accurate LOS predictions in days would necessitate incorporating a wider range of granular data points. Focusing on consistently available preoperative posted data, our model provides a practical tool for hospital administrators to initiate discharge planning. Second, LOS prediction should reasonably become more accurate during later stages of stay using new and updated data; however, there are tradeoffs between the accuracy and timing of LOS prediction. Although accurate prediction of LOS and DD is important, the timing of such predictions is equally important because earlier prediction provides more opportunity for resource planning and patient family awareness and counseling that could directly affect care outcomes and increase patient satisfaction.[42–44] Predicting LOS before admission could help hospitals' admission and bed management to monitor the quality of inpatient care, identify abnormalities or potential risks that require additional resources, and anticipate any bottleneck in bed availability before patients are admitted.[14] Lequertier et al[18] systematically reviewed LOS literature showing only 6% of reviewed studies predicted LOS before patient admission indicating the challenge and the need to predict LOS before admission. It has been shown that clinical factors were not among the top predictors, and demographics, service, and procedural variables were most predictive of LOS indicating that data available at case posting time could be leveraged to create a predictive LOS model.[27] It is important to appreciate the inherent limitation of case posting data as elective cases could be posted days or even months before the procedure. LOS prediction is quite a complex and challenging task considering a variety of players such as clinical, operational, and socioeconomic factors, and using the restricted case posting data to make such prediction renders it even more difficult. We did not include other data such as previous admissions, preexisting comorbidities, and insurance status, which may not be complete or reliably available at case posting time. These data are available at our hospital but require thorough work to collect and record them in a suitable format and expert input on data quality and relevance.[2] Our model potentially sacrifices accuracy by using the case posting data in exchange for the clinical and operational utility of the predicted LOS before the time of admission. We also decided to use the model with applied sample weight to achieve higher accuracy in short and prolonged classes at the expense of reduced accuracy for medium stays, which may not be operationally as helpful as short and prolonged LOS prediction. Third, there is no gold standard to predict LOS and benchmark against our model.[22] Future evaluation is required if such a standard model is available. Fourth, our model was created using only DUHS data and should be evaluated using other health systems' data. We acknowledge that our LOS classes may not generalize to other institutions. However, we hypothesize that our approach should perform comparably the same in other institutions using similar data, which would make the models replicable and adoptable at other health systems. Fifth, this study was a retrospective analysis. The models need to be implemented, and operational variations due to the use of the models should be evaluated. Sixth, there is an inherent limitation in CPT codes of some procedures such as cervical, thoracic, and lumbar spinal constructs that render separation of cases and, therefore, their LOS impossible using their CPT codes.

CONCLUSIONS

We developed a machine learning model to predict, at the time of case posting, the postsurgical LOS for adult elective cases in multiple services and separate them into 3 categories, namely, short, medium, and prolonged LOSs. Adding the work RVU and average LOS, calculated by a similarity cascade, resulted in about 73% to 74% accuracy for the short and prolonged groups, the 2 most operationally important groups. Using the same approach, we created another model to predict DD, which achieved an AUC of 0.88. These models could be leveraged to assist case managers and hospital administrators in initiating preemptive measures and earlier discharge planning to reduce LOS for surgical patients.

REFERENCES

1. Gabriel RA, Sharma BS, Doan CN, et al. A predictive model for determining patients not requiring prolonged hospital length of stay after elective primary total hip arthroplasty. *Anesth Analg*. 2019;129:43–50.
2. Stone K, Zwiggelaar R, Jones P, et al. A systematic review of the prediction of hospital length of stay: towards a unified framework. *PLOS Digit Health*. 2022;1:e0000017.
3. Alsinglawi B, Alshari O, Alorjani M, et al. An explainable machine learning framework for lung cancer hospital length of stay prediction. *Sci Rep*. 2022;12:607.
4. Liu YX, Qin SW. An interpretable machine learning approach for predicting hospital length of stay and readmission. Advanced Data Mining and Applications, Adma 2021, Pt I 2022; 13087:73–85.
5. Baek H, Cho M, Kim S, et al. Analysis of length of hospital stay using electronic health records: a statistical and data mining approach. *PLoS One*. 2018;13:e0195901.
6. El-Bouri R, Taylor T, Youssef A, et al. Machine learning in patient flow: a review. *Prog Biomed Eng (Bristol)*. 2021;3:022002.
7. Jaotombo F, Pauly V, Fond G, et al. Machine-learning prediction for hospital length of stay using a French medico-administrative database. *J Mark Access Health Policy*. 2023;11:2149318.
8. Muhlestein WE, Akagi DS, Davies JM, et al. Predicting inpatient length of stay after brain tumor surgery: developing machine learning ensembles to improve predictive performance. *Neurosurgery*. 2019;85: 384–393.
9. Barnes S, Hamrock E, Toerper M, et al. Real-time prediction of inpatient length of stay for discharge prioritization. *J Am Med Inform Assoc*. 2016;23(e1):e2–e10.
10. Stonko DP, Weller JH, Gonzalez Salazar AJ, et al. A pilot machine learning study using trauma admission data to identify risk for high length of stay. *Surg Innov*. 2022;30:356–365.
11. Masum S, Hopgood A, Stefan S, et al. Data analytics and artificial intelligence in predicting length of stay, readmission, and mortality: a population-based study of surgical management of colorectal cancer. *Discov Oncol*. 2022;13:11.
12. Goshtasbi K, Yasaka TM, Zandi-Toghani M, et al. Machine learning models to predict length of stay and discharge destination in complex head and neck surgery. *Head Neck*. 2021;43:788–797.
13. Awad A, Bader-El-Den M, McNicholas J. Patient length of stay and mortality prediction: a survey. *Health Serv Manage Res*. 2017;30:105–120.
14. Tsai PF, Chen PC, Chen YY, et al. Length of hospital stay prediction at the admission stage for cardiology patients using artificial neural network. *J Healthc Eng*. 2016;2016.
15. Li H, Jiao J, Zhang S, et al. Construction and comparison of predictive models for length of stay after total knee arthroplasty: regression model and machine learning analysis based on 1,826 cases in a single Singapore center. *J Knee Surg*. 2022;35:7–14.
16. Morris RS, Tignanelli CJ, deRoon-Cassini T, et al. Improved prediction of older adult discharge after trauma using a novel machine learning paradigm. *J Surg Res*. 2022;270:39–48.
17. Arora A, Lituiev D, Jain D, et al. Predictive models for length of stay and discharge disposition in elective spine surgery: development, validation, and comparison to the ACS NSQIP risk calculator. *Spine (Phila Pa 1976)*. 2023;48:E1–E13.
18. Lequertier V, Wang T, Fondrevelle J, et al. Hospital length of stay prediction methods: a systematic review. *Med Care*. 2021;59:929–938.
19. Kramer AA, Zimmerman JE. A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay. *BMC Med Inform Decis Mak*. 2010;10:27.
20. Medeiros NB, Fogliatto FS, Rocha MK, et al. Forecasting the length-of-stay of pediatric patients in hospitals: a scoping review. *BMC Health Serv Res*. 2021;21:938.
21. Symum H, Zayas-Castro JL. Prediction of chronic disease-related inpatient prolonged length of stay using machine learning algorithms. *Healthc Inform Res*. 2020;26:20–33.
22. Austin PC, Rothwell DM, Tu JV. A comparison of statistical modeling strategies for analyzing length of stay after CABG surgery. *Health Serv Outcomes Res Methodol*. 2002;3:107–133.
23. Castineira D, Schlosser KR, Geva A, et al. Adding continuous vital sign information to static clinical data improves the prediction of length of stay after intubation: a data-driven machine learning approach. *Respir Care*. 2020;65:1367–1377.
24. Zaribafzadeh H, Webster WL, Vail CJ, et al. Development, deployment, and implementation of a machine learning surgical case length prediction model and prospective evaluation. *Ann Surg*. 2023;278: 890–895.
25. Deng Y, Liu S, Wang Z, et al. Explainable time-series deep learning models for the prediction of mortality, prolonged length of stay and 30-day readmission in intensive care patients. *Front Med (Lausanne)*. 2022;9:933037.
26. Laios A, De Freitas DLD, Saalmink G, et al. Stratification of length of stay prediction following surgical cytoreduction in advanced high-grade serous ovarian cancer patients using artificial intelligence; the Leeds L-AI-OS score. *Curr Oncol*. 2022;29:9088–9104.
27. Goldstein BA, Cerullo M, Krishnamoorthy V, et al. Development and performance of a clinical decision support tool to inform resource utilization for elective operations. *JAMA Netw Open*. 2020;3:e2023547.

28. Li Y, Wang H, Luo Y. Improving fairness in the prediction of heart failure length of stay and mortality by integrating social determinants of health. *Circ Heart Fail*. 2022;15:e009473.
29. Garside N, Zaribafzadeh H, Henao R, et al. CPT to RVU conversion improves model performance in the prediction of surgical case length. *Sci Rep*. 2021;11:14169.
30. Zhang AS, Veeramani A, Quinn MS, et al. Machine learning prediction of length of stay in adult spinal deformity patients undergoing posterior spine fusion surgery. *J Clin Med*. 2021;10:4074.
31. Wang K, Yan LZ, Li WZ, et al. Comparison of four machine learning techniques for prediction of intensive care unit length of stay in heart transplantation patients. *Front Cardiovasc Med*. 2022;9:863642.
32. Chen TQ, Guestrin C. XGBoost: A Scalable Tree Boosting System. Kdd'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016:785–794.
33. Fernandez GA, Vatcheva KP. A comparison of statistical methods for modeling count data with an application to hospital length of stay. *BMC Med Res Methodol*. 2022;22:211.
34. Abbas A, Mosseri J, Lex JR, et al. Machine learning using preoperative patient factors can predict duration of surgery and length of stay for total knee arthroplasty. *Int J Med Inform*. 2022;158:104670.
35. Zeng X. Length of stay prediction model of indoor patients based on light gradient boosting machine. *Comput Intell Neurosci*. 2022;2022:9517029.
36. Zhong H, Wang B, Wang D, et al. The application of machine learning algorithms in predicting the length of stay following femoral neck fracture. *Int J Med Inform*. 2021;155:104572.
37. Ramkumar PN, Navarro SM, Haeberle HS, et al. Development and validation of a machine learning algorithm after primary total hip arthroplasty: applications to length of stay and payment models. *J Arthroplasty*. 2019;34:632–637.
38. Flanigan TL, Kiskaddon EM, Rogozinski JA, et al. Predictive factors of extended length of hospital stay following total joint arthroplasty in a veterans affairs hospital population. *J Arthroplasty*. 2021;36: 1527–1532.
39. Watkins JR, Soto JR, Bankhead-Kendall B, et al. What's the hold up? Factors contributing to delays in discharge of trauma patients after medical clearance. *Am J Surg*. 2014;208:969–73; discussion 972.
40. Hwabejire JO, Kaafarani HMA, Imam AM, et al. Excessively long hospital stays after trauma are not related to the severity of illness. *JAMA Surg*. 2013;148:956–961.
41. Lee S, Reddy Mudireddy A, Kumar Pasupula D, et al. Novel machine learning approach to predict and personalize length of stay for patients admitted with syncope from the emergency department. *J Pers Med*. 2022;13:7.
42. Yang CS, Wei CP, Yuan CC, et al. Predicting the length of hospital stay of burn patients: comparisons of prediction accuracy among different clinical stages. *Decis Support Syst*. 2010;50:325–335.
43. Hu Z, Qiu H, Wang L, et al. Network analytics and machine learning for predicting length of stay in elderly patients with chronic diseases at point of admission. *BMC Med Inform Decis Mak*. 2022;22:62.
44. Bednarski BP, Singh AD, Zhang W, et al. Temporal convolutional networks and data rebalancing for clinical length of stay and mortality prediction. *Sci Rep*. 2022;12:21247.