Explainable Machine Learning Approach to Prediction of Prolonged Intensive Care Unit Stay in Adult Spinal Deformity Patients: Machine Learning Outperforms Logistic Regression

Bashar Zaidat, BS[1] ⓘ, Mark Kurapatti, BS[1], Jonathan S. Gal, MD[1], Samuel K. Cho, MD[1] ⓘ, and Jun S. Kim, MD[1]

Abstract

Study Design: Retrospective cohort study.

Objectives: Prolonged ICU stay is a driver of higher costs and inferior outcomes in Adult Spinal Deformity (ASD) patients. Machine learning (ML) models have recently been seen as a viable method of predicting pre-operative risk but are often 'black boxes' that do not fully explain the decision-making process. This study aims to demonstrate ML can achieve similar or greater predictive power as traditional statistical methods and follows traditional clinical decision-making processes.

Methods: Five ML models (Decision Tree, Random Forest, Support Vector Classifier, GradBoost, and a CNN) were trained on data collected from a large urban academic center to predict whether prolonged ICU stay would be required post-operatively. 535 patients who underwent posterior fusion or combined fusion for treatment of ASD were included in each model with a 70-20-10 train-test-validation split. Further analysis was performed using Shapley Additive Explanation (SHAP) values to provide insight into each model's decision-making process.

Results: The model's Area Under the Receiver Operating Curve (AUROC) ranged from 0.67 to 0.83. The Random Forest model achieved the highest score. The model considered length of surgery, complications, and estimated blood loss to be the greatest predictors of prolonged ICU stay based on SHAP values.

Conclusions: We developed a ML model that was able to predict whether prolonged ICU stay was required in ASD patients. Further SHAP analysis demonstrated our model aligned with traditional clinical thinking. Thus, ML models have strong potential to assist with risk stratification and more effective and cost-efficient care.

Keywords

adult spinal deformity, adult spinal deformity, machine learning, artificial intelligence, intensive care unit stay, shapley additive explanation

[1] Department of Orthopaedics, The Mount Sinai Hospital, New York, NY, USA

Corresponding Author:
Jun S. Kim, MD, Department of Orthopaedic Surgery, Icahn School of Medicine at Mount Sinai, The Mount Sinai Hospital, 1468 Madison Ave, New York, NY 10029, USA.
Email: Jun.Kim@mountsinai.org

Introduction

As health care costs continue to rise in the US,[1] cost optimization has become increasingly essential. The treatment of adult spinal deformity (ASD) is especially costly, with approximately US$290 billion spent on fusion-related spine procedures in the United States between 2000 and 2010.[2] Further, the annual number of ASD surgeries appears to be increasing.[3] A major contributor to ASD surgery cost is prolonged postoperative hospital length of stay (LOS), which often includes expensive intensive care unit (ICU) stays and can exceed US$38,000.[4] Prior studies have sought to assist in cost optimization by using regression modeling to predict the impact of various variables on likelihood of hospital LOS following ASD surgery.[5,6] However, regression modeling is often an oversimplification failing to account for the interactions between variables that occur in real situations.

As such, there has been recent interest in the use of machine learning (ML) to predict discharge following spine surgery more accurately. However, ML models often derive very complex algorithms based on data resulting in 'black boxes' where the decision-making process is unknown or cannot be easily explained. Recent spine research has demonstrated the addition of Shapley Additive Explanation (SHAP) values can clearly explain the model's decision-making process while maintaining high predictive power. For example, three prior studies utilized ML to examine surgical spine cohorts and consistently determined using ML models with SHAP values were highly accurate in predicting postoperative prolonged stay or non-home discharge (NHD).[7-9] One study demonstrated the explainable "SHAP-derived features" model outperformed each individual feature domain model and had a comparable predictive capacity to the more complicated "full feature" model.[7] Importantly, these studies emphasized and explained risks of prolonged stay or NHD were multifactorial, with variables like intraoperative and sociodemographic factors having bidirectional influences on risk and other variables (such as older age and multiple fusion levels) having a compounded positive impact on risk.[7-9]

Prior studies have demonstrated the use of ML models with SHAP-derived features successfully predicts NHD or prolonged LOS while elucidating the effects of variable interactions in spine surgeries. However, to our knowledge, the current study is the first to utilize ML models to predict whether ASD patients will require prolonged ICU stay. We aim to apply five ML models with SHAP analysis to an ASD cohort to make predictions more in line with clinical reasoning in an effort to enhance cost-related efficiency, risk stratification, and care effectiveness. The main problem that this predictive model seeks to address is to serve as an example of a risk stratification model particularly for prolonged ICU stay in ASD patients.[10,11]

Methods

Patient Data

All data was retrospectively gathered from a large, tertiary, urban academic center. 535 patients who underwent posterior fusion or combined anterior and posterior fusion were isolated using Current Procedural Terminology (CPT) codes. An institutional dataset of spine deformity cases was created using CPT codes 22800, 22802, 22804, 22614, or 22634 where patients had 4 or more segments fused. Patients less than 18 years old were excluded and adult spinal deformity cases were confirmed via manual chart review. IRB approval was granted by the Icahn School of Medicine at Mount Sinai (IRB approval number: STUDY-17-00660). Request for waiver of informed consent was approved for this study.

Model Selection and Training

Five different machine learning models were tested and trained on the data with a 70-20-10 train-test-validation split: a Decision Tree, a Random Forest Classifier, a Support Vector Classifier, a Gradient Boost (GradBoost) model, and a CNN. These were compared to a logistic regression model. The demographic features included in the models were age, gender, ethnicity, BMI, insurance type, and ASA status. The perioperative features included in the models were length of surgery, estimated blood loss, segments operated, combined posterior/anterior approach, and incidence of complications.

Shapley Analysis

To achieve greater insight into the decision-making process of the best-performing model, a Shapley Additive Explanation (SHAP) analysis was performed. A SHAP analysis determines the effect of each feature input to the model on the outcome, in this case, the effect of each demographic and perioperative feature on predicting whether prolonged ICU stay was required for the patient. In addition to calculating the mean absolute SHAP values for each feature, beeswarm plots were generated to examine the magnitude and direction of impact of each feature and dependency plots were generated to examine the interaction effects between each pair of features. Furthermore, waterfall plots were generated for a sample of patients to investigate the model's decision-making process for individual patients. These plots depict the direction and magnitude of impact of each feature value on predicting the patient's outcome.

Results

Patient demographics are summarized in Table 1. Of the 535 patients included in this analysis who underwent posterior fusion or combined fusion for treatment of ASD, 110 (20.6%) required a prolonged ICU stay, defined as the top 20th

Table 1. Patient Demographics.

	Non-Prolonged ICU Stay (n = 425)	Prolonged ICU Stay (n = 110)	P-Value
Age, yr. (SD[a])	59.7 (12.8)	59.3 (15.4)	0.79
Length of surgery, min[b] (SD)	239.1 (101.9)	360.1 (153.7)	<0.001
Estimated blood loss, mL (SD)	494.5 (523.1)	1165.2 (1552.2)	<0.001
Sex, no. (%)			0.19
Female	206 (48.5)	45 (40.9)	
Male	219 (51.5)	65 (59.1)	
ASA status, no. (%)			0.32
1	11 (2.6)	5 (4.5)	
2	223 (52.5)	48 (43.6)	
3	174 (40.9)	51 (46.4)	
4	17 (4.0)	6 (5.5)	
Race, no. (%)			0.01
Asian	26 (6.1)	2 (1.8)	
Black	46 (10.8)	14 (12.7)	
White	246 (57.9)	79 (71.8)	
Other	107 (25.2)	15 (13.6)	
Segments, no. (SD)	5.21 (1.88)	5.79	0.007
Insurance, no. (%)			0.147
Medicaid	28 (6.6)	10 (9.1)	
Medicare	163 (38.4)	40 (36.4)	
Private	187 (44.0)	55 (50.0)	
Other	47 (11.1)	5 (4.5)	
BMI (SD)	29.83 (16.34)	31.59 (21.9)	0.351
Complications	117 (27.5)	64 (58.2)	<0.001
Combined posterior/Anterior approach	11 (2.6)	9 (8.2)	0.013

[a]SD = Standard Deviation;
[b]min = minute(s);
P-value α = 0.05.

percentile of the cohort with an average ICU stay of 2.3 days. Patients who required a prolonged ICU stay tended to have longer lengths of surgery, higher estimated blood loss, and more segments fused. The average length of surgery of the required prolonged ICU stay cohort was two hours longer than the non-prolonged ICU stay cohort, and the average estimated blood loss of the required prolonged ICU stay cohort was more than double the non-prolonged ICU stay cohort. The required prolonged ICU stay cohort was also more likely to experience complications compared to patients who did not require prolonged ICU stay, and the racial composition of the two cohorts was significantly different.

Comparison of Different Models

Model performances are summarized in Figure 1. The models' AUROC values ranged from 0.67 to 0.83. From lowest to highest AUROC, the Grid Search SVM model achieved an AUROC of 0.67, the Neural Net achieved 0.748, the Decision Tree Classifier achieved 0.78, the Gradient Boosting Classifier and the Logistic Regression models both achieved 0.79, and the Random Forest Classifier achieved the highest AUROC value of 0.83.

Most Important Features Based on SHAP Values

The SHAP analysis revealed which demographic, perioperative, and postoperative features had the largest relative impact on whether or not a patient would require a prolonged ICU stay. A SHAP value was calculated for each of the eleven features for each patient. For a binary prediction in our case – prolonged ICU stay required or not – the values take the form of a log odds, or the natural log of the probability of required prolonged ICU stay occurring. Averaging these (absolute) values for each feature across all patients gives the mean SHAP value, where larger values indicate the feature had a greater impact on prediction of prolonged ICU stay. Length of surgery had the largest impact on predicting prolonged ICU stay, followed by complications, estimated blood loss, BMI, Age, and ASA status, which all had a SHAP value greater than 0.5, while ethnicity, gender, number of segments, insurance, and type of approach did not have a significant impact on predicting prolonged ICU stay (Figure 2).

A beeswarm plot reveals the relationship between each feature and its impact on ICU stay by visualizing all SHAP values from the dataset. In Figure 3, the features along the
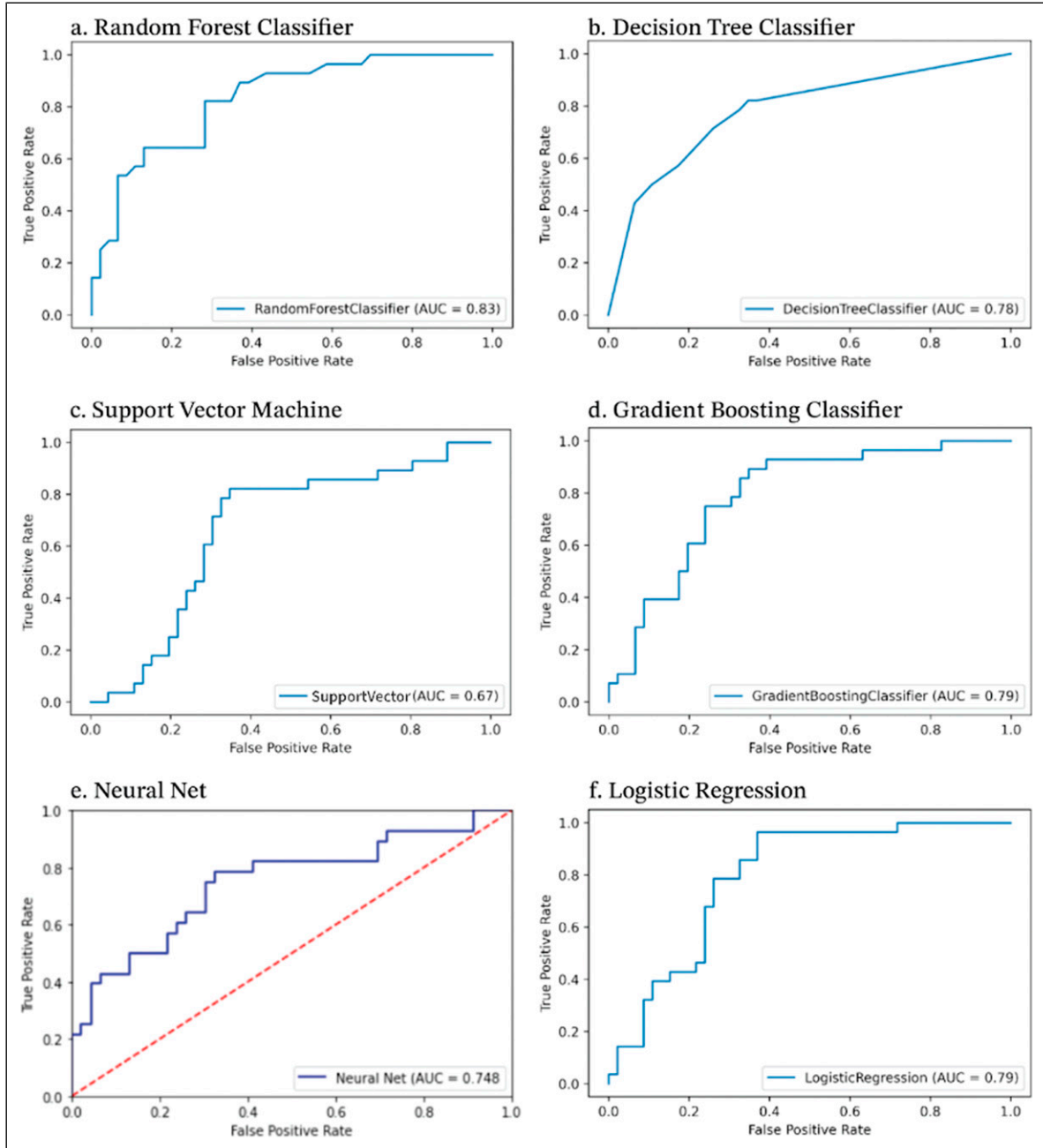
Figure 1. Model performance.

y-axis are ordered by their overall impact on ICU stay as in Figure 2, and the color of dots indicates whether the value of the feature is high (red) or low (blue). For binary features, such as female gender, red indicates the feature is true (female) or false (male). The *x*-axis shows the SHAP value, which indicates the impact of each feature value on predicting prolonged ICU stay. Longer length of surgery, presence of complications, and higher estimated blood loss were strong predictors of required prolonged ICU stay as shown by the prevalence of red dots in the long right-hand tail; conversely, shorter lengths of surgery, the absence of complications, and lower estimated blood loss had a smaller impact on predicting non-prolonged ICU stay, as shown by the short, blue, left-hand tails. The relationship between BMI and age with prolonged ICU stay prediction is unclear, as shown by the lack of an obvious division of red and blue values. Higher ASA status, white ethnicity, and male gender were also associated with prolonged ICU stay, and these three features were more balanced – ie, lower ASA status, non-white ethnicity, and female gender were associated with non-prolonged ICU stay with similar predictive power as their
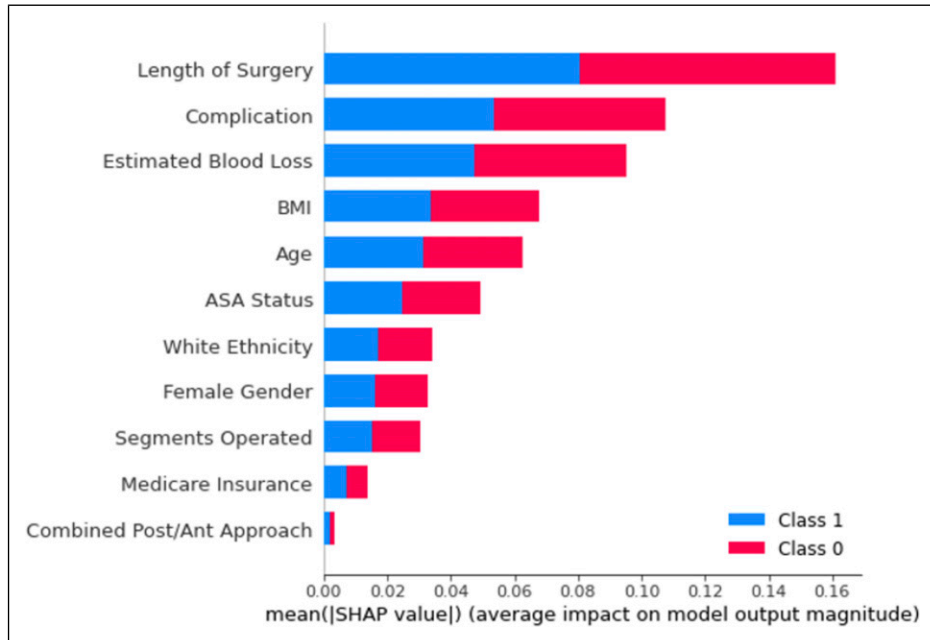
Figure 2. Random forest mean SHAP values.
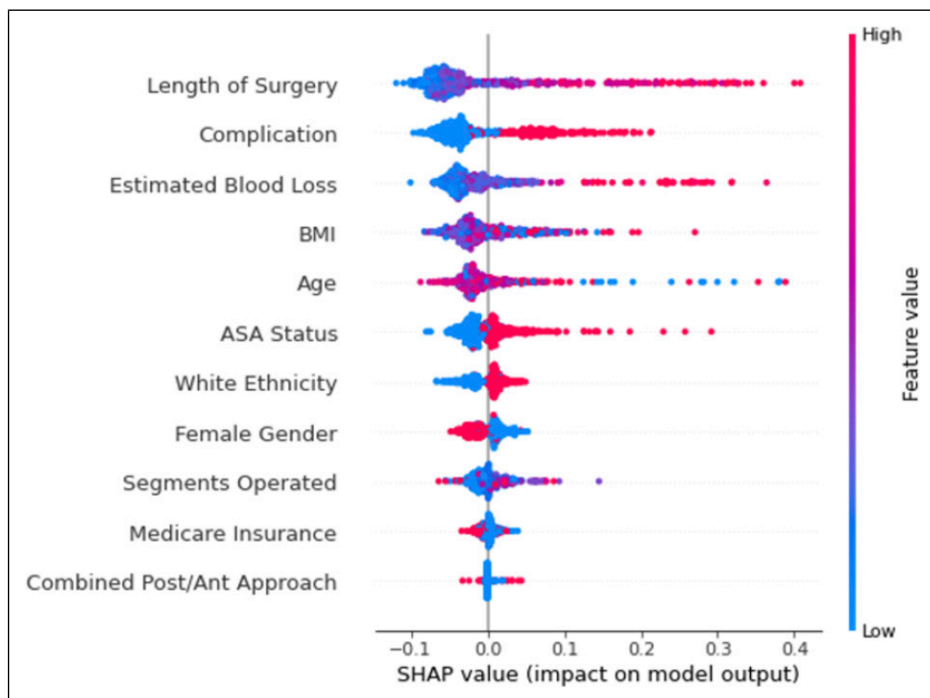
Figure 3. Random forest beeswarm plot.

complements. Finally, the number of segments fused, medicare insurance, and type of approach had little predictive power in prolonged ICU stay.

Waterfall plots can be made for each individual patient to demonstrate the predictive power of each feature on requiring prolonged ICU stay. E [f(x)] is the average predicted log odds of prolonged ICU stay across all patients, which in this analysis was 0.14, while f(x) is the predicted log odds. The horizontal bars then show how each feature value contributed to the deviance of the predicted log odds from the average. In Figure 4(A), f(x) = 1, meaning the model predicted that this specific patient
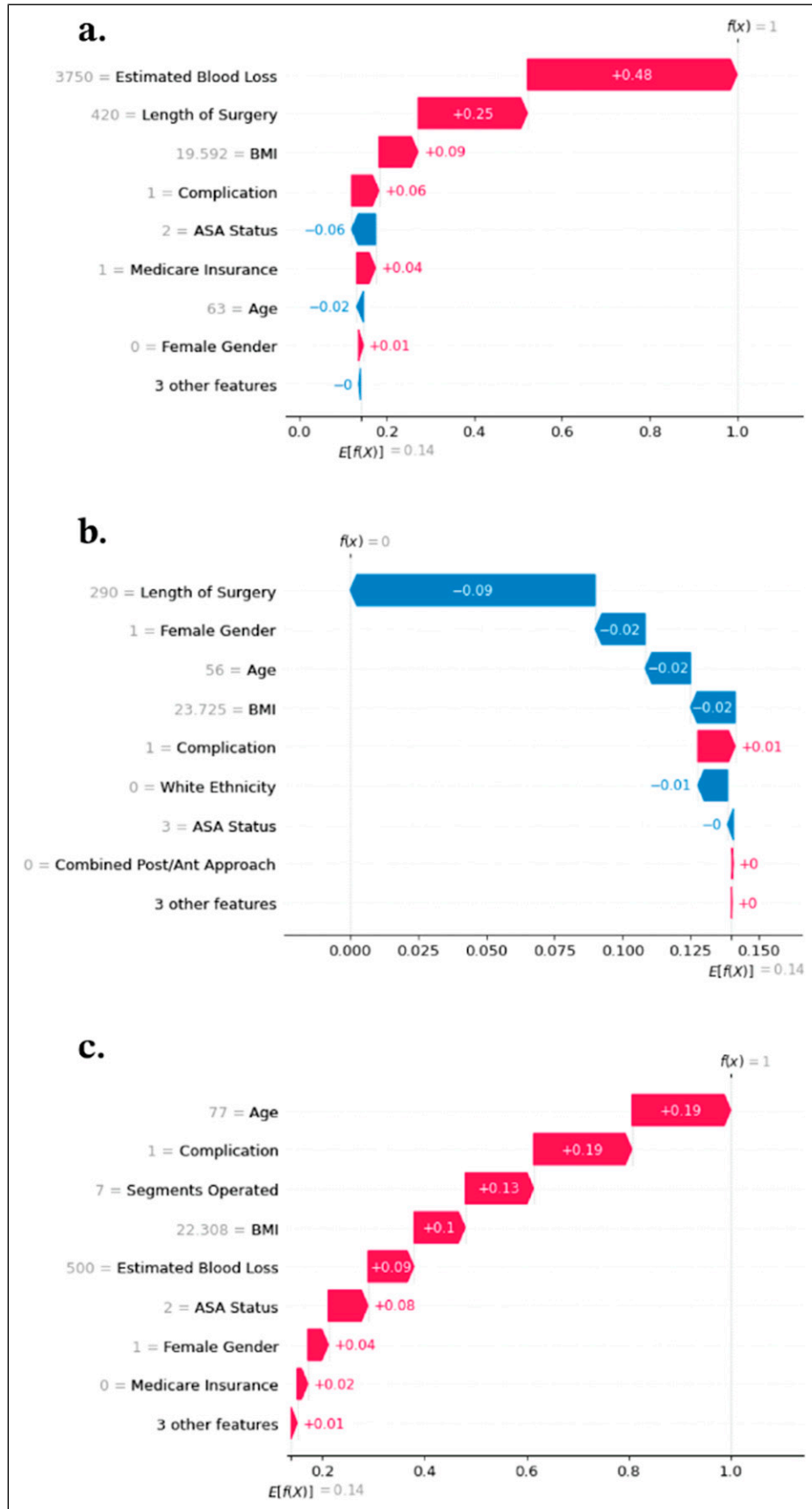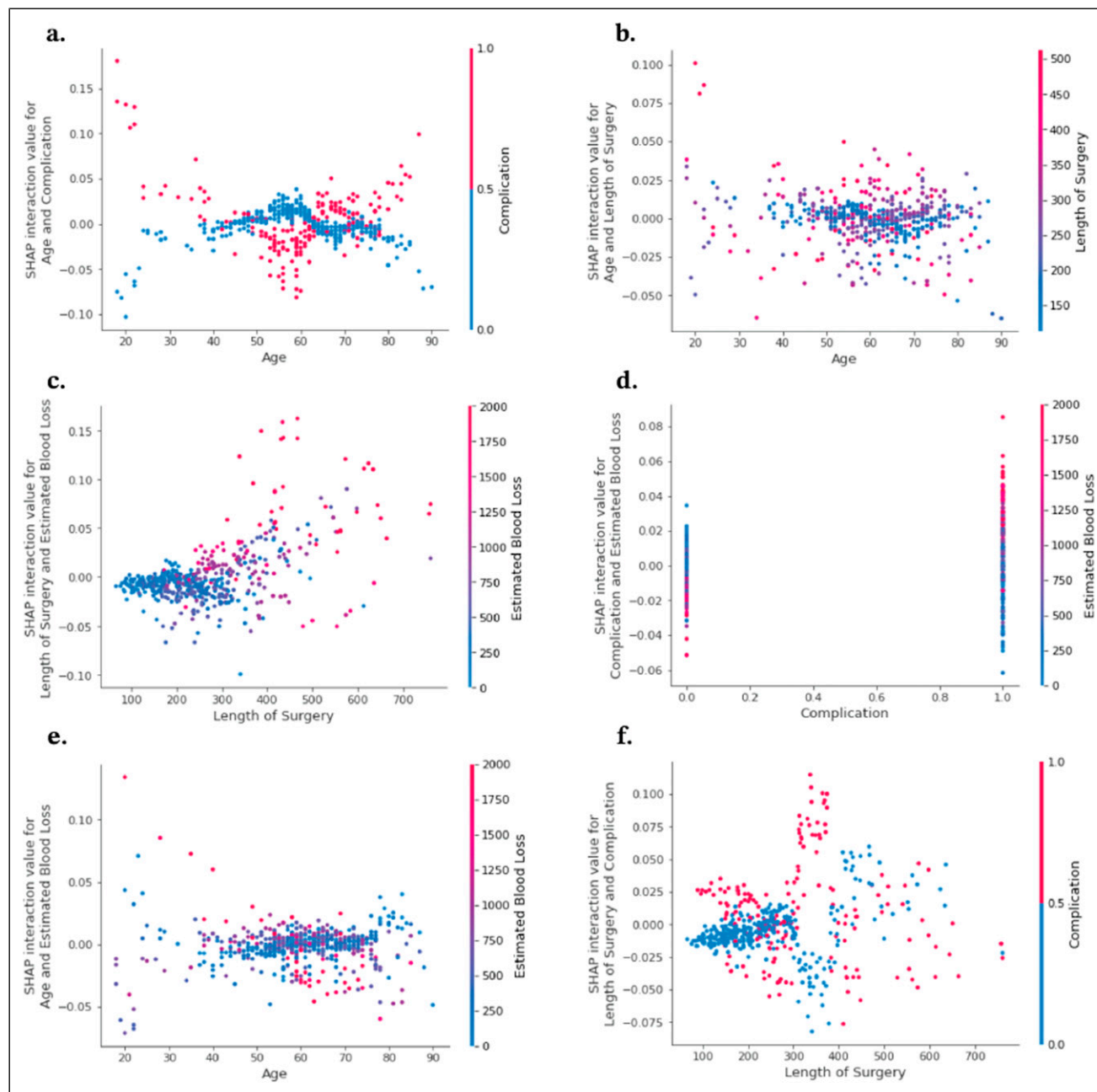
Figure 4. Sample waterfall plots for prediction of prolonged ICU stay.

Figure 5. Intervariable interactions for top performing variables.

was more likely to require prolonged ICU stay than the average patient. Estimated blood loss had the highest impact on this prediction, increasing the log odds by 0.48, followed by length of surgery, which increased the log odds by 0.25. BMI also had a small positive effect on increasing prolonged ICU stay probability, while the other eight features had smaller impacts. In Figure 4(B), we see a different patient for whom the model predicted prolonged ICU stay was much less likely compared to the cohort average, with length of surgery having the greatest negative impact on the log odds. Finally, Figure 4(C) shows a

patient for whom each feature increased the probability of required prolonged ICU stay.

Variable Interactions and Dependence

Dependency plots provide deeper information about how different features interact to influence the likelihood of prolonged ICU stay. A feature's effects on this prediction can then be broken into its main effect, described above, and its interaction effects. The dependency plots in Figure 5 show the interaction effects for six pairs of features. For example, when

a complication is present (red), there is a negative interaction effect with age until age = 60, after which there is a positive interaction effect, and the opposite relationship exists when a complication is not present (blue) (Figure 5(A)). The main effect of complications is positive, that is, the presence of a complication increases the probability that ICU is required (Figure 3). When the main and interaction effects align, the predictive power of those features is enhanced, and vice versa. So, the dependency plot shows that for patients under 60, the presence of complications did not have a strong effect on whether prolonged ICU stay would be required, while for patients over 60, the presence of complications did have a strong effect on prolonged ICU stay. The presence of complications also had a strong positive interaction with estimated blood loss, meaning high estimated blood loss was an even stronger predictor of prolonged ICU stay if the patient also experienced a complication (Figure 5(D)). No strong interaction effects were observed between length of surgery and age (Figure 5(B)), length of surgery and estimated blood loss (Figure 5(C)), or estimated blood loss and age (Figure 5(E)).

Discussion

This study serves to design a machine learning model to explain drivers of prolonged postoperative ICU stay and explore the model's decision-making using a novel analysis. The results demonstrate that length of surgery, complications, estimated blood loss, BMI, and age are predictors of prolonged ICU stay. Moreover, intervariable interaction analysis reveals varying additive effects between variables in the model. To the author's knowledge this is the first study to analyze drivers of prolonged ICU stay in adult spinal deformity patients.

Prior efforts to improve ASD surgery have primarily focused on minimizing morbidity and mortality, and more recently improving quality of life.[12,13] However, the financial burden on the patient is another crucial variable to consider. Many ASD patients enter the ICU postoperatively for pain management, blood volume and pressure management, and airway precaution.[14,15] Unfortunately, the cost of ASD surgery is higher in part due to higher costs of ICU stay.[16,17] A cost analysis of two surgical ICUs estimates the median cost of postoperative ICU stay to be US\$2636 (interquartile range: US\$1834–US\$4282) for the first day alone.[18] Thus, the length of postoperative ICU stay is a critical factor in ASD surgery as it reflects the duality of the medical and financial burden on the recovering ASD surgical patient. Rafael De la Garza-Ramos et al examined predictors of ICU-level complications in 826 ASD long-segment fusion patients using multivariate regression. Older age, diabetes, being dependent on others for activities of daily living, or having combined approaches ($P = 0.044$, $P = 0.048$, $P = 0.004$, $P = 0.023$, respectively) were risk factors for postoperative ICU-level complications, which occurred in 5.4% of patients across the cohort.[19] Similarly, Amin et al[20] reviewed 244 ASD surgical patients and

determined class II/III obesity (BMI ≥35, n = 17) were at significantly higher risk for ICU stay (>2 days, $P = 0.001$) and high episode-of-care costs ($P = 0.013$). Higher costs-of-care and length of ICU stay are likely closely interrelated and thus the present study focuses on delineating predictors of prolonged ICU stay using an explainable ML approach.

Traditional statistical models rely on top-down approaches to modeling the data. For example, an assumption of linear regression models is that the model output and input is characterized by a linear relationship. Machine learning models present a unique opportunity in that they do not rely on the statistician to come up with a model they believe best fits the data. Instead these models offer the distinct advantage of fitting themselves to the data in a "bottom-up" approach, enabling them to be flexible and handle more elaborate data.[21] SHAP analysis further benefits these models as it assists in explainability, and may help with the hesitancy surrounding application of ML models in the clinical setting.

Recent spine studies using ML demonstrate its clear benefits over the traditional logistic regression approach in its ability to account for and explain inter-variable interactions. For example, Martini et al examined a cohort of 11,150 spine patients, 29.7% who had prolonged LOS using ML. They demonstrated the explainable "SHAP-derived features" model (C-statistic [C] = 0.87) outperformed each individual feature domain model ("demographic": C = 0.77; "perioperative": C = 0.84; "postoperative": C = 0.72) and had a comparable predictive capacity to the more complicated "full feature" model (C = 0.89).[7] Importantly, this study emphasized prolonged stay risk was multifactorial, with intraoperative and sociodemographic factors having bidirectional influences on risk.[7] This aligns with another study by Valliani et al[8] which used SHAP values on ML prediction of NHD following 492,312 thoracolumbar spine surgery cases, in which the model achieved an area under the receiver operating characteristic curve (AUROC) of 0.77.[8] Age, total Elixhauser comorbidities, Medicare insurance, weighted Elixhauser score, and female sex were among the most important predictors of NHD. Furthermore, a 2023 study on ML prediction of NHD following 2227 anterior cervical discectomy and fusion (ACDF) cases also achieved a high AUROC of 0.83.[9] ASA scores had positive interaction effects with female sex, levels fused and body-mass index (BMI), and female patients over 65 with greater fusion levels were more likely to undergo NHD.[9] The issues with ML models lie in their lack of explainability, as they are often described as "black-box" models. Thus, we expand on the growing literature by including SHAP analysis.

Our overall complication rate was 33.8%, which is well-aligned with complication rates in prior ASD surgery studies with rates in the range of approximately 27%–63%.[22-25] At the core of our analysis was the use of SHAP values for important feature identification, which revealed length of surgery had the largest impact on predicting prolonged ICU stay, followed by complications, estimated blood loss, BMI, Age, and ASA status. While no prior ML modeling studies to our knowledge

have analyzed predictors of postoperative ICU stay duration, these results align with a prior univariate analysis of 127 spinal fusions patients with cerebral palsy (CP), which concluded having at least one perioperative complication more than doubled ICU stay duration (7.8 vs 3.2 days, $P < 0.05$).[26] Furthermore, multivariate regression analysis revealed an increased estimated blood loss was independently associated with a major perioperative complication ($P < 0.05$). Our results on blood loss as an important predictor of ICU stay duration also aligns with a prior linear regression study on 103 spine surgery patients which found intraoperative fluid management (including total blood loss) was correlated with prolonged (>1 day) ICU stay.[27] While comparison to the literature is limited by lack of focus on prior ML application to ICU stay duration, the use of SHAP analysis in the current study uniquely enabled more sophisticated delineation of intervariable interactions in relation to risk of prolonged ICU stay.

In addition, the current study's five ML models achieved relatively high AUROCs (0.67 to 0.83) and the best performing models outperformed logistic regression. This extends prior research on the superiority of ML in the prediction of ASD correction outcomes.[7-9,28] For example, a prior ML study on 4073 ASD patients determined artificial neural networks (ANN) and logistic regression significantly outperformed ASA scoring in predicting every examined postoperative complication (cardiac complications, wound complications, venous thromboembolism (VTE), and mortality).[28] Further, the ML algorithms outperformed logistic regression in predicting individual risk for all complications except VTE.[28]

Traditional analysis highlights the interactions of single variables, but less is understood of the interaction between these variables. SHAP intervariable interaction analysis reveals key insights into drivers of prolonged ICU stay as well. The results demonstrate an additive effect when complications are associated with higher estimated blood loss (Figure 5(D)). Previous studies have demonstrated that higher estimated blood loss is a predictor of longer ICU stay in patients undergoing spine surgery.[27] This interaction suggests that complications that are associated with greater intraoperative bleeding may be the complications driving longer ICU stays. Another interesting interaction was between age and complications. The results showed that either the absence or presence of complication was more important for the model's decision making at younger ages and older ages, but less impactful for middle-aged patients (Figure 5(A)). Age and length of surgery seem to have little interaction effects and thus suggest that their effects function separately, as did age and estimated blood loss. Better understanding of these variable interactions will assist surgeons in developing more robust risk profiles for prolonged ICU stay.

Specifically, the results show that the presence of a complication increased the likelihood of prolonged ICU stay for patients less than 40 years old and more than 60 years old,

while this effect was much smaller for middle aged patients. Similarly, the absence of a complication decreased the likelihood of prolonged ICU stay for patients less than 40 years old and more than 60 years old. Dependency plots also revealed that estimated blood loss had a positive interaction effect with length of surgery but only after length of surgery exceeded around 250 min. This suggests a combined effect where both high blood loss and prolonged surgical time significantly increase the risk, or more likely that there is a synergistic interaction between longer operative times and higher estimated blood loss on prolonged ICU stay. Finally, complications increased the SHAP values for estimated blood loss, indicating that complications amplify the effect of blood loss on the risk of prolonged ICU stay as well.

This study is valuable in demonstrating the successful application of ML models to illuminate predictors of prolonged ICU following ASD surgery. Prior research suggests ICU stay is an understudied metric that can be a large contributor to both high costs of surgery and an indirect indicator of suboptimal perioperative recovery.[16,17] Such results may be useful in tailoring risk stratification and highlight the importance of perioperative care programs. For example, Dagal et al[29] described the one-year-long implementation of an institutional enhanced perioperative care (EPOC) program tailored to timely and cost-effective care for spine surgery patients and reported significant decreases in postoperative ICU admissions and overall costs. For those who are admitted to the ICU postoperatively, our results suggest further emphasis on resource optimization to reduce surgical time, complications, and blood loss may be especially useful in decreasing the duration of ICU stay, overall costs, and optimizing recovery. This study paves the way for further exploration of the role of ML in ASD surgery recovery including the use of new ML models and examination of additional predictive variables and/or indicators of suboptimal recovery.

Currently, machine learning models are being used to develop a number of web-based risk stratification calculators.[30-32] These calculators based on these algorithms allow health care providers to quantify what the most contributory risk factors are for a given individual patient. Using these calculators, surgeons can isolate variables that are affecting outcomes, such as prolonged ICU stay, and effectively address them to the benefit of their patients. For example, similar machine learning algorithms were used to stratify patients who were at higher risk of venous thromboembolism after posterior lumbar fusion.[33] Using the results of our study, ML models such as the Random Forest Classifier can be applied to larger sample sizes, such that patients can be better fit by demographic variables to a specific risk profile. This paves the way for more robust ML calculators that surgeons and their teams can use for optimal resource allocation given an individual patient's highest contributory risk factors, and better improve preoperative and postoperative planning for both patients and providers.

Clinical Significance

Medical and research institutions around the world are now utilizing machine learning models to aid in risk stratification and postoperative management for patients undergoing a surgical operation or other medical interventions. In this way, the utilization of ML is clinically important on a global scale to patient care and can improve patient outcomes. For example, Shah et al[34] developed a ML model to predict major complications and readmissions after lumbar spinal fusion. Notably, this model outperformed traditional logistic regression. Unlike conventional regression techniques, ML models can detect complex nonlinear or logistic relationships as well as better explain factor-factor interactions. ML models have the unique ability to delineate and account for variables with bidirectional influences on risk as well as other variables that may have a compounded impact on risk. These ML models are increasingly being employed in the spine surgery field to predict outcomes including discharge disposition, surgical site infection, vertebral compression fracture, and mortality in metastatic disease and infection.[34-37] Moreover, ML models can be designed for different outcome measures and as such there is a benefit to global institutions in seeing how this model is designed and which types of models performed best.

While the primary predictors of prolonged ICU stay identified in our study are indeed based on operative and postoperative data in addition to preoperative risk factors, their implications can be communicated preoperatively to enhance patient understanding and involvement. Surgeons can use this information to discuss potential risks and strategies to mitigate these risks, such as optimizing surgical techniques to minimize duration and blood loss. This proactive approach can demystify the surgical process and engage patients more deeply in their care journey, promoting informed consent and collaborative decision making. By outlining the factors that influence outcomes, this can allow patients to participate meaningfully in discussions about their surgical options and expectations.

Limitations

This study was not without limitations. First, the data collected was from a single, large, urban institution and as a result may not be generalizable to other hospital systems or the general population. This is a North American institution and as such our postoperative measure may not be universally applicable. Greater confidence in these results would require additional cross validation with other external datasets. Furthermore, as a retrospective analysis, this study is subject to potential biases based on patient selection, the time frame of analysis, and other factors. It is also important to note that the three strongest predictors in this model are operative and postoperative variables, which is not relevant to preoperative risk stratification. In addition, our dataset had a class imbalance, with the no required ICU stay cohort being almost four times as large as the required ICU stay cohort, which may have impacted the model performance. Finally, it must also be noted that despite

the limitation of the logistic regression's top-down approach, it still managed to perform the second-best of the five tested models. This highlights not only the simple robustness of the classic logistic regression model but also the fact that many machine learning models require large amounts of data before they can be adequately trained to out-perform the logistic regression, as did the Random Forest model in this study.

Conclusion

The results of this study affirm the validity of machine learning models for understanding drivers of prolonged ICU stay in ASD patients. SHAP analysis allowed these models to be explainable and assisted in determining that the findings were in alignment with previous studies. The application of such models in the clinical setting may allow providers to better categorize patient risk profiles, to optimize cost-saving measures and avoid prolonged ICU stay.

Declaration of Conflicting Interests

Funding

ORCID iDs

Bashar Zaidat   https://orcid.org/0000-0002-8823-720X
Samuel K. Cho   https://orcid.org/0000-0001-7511-2486

References

1. National and Surgical Health Care Expenditures. Ann Surgals of Surgery. 2005. https://journals.lww.com/annalsofsurgery/abstract/2010/02000/national_and_surgical_health_care_expenditures,.2.aspx. Accessed 5 December 2023.
2. Goz V, Weinreb JH, Lafage V, Errico TJ. Perioperative complications and mortality after spinal fusions: analysis of trends and risk factors. *Spine J*. 2013;13(9):S105-S106. doi:10.1016/j.spinee.2013.07.283.
3. Analysis of national rates, cost, and sources of cost variat. https://journals.lww.com/neurosurgery/abstract/2018/03000/analysis_of_national_rates,_cost,_and_sources_of.26.aspx. Accessed December 5, 2023.
4. Yeramaneni S, Robinson C, Hostin R. Impact of spine surgery complications on costs associated with management of adult spinal deformity. *Curr Rev Musculoskelet Med*. 2016;9(3):327-332. doi:10.1007/s12178-016-9352-9.

5. Klineberg EO, Passias PG, Jalai CM, et al. Predicting extended length of hospital stay in an adult spinal deformity surgical population. *Spine*. 2016;41(13):E798. doi:10.1097/BRS.0000000000001391.

6. Schupper AJ, Shuman WH, Baron RB, et al. Utilization of the American Society of Anesthesiologists (ASA) classification system in evaluating outcomes and costs following deformity spine procedures. *Spine Deform*. 2021;9(1):185-190. doi:10.1007/s43390-020-00176-4.

7. Martini ML, Neifert SN, Gal JS, Oermann EK, Gilligan JT, Caridi JM. Drivers of prolonged hospitalization following spine surgery: a game-theory-based approach to explaining machine learning models. *JBJS*. 2021;103(1):64. doi:10.2106/JBJS.20.00875.

8. Valliani AA, Kim NC, Martini ML, et al. Robust prediction of non-home discharge after thoracolumbar spine surgery with ensemble machine learning and validation on a nationwide cohort. *World Neurosurg*. 2022;165:e83-e91. doi:10.1016/j.wneu.2022.05.105.

9. Geng EA, Gal JS, Kim JS, et al. Robust prediction of nonhome discharge following elective anterior cervical discectomy and fusion using explainable machine learning. *Eur Spine J*. 2023; 32(6):2149-2156. doi:10.1007/s00586-023-07621-8.

10. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: decide-ai. *Nat Med*. 2022; 28(5):924-933. doi:10.1038/s41591-022-01772-9.

11. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open*. 2021;11(7):e048008. doi:10.1136/bmjopen-2020-048008.

12. Yoshida G, Boissiere L, Larrieu D, et al. Advantages and disadvantages of adult spinal deformity surgery and its impact on health-related quality of life. *Spine*. 2017;42(6):411. doi:10.1097/BRS.0000000000001770.

13. Youssef JA, Orndorff DO, Patty CA, et al. Current status of adult spinal deformity. *Global Spine J*. 2013;3(1):051-062. doi:10.1055/s-0032-1326950.

14. Scheer JK, Mundis GM, Klineberg E, et al. Postoperative recovery after adult spinal deformity surgery: comparative analysis of age in 149 patients during 2-year follow-up. *Spine*. 2015; 40(19):1505-1515. doi:10.1097/BRS.0000000000001062.

15. Glassman SD, Hamill CL, Bridwell KH, Schwab FJ, Dimar JR, Lowe TG. The impact of perioperative complications on clinical outcome in adult deformity surgery. *Spine*. 2007;32(24):2764. doi:10.1097/BRS.0b013e31815a7644.

16. Norris C, Jacobs P, Rapoport J, Hamilton S. ICU and non-ICU cost per day. *Can J Anaesth*. 1995;42(3):192-196. doi:10.1007/BF03010674.

17. Zilberberg MD. Understanding cost-effectiveness in the ICU. *Semin Respir Crit Care Med*. 2010;31(1):13-18. doi:10.1055/s-0029-1246282.

18. Gershengorn HB, Garland A, Gong MN. Patterns of daily costs differ for medical and surgical intensive care unit patients. *Ann Am Thorac Soc*. 2015;12(12):1831-1836. doi:10.1513/AnnalsATS.201506-366BC.

19. De la Garza-Ramos R, Nakhla J, Gelfand Y, et al. Predicting critical care unit-level complications after long-segment fusion procedures for adult spinal deformity. *J Spine Surg*. 2018;4(1): 55-61. doi:10.21037/jss.2018.03.15.

20. Amin RM, Raad M, Jain A, Sandhu KP, Frank SM, Kebaish KM. Increasing body mass index is associated with worse perioperative outcomes and higher costs in adult spinal deformity surgery. *Spine*. 2018;43(10):693-698. doi:10.1097/BRS.0000000000002407.

21. Ley C, Martin RK, Pareek A, Groll A, Seil R, Tischer T. Machine learning and conventional statistics: making sense of the differences. *Knee Surg Sports Traumatol Arthrosc*. 2022; 30(3):753-757. doi:10.1007/s00167-022-06896-6.

22. Soroceanu A, Burton DC, Oren JH, et al. Medical complications after adult spinal deformity surgery: incidence, risk factors, and clinical impact. *Spine*. 2016;41(22):1718. doi:10.1097/BRS.0000000000001636.

23. Bhagat S, Vozar V, Lutchman L, Crawford RJ, Rai AS. Morbidity and mortality in adult spinal deformity surgery: norwich Spinal Unit experience. *Eur Spine J*. 2013;22(1):42-46. doi:10.1007/s00586-012-2627-y.

24. Smith JS, Klineberg E, Lafage V, et al. Prospective multicenter assessment of perioperative and minimum 2-year postoperative complication rates associated with adult spinal deformity surgery. *J Neurosurg Spine*. 2016;25(1):1-14. doi:10.3171/2015.11.SPINE151036.

25. Uribe JS, Deukmedjian AR, Mummaneni PV, et al. Complications in adult spinal deformity surgery: an analysis of minimally invasive, hybrid, and open surgical techniques. *Neurosurg Focus*. 2014;36(5):E15. doi:10.3171/2014.3.FOCUS13534.

26. Samdani AF, Belin EJ, Bennett JT, et al. Major perioperative complications after spine surgery in patients with cerebral palsy: assessment of risk factors. *Eur Spine J*. 2016;25(3):795-800. doi:10.1007/s00586-015-4054-3.

27. Nahtomi-Shick O, Kostuik JP, Winters BD, Breder CD, Sieber AN, Sieber FE. Does intraoperative fluid management in spine surgery predict intensive care unit length of stay? *J Clin Anesth*. 2001;13(3):208-212. doi:10.1016/S0952-8180(01)00244-6.

28. Kim JS, Arvind V, Oermann EK, et al. Predicting surgical complications in patients undergoing elective adult spinal deformity procedures using machine learning. *Spine Deform*. 2018;6(6):762-770. doi:10.1016/j.jspd.2018.03.003.

29. Dagal A, Bellabarba C, Bransford R, et al. Enhanced perioperative care for major spine surgery. *Spine*. 2019;44(13): 959-966. doi:10.1097/BRS.0000000000002968.

30. Lubelski D, Feghali J, Ehresman J, et al. Web-based calculator predicts surgical-site infection after thoracolumbar spine surgery. *World Neurosurg*. 2021;151:e571-e578. doi:10.1016/j.wneu.2021.04.086.

31. Shi X, Cui Y, Wang S, Pan Y, Wang B, Lei M. Development and validation of a web-based artificial intelligence prediction model to assess massive intraoperative blood loss for metastatic spinal disease using machine learning techniques. *Spine J*. 2024;24(1): 146-160. doi:10.1016/j.spinee.2023.09.001.

32. Lee CH, Jo DJ, Oh JK, et al. Development and validation of an online calculator to predict proximal junctional kyphosis after

adult spinal deformity surgery using machine learning. *Neurospine*. 2023;20(4):1272-1280. doi:10.14245/ns.2342434.217.

33. Wang KY, Ikwuezunma I, Puvanesarajah V, et al. Using predictive modeling and supervised machine learning to identify patients at risk for venous thromboembolism following posterior lumbar fusion. *Global Spine J*. 2023;13(4):1097-1103. doi:10.1177/21925682211019361.

34. Shah AA, Devana SK, Lee C, et al. Prediction of major complications and readmission after lumbar spinal fusion: a machine learning–driven approach. *World Neurosurg*. 2021;152:e227-e234. doi:10.1016/j.wneu.2021.05.080.

35. Karhade AV, Ogink P, Thio Q, et al. Development of machine learning algorithms for prediction of discharge disposition after elective inpatient surgery for lumbar degenerative disc disorders. *Neurosurg Focus*. 2018;45(5):E6. doi:10.3171/2018.8.FOCUS18340.

36. Liu WC, Ying H, Liao WJ, et al. Using preoperative and intraoperative factors to predict the risk of surgical site infections after lumbar spinal surgery: a machine learning–based study. *World Neurosurg*. 2022;162:e553-e560. doi:10.1016/j.wneu.2022.03.060.

37. Ma Y, Lu Q, Yuan F, Chen H. Comparison of the effectiveness of different machine learning algorithms in predicting new fractures after PKP for osteoporotic vertebral compression fractures. *J Orthop Surg*. 2023;18(1):62. doi:10.1186/s13018-023-03551-9.