Check for updates

Prediction of Major Complications and Readmission After Lumbar Spinal Fusion: A Machine Learning—Driven Approach

Akash A. Shah[1], Sai K. Devana[1], Changhee Lee[2], Amador Bugarin[1], Elizabeth L. Lord[1], Arya N. Shamie[1], Don Y. Park[1], Mihaela van der Schaar[2,3], Nelson F. SooHoo[1]

■ BACKGROUND: Given the significant cost and morbidity of patients undergoing lumbar fusion, accurate preoperative risk-stratification would be of great utility. We aim to develop a machine learning model for prediction of major complications and readmission after lumbar fusion. We also aim to identify the factors most important to performance of each tested model.

■ METHODS: We identified 38,788 adult patients who underwent lumbar fusion at any California hospital between 2015 and 2017. The primary outcome was major perioperative complication or readmission within 30 days. We build logistic regression and advanced machine learning models: XGBoost, AdaBoost, Gradient Boosting, and Random Forest. Discrimination and calibration were assessed using area under the receiver operating characteristic curve and Brier score, respectively.

■ RESULTS: There were 4470 major complications (11.5%). The XGBoost algorithm demonstrates the highest discrimination of the machine learning models, outperforming regression. The variables most important to XGBoost performance include angina pectoris, metastatic cancer, teaching hospital status, history of concussion, comorbidity burden, and workers' compensation insurance. Teaching hospital status and concussion history were not found to be important for regression.

■ CONCLUSIONS: We report a machine learning algorithm for prediction of major complications and readmission after lumbar fusion that outperforms logistic regression. Notably, the predictors most important for XGBoost differed from those for regression. The superior performance of XGBoost may be due to the ability of advanced machine learning methods to capture relationships between variables that regression is unable to detect. This tool may identify and address potentially modifiable risk factors, helping risk-stratify patients and decrease complication rates.

INTRODUCTION

With more than 450,000 cases performed per year, spinal fusion is one of the most commonly performed surgeries in the United States and is rapidly growing in prevalence.[1-3] There has been an increase in all spinal fusions across all ages between 2001 and 2010, with the largest increase (74%) observed for lumbar fusion. During the same time period, the average total charges increased by over 100%.[2] Complications with associated unplanned readmissions are a driver of increased cost and morbidity.[3] The average age and comorbidity burden of patients undergoing lumbar fusion have also increased.[2] Older with more medical comorbidities, these patients are at elevated risk of perioperative complications. It is thus important to

preoperatively assess which patients are at increased risk of major perioperative complications.

Numerous studies have used logistic regression (LR) to analyze risk factors and create prediction models for outcomes after lumbar fusion.[1-8] Advanced machine learning (ML) methods have grown in popularity because of their ability to recognize complex nonlinear relationships, often outperforming LR.[4-7] ML has been increasingly employed for degenerative, neoplastic, and infectious spinal pathology.[8-12] Yet ML methods have been sparingly used to specifically predict outcomes after lumbar fusion. Using a large patient discharge database, we primarily aim to build an ML model for prediction of major complication or readmission after lumbar fusion. Our secondary aim is to compare its performance against LR. We hypothesize that the optimized model will outperform LR as well as identify novel risk factors for major complication or readmission after lumbar fusion.

MATERIAL AND METHODS

Study Design and Subjects

This study is a retrospective review of patients undergoing lumbar fusion using the California Office of Statewide Health and Planning and Development (OSHPD) Patient Discharge Database (PDD). The PDD is a mandatory statewide discharge database containing admissions data for all nonfederal hospital admissions in California. Patients in this database are assigned a unique record linkage number that allows them to be tracked longitudinally for complications as long as future admissions occur at a hospital that is included in the PDD. We identified patients undergoing lumbar fusion between 2015 and 2017. We included patients 18 years or older who underwent lumbar fusion using International Classification of Diseases, Tenth Revision (ICD-10) procedure codes to identify patients (Supplementary Table 1).

Outcome and Other Variables

The primary outcome was any major complication or readmission after index lumbar fusion. Complications were identified using ICD-10 codes adapted from performance measures developed by the Centers for Medicare and Medicaid (CMS) for total joint replacement.[13] These include acute myocardial infarction, pneumonia, sepsis, pulmonary embolism, surgical site bleeding, and wound infection. Myocardial infarction, pneumonia, and sepsis were included if the complication occurred during the index admission or within 7 days of index admission. Pulmonary embolism was included if it occurred during the index admission or within 30 days of admission. Surgical site bleeding and wound infection were included during the index admission or within 90 days. Readmission for any cause within 30 days of index lumbar fusion was included as an outcome. The ICD-10 diagnosis and procedure codes used to identify surgical site bleeding and wound complications are specific to lumbar spine surgery.

Explanatory features collected for the cohort include patient demographic characteristics (e.g., age, sex, insurance type), hospital characteristics (e.g., volume, teaching institution), and patient medical comorbidities using the Condition Categories as defined by the CMS Hierarchical Condition Category risk adjustment model (e.g., malignancy, coronary atherosclerosis, renal failure, diabetes).

Model Development and Evaluation

We built 5 standard ML benchmark models spanning different classes of ML modeling approaches: LR (a linear classifier), random forest (a tree-based ensemble classifier), AdaBoost, gradient boosting machines, and XGBoost (boosting ensemble classifiers).[14-17] We implemented LR, random forest, AdaBoost, and gradient boosting machines using the *scikit-learn* Python library.[18] XGBoost was built using the *xgboost* Python library.[17]

We evaluated the discrimination and calibration performances of the prognostic models using 5-fold stratified cross-validation to avoid overfitting. In each cross-validation fold, the training cohort (80% of the study population) was used to derive the models. A hold-out testing cohort (20% of the population) was used for performance evaluation. We report the mean and 95% confidence interval for all models.

Assessed by area under the receiver operating characteristic curve (AUROC), discrimination determines how well a model distinguishes patients who developed complications from those who did not. AUROC represents the probability that a randomly selected patient who experienced the outcome was assigned a higher risk by the classifier than a patient who did not experience the outcome. An AUROC of 0.5 indicates random prediction, whereas an AUROC of 1 indicates perfect prediction.[19,20] Calibration signifies the agreement between the model's predictions and observed outcomes in the study population. The Brier score is the mean squared error between the observed values and the predicted probabilities; it is a measure of discrimination and calibration. Brier scores closer to zero indicate a more accurate model.[19]

The area under the precision-recall curve (AUPRC) is a useful performance metric when analyzing a dataset in which negative cases far outnumber positive cases. The precision-recall curve is constructed by plotting positive predictive value (precision) versus the sensitivity (recall). Ignoring true negatives, the precision-recall curve depicts the model's ability to correctly identify positive cases.[21,22] Unlike AUROC, the baseline AUPRC is the proportion of true positive cases in the cohort. Random prediction will result in the baseline AUPRC. The higher the AUPRC compared with the baseline value, the better the model handles positive cases. An AUPRC of 1 suggests a classifier with perfect recall and precision.

Feature Importance

We use the partial dependence function described by Friedman to measure importance of an individual feature by assessing the average effect in predicted risks when its value is altered.[16] Specifically, x_c is a chosen target feature in the set of input features $\mathscr{X}$ and $\mathscr{X}_{\setminus c}$ is its complement (i.e., $\mathscr{X} = \mathscr{X}_{\setminus c} \cup x_c$), and $r(\mathscr{X}) = r(\mathscr{X}_{\setminus c}, x_c)$ is the predicted risk by our trained model. We define the feature importance score for an individual feature x_c by averaging $r(\mathscr{X}_{\setminus c}, x_c = 1) - r(\mathscr{X}_{\setminus c}, x_c = 0)$ for binary features and $r(\mathscr{X}_{\setminus c}, x_c = \max(x_c)) - r(\mathscr{X}_{\setminus c}, x_c = \min(x_c))$, where $\max(x_c)$ and $\min(x_c)$ are the maximum and minimum of feature x_c for continuous features, respectively. For categorical variables, we define feature importance of category $b \in \{1,...,B\}$ as $r(\mathscr{X}_{\setminus c}, x_c = b) - r(\mathscr{X}_{\setminus c}, x_c = \text{mode}(x_c))$, where $\text{mode}(x_c)$ indicates the most frequency category of feature x_c.

RESULTS

Baseline Characteristics

A total of 38,788 patients met inclusion criteria for this study. The median age of the cohort was 64 years, with 18,021 males (46.5%). The most common medical comorbidity was diabetes mellitus (7.1%), followed by coronary atherosclerosis (5.7%) and chronic obstructive pulmonary disease (5.4%). The mean number of Condition Categories that a patient fell under was 0.34. A complete description of the cohort demographics is provided in Table 1. A total of 4470 patients (11.5%) had at least 1 complication or readmission. There were 3354 patients (8.6%) who required readmission within 30 days. The most common complications were pneumonia, sepsis, and pulmonary embolism (Table 2).

Model Performance

Algorithms predicting the risk of major complications or readmission after lumbar fusion were built with LR and 4 benchmark ML models (XGBoost, Gradient Boosting, AdaBoost, Random Forest). In the overall cohort, XGBoost demonstrates higher discrimination (AUROC: 0.687 ± 0.01) compared with LR (AUROC: 0.675 ± 0.01). It also outperforms the standard benchmark ML models. The XGBoost model is well calibrated with a Brier score of 0.094 ± 0.001. The LR and benchmark ML models are similarly well calibrated, with the exception of AdaBoost. The AUPRC of the XGBoost model is 0.284; a random classifier would result in AUPRC of 0.115 (Table 3). The receiver operating characteristic and precision-recall curves of the XGBoost and LR models are depicted in Figures 1 and 2, respectively.

Feature Importance

The relative importance of each feature to model performance for XGBoost and LR is displayed in Table 4. The features most important for risk prediction for XGBoost include angina pectoris, metastatic cancer, musculoskeletal infection, other malignancy, and history of cerebral hemorrhage. The most important continuous variable for both XGBoost and LR is number of medical comorbidities as defined by the CMS Condition Categories. Workers' compensation insurance status is the most important insurance category toward risk prediction for XGBoost. The variables most important for XGBoost differ from those for LR.

DISCUSSION

The prevalence of lumbar fusion is projected to increase substantially in the coming decades.[3] The cost of caring for lumbar pathology exceeds $100 billion per year, almost equal to the cost of treating all malignancies combined.[23-25] Given the morbidity and cost incurred by perioperative complications and unplanned readmissions, accurate prediction of a patient's complication risk after lumbar fusion would be useful. Accurate prognostic information can inform preoperative counseling and management decisions. Furthermore, identification of high-risk patients before surgery provides an opportunity to preoperatively address potentially modifiable risk factors.

Numerous studies have used multivariate LR modeling for prediction of complications after lumbar spinal surgery.[26-33]

Table 1. Baseline Cohort Demographics

Variable	All Patients (n = 38,788)
Demographics	
Age (years), median (IQR)	64 (54–72)
Hospital volume,* median (IQR)	412 (222–681)
Male	18,021 (46.46)
Race	
White	31,223 (80.50)
Black	1834 (4.73)
Asian/Pacific Islander	1746 (4.50)
Native American	145 (0.37)
Other	3515 (9.06)
Unknown	325 (0.84)
Ethnicity	
Non-Hispanic	32,117 (82.80)
Hispanic	6196 (15.97)
Unknown	475 (1.22)
Insurance	
Medicare	18,810 (48.49)
Private	12,449 (32.09)
Medi-Cal	3343 (8.62)
Workers' compensation	3259 (8.40)
Other	927 (2.39)
Medical comorbidities	
Diabetes mellitus	2764 (7.13)
Coronary atherosclerosis	2219 (5.72)
Angina pectoris	1600 (4.12)
COPD	2080 (5.36)
Metastatic cancer or acute leukemia	1685 (4.34)
Other major cancer	1630 (4.20)
Protein-calorie malnutrition	1740 (4.49)
Osteoporosis	2012 (5.19)
Osteoarthritis	1876 (4.84)
Chronic kidney disease requiring dialysis	1612 (4.16)
Bone, joint, or muscle infection	1812 (4.67)
Vertebral fracture without spinal cord injury	1871 (4.82)
Concussion or unspecified head injury	1572 (4.05)
Complications of implants	1810 (4.67)
Other complications of medical care	1957 (5.05)
Number of medical comorbidities, mean	0.34

Data represent n (%) unless otherwise specified.
COPD, chronic obstructive pulmonary disease; IQR, interquartile range.
*Cases of lumbar fusions performed between 2015 and 2017.

Table 2. Major Complications and Readmission

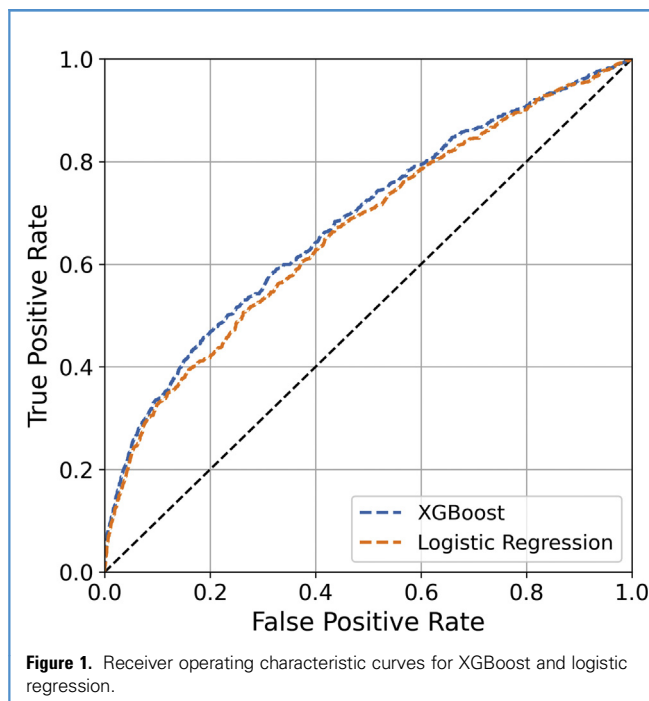| Complications	All Patients (n = 38,788), n (%)
At least 1 complication or readmission	4470 (11.52)
Readmission within 30 days	3354 (8.65)
Pneumonia	724 (1.87)
Sepsis	701 (1.81)
Pulmonary embolism	303 (0.78)
Acute myocardial infarction	123 (0.32)
Surgical site bleeding	86 (0.22)
Wound infection	15 (0.04)

ML methods have grown in popularity in recent years in the neurosurgery and orthopedic surgery literature. A subset of artificial intelligence, ML represents a set of techniques that allow machines to learn and perform classification tasks by recognizing underlying patterns in data. In contrast to conventional regression techniques, advanced ML methods can detect complex nonlinear relationships as well as factor-factor interactions.[34,35] Indeed, ML models have been shown to outperform LR in many cases.[5-7] ML has been increasingly employed in spinal surgery for prediction of outcomes including discharge disposition, surgical site infection, presence of vertebral compression fracture, and mortality in metastatic disease and infection.[10,36-39]

ML has been sparingly used to predict outcomes after lumbar fusion. With 22,629 patients from the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP), Kim et al[40] developed an artificial neural network to predict complications after posterior lumbar fusion with an AUROC of 0.641. Goyal et al[3] queried ACS-NSQIP for cervical and lumbar fusions, analyzing 59,145 cases for prediction of unplanned readmission. They developed ML models predicting readmission with AUROC ranging between 0.63 and 0.66.[3] Most recently, Jain et al[9] analyzed 37,852 patients who underwent posterior lumbar fusion of 3 or more levels to predict 90-day major complication and 90-day readmissions. ML models achieved an AUROC of 0.69 and 0.63 for prediction of major complications and readmission, respectively.[9] ML methods underperformed LR in their analysis.

Table 3. Discrimination, Calibration, and AUPRC of Logistic Regression and Advanced ML Models

Model	AUROC	Brier Score	AUPRC
XGBoost	0.687 ± 0.01	0.094 ± 0.01	0.284 ± 0.014
Logistic Regression	0.675 ± 0.01	0.095 ± 0.01	0.265 ± 0.011
Gradient Boosting	0.686 ± 0.009	0.094 ± 0.009	0.283 ± 0.009
AdaBoost	0.686 ± 0.011	0.248 ± 0.011	0.278 ± 0.015
Random Forest	0.629 ± 0.006	0.112 ± 0.006	0.197 ± 0.007

AUPRC, area under the precision-recall curve; AUROC, area under the receiver operating characteristic curve; ML, machine learning.

Figure 1. Receiver operating characteristic curves for XGBoost and logistic regression.

With a cohort of 38,788 patients, we report a boosting model (XGBoost) for prediction of major perioperative complications and 30-day readmission. With an AUROC of 0.687, this model represents fair accuracy comparable or superior to currently available ML algorithms predicting complications after lumbar fusion.[3,9,40] This model also outperforms the ACS-NSQIP calculator and the Risk Assessment Tool developed by Veeravagu et al[33] for prediction of complications after spinal surgery. Our reported model is well calibrated and
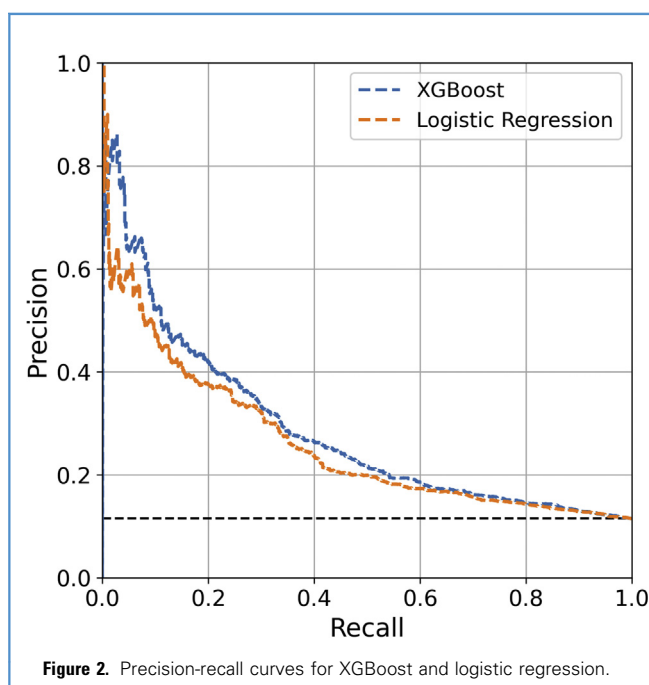
Figure 2. Precision-recall curves for XGBoost and logistic regression.

Table 4. Relative Feature Importance for Complications After Lumbar Fusion

Feature	Rank in XGBoost (Rank in Logistic Regression)	Change to Risk Prediction
Binary features		
Angina pectoris	1 (4)	−0.0535
Metastatic cancer or leukemia	2 (1)	0.0463
Musculoskeletal infection	3 (3)	0.0321
Malignancy	4 (7)	−0.0259
Cerebral hemorrhage	5 (8)	−0.0235
Diabetes mellitus	6 (6)	0.0499
Complications of medical care	7 (5)	0.0508
Teaching hospital	8 (34)	0.0186
Implant complication	9 (2)	0.0183
Concussion	10 (17)	−0.0174
Continuous features		
Number of comorbidities	1 (1)	0.12222
Age	2 (2)	0.0701
Hospital volume	3 (3)	0.0301
Insurance status		
Medicare	Reference	0
Workers' compensation	1 (2)	−0.0284
Personal	2 (1)	−0.0259
Other	3 (4)	−0.0201
Medi-Cal	4 (3)	0.0193

demonstrates superior discrimination compared with LR. With an AUPRC of 0.284, the XGBoost model shows good performance in this relatively imbalanced dataset and far exceeds random prediction. AUPRC is a useful performance metric to evaluate the predictive performance of ML models in imbalanced datasets.[21,22]

We also report the importance of each feature to performance of the XGBoost model. Although these features are not necessarily causative, their inclusion increases the predictive performance of the model. The most important binary feature identified is angina pectoris. A possible sign of symptomatic coronary artery disease (CAD), angina has not specifically been implicated as an important feature for poor outcomes after lumbar fusion. CAD has been reported as a predictor for readmission after elective lumbar spinal surgery.[32] Associated with reduced immunity, diabetes mellitus is an established risk factor for poor outcomes after spinal surgery including wound complications and infection. Diabetes has been shown to be a predictor for pneumonia, mortality, sepsis, wound complications, and readmission after lumbar fusion.[28,30,41] History of malignancy, musculoskeletal infection, and cerebral hemorrhage are comorbidities that are important contributors to XGBoost

model performance. A marker of low physiologic reserve and immunosuppressed state, metastatic cancer has been shown to be an important predictor of mortality after spinal fusion.[2] Cerebral hemorrhage is associated with the development of multiple medical conditions including renal disease, cognitive impairment, and dementia.[42-44] These associated comorbidities are risk factors for poor outcomes after lumbar spine surgery.[45-47] Number of CMS Condition Categories is the most important continuous feature to model performance, suggesting that overall comorbidity burden may contribute to poor outcomes after lumbar fusion.

We additionally find that a history of complications in past admissions (e.g., failed/difficult intubation, excessive transfusion, implant infection) is important to model performance. Similarly, a history of implant-related complications is an important feature for XGBoost—potentially a sign of revision surgery. Although these features encompass a wide range of diagnosis codes due to constraints of the OSHPD dataset, they can be interpreted as markers of past complications or mechanical complication. We also find that workers' compensation insurance status is an important categorical feature. Although workers' compensation has been implicated extensively in the elective spinal surgery literature, it has not been identified as an important factor in ML analysis to our knowledge.[48-50] A proposed hypothesis for the link between workers' compensation and poor outcomes is secondary gain for the patient (e.g., settling claims from civil litigation).[51] Furthermore, patients with active workers' compensation claims are more likely to have positive smoking history and elevated body mass index, both of which are linked to adverse outcomes after spine surgery.[49-51]

Notably, 2 features were found to be extremely important to XGBoost but markedly less so for LR. History of concussion is the 10th most important feature for XGBoost but the 17th most important for LR; this may indicate prior head trauma or neurologic injury. Concussion and traumatic brain injury have been shown to be associated with the development of dementia and psychiatric disorders.[52-54] Both dementia and psychiatric comorbidities are risk factors for adverse outcomes after lumbar fusion.[45,55] Interestingly, teaching hospital status is the 8th most important feature for XGBoost but the 34th most important for LR. Although this has not been shown before in an ML analysis, a study by Durand et al[56] suggests that teaching hospitals may have suboptimal resource allocation compared with nonteaching hospitals when it comes to lumbar fusion. Resident physician involvement is independently associated with increased complication risk after spinal fusion, although this finding may be confounded by increased case complexity at an academic teaching hospital.[26]

Differences in feature importance for XGBoost versus LR must be interpreted with caution. For example, metastatic cancer is an important feature for both models—it is the most important feature for LR and the second most important feature for XGBoost. This does not imply that metastatic cancer is more correlated with complications than teaching hospital status, a feature markedly different in importance between XGBoost and LR. In fact, XGBoost is unable to determine correlation at all; it is designed for classification and not for statistical inference. Rather, this finding underlines that XGBoost and LR analyze the same features quite differently. The superior performance of XGBoost in this dataset is due to its ability to capture relationships between variables that regression is unable to detect.

This study has limitations, first of which is its retrospective design. The use of a de-identified administrative database limits the granularity of the collected variables and outcomes. The reliance on diagnosis codes to assign complications may underestimate complication rates and is less comprehensive compared with chart review. Similarly, medical comorbidities may not always be coded accurately. Specific laboratory values (e.g., hemoglobin A1c) are not available for manual look-up, requiring reliance on CMS-Hierarchical Condition Category Condition Categories comprising groups of ICD-10 codes to determine comorbidities. Although we aim to only capture complications associated with surgery by limiting inclusion of complications in the immediate perioperative period, we cannot exclude the possibility that a small number of medical complications that occur in the perioperative period are unrelated to index lumbar fusion. Furthermore, this database does not contain data on mortality or patient-reported functional outcomes. In addition, with any predictive model there exists a concern for overfitting. In overfitting, the algorithm has good performance on the development set but generalizability to new cohorts is diminished because the algorithm is fit to the idiosyncrasies of the development cohort. Although we aim to protect against overfitting with our model development and validation strategy, future studies in which this algorithm is validated on external cohorts are necessary. Finally, it should be noted that systematic biases present in the dataset may be amplified by ML, potentially propagating treatment biases adversely affecting under-represented groups such as patients of lower socioeconomic status and ethnic minorities.[35]

CONCLUSIONS

We report a boosting classifier that predicts major complications and readmission after lumbar fusion with superior accuracy compared with LR. In addition, it identifies novel features important to model performance. By including all lumbar fusions regardless of indication or approach, we hope to enhance the generalizability of this model. Furthermore, by providing accurate prognostic information, this tool may facilitate preoperative shared decision-making and aid with appropriate patient selection. No predictive model is a substitute for clinical judgment. The determination of an acceptable risk-to-benefit ratio remains solely with the patient and his or her surgeon; we simply aim to provide an accurate estimation of perioperative risk. Furthermore, this tool can be used to identify and address potentially modifiable risk factors for complications. Preoperatively addressing issues such as diabetes and CAD may reduce overall health care costs by decreasing the likelihood of complications after lumbar fusion.

REFERENCES

1. Fingar K, Stocks C, Weiss A, Steiner C. *Most frequent operating room procedures performed in U.S. hospitals, 2003-2012: Statistical brief #186. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs.* Rockville, MD: Agency for Healthcare Research and Quality; 2014.

2. Goz V, Weinreb J, McCarthy I, Schwab F, Lafage V, Errico T. Perioperative complications and mortality after spinal fusions: analysis of trends and risk factors. *Spine (Phila Pa 1976).* 2013;38:1970-1976.

3. Goyal A, Ngufor C, Kerezoudis P, McCutcheon B, Storlie C, Bydon M. Can machine learning algorithms accurately predict discharge to nonhome facility and early unplanned readmissions following spinal fusion? Analysis of a national surgical registry. *J Neurosurg Spine.* 2019;31:568-578.

4. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. *Front Bioeng Biotechnol.* 2018;6:1-20.

5. Esteva A, Kuprel B, Novoa R, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542:686.

6. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *J Am Med Assoc.* 2016;316:2402-2410.

7. Menden M, Iorio F, Garnett M, et al. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One.* 2013;8:e61318.

8. Hopkins B, Yamaguchi J, Garcia R, et al. Using machine learning to predict 30-day readmissions after posterior lumbar fusion: an NSQIP study involving 23,264 patients. *J Neurosurg Spine.* 2020;32: 399-406.

9. Jain D, Durand W, Burch S, Daniels A, Berven S. Machine learning for predictive modeling of 90-day readmission, major medical complication, and discharge to a facility in patients undergoing long segment posterior lumbar spine fusion. *Spine (Phila Pa 1976).* 2020;45:1151-1160.

10. Karhade AV, Thio QC, Ogink PT, et al. Development of machine learning algorithms for prediction of 30-day mortality after surgery for spinal metastasis. *Neurosurgery.* 2019;85:E83-E91.

11. Karhade AV, Thio QCBS, Ogink PT, et al. Predicting 90-day and 1-year mortality in spinal metastatic disease: development and internal validation. *Neurosurgery.* 2019;85:E671-E681.

12. Shah AA, Karhade AV, Bono CM, Harris MB, Nelson SB, Schwab JH. Development of a machine learning algorithm for prediction of failure of nonoperative management in spinal epidural abscess. *Spine J.* 2019;19:1657-1665.

13. Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation. *2017 Procedure-specific measure updates and specifications report hospital-level risk-standardized complication measure: elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA)—version 6.0.* Baltimore, MD: Centers for Medicare & Medicaid Services; 2017.

14. Breiman L. Random forests. *Mach Learn.* 2001;45: 5-32.

15. Ratsch G, Onoda T, Muller K. Soft margins for AdaBoost. *Mach Learn.* 2001;42:287-320.

16. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat.* 2001;29: 1189-1232.

17. Chen T, 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.

18. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.

19. Manning DW, Edelstein AI, Alvi HM. Risk prediction tools for hip and knee arthroplasty. J Am Acad Orthop Surg. 2016;24:19-27.

20. Harris AHS, Kuo AC, Bozic KJ, et al. American joint replacement registry risk calculator does not predict 90-day mortality in veterans undergoing total joint replacement. Clin Orthop Relat Res. 2018; 476:1869-1875.

21. Ozenne B, Subtil F, Maucort-Boulch D. The precision-recall curve overcame the optimism of the receiver operaitng characteristic curve in rare diseases. J Clin Epidemiol. 2015;68:855-859.

22. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015;10:e0118432.

23. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J. 2008;8:8-20.

24. Martin B, Deyo R, Mirza S, et al. Expenditures and health status among adults with back and neck problems. JAMA. 2008;299:656-664.

25. Mummaneni P, Whitmore R, Curran J, et al. Cost-effectiveness of lumbar discectomy and single-level fusion for spondylolisthesis: experience with the NeuroPoint-SD registry. Neurosurg Focus. 2014;6:E3.

26. Schoenfeld A, Carey P, Cleveland A, Bader J, Bono C. Patient factors, comorbidities, and surgical characteristics that increase mortality and complication risk after spinal arthrodesis: a prognostic study based on 5,887 patients. Spine J. 2013;13:1171-1179.

27. Bohl D, Ahn J, Tabaraee E, et al. Urinary tract infection following posterior lumbar fusion procedures: an American College of Surgeons National Surgical Quality Improvement Program study. Spine (Phila Pa 1976). 2015;40:1785-1791.

28. Bohl D, Mayo B, Massel D, et al. Incidence and risk factors for pneumonia after posterior lumbar fusion procedures: an ACS-NSQIP study. Spine (Phila Pa 1976). 2016;41:1058-1063.

29. Di Capua J, Somani S, Kim J, et al. Analysis of risk factors for major complications following elective posterior lumbar fusion. Spine (Phila Pa 1976). 2017; 42:1347-1354.

30. Lee N, Kothari P, Phan K, et al. Incidence and risk factors for 30-day unplanned readmissions after elective posterior lumbar fusion. Spine (Phila Pa 1976). 2017;43:41-48.

31. Khormaee S, Samuel A, Schairer W, et al. Discharge to inpatient facilities after lumbar fusion surgery is associated with increased postoperative venous thromboembolism and readmissions. Spine J. 2019;19:430-436.

32. Sivaganesan A, Zuckerman S, Khan I, et al. Predictive model for medical and surgical readmissions following elective lumbar spine surgery: a national study of 33,674 patients. Spine (Phila Pa 1976). 2019;44:588-600.

33. Veeravagu A, Li A, Swinney C, et al. Predicting complication risk in spine surgery: a prospective analysis of a novel risk assessment tool. J Neurosurg Spine. 2017;27:81-91.

34. Chen J, Asch S. Machine learning and predition in medicine—beyond the peak of inflated expectations. N Engl J Med. 2017;376:2507-2509.

35. Hashimoto D, Rosman G, Rus D, Meireles O. Artificial intelligence in surgery: promises and perils. Ann Surg. 2018;268:70-76.

36. Ogink P, Karhade A, Thio Q, et al. Predicting discharge placement after elective surgery for lumbar spinal stenosis using machine learning methods. Eur Spine J. 2019;28:1433-1440.

37. Hopkins B, Mazmudar A, Driscoll C, Svet M, Goergene J, Kelsten M. Using artificial intelligence (AI) to predict postoperative surgical site infection: a retrospective cohort of 4046 posterior spinal fusions. Clin Neurol Neurosurg. 2020;192:105718.

38. Burns J, Yao J, Summers R. Vertebral body compression fractures and bone density: automated detection and classification on CT images. Radiology. 2017;284:788-797.

39. Karhade AV, Shah AA, Bono CM, et al. Development of machine learning algorithms for prediction of mortality in spinal epidural abscess. Spine J. 2019;19:1950-1959.

40. Kim J, Merrill R, Arvind V, et al. Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. Spine (Phila Pa 1976). 2018;43:853-860.

41. Golinvaux N, Varthi A, Bohl D, Basquees B, Grauer J. Complication rates following elective lumbar fusion in patients with diabetes: insulin dependence makes the difference. Spine (Phila Pa 1976). 2014;39:1809-1816.

42. Kazim S, Ogulnick J, Robinson M, et al. Cognitive impairment after intracerebral hemorrhage: a systematic review and meta-analysis. World Neurosurg. 2021;148:141-162.

43. Donnellan C, Werring D. Cognitive impairment before and after intracerebral hemorrhage: a systematic review. Neurol Sci. 2020;41:509-527.

44. Zhang C, Xia J, Ge H, et al. Long-term mortality related to acute kidney injury following intracerebral hemorrhage: a 10-year (2010-2019) retrospective study. J Stroke Cerebrovasc Dis. 2021;30:105688.

45. Jiminez-Almonte J, Hautala G, Abbenhaus E, et al. Spine patients demystified: what are the predictive factors of poor surgical outcome in patients after elective cervical and lumbar spine surgery. Spine J. 2020;20:1529-1534.

46. Taree A, Mikhail C, Markowitz J, et al. Risk factors for 30- and 90-day readmissions due to surgical site infection following posterior lumbar fusion. Clin Spine Surg. 2021;34:E216-E222.

47. Ilyas H, Golubovsky J, Chen J, Winkelman R, Mroz T, Steinmetz M. Risk factors for 90-day reoperation and readmission after lumbar surgery for lumbar spinal stenosis. J Neurosurg Spine. 2019;31:20-26.

48. Daniels A, Kuris E, Kleinhenz D, Palumbo M. Spine surgery outcomes in workers' compensation patients. J Am Acad Orthop Surg. 2017;25:e225-e234.

49. Gum J, Glassman S, Carreon L. Is type of compensation a predictor of outcome after lumbar fusion? Spine (Phila Pa 1976). 2013;38:443-448.

50. Carreon L, Glassman S, Kantamneni N, Mugavin M, Djurasovic M. Clinical outcomes after posterolateral lumbar fusion in workers' compensation patients. A case-control study. Spine (Phila Pa 1976). 2010;35:1812-1817.

51. Cheriyan T, Harris B, Cheriyan J, et al. Association between compensation and outcomes in spine surgery: a meta-analysis of 31 studies. Spine J. 2015;15:2564-2573.

52. Kerr Z, Marshall S, Harding H, Guskiewicz K. Nine-year risk of depression diagnosis increases with increasing self-reported concussions in retired professional football players. Am J Sport Med. 2012;40:2206-2212.

53. Fann J, Burington B, Leonetti A, Jaffe K, Katon W, Thompson R. Psychiatric illness following traumatic brain injury in an adult health maintenance organization population. Arch Gen Psychiatry. 2004;61:53-61.

54. Fann J, Ribe A, Pedersen H, et al. Long-term risk of dementia among people with traumatic brain injury in Denmark: a population-based observational cohort study. Lancete Psychiatry. 2018;5:424-431.

55. Jackson K, Rumley J, Griffith M, Agochukwu U, DeVine J. Correlating psychological comorbidities and outcomes after spine surgery. Glob Spine J. 2020;10:929-939.

56. Durand W, Johnson J, Li N, et al. Hospital competitive intensity and perioperative outcomes following lumbar spinal fusion. Spine J. 2018;18:626-631.

SUPPLEMENTARY DATA

Supplementary Table 1. ICD-10 Procedure Codes for Lumbar Spinal Fusion
ICD-10 procedure codes
Fusion of the thoracolumbar vertebral joint
0RGA070, 0RGA071, 0RGA07J, 0RGA0A0, 0RGA0AJ, 0RGA0J0, 0RGA0J1, 0RGA0JJ, 0RGA0K0, 0RGA0K1, 0RGA0KJ, 0RGA370, 0RGA371, 0RGA37J, 0RGA3A0, 0RGA3AJ, 0RGA3J0, 0RGA3J1, 0RGA3JJ, 0RGA3K0, 0RGA3K1, 0RGA3KJ, 0RGA470, 0RGA41, 0RGA47J, 0RGA4A0, 0RGA4AJ, 0RGA4J0, 0RGA4J1, 0RGA4JJ, 0RGA4K0, 0RGA4K1, 0RGA4KJ
Fusion of the lumbar vertebral joint
0SG0070, 0SG0071, 0SG007J, 0SG00A0, 0SG00AJ, 0SG00J0, 0SG001J, 0SG00JJ, 0SG00K0, 0SG00K1, 0SG00KJ, 0SG0370, 0SG0371, 0SG037J, 0SG03A0, 0SG03AJ, 0SG03J0, 0SG03J1, 0SG30JJ, 0SG03K0, 0SG03K1, 0SG03KJ, 0SG0470, 0SG0471, 0SG047J, 0SG04A0, 0SG04AJ, 0SG04J0, 0SG04J1, 0SG04JJ, 0SG04K0, 0SG04K1, 0SG04KJ
Fusion of 2 or more lumbar vertebral joints
0SG1070, 0SG1071, 0SG107J, 0SG10A0, 0SG10AJ, 0SG10J0, 0SG10J1, 0SG10JJ, 0SG10K0, 0SG10K1, 0SG10KJ, 0SG1370, 0SG1371, 0SG137J, 0SG13A0, 0SG13AJ, 0SG13J0, ASG13J1, 0SG13JJ, 0SG13K0, 0SG13K1, 0SG13KJ, 0SG1470, 0SG1471, 0SG147J, 0SG14A0, 0SG14AJ, 0SG14J0, 0SG14J1, 0SG14JJ, 0SG14JK0, 0SG14K1, 0SG14KJ
Fusion of the lumbosacral joint
0SG3070, 0SG3071, 0SG307J, 0SG30A0, 0SG30AJ, 0SG30J0, 0SG30J1, 0SG30JJ, 0SG30K0, 0SG30K1, 0SG30KJ, 0SG3370, 0SG3371, 0SG337J, 0SG33A0, 0SG33AJ, 0SG33J0, 0SG33J1, 0SG33JJ, 0SG33K0, 0SG33K1, 0SG33KJ, 0SG3470, 0SG3471, 0SG347J, 0SG34A0, 0SG34AJ, 0SG34J0, 0SG34J1, 0SG34JJ, 0SG34K0, 0SG34K1, 0SG34KJ
ICD-10, International Classification of Diseases, Tenth Revision.