J Shoulder Elbow Surg (2019) 28, e410–e421

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty

Anirudh K. Gowd, MD[a,*], Avinesh Agarwalla, MD[b], Nirav H. Amin, MD[c], Anthony A. Romeo, MD[d], Gregory P. Nicholson, MD[e], Nikhil N. Verma, MD[e], Joseph N. Liu, MD[f]

[a]*Wake Forest University Baptist Medical Center, Winston-Salem, NC, USA*
[b]*Westchester Medical Center, Valhalla, NY, USA*
[c]*Veterans Affairs Loma Linda, Loma Linda, CA, USA*
[d]*The Rothman Institute, New York, NY, USA*
[e]*Rush University Medical Center, Chicago, IL, USA*
[f]*Loma Linda University Medical Center, Loma Linda, CA, USA*

Background: We aimed to demonstrate that supervised machine learning (ML) models can better predict postoperative complications after total shoulder arthroplasty (TSA) than comorbidity indices.
Methods: The American College of Surgeons–National Surgical Quality Improvement Program database was queried from 2005-2017 for TSA cases. Training and validation sets were created by randomly assigning 80% and 20% of the data set. Included variables were age, body mass index (BMI), operative time, smoking status, comorbidities, diagnosis, and preoperative hematocrit and albumin. Complications included any adverse event, transfusion, extended length of stay (>3 days), surgical site infection, return to the operating room, deep vein thrombosis or pulmonary embolism, and readmission. Each SML algorithm was compared with one another and to a baseline model using American Society of Anesthesiologists (ASA) classification. Model strength was evaluated by calculating the area under the receiver operating characteristic curve (AUC) and the positive predictive value (PPV) of complications.
Results: We identified a total of 17,119 TSA cases. Mean age, BMI, and length of stay were 69.5 ± 9.6 years, 31.1 ± 6.8, and 2.0 ± 2.2 days. Percentage hematocrit, BMI, and operative time were of highest importance in outcome prediction. SML algorithms outperformed ASA classification models for predicting any adverse event (71.0% vs. 63.0%), transfusion (77.0% vs. 64.0%), extended length of stay (68.0% vs. 60.0%), surgical site infection (65.0% vs. 58.0%), return to the operating room (59.0% vs. 54.0%), and readmission (64.0% vs. 58.0%). SML algorithms demonstrated the greatest PPV for any adverse event (62.5%), extended length of stay (61.4%), transfusion (52.2%), and readmission (10.1%). ASA classification had a 0.0% PPV for complications.

IRB approval was waived for this investigation by the Rush University Medical Center because of the provision for responders' anonymity in the NSQIP-ACS database.

*Reprint requests: Anirudh K. Gowd, MD, 1 Medical Center Blvd, Winston-Salem, NC 27517, USA.
E-mail address: anirudhkgowd@gmail.com (A.K. Gowd).

https://doi.org/10.1016/j.jse.2019.05.017

Conclusion: With continued validation, intelligent models could calculate patient-specific risk for complications to adjust perioperative care and site of surgery.
Level of evidence: Level IV; Case-Control Design; Diagnostic Study

Keywords: Total shoulder arthroplasty; anatomic total shoulder arthroplasty; reverse total shoulder arthroplasty; complication rate; risk assessment; machine learning; neural networks

The Bundled Payments for Care Improvement (BPCI) Initiative (launched by the Center for Medicare & Medicaid Innovation) emphasizes providing value-based care to optimize surgical outcomes with lower operating costs.[21] However, allocating appropriate resources is highly intricate and dependent on many variables. Total shoulder arthroplasty, in particular, is a rapidly developing surgical procedure that can be performed in an inpatient or outpatient setting to effectively manage end-stage arthritis in a cost-effective manner.[2-4,6,24] Predicting short-term complications following this surgery is challenging but is an important step toward performing this procedure safely and minimizing hospital length of stay.

Machine learning (ML) is the science of utilizing computerized neural networks to adapt to complicated data. Regression algorithms can thereby be optimized to create predictive algorithms for use, based on collected variables. ML has a wide range of clinical applications ranging from computer-aided diagnosis, drug discovery, and modeling outcome data sets.[8,9,36] Recently, orthopedic surgery specialists have adapted this technology toward predicting complications.[16,19,22]

Both anatomic and reverse total shoulder arthroplasty (aTSA and rTSA) are safe, elective procedures with low overall morbidity.[2,30] Despite this, several patient factors have been identified that increase the risk of complications.[18,25] To provide value-based care, patient selection is vital for identifying these factors and allocating resources appropriately to minimize the economic burden when providing care.[12,26] A generalizable and predictive model specific to elective shoulder arthroplasty could be valuable in helping physicians select candidates for surgery, determine sites of operation, and appropriately allocate resources to ensure safe performance of the procedure.

The purpose of this study was to evaluate machine learning algorithms in predicting postoperative complications after total shoulder arthroplasty, determine variables most frequently used within the model, and compare the effectiveness of these models to one based solely on the American Society of Anesthesiologists (ASA) physical status classification. We hypothesize that supervised machine learning (ML) models will better predict postoperative complications after TSA than would comorbidity indices.

Methods

Data collection

Data for the present study were collected from the American College of Surgeons–National Surgical Quality Improvement Program (ACS-NSQIP). This program prospectively collects data on 274 variables from a growing network of 700+ hospitals.[20] Surgical outcomes in the database are all recorded within 30 days of the indexed surgery. Institutional participation is voluntary and exclusive to the United States. To ensure that data quality is maintained, participating institutions must employ a clinical surgical reviewer with a medical background to oversee this data collection. This database has been externally validated.[20,34] In addition, the ACS routinely performs internal audits for quality assurance.

Patient population

The ACS-NSQIP database was retrospectively queried between the years 2005-2017 for all patients who were listed under the CPT code 23472, which is used for shoulder replacement procedures that include both the humerus and glenoid, that is, total shoulder arthroplasty. The same code is used whether the patient had an anatomic total shoulder or a reverse total shoulder. The International Classification of Diseases (ICD) codes were reviewed manually. Both ICD-9 and ICD-10 codes were categorized into osteoarthritis, rheumatoid arthritis, posttraumatic/instability arthropathy, avascular necrosis, and fracture (Appendix 1). Cases that could not be grouped into any of these categories were excluded, for example, the codes corresponding to malignancy, infection, or prosthetic loosening. All available preoperative and intraoperative variables were individually analyzed for appropriateness to be included in the model. Correlations of features to one another were evaluated to determine any bias from feature selection. Features selected for analysis included age, body mass index, dependent functional status, smoking history, diabetes, hematocrit and albumin laboratory levels, history of congestive heart failure, dialysis, ascites, disseminated cancer, hypertension requiring medication, dyspnea, steroid use, renal failure, bleeding disorder, and recent weight loss (within 6 months). Cases with missing features were excluded from the study.

Predictive model design

Designated labels were extended length of stay (>3 days), surgical site infection, anemia requiring transfusion, deep vein thrombosis

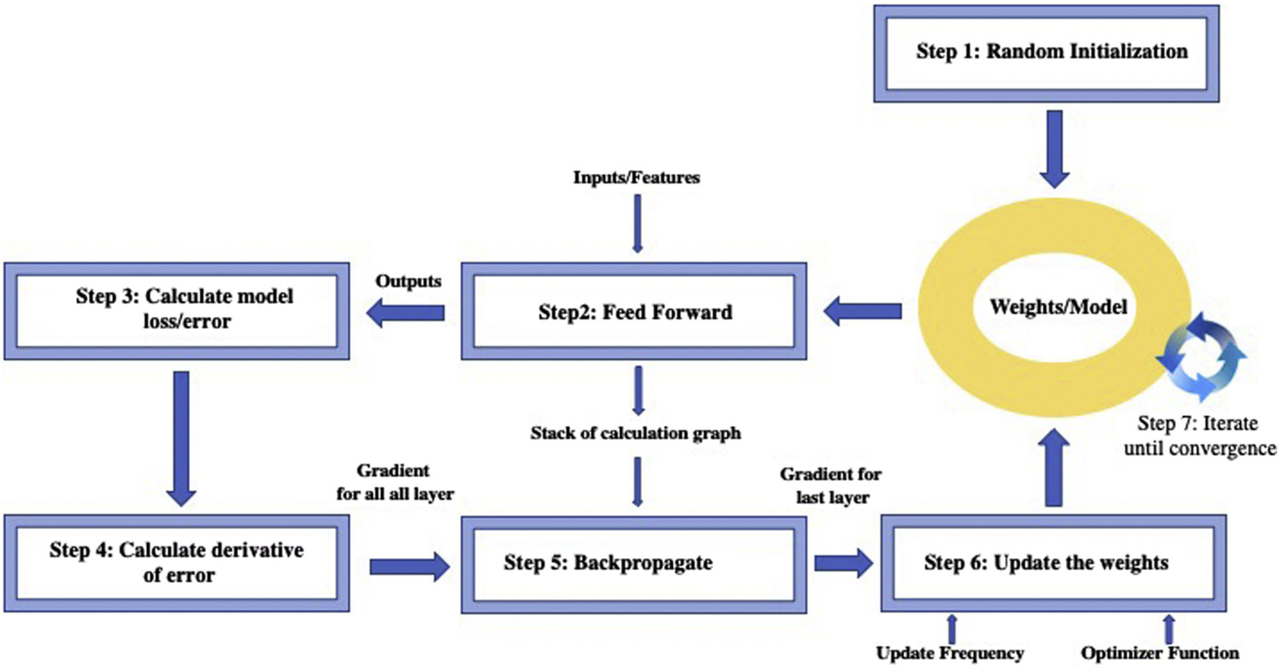

Figure 1 Generalized process of machine learning algorithms in calculation of complex models through minimization of error and optimization of accuracy.

or pulmonary embolism, return to the operating room (OR), and any adverse event. The cutoff for extended length of stay was determined as the integer greater than the 75% quartile.[13] Any adverse event included the incidence of 1 or more of the following complications within 30 days from surgery: surgical site infection, anemia requiring transfusion, deep vein thrombosis or pulmonary embolism, urinary tract infection, acute renal failure/compensation, sepsis, intubation-related event, pneumonia, myocardial infarction, cerebrovascular event, cardiac arrest, and return to the OR. Of note, incidence of readmission was recorded by the ACS-NSQIP only from 2013 onward. Preoperative and intraoperative features are as noted above. ASA classification was stratified by the commonly used 6-level classification, and Frailty Index was classified by the 5-factor modified index.[13,17]

Statistical analysis

Machine learning algorithms were created using the Scikit-Learn package of Python, version 3.7.2.[29] The overall data set was split into a training subset to design the predictive model and a validation subset to evaluate the metrics of the model. The training subset consisted of a randomly assorted 80% of the overall data set, whereas the validation set consisted of the remaining 20%. Input features were selected as above and modeled per different algorithms. Machine learning is employed through forward propagation and calculation of loss/error within the training set, and utilization of backpropagation to update model weights for each adverse event. Several layers, or neural networks, are used to increase the speed of calculation of this basic process. Finally, this general process is reiterated until an optimum accuracy is reached, and the accuracy of the model is evaluated using the validation subset (Fig. 1). The following machine learning algorithms for binary classifications were implemented and compared: logistic regression, K-nearest neighbor, random forest, Naive-Bayes, decision tree, and gradient boosting trees. Optimal hyperparameters were iteratively selected from a random pool of candidates based on the accuracy of the model. Ordinal features were incorporated into the model using 1-hot-encoding, whereas nominal variables were bucketed into 10 equally sized bins to minimize bias. Final outcomes were model accuracy, area under the receiver operating characteristic curve (AUC), and positive predictive value (PPV) in predicting every single adverse event of each machine learning algorithm with respect to all of the aforementioned labels. Accuracy was defined as the sum of true-positive (correctly predicted adverse events) and true-negative (currently predicted nonevents) predictions divided by the sample size.

Results

An initial query of the database yielded 19,185 cases of total shoulder arthroplasty. On applying the inclusion/exclusion criteria, 17,119 cases remained. The baseline characteristics of the resulting population is represented in Table I.

The average length of stay was 2.0 ± 2.2 days. The most common complication was anemia requiring transfusion (n = 471, 3%). Adverse events are further summarized in Table II.

Training and validation sets consisted of 13,697 and 3,422 cases, respectively. A total of 22 features were used to determine the optimal model in the prediction of 6 separate labels. A correlation matrix between features was

Table I Preoperative and intraoperative characteristics of the included patient population

	Mean/ Count	Standard deviation
Demographics		
Age, yr	69.5	9.6
BMI	31.1	6.8
Gender		—
Male	9626	
Female	7493	
Functional status		—
Independent	16,682	
Partially dependent	418	
Totally dependent	19	
Diagnosis		
Osteoarthritis	13,725	—
Rotator cuff arthropathy	1677	
Fracture	856	—
Posttraumatic/ dislocation arthropathy	428	—
Avascular necrosis	338	—
Rheumatoid	95	—
Comorbidities		
Smoking		—
Yes	1786	
No	15,333	
Diabetes		—
No	14,110	
Non-insulin	2148	
Insulin	861	
Congestive heart failure		—
Yes	87	
No	17,032	
Dialysis		—
Yes	63	
No	17,056	
Hypertension		—
Yes	11,615	
No	5504	
Ascites		—
Yes	2	
No	17,117	
Cancer		—
Yes	29	
No	17,090	
Dyspnea		—
At rest	58	
Exertional	1,123	
No	15,938	
History of oral steroid use		—
Yes	848	
No	16,271	
Bleeding disorder		—
Yes	462	
No	16,657	
Weight loss		—
Yes	29	
No	17,090	
Laboratory values		
Hematocrit (%)	40.5	4.3
Albumin (g/dL)	4.0	0.3
Intraoperative		
ASA classification		—
1	279	
2	7443	
3	8940	
4	456	
5	1	
Frailty Index		
0	4920	
1	8640	
2	2557	
3	709	
4	273	
5	20	
Operative time (min)	111.2	44.0
Anesthesia		—
General	16,544	
Regional	351	
MAC/IV	145	
Other	79	

BMI, body mass index; *ASA*, American Society of Anesthesiologists; MAC/IV, monitored anesthesia care/intravenous.

Table II Incidence of adverse events following total shoulder arthroplasty cases performed between 2005 and 2017

	Count (n)	Incidence (%)
Surgical site infection	63	0.37
Anemia requiring transfusion	471	2.75
Pneumonia	82	0.48
Deep vein thrombosis	59	0.34
Pulmonary embolism	55	0.32
Urinary tract infection	128	0.75
Cerebrovascular accident	18	0.11
Cardiac arrest	13	0.08
Myocardial infarction	38	0.22
Return to OR	187	1.09
Any adverse event*	873	5.10
Extended length of stay†	3,641	21.27
Readmission‡	408	2.85

OR, operating room.
* Count of any adverse event were based on patients so that 1 patient with multiple adverse events during admission will still be counted once.
† Extended length of stay defined as 3 days or greater.
‡ Readmission only collected from 2013 onward.

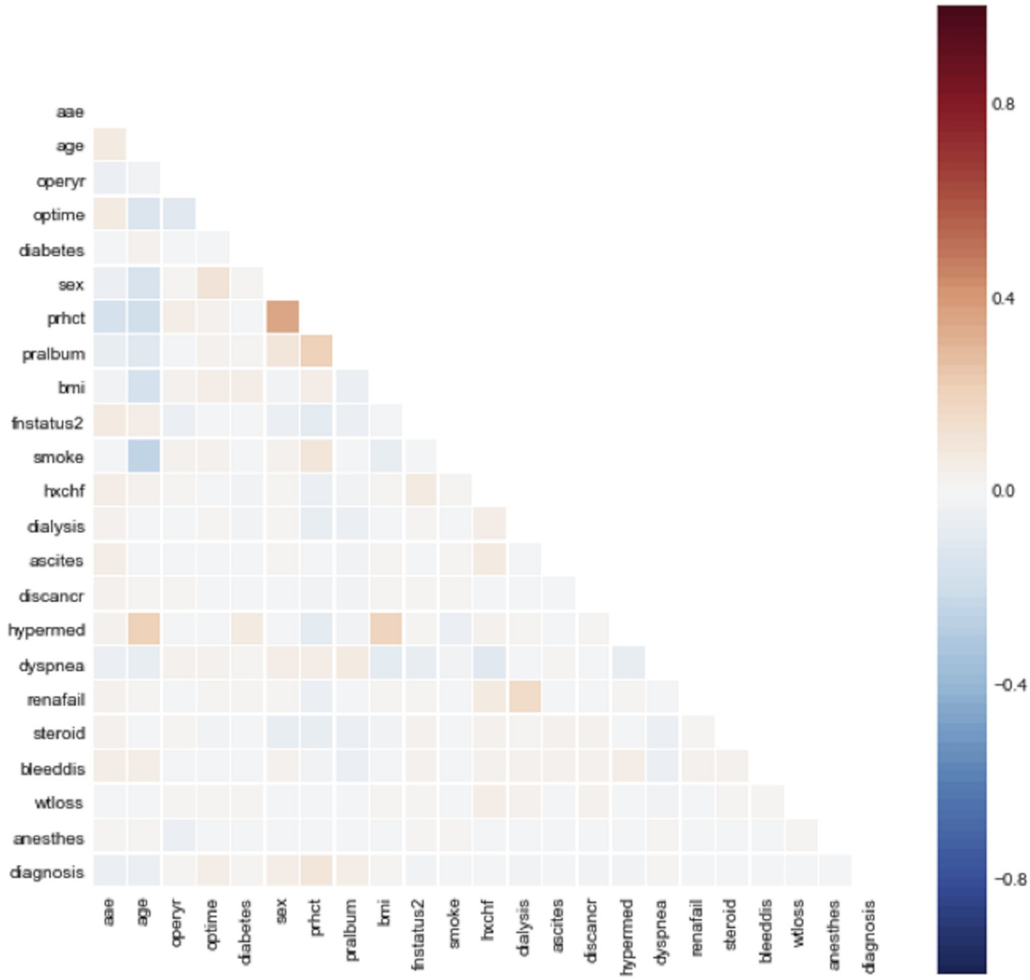

Figure 2 Correlation matrix of features included within machine learning algorithms. No features were excluded as a result of correlation. Parallel correlation of variables is denoted by reddish hue, minimal correlation is noted by gray, and antiparallel correlation is noted by blue.

built to ensure the selected features were not intercorrelated (Fig. 2).

Each machine learning algorithm was constructed with the training data set and cross-validated with the validation set. Algorithms were trained to minimize error and adjusted corresponding feature weights to improve accuracy. Importance of model weights for each label are ordered in Figure 3. Percentage hematocrit, body mass index, and operative time were weighted for highest importance in the predictive models of all 6 labels. Percentage hematocrit was the highest weighted for prediction of any adverse event, extended length of stay, and anemia requiring transfusion. Body mass index was the highest-weighted feature for prediction of surgical site infection, return to OR, and deep vein thrombosis or pulmonary embolism.

The accuracy and the AUC were evaluated with respect to each model. For both any adverse event and transfusion rate, the random forest classifier had the greatest accuracy (95.4% and 95.6%, respectively), whereas logistic regression had the greatest AUC (71.0% and 77%, respectively). For extended length of stay, logistic regression classifier had the greatest accuracy (82.3%) and gradient boosting trees had the greatest AUC (68%). The accuracies of the models predicting surgical site infections, return to OR, and deep vein thrombosis or pulmonary embolism (DVT/PE) were not distinguishable (Table III). Logistic regression classifier had the greatest AUC for the prediction of surgical site infection and return to OR (65.0% and 59.0%, respectively). Random forest classifier had the greatest AUC for prediction of DVT/PE (58.0%) (Fig. 4). Logistic regression classifier had the greatest AUC for prediction of unplanned readmission (64.0%). Models constructed from ASA classification alone had an AUC of 63.0% for any adverse event, 60.0% for extended length of stay, 64.0% for transfusion, 58.0% for surgical site infection, 54.0% for return to OR, and 60.0% for DVT/PE. Models constructed from Frailty Index alone had an AUC of 58.0% for any adverse event, 58.0% for extended length of stay, 58.0% for transfusion, 41.0% for surgical site infection, 54.0% for return to OR, and 43.0% for DVT/PE.

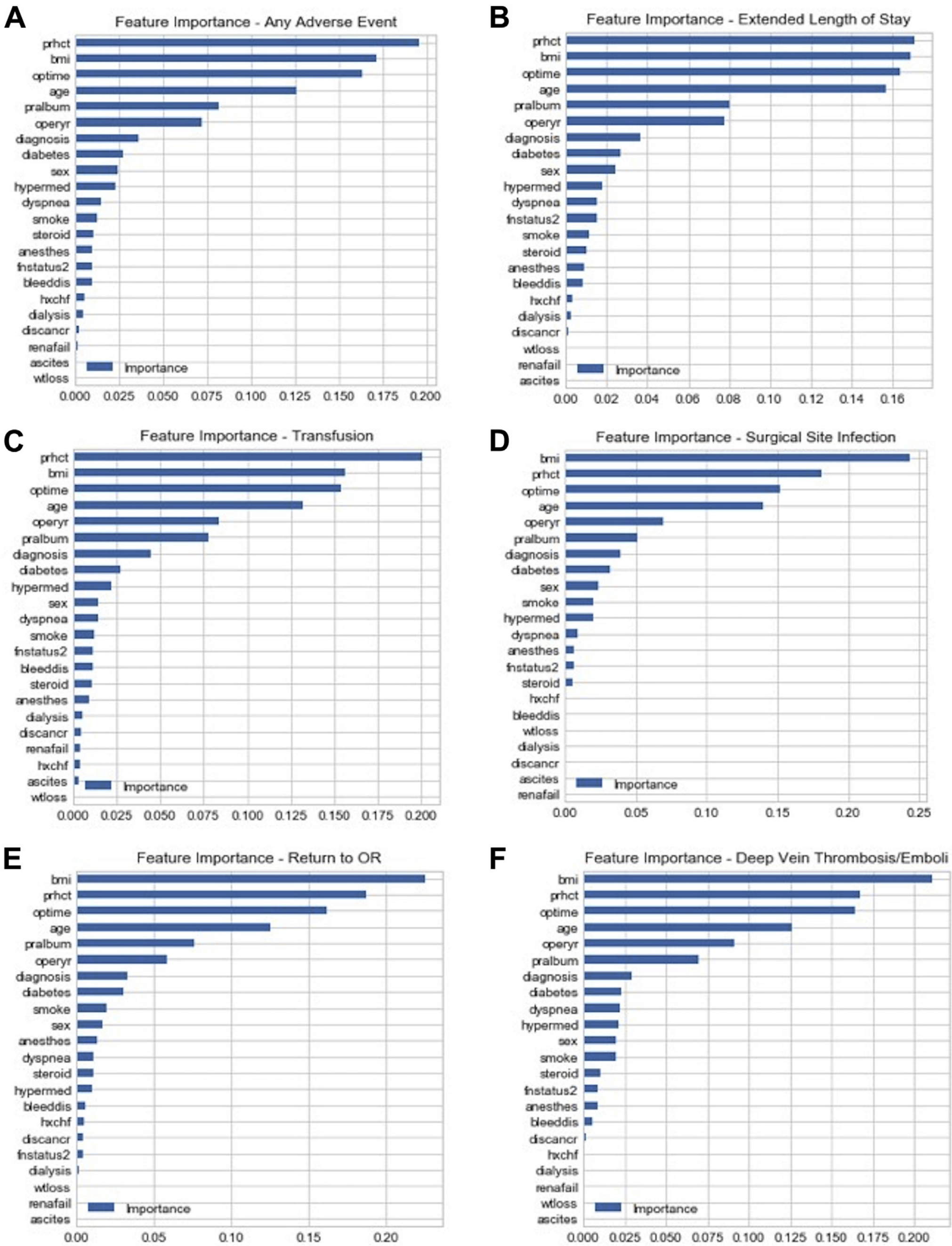

Figure 3 Feature importance as determined through a random forest classifier for prediction of (**A**) any adverse event, (**B**) extended length of stay, (**C**) anemia requiring transfusion, (**D**) surgical site infection, (**E**) return to the OR, and (**F**) deep vein thrombosis or pulmonary embolism. Prhct: preoperative hematocrit; Bmi: body mass index; optime: operative time; operyr: year of indexed surgery; pralbum: preoperative albumin; hypermed: hypertension requiring medication; fnstatus2: functional dependency; anesthes: anesthesia modality; bleeddis: bleeding disorder; discancr: disseminated cancer; hxchf: history of congestive heart failure; renafail: renal failure; wtloss: greater than 10% weight loss within 6 months. *OR*, operating room.

The models were further evaluated through confusion matrices (Fig. 5). Of note, both the random forest and AUC alone model predicted 0 adverse events (true positives) within our patient population. The decision tree model had the fewest false negatives (222), followed by gradient-based trees (236), and logistic regression (237).

Acceptable models (at least 1 model with an AUC >70%) were evaluated further for the PPV of the model. Logistic regression had the greatest PPV for prediction of any adverse event, random forest for extended length of stay, and gradient-based trees for anemia requiring transfusion (Table IV). Conversely, ASA classification and

Table III Evaluation of machine learning algorithm accuracy (correctly predicted adverse events and nonevents divided by sample size) in prediction of each label

	AAE	Extended	Transfus.	SSI	Ret. OR	DVT/PE
Logistic regression (%)	95.3	**82.3**	95.3	99.6	99.2	99.4
Gradient-boosting trees (%)	95.3	82.1	95.0	99.5	99.2	99.4
Random forest (%)	**95.4**	82.1	**95.6**	99.6	99.2	99.4
K-nearest neighbors (%)	95.1	78.5	95.5	99.6	99.2	99.4
Decision tree (%)	92.1	74.9	87.8	99.2	99.2	99.2
Naïve Bayes (%)	6.0	20.8	6.4	13.0	36.0	38.1

AAE, any adverse event; *Transfus*, transfusion; *SSI*, surgical site infection; *Ret. OR*, return to operating room; *DVT/PE*, deep vein thrombosis or pulmonary embolism.
Boldface indicates model with greatest accuracy for each label.

Frailty Index alone had no predictive value for any of the previously mentioned events.

Discussion

Using variables that are routinely collected by the participating hospitals of the ACS-NSQIP database, our machine learning algorithms were able to create accurate models for the prediction of any adverse event, anemia requiring transfusion, and extended length of stay. Although the present models are limited in their prediction by the features collected by this national database, the best model for each of the above complications had an AUC greater than 70% and a PPV of 50% or greater. The remaining complication rates (surgical site infection, DVT/PE, and reoperation) had limited predictive value from these models, which may be due to the limited incidence of each complication or contributing factor outside the collected features. In comparison, the ASA comorbidity index provided close to acceptable models (based on AUC analysis) for any adverse event, transfusion, and extended length of stay, but had no PPV of any of the above events. Given the low incidence of complication, acceptable and accurate models were achieved by predicting no adverse event consistently; however, the correct prediction of adverse events (greater PPV) was only achieved by ML models using all data features. By implementing this ML algorithm in preoperative evaluation, surgeons can gain an acceptable representation of operative risk. Although not comprehensive, this quick risk calculation may provide information that can be used to make clinically informed decisions regarding short-term complication risks in patients undergoing TSA. For example, the operative risk profile can be used to assist in matching patients to the most appropriate cost-effective site of care for total shoulder arthroplasty: hospital inpatient, hospital outpatient, or ambulatory surgery center.

Performance of TSA within outpatient ambulatory surgical centers has become increasingly popular.[6] When performed in the outpatient setting, cost is significantly decreased from the patient perspective.[14] However, these benefits must be weighed with the greater risk burden should complications arise without the infrastructure of an inpatient hospital. In a recent survey of ASES members, Brolin et al demonstrated that the largest barrier to performing surgery in ambulatory surgical centers was due to the concern for development of medical complications.[4] Additionally, surgeons incur additional financial risk with partial ownership of the ambulatory surgical center should a complication occur. Patient selection and optimization has therefore become increasingly important for valuable outcomes following surgery.[12] Risk stratification remains elusive to both surgeons practicing in the inpatient setting and those in the outpatient setting, and there is a need to improve upon this as the Bundled Payments for Care Initiative policy spreads.[5] Comorbidity indices, including the ASA classification, Charlson Comorbidity Index, and Frailty Index remain common tools associated with increased rates of complication; however, they do not possess the sensitivity to predict when adverse events may occur.[13,17,23] As the present study demonstrates, ML has the advantage of utilizing all collected preoperative variables to formulate accurate positive predictions of adverse events despite those events' low occurrence.

The present study emphasizes anemia requiring transfusion as the most common cause of short-term complication following TSA. Heavy blood loss, subsequent anemia, and transfusion will also increase patient length of stay because of the need to monitor for acute reactions. Previous research has cited a growing incidence of transfusion rate in TSA between 2005 and 2009 from 4.9% to 7.1% using the National Inpatient Sample database while an institutional database found incidence to be 11.3%.[1,32] Grier and coauthors matched patients receiving transfusion to a regular TSA cohort based on age and gender, and found that the transfusion cohort had statistically greater rates of myocardial infarction, pneumonia, systemic inflammatory response syndrome or sepsis, venous thromboembolic events, and cerebrovascular accidents at all time points, and periprosthetic infection and mechanical complications up to

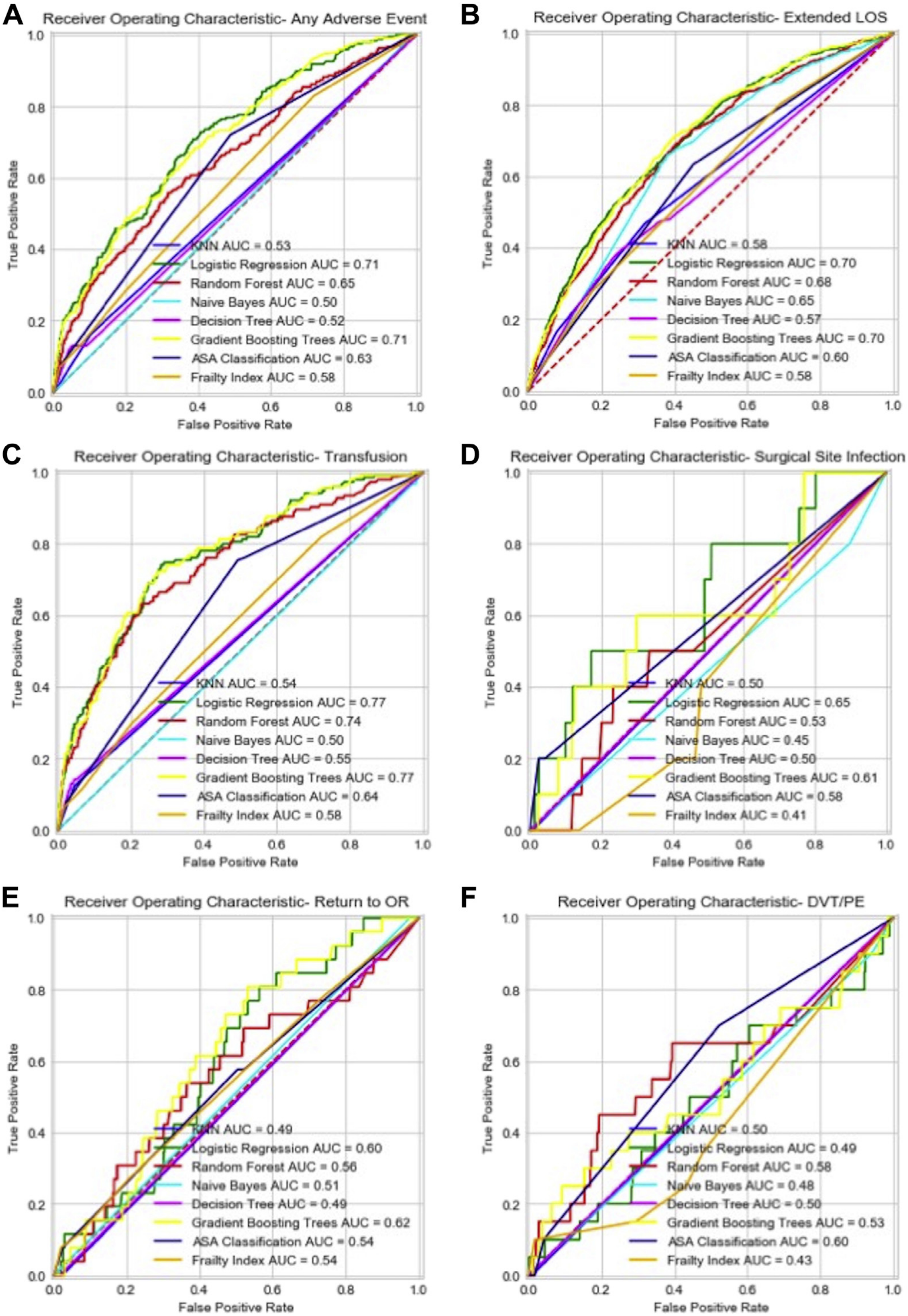

Figure 4 Receiver operating characteristic (ROC) curve with area under curve calculation (legend) of each machine learning algorithm in the prediction of (**A**) any adverse event, (**B**) extended LOS, (**C**) anemia requiring transfusion, (**D**) surgical site infection, (**E**) return to OR, and (**F**) DVT/PE. *KNN*, K-nearest neighbor; *AUC*, area under the ROC curve; *ASA*, American Society of Anesthesiologists; *OR*, operating room; *LOS*, length of stay; *DVT/PE*, deep vein thrombosis or pulmonary embolism.

2 years postoperatively.[10,15] In corroboration with these findings, it is not surprising that percentage hematocrit was a heavily weighted feature to predict any adverse event, extended length of stay, and transfusion within the present study. Additionally, Padegimas et al classified a threshold hematocrit of less than 39.6% that had a 90% sensitivity of predicting need for intraoperative transfusion and an 11.0% incidence of transfusion.[28] Potentially biasing this finding

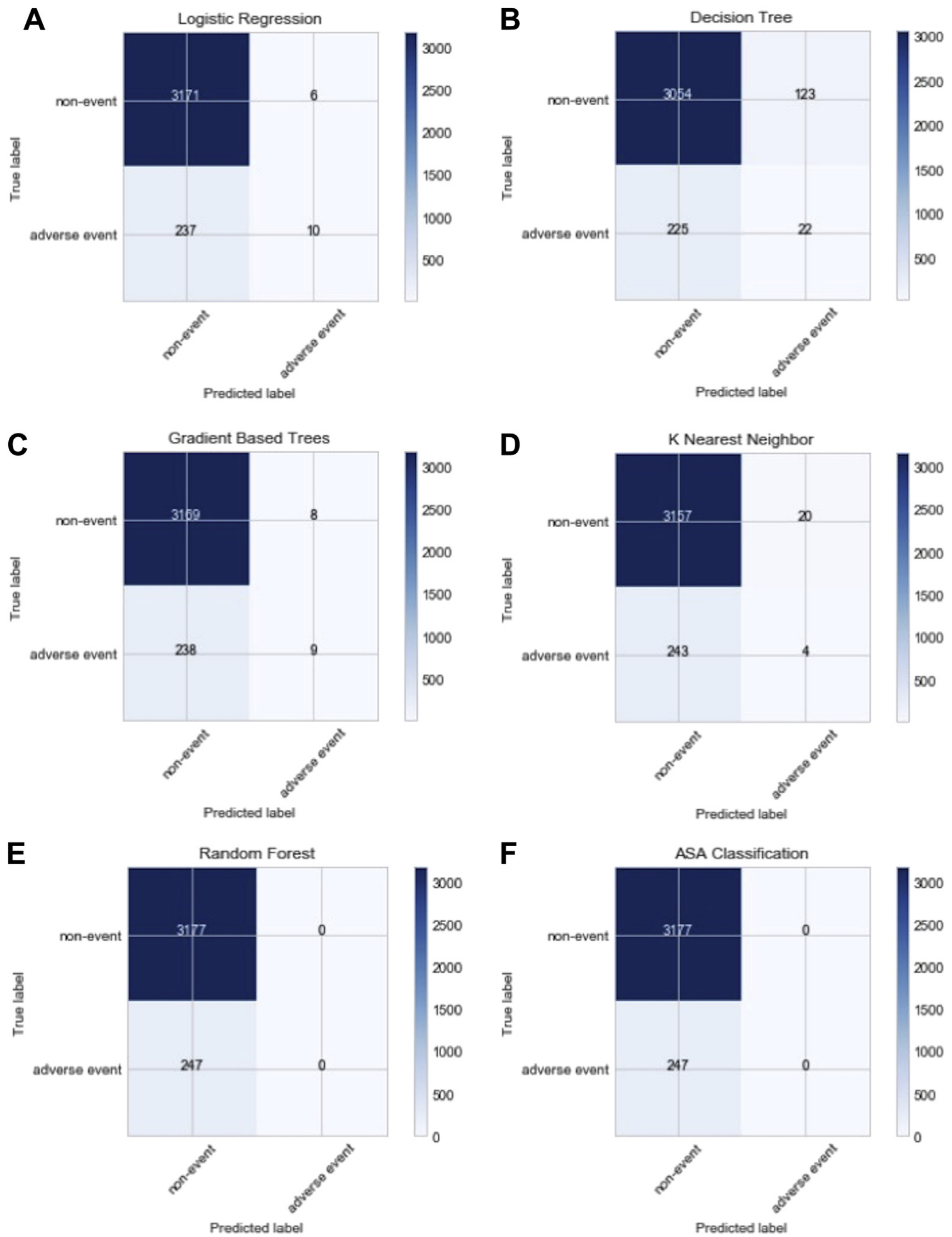

Figure 5 Confusion matrix demonstrating true-positive (bottom right), true-negative (top left), false-positive (top right), and false-negative (bottom left) predictions of any adverse event using each trained model on the validation data set using (**A**) logistic regression, (**B**) decision tree, (**C**) gradient-based trees, (**D**) K-nearest neighbor, (**E**) random forest, and (**F**) logistic regression with ASA classification alone. Within each matrix, the top left represents correct prediction of no adverse event (true negative), the bottom right represents correct prediction of adverse event (true positive), the top right is incorrect prediction of adverse event (false positive), and the bottom left is incorrect prediction of no adverse event (false negative). *ASA*, American Society of Anesthesiologists.

is the low percentage hematocrit of this population cohort (40.5% $\pm$ 4.5%). ML offers the additional advantage of combining other patient comorbidities and demographic variables to stratify risk for transfusion. With this information, surgeons may make better-informed decisions on surgical site or need for interventions in limiting post-operative bleeding. Patient optimization or use of tranexamic acid may be a cost conscious method of counteracting high risk in patients with greater likelihood of transfusion.[7,33,35]

Table IV Positive predictive value in prediction of adverse events in acceptable models toward prediction of any adverse event (AAE), extended length of stay (LOS), and need for transfusion

	AAE	Extended LOS	Transfusion
Logistic regression	**62.5**	61.0	40.0
Decision tree	17.0	40.2	12.7
Gradient-based trees	57.9	59.4	**52.2**
Gaussian Bayes	7.20	30.3	4.5
K-nearest neighbors	16.7	45.2	22.2
Random forest	0	**61.4**	0
ASA class alone	0	0	0
Frailty Index alone	0	0	0

ASA, American Society of Anesthesiologists.
Boldface indicates greatest positive predicted value for respective adverse event.

The ASA classification is routinely used to assess operative risk and has been significantly correlated with complication rates following TSA.[11,25] The ASA classification is also routinely used as part of the assessment for the eligibility of navigating the patient to an outpatient setting. Although it is logical that greater comorbidities carry greater risks of complication, the present study found that the ASA classification lacked the granularity by which to predict adverse events as did a model constructed from numerous features. Although the ASA classification is simple and heavily used, the reliability of this metric is only fair by interclass correlation within anesthetists.[27] Recently, Fu et al corroborated similar findings that existing comorbidity indices could only provide moderate associations to complication rate.[13] Alternatively, the Elixhauser Comorbidity Measure, constructed from ICD diagnosis codes, has been shown to be an excellent model of prediction for death, extended length of stay, and nonroutine discharge.[23] The generalizability of this model may be limited by correct and routine classification of diagnosis codes. In addition, the count of many severe comorbidities (congestive heart failure, dialysis, cancer, ascites, dyspnea at rest, and weight loss within 6 months) was less than 100 within this population cohort. Therefore, these comorbidities had limited utility to the model, but that does not necessarily indicate that comorbidities are not associated with adverse events, as has been previously demonstrated.[23,31] In addition, predictive models from comorbidity indices take advantage of the low complication rate following TSA. These indices may simply predict lack of complication with reasonable accuracy yet have no positive prediction of adverse events. Comorbidity indices, therefore, have limited utility in risk assessment. Machine learning algorithms, though in their infancy, demonstrate an appreciable method to correctly predict the incidence of these adverse events.

ML algorithms are increasingly becoming a recognizable method for improved assessment of preoperative risk within orthopedics.[16,19,22] Neural network models used for prediction of adverse events in hip and knee arthroplasty have achieved acceptable AUC values (>70%) for prediction of renal, cardiac, and death.[16] Although the current population cohort lacked the count of these severe complications, the present optimal model achieved a greater AUC for the prediction of any adverse event (71% vs. 64%). Using a gradient-boosting algorithm, another study was able to predict readmissions following laminectomy with an AUC of 80.6%.[19] In comparing neural network models to non-ML logistic regression and ASA classification alone, Kim et al also found superiority within the neural network model in prediction of complication rates.[22] With respect to recent ML studies, the strength of the present study is within the selection of several ML algorithms to determine which model best predicts postoperative complication risk. Although a logistic regression algorithm was best fitted for prediction of any adverse event and extended length of stay, the gradient-based trees model had superior PPV for prediction of transfusion. Continued validation and research of ML models may allow for the implementation of risk calculation to make more informed decisions regarding the optimization of the patient's health status before the surgical procedure, and the appropriateness of patient-matching to the outpatient surgical environment.

The ability of the present study to identify all possible key indicators of adverse events is largely limited by the features collected by the ACS-NSQIP database. The ACS-NSQIP is largely deidentified and, therefore, several variables are unavailable for inclusion in the model, such as use of tranexamic acid, surgeon experience, implants used, patient medications, and/or changes in surgical techniques. Importantly, complications and clinically relevant variables were restricted to those reported by this database. Additionally, anatomic TSA was not able to be differentiated from reverse TSA because of overlapping CPT codes, as were revision cases prior to 2013. To limit bias from coding, diagnosis codes were reviewed, and patient cases were excluded if not belonging to the above categories. In addition, the present study is retrospective in nature, which limits the control of variables. Institutions included within the ACS-NSQIP database must staff clinical reviewers for quality assurance of the data. Smaller, independent surgical centers are unlikely to be represented within the database. Selection bias therefore exists from patient cases selected to be a part of the ACS-NSQIP database.

Conclusion

Machine learning accurately predicted postoperative complications in a random sample of a national cohort based on routinely collected preoperative variables and outperformed models by comorbidity indices alone.

With continued validation, intelligent models may be used to calculate patient risk of complication and adjust perioperative care and site of surgery accordingly.

Disclaimer

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jse.2019.05.017

References

1. Ahmadi S, Lawrence TM, Sahota S, Schleck CD, Harmsen WS, Cofield RH, et al. The incidence and risk factors for blood transfusion in revision shoulder arthroplasty: our institution's experience and review of the literature. J Shoulder Elbow Surg 2014;23:43-8. https://doi.org/10.1016/j.jse.2013.03.010
2. Basques BA, Erickson BJ, Leroux T, Griffin JW, Frank RM, Verma NN, et al. Comparative outcomes of outpatient and inpatient total shoulder arthroplasty: an analysis of the Medicare dataset. Bone Joint J 2017;99-B:934-8. https://doi.org/10.1302/0301-620X.99B7.BJJ-2016-0976.R1
3. Bhat SB, Lazarus M, Getz C, Williams GRJ, Namdari S. Economic decision model suggests total shoulder arthroplasty is superior to hemiarthroplasty in young patients with end-stage shoulder arthritis. Clin Orthop Relat Res 2016;474:2482-92. https://doi.org/10.1007/s11999-016-4991-0
4. Brolin TJ, Cox RM, Zmistowski BM, Namdari S, Williams GR, Abboud JA. Surgeons' experience and perceived barriers with outpatient shoulder arthroplasty. J Shoulder Elbow Surg 2018;27:S82-7. https://doi.org/10.1016/j.jse.2018.01.018
5. Bundled Payments for Care Improvement (BPCI) Initiative: general information. https://innovation.cms.gov/initiatives/bundled-payments/. Accessed November 4, 2019
6. Cancienne JM, Brockmeier SF, Gulotta LV, Dines DM, Werner BC. Ambulatory total shoulder arthroplasty: a comprehensive analysis of current trends, complications, readmissions, and costs. J Bone Joint Surg Am 2017;99:629-37. https://doi.org/10.2106/JBJS.16.00287
7. Cvetanovich GL, Fillingham YA, O'Brien M, Forsythe B, Cole BJ, Verma NN, et al. Tranexamic acid reduces blood loss after primary shoulder arthroplasty: a double-blind, placebo-controlled, prospective, randomized controlled trial. JSES Open Access 2018;2:23-7. https://doi.org/10.1016/j.jses.2018.01.002
8. El Naqa I, Ruan D, Valdes G, Dekker A, McNutt T, Ge Y, et al. Machine learning and modeling: data, validation, communication challenges. Med Phys 2018;45:e834-40. https://doi.org/10.1002/mp.12811
9. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. Radiographics 2017;37:505-15. https://doi.org/10.1148/rg.2017160130
10. Everhart JS, Bishop JY, Barlow JD. Medical comorbidities and perioperative allogeneic red blood cell transfusion are risk factors for surgical site infection after shoulder arthroplasty. J Shoulder Elbow Surg 2017;26:1922-30. https://doi.org/10.1016/j.jse.2017.04.006
11. Foruria AM, Sperling JW, Ankem HK, Oh LS, Cofield RH. Total shoulder replacement for osteoarthritis in patients 80 years of age and older. J Bone Joint Surg Br 2010;92:970-4. https://doi.org/10.1302/0301-620X.92B7.23671
12. Fournier MN, Brolin TJ, Azar FM, Stephens R, Throckmorton TW. Identifying appropriate candidates for ambulatory outpatient shoulder arthroplasty: validation of a patient selection algorithm. J Shoulder Elbow Surg 2019;28:65-70. https://doi.org/10.1016/j.jse.2018.06.017
13. Fu MC, Ondeck NT, Nwachukwu BU, Garcia GH, Gulotta LV, Verma NN, et al. What Associations exist between comorbidity indices and postoperative adverse events after total shoulder arthroplasty? Clin Orthop Relat Res 2019;477:881-90. https://doi.org/10.1097/CORR.0000000000000624
14. Gregory JM, Wetzig AM, Wayne CD, Bailey L, Warth RJ. Quantification of patient-level costs in outpatient total shoulder arthroplasty. J Shoulder Elbow Surg 2019;28:1066-73. https://doi.org/10.1016/j.jse.2018.10.006
15. Grier AJ, Bala A, Penrose CT, Seyler TM, Bolognesi MP, Garrigues GE. Analysis of complication rates following perioperative transfusion in shoulder arthroplasty. J Shoulder Elbow Surg 2017;26:1203-9. https://doi.org/10.1016/j.jse.2016.11.039
16. Harris AHS, Kuo AC, Weng Y, Trickey AW, Bowe T, Giori NJ. Can machine learning methods produce accurate and easy-to-use prediction models of 30-day complications and mortality after knee or hip arthroplasty? Clin Orthop Relat Res 2019;477:452-60. https://doi.org/10.1097/CORR.0000000000000601
17. Holzgrefe RE, Wilson JM, Staley CA, Anderson TL, Wagner ER, Gottschalk MB. Modified frailty index is an effective risk-stratification tool for patients undergoing total shoulder arthroplasty. J Shoulder Elbow Surg 2019;28:1232-40. https://doi.org/10.1016/j.jse.2018.12.004
18. Jiang JJ, Toor AS, Shi LL, Koh JL. Analysis of perioperative complications in patients after total shoulder arthroplasty and reverse total shoulder arthroplasty. J Shoulder Elbow Surg 2014;23:1852-9. https://doi.org/10.1016/j.jse.2014.04.008
19. Kalagara S, Eltorai AEM, Durand WM, DePasse JM, Daniels AH. Machine learning modeling for predicting hospital readmission following lumbar laminectomy. J Neurosurg Spine 2018;30:344-52. https://doi.org/10.3171/2018.8.SPINE1869
20. Khuri SF. The NSQIP: a new frontier in surgery. Surgery 2005;138:837-43. https://doi.org/10.1016/j.surg.2005.08.016
21. Kim K, Iorio R. The 5 clinical pillars of value for total joint arthroplasty in a bundled payment paradigm. J Arthroplasty 2017;32:1712-6. https://doi.org/10.1016/j.arth.2017.02.012
22. Kim JS, Merrill RK, Arvind V, Kaji D, Pasik SD, Nwachukwu CC, 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;43:853-60. https://doi.org/10.1097/BRS.0000000000002442
23. Kim C-Y, Sivasundaram L, LaBelle MW, Trivedi NN, Liu RW, Gillespie RJ. Predicting adverse events, length of stay, and discharge disposition following shoulder arthroplasty: a comparison of the Elixhauser Comorbidity Measure and Charlson Comorbidity Index. J Shoulder Elbow Surg 2018;27:1748-55. https://doi.org/10.1016/j.jse.2018.03.001
24. Leroux TS, Basques BA, Frank RM, Griffin JW, Nicholson GP, Cole BJ, et al. Outpatient total shoulder arthroplasty: a population-based study comparing adverse event and readmission rates to inpatient total shoulder arthroplasty. J Shoulder Elbow Surg 2016;25:1780-6. https://doi.org/10.1016/j.jse.2016.04.006
25. Leschinger T, Raiss P, Loew M, Zeifang F. Total shoulder arthroplasty: risk factors for intraoperative and postoperative complications in

patients with primary arthritis. J Shoulder Elbow Surg 2017;26:e71-7. https://doi.org/10.1016/j.jse.2016.08.001

26. Luzzi AJ, Fleischman AN, Matthews CN, Crizer MP, Wilsman J, Parvizi J. The "Bundle Busters": incidence and costs of postacute complications following total joint arthroplasty. J Arthroplasty 2018; 33:2734-9. https://doi.org/10.1016/j.arth.2018.05.015
27. Mak PHK, Campbell RCH, Irwin MG. The ASA Physical Status Classification: inter-observer consistency. American Society of Anesthesiologists. Anaesth Intensive Care 2002;30:633-40. https://doi.org/10.1177/0310057X0203000516
28. Padegimas EM, Clyde CT, Zmistowski BM, Restrepo C, Williams GR, Namdari S. Risk factors for blood transfusion after shoulder arthroplasty. Bone Joint J 2016;98-B:224-8. https://doi.org/10.1302/0301-620X.98B2.36068
29. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-30.
30. Ponce BA, Oladeji LO, Rogers ME, Menendez ME. Comparative analysis of anatomic and reverse total shoulder arthroplasty: in-hospital outcomes and costs. J Shoulder Elbow Surg 2015;24:460-7. https://doi.org/10.1016/j.jse.2014.08.016
31. Rosas S, Sabeh KG, Buller LT, Law TY, Kalandiak SP, Levy JC. Comorbidity effects on shoulder arthroplasty costs analysis of a nationwide private payer insurance data set. J Shoulder Elbow Surg 2017;26:e216-21. https://doi.org/10.1016/j.jse.2016.11.044
32. Ryan DJ, Yoshihara H, Yoneoka D, Zuckerman JD. Blood transfusion in primary total shoulder arthroplasty: incidence, trends, and risk factors in the United States from 2000 to 2009. J Shoulder Elbow Surg 2015;24:760-5. https://doi.org/10.1016/j.jse.2014.12.016
33. Sun C-X, Zhang L, Mi L-D, Du G-Y, Sun X-G, He S-W. Efficiency and safety of tranexamic acid in reducing blood loss in total shoulder arthroplasty: A systematic review and meta-analysis. Medicine (Baltimore) 2017;96:e7015. https://doi.org/10.1097/MD.0000000000007015
34. Trickey AW, Wright JM, Donovan J, Reines HD, Dort JM, Prentice HA, et al. Interrater reliability of hospital readmission evaluations for surgical patients. Am J Med Qual 2017;32:201-7. https://doi.org/10.1177/1062860615623854
35. Vara AD, Koueiter DM, Pinkas DE, Gowda A, Wiater BP, Wiater JM. Intravenous tranexamic acid reduces total blood loss in reverse total shoulder arthroplasty: a prospective, double-blinded, randomized, controlled trial. J Shoulder Elbow Surg 2017;26:1383-9. https://doi.org/10.1016/j.jse.2017.01.005
36. Zhang L, Tan J, Han D, Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug Discov Today 2017;22:1680-5. https://doi.org/10.1016/j.drudis.2017.08.010