

# Development of a Machine Learning Algorithm for Prediction of Complications and Unplanned Readmission Following Reverse Total Shoulder Arthroplasty

Journal of Shoulder and Elbow  
Arthroplasty  
Volume: 1–7  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI:10.1177/24715492211038172  
journals.sagepub.com/home/sea



Sai K. Devana, MD<sup>1</sup> , Akash A. Shah, MD<sup>1</sup>, Changhee Lee, BS<sup>2</sup> ,  
Varun Gudapati, MD<sup>1</sup>, Andrew R. Jensen, MD<sup>1</sup>, Edward Cheung,  
MD<sup>1</sup>, Carlos Solorzano, BS<sup>1</sup> , Mihaela van der Schaar, PhD<sup>2,3</sup> and  
Nelson F. SooHoo, MD<sup>1</sup>

## Abstract

**Background:** Reverse total shoulder arthroplasty (rTSA) offers tremendous promise for the treatment of complex pathologies beyond the scope of anatomic total shoulder arthroplasty but is associated with a higher rate of major postoperative complications. We aimed to design and validate a machine learning (ML) model to predict major postoperative complications or readmission following rTSA.

**Methods:** We retrospectively reviewed California's Office of Statewide Health Planning and Development database for patients who underwent rTSA between 2015 and 2017. We implemented logistic regression (LR), extreme gradient boosting (XGBoost), gradient boosting machines, adaptive boosting, and random forest classifiers in Python and trained these models using 64 binary, continuous, and discrete variables to predict the occurrence of at least one major postoperative complication or readmission following primary rTSA. Models were validated using the standard metrics of area under the receiver operating characteristic (AUROC) curve, area under the precision–recall curve (AUPRC), and Brier scores. The key factors for the top-performing model were determined.

**Results:** Of 2799 rTSAs performed during the study period, 152 patients (5%) had at least 1 major postoperative complication or 30-day readmission. XGBoost had the highest AUROC and AUPRC of 0.681 and 0.129, respectively. The key predictive features in this model were patients with a history of implant complications, protein-calorie malnutrition, and a higher number of comorbidities.

**Conclusion:** Our study reports an ML model for the prediction of major complications or 30-day readmission following rTSA. XGBoost outperformed traditional LR models and also identified key predictive features of complications and readmission.

## Keywords

reverse total shoulder arthroplasty, machine learning, complications prediction model

received 7 May 2021; revised received 21 June 2021; accepted 20 July 2021

## Introduction

Reverse total shoulder arthroplasty (rTSA) uses a spherical glenoid and concave humeral prosthesis to provide the deltoid with a mechanical advantage in patients with a deficient or damaged rotator cuff.<sup>1</sup> The altered biomechanics of rTSA enable the treatment of complex shoulder pathologies including primary glenohumeral arthritis with significant

<sup>1</sup>David Geffen School of Medicine UCLA, Los Angeles, CA, USA

<sup>2</sup>University of California, Los Angeles, CA, USA

<sup>3</sup>University of Cambridge, Cambridge, UK

### Corresponding Author:

Sai K Devana, Department of Orthopaedic Surgery, David Geffen School of Medicine UCLA, 10833 LeConte Ave. (76-119 CHS), Los Angeles, CA 90095-6902, USA.

Email: skdevana@gmail.com



glenoid deformity, rotator cuff arthropathy, pseudoparalysis due to rotator cuff tear, proximal humerus fractures, tumor and rheumatoid arthritis.<sup>1-4</sup> The incidence of rTSA has increased exponentially over the past decade with some models projecting a 90% to 350% growth in rTSA volume between 2017 and 2025.<sup>5</sup> Though rTSA has enabled the treatment of conditions beyond the scope of traditional anatomic total shoulder arthroplasty (aTSA), the surgery is associated with a particularly high rate of postoperative complications, often reported as greater than 15%.<sup>6-11</sup> As the incidence and indications for rTSA continue to rise and expand, research efforts have been directed towards identifying preoperative risk factors that may help identify patients at risk for perioperative complications. Despite successful attempts to study singular factors such as operative indication or surgeon experience, there is still a need for accurate comprehensive risk stratification to meaningfully improve safety and lower costs associated with rTSA.<sup>7,12,13</sup>

The subjective and multifactorial nature of surgical outcomes has long been a challenge for clinical research in orthopedic surgery. Machine learning (ML) methods offer a means of accounting for many variables and identifying nonlinear relationships between these factors that can overwhelm traditional regression techniques. Following early adoption for radiographic purposes, ML has been increasingly applied to clinical questions.<sup>14,15</sup> The growing availability of accessible data sets and ML models, including a variety of ensemble methods that promise greater accuracy and efficiency, offers an opportunity to analyze complex problems like postoperative outcomes from a novel and holistic perspective.<sup>16-20</sup>

The aims of this study are to (1) implement an ML model to predict patients at risk of at least one major postoperative complication or 30-day readmission for any cause following rTSA, (2) compare the performance of our model to a traditional logistic regression (LR) model, and (3) compare which features have the most predictive power between our most accurate ML model and LR. We hypothesized that an ensemble ML model would outperform a traditional LR model and that feature analysis would reveal novel variables that correlate with the risk of complications.

## Methods

### Data

We retrospectively reviewed California's Office of Statewide Health Planning and Development (OSHPD) database, which contains longitudinal patient and inpatient procedure information across all licensed nonfederal hospitals in California. Inclusion criteria were patients older than 18 from 1 October 2015 to 13 December 2017 who underwent primary rTSA based on International Classification of Diseases, Tenth Revision, (ICD-10-CPS) codes. The principal inclusion codes were: 0RRJ00Z, 0RRK00Z. Per the

2020 Procedure-Specific Complication Measure Updates and Specifications Report, patients were excluded if they had discharge diagnosis codes for fracture of the upper extremity/shoulder girdle, concurrent revision, resurfacing, or implanted device/prosthesis removal, mechanical complications, malignant neoplasm of the upper extremities/shoulder girdle, bone/bone marrow or a disseminated malignant neoplasm.

Following the selection of patients in the database who underwent primary rTSA, we identified the incidence of readmission for major complications: myocardial infarction, pneumonia, sepsis, pulmonary embolism, wound infection, surgical site bleeding, mechanical complication, or 30-day readmission. Myocardial infarction, pneumonia, and sepsis were included if they occurred during the initial admission or within 7 days. Pulmonary embolism was included if it occurred during the initial admission or within 30 days and wound infection and surgical site bleeding were included if they occurred within 90 days. Readmission for any cause within 30 days following index rTSA was also included. These complications and timeframes were identified using ICD-10 codes adapted from the performance by the Centers for Medicare and Medicaid (CMS) for total joint replacement.<sup>21</sup>

The patient features (explanatory variables) available in the OSHPD data included demographic characteristics (age, gender, race, ethnicity, and insurance type) while medical comorbidities were defined using the CMS hierarchical condition category risk adjustment model. Hospital characteristics included academic teaching status and hospital volume of rTSA. These variables served as features for our ML models.

### ML Modeling

We utilized 5 publicly accessible ML methods including LR and 4 benchmark ML methods—random forest,<sup>22</sup> adaptive boosting (AdaBoost),<sup>23</sup> gradient boosting machines (Gradient Boosting),<sup>24</sup> and extreme gradient boosting (XGBoost).<sup>25</sup> We implemented LR, random forest, AdaBoost, and Gradient Boosting using the *scikit-learn* Python library<sup>26</sup> and XGBoost using the *xgboost* python library.<sup>25</sup> The hyperparameters (which define the mathematical limits of an ML algorithm) of each model were selected via grid search: for LR, the coefficient for L2 regularization was chosen from a set of values in a logarithmic scale between  $1 \times 10^{-3}$  to  $1 \times 10^3$ ; for Random Forest, Adaboost, Gradient Boosting, and XGBoost the number of trees and the maximum depth of each tree were selected from (50, 100, 200, and 300) and (2, 3, 4, and 5), respectively.

### Validation

We performed 5-fold stratified cross validation, using 80% of the data set for training and 20% of the data set for

**Table 1.** Baseline Cohort Demographics.

Variable	All patients (n = 2799)
	Median (IQR)
Age (years)	69 (12)
Hospital volume <sup>a</sup>	102 (120)
	Number (%)
Male	1430 (51.09)
Race	
White	2455 (87.71)
Black	107 (3.82)
Asian/Pacific Islander	51 (1.82)
Native American	14 (0.50)
Other	140 (5.00)
Unknown	32 (1.14)
Ethnicity	
Non-Hispanic	2534 (90.53)
Hispanic	222 (7.93)
Unknown	43 (1.54)
Insurance	
Medicare	1777 (63.49)
Private	733 (26.19)
Medi-Cal	120 (4.29)
Workers' compensation	136 (4.86)
Other	33 (1.18)
Medical comorbidities	
Diabetes mellitus without complications	212 (7.57)
Diabetes mellitus with chronic complications	192 (6.86)
Coronary atherosclerosis	212 (7.57)
Morbid obesity	207 (7.40)
COPD	199 (7.11)
Chronic kidney disease, mild	198 (7.07)
Chronic kidney disease, moderate	189 (6.75)
Chronic kidney disease, severe	176 (6.29)
Chronic kidney disease requiring dialysis	176 (6.29)
Vascular disease	199 (7.11)
Other circulatory disease	186 (6.65)
Acute renal failure	185 (6.61)
Cardio-respiratory failure	183 (6.54)
Major depressive or bipolar disorder	205 (7.32)
Major fracture (except skull)	179 (6.40)
Hip fracture or dislocation	176 (6.29)
Protein-calorie malnutrition	183 (6.54)
Metastatic cancer or leukemia	176 (6.29)
Complications of implants	198 (7.07)
History of prior complications	188 (6.72)
Osteoarthritis of hip or knee	231 (8.25)
Osteoporosis	208 (7.43)
History of bone/joint/muscle infection	184 (6.57)
	Mean (SD)
Number of comorbidities	0.23 (0.93)

Abbreviations: COPD, chronic obstructive pulmonary disease; IQR, interquartile range; SD, standard deviation; rTSA, reverse total shoulder arthroplasty.

<sup>a</sup>Cases of primary rTSA performed between 1 October 2015 and 13 December 2017.

testing in each iteration. For each ML model, we calculated the area under receiver operating characteristic (AUROC) curve, area under the precision-recall curve (AUPRC), and Brier score, reported as a mean and standard deviation. The AUROC is used extensively across all ML applications to measure classifier performance. The receiver operating characteristic (ROC) curve plots the true positive rate against the false-positive rate, and the AUROC reflects the predictive power of the model—an AUROC of 1 suggests perfect classification and an AUROC of 0.5 indicates no discriminative power.<sup>27</sup> Precision-recall curves plot sensitivity against the positive predictive value and are used for imbalanced or skewed data sets with small numbers of positive cases. Unlike the AUROC, the baseline for AUPRC is equivalent to the proportion of positive cases in the data set. Predictive power is reflected by an AUPRC greater than this baseline value, so in an imbalanced data set with few positives, an AUPRC well >1 but well above the baseline may still suggest a strong model.<sup>28</sup> Finally, the Brier score represents a cost function based on the probability predicted by the model and the actual outcome. A well-calibrated model, which has probabilistic confidence comparable to its prediction accuracy, will have a Brier score closer to 0, while a poorly calibrated model will have a Brier score closer to 1.<sup>29</sup> The performance scores (AUROC, AUPRC and Brier scores) were generated to the 15th decimal but rounded to the third decimal in this study for brevity.

### Feature Analysis

To determine what features were most important to the top-performing ML benchmark and LR algorithms, we applied Friedman's partial dependence function,<sup>24</sup> which calculates the marginal effect of each variable on the model's prediction. The continuous variables were standardized to zero mean and unit variance, and the categorical variables were one-hot encoded.

**Table 2.** Major Complications and Readmission.

Complications	All patients (n = 2799)
	Number (%)
At least one complication or readmission	142 (5.07)
Readmission within 30 days	75 (2.68)
Wound infection	22 (0.79)
Sepsis	5 (0.18)
Mechanical complication	1 (0.04)
Pneumonia	15 (0.54)
Pulmonary embolism	11 (0.39)
Surgical site bleeding	8 (0.29)
Acute myocardial infarction	5 (0.18)

**Table 3.** Discrimination and Calibration.

Model	AUROC	AUPRC	Brier score
XGBoost	0.681 ± 0.064	0.129 ± 0.049	0.037 ± 0.002
Logistic regression	0.637 ± 0.046	0.105 ± 0.051	0.038 ± 0
Gradient boosting	0.638 ± 0.096	0.104 ± 0.042	0.043 ± 0.005
AdaBoost	0.568 ± 0.097	0.082 ± 0.011	0.170 ± 0.063
Random forest	0.667 ± 0.050	0.075 ± 0.018	0.044 ± 0.002

Abbreviations: AUROC, area under the receiver operating characteristic; AUPRC, area under the precision–recall curve; AdaBoost, adaptive boosting;

## Results

### Demographics

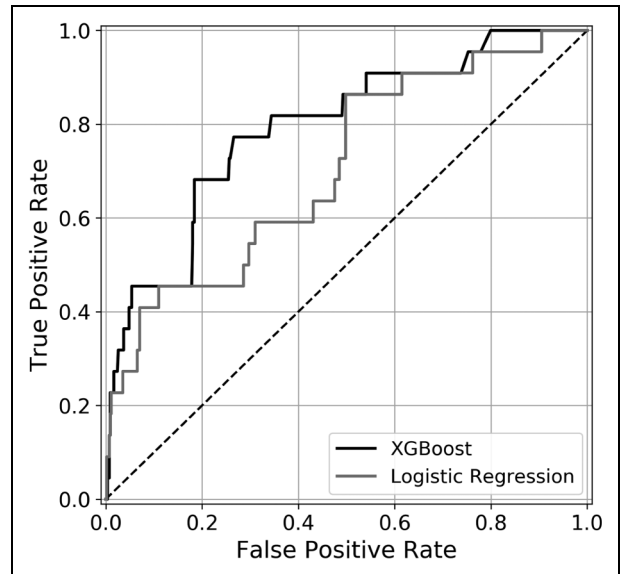
A total of 2799 rTSAs were performed during the study period and met inclusion and exclusion criteria. In total, 51% of the cohort was male with median age of 69 and an interquartile range of 12. A summary of demographics and patient features is provided in Table 1. We identified 142 (5.1%) patients with at least 1 major postoperative complication or readmission within 30 days of the index surgery for any cause. The all-cause 30-day readmission rate was 2.7% and wound infection was the most common major complication (0.8%). The incidence of each complication is documented in Table 2.

### Model Performance

Based on the AUROC and AUPRC, XGBoost had the greatest predictive power with an AUROC of 0.681 and AUPRC of 0.129. Comparatively, LR had AUROC of 0.637 and AUPRC of 0.105. The data set baseline value used for AUPRC reference was 0.051. Random forest and gradient boosting methods had comparable AUROCs of 0.667 and 0.638 but lower AUPRCs of 0.075 and 0.104. XGBoost, random forest, gradient boosting, and LR were well-calibrated with Brier scores of 0.037, 0.044, 0.043, and 0.038, respectively. AdaBoost had an AUROC of 0.568, AUPRC of 0.082, and Brier score of 0.170. The validation results of each model are summarized in Table 3. The ROC curves and precision–recall curves of the XGBoost and LR models are depicted in Figures 1 and 2, respectively.

### Feature Analysis

Given the performance results above, XGBoost was selected for feature comparison to the traditional LR model. The 6 features with the most discriminating power for each of the 2 models, as determined by the partial dependence function, are ranked in Table 4. A patient history of implant complications was the most significant feature for both models. Hospital-type was second most important for XGBoost and

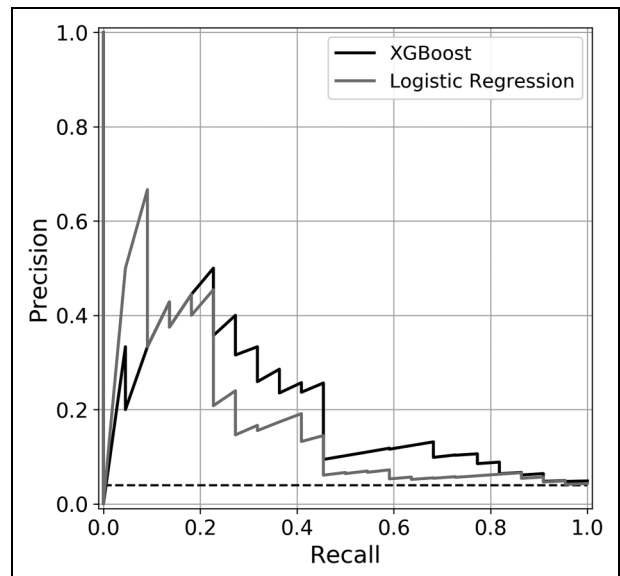


**Figure 1.** Area under receiver operating curve. Receiver operating characteristic curves for extreme gradient boosting (XGBoost) and logistic regression.

third most important for LR, while protein-calorie malnutrition was the second most important predictive variable for LR and third most important for XGBoost. The remainder of the listed features was distinct for each of the 2 models.

## Discussion

Much of the recent increase in TSA volume can be attributed to the exponential growth in the use of reverse TSA.<sup>5</sup> The



**Figure 2.** Area under precision–recall curve. Precision–recall curves of extreme gradient boosting (XGBoost) and logistic regression.

**Table 4.** Relative Feature Importance for Complications or Readmission Following Primary rTSA.

Feature	Rank in XGBoost (rank in logistic regression)	Change to risk prediction
<b>Binary features</b>		
History of implant complication	1 (1)	0.032
Teaching hospital	2 (3)	0.024
Protein calorie malnutrition	3 (2)	0.011
Osteoporosis	4 (17)	-0.007
Male sex	5 (28)	-0.005
Coronary atherosclerosis	6 (64)	-0.002
<b>Continuous features</b>		
Number of medical comorbidities	1 (1)	0.024
Hospital volume	2 (2)	-0.010
Age	3 (3)	-0.009
<b>Insurance status</b>		
Medicare	Reference	0
Private	1 (1)	-0.004
Medical	2 (2)	0
Workers comp	2 (2)	0
Other	2 (2)	0
<b>Race</b>		
White	Reference	0
Asian/Pacific Islander	1 (1)	0.008
Black	2 (2)	0
Other	2 (3)	0
Native American	2 (4)	0
Unknown	2 (5)	0

Abbreviations: rTSA, reverse total shoulder arthroplasty; XGBoost, extreme gradient boosting.

expansion of surgical indications and the relatively high complication rate of rTSA when compared to its alternatives create an especially pressing need to better understand the risk factors for complications that will only increase as demand for the surgery continues to increase. ML offers a unique opportunity to process large volume multivariable data sets to generate more accurate predictive models than traditional methods.<sup>16,19,30</sup> The purpose of this study was to create an ML algorithm to predict postoperative complications following rTSA using a statewide retrospective database. We found that XGBoost produced the most accurate predictive model with a patient history of prior implant complication being the most important patient feature to the prediction model.

Comparing multiple standard ML methods enables us to understand how different algorithms handle the data set and paves the way for the continued improvement of predictive accuracy. XGBoost was the top-performing model by AUROC and AUPRC metrics. LR has traditionally been

used in clinical studies for outcome prediction, but the results of this study suggest that ensemble ML methods like XGBoost may be better equipped to handle the complex multifactorial relationships between features and postoperative complications. Though the AUPRC is more difficult to interpret than the AUROC, it is an important metric given the imbalanced data set with a minimal incidence of positive cases. Compared to the baseline reference of 0.051 the XGBoost AUPRC of 0.129 supports the findings that XGBoost provided more accurate predictions than the other methods. Compared to gradient boosting and AdaBoost, XGBoost uses model formalization to control overfitting and improve computational efficiency, which may explain the performance benefits observed in the study. All methods were well-calibrated per the Brier scores.

A few groups have previously applied ML techniques to study outcomes of TSA.<sup>31</sup> Kumar et al<sup>32</sup> analyzed 3621 primary rTSA patients using XGBoost to predict whether patients experienced a minimal or substantial clinical benefit after TSA. Their model reported AUROCs ranging from 0.70 to 0.94 depending on which outcome they were predicting and whether or not they were using an abbreviated feature set. Despite this study of clinical improvement following rTSA using ML, there is still a need to understand the specific predictors of postoperative complications. The American College of Surgeons National Surgical Quality Improvement Database database uses the same code for aTSA and rTSA, patients who underwent either procedure were treated as a single group. Utilizing this database, Gowd et al<sup>33</sup> and Arvind et al<sup>34</sup> applied ML methods using data to predict short-term postoperative complications and unplanned readmissions following shoulder replacement surgery. The AUROC of the random forest classifiers used in these studies ranged from 0.74 to 0.77.

Our study uniquely examines complications of rTSA independently, which is valuable because rTSA has different indications and a significantly greater complication rate than aTSA. The AUROC of 0.681 of our XGBoost model is comparable to the above studies, but the limited breadth of features used in the OSHPD database may explain the slightly lower value when compared to models that were trained with alternative data sets.

Both XGBoost and LR identified a patient history of implant-related complications as the feature with the most discriminating power, with teaching hospitals and protein-calorie malnutrition in either order being the second and third most significant binary features. The nature of these models does not allow us to infer causality, but it is plausible to hypothesize relationships to postoperative complications. A history of implant complications suggests a patient may be prone to further complications related to the failure of existing prostheses. Teaching hospitals may be more likely to encounter complex patients and pathologies that are more prone to complications, and patients with protein-calorie malnutrition may experience delayed recovery,

increasing the likelihood of medical complications. Among the continuous variables, the number of patient comorbidities was the strongest predictor for both XGBoost and LR, which is in line with the rationale that patients with a more extensive medical history may be at greater risk of complications. XGBoost also identified increased age and lower hospital volume as having associations with postoperative complications. Walch et al<sup>13</sup> previously identified that less surgeon experience, which may correlate with lower hospital volumes, was associated with a greater rate of postoperative complications following rTSA.

Of note, the top features in the XGBoost model consistently had a greater influence on risk prediction than those in the LR model, suggesting that XGBoost was better able to identify and process significant features. Quantifying the impact of each feature using the partial dependence function provides key insights that can be applied to the design of future novel predictive models.

Though the design of our study enabled analysis of a large cohort, it does have limitations. The variables serving as features are dependent on ICD-10 codes and were selected from an administrative data set. This strategy enables us to process a wide range of variables but likely limits the accuracy of ML models due to the dependence on coding accuracy. Prior studies have found more specific variables such as surgeon experience and the indication for surgery to be predictors of complications in rTSA, and the data set does not allow us to directly incorporate these features into our model as a systemic chart review would.<sup>11,13</sup> Similarly, we recognize that predicting orthopedic/implant-related complications would be clinically useful, however, were unable to do so with this particular data set. Despite the substantial size of the data set, the number of patients found to have major postoperative complications was small, reducing the number of positive cases available to train the algorithms. We consequently analyzed our validation data using AUPRC, but the lack of complications may have limited the predictive capabilities of our models. Another challenge of working with ML is that methods often operate as a black box, making it difficult to interpret the relationships that the algorithms build between variables.<sup>35</sup> We can intuitively explain the directionality of relationships between many features and outcomes but cannot conclusively infer causality.

## Conclusion

We achieved the aims of building an ML model for the prediction of postoperative outcomes, showing the superiority of XGBoost over LR, and determining which features had the greatest discriminatory power. This model and identified prognostic features have the potential for improving preoperative decision making and the informed consent process. Additionally, this tool may hold value with risk adjustment of outcome-based performance measures and reimbursement programs. Further studies can continue to improve feature

selection, aggregating the results from our feature analysis with other studies that have identified singular risk factors in rTSA to improve the accuracy of our top-performing models. Using data sets built from chart review would enable greater control of input variables at the expense of volume and offer additional insights into the relationships between patient factors and outcomes. Going forward, as we apply new techniques to improve our predictive accuracy, we hope that this study's novel application of accessible ML methods to rTSA complications offers a foundation and provides insights into ultimately helping surgeons improve patient outcomes.


## Declaration of Conflicting Interests


The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported in part by the H H Lee Resident Research Grant [HHLEE FAU 4414893H-62252].

## ORCID iDs

Sai K. Devana  <https://orcid.org/0000-0002-7264-2307>

Changhee Lee  <https://orcid.org/0000-0002-8681-4739>

Carlos Solorzano  <https://orcid.org/0000-0003-2032-980X>

## References

- Gerber C, Pennington SD, Nyffeler RW. Reverse total shoulder arthroplasty. *J Am Acad Orthop Surg.* 2009;17(5):284–295. doi:10.5435/00124635-200905000-00003
- Burden EG, Batten TJ, Smith CD, Evans JP. Reverse total shoulder arthroplasty. *Bone Joint J.* 2021;103-B(5):813–821.
- Drake GN, O'Connor DP, Edwards TB. Indications for reverse total shoulder arthroplasty in rotator cuff disease. *Clin Orthop Relat Res.* 2010;468(6):1526–1533. doi:10.1007/s11999-009-1188-9
- Berliner JL, Regalado-Magdos A, Ma CB, Feeley BT. Biomechanics of reverse total shoulder arthroplasty. *J Shoulder Elb Surg.* 2015;24(1):150–160. doi:10.1016/j.jse.2014.08.003
- Wagner ER, Farley KX, Higgins I, Wilson JM, Daly CA, Gottschalk MB. The incidence of shoulder arthroplasty: rise and future projections compared with hip and knee arthroplasty. *J Shoulder Elb Surg.* 2020;29(12):2601–2609. doi:10.1016/j.jse.2020.03.049
- Carducci MP, Zimmer ZR, Jawa A. Predictors of unsatisfactory patient outcomes in primary reverse total shoulder arthroplasty. *J Shoulder Elb Surg.* 2019;28(11):2113–2120. doi:10.1016/j.jse.2019.04.009
- Cheung E, Willis M, Walker M, Clark R, Frankle MA. Complications in reverse total shoulder arthroplasty. *J Am Acad Orthop Surg.* 2011;19(7):439–449. doi:10.5435/00124635-201107000-00007
- Fu MC, Ondeck NT, Nwachukwu BU, et al. What associations exist between comorbidity indices and postoperative adverse

- events after total shoulder arthroplasty? *Clin Orthop Relat Res*. 2019;447(4):881–890. doi:10.1097/CORR.0000000000000624
9. Gerber C, Canonica S, Catanzaro S, Ernstbrunner L. Longitudinal observational study of reverse total shoulder arthroplasty for irreparable rotator cuff dysfunction: results after 15 years. *J Shoulder Elb Surg*. 2018;21(5):831–838. doi:10.1016/j.jse.2017.10.037
  10. Simovitch R, Flurin PH, Marczuk Y, et al. Rate of improvement in clinical outcomes with anatomic and reverse total shoulder arthroplasty. *Bull Hosp Joint Dis*. 2015;73(Suppl 1):S111–S117.
  11. Barco R, Savvidou OD, Sperling JW, Sanchez-Sotelo J, Cofield RH. Complications in reverse shoulder arthroplasty. *EFORT Open Rev*. 2016;1(3):72–80. doi:10.1302/2058-5241.1.160003
  12. Wall B, Nové-Josserand L, O'Connor DP, Edwards TB, Walch G. Reverse total shoulder arthroplasty: a review of results according to etiology. *J Bone Jt Surg—Ser A*. 2007;89(7):1476–1485. doi:10.2106/JBJS.F.00666
  13. Walch G, Bacle G, Lädermann A, Nové-Josserand L, Smithers CJ. Do the indications, results, and complications of reverse shoulder arthroplasty change with surgeon's experience? *J Shoulder Elb Surg*. 2012;21(11):1470–1477. doi:10.1016/j.jse.2011.11.010
  14. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920–1930. doi:10.1161/CIRCULATIONAHA.115.001593
  15. Baştanlar Y, Özuysal M. *Methods Mol Biol*. vol 1107. Humana Press, Totowa, NJ; 2014. URL: [https://doi.org/10.1007/978-1-62703-748-8\\_7](https://doi.org/10.1007/978-1-62703-748-8_7)
  16. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. *Front Bioeng Biotechnol*. 2018;6(1):75. doi:10.3389/fbioe.2018.00075
  17. Groot OQ, Ogink PT, Lans A, et al. Machine learning prediction models in orthopedic surgery: a systematic review in transparent reporting. *J Orthop Res*. 2021. doi:10.1002/jor.25036 Online ahead of print.
  18. Han XG, Tian W. Artificial intelligence in orthopedic surgery: current state and future perspective. *Chin Med J (Engl)*. 2019;132(21):2521–2523. doi:10.1097/CM9.0000000000000479
  19. Maffulli N, Rodriguez HC, Stone IW, et al. Artificial intelligence and machine learning in orthopedic surgery: a systematic review protocol. *J Orthop Surg Res*. 2020;15(1):478. doi:10.1186/s13018-020-02002-z
  20. Polce EM, Kunze KN, Fu MC, et al. Development of supervised machine learning algorithms for prediction of satisfaction at 2 years following total shoulder arthroplasty. *J Shoulder Elb Surg*. 2021;30(6):e290–e299. doi:10.1016/j.jse.2020.09.007
  21. (YNHSC/CORE) YNHSC-C for OR and E. 2020 Procedure-Specific Complication Measure Updates and Specifications Report: Elective Primary Total Hip Arthroplasty (THA) and/or Total knee Arthroplasty (TKA)—Version 9.0. 2020.
  22. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
  23. Rätsch G, Onoda T, Müller KR. Soft margins for AdaBoost. *Mach Learn*. 2001;42(3):287–320.
  24. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189–1232.
  25. Chen T, Guestrin C. XGBoost: a scalable tree boosting system ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, New York, NY, USA.
  26. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12(1):2825–2830.
  27. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Casp J Intern Med*. 2013;4(2):627–635.
  28. Ozenne B, Subtil F, Maucourt-Boulch D. The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol*. 2015;68(8):855–859.
  29. Ikeda M, Ishigaki T, Yamauchi K. Relationship between Brier score and area under the binormal ROC curve. *Comput Methods Programs Biomed*. 2002;67(3):187–194. doi:10.1016/S0169-2607(01)00157-2
  30. Kumar V, Roche C, Overman S, et al. What is the accuracy of three different machine learning techniques to predict clinical outcomes after shoulder arthroplasty? *Clin Orthop Relat Res*. 2020;478(10):2351–2363. doi:10.1097/CORR.0000000000001263
  31. Roche C, Kumar V, Overman S, et al. Validation of a machine learning derived clinical metric to quantify outcomes after TSA. *J Shoulder Elb Surg*. 2021;30(10):2211–2224. doi:10.1016/j.jse.2021.01.021
  32. Kumar V, Roche C, Overman S, et al. Using machine learning to predict clinical outcomes after shoulder arthroplasty with a minimal feature set. *J Shoulder Elb Surg*. 2021;30(5):e225–e236. doi:10.1016/j.jse.2020.07.042
  33. Gowd AK, Agarwalla A, Amin NH, et al. Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. *J Shoulder Elb Surg*. 2019;28(12):e410–e421. doi:10.1016/j.jse.2019.05.017
  34. Arvind V, London DA, Cirino C, Keswani A, Cagle PJ. Comparison of machine learning techniques to predict unplanned readmission following total shoulder arthroplasty. *J Shoulder Elb Surg*. 2021;30(2):e50–e59. doi:10.1016/j.jse.2020.05.013
  35. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol*. 2019;20(5):e262–e273. doi:10.1016/S1470-2045(19)30149-4