International Journal of Research in Orthopaedics
Dasari P et al. Int J Res Orthop. 2024 Mar;10(2):227-234
http://www.ijoro.org

DOI: https://dx.doi.org/10.18203/issn.2455-4510.IntJResOrthop20240401

Original Research Article

Application of machine learning constructs to predict post-operative complications and adverse events following shoulder hemiarthroplasty

Prathyusha Dasari[1]*, Richard C. Rice[2], Andrew Cabrera[3], Mikayla Kricfalusi[4], Trevor Case[4], Matthew T. Gulbrandsen[2], Jeremy Brown[2], Wesley P. Phipatanakul[2]

[1]University of California Riverside School of Medicine, Riverside, CA, USA
[2]Department of Orthopaedic Surgery, Loma Linda University School of Medicine, Loma Linda, CA, USA
[3]Loma Linda University School of Medicine, Loma Linda, CA, USA
[4]California University of Science and Medicine, Colton, CA, USA

Received: 23 January 2024
Accepted: 19 February 2024

***Correspondence:**
Dr. Prathyusha Dasari,
E-mail: pdasari09@gmail.com

ABSTRACT

Background: Artificial intelligence (AI) constructs and machine learning (ML) algorithms have demonstrated utility in predicting various clinical, surgical, and financial outcomes. In this study, we applied AI to shoulder hemiarthroplasty (HA) to predict various post-operative complications.
Methods: The sample was queried from the American college of surgeons-national surgical quality improvement program (ACS-NSQIP) database for all shoulder HA cases from 2008-2018. Six ML algorithms-random forest classifier, gradient boosting classifier, decision tree classifier, SVM classifier-tuned model, Gaussian Naïve Bayes classifier, multi-layer perception-analyzed the sample dataset. Postoperative complications included extended length of stay, non-home discharge destination, transfusion, and any adverse event. Each ML model was compared to logistic regression (LR), and model strength was evaluated.
Results: We identified a total of 1585 shoulder HA cases. Mean age, BMI, operative time, and length of stay were 66±12 years, 31±8 kg/m^2, 114±61 minutes, and 2.93±6.61 days. Preop hematocrit, longer operative time, and older age were most predictive of extended length of stay. Preop hematocrit, operative time, and ASA class had the highest importance in any adverse events (AAE) prediction. ML models outperformed traditional comorbidity indices, LR, for predicting extended length of stay (79% vs. 66%), non-home discharge destination (79% vs. 65%), any adverse event (78% vs. 66%), and transfusion requirement (82% vs. 63%).
Conclusions: ML algorithms predicted post-surgical outcomes of interest following shoulder HA at a higher rate to conventional LR and can assist orthopedic surgeons in decision making.

Keywords: ML, Shoulder HA, Postoperative complications, Outcomes, AI

INTRODUCTION

Hemiarthroplasty (HA) is traditionally employed as an intervention for younger patients with primary glenohumeral arthritis and complex proximal humerus fractures. Estimated 1-year cost of performing a primary HA procedure for a proximal humerus fracture is estimated at approximately $45,000.[1] Moreover, approximately 38% require revision to shoulder arthroplasty in their lifetime, which is estimated to add an additional $23,377.[1] These findings are corroborated by findings that demonstrate patients experience longer lengths of stay, higher surgical costs, and increased opioid consumption following HA compared to osteosynthesis, closed reduction and percutaneous pinning, and conservative management for proximal humerus fractures.[2]

Recently, technological advancements to AI have revolutionized contemporary data analysis. Applying its predictive abilities perioperatively to patients undergoing

HA may represent a solution for practicing quality, cost-effective medicine.[3,4] ML, a branch of AI that utilizes data and algorithms to imitate the way in which humans learn, has vastly expanded our predictive capabilities within the healthcare field, and is especially efficacious when applied to retrospective, outcome-driven cohort studies. For example, Gowd et al. recently employed ML to validate a construct and predict short-term postoperative complications following total shoulder arthroplasty (TSA).[5] Various outcomes, including need for transfusion, extended length of stay, surgical site infection, and any adverse event (amongst others), were predicted with higher affinity than conventional comorbidities indices, including American society of anesthesiologists (ASA) class, modified Charlson comorbidity index (mCCI), and frailty index. In this study, we utilized a ML construct that incorporates various patient characteristics, surgical factors, and perioperative variables to predict outcomes of interest following HA.

METHODS

In this study, we utilized the ACS-NSQIP database and received an exempt status from our institutional review board. The ACS-NSQIP database was queried using R studio (RStudio, PBC, Boston, MA) to identify adult patients undergoing shoulder HA from 2008 to 2018. Patients who underwent shoulder HA in the ACS-NSQIP database were identified using current procedural terminology (CPT) code 23470. Patients with malignancy and those containing missing data were excluded from analyses within our study.

Patients that met criteria were analyzed by six ML algorithms: random forest classifier (RF), gradient boosting classifier (GB), decision tree classifier (DT), support vector machine classifier (SVM), Gaussian Naive Bayes classifier (GNB), and multi-layer perceptron classifier (MLP). We utilized each ML algorithm to predict the extended length of stay, non-home discharge, transfusion, and any adverse event using the SciKit-learn library in the Python programming language.[6,7] As coded in ACS-NSQIP, non-home discharge was defined as discharge to "skilled care," "rehabilitation facility," "separate acute care," "unskilled facility not home," or "multi-level senior community". Home discharge was defined as "home," "facility which was home," or "against medical advice". AAE were defined as those having any one or multiple of the following: surgical site infection, renal complications, sepsis, intubation, transfusion, pneumonia, deep vein thrombosis (DVT), urinary tract infection (UTI), cerebrovascular accidents, cardiac arrest, myocardial infarction (MI), return to operating room, or death.

Patient variables included demographic information, preoperative lab values, comorbidities, and operative time (Table 1). Patient variables were preprocessed using SciKit-Learn's StandardScaler and subsequently a 70:30 train_test_split method was applied.[6,8] This split our population into a training dataset consisting of 70% of the population and a testing dataset consisting of the remaining 30%. The 30% testing dataset would be later utilized for testing model performance. For each model, GridSearchCV was used to determine optimal hyperparameters for individual predictions.[6,9] Along with GridSearchCV a stratified five-fold cross validation was used to ensure generalizability and prevent overfitting.[6,9] Once the appropriate hyperparameters were chosen, the final models were evaluated using the 30% testing subset to determine the model's performance.

The performance of the six ML models was then evaluated by a series of metrics including classification accuracy, sensitivity, specificity, and area under receiver operating characteristics curve (ROC AUC). Additionally, negative likelihood ratios (NLR) and positive likelihood ratios (PLR) were calculated using sensitivity and specificity. PLRs above 10 and NLRs below 0.1 are considered to provide strong evidence to rule in or rule out diagnoses in most circumstances, respectively.[10] Subsequently, each model was categorized as acceptable, excellent, or outstanding based on AUC ranges of 0.70-0.79, 0.80-0.89, or 0.90 and greater, respectively.[11] The graphical visualization of the ROCs produced by each of the models was accomplished through utilization of the Matplotlib library in python.[12] For the highest-performing model in each prediction, the importance of each variable was quantified based on permutation feature importance (PFI) using the ELI5 library (version 0.11.0). PFI is derived by permuting a single feature through random shuffling or removal, which disrupts its association with predicted outcome. Consequently, a decrease in model's performance indicates the degree to which the model relies on that particular variable for making accurate predictions.[13,14]

All subsequent statistical analyses were implemented using SPSS version 29 (IBM Corporation, 2021, Armonk, NY, USA) with a statistical significance defined as $p<0.05$. Categorical variables were analyzed using Pearson's chi-square test; numerical variables were analyzed using independent sample t tests.

RESULTS

The patient dataset contained 1585 patients (906 male, 679 female) with a mean age of 66 years (Table 1). The majority of patients (94%) were functionally independent upon admission and 58% had an ASA of 3. A frailty index of 1 was most common, followed by 0, accounting for 46% and 31% of patients respectively. The average LOS was 2.9 days, with 19% of patients having an extended length of stay (>7 days).

Non-home discharge was required in 7% of patients, and 7% of patients required transfusion. Postop adverse events occurred in 10% of patients, with transfusion being the most common at 7%, followed by return to operating room at 2% (Table 2).

AUC, accuracy, sensitivity, specificity, negative likelihood ratio, positive likelihood ratio

We compared the performance of six different ML algorithms with LR in predicting various postoperative complications, including extended length of stay (LOS), non-home discharge, any adverse event (AAE), and transfusion requirement (Table 3 and Figure 1).

For transfusion requirement, extended LOS, and AAE, all ML algorithms outperformed the LR model. For NHD, five of the ML models-RF, GB, GNB, MLP, SVM - demonstrated superior predictive ability than LR; while DT was the only ML model that was outperformed by traditional LR. The most predictive models for each outcome had acceptable or excellent AUCs. RF was the highest performing model for transfusion requirement, extended LOS, and AAE with AUCs of 0.90, 0.79, 0.78; specificities of 0.97, 0.77, 0.88; and accuracies of 93%, 75%, 83%, respectively. GB was the most predictive model for NHD with an AUC of 0.79, specificity of 1.00, and accuracy of 93%.

Notably, for transfusion requirement, the RF and GB models demonstrated PLRs of 13 and negative infinity, respectively. These PLR values indicate that any patient flagged by this algorithm to the need a transfusion will almost always experience that outcome.

Table 1: Preoperative characteristics of study population with shoulder HA, n=1585.

Variables	Mean (± SD) or n (%)
Demographics	
Age (In years)	66.25±12.45
BMI (Kg/m^2)	31.02±7.81
Gender	
Male	906 (57)
Female	679 (43)
Race	
American Indian or Alaska native	10 (0.6)
Asian	25 (1.6)
Black or African American	84 (5.3)
Native Hawaiian or Pacific Islander	6 (0.4)
White	1268 (80)
Unknown/ not reported	192 (12.1)
Ethnicity Hispanic	75 (5)
Functional status	
Independent	1496 (94)
Partially dependent	75 (5)
Totally dependent	14 (1)
Comorbidities	
Smoking	240 (15)
Diabetes	
Non-insulin dependent	218 (14)
Insulin dependent	108 (7)
Congestive heart failure	9 (0.6)
COPD	119 (8)
Dialysis	8 (0.5)
Hypertension requiring medication	1029 (65)
Dyspnea	
Moderate exertion	118 (7)
At rest	5 (0.3)
History of oral steroid use	85 (5)
Bleeding disorder	54 (3)
Weight loss	4 (0.3)
Laboratory values	
Preoperative HCT	39.42±5.10
Preoperative creatinine	0.93±0.51
Preoperative BUN	17.79±8.10
Preoperative WBC	7.67±2.62
Preoperative platelet count	249.96±81.02
Preoperative sodium	138.82±3.16

Continued.

Variables	Mean (± SD) or n (%)
ASA classification	
1-No disturb	32 (2)
2-Mild disturb	563 (36)
3-Severe disturb	914 (58)
4-Life threat	76 (5)
Fragility index	
0	483 (30)
1	731 (46)
2	272 (17)
3	63 (4)
4	34 (2)
5	2 (0.1)
Operative time	113.76±61.22

Table 2: Adverse events recorded, (n=1585).

Variables	Percentage (%)
Extended length of stay	302 (19)
Non-home discharge destination	109 (7)
Transfusion	107 (7)
AAE	166 (10)
Surgical site infection	14 (0.9)
Renal complications	5 (0.3)
Sepsis	11 (0.7)
Intubation	10 (0.6)
Transfusion	107 (7)
Pneumonia	15 (0.9)
DVT	12 (0.8)
UTI	21 (1)
Cerebrovascular accident	1 (0.06)
Cardiac arrest	3 (0.2)
MI	7 (0.4)
Return to OR	31 (2)
Death	7 (0.4)

Table 3: Performance metrics for ML algorithms. AUC, area under operator curve; PLR; NLR, PLR and NLR recorded as N/A when unable to be calculated (Dr. Phip rec table).

Outcomes and algorithms	ROC AUC	Specificity	Sensitivity	Accuracy	PLR	NLR
Transfusion						
Random forest classifier	0.899	0.97	0.38	93.1%	12.53	0.64
Multi-layer perceptron	0.881	0.93	0.43	92.7%	6.34	0.61
Gradient boosting classifier	0.876	1.00	0.19	94.6%	-∞	0.81
SVM classifier	0.861	0.83	0.71	82.6%	4.32	0.34
Gaussian Naive Bayes classifier	0.830	0.93	0.43	90.0%	6.34	0.61
Decision tree classifier	0.759	0.76	0.71	75.7%	2.98	0.38
LR	0.632	0.57	0.57	57.4%	1.34	0.75
Non-home discharge						
Gradient boosting classifier	0.787	1.00	0.00	93.1%	N/A	1
Random forest classifier	0.782	0.91	0.36	86.8%	3.83	0.70
Multi-layer perceptron	0.771	0.94	0.14	92.7%	2.37	0.92
SVM classifier	0.760	0.77	0.59	75.7%	2.56	0.53
Gaussian Naive Bayes classifier	0.732	0.94	0.14	88.6%	2.37	0.92
LR	0.646	0.51	0.73	52.7%	1.49	0.53
Decision tree classifier	0.631	0.83	0.36	79.5%	2.10	0.77

Continued.

Outcomes and algorithms	ROC AUC	Specificity	Sensitivity	Accuracy	PLR	NLR
Extended length of stay						
Random forest classifier	0.790	0.77	0.70	75.4%	3.00	0.39
SVM classifier	0.790	0.75	0.70	74.1%	2.81	0.40
Gradient boosting classifier	0.789	0.97	0.27	83.9%	9.79	0.75
Gaussian Naive Bayes classifier	0.765	0.92	0.25	79.5%	3.21	0.81
Multi-layer perceptron	0.758	0.92	0.25	82.6%	3.21	0.81
Decision tree classifier	0.685	0.64	0.67	64.7%	1.86	0.52
LR	0.655	0.60	0.63	60.3%	1.57	0.62
Any adverse event						
Random forest classifier	0.776	0.88	0.39	83.3%	3.39	0.69
Gradient boosting classifier	0.766	0.98	0.15	89.3%	7.17	0.87
Gaussian Naive Bayes classifier	0.766	0.93	0.30	86.1%	4.10	0.75
Multi-layer perceptron	0.751	0.93	0.30	88.3%	4.10	0.75
SVM classifier	0.737	0.82	0.64	80.0%	3.48	0.45
Decision tree classifier	0.734	0.70	0.76	70.3%	2.50	0.35
LR	0.658	0.61	0.58	60.3%	1.46	0.70

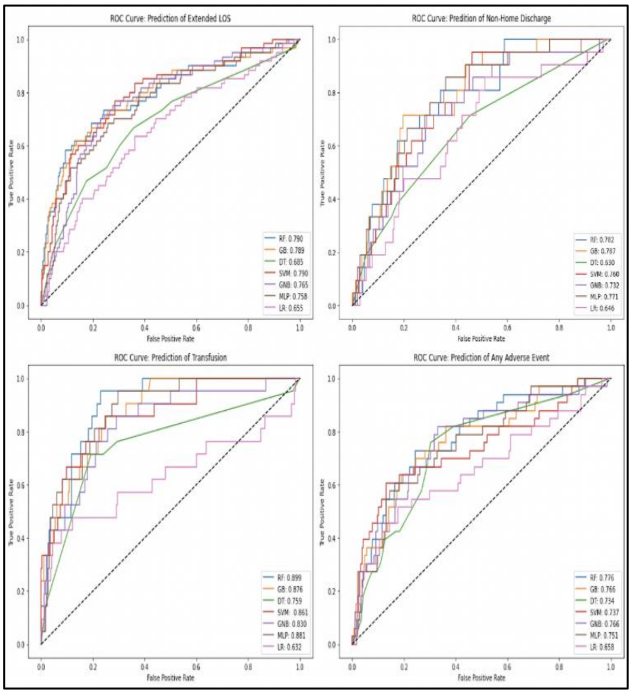

Figure 1: ROC of various ML algorithms applied to the outcome variables of interest: extended length of stay, non-home discharge, transfusion, and any adverse event.

Permutation importance

We utilized permutation factor importance (PFI) on the highest performing models to determine the most influential variables for outcome prediction of extended length of stay (LOS), any adverse event (AAE), transfusion requirement, and non-home discharge (NHD). The summary table containing PFIs and associated statistical significance utilizing p-values is provided in Table 4. For extended length of stay (LOS), we employed the RF model, which identified preoperative hematocrit emerged as the variable with the greatest importance (PFI=0.06, p<0.001), followed by operative time (PFI=0.03, p<0.001) and age (PFI=0.018, p<0.001). In predicting AAE, the RF model again identified preoperative hematocrit as the most important variable (PFI=0.08, p<0.001), followed by operative time (PFI=0.05, p<0.001) and ASA Class (PFI=0.02, p<0.001). Transfusion requirement was best predicted by age (PFI=0.09, p<0.001), preoperative hematocrit (PFI=0.06, p<0.001), and BMI (0.03, p<0.001). For NHD, the GB model exhibited the highest performance, and its permutation importance analysis indicated age as the variable with the most significant impact on outcome prediction (PFI=0.09, p<0.001). Additionally, preoperative hematocrit (PFI=0.06, p<0.001) and BMI (PFI=0.03, p<0.001) were identified as influential factors.

Table 4: PFI of variables predictive of outcomes following shoulder HA, as determined by the highest performing algorithm. p<0.05 considered significant, (n=1585).

Outcomes and variables (ML model)	Experienced adverse event (Count or average ± SD)	Did not experience event (Count or average ± SD)	PFI	P value
Length of stay (RF)	**Extended length of stay**	**Average length of stay**		
Preoperative hematocrit	36.25±5.47%	40.16±4.72%	0.059	<0.001
Operative time	141.17±90.02 minutes	107.30±50.10 minutes	0.033	<0.001
Age (In years)	70.74±12.91	65.20±12.12	0.018	<0.001
Discharge destination (GB)	**Non-home discharge (1)**	**Home discharge (0)**		
Age (In years)	72.90±12.28	65.76±12.33	0.091	<0.001
Preoperative hematocrit	36.67±4.94%	39.62±5.06%	0.063	<0.001
Body mass index (kg/m^2)	27.34±7.80	31.29±7.75	0.026	<0.001

Continued.

Outcomes and variables (ML model)	Experienced adverse event (Count or average ± SD)	Did not experience event (Count or average ± SD)	PFI	P value
Transfusion requirement (RF)	**Received transfusion**	**Did not receive transfusion**		
Age (In years)	71.97±11.30	65.84±12.44	0.091	<0.001
Preoperative hematocrit	33.07±5.16%	39.88±4.79%	0.063	<0.001
Body mass index (kg/m^2)	28.25±7.67	31.22±7.79	0.026	<0.001
Any adverse event (RF)	**Experienced**	**Did not experience**		
Preoperative hematocrit	34.90±5.72%	39.95±4.76%	0.078	<0.001
Operative time	146.02±85.98 minutes	109.98±56.49 minutes	0.045	<0.001
ASA class 4	28 (1.77%)	48 (3.03%)	0.016	<0.001

DISCUSSION

In this study, we employed six ML algorithms to analyze various patient characteristics, surgical variables, and other values to predict post-operative outcomes of interest following HA. ML algorithms outperformed the LR model for transfusion requirement, extended LOS, and AAE. RF was the highest performing algorithm overall with AUCs of 0.90, 0.79, 0.78; specificities of 0.97, 0.77, 0.88; and accuracies of 93%, 75%, 83%, respectively. For NHD, five of the ML models-RF, GB, GNB, MLP, SVM-demonstrated superior predictive ability than the LR model; DT was the only ML model that was outperformed by traditional LR. GB was the most predictive model for NHD with an AUC=0.79, specificity of 1.00, and accuracy of 93%. Preoperative hematocrit was identified as the most influential variable across all four outcomes on permutation importance analysis with a PFI=0.06 (p<0.001) for extended LOS, PFI=0.08 (p<0.001) for AAE, PFI=0.06 (p<0.001) for transfusion requirement, and PFI=0.06 (p<0.001) for NHD.

Multiple studies have previously applied traditional statistical measures to evaluate outcomes in shoulder arthroplasty. Khazzam et al used logistic multivariate analysis to implicate kidney injury (GFR<60 mL/min), anemia, and coagulopathy as individual factors associated with an increased 30-day risk of several postoperative complications.[15] Similarly, Koh et al also identify numerous peri-operative variables, including age>80, cardiovascular disease, concomitant periprosthetic fracture, and revision surgery, as mediators of acute postoperative adverse events which included complications, readmission, thromboembolic events, need for blood transfusion, mortality, and need for revision surgery.[16] Our application of ML to outcome-driven, predictive modeling in shoulder HA yielded results that largely reflected prior literature. In our study, hematocrit was identified as the most important variable in predicting extended length of stay and any adverse event. And age, not hematocrit, that most strongly correlated with need for transfusion as well as non-home discharge, although hematocrit and BMI were statistically significant predictive variables for both as well. Other variables implicated in adverse outcomes include operative time and ASA classification. These findings align with previous literature emphasizing the significance of preoperative hematocrit (p<0.01 for LOS, AAE, NHD, and transfusion requirement) in predicting adverse outcomes, as well as the influence of age on the complications, as supported by the studies conducted by Khazzam et al and the Koh et al.[15,16]

In addition to validating previous literature, our ML approach offers a practical and versatile application of procedure-specific information that can greatly benefit clinical practice across various disciplines. One notable advantage is the potential integration of our predictive model with traditional preoperative risk assessment and stratification methods. As highlighted by Hill et al one of the key advantages of ML is its automated nature and ability to leverage information from electronic medical records.[17] By employing an automated tool that analyzes readily available clinical data within the patient's chart, we can supplement traditional risk stratification indices, such as the ASA classification and fragility index. Utilizing objective clinical data that is readily accessible before or upon admission, ML algorithms can provide valuable insights to better characterize preoperative risk.[17-21] For instance, our analysis consistently identified preoperative hematocrit as a significant variable, which aligns with previous studies demonstrating the importance of addressing low hematocrit levels in patients undergoing reverse TSA and HA following a proximal humerus fracture to mitigate the risk of 30-day mortality.[15] This approach utilizes both subjective variables that are present in current preoperative risk classifications and objective variables such as comorbidities, functional status, and overall health, enhancing the accuracy and comprehensiveness of risk assessment.

This ML tool serves as a valuable supplement in clinical decision-making and pre-surgical discussions with patients. By doing so, healthcare professionals can engage in open conversations with both the surgical team and the patient, ensuring that appropriate measures are taken to minimize risks and maximize positive outcomes. With the added weight of cuts to Medicare reimbursement and the bundled payment care initiative (BPCI), it becomes imperative to have a reliable tool that can assist surgeons and patients in increasing postoperative success and reducing potential complications. By harnessing the power of ML algorithms, we not only aim to improve patient outcomes but also to address the issue of healthcare expenditure in a more efficient and the effective manner.[9,22-24]

In addition to serving as a tool for risk stratification and a supplement to help with alleviating healthcare expenditure, it also contributes to the ongoing literature that aims to validate ML and AI as tools to aid in clinical decision-making. This body of literature includes those that have created valid ML algorithms to predict outcomes after hip fractures, total knee arthroplasty, total hip arthroplasty, and chondrosarcoma.[25,26] Our results in this study are comparative to studies that utilized ML for other orthopedic operations and can serve as a validated construct to provide results based on objective data.

Limitations of this study include the ACS-NSQIP database, which is de-identified, so several factors are not reported and could not be used in the analysis such as surgeon experience, changes in surgical technique, patient medications, and preventive measures that had already been taken preoperatively. Of note, complications and variables that were analyzed were those restricted to the confines of this dataset. Additionally, the ACS-NSQIP database only records data up to 30 days after a procedure; therefore, any complications that occurred after 30 days will not be noted in this dataset. This study was also retrospective in nature, and this limits the control of variables. Finally, the ACS-NSQIP data comes from institutions that are capable of staffing clinical reviewers for quality assurance of the data. Therefore, smaller surgical centers and institutions that are not capable of staffing clinical reviewers are not represented within this database and limits the applicability of our results. Additionally, PFI provides insight into the relative importance of variables in ML models.[14] While PFI can indicate important variables, it should not be interpreted as absolute importance.

For example, comparing two equally important variables in a smaller subset may lead to overestimating their importance. Therefore, PFI does not directly predict risk factors but instead offers perspective for clinical correlation of different variables.

CONCLUSION

AI models outperformed LR in predicting non-home discharge, need for transfusion, extended length of stay, and any adverse event with high accuracy. Preoperative hematocrit, followed by age, then body mass index were highlighted as variables of high importance in predicting postoperative complications. These findings demonstrate a tool that can help supplement the traditional preoperative risk stratification process, provide postoperative complication risks based on objective data, and help diminish the expenditure associated with the musculoskeletal disorders.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Politzer CS, Bala A, Seyler TM, Bolognesi MP, Garrigues GE. Use and Cost of Reverse Shoulder Arthroplasty Versus Hemiarthroplasty for Acute Proximal Humerus Fractures. Orthopedics. 2020;43(2):119–25.
2. London DA, Cagle PJ, Parsons BO, Galatz LM, Anthony SG, Zubizarreta N et al. Impact of Increasing Comorbidity Burden on Resource Utilization in Patients with Proximal Humerus Fractures. J Am Academy Orthop Surg. 2020;28(21):e954.
3. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. RadioGraphics. 2017;37(2):505-15.
4. Zhang L, Tan J, Han D, Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug Discovery Today. 2017;22(11):1680-5.
5. Gowd AK, Agarwalla A, Amin NH, Romeo AA, Nicholson GP, Verma NN et al. Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. Journal of Shoulder and Elbow Surgery. 2019;28(12):e410-21.
6. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al. Scikit-learn: Machine Learning in Python. J Machine Learning Res. 2011;12:2825-30.
7. Rossum FLJDGV. The Python Language Reference Manual.
8. Gholamy A, Kreinovich V, Kosheleva O. Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. Scholarworks.utep. 2018.
9. Testa EJ, Haglin JM, Li NY, Moore ML, Gil JA, Daniels AH et al. Temporal and Geographic Trends in Medicare Reimbursement of Primary and Revision Shoulder Arthroplasty: 2000 to 2020. J Am Academy Orthop Surgeons. 2021;29(24):e1396.
10. Jaeschke R, Guyatt GH, Sackett DL, Guyatt G, Bass E, Brill-Edwards P et al. Users' Guides to the Medical Literature: III. How to Use an Article About a Diagnostic Test B. What Are the Results and Will They Help Me in Caring for My Patients? JAMA. 1994;271(9):703-7.
11. Hosmer DW, Lemeshow S. Applied Logistic Regression: Hosmer/Applied Logistic Regression. Hoboken, NJ, USA: John Wiley and Sons, Inc. 2000.
12. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing Sci Engineering. 2007;9(3):90-5.
13. Kaneko H. Cross-validated permutation feature importance considering correlation between features. Analytical Sci Adv. 2022;3(9-10):278-87.
14. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics. 2009;10(1):213.

15. Khazzam M, Ahn J, Sager B, Gates S, Sorich M, Boes N. 30-Day Postoperative Complications After Surgical Treatment of Proximal Humerus Fractures: Reverse Total Shoulder Arthroplasty Versus Hemiarthroplasty. JAAOS Global Res Rev. 2023;7(3):e22.00174.
16. Koh J, Galvin JW, Sing DC, Curry EJ, Li X. Thirty-day Complications and Readmission Rates in Elderly Patients After Shoulder Arthroplasty. J Am Acad Orthop Surg Glob Res Rev. 2018;2(11):e068.
17. Hill BL, Brown R, Gabel E, Rakocz N, Lee C, Cannesson M et al. An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data. Brit J Anaesthesia. 2019;123(6):877-86.
18. Fritz BA, Cui Z, Zhang M, He Y, Chen Y, Kronzer A et al. Deep-learning model for predicting 30-day postoperative mortality. Brit J Anaesthesia. 2019;123(5):688-95.
19. Hofer IS, Kupina M, Laddaran L, Halperin E. Integration of feature vectors from raw laboratory, medication and procedure names improves the precision and recall of models to predict postoperative mortality and acute kidney injury. Sci Rep. 2022;12(1):10254.
20. Kowadlo G, Mittelberg Y, Ghomlaghi M, Stiglitz D, Kishore K, Guha R et al. Development and Validation of ‘Patient Optimizer’ (POP) Algorithms for Predicting Surgical Risk with Machine Learnin. 2022. Available at: https://www.medrxiv.org/content/10.1101/2022.10.03.22280539v2. Accessed on 12 January, 2024.
21. Russell M, Russell D, Corizzo R, Japkowicz N. Machine Learning for Surgical Risk Assessment Decision Systems. In: 2022 International Joint Conference on Neural Networks (IJCNN). 2022;1-8.
22. Kim K, Iorio R. The 5 Clinical Pillars of Value for Total Joint Arthroplasty in a Bundled Payment Paradigm. J Arthroplasty. 2017;32(6):1712-6.
23. Mayfield CK, Haglin JM, Levine B, Della Valle C, Lieberman JR, Heckmann N. Medicare Reimbursement for Hip and Knee Arthroplasty From 2000 to 2019: An Unsustainable Trend. J Arthroplasty. 2020;35(5):1174-8.
24. Wilensky GR. Will MACRA Improve Physician Reimbursement? N Engl J Med. 2018;378(14):1269-71.
25. Karnuta JM, Navarro SM, Haeberle HS, Billow DG, Krebs VE, Ramkumar PN. Bundled Care for Hip Fractures: A Machine-Learning Approach to an Untenable Patient-Specific Payment Model. J Orthopaed Trauma. 2019;33(7):324.
26. Thio QCBS, Karhade AV, Ogink PT, Raskin KA, De Amorim Bernstein K, Lozano Calderon SA et al. Can Machine-learning Techniques Be Used for 5-year Survival Prediction of Patients with Chondrosarcoma? Clin Orthop Relat Res. 2018;476(10):2040-8.

Cite this article as: Dasari P, Rice RC, Cabrera A, Kricfalusi M, Case T, Gulbrandsen MT et al. Application of machine learning constructs to predict post-operative complications and adverse events following shoulder hemiarthroplasty. Int J Res Orthop 2024;10:227-34.