**Using Machine Learning to Predict Internal Rotation after Anatomic and Reverse Total Shoulder Arthroplasty**

**Abstract**

**Introduction**

Internal rotation (IR) after anatomic (aTSA) and reverse (rTSA) total shoulder arthroplasty is unpredictable, with rTSA patients often reported to experience more variable and more limited IR improvements than aTSA patients. Numerous efforts have attempted to identify predictors of poor IR after shoulder arthroplasty; however, none of these findings to date are sufficient to permit pre-operative identification of patients who will fail to achieve clinically meaningful IR improvements ~~after aTSA or rTSA~~. ~~To that end,~~ the purpose of this study is to use machine learning to predict internal rotation after aTSA and rTSA at multiple post-operative timepoints.

**Methods**

Clinical data from 2,270 aTSA and 4,198 rTSA patients were analyzed using 3 supervised machine learning techniques: linear regression, XGBoost, and Wide and Deep, to create predictive models for internal rotation as measured by the IR score at 6 different post-operative timepoints ~~using a full input feature set and the 2 different minimal feature sets~~. Mean absolute errors (MAE) quantified the difference between actual and predicted IR scores for each model at each post-operative timepoint, and the performance for each model was also quantified for prediction of clinical improvement greater than the minimal clinically important difference (MCID) and the substantial clinical benefit (SCB) patient satisfaction thresholds for IR score at 2-3 years after surgery.

**Results**

rTSA patients had significantly lower mean IR score and significantly less mean IR score improvement than aTSA patients at each post-operative timepoint analyzed in this study. Despite differences in post-operative IR scores, both aTSA and rTSA patients experienced significant improvements in their ability to perform ADLs; however, aTSA patients were significantly more likely to perform these ADLs than rTSA patients. Despite IR score outcomes being more variable with rTSA than aTSA, our machine learning predictive models provided equivalent MAE when predicting IR score for both aTSA (0.92-1.18) and rTSA (1.03-1.25) from 3 months to >5 years after surgery, using the minimal feature set of inputs. Furthermore, these predictive algorithms can identify with 90% accuracy for aTSA and 85% accuracy for rTSA which patients will achieve MCID IR score improvement and predict with 85% accuracy for aTSA patients and 77% accuracy for rTSA which patients will achieve SCB IR score improvement at 2 years after surgery.

**Discussion**

Our machine learning outcome study demonstrated that active internal rotation can be accurately predicted after aTSA and rTSA at multiple post-operative timepoints ~~using a minimal feature set of pre-operative inputs~~. These predictive algorithms identified which patients will, and will not achieve clinical improvement in IR score that exceeds the MCID (90% accuracy for aTSA and 85% accuracy for rTSA) and SCB (85% for aTSA and 77% for rTSA) patient satisfaction thresholds. Future work will quantify if these accurate predictive algorithms can be utilized to effectively counsel patients undergoing elective primary TSA and better inform them of their expected outcome.

**Introduction**

Internal rotation (IR) after anatomic (aTSA) and reverse (rTSA) total shoulder arthroplasty is unpredictable, with rTSA patients often reported to experience more variable and more limited IR improvements than aTSA patients. (reference Allem, Cox, Triplet, Friedman, and Wright) Restricted IR can negatively impact a patient’s ability to perform activities of daily living (ADLs), as IR is necessary to perform recreational activities, wash back, fasten bra, comb hair, put on a coat, and perform toilet and personal hygiene needs. (Triffitt et al) Given the functional importance of IR, many authors have attempted to identify risk factors associated with poor IR, including: higher BMI, <6mm inferior glenosphere offset, superior glenoid tilt, thinner glenospheres, male gender, diagnosis of heart disease, diagnosis of diabetes, history of tobacco use, use in dominant arm, not repairing the subscapularis with rTSA, lower pre-operative clinical outcome scores, lower pre-operative forward elevation, and lower pre-operative IR score. (Eichinger, Kim Jpn Ortho Assoc, Rol, Friedman 1, Friedman 2, Werner, Collin, Levy) However to date, none of these findings are sufficient to permit pre-operative identification of patients who will fail to achieve clinically meaningful IR improvements after aTSA or rTSA.

Recently, machine learning based predictive models have been introduced that use pre-operative data to accurately predict post-operative range of motion (active forward elevation, abduction, and external rotation), subjective pain and function scores, clinical outcome scores, patient satisfaction, complications, and risk for reoperation after shoulder arthroplasty. (Polce, Gowd, Kumar 1-3, Roche, McLendon) By sharing these post-operative outcomes predictions with patients prior to surgery, surgeons could more accurately manage patient expectations for pain relief and functional improvement after elective shoulder arthroplasty. We are unaware of any predictive algorithm trained to predict IR after aTSA or rTSA. The purpose of this study is to use machine learning to predict internal rotation after aTSA and rTSA. Specifically, we will utilize 3 different feature sets of pre-operative inputs and 3 different machine learning techniques on a clinical database of one platform shoulder prosthesis to quantify the accuracy of these algorithms to predict IR as measured by the IR score (Flurin) at multiple post-operative timepoints after aTSA and rTSA. Additionally, we quantify the patient satisfaction based minimal clinically important difference (MCID) and substantial clinical benefit (SCB) improvement thresholds for IR score and make predictions for each threshold at 2-3 years after aTSA and rTSA.

**Methods**

Three supervised machine learning techniques were utilized, including: linear-regression-based, tree-based, deep-learning-based techniques, to analyze a multi-center clinical outcomes database of shoulder arthroplasty patients who received a single platform shoulder prosthesis (Equinoxe, Exactech, Inc, Gainesville, FL) between November 2004 and May 2020. Every patient consented and all data was collected using standardized forms according to an IRB approved protocol. To ensure a homogenous dataset, patients with revisions, humeral fractures, endoprostheses, and hemiarthroplasty were excluded, along with patients with <3 months follow-up. Active internal rotation was quantified by the IR score, an 8 point numeric scale with the following discreet assignments based on motion to vertebral segments: 0º = 0, hip = 1, buttocks = 2, sacrum = 3, L5 – L4 = 4, L3 – L1 = 5, T12 – T8 = 6, T7 or higher = 7. (Flurin) Given the recent report that a patient’s ability to perform various functional activities may be more accurate representation of IR capability than a vertebral level measurement, (Southard) we also quantified the ability of aTSA and rTSA patients to perform numerous ADLs >2 years after surgery. Similar to the methodology utilized by Simovitch et al, we quantify MCID (Simovitch 1) as the mean difference in improvement between aTSA/rTSA patients who described themselves as being “better” at each 2-year minimum follow-up visit as compared to patients who described themselves as “worse” and also “unchanged” and we quantify SCB (Simovitch 2) as the mean difference in improvement between “much better” patients as compared to “worse” and also “unchanged” patients.

Pre-operative, intra-operative, and post-operative data from 6,468 patients with 18,892 post-operative follow-up visits were analyzed to create algorithms that predict the IR score for aTSA and rTSA at multiple timepoints after surgery, including: 3-6 months, 6-9 months, 1 year [9-18 months], 2-3 years [18-36 months], 3-5 years [36-60 months], and 5+ years [60+ months]. These predictive algorithms were constructed using demographic data, diagnoses, comorbidities, implant size/type, pre-operative range of motion, pre-operative radiographic findings, and pre-operative clinical outcome metric scores, including the individual questions used to derive each score; in total, 291 labeled features were utilized in the full feature input dataset. IR score predictions were also calculated using a minimal feature set of 19 different pre-operative inputs, and the minimal feature set when supplemented with implant size/type data, and measurements of native glenoid version and inclination (i.e. beta angle), to simulate the additional predictive accuracy that could be gained through utilization of data from CT based pre-operative planning software.

The clinical data from 2,270 primary aTSA patients with 7,665 visits and 4,198 primary rTSA patients with 11,227 visits was split 2:1 into mutually exclusive datasets to build and test the IR score predictive models. Similar to our previous work, (Kumar CORR, Kumar JSES, Kumar Seminars JSES) a random selection of 66.7% of the data defined the training cohort and the remaining 33.3% defined the validation test cohort. Predictive models were created using 3 supervised machine learning techniques: 1) linear regression, (Stigler 1986) 2) XGBoost(Chen Torley 2017), and 3) Wide and Deep(Cheng and Zheng 2017). The performance of the full model, the minimal feature set, and the minimal feature set supplemented with the implant size/type data and CT measurement inputs to predict post-operative IR score after aTSA and rTSA was quantified by the Mean Absolute Error (MAE) between the actual and predicted values in IR score at each post-operative timepoint in the 33.3% validation test cohort. To evaluate the relative learning ability of the machine learning models, we also conducted a baseline average analysis as the study control. To aid in model interpretability, (Ahmad 2018 and Lipton 2018) the features with the highest F-score (Torley) from the XGBoost models identified the top 10 most-meaningful features for each full and minimal feature set IR score model, these F-scores were then compared to Pearsons correlations for each parameter to contrast findings to more traditional statistical techniques. The performance of the full model and each minimal feature set model to predict if a patient will achieve the MCID and SCB patient-satisfaction based improvement thresholds were quantified using the classification metrics of precision or positive predictive value (e.g. quantifying the ability of a model to not identify a negative as positive), recall or sensitivity (e.g. quantifying the ability of a model to identify a positive as a positive), F1-score, and by the concordance (c) statistic as the area under the receiver operating curve (AUROC),which is a measure of sensitivity (i.e. true positive rate) vs. 1-sensitivity (i.e. false positive rate). (Gortmaker 1994, Cabitza 2018, Hosmer 2013, Steyerberg 2010)

**Results**

The clinical data from 2,270 primary aTSA patients (7,665 visits; average follow-up = 32.5 ± 31.4 months) and 4,198 primary rTSA patients (11,227 visits; average follow-up = 22.8 ± 23.0 months) was used to train and test predictive models of the IR score at each post-operative timepoint: 3-6 months (aTSA=1343 and rTSA=2578 visits), 6-9 months (aTSA=684 and rTSA=1342 visits), 1 year (aTSA=1509 and rTSA=2725 visits), 2-3 years (aTSA=1371 and rTSA=2010 visits), 3-5 years (aTSA=1364 and rTSA=1566 visits), and 5+ years (aTSA=1378 and rTSA=1006 visits). A summary of demographics, diagnoses, and comorbidities for the aTSA and rTSA patient cohorts are presented in Table 1. Table 2 compares the pre-operative, post-operative, and pre-to-post-operative improvement in IR score for aTSA and rTSA patients at each follow-up duration and demonstrates that both male and female rTSA patients had significantly lower mean IR score and significantly less mean IR score improvement than both male and female aTSA patients at each post-operative timepoint. The frequency distribution between pre-operative and post-operative IR score is described in Figure 1 for aTSA patients and Figure 2 for rTSA patients and demonstrates a broad distribution in post-operative IR score for both aTSA and rTSA patients with a IR score of 6 being most common and achieved by 41% of aTSA and 31% of rTSA patients. Figure 1 also describes that cumulatively ~6.5% of aTSA patients and ~17% of rTSA patients had a lower IR score at >2yrs followup than they had pre-operatively. Table 3 compares the ability of aTSA and rTSA patients to perform ADLs before surgery and after 2 years and demonstrates that both aTSA and rTSA patients experienced significant improvements in their ability to perform these ADLs. However, aTSA patients were significantly more likely to be able to perform these ADLs, with 3 exceptions: comfort of sleep, comfort with arm at side, and ability to put on an undershirt. Interestingly, ADL tasks for aTSA patients were in every case more linearly correlated (as quantified by the Pearsons coefficient) to post-operative IR score than those same activities for rTSA patients.

The accuracy of the 3 different machine learning techniques to predict IR score using the full feature set of 291 inputs after aTSA and rTSA is presented in Table 4. ~~As described in Table 4,~~ the XGBoost and Wide and Deep models provided equivalent MAE in IR score for both aTSA (±1.0) and rTSA (±1.1) across the post-operative timepoints. Despite differences in accuracy ~~between these 3 machine learning techniques,~~ all 3 techniques had lower MAE than the baseline average model, demonstrating that each technique performed better than the average choice. Table 5 demonstrates that this level of prediction accuracy (MAE of ±1.1) was maintained across the 3 different model input categories for the minimal feature with and without implant/x-ray data at each prediction timepoint. To aid in model interpretability, Table 6 presents the top 10 most meaningful input features driving the XGBoost IR score prediction models for each of the full feature set and each version of the minimal feature set and demonstrates reasonably-good agreement in the features driving the predictions. Table 7 describes the relationship between the F-score and Pearsons correlation coefficient with the post-operative IR score for numerous pre-operative parameters, and illustrates the ability of these machine learning algorithms to differentiate features based upon predictive validity despite traditional statistical methods demonstrating little-to-no correlations.

The patient satisfaction anchor-based MCID thresholds for IR score were calculated to be 0.2 for the combined cohort of aTSA and rTSA patients, 0.6 for aTSA patients, and -0.1 for rTSA patients. As described in Table 8, 76% of aTSA patients and 74.7% of rTSA achieved MCID improvement in IR score at 2-3 years follow-up. The patient satisfaction anchor-based SCB thresholds for IR score were calculated to be 0.9 for the combined cohort of aTSA and rTSA patients, 1.1 for aTSA patients, and 0.6 for rTSA patients. As described in Table 8, 61.6% of aTSA patients and 58.1% of rTSA patients achieved SCB improvement at 2-3 years follow-up. The MCID and SCB internal rotation predictions for aTSA and rTSA patient at 2-3 years follow-up for the 3 different input models are presented in Table 8. For aTSA patients, the XGBoost IR score predictive models achieved 90% accuracy in MCID with an AUROC of 0.86 and 85% accuracy in SCB with an AUROC of 0.83 using the full input dataset. For rTSA patients, the XGBoost IR score predictive models achieved 85% accuracy in MCID with an AUROC of 0.84 and 77% accuracy in SCB with an AUROC of 0.77 using the full input dataset. For both aTSA and rTSA minimal features set models, similar accuracy and AUROC values were achieved as reported for the full feature set model.

**Discussion**

The results of this machine learning study demonstrate that active internal rotation after aTSA and rTSA can be accurately predicted at multiple post-operative timepoints using a minimal feature set of pre-operative inputs. Despite IR score outcomes being more variable with rTSA than aTSA, the predictive accuracy for aTSA and rTSA patients were equivalent with an error range of 0.92-1.18 for aTSA and 1.03-1.25 from 3 months to >5 years after surgery with the minimal feature set of inputs. Furthermore, these predictive algorithms identified which patients will, and will not achieve clinical improvement in IR score that exceeds the MCID (90% accuracy for aTSA and 85% accuracy for rTSA) and SCB (85% for aTSA and 77% for rTSA) patient satisfaction thresholds.

Improving IR after aTSA and rTSA is necessary for patients to return to ADLs. (Puzzitiello) Our clinical results demonstrate that the majority of aTSA and rTSA patients experience both MCID (76.0% of aTSA and 74.7% of rTSA) and SCB (61.6% of aTSA and 58.7% of rTSA) levels of IR score improvement, and also a significantly greater ability to perform numerous ADLs 2yrs after surgery. However, our results also demonstrate that 6% of aTSA and 17% of rTSA patients may lose IR after surgery. Furthermore, we observed that 29.4% of aTSA patients and 40.1% of rTSA patients found it “very difficult” or that they were “unable” to wash back/fasten bra 2 years after surgery and 13.8% of aTSA patients and 22.2% of rTSA patients found it “very difficult” or that they were “unable” to tuck their shirt behind their back 2 years after surgery. Loss of IR motion may result in failure to perform ADLs and patient dissatisfaction, particularly if these patients were expecting improvement. (Monir) For those patients in particular, the predictive capability of our machine learning algorithms can be utilized to effectively counsel patients undergoing primary total shoulder arthroplasty to better inform them of their expected outcome. Better alignment of patient’s expectations with actualized results may led to improved satisfaction with these elective procedures.

Despite more limited IR improvement relative to aTSA patients, rTSA patients often remain satisfied with their function. In one study of 215 primary rTSA shoulders, Southard et al. quantified their patients ability to perform functional IR activities relative to physical IR measures. (Southard) They reported significant improvements in performing those functional IR activities while also observing non-significant changes in IR physical measurements. It is important to note that Southard et al. did not measure IR by the most common vertebral level method due to concerns of measurement error and reliability, as have been previously reported by Rojas et al. and Edwards et al. (Rojas and Edwards) Instead, Southard et al. proposed a new technique which quantifies IR by one of 3 types (type 1: could not reach behind back, type 2: able to reach to waist level, and type 3: able to reach to at a minimum waist level in an uninterrupted fashion), the primary advantage of this new method is that it permits patients to self-report IR motion. (Southard) Unfortunately this measurement proposal has low resolution, having only 3 types, which is the most likely reason for their non-statistically significant changes in IR after rTSA. We share their measurement error and reliability concerns related to the vertebral level method, particularly the 6 level IR metric used in the Constant score. (Constant) However, reducing measurement resolution from 6 levels to 3 types cannot possibly improve accuracy for any objective measurement. Alternatively, Flurin et al. proposed an 8 level IR score to modify the Constant vertebral level method by adding 2 additional scoring levels and gain greater resolution in the vertebral regions achieved by the majority of total shoulder arthroplasty patients. (Flurin) Eichinger has recently demonstrated that the IR score has sufficient resolution to distinguish between aTSA and rTSA patient difficulty levels (normal, slightly difficult, very difficult, and unable) for numerous ADLs. (Eichinger) Friedman et al. also used the IR score to distinguish between internal rotation for rTSA patients with subscapularis repair and those who did not have their subscapularis repaired. (Friedman JSES)

In explaining their findings related to significantly improved IR functional activities despite nonsignificant changes in IR physical measurements, Southard et al. suggested that patients are able to compensate beyond what a physical exam evaluates because rTSA reliably provides pain relief and patients have a greater ability to compensate for ADLs if they are not in pain. (Southard) This is an astute observation that helps explain our own findings related to a slightly negative (-0.1) MCID IR score for rTSA patients, as rTSA patients likely had other requirements to be satisfied with their rTSA procedure, such as pain relief or restoration of stability. In our study we observed that patient satisfaction based MCID and SCB thresholds for IR improvement were greater for aTSA patients than rTSA patients. Clearly, patient satisfaction is required for a successful clinical outcome following any orthopedic surgery. (Swarup) Patients have multiple different requirements and expectations to be met in order to be satisfied with their procedure; Henn et al. reported that younger patients, and those with higher preoperative function have higher expectations of total shoulder arthroplasty. (Henn) Similarly, Simovitch et al. reported different MCID and SCB thresholds for pain relief, function, clinical outcome measure scores, and range of motion measurements for aTSA and rTSA patients of different gender and age. (Simovitch 1 and 2) The IR score predictive algorithms used in our study are commercially available in Predict+ (Exactech Inc, Gainesville, FL); future research is necessary to quantify how this clinical decision support tool influences patient satisfaction and their confidence in the decision to pursue aTSA and/or rTSA surgery.

Internal rotation after total shoulder arthroplasty is highly variable and difficult to predict. (Allem, Cox, Triplet, Friedman, and Wright) Multiple shoulder arthroplasty outcome studies have used traditional statistical techniques to identify numerous pre-operative risk factors that have weak correlations to post-operative IR improvement. (Eichinger, Kim Jpn Ortho Assoc, Rol, Friedman 1, Friedman 2, Werner, Collin, Levy) Collectively these findings suggest IR improvement after total shoulder arthroplasty is multifactorial. The pre-operative features analyzed in our study with the highest linear correlation to the post-operative IR score were the pre-operative IR score (r=0.260) and the pre-operative composite range of motion score (r=0.240). In alignment, Friedman et al. previously utilized multiple linear regression with backward stepwise selection to identify the pre-operative IR score as a significant predictor of post-operative IR score for aTSA and rTSA and Kumar et al. previously utilized machine learning to identify the composite range of motion score as one of the most powerful pre-operative prediction variables, being highly predictive of post-operative pain, function, clinical outcome scores, and also range of motion for aTSA and rTSA. (Friedman et al and Kumar et al) Notably, Pearsons correlation coefficients of 0.24-0.26 are considered weak and suggest that there is no dominant predictive factor for the IR score. Greater clarity of the predictors of IR is clinically important, especially now that rTSA in being utilized in patients with an intact rotator cuff and without glenoid bone loss. (Wright)

Machine learning based clinical decision support tools, ~~like Predict+,~~ offer the potential to extend traditional statistical methods and evolve from a simple risk-factor identification to a clinical prediction models that can be used to optimize outcomes. By better identifying and objectively ranking the pre-operative factors with the greatest predictive validity, machine learning seeks to create algorithms that utilize the most clinically relevant parameters to minimize predictive error. (Kumar 1) To that end, a review of the top pre-operative features influencing our predictive models for each of the full feature set and the 2 different minimal features sets is informative. Interestingly, the pre-operative features in our predictive models with the largest F-scores were observed to have relatively low Pearson correlation coefficients to the post-operative IR score. This finding suggests that the relationship between these pre-operative parameters and post-operative IR score is individually weak for aTSA and rTSA, and also may not be linear. Furthermore, this finding also highlights the advantage of XGBoost, which is an ensemble decision tree-based machine learning algorithm that considers groupings of multiple trees and parameters, and may not always be linear.

Our machine learning based prediction study of internal rotation after aTSA and rTSA has several limitations. First, there are well-document measurement accuracy and reliability concerns associated with the vertebral level method to quantify internal rotation; (Rojas and Edwards) ~~however, as recently described by in a meta-analysis by Rojas et al., the vertebral level method is simple and widely-used and has been utilized in 86% of the literature quantifying internal rotation.~~ (Rojas) Second, as with any machine learning based study, the generalizability of trained predictive models are limited by the quality of the underlying data and any biases during data collection or the model creation process will limit the predictive accuracy of the resulting algorithms. (Ahmad, Cabitza, Lipton, Obermeyer) Third, our clinical database, while large, is just a sample of the patients that may pursue shoulder arthroplasty and consists primarily of elderly, non-Hispanic, Caucasians of European descent. As such, predictive models derived from this dataset may not be representative of the outcomes achieved by patients of different demographics, regions, or ethnicity/race. Fourth, our database only includes patients who elected to undergo shoulder arthroplasty; as such, model predictions may be biased against patients too sick to safely undergo the procedure or patients whose condition was not sufficiently symptomatic to have the procedure. Fifth, our models were developed from a dataset of primary aTSA and primary rTSA patients using one platform shoulder prosthesis, where patients with revisions, humeral fractures, endoprostheses, or hemiarthroplasty were excluded; therefore, model predictions may not be appropriate for patients with other indications or other implant types. Sixth, the feature set of inputs used by our models is not exhaustive of all possible pre-operative parameters that may influence internal rotation; future clinical research efforts should continue to collect additional relevant parameters to further improve the algorithm predictions. Finally, the MCIDand SCBIR score improvement thresholds identified in this study are based on a patient satisfaction based anchor-question and that anchor question is not representative of every patient’s needs as it relates to internal rotation after shoulder arthroplasty.

**Conclusion**

We utilized 3 different machine learning techniques in this study of 2,270 aTSA and 4,198 rTSA patients to create aTSA and rTSA internal rotation predictive models at multiple post-operative timepoints. Despite IR score outcomes being more variable with rTSA patients, the predictive accuracy of these algorithms were equivalent for aTSA and rTSA patients, with an error range of 0.92-1.18 for aTSA and 1.03-1.25 from 3 months to >5 years after surgery with the minimal feature set of inputs. Additionally, we quantified the most-meaningful features driving IR score predictions and compared those findings to traditional statistical correlation techniques which were unable to distinguish these predictive pre-operative features. Finally, we predicted which patients will, and will not achieve clinical improvement in IR score that exceeds the MCID (90% accuracy for aTSA and 85% accuracy for rTSA) and SCB (85% for aTSA and 77% for rTSA) patient satisfaction thresholds. Future work will quantify if these accurate predictive algorithms can be utilized to effectively counsel patients undergoing elective primary TSA and better inform them of appropriate expectations through the post-operative course.

**References**

1. Ahmad MA, Eckert C, Teredesai A. Interpretable machine learning in healthcare. IEEE Intelligent Informatics Bulletin, Aug: Vol. 19, #1: 1-6. 2018.
2. Aleem AW, Chamberlain AM, Keener JD. The functional internal rotation scale: a novel shoulder arthroplasty outcome measure. JSES Int. 2019 Nov 27;4(1):202-206. doi: 10.1016/j.jses.2019.10.002.
3. Cabitza F, Banfi G. Machine learning in laboratory medicine: waiting for the flood? Clin Chem Lab Med. 2018 Mar 28;56(4):516-524. doi: 10.1515/cclm-2017-0287.
4. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794). ACM.
5. Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., & Anil, R. (2016, September). Wide & deep learning for recommender systems. In Proceedings of the 1st workshop on deep learning for recommender systems (pp. 7-10). ACM.
6. Collin P, Matsukawa T, Denard PJ, Gain S, Lädermann A. Pre-operative factors influence the recovery of range of motion following reverse shoulder arthroplasty. Int. Orthop. 2017;41(10):2135–2142. doi:10.1007/s00264-017-3573-4
7. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop Relat Res. 1987 Jan;(214):160-4.
8. Cox RM, Padegimas EM, Abboud JA, Getz CL, Lazarus MD, Ramsey ML, et al. Outcomes of an anatomic total shoulder arthroplasty with a contralateral reverse total shoulder arthroplasty. J. Shoulder Elbow Surg. 2018 Jun;27(6):998–1003. doi:10.1016/j.jse.2017.12.005
9. Edwards TB, Bostick RD, Greene CC, Baratta RV, Drez D. Interobserver and intraobserver reliability of the measurement of shoulder internal rotation by vertebral level. J Shoulder Elbow Surg. 2002 Jan-Feb;11(1):40-2. doi: 10.1067/mse.2002.119853.
10. Eichinger JK, Rao MV, Lin JJ, Goodloe JB, Kothandaraman V, Barfield WR, et al. The effect of body mass index on internal rotation and function following anatomic and reverse total shoulder arthroplasty. J. Shoulder Elbow Surg. 2021 Feb;30(2):265–272. doi:10.1016/j.jse.2020.06.008
11. Friedman RJ, Eichinger J, Schoch B, Wright T, Zuckerman J, Flurin PH, Bolch C, Roche C. Preoperative parameters that predict postoperative patient-reported outcome measures and range of motion with anatomic and reverse total shoulder arthroplasty. JSES Open Access. 2019 Nov 18;3(4):266-272. doi: 10.1016/j.jses.2019.09.010.
12. Friedman RJ, Flurin PH, Wright TW, Zuckerman JD, Roche CP. Comparison of reverse total shoulder arthroplasty outcomes with and without subscapularis repair. J Shoulder Elbow Surg. 2017 Apr;26(4):662-668. doi: 10.1016/j.jse.2016.09.027.
13. Flurin PH, Marczuk Y, Janout M, Wright TW, Zuckerman J, Roche CP. Comparison of outcomes using anatomic and reverse total shoulder arthroplasty. Bull Hosp Jt Dis (2013). 2013;71 Suppl 2:101-7.
14. Gortmaker SL, Hosmer DW, Lemeshow S. Applied logistic regression. Contemp Sociol. 1994;23:159.
15. Gowd AK, Agarwalla A, Amin NH, Romeo AA, Nicholson GP, Verma NN, Liu JN. Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. J Shoulder Elbow Surg. 2019 Aug 2. pii: S1058-2746(19)30352-0. doi: 10.1016/j.jse.2019.05.017.
16. Henn RF, Ghomrawi H, Rutledge JR, Mazumdar M, Mancuso CA, Marx RG. Preoperative patient expectations of total shoulder arthroplasty. J. Bone Joint Surg. Am. 2011 Nov 16;93(22):2110–2115. doi:10.2106/JBJS.J.01114
17. Hosmer Jr, DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. Hoboken, NJ: John Wiley & Sons; 2013:177.
18. Kim MS, Jeong HY, Kim JD, Ro KH, Rhee S-M, Rhee YG. Difficulty in performing activities of daily living associated with internal rotation after reverse total shoulder arthroplasty. J. Shoulder Elbow Surg. 2020 Jan;29(1):86–94. doi:10.1016/j.jse.2019.05.031
19. Kim SC, Lee JE, Lee SM, Yoo JC. Factors affecting internal rotation after reverse shoulder arthroplasty. J. Orthop. Sci. Off. J. Jpn. Orthop. Assoc. 2021 Jan 23;doi:10.1016/j.jos.2020.11.012
20. Kumar, V, Roche C, Overman S, Simovitch R, Flurin PH, Wright T, et al. What is the accuracy of Three Different Machine Learning Techniques to Predict Clinical Outcomes After Shoulder Arthroplasty. CORR. In press. 2020.
21. Kumar, V, Roche C, Overman S, Simovitch R, Flurin PH, Wright T, et al. Using Machine Learning to Predict Clinical Outcomes After Shoulder Arthroplasty with a Minimal Feature Set. J Shoulder Elbow Surg. In press. 2020.
22. Kumar, V, Roche C, Overman S, Simovitch R, Flurin PH, Wright T, et al. Use of Machine Learning to Assess the Predictive Value of 3 Commonly Used Clinical Measures to Quantify Outcomes After Total Shoulder Arthroplasty. Seminars in Arthroplasty: JSES. In print. 2020.
23. Levy JC, Ashukem MT, Formaini NT. Factors predicting postoperative range of motion for anatomic total shoulder arthroplasty. J. Shoulder Elbow Surg. 2016 Jan;25(1):55–60. doi:10.1016/j.jse.2015.06.026
24. Lipton ZC. The mythos of model interpretability. arXiv:1606.03490, 2016. <https://arxiv.org/abs/1606.03490>
25. McLendon PB, Christmas KN, Simon P, Plummer OR, Hunt A, Ahmed A, Mighell MA, Frankle MA. Machine Learning Can Predict Level of Improvement in Shoulder Arthroplasty. JBJS Open Access. 2021: e20.00128.
26. Monir JG, Tams C, Wright TW, Parsons M, King JJ, Schoch BS. Preoperative factors associated with loss of range of motion after reverse shoulder arthroplasty. J. Shoulder Elbow Surg. [Internet]. 2021 Mar 4 [cited 2021 Jun 23];Available from: https://www.sciencedirect.com/science/article/pii/S1058274621001518doi:10.1016/j.jse.2021.02.010
27. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016 Sep 29;375(13):1216-9. doi:10.1056/NEJMp1606181.
28. Polce EM, Kunze KN, Fu MC, Garrigues GE, Forsythe B, Nicholson GP, et al. Development of supervised machine learning algorithms for prediction of satisfaction at 2 years following total shoulder arthroplasty. J. Shoulder Elbow Surg. 2021 Jun;30(6):e290–e299. doi:10.1016/j.jse.2020.09.007
29. Puzzitiello RN, Nwachukwu BU, Agarwalla A, Cvetanovich GL, Chahla J, Romeo AA, et al. Patient Satisfaction After Total Shoulder Arthroplasty. Orthopedics. 2020 Nov 1;43(6):e492–e497. doi:10.3928/01477447-20200812-03
30. Roche C, Kumar**,** V, Overman S, Simovitch R, Flurin PH, Wright T, et al. Validation of a Machine Learning Derived Clinical Metric to Quantify Outcomes after TSA. J Shoulder Elbow Surg. In press. 2021.
31. Roche C, Simovitch R, Flurin, PH, et al. Comparison of the Accuracy Associated with Three Different Machine Learning Models to Predict Outcomes After Anatomic Total Shoulder Arthroplasty and Reverse Total Shoulder Arthroplasty. Orthopaedic Proceedings. Vol. 102-B. No. SUPP\_1. Feb 2020.
32. Rojas J, Joseph J, Srikumaran U, McFarland EG. How internal rotation is measured in reverse total shoulder arthroplasty: a systematic review of the literature. JSES Int. 2019 Dec 20;4(1):182-188. doi: 10.1016/j.jses.2019.10.109.
33. Rol M, Favard L, Berhouet J, la Société d’orthopédie de l’Ouest (SOO). Factors associated with internal rotation outcomes after reverse shoulder arthroplasty. Orthop. Traumatol. Surg. Res. OTSR. 2019 Dec;105(8):1515–1519. doi:10.1016/j.otsr.2019.07.024
34. Simovitch R, Flurin PH, Wright T, Zuckerman JD, Roche CP. Quantifying success after total shoulder arthroplasty: the minimal clinically important difference. J Shoulder Elbow Surg. 2018 Feb;27(2):298-305. doi: 10.1016/j.jse.2017.09.013.
35. Simovitch R, Flurin PH, Wright T, Zuckerman JD, Roche CP. Quantifying success after total shoulder arthroplasty: the substantial clinical benefit. J Shoulder Elbow Surg. 2018 May;27(5):903-911. doi: 10.1016/j.jse.2017.12.014.
36. Southard EJ, Ode G, Simon P, Christmas KN, Pamic D, Collin P, et al. Comparing patient-reported outcome measures and physical examination for internal rotation in patients undergoing reverse shoulder arthroplasty: does surgery alter patients’ perception of function? J. Shoulder Elbow Surg. 2021 Feb 16;doi:10.1016/j.jse.2021.01.020
37. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010 Jan;21(1):128-38. doi: 10.1097/EDE.0b013e3181c30fb2.
38. Stigler SM. (1986). The History of Statistics: The Measurement of Uncertainty before 1900. Cambridge: Harvard. 1986. ISBN 0-674-40340-1.
39. Swarup I, Henn CM, Gulotta LV, Henn RF. Patient expectations and satisfaction in orthopaedic surgery: A review of the literature. J. Clin. Orthop. Trauma [Internet]. 2019 [cited 2021 Jun 22];10(4):755–760. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6611830/doi:10.1016/j.jcot.2018.08.008
40. Torlay L, Perrone-Bertolotti M, Thomas E, Baciu M. Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. Brain Inform. 2017;4:159-169.
41. Triffitt PD. The relationship between motion of the shoulder and the stated ability to perform activities of daily living. J Bone Joint Surg Am. 1998 Jan;80(1):41-6. doi: 10.2106/00004623-199801000-00008.
42. Triplet JJ, Everding NG, Levy JC, Moor MA. Functional internal rotation after shoulder arthroplasty: a comparison of anatomic and reverse shoulder arthroplasty. J. Shoulder Elbow Surg. 2015 Jun;24(6):867–874. doi:10.1016/j.jse.2014.10.002
43. Werner BC, Lederman E, Gobezie R, Denard PJ. Glenoid lateralization influences active internal rotation after reverse shoulder arthroplasty. J. Shoulder Elbow Surg. 2021 Mar 19;doi:10.1016/j.jse.2021.02.021
44. Wright MA, Keener JD, Chamberlain AM. Comparison of Clinical Outcomes After Anatomic Total Shoulder Arthroplasty and Reverse Shoulder Arthroplasty in Patients 70 Years and Older With Glenohumeral Osteoarthritis and an Intact Rotator Cuff. J. Am. Acad. Orthop. Surg. 2020 Mar 1;28(5):e222–e229. doi:10.5435/JAAOS-D-19-00166
45. Zheng Z, Yang Y, Niu X, Dai HN, Zhou Y. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. *IEEE Transactions on Industrial Informatics*, 2017;14:1606-1615.