The Impact of Deltoid CT Image Data on the Accuracy of Machine Learning Clinical Decision Support Tool Predictions of Clinical Outcomes After Anatomic and Reverse Total Shoulder Arthroplasty

Kumar V, Hamid O, Berry D, Singh A, Schoch B, Polakovic S, Roche C, Elwell J

Disclosures: Kumar V: 3A Exactech, Inc. Hamid O: 3A Exactech, Inc. Berry D: Nothing to Disclose. Singh A: Nothing to Disclose. J. Schoch B: 3B: Exactech, Inc. Polakovic S: 3A Exactech, Inc. Elwell: 3A; Exactech, Inc; Roche C: 3A; Exactech, Inc. 4: Exactech, Inc

INTRODUCTION: Clinical decision support tools (CDSTs) are software that generate patient-specific assessments that can be used to better inform healthcare provider decision making. Machine learning (ML) based CDST have been developed for predicting clinical outcomes after anatomic and reverse total shoulder arthroplasty (aTSA and rTSA). These CDSTs can provide health care providers with personalized predictions of post-operative outcomes based upon input of patient specific preoperative data, like demographics, diagnoses, patient reported outcome measures, including measures of pain, function, and active range of motion. In one particular ML based CDST, Predict+ (Exactech, Inc, Gainesville, FL), the specific preoperative inputs utilized by these ML algorithms were selected from a larger pool of data by rank-ordering the relative "preoperative input feature importance" to the ML model to identify the specific preoperative inputs that were most predictive of clinical outcomes after aTSA and rTSA.[1] Currently, this list of Predict+ preoperative input requirements, termed the minimal feature set, does not include any direct CT image data or any data related to the quality of a patients soft-tissue or bone.[1] Because of the wide-spread clinical use of CT-based pre-operative planning software tools for both aTSA and rTSA, ample CT data is readily available in existing clinical workflows to enhance CDST predictions and further support clinical decision making by including each patient's CT image information related to their bones and soft-tissue. However, it is currently unknown specifically which CT image data is useful to improve the accuracy of the ML based CDST outcome predictions. Therefore, the aim of this pilot study is to quantify the "feature importance" of various CT image data describing the deltoid and evaluate the addition of this CT image data to improve the accuracy of 2-year clinical outcomes predictions after aTSA and rTSA.

METHODS: 224 patients (55 aTSA and 169 rTSA) from a multi-center, IRB-approved, international database of a single platform shoulder arthroplasty system were analyzed in this pilot study. Each patient had pre-operative 3D CT scans, preoperative data to satisfy the minimal feature set requirements, along with 2-year minimum clinical data for various clinical outcome measures including, active range of motion: abduction, forward elevation, external rotation, and internal rotation score, and patient reported outcome measures, including: VAS pain, global shoulder function, and the shoulder arthroplasty smart (SAS) score. Pre-operative deltoid characteristics were quantified by employing a CT-based machine-learning algorithm which automatically segments the anterior, middle, and posterior deltoid into a single 3D model. A quantification algorithm extracts numerous information about the deltoid, including deltoid volume, which is normalized to the volume of the scapular bone volume, the percentage of deltoid fat, which describes the percentage of voxels in the deltoid that are within the Hounsfield unit range of fat, and normalized deltoid atrophy, which is a normalized volume divided by the normalized volume of a reference group of patients of similar age and gender. Additional radiomic features were extracted using an open-source python package, PyRadiomics, and included the following first order features related to the voxel intensity values: mean, 90th percentile and skewness, a measure of the asymmetry of the distribution about the mean. XGBoost was utilized to predict 2-year clinical outcomes for each outcome measure both before and after addition of the aforementioned CT derived image data. The relative importance of each preoperative input data to predict each 2-year clinical outcome measure was quantified by the F-score and the accuracy of each 2-year prediction was quantified by the Mean Absolute Error (MAE), which describes the mean absolute difference between the actual and predicted values of each 2-year outcome measure.

RESULTS: The addition of the CT-based deltoid image data was useful for development of each 2-year minimum clinical outcome predictive model. By rank-ordering the F-scores associated with each pre-operative input, for each of the 2-year minimum clinical outcomes, it was observed that all 6 of the deltoid CT-image measurements had a feature importance rank in the top 20 of the 47 overall preoperative inputs. As described in Table 1, the deltoid image measurements with the highest range feature importance rank were the deltoid voxel skewness (top 5 feature in 5 of 7 models), followed by the normalized deltoid volume (top 5 feature in 5 of 7 models), and deltoid fat percentage (top 5 feature in 3 of 7 models). The next 3 deltoid image measurements were the mean deltoid voxel (top 10 feature in 4 of 7 models, and the #1 ranked input for the SAS outcome model), deltoid atrophy (top 10 feature in 4 of 7 models), and finally the 90th percentile deltoid voxel (top 10 feature in 1 of the 7 models). Generally, the addition of the deltoid image data slightly improved the accuracy of the 2-year minimum clinical outcome prediction models. Where specifically, the addition of the deltoid measurements was associated with a reduction in the MAE for prediction of global shoulder function (4.3% improvement), external rotation (3.0% improvement), abduction (2.4% improvement), VAS pain (0.5% improvement), and forward elevation (0.2% improvement). However, the models with the deltoid image data were found to be less accurate for prediction of internal rotation (4.4% worse) and the SAS score (1.7% worse). Interestingly, the addition of the deltoid image data seemed to be more helpful to improve the prediction of aTSA outcomes, most notably a 26.6% improvement in external rotation and a 18.7% improvement in SAS score.

DISCUSSION: The results of this pilot study demonstrate that CT image data can be used to develop ML-based 2-year minimum clinical outcome predictions after aTSA and rTSA. By rank ordering the input feature importance of each clinical outcome model, it was demonstrated that all 6 of the evaluated deltoid measurements were useful data to make predictions. Additionally, the use of this deltoid image data was demonstrated to improve ML model accuracy for the majority of the 2-year minimum clinical outcome models. This pilot study has numerous limitations. First, 224 patients is a relatively small cohort and there was an unequal distribution of patients by prosthesis type, age, gender, image kernel, and other potentially relevant variables. Second, this pilot study only characterized the deltoid using 6 measurements. Future work should continue to assess CT based deltoid image measurements and identify the parameters which are most predictive of clinical outcomes after aTSA and rTSA. Third, deltoid image data may not be the most useful data available in these CT scans to predict clinical outcomes after aTSA and rTSA. Future work should assess the value of other CT based image measurements, and specifically evaluate the utility of rotator cuff CT image data on aTSA and rTSA clinical outcomes and also evaluate the utility of CT based bone measurements of the proximal humerus, glenoid vault, acromion, and scapular spine.

SIGNIFICANCE/CLINICAL RELEVANCE: This is the first orthopedic study to evaluate the efficacy of CT image based muscle data to develop machine learning based clinical outcome predictions. Our results demonstrated that several deltoid image measurements were useful for machine learning model development and the inclusion of this muscle data resulted in an improvement of accuracy for the majority of clinical outcome predictions after aTSA and rTSA.

REFERENCES:

1. Kumar V, Roche C, Overman S, Simovitch R, Flurin P-H, Wright T, et al. Using machine learning to predict clinical outcomes after shoulder arthroplasty with a minimal feature set. J Shoulder Elbow Surg. 2021 May;30(5):e225–e236.

Table 1. Comparison of the Mean Absolute Error (MAE) and the Relative Feature Importance of Each CT-based Deltoid Image Parameter to Predict 2-year Clinical Outcomes after aTSA and rTSA

Clinical Outcomes Prediction	**MAE of Predict+ w/o Deltoid Image Data (aTSA, rTSA)**	**MAE of Predict+ with Deltoid Image Data**	**Percent Difference in MAE**	**F Score Rank: Mean Deltoid Voxel**	**F Score Rank: Skewness of Deltoid Voxels**	**F Score Rank: Deltoid Fat Percentage**	**F Score Rank: Normalized Deltoid Volume**	**F Score Rank: Deltoid Atrophy (Normalized by Age & Gender)**	**F Score Rank: 90 percentile Deltoid Voxel**
2yr Abduction	18.51 (11.7, 20.5)	18.07 (12.8, 20.0)	2.4% (-9.9%, 2.5%)	15	3	9	5	13	14
2yr External Rotation	12.27 (11.8, 12.4)	11.90 (8.6, 12.8)	3.0% (26.6%, -2.7%)	12	3	9	4	8	19
2yr Forward Elevation	13.07 (9.8, 14.0)	13.05 (9.2, 14.5)	0.2% (6.0%, -3.7%)	8	6	2	3	13	16
2yr Internal Rotation Score	1.28 (1.1, 1.4)	1.34 (1.1, 1.5)	-4.4% (1.2%, -7.5%)	11	5	4	6	15	8
2yr SAS	8.30 (8.6, 8.2)	8.45 (7.0, 8.8)	-1.7% (18.7%, -7.0%)	1	4	7	9	10	15
2yr Global Shoulder Function	1.44 (1.4, 1.5)	1.37 (1.4, 1.4)	4.3% (2.8%, 4.6%)	9	6	7	5	10	18
2yr VAS Pain	1.66 (1.6, 1.7)	1.65 (1.7, 1.6)	0.5% (-4.2%, 1.7%)	10	5	2	4	3	15
			Average F Score Rank (out of 47):	9.4	4.6	5.7	5.1	10.3	15