


## Research Article

# Effect of machine learning prediction on surgical decision making for shoulder arthroplasty: a multi-surgeon study

Dave R. Shukla, MD<sup>1</sup><sup>a</sup>, Brian J. Rebolledo, MD<sup>2</sup><sup>b</sup>, Alexander W. Aleem, MD<sup>3</sup><sup>c</sup>, Adrien Jacquot<sup>4,5</sup><sup>d</sup>, Jean-David Werthel<sup>6,7</sup><sup>e</sup>, Diego Villacis, MD<sup>8</sup><sup>f</sup>, Manuel Urvoy<sup>9</sup><sup>g</sup>

<sup>a</sup> Dr. Dave Shukla is an Assistant Professor of Orthopedic Surgery with subspecialty training in Shoulder and Elbow Surgery. Dr. Shukla completed his undergraduate degree at the University of California Berkeley and medical degree at University College Dublin. His orthopedic surgical training started at the Mayo Clinic where he completed his General Surgery Internship. He then completed a two-year Biomechanics Research Fellowship at Mayo Clinic, focusing on traumatic elbow injuries. He completed Orthopedic Surgery Residency at the Icahn School of Medicine at Mount Sinai in New York, and then returned to the Mayo Clinic for a one-year Subspecialty Fellowship in Adult Reconstruction of the Upper Extremity (Shoulder and Elbow surgery). After fellowship he joined a multi-specialty group in Southern California where he became a partner, and was involved in hospital Quality Control and developed an Enhanced Recovery Protocol for Shoulder Arthroplasty. He has been very active in research and has maintained an interest in advancing the field of shoulder and elbow surgery.

He is committed to practicing evidence-based medicine, and is an author on over 50 peer-reviewed publications. He has had 24 presentations at national and international meetings, and serves as a consultant editorial reviewer for four orthopedic journals. Dr. Shukla is a member of the American Shoulder and Elbow Surgeons, a Fellow of the American Academy of Orthopedic Surgeons and is Board Certified by the American Board of Orthopedic Surgery.

[Visit Dr. Shukla's Website](#)

[Connect with Visit Dr. Shukla on LinkedIn](#)

[Visit the Open Payments Data Page for Dr. Shukla](#)

[Conflicts of Interest Statement for Dr. Shukla](#)

<sup>b</sup> Dr. Rebolledo graduated with Honors from Weill Cornell Medical College and received the T. Campbell Thompson Award for Excellence in Orthopaedic Surgery. He completed his residency at the renowned Hospital for Special Surgery in New York City, where he was recognized with the Thomas P. Sculco, MD Award for outstanding patient care, compassion, and professionalism. Dr. Rebolledo further specialized in sports medicine through a fellowship at Stanford University, during which he worked as an assistant team physician for the San Francisco 49ers. He has pursued additional training opportunities nationwide and was selected as a Global Traveling Fellow by the International Society of Arthroscopy, Knee Surgery, and Orthopaedic Sports Medicine (ISAKOS). Regarding executive education, Dr. Rebolledo completed the Surgical Leadership Program at Harvard Medical School and is pursuing an Executive MBA at The Wharton School at The University of Pennsylvania.

Dr. Rebolledo's commitment to research is evident through his extensive publications, including over 50 scientific papers and book chapters. He also contributes as a journal reviewer for prominent publications such as the American Journal of Sports Medicine, Orthopaedic Journal of Sports Medicine, and HSS Journal.

[Visit Dr. Rebolledo's Website](#)

[Connect with Visit Dr. Rebolledo on LinkedIn](#)

[Visit the Open Payments Data Page for Dr. Rebolledo](#)

[Conflicts of Interest Statement for Dr. Rebolledo](#)

<sup>c</sup> Dr. Aleem is an Associate Professor, Residency Program Director and Vice Chair of Education at Washington University in St. Louis, Department of Orthopaedic Surgery. His clinical interests range across the spectrum of diseases in shoulder and elbow. He has particular interest in shoulder arthroplasty innovations. Academically he is also involved in new ways to educate surgical trainees.

[Visit Dr. Aleem's Website](#)

[Connect with Dr. Aleem on LinkedIn](#)

[Visit the Open Payments Data Page for Dr. Aleem](#)

[Conflicts of Interest Statement for Dr. Aleem](#)

<sup>d</sup> Dr. Adrien Jacquot completed his entire medical course at the University of Nancy, France, where he received his Medical Doctorate and Diploma of Orthopaedic and Traumatologic Surgery in 2013. He spent 4 years as a Fellow in the Orthopaedic, Traumatologic and Sport Surgery Department of Pr. Daniel Molé in Nancy.

He works as a Shoulder surgeon in the Artic.S Joints and Sports Surgery

Center in Nancy, with Pr. Daniel Molé, Dr. Thomas Goetzmann and Dr. Frank Wein, since 2017. He has a special interest for shoulder arthroplasty and more particularly for 3D imaging and pre-operative planning, and for shoulder instability, particularly the development and promotion of arthroscopic Latarjet. He is also specialized in Hip Arthroscopy.

Dr. Jacquot is a member of the French Society of Orthopaedic and Traumatologic Surgery (SOFOT), of the French College of Orthopaedic and Traumatologic Surgery (CFCOT), and of the French Society of Arthroscopy (SFA), and board member of CAOS France (Computer Assisted Orthopaedic Surgery)

[Visit Dr. Jacquot's Website](#)

[Connect with Visit Dr. Jacquot on LinkedIn](#)

[Conflicts of Interest Statement for Dr. Jacquot](#)

<sup>1</sup> Orthopedic Surgery, Icahn School of Medicine at Mount Sinai, <sup>2</sup> Orthopedic Surgery, Scripps Clinic, <sup>3</sup> Orthopedic Surgery, Washington University, <sup>4</sup> Orthopedic Surgery, SAS Clinique Louis Pasteur, <sup>5</sup> Orthopedic Surgery, Chirurgie des Articulations et du Sport, <sup>6</sup> Orthopedic Surgery, Hôpital Ambroise-Paré, <sup>7</sup> Orthopedic Surgery, Ramsay Generale de Sante, Clinique La Montagne, <sup>8</sup> Orthopedic Surgery, Aurora Medical Center, <sup>9</sup> Imascap

Keywords: artificial intelligence, machine learning, shoulder arthroplasty, reverse shoulder replacement, anatomic shoulder replacement

<https://doi.org/10.60118/001c.84084>

---

## Journal of Orthopaedic Experience & Innovation

Vol. 4, Issue 2, 2023

---

### Background

Machine learning algorithms for surgical decision making in shoulder arthroplasty has not been reported. Though there are recommendations based on available literature regarding the selection of anatomic versus reverse shoulder replacement, there are no clear guidelines on how this decision should be made. Our aim was to assess the viability of machine learning for this application by evaluating the agreement of the algorithm's recommendation on type of arthroplasty versus that recommended by six shoulder surgeons.

### Methods

There were 84 cases of patients with glenohumeral osteoarthritis planned using a three-dimensional CT-based software. Half of the cases were planned without any ML algorithm-based recommendation available, and half were planned with the recommendation available, and kappa coefficients were calculated to determine agreement.

### Results

In 78% of cases, the software's surgical recommendation on arthroplasty type completely aligned with that of the surgeons. The Cohen's kappa coefficients for surgeons' versus software's recommendations were 0.56 and 0.61 for rounds 1 and 2, respectively, while the Fleiss kappa coefficients (inter-surgeon agreement) were 0.87 and 0.77.

### Discussion

Machine learning for the application of guiding surgeons on which type of shoulder arthroplasty to select has demonstrated viability in this study, with further research needed to refine this system.

---

<sup>e</sup> Jean-David Werthel is associate professor at Hôpital Ambroise Paré where he is responsible for the shoulder unit. He completed his training in Paris and spent a year as a research fellow in the biomechanics lab of the Mayo Clinic. He also completed a PhD in computer science at the IMT Atlantique in partnership with Imascap.

His interests are in complex reconstructive surgery of the shoulder and computer-assisted surgery as well as in basic science research related to the biomechanics of the shoulder and of shoulder implants. He is involved in the design of implants for shoulder replacements and has presented his research in numerous occasions at national and international meetings. Jean-David has published over one hundred scientific articles and is a reviewer for shoulder articles and corresponding member in several journals.

[Conflicts of Interest Statement for Jean-David Werthel](#)

<sup>f</sup> [Visit Dr. Villacis's Website](#)

[Visit the Open Payments Data Page for Dr. Villacis](#)

[Conflicts of Interest Statement for Dr. Villacis](#)

<sup>g</sup> [Connect with Manuel Urvoy on LinkedIn](#)

[Conflicts of Interest Statement for Manuel Urvoy](#)

## INTRODUCTION

Pre-operative planning for shoulder arthroplasty has undergone an evolution from the use of two-dimensional computed tomography (CT) scans to three-dimensional CT scans which can be modulated within templating software for accurate implant positioning. Glenoid component placement within optimal parameters of version and inclination dramatically affects longevity in both anatomic shoulder arthroplasty (aTSA) and reverse shoulder arthroplasty (rTSA). Studies have demonstrated that three-dimensional templating software can improve the accuracy of glenoid component placement in both aTSA and rTSA (Iannotti et al. 2014, 2015; Walch et al. 2015; Venne et al. 2015). The software can also allow for the manufacturing of patient-specific instrumentation, the use of which can increase operative reproducibility relative to a surgeon's initial plan (Hendel et al. 2012; Levy et al. 2014; Throckmorton et al. 2015). Planning software influences implant positioning accuracy, as well as selection of the appropriate implant type (i.e. aTSA versus rTSA), both of which can affect outcomes and survival rates of shoulder arthroplasty.

A surgeon's decision of whether to perform an aTSA versus an rTSA is based on one's assessment of a patient's lifestyle and demographic background, as well as specific anatomic features. The presence of a massive rotator cuff tear or significant glenoid deformity are generally accepted as guidelines in support of an rTSA. However, surgeons are often faced with borderline cases for which no clear guidelines exist. For example, in a patient with an intact but degenerated rotator cuff, decision-making is driven by a surgeon's intuition and based on evolving data, with recent trends supporting the use of an rTSA in elderly cuff-intact patients. A recent study demonstrated that the rate of secondary rotator cuff dysfunction at ten years was 16.8%, with a 55% risk of rotator cuff dysfunction or revision surgery at 15 years (Young et al. 2012). Furthermore, recently published data have shown that the outcomes following both types of arthroplasties may not be as markedly different as once thought. Shoch et al. reported greater satisfaction, fewer complications and fewer reoperations following rTSA compared to aTSA, with similar patient reported outcomes between the two (Schoch et al. 2020).

Machine learning (ML) algorithms are gaining popularity within orthopedic surgery. Machine learning is also known, and commonly referred to, as 'Artificial Intelligence' (AI). Reported uses include the prediction of outcomes after total joint arthroplasty (Fontana et al. 2019; Huber, Kurz, and Leidl 2019; Kunze, Polce, Patel, et al. 2021; Kunze et al. 2020), hip arthroscopy (Kunze, Polce, Clapp, et al. 2021; Kunze, Polce, Nwachukwu, et al. 2021), readmission rates (Arvind et al. 2021), and post-operative complications (Gowd et al. 2019) among other things. ML models have also been applied to assist in patient selection (Biron et al. 2019) as well as in the prediction of unplanned readmissions (Arvind et al. 2021) and complications (Gowd et al. 2019) following shoulder arthroplasty. To our knowledge, the application of ML algorithms in the prediction of shoulder arthroplasty type, specifically regarding whether to per-

form an anatomic versus reverse shoulder arthroplasty, has not been studied.

The primary objective of this study was to determine whether an ML-based software's prediction of optimal shoulder arthroplasty type based on CT imaging would affect overall agreement amongst a group of expert surgeons. We hypothesized that by providing an 'artificial intelligence' software's surgical prediction, agreement among six surgeons would increase compared to overall agreement in the absence of an algorithmic prediction. The secondary objective of this study was to evaluate the accuracy of the prediction of the ML algorithm compared to an experienced surgeon. We hypothesized that the artificial intelligence software could accurately predict an experienced surgeon's decision regarding procedure type.

## METHODS

Eighty-four shoulders with a diagnosis of primary glenohumeral osteoarthritis were used in this study. The mean age of the 84 patients was 68 (34 to 89) years. The images were obtained from different organizations, using a standardized protocol that resulted in uniform output from a previously-validated three-dimensional planning software Glenosys (Imascap, Brest, France) (Walch et al. 2015).

The 84 cases were divided into two subsets of 42 cases. Two groups of three surgeons were created randomly. Each of the six surgeons is considered a high-volume arthroplasty surgeon at each of their respective institutions, and all were within the first ten years of their respective practices. Four of the surgeons practice in the United States and two practice in France. None of the surgeons were involved in the conception or design of this software.

Surgeons number 1, 2 and 3 were in Group I and surgeons number 4, 5 and 6 were in Group II. Each group of surgeons received one subset of 42 cases. During a first round, each surgeon of each group had to plan the 42 cases with software's prediction blinded, resulting in a total of 252 plans without prediction (2 groups of 3 surgeons, each surgeon planning 42 cases). In a second round, each surgeon of each group had to plan the 42 cases with the prediction visible prior to surgical decision making, resulting in a total of 252 plans with prediction. There was a five-month interval between the planning of each round of the same cases to reduce recall bias.

The software provided a prediction based on several factors: Age, Gender, Glenoid Orientation, Glenoid Sphere Radius, Glenoid Version, Glenoid Inclination, Humeral Subluxation, Glenoid Direction, Glenoid Area. It was previously trained to predict the majority vote of a group of shoulder specialists. The software was initially developed using 210 cases that were not used during software training. The cases were performed by four surgeons who are considered international experts in the field of shoulder arthroplasty, none of whom were involved in this study. The software analyzed those factors mentioned above and learned which combination of values typically resulted in the decision for an aTSA versus an rTSA. The agreement between the prediction and the majority vote of the shoulder specialists was



©2022 Pacira Pharmaceuticals, Inc., a wholly owned subsidiary of Pacira Biosciences, Inc. All rights reserved. PP-NP-US-1328 05/22

\*Advertisement

[Click here to learn more about Pacira](#)

measured using Cohen's kappa. A value of 0.71 was found: there was a substantial agreement between the algorithmic prediction and the majority vote of the surgeons.

#### PRE-OPERATIVE PLANNING WITHOUT ML-PREDICTION

From November to December 2020 (round 1), planning without AI prediction display was performed, as one would typically do in clinical practice currently. The decision of whether to perform an anatomic or reverse shoulder arthroplasty was based off the automated three-dimensional segmented image, the glenoid measurements and CT-based assessment of rotator cuff integrity. The details regarding the measurements generated automatically by the software (i.e. glenoid version, glenoid inclination and humeral head subluxation) have been previously described in the peer-reviewed literature (Boileau et al. 2018; Shukla, McLaughlin, Lee, Nguyen, et al. 2019). In general, a reverse shoulder arthroplasty was selected in shoulders with Goutallier grade 3 and 4 fatty infiltration (Goutallier et al. 1994), superior glenoid inclination over 10 degrees given the potential for eccentric loading-related early implant failure (Kandemir et al. 2006; Walch, Young, et al. 2012), glenoid biconcavity with glenoid retroversion over 27 degrees (or if an aTSA polyethylene glenoid could not be positioned with less than 10 degrees of retroversion) or humeral head subluxation over 80% (Mizuno et al. 2013; Walch, Moraga, et al. 2012).

#### PRE-OPERATIVE PLANNING WITH ML-PREDICTION

From May to June 2021 (round 2), a software-generated prediction on whether an anatomic or reverse shoulder arthroplasty would be most appropriate was visible to each surgeon at the time of case planning. This prediction was expressed as a percentage of probability (Figure 2) in favor of aTSA or rTSA. Each case was then planned at the surgeon's discretion.

#### STATISTICAL ANALYSIS

Statistical analysis was performed on MedCalc® Statistical Software version 19.6.4 (MedCalc Software Ltd, Ostend, Belgium; <https://www.medcalc.org>; 2021). The interpretation of Cohen's and Fleiss' kappa values were made as follows, according to previously published boundaries(26): < 0 (less than chance agreement); 0.01 – 0.20 (slight agreement); 0.21 – 0.40 (fair agreement); 0.41 – 0.60 (moderate agreement); 0.61 – 0.80 (substantial agreement); 0.81 – 0.99 (almost perfect agreement). The kappa coefficient has been routinely used to assess agreement within the ortho-

pedic literature (Atoun et al. 2017; Heuer et al. 2014; Parsons et al. 2020; Ricchetti et al. 2021; Shukla, McLaughlin, Lee, Cofield, et al. 2019).

#### AGREEMENT BETWEEN EACH SURGEON AND ML PREDICTION

The Cohen's kappa was used to quantify the agreement between each surgeon and ML prediction during round 1 (no prediction displayed) and round 2 (prediction displayed).

#### AGREEMENT BETWEEN THE MAJORITY VOTE OF THE SURGEONS AND ML PREDICTION

The aTSA vs rTSA surgeons' majority vote was registered for each of the two rounds and for each case. During round 1, aTSA or rTSA majority vote was obtained on each of the 84 cases without software's prediction displayed. During round 2, aTSA or rTSA majority vote was obtained on each of the same 84 cases with software's prediction displayed. The Cohen's kappa was calculated for each round, between the majority vote of the surgeons and the ML prediction. This was done to evaluate how the display of AI prediction could affect the majority vote among a group of surgeons.

#### AGREEMENT BETWEEN THE SURGEONS OF EACH GROUP

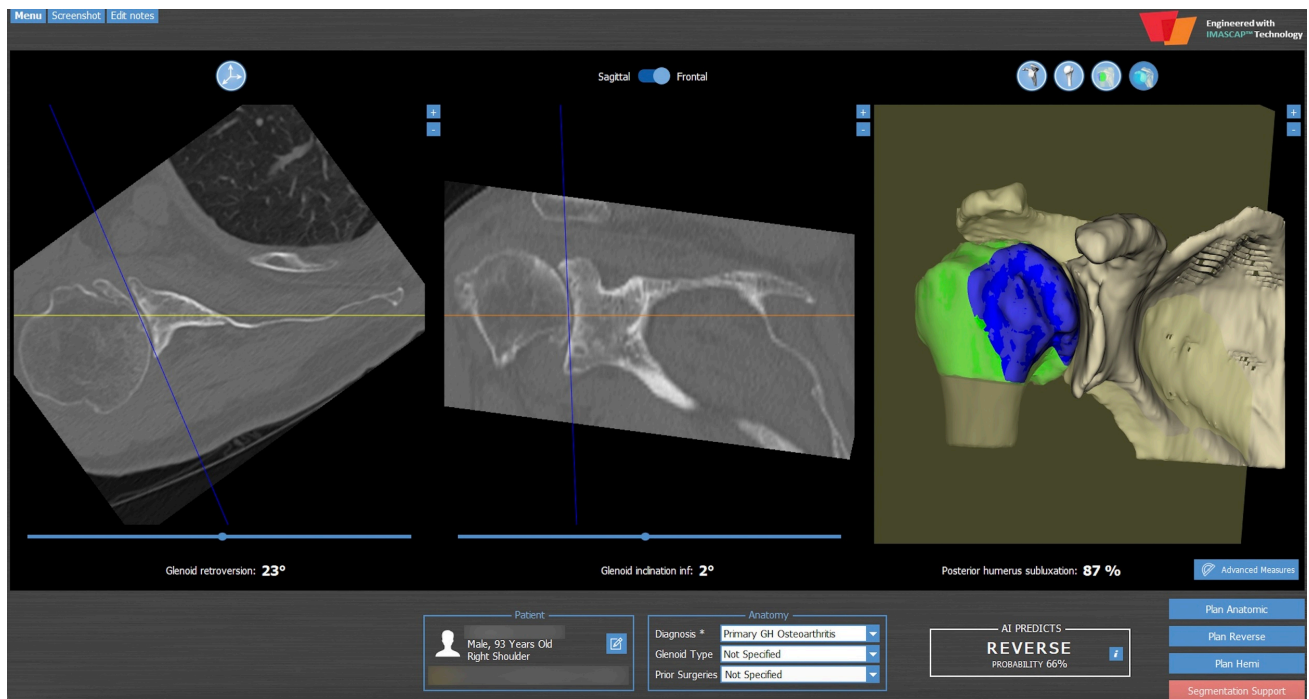
Regarding each group of surgeons (group 1 and group 2), the Fleiss' kappa was used to quantify the agreement between the surgeons of the group, with (round 1) and without (round 2) ML prediction displayed.

## RESULTS

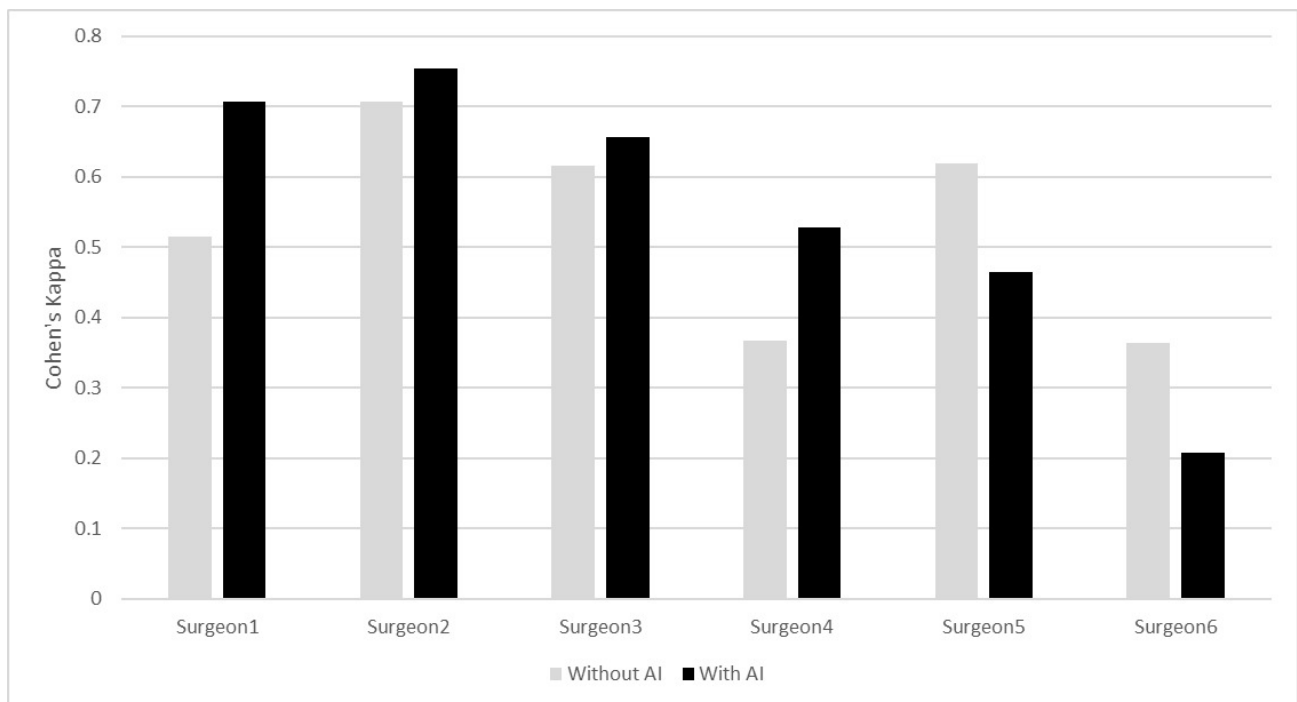
#### MACHINE LEARNING PREDICTABILITY

The mean age of the 84 patients was 68 (range: 34 to 89) years. In 395 / 504 (78%) plans there was perfect agreement between the software-predicted surgery type and the surgeon's decision, with disagreement being present in 109 / 504 (22%). In round I (prediction blinded) there was perfect agreement between the surgeon and algorithm in 194 / 252 cases (77%), and in round II, there was perfect agreement in 201 / 252 cases (80%). For four of the six surgeons, the kappa value increased when cases were planned with software prediction (WP) compared to having no prediction (NP) available (Figure 1), while the kappa values decreased for the other two surgeons.

There were two cases for which there was complete disagreement with the software both with and without the software prediction, meaning that all six surgeons dis-



**Figure 1.** Sample image of the Blueprint pre-operative planning software, with the 'AI' prediction visible in the bottom right corner.



**Figure 2.** Bar graph demonstrating the agreement amongst surgeons without AI (Artificial Intelligence) prediction and with the AI prediction available.

agreed with the proposed plan. In both cases, an rTSA was preferred over the proposed plan of an aTSA.

This study showed that the availability of a software-generated ML prediction increased the agreement among four out of six surgeons, though these increases in agreement were modest, as no increases in more than one grade

(i.e. 'moderate' to 'substantial', or 'substantial' to 'almost perfect') were observed.



BE A PART OF **BETTER** ORTHOPAEDIC CARE.  
CLICK HERE TO LEARN HOW.

**THE ORTHOPAEDIC IMPLANT COMPANY**



\*Advertisement

[Click here to learn more about The Orthopaedic Implant Company](#)

	Confusion Matrix				Cohen's Kappa	95% CI
Without AI prediction displayed (round 1)	AI Prediction	Majority Vote			0.562	0.388 to 0.735
		RSA	TSA			
	RSA	41	4	45		
	TSA	14	25	39		
With AI prediction displayed (round 2)	AI Prediction	Majority Vote			0.608	0.445 to 0.770
		RSA	TSA			
	RSA	44	1	45		
	TSA	15	24	39		
		55	59	84		
		59	25	84		

Table 1. Kappa's cohen values were determined for the surgeons' majority votes versus the AI (artificial intelligence) software's prediction.

Table 2. Fleiss Kappa values demonstrating the agreement within each group, with and without the software's prediction displayed.

Group Agreement (Fleiss Kappa)	Group 1	Group 2
Without AI prediction displayed (round 1)	k = 0.872	k = 0.765
With AI prediction displayed (round 2)	k = 0.935	k = 0.738

#### AGREEMENT BETWEEN THE MAJORITY VOTE OF THE SURGEONS AND THE ML PREDICTION ([TABLE I](#))

The agreement between majority votes and the software's prediction evolved from moderate ( $k = 0.56$ ; 95% CI 0.45 to 0.77) in round 1 (no prediction displayed) to substantial ( $k = 0.61$ ; 95% CI 0.45 to 0.77) in round 2 (prediction displayed).

#### AGREEMENT BETWEEN THE SURGEONS OF EACH GROUP ([TABLE II](#))

During round 1, there was an 'almost perfect' agreement in group 1 ( $k = 0.87$ ) and 'substantial' agreement in group 2 ( $k = 0.77$ ). Similar results were found during round 2, with an increase of the kappa in both groups (0.94 for group 1 and 0.74 for group 2).



[Click here to learn more about Mako Total Knee 2.0](#)

#### DISCUSSION

The ML algorithm evaluated in this study provided a recommendation on shoulder arthroplasty type with 78% accuracy as compared with the surgeons' recommendation. Additionally, the algorithm's prediction influenced surgical decision-making in 10% of cases. Though the presence of a

prediction only resulted in a modest change in agreement between rounds of planning, the data suggest that the software's ability to predict the most appropriate surgery, or that surgery that best aligned with the preferences of the shoulder specialists in this study, was very good. These data also highlight the persistent variation in decision-making amongst surgeons, as consistent with data reported by Parsons et al (Parsons et al. 2020), who demonstrated substantial variation within and amongst surgeons planning rTSA. The authors emphasized that surgeons largely rely on intuition and experience. The primary hypothesis was validated, in that the ML prediction did increase agreement among the surgeons as compared to overall agreement in the absence of the algorithm's prediction. The secondary hypothesis was validated as well, in that there was very strong agreement between the software's predicted surgery type and the surgeons' preference, as there was perfect agreement between the algorithm and surgeon in 78% of cases.

These findings highlight the potential utility of ML for surgical decision-making prior to shoulder arthroplasty. The use of ML for this purpose within orthopedic surgery has yet to be reported to our knowledge, and represents a contribution to the body of literature on this evolving subject.

The use of ML in orthopedic surgery has been reported with increasing frequency (Fontana et al. 2019; Huber, Kurz, and Leidl 2019; Kunze, Polce, Patel, et al. 2021; Kunze et al. 2020; Kunze, Polce, Clapp, et al. 2021; Kunze, Polce, Nwachukwu, et al. 2021; Arvind et al. 2021; Gowd et al. 2019). However, there is a relative paucity of literature on the use of ML with respect to total shoulder arthroplasty. Reddy et al. identified that the maximum areas for influence of ML in healthcare are administration, clinical decision support, patient monitoring and healthcare interventions (Reddy, Fox, and Purohit 2018). Machine learning categories include supervised machine learning, unsupervised machine learning, semi-supervised machine learning and artificial neural networks/deep learning (33).

Arvind et al. reported that predictive analytics algorithms could acceptably predict unplanned readmission following shoulder arthroplasty (Arvind et al. 2021). Biron et al developed an ML model to predict suitable short stay or outpatient shoulder arthroplasty candidates (Biron et al. 2019). Gowd et al recommended continued validation efforts towards the refinement of intelligent models such as for the calculation of patient-specific risk for complications, which was studied in their work (Gowd et al. 2019).

The type of supervised ML used in this study was support vector machine (SVM). Several studies have demonstrated the applicability of this type of ML. Polce et al demonstrated that the SVM algorithm, which was the same type of ML algorithm used in this study, performed best amongst five ML algorithms employed to predict satisfaction following TSA (Polce et al. 2021).

Surgical decision-making is dependent on many clinical features that were not included in this algorithm's analysis, such as patient activity and physiologic age. Additionally, the value of surgeon experience and comfort level could not

be studied here and are critically influential in planning. While a tentative surgical plan can be created pre-operatively, the final decision is sometimes not known until the time of surgery. Challenging cases are often discussed with colleagues or mentors. There is no software-derived algorithm that can or should supplant this process or claim to definitively direct a surgeon towards one arthroplasty type. The ML prediction is rather another data-point that one can use, if he or she chooses, to augment the decision-making process. It can provide surgeons with an additional measure of confidence during planning as the machine-learned prediction is embedded in the planning software, though only if the surgeon chooses to use it.

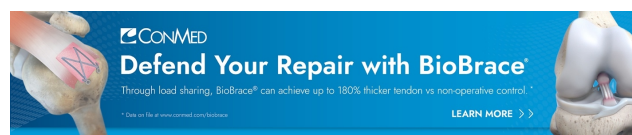
The decision on whether or not to select an aTSA versus an rTSA can be challenging, particularly in borderline cases. The decision is further complicated by the fact that it is becoming less clear whether or not the results of an rTSA are truly inferior to outcomes after an aTSA, such in patients with an intact rotator cuff (Schoch et al. 2020). Cox et al. reported high satisfaction rates and similar outcomes in patients with an aTSA and contralateral rTSA (Cox et al. 2018). Shoch et al. reported higher patient satisfaction, fewer reoperations and fewer complications following rTSA versus aTSA (Schoch et al. 2020). As outcomes and complication rates of the two arthroplasty types continue to approximate each other, surgeons consider other parameters. This study demonstrates that ML can help us to better understand the parameters that have been found to influence surgeons in their choice of implant. The ML algorithm studied here provided a recommendation based on several factors: age, gender, glenoid orientation, glenoid sphere radius, glenoid direction, glenoid area, glenoid inclination and version, and humeral head subluxation value. Though there are other considerations that influence one's decision, the algorithm studied here can provide some measure of guidance based on the quantifiable parameters that were incorporated, and can help to guide this decision with further refining of the software. This software-generated prediction feature is embedded within a widely-available pre-operative planning software. When the surgeon initiates case-planning, the prediction is auto-generated, so that no alteration in planning work-flow is required.

One strength of this study was that multiple surgeons planned the same cases. We observed similar benefits from this as have other multi-surgeon studies (Throckmorton et al. 2015; Parsons et al. 2020). The heterogeneity in results as regards the inability of the software's prediction to uniformly influence surgeon agreement increases the validity of our methodology. Importantly, it also provides a degree of confidence in the software and its ability to align with the surgeons' decisions in 78% of cases.

## LIMITATIONS

One limitation of this study was that surgical decisions were based on CT imaging only, without any potentially influential clinical data, as discussed above. As a natural evolution of this, patient-specific functional parameters and clinical outcomes data will need to be incorporated into

the algorithm, which is an area of future development and study. Additionally, more precise analysis of the pre-operative soft-tissue quality would greatly augment the software's value to surgeons, which is an additional area of future research. Despite these current deficiencies in the algorithm's dataset, this study demonstrated that machine-learning can potentially be used to provide some measure of guidance in surgical-decision for shoulder arthroplasty, with the understanding that ultimately this choice is a matter of experience and of preference. Additionally, though there was the potential for recall bias during the second round of case planning, measures were taken to minimize this risk including the implementation of a five-month interval between planning of the same cases, case order alteration and anonymization through the use of alpha-numeric case identifiers. Another limitation was that there were no defined parameters to which each surgeon adhered when selecting an anatomic versus reverse shoulder arthroplasty. However, this limitation potentially results in a more clinically-realistic or translatable result. As the indications for reverse shoulder arthroplasty continue to evolve, it is likely that surgeons are less-inclined to make decisions solely based on strict anatomic parameters. General guidelines do exist in the literature, and it has been purported that a reverse shoulder arthroplasty should be considered in the presence of Goutallier grade 3 and 4 fatty infiltration (Goutallier et al. 1994), superior inclination over 10 degrees (Kandemir et al. 2006; Walch, Young, et al. 2012), and glenoid biconcavity with retroversion over 27 degrees or humeral head subluxation over 80% (Mizuno et al. 2013; Walch, Moraga, et al. 2012). Though these factors were taken into consideration, each surgeon independently used his own discretion. Each surgery was planned by a high-volume surgeon with similar experience and expertise in this particular field. Though this algorithm had little influence on the choice of high-volume surgeons, it might have a greater influence on those surgeons who perform a lower



[Click here to learn more about ConMed](#)

number of shoulder arthroplasties, and who could therefore greatly benefit from this guidance.

ML within orthopedic surgery is in its infancy in terms of development and refinement, and ML-based models are not widely available for many orthopedic surgeons as of yet. However, available data are increasing, and it is likely that ML will be integrated into the clinical workflow in the near future. Therefore, despite each limitation outlined above, these data are useful so that we can determine which clinical applications are appropriate for ML, identify areas of improvement and focus future research efforts.

## CONCLUSION

The ML software studied here demonstrated that it could reliably predict with 78% accuracy whether an aTSA or an rTSA should be used in select cases, compared to a standard of the majority vote of six different shoulder arthroplasty specialists. However, these surgeons with expertise in the field were not significantly influenced by the software's prediction. Nonetheless, this demonstrates that machine learning is valuable in this application, and can help to guide surgeons who might perform shoulder arthroplasty infrequently, particularly as there are no clear guidelines on arthroplasty type in many cases.

This article will be in press until the end of 2023.

Submitted: March 29, 2023 EDT, Accepted: July 11, 2023 EDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-ND-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc-nd/4.0> and legal code at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode> for more information.



## REFERENCES

- Arvind, Varun, Daniel A. London, Carl Cirino, Aakash Keswani, and Paul J. Cagle. 2021. "Comparison of Machine Learning Techniques to Predict Unplanned Readmission Following Total Shoulder Arthroplasty." *Journal of Shoulder and Elbow Surgery* 30 (2): e50–59. <https://doi.org/10.1016/j.jse.2020.05.013>.
- Atoun, Ehud, Ron Gilat, Alexander van Tongel, Riten Pradhan, Ornit Cohen, Ehud Rath, and Ofer Levy. 2017. "Intraobserver and Interobserver Reliability of the Copeland-Levy Classification for Arthroscopic Evaluation of Subacromial Impingement." *Journal of Shoulder and Elbow Surgery* 26 (12): 2167–72. <https://doi.org/10.1016/j.jse.2017.07.018>.
- Biron, Dustin R., Ishan Sinha, Justin E. Kleiner, Dilum P. Aluthge, Avi D. Goodman, I. Neil Sarkar, Eric Cohen, and Alan H. Daniels. 2019. "A Novel Machine Learning Model Developed to Assist in Patient Selection for Outpatient Total Shoulder Arthroplasty." *Journal of the American Academy of Orthopaedic Surgeons* 28 (13): e580–85. <https://doi.org/10.5435/jaaos-d-19-00395>.
- Boileau, Pascal, Damien Cheval, Marc-Olivier Gauci, Nicolas Holzer, Jean Chaoui, and Gilles Walch. 2018. "Automated Three-Dimensional Measurement of Glenoid Version and Inclination in Arthritic Shoulders." *Journal of Bone and Joint Surgery* 100 (1): 57–65. <https://doi.org/10.2106/jbjs.16.01122>.
- Cox, Ryan M., Eric M. Padegimas, Joseph A. Abboud, Charles L. Getz, Mark D. Lazarus, Matthew L. Ramsey, Gerald R. Williams Jr, and John G. Horneff III. 2018. "Outcomes of an Anatomic Total Shoulder Arthroplasty with a Contralateral Reverse Total Shoulder Arthroplasty." *Journal of Shoulder and Elbow Surgery* 27 (6): 998–1003. <https://doi.org/10.1016/j.jse.2017.12.005>.
- Fontana, Mark Alan, Stephen Lyman, Gourab K. Sarker, Douglas E. Padgett, and Catherine H. MacLean. 2019. "Can Machine Learning Algorithms Predict Which Patients Will Achieve Minimally Clinically Important Differences From Total Joint Arthroplasty?" *Clinical Orthopaedics & Related Research* 477 (6): 1267–79. <https://doi.org/10.1097/corr.0000000000000687>.
- Goutallier, Daniel, Jean-Marie Postel, Jacques Bernageau, Laurent Lavau, and Marie-Catherine Voisin. 1994. "Fatty Muscle Degeneration in Cuff Ruptures." *Clinical Orthopaedics and Related Research* 304 (304): 78–83. <https://doi.org/10.1097/00003086-199407000-00014>.
- Gowd, Anirudh K., Avinash Agarwalla, Nirav H. Amin, Anthony A. Romeo, Gregory P. Nicholson, Nikhil N. Verma, and Joseph N. Liu. 2019. "Construct Validation of Machine Learning in the Prediction of Short-Term Postoperative Complications Following Total Shoulder Arthroplasty." *Journal of Shoulder and Elbow Surgery* 28 (12): e410–21. <https://doi.org/10.1016/j.jse.2019.05.017>.
- Hendel, Michael D., Jason A. Bryan, Wael K. Barsoum, Eric J. Rodriguez, John J. Brems, Peter J. Evans, and Joseph P. Iannotti. 2012. "Comparison of Patient-Specific Instruments with Standard Surgical Instruments in Determining Glenoid Component Position: A Randomized Prospective Clinical Trial." *Journal of Bone and Joint Surgery* 94 (23): 2167–75. <https://doi.org/10.2106/jbjs.k.01209>.
- Heuer, Hinrich J.D., Robert E. Boykin, Charles J. Petit, Juliane Hardt, and Peter J. Millett. 2014. "Decision-Making in the Treatment of Diaphyseal Clavicle Fractures: Is There Agreement among Surgeons? Results of a Survey on Surgeons' Treatment Preferences." *Journal of Shoulder and Elbow Surgery* 23 (2): e23–33. <https://doi.org/10.1016/j.jse.2013.04.016>.
- Huber, Manuel, Christoph Kurz, and Reiner Leidl. 2019. "Predicting Patient-Reported Outcomes Following Hip and Knee Replacement Surgery Using Supervised Machine Learning." *BMC Medical Informatics and Decision Making* 19 (1): 3. <https://doi.org/10.1186/s12911-018-0731-6>.
- Iannotti, Joseph P., Justin Baker, Eric Rodriguez, John Brems, Eric Ricchetti, Mena Mesiha, and Jason Bryan. 2014. "Three-Dimensional Preoperative Planning Software and a Novel Information Transfer Technology Improve Glenoid Component Positioning." *Journal of Bone and Joint Surgery* 96 (9): e71. <https://doi.org/10.2106/jbjs.l.01346>.
- Iannotti, Joseph P., Scott Weiner, Eric Rodriguez, Naveen Subhas, Thomas E. Patterson, Bong Jae Jun, and Eric T. Ricchetti. 2015. "Three-Dimensional Imaging and Templating Improve Glenoid Implant Positioning." *Journal of Bone and Joint Surgery* 97 (8): 651–58. <https://doi.org/10.2106/jbjs.n.00493>.
- Kandemir, U., R. B. Allaire, J. T. Jolly, R. E. Debski, and P. J. McMahon. 2006. "The Relationship between the Orientation of the Glenoid and Tears of the Rotator Cuff." *The Journal of Bone and Joint Surgery. British Volume* 88-B (8): 1105–9. <https://doi.org/10.1302/0301-620x.88b8.17732>.
- Kunze, Kyle N., Evan M. Polce, Ian Clapp, Benedict U. Nwachukwu, Jorge Chahla, and Shane J. Nho. 2021. "Machine Learning Algorithms Predict Functional Improvement After Hip Arthroscopy for Femoroacetabular Impingement Syndrome in Athletes." *Journal of Bone and Joint Surgery* 103 (12): 1055–62. <https://doi.org/10.2106/jbjs.20.01640>.
- Kunze, Kyle N., Evan M. Polce, Benedict U. Nwachukwu, Jorge Chahla, and Shane J. Nho. 2021. "Development and Internal Validation of Supervised Machine Learning Algorithms for Predicting Clinically Significant Functional Improvement in a Mixed Population of Primary Hip Arthroscopy." *Arthroscopy* 37 (5): 1488–97. <https://doi.org/10.1016/j.arthro.2021.01.005>.

- Kunze, Kyle N., Evan M. Polce, Arpan Patel, P. Maxwell Courtney, and Brett R. Levine. 2021. "Validation and Performance of a Machine-Learning Derived Prediction Guide for Total Knee Arthroplasty Component Sizing." *Archives of Orthopaedic and Trauma Surgery* 141 (12): 2235–44. <https://doi.org/10.1007/s00402-021-04041-5>.
- Kunze, Kyle N., Evan M. Polce, Alexander J. Sadauskas, and Brett R. Levine. 2020. "Development of Machine Learning Algorithms to Predict Patient Dissatisfaction After Primary Total Knee Arthroplasty." *The Journal of Arthroplasty* 35 (11): 3117–22. <https://doi.org/10.1016/j.arth.2020.05.061>.
- Levy, Jonathan C., Nathan G. Everding, Mark A. Frankle, and Louis J. Keppeler. 2014. "Accuracy of Patient-Specific Guided Glenoid Baseplate Positioning for Reverse Shoulder Arthroplasty." *Journal of Shoulder and Elbow Surgery* 23 (10): 1563–67. <https://doi.org/10.1016/j.jse.2014.01.051>.
- Mizuno, Naoko, Patrick J Denard, Patric Raiss, and Gilles Walch. 2013. "Reverse Total Shoulder Arthroplasty for Primary Glenohumeral Osteoarthritis in Patients with a Biconcave Glenoid." *The Journal of Bone and Joint Surgery-American Volume* 95 (14): 1297–1304. <https://doi.org/10.2106/jbjs.l.00820>.
- Parsons, Moby, Alex Greene, Sandrine Polakovic, Ian Byram, Emilie Cheung, Richard Jones, Rick Papandrea, et al. 2020. "Assessment of Surgeon Variability in Preoperative Planning of Reverse Total Shoulder Arthroplasty: A Quantitative Comparison of 49 Cases Planned by 9 Surgeons." *Journal of Shoulder and Elbow Surgery* 29 (10): 2080–88. <https://doi.org/10.1016/j.jse.2020.02.023>.
- Polce, Evan M., Kyle N. Kunze, Michael C. Fu, Grant E. Garrigues, Brian Forsythe, Gregory P. Nicholson, Brian J. Cole, and Nikhil N. Verma. 2021. "Development of Supervised Machine Learning Algorithms for Prediction of Satisfaction at 2 Years Following Total Shoulder Arthroplasty." *Journal of Shoulder and Elbow Surgery* 30 (6): e290–99. <https://doi.org/10.1016/j.jse.2020.09.007>.
- Reddy, Sandeep, John Fox, and Maulik P Purohit. 2018. "Artificial Intelligence-Enabled Healthcare Delivery." *Journal of the Royal Society of Medicine* 112 (1): 22–28. <https://doi.org/10.1177/0141076818815510>.
- Ricchetti, Eric T., Michael S. Khazzam, Patrick J. Denard, David M. Dines, T. Bradley Edwards, Vahid Entezari, Richard J. Friedman, et al. 2021. "Reliability of the Modified Walch Classification for Advanced Glenohumeral Osteoarthritis Using 3-Dimensional Computed Tomography Analysis: A Study of the ASES B2 Glenoid Multicenter Research Group." *Journal of Shoulder and Elbow Surgery* 30 (4): 736–46. <https://doi.org/10.1016/j.jse.2020.07.013>.
- Schoch, Bradley S., Joseph J King, Joseph Zuckerman, Thomas W Wright, Chris Roche, and Pierre-Henri Flurin. 2020. "Anatomic versus Reverse Shoulder Arthroplasty: A Mid-Term Follow-up Comparison." *Shoulder & Elbow* 13 (5): 518–26. <https://doi.org/10.1177/1758573220921150>.
- Shukla, Dave R., Richard J. McLaughlin, Julia Lee, Robert H. Cofield, John W. Sperling, and Joaquin Sánchez-Sotelo. 2019. "Intraobserver and Interobserver Reliability of the Modified Walch Classification Using Radiographs and Computed Tomography." *Journal of Shoulder and Elbow Surgery* 28 (4): 625–30. <https://doi.org/10.1016/j.jse.2018.09.021>.
- Shukla, Dave R., Richard J McLaughlin, Julia Lee, Ngoc Tram V Nguyen, and Joaquin Sanchez-Sotelo. 2019. "Automated Three-Dimensional Measurements of Version, Inclination, and Subluxation." *Shoulder & Elbow* 12 (1): 31–37. <https://doi.org/10.1177/1758573218825480>.
- Throckmorton, Thomas W., Lawrence V. Gulotta, Frank O. Bonnarens, Stephen A. Wright, Jeffrey L. Hartzell, William B. Rozzi, Jason M. Hurst, Simon P. Frostick, and John W. Sperling. 2015. "Patient-Specific Targeting Guides Compared with Traditional Instrumentation for Glenoid Component Placement in Shoulder Arthroplasty: A Multi-Surgeon Study in 70 Arthritic Cadaver Specimens." *Journal of Shoulder and Elbow Surgery* 24 (6): 965–71. <https://doi.org/10.1016/j.jse.2014.10.013>.
- Venne, Gabriel, Brian J. Rasquinha, David Pichora, Randy E. Ellis, and Ryan Bicknell. 2015. "Comparing Conventional and Computer-Assisted Surgery Baseplate and Screw Placement in Reverse Shoulder Arthroplasty." *Journal of Shoulder and Elbow Surgery* 24 (7): 1112–19. <https://doi.org/10.1016/j.jse.2014.10.012>.
- Walch, Gilles, Claudio Moraga, Allan Young, and Juan Castellanos-Rosas. 2012. "Results of Anatomic Nonconstrained Prosthesis in Primary Osteoarthritis with Biconcave Glenoid." *Journal of Shoulder and Elbow Surgery* 21 (11): 1526–33. <https://doi.org/10.1016/j.jse.2011.11.030>.
- Walch, Gilles, Peter S. Vezzeridis, Pascal Boileau, Pierrick Deransart, and Jean Chaoui. 2015. "Three-Dimensional Planning and Use of Patient-Specific Guides Improve Glenoid Component Position: An in Vitro Study." *Journal of Shoulder and Elbow Surgery* 24 (2): 302–9. <https://doi.org/10.1016/j.jse.2014.05.029>.
- Walch, Gilles, Allan A Young, Pascal Boileau, Markus Loew, Dominique Gazielly, and Daniel Molé. 2012. "Patterns of Loosening of Polyethylene Keeled Glenoid Components after Shoulder Arthroplasty for Primary Osteoarthritis: Results of a Multicenter Study with More than Five Years of Follow-Up." *The Journal of Bone and Joint Surgery-American Volume* 94 (2): 145–50. <https://doi.org/10.2106/jbjs.j.00699>.
- Young, Allan A., Gilles Walch, Guido Pape, Frank Gohlke, and Luc Favard. 2012. "Secondary Rotator Cuff Dysfunction Following Total Shoulder Arthroplasty for Primary Glenohumeral Osteoarthritis: Results of a Multicenter Study with More than Five Years of Follow-Up." *Journal of Bone and Joint Surgery* 94 (8): 685–93. <https://doi.org/10.2106/jbjs.j.00727>.