JSES Reviews, Reports, and Techniques 3 (2023) 189–200

Contents lists available at ScienceDirect

JSES Reviews, Reports, and Techniques

journal homepage: www.jsesreviewsreportstech.org

Artificial intelligence-based applications in shoulder surgery leaves much to be desired: a systematic review

Puneet Gupta, BS[a,b], Heather S. Haeberle, MD[b], Zachary R. Zimmer, MD[a], William N. Levine, MD[b], Riley J. Williams, MD[c], Prem N. Ramkumar, MD, MBA[c,d,*]

[a]*Department of Orthopaedic Surgery, George Washington University School of Medicine and Health Sciences, Washington, DC, USA*
[b]*Department of Orthopaedic Surgery, Columbia University Irving Medical Center, New York, NY, USA*
[c]*Institute for Cartilage Repair, Hospital for Special Surgery, New York, NY, USA*
[d]*Long Beach Orthopaedic Institute, Long Beach, CA, USA*

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
Shoulder surgery
Shoulder
Technology
Orthopedics

Level of Evidence: Level IV; Systematic Review

Background: Artificial intelligence (AI) aims to simulate human intelligence using automated computer algorithms. There has been a rapid increase in research applying AI to various subspecialties of orthopedic surgery, including shoulder surgery. The purpose of this review is to assess the scope and validity of current clinical AI applications in shoulder surgery literature.

Methods: A systematic literature review was conducted using PubMed for all articles published between January 1, 2010 and June 10, 2022. The search query used the terms as follows: *(artificial intelligence OR machine learning OR deep learning) AND (shoulder OR shoulder surgery OR rotator cuff)*. All studies that examined AI application models in shoulder surgery were included and evaluated for model performance and validation (internal, external, or both).

Results: A total of 45 studies were included in the final analysis. Eighteen studies involved shoulder arthroplasty, 13 rotator cuff, and 14 other areas. Studies applying AI to shoulder surgery primarily involved (1) automated imaging analysis including identifying rotator cuff tears and shoulder implants (2) risk prediction analyses including perioperative complications, functional outcomes, and patient satisfaction. Highest model performance area under the curve ranged from 0.681 (poor) to 1.00 (perfect). Only 2 studies reported external validation.

Conclusion: Applications of AI in the field of shoulder surgery are expanding rapidly and offer patient-specific risk stratification for shared decision-making and process automation for resource preservation. However, model performance is modest and external validation remains to be demonstrated, suggesting increased scientific rigor is warranted prior to deploying AI-based clinical applications.

Applications of artificial intelligence (AI) are growing in medicine and surgery. AI is a broad term used to describe the development and application of computer systems that simulate human intelligence, with machine learning (ML) considered a subset of AI. AI involves the development of algorithms that can be "trained" to classify or predict a certain output based off certain inputs. To develop these models, they must first be "trained" using a large dataset by feeding the AI model the inputs and outputs of a large dataset, from which it identifies any possible patterns and relationships using human-like intelligence. After being "trained", a new "testing set" of data is used to assess the accuracy of the model by comparing the outputs that the model predicts to the known outcomes. As such, the accuracy can be quantified and the model refined, and new data can continue to be inputted to further strengthen the model's power. Once deemed to be of optimal accuracy, the AI model can be applied to the real-world setting. Further information regarding the technical essentials of AI for surgeons has been reviewed extensively.[3,18,41,48,53,66]

AI is integrated in medicine for various purposes, including clinical decision support systems, personalizing treatments, disease diagnostics, drug development, and health monitoring.[1] In surgery, this is particularly important for improving patient outcomes and reducing both health system and patient costs. Such significant utility has inspired many research studies applying AI to orthopedic surgery. For example, several studies have developed AI models that are able to predict postoperative outcomes, length of stay (LOS), and costs for various procedures including total hip

Institutional review board approval was not required for this review article.

*Corresponding author: Prem N. Ramkumar, MD, MBA, Hospital for Special Surgery, 535 E 70th St, New York, NY 10021, USA.
E-mail address: premramkumar@gmail.com (P.N. Ramkumar).

https://doi.org/10.1016/j.xrrt.2022.12.006
2666-6391/

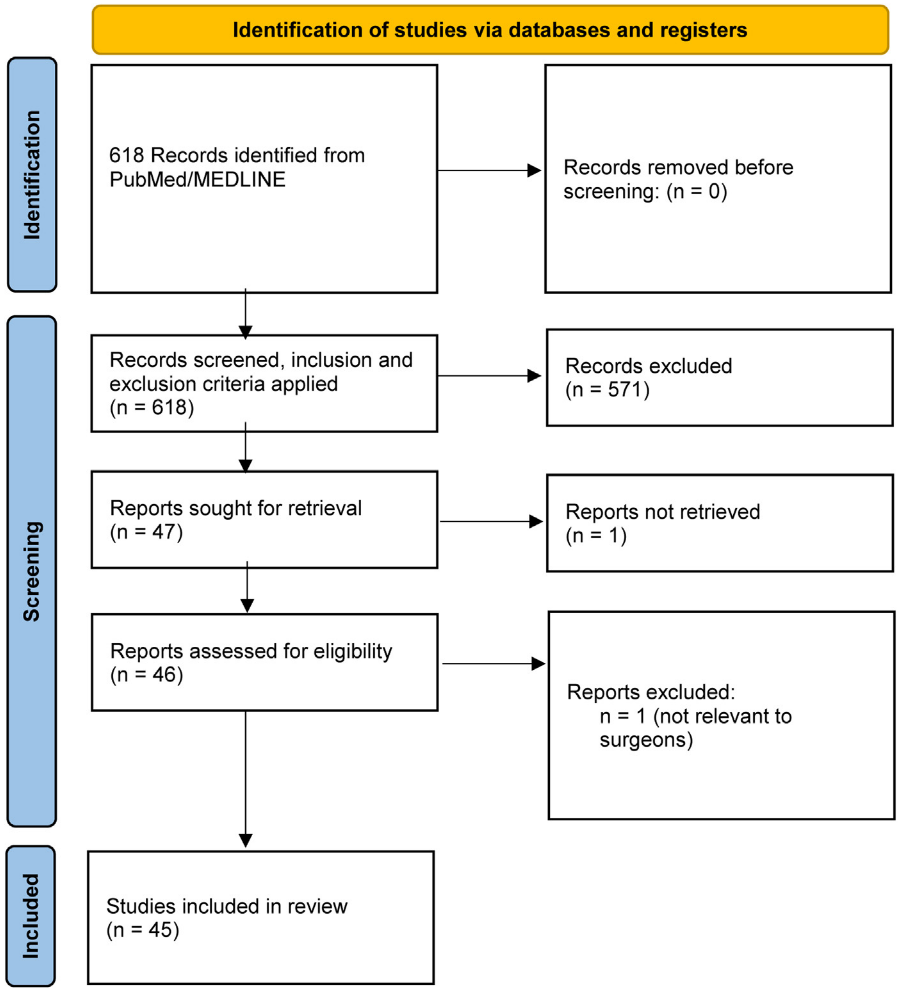

Figure 1 PRISMA diagram

arthroplasty, total knee arthroplasty, and lumbar spinal fusion.[14,20,25,27,52,57] Likewise, several studies have developed AI models capable of successfully identifying and differentiating between implants used in hip and knee arthroplasty using plain radiographs.[23,26,47,68] Moreover, AI has also been shown to be able to predict patient satisfaction or dissatisfaction following lower extremity arthroplasty.[12,16,32,35,36]

Integration of AI in the realm of shoulder surgery is already underway as well. For example, Ramkumar et al evaluated an iPhone application that uses a ML software development kit for measuring a patient's shoulder range of motion (ROM) in 4 arcs, which was clinically applied to successfully learn and analyze complex spatial motions of the shoulder joint, as indicated by no significant difference in the ROM angle measurements between the software development kit and manual goniometer.[51] Similarly, Predict+ (Exactech, Gainesville, FL, USA), a clinical decision support tool was recently released, which uses AI to predict patient-specific outcomes following shoulder arthroplasty.[11] Likewise, Food and Drug Administration-approval was recently obtained for PreView Shoulder (Genesis Software Innovations, Grand Rapids, MI, USA), an AI-based software that creates 3D models specific to a patient's shoulder anatomy to assist in preoperative surgical planning.[21]

Due to the rapid advancements being made in applying and integrating AI in surgery, it is essential for all orthopedic surgeons to be aware of the technologies being adopted and readiness for broad incorporation. In particular, there have been many advancements in understanding how AI can be applied to shoulder surgery. However, a thorough review of the current advancements made in this area has not yet been performed. Thus, the purpose of this review is to assess the scope and validity of current AI applications in the shoulder surgery literature.

Materials and methods

Search strategy

We performed a systematic literature review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. One reviewer independently completed structured searches using the PubMed database on June 10, 2022 to search for all available articles published between January 1, 2010 and June 10, 2022. The search query used the terms as follows: *(artificial intelligence OR machine learning OR deep learning) AND (shoulder OR shoulder surgery OR rotator cuff)*.

Two reviewers independently screened all titles, abstracts, and full-text articles. The reference lists of the final articles were also reviewed and cross-referenced to identify any other additional pertinent studies that were not found from the keyword search. The search strategy used in this study is displayed in Figure 1.

Eligibility criteria

Standardized inclusion and exclusion criteria were used to determine study eligibility. Any disagreements or discrepancies were resolved by consensus. Inclusion criteria were as follows: (1) involve shoulder surgery; (2) involve AI or ML; (3) clinically or operatively relevant to orthopedic surgeons; (4) published in English; (5) published from January 1, 1990 to June 10, 2022; (6) original studies with level I-IV evidence; and (7) provide extractable outcome data. Exclusion criteria were as follows: (1) no original, extractable clinical data (ie, review articles, commentaries, letters to the editor) and (2) no full-text available. Articles were grouped in the categories of "shoulder arthroplasty", "rotator cuff", or "other". Studies not pertaining to shoulder arthroplasty or the rotator cuff were grouped together as "other".

Data items

For any predictive modeling study (clinical or imaging), performance metrics were evaluated. For studies in which the dependent variable was categorical, area under the curve (AUC) of a receiver operating characteristic curve was used as the primary metric for evaluating model performance. Accuracy was recorded if AUC was not available. When both AUC and accuracies were reported, the AUC was recorded. For studies in which the dependent variable was continuous, R^2 (coefficient of determination) was recorded in order to allow for comparisons of model performance between studies. For imaging studies, if AUC or accuracy were not available, then the highest Dice (F1) scores obtained were recorded.

A receiver operating characteristic curve is a plot of a model's sensitivity (true positive) on the y-axis and its 1-specificity (false positive) on the x-axis, thereby allowing assessment of a model's performance. A perfect model would have 100% sensitivity and 100% specificity and thus an AUC of 1.0. A model with an AUC of 1.0 was considered a perfect discriminator, 0.90-0.99 was considered excellent, 0.80-0.89 was good, 0.70-0.79 was fair, and 0.51-0.69 was considered poor.[38] Accuracy is determined by the sum of the number of true positives and true negatives divided by the number of total cases. Dice similarity coefficient (Dice score, F1 score) is a broad statistical measure used to assess overlap between two different datasets. In the context of AI, it is frequently utilized for evaluating the accuracy and reproducibility of image segmentation by models by comparing the overlap with the ground-truth.[69] Dice scores range from 0 to 1, with a value of 1 indicating a perfect match.

For clinical classification studies, model performances against logistic regression were evaluated by comparing AUC or accuracy. For clinical studies with continuous variables, performance against linear regression or other traditional statistics models was compared by evaluating multiple metrics including root mean squared error, calibration slope, and R^2.

The number of types of models studied were also recorded, including regression models. If multiple models of the same type were studied, then the count total was recorded as 1. For example, if three different deep convolutional neural networks were studied such as Xception, ResNet50, and InceptionV3, the total number of types was recorded as 1. External validation was defined as comparing the performance of the algorithm when applied to an external cohort, such as that from a different institution or national database. Studies in which data from a single population was split into training, validation, and independent test sets were not considered to have externally validated their models. Additionally, temporal validation, or using data from the same population acquired during a different time period, was not considered as external validation. For example, if a single dataset has data from the years 2015-2020 but training and validation is used with 2015-2018 data while testing is used with 2019-2020 data, then it is considered temporal validation and was not counted as external validation in this systematic review.

Results

A total of 618 studies were identified from the primary search. No additional articles were identified after cross-referencing and reviewing the reference lists. A total of 45 articles were included in the final analysis.

Shoulder arthroplasty

A total of 18 studies applied AI to shoulder arthroplasty, including 178,237 patients or images (Table I). AI was applied in 12 of the 18 studies for risk prediction (7 for postoperative complications, 3 for postoperative functional outcomes, and 1 for patient satisfaction), 3 of the 18 studies for imaging (implant identification), and 3 of the studies for other purposes. Among the 18 studies, 15 studies were about clinical or imaging prediction/classification, of which 14 of the studies reported performance in AUC or accuracy. AUCs ranged from 0.689 to 1.00. One study reported only the accuracy, which was 85.92%. In an analysis of studies comparing more than 1 model for predicting outcomes (9 studies), the model that performed the best most frequently was XGBoost (5 out of 9 studies; 55.6%). In an analysis of imaging studies only for implant identification, only deep learning neural networks performed best (3 of 3 studies). Of the 12 clinical prediction studies, 7 studies compared their model performance to traditional statistical models (Table II). Seven out of the 7 studies (100%) reported an AUC higher than that of traditional models. Zero of the 16 eligible studies were externally validated.

Predicting surgical and medical complications

Five of the studies used AI to predict surgical and medical complications, including readmissions and nonhome discharge destination, following shoulder arthroplasty. Gowd et al used the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) database to analyze the potential utility of AI in predicting complications such as surgical site infections, return to operation room, and deep vein thrombosis/pulmonary embolism following anatomic total shoulder arthroplasty (aTSA) or reverse total shoulder arthroplasty (rTSA).[13] Similarly, Biron et al used 4500 patients from the ACS-NSQIP database that underwent elective TSA to develop ML models for predicting an extended LOS, defined as 3 or more days.[4] By using short LOS as a proxy for outpatient TSA, the authors believe AI could be used in the future to assist in identifying optimal candidates for outpatient TSA.[4] Likewise, Arvind et al also used 9043 primary TSA patients from the ACS-NSQIP database to evaluate five different AI algorithms for predicting readmission within 30 days postoperatively.[2] Additionally, Lopez et al also selected 21,544 elective TSA patients from the ACS-NSQIP database to develop and test two AI models for predicting nonhome discharge and 1 or more postoperative complications.[37] Furthermore, Karnuta et al used 111,147 patients who underwent aTSA and rTSA from the National Inpatient Sample database and evaluated artificial neural networks for predicting LOS, discharge disposition, and inpatient costs.[24]

Predicting functional outcomes

Three studies have applied ML to predict functional clinical outcomes following shoulder arthroplasty.[30,31,42] Kumar et al used 4782 primary aTSA or rTSA patients from a multicenter database to build and evaluate the performance of three ML models in predicting functional outcomes, such as American Shoulder and Elbow

Table I
Artificial intelligence-based studies in shoulder arthroplasty included for analysis.

Author	Year	Title	Subject area	Subtopic	Number of subjects	Median or average age (yr)	% Males	Number of types of models evaluated	External validation?
Devana SK	2022	Development of a machine learning algorithm for prediction of complications and unplanned readmission following primary anatomic total shoulder replacements	Shoulder Arthroplasty	Clinical Predictions	10,302 patients	71	45.88	5	No
Kumar V	2022	Using machine learning to predict internal rotation after anatomic and reverse total shoulder arthroplasty	Shoulder Arthroplasty	Clinical Predictions	6468 patients	70.4	38.8	3	No
Arvind V	2021	Comparison of machine learning techniques to predict unplanned readmission following total shoulder arthroplasty	Shoulder Arthroplasty	Clinical Predictions	9043 patients	69.4	43.6	5	No
Devana SK	2021	Development of a machine learning algorithm for prediction of complications and unplanned readmission following reverse total shoulder arthroplasty	Shoulder Arthroplasty	Clinical Predictions	2799 patients	69	51	5	No
Kumar V	2021	Using machine learning to predict clinical outcomes after shoulder arthroplasty with a minimal feature set	Shoulder Arthroplasty	Clinical Predictions	5774 patients	70.1	39.3	1	No
Lopez CD	2021	Using machine learning methods to predict nonhome discharge after elective total shoulder arthroplasty	Shoulder Arthroplasty	Clinical Predictions	21,544 patients	69.1	44.7	2	No
McLendon PB	2021	Machine learning can predict level of improvement in shoulder arthroplasty	Shoulder Arthroplasty	Clinical Predictions	472 patients	68	56	3	No
Polce EM	2021	Development of supervised machine learning algorithms for prediction of satisfaction at 2 years following total shoulder arthroplasty	Shoulder Arthroplasty	Clinical Predictions	413 patients	66	58.6	5	No
Kumar V	2020	What is the accuracy of three different machine learning techniques to predict clinical outcomes after shoulder arthroplasty?	Shoulder Arthroplasty	Clinical Predictions	4782 patients	70	39.9	3	No
Karnuta JM	2020	The value of artificial neural networks for predicting length of stay, discharge disposition, and inpatient costs after anatomic and reverse shoulder arthroplasty	Shoulder Arthroplasty	Clinical Predictions	90,792 patients	69	40.8	1	No
Biron DR	2020	A novel machine learning model developed to assist in patient selection for outpatient total shoulder arthroplasty	Shoulder Arthroplasty	Clinical Predictions	3128 patients	69.4	44.9	1	No
Gowd AK	2019	Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty	Shoulder Arthroplasty	Clinical Predictions	17,119 patients	69.5	56.2	6	No
Sultan H	2021	Artificial intelligence-based recognition of different types of shoulder implants in X-ray scans based on dense residual ensemble-network for personalized medicine	Shoulder Arthroplasty	Imaging	597 images	N/A	N/A	1	No
Yi PH	2020	Automated detection and classification of shoulder arthroplasty models using deep learning	Shoulder Arthroplasty	Imaging	482 images	N/A	N/A	1	No
Urban G	2020	Classifying shoulder implants in X-ray images using deep learning	Shoulder Arthroplasty	Imaging	597 images	N/A	N/A	5	No
Roche C	2021		Shoulder Arthroplasty	Other	3667 patients	N/A	N/A	N/A	N/A

Table I (*continued*)

Author	Year	Title	Subject area	Subtopic	Number of subjects	Median or average age (yr)	% Males	Number of types of models evaluated	External validation?
		Validation of a machine learning-derived clinical metric to quantify outcomes after total shoulder arthroplasty							
Menendez ME	2019	Negative patient-experience comments after total shoulder arthroplasty	Shoulder Arthroplasty	Other	186 patients	69.6	32	N/A	N/A
Tschannen M	2016	Regression forest-based automatic estimation of the articular margin plane for shoulder prosthesis planning	Shoulder Arthroplasty	Other	72 images	N/A	N/A	1	No

Table II
Performance of deep learning models in shoulder arthroplasty reports analyzed.

Author	Year	Subject area	Best/highest performing model (or model studied)	Best AUC or R^2	AI outperform traditional statistics?
Devana SK	2022	Shoulder Arthroplasty	XGBoost	AUC 0.689	Yes
Kumar V	2022	Shoulder Arthroplasty	XGBoost	AUC 0.86	Yes
Arvind V	2021	Shoulder Arthroplasty	Random Forest	AUC 0.74	Yes
Devana SK	2021	Shoulder Arthroplasty	XGBoost	AUC 0.681	Yes
Kumar V	2021	Shoulder Arthroplasty	XGBoost	AUC 0.98	Did not compare
Lopez CD	2021	Shoulder Arthroplasty	Artificial Neural Network	AUC 0.851	Did not compare
McLendon PB	2021	Shoulder Arthroplasty	Unspecified	Did not report AUC	Did not compare
Polce EM	2021	Shoulder Arthroplasty	Support Vector Machine	AUC 0.8	Yes
Kumar V	2020	Shoulder Arthroplasty	XGBoost	AUC 0.97	Yes
Karnuta JM	2020	Shoulder Arthroplasty	Artificial Neural Network	AUC 0.89	Did not compare
Biron DR	2020	Shoulder Arthroplasty	Random Forest	AUC 0.77	Did not compare
Gowd AK	2019	Shoulder Arthroplasty	XGBoost	AUC 0.77	Yes
Lu Y	2022	Rotator Cuff	Ensemble	R2 0.53	Yes
Vassalou EE	2022	Rotator Cuff	XGBoost	AUC 0.692	No
Lu Y	2021	Other	Gradient-boosted ensemble	AUC 0.86	Yes
Bullock GS	2022	Other	Linear Regression	Only reported regression R^2 of 0.41	No
Nicholson KF	2021	Other	Gradient boosting	Did not report R^2	Yes

AUC, area under the curve; R^2, coefficient of determination.

Surgeons (ASES) scores, global shoulder function scores, active abduction, and external rotation at multiple time points postoperatively.[30] Moreover, Kumar et al built upon this study using 5774 aTSA and rTSA patients from a multicenter database by evaluating if an ML model with a minimal feature set of only 19 input parameters is as accurate as a full feature set with hundreds of inputs in predicting functional clinical outcomes postoperatively.[31] Additionally, McLendon et al evaluated the performance of three AI models in predicting ASES scores at a minimum of 2 years after shoulder arthroplasty for 472 patients with primary glenohumeral osteoarthritis.[42]

Predicting patient satisfaction

Polce et al used 413 eligible patients who underwent primary aTSA or rTSA at a single tertiary referral center to build and evaluate the performance of five different AI models in predicting patient satisfaction at a minimum of two years postoperatively.[50] Additionally, the authors found that the five most important factors for predicting higher patient satisfaction with the support vector machine model were higher baseline Single Assessment Numeric Evaluation score, greater baseline exercise and activity, non-workers' compensation insurance status, diagnosis of primary glenohumeral arthritis, and preoperative duration of symptoms exceeding two years.[50]

Implant identification

Three studies applied AI in shoulder arthroplasty to identify implant types and models.[60,67] Yi et al used 482 radiographic images from publicly available repositories to test deep convolutional neural networks in detecting and classifying five different shoulder arthroplasty implants and differentiating between aTSA and rTSA.[67] Similarly, Sultan et al used 597 publicly available plain radiographs to evaluate the performance of an AI model in classifying shoulder implants into one of the four manufacturers.[60] Likewise, Urban et al[70] used 597 plain radiographs to evaluate the performance of more than 10 different AI models in classifying shoulder implants of four manufacturers directly from plain radiographs.

Other

Three studies applying AI in shoulder arthroplasty in other manners were identified.[43,56,62] Tschannen et al used 72 whole body computed tomography (CT) scans from a single institution to develop and test a regression forest-based model utilizing CT images of the upper arm to estimate a relatively accurate articular margin plane, an important resection plane used for shoulder arthroplasty procedures.[62] Furthermore, previous AI literature was applied to develop a new clinical outcome measure, the Shoulder Arthroplasty Smart score, for patients undergoing shoulder arthroplasty.[56] Using prior data to identify preoperative measures highly predictive of postoperative measures, Roche et al developed the Shoulder Arthroplasty Smart score that requires only 3 subjective and 3 objective measures and compared its performance to other similar TSA scores such as the ASES score, Constant score, and the Simple Shoulder Test score.[56] Additionally, Menendez et al used a machine-learning-based natural language processing software to

Table III
Artificial intelligence-based studies in shoulder arthroscopy included for analysis.

Author	Year	Title	Subject area	Subtopic	Number of subjects	Median or average age (yr)	% Males	Number of types of models evaluated	External validation?
Lu Y	2022	Identifying modifiable and non-modifiable cost drivers of ambulatory rotator cuff repair: a machine learning analysis	Rotator Cuff	Clinical Predictions	33,976 patients	58	57	7	No
Vassalou EE	2022	Predicting long-term outcomes of ultrasound-guided percutaneous irrigation of calcific tendinopathy with the use of machine learning	Rotator Cuff	Clinical Predictions	100 patients	46	31	2	Yes
Ho TT	2022	Classification of rotator cuff tears in ultrasound images using deep learning models	Rotator Cuff	Imaging	103 patients (194 images)	59.4	34.9	1	No
Ro K	2021	Deep-learning framework and computer assisted fatty infiltration analysis for the supraspinatus muscle in MRI	Rotator Cuff	Imaging	240 patients	N/A	N/A	1	No
Kang Y	2021	Evaluating subscapularis tendon tears on axillary lateral radiographs using deep learning	Rotator Cuff	Imaging	3746 patietns	60.9	56.7	1	No
Lee K	2021	Imbalanced loss-integrated deep-learning-based ultrasound image analysis for diagnosis of rotator-cuff tear	Rotator Cuff	Imaging	35 patients (1400 images)	N/A	N/A	1	No
Shim E	2020	Automated rotator cuff tear classification using 3D convolutional neural network	Rotator Cuff	Imaging	2124 patients	N/A	N/A	1	No
Medina G	2021	Deep learning method for segmentation of rotator cuff muscles on MR images	Rotator Cuff	Imaging	1456 scans (>1030 patients)	56.1	N/A	1	Yes
Taghizadeh E	2021	Deep learning for the rapid automatic quantification and characterization of rotator cuff muscle degeneration from shoulder CT datasets	Rotator Cuff	Imaging	95 patients (103 scans)	70.5	34.7	1	No
Kim Y	2020	Ruling out rotator cuff tear in shoulder radiograph series using deep learning: redefining the role of conventional radiograph	Rotator Cuff	Imaging	7888 patients	59.1	38.2	1	No
Kim JY	2019	Development of an automatic muscle atrophy measuring algorithm to calculate the ratio of supraspinatus in supraspinous fossa using deep learning	Rotator Cuff	Imaging	240 patients (240 images)	N/A	N/A	1	No
Lin CC	2014	Combined image enhancement, feature extraction, and classification protocol to improve detection and diagnosis of rotator-cuff tears on MR imaging	Rotator Cuff	Imaging	48 patients	65.2	43.8	1	No
Wang TF	2021	Unsupervised machine learning-based analysis of clinical features, bone mineral density features and medical care costs of rotator cuff tears	Rotator Cuff	Other	53 patients	55.5	49.1	1	No

MRI, magnetic resonance imaging; *CT*; computed tomography; *3D*, three dimensional.

evaluate and classify patient-experience comments as positive, negative, mixed, or neutral after primary TSA and then subsequently explored whether any patient-level factors or perioperative outcomes were associated with negative comments.

Rotator cuff

More than 49,678 patients were included from the 13 studies involving the rotator cuff (Table III). AI was applied in 10 of the 13 studies for imaging (2 for ultrasound, 5 for magnetic resonance imaging (MRI), 2 for plain radiographs, and 1 for CT), 2 for clinical outcomes, and 1 for other purposes. Six studies reported AUCs, which ranged from 0.69 to 0.92. Three studies reported accuracies, which ranged from 93.52% to 99.90%. Dice (F1) scores of two studies were 0.91 and 0.994. One study reported an R^2 of 0.53. 90% (9 out of 10) of the imaging studies evaluated deep learning models including convolutional neural networks (Table IV). Both of the clinical prediction studies compared their model performance to

Table IV
Performance of deep learning models in shoulder arthroscopy reports analyzed.

Author	Year	Subject area	Sub-Topic	Best/Highest performing model (or model studied)	Highest AUC, accuracy, Dice achieved
Sultan H	2021	Shoulder Arthroplasty	Imaging	Dense-residual ensemble network	Accuracy 85.92%
Yi PH	2020	Shoulder Arthroplasty	Imaging	Deep convolutional neural network	AUC 1.00
Urban G	2020	Shoulder Arthroplasty	Imaging	Deep convolutional neural network	AUC 0.94
Ho TT	2022	Rotator Cuff	Imaging	Convolutional neural network	AUC 0.845
Ro K	2021	Rotator Cuff	Imaging	Convolutional neural network	Accuracy 99.89%
Kang Y	2021	Rotator Cuff	Imaging	Multimodal deep learning model	AUC 0.83
Lee K	2021	Rotator Cuff	Imaging	Convolutional neural network	Accuracy 93.52%
Shim E	2020	Rotator Cuff	Imaging	Convolutional neural network	AUC 0.92
Medina G	2021	Rotator Cuff	Imaging	Convolutional neural network	Dice score 0.994
Taghizadeh E	2021	Rotator Cuff	Imaging	Convolutional neural network	Dice score 0.91
Kim Y	2020	Rotator Cuff	Imaging	Deep neural network	AUC 0.91
Kim JY	2019	Rotator Cuff	Imaging	Fully convolutional neural network	Accuracy 99.90%
Lin CC	2014	Rotator Cuff	Imaging	Support Vector Machine	AUC 0.844
Mu X	2021	Other	Imaging	Convolutional neural network	Accuracy 97.9%
Wang TF	2021	Other	Imaging	Convolutional neural network	Accuracy 97%
Rodrigues TC	2021	Other	Imaging	Convolutional neural network	Dice score 0.95
Lin B-S	2020	Other	Imaging	Convolutional neural network	Accuracy 94%
Chung SW	2018	Other	Imaging	Convolutional neural network	AUC 0.996
Minelli M	2022	Other	Imaging	Convolutional neural network	Did not report
Hahn S	2022	Other	Imaging	Deep-learning based reconstruction (convolutional neural network)	Accuracy 92.6%
Grauhan NF	2022	Other	Imaging	Convolutional neural network	AUC 1.00
Jiang H	2021	Other	Imaging	Not specified	AUC 0.789

AUC, area under the curve.

traditional statistical models, in which 50% (1 out of 2) outperformed traditional statistical models. External validation was performed in 2 of the 13 studies.

Rotator cuff in ultrasound images

Two of the studies utilized AI for assessing the rotator cuff in ultrasound images. Ho et al used 194 ultrasound images from a single institution to train and test five different convolutional neural networks (CNNs) in visualizing randomized controlled trial (RCT) location and classify them directly from ultrasound images.[19] Similarly, Lee et al developed and trained a novel AI model using a dataset of 1400 ultrasound images with RCTs of different sizes (massive, large, medium, and small) for the purpose of segmenting RCTs from ultrasound images.[33]

Rotator cuff in magnetic resonance images

Five studies used AI for improving RCT diagnosis from MRI images. Medina et al[71] developed and validated two CNN models for automated segmentation of rotator cuff muscles from shoulder MRI images.[58] The first model performed well in selecting a scapular Y-view from a routine sagittal T1-weighted shoulder MRI, and the second model performed well in segmenting the subscapularis, supraspinatus, and infraspinatus/teres minor muscles on a Y-view.[58] Similarly, Kim et al developed, trained, and tested an AI model using 240 MRIs for automated detection of the supraspinatus muscle and the fossa region and calculation of the occupation ratio from MRI images.[28] Likewise, Ro et al also developed and tested an AI model using 240 shoulder MRIs for automated segmentation of the supraspinatus muscle and fossa from MRI images to assist in calculating the occupation ratio.[55] Additionally, Lin et al developed and tested a support vector machine AI model using 48 patients in detecting supraspinatus injury from shoulder MRI images that have undergone enhanced image processing.[34] Furthermore, Shim et al used 2214 MRI data to develop an AI model and evaluate its performance in diagnosing the presence or absence of an RCT, classifying the tear size, and providing 3D visualization of the tear location.[58]

Rotator cuff in plain radiographs

Kim et al developed, trained, and validated a CNN using 7888 patient cases for ruling out significant RCTs from conventional shoulder radiographs.[29] Similarly, Kang et al used 2779 axillary lateral shoulder radiographs and clinical information (age, sex, and so on) to develop and test an AI model's performance in identifying subscapularis tears directly from radiographs in patients undergoing arthroscopic surgery.[17]

Rotator cuff in CT images

Taghizadeh et al used 103 shoulder CT scans to develop and test an AI model's performance in automatic quantification and characterization of the overall level of degeneration of rotator cuff muscles, including muscle atrophy and fatty infiltration, directly from conventional shoulder CT scans. [61]

Outcomes

Lu et al used 33,976 patients from a single New York database to apply AI for predicting total costs after ambulatory arthroscopic rotator cuff repair and for identifying important contributors to total charges.[39]

Vassalou studied 100 patients who underwent ultrasound-guided percutaneous irrigation for rotator cuff calcific tendinopathy and applied AI for predicting which patients would obtain complete pain resolution at 1-year follow-up.[63]

Other

Wang et al aimed to apply AI for investigating the clinical features (eg, age and sex), bone mineral density features (eg, T-score and Z-score of lumbar), and medical care costs of RCTs.[65] The authors conducted an unsupervised ML-based analysis of 53 patients with RCTs to better elucidate any underlying relationships between the three aforementioned areas.[65] The AI algorithm divided the input dataset into four subgroups and highlighted several characteristics of each.[65] For example, one subgroup had the highest frequency of osteoporosis, infraspinatus tears, and subscapularis tendon tears, suggesting that decreased bone mineral density may be directly

Table V
Other AI-related studies that pertain to the shoulder included in analysis.

Author	Year	Title	Subject area	Subtopic	Number of subjects	Median or average age (yr)	% Males	Number of types of models evaluated	External validation?
Mu X	2021	In-depth learning of automatic segmentation of shoulder joint magnetic resonance images based on convolutional neural networks	Other	Imaging	800 images	N/A	N/A	1	No
Wang TF	2021	Convolutional neural network for automatically segmenting magnetic resonance images of the shoulder joint	Other	Imaging	800 images	N/A	N/A	1	No
Rodrigues TC	2021	Three-dimensional MRI bone models of the glenohumeral joint using deep learning: evaluation of normal anatomy and glenoid bone loss	Other	Imaging	185 patients	38.9 years	64%	1	No
Lin B-S	2020	Using deep learning in ultrasound imaging of bicipital peritendinous effusion to grade inflammation severity	Other	Imaging	3801 images	N/A	N/A	1	No
Chung SW	2018	Automated detection and classification of the proximal humerus fracture by using deep learning algorithm	Other	Imaging	1891 images	65	31.3	1	No
Minelli M	2022	Measuring the critical shoulder angle on radiographs: an accurate and repeatable deep learning model	Other	Imaging	8467 images	N/A	N/A	1	No
Hahn S	2022	Image quality and diagnostic performance of accelerated shoulder MRI with deep learning-based reconstruction	Other	Imaging	110 images/105 patients	57.6	42.90%	1	N/A
Grauhan NF	2022	Deep learning for accurately recognizing common causes of shoulder pain on radiographs	Other	Imaging	3644 radiographs (2442 patients)	N/A	N/A	1	No
Lu Y	2021	Understanding anterior shoulder instability through machine learning: new models that predict recurrence, progression to surgery, and development of arthritis	Other	Clinical Predictions	654 patients	21.7	76.5	7	No
Burns DM	2018	Shoulder physiotherapy exercise recognition: machine learning the inertial signals from a smartwatch	Other	Other	20 patients	28.9	30	5	N/A
Ramkumar PN	2018	Mobile technology and telemedicine for shoulder range of motion: validation of a motion-based machine-learning software development kit	Other	Other	10 patients	27	50%	N/A	N/A
Bullock GS	2022	Machine learning does not improve humeral torsion prediction compared to regression in baseball pitchers	Other	Clinical Predictions	407 patients	23.2	100%	5	No
Jiang H	2021	Machine learning-based ultrasomics for predicting subacromial impingement syndrome stages	Other	Imaging	324 patients	58	38.60%	1	No
Nicholson KF	2021	Machine learning and statistical prediction of pitching arm kinetics	Other	Clinical Predictions	168 patients	16.7	100%	5	No

MRI, magnetic resonance imaging.

contributing to infraspinatus and subscapularis tendon tears.[65] Consistent with those findings, another subgroup that had the lowest rate of osteoporosis had the lowest infraspinatus and subscapularis tendon tears.[65]

Other

Fourteen studies were included in this category, 9 of which were related to imaging, 3 were related to prediction models, and 2 were related to technology (Table V). Vast topics were covered including predicting humeral torsion in baseball pitchers, grading inflammation severity in bicipital peritendinous effusions, and automating detection and classification of proximal humerus fractures from images. AUCs were reported in 4 studies and accuracy was reported in 4 studies. AUCs ranged from 0.789 to 1.00, and accuracy ranged from 92.6% to 97.9%. Dice (F1) score of one study was 0.95. One study reported an R^2 of 0.41 but for only one model. Deep learning neural networks were the only type of model studied in the imaging studies. No study was externally validated.

Several studies applied AI to MRI images specifically. Mu et al used approximately 800 MRI images to test and evaluate the performance of AI models in automatically segmenting different bony regions of interest from shoulder MRIs.[46] Similarly, Wang et al also used approximately 800 MRI images to develop and test an AI model for automated detection, classification, and segmentation of bone regions of interest in shoulder MRIs.[64] Likewise, Rodrigues et al used over 100 patients to apply AI models for fully automated segmentation of the glenohumeral joint and quantification of glenoid anatomy, glenoid bone loss, and humeral anatomy using shoulder MRIs.[7] Additionally, Hahn et al used 110 3-T shoulder MRIs and aimed to improve the image quality and diagnostic performance of accelerated shoulder MRIs using AI-based reconstruction techniques and compared it to standard and accelerated sequences reconstructed conventionally.[17]

Several studies also applied AI to ultrasound imaging and plain radiographs. Lin et al applied AI for automated classification of inflammation severity (normal, mild, moderate, and severe) in bicipital peritendinous effusions from ultrasound images. Similarly, Jiang et al applied an ultrasound-based AI model for classifying stages of subacromial impingement syndrome.[22] In regards to plain radiographs, Chung et al applied AI for automated detection and classification of proximal humerus factures into four types using Neer's classification directly from plain anteroposterior shoulder radiographs.[8] Similarly, Minelli et al applied AI for automated identification of the three landmarks used to calculate the critical shoulder angle on anteroposterior radiographs of the shoulder.[44] Likewise, Grauhan et al trained and tested an AI model for detecting common causes of shoulder pain, including fractures, osteoarthritis, osteosynthesis, endoprosthesis, calcification, and dislocations, directly from plain radiographs.[15]

Three AI articles were related to nonimaging prediction models. Nicholson et al applied AI for creating prediction models for elbow valgus torque and shoulder distraction force in baseball players using biomechanical variables such as pitch velocity, stride length, and maximum shoulder external rotation.[49] Similarly, Bullock et al developed and tested several ML models for predicting humeral torsion in professional baseball players using input variables such as player demographics, injury history, and shoulder ROM (external rotation, internal rotation, and horizontal adduction).[5] Additionally, Lu et al evaluated several AI models in anterior shoulder instability patients for predicting recurrent instability, progression to surgery, and the development of symptomatic osteoarthritis.[40]

Two articles were related to technology. Burns et al built AI models for classifying shoulder physiotherapy exercises from sensor data from a smartwatch worn on the person.[6] Ramkumar et al applied AI to assist with measuring shoulder ROM in four different arcs (abduction, forward flexion, internal rotation, and external rotation) through a motion-based ML software development kit.[51]

Discussion

AI may have the potential to change the way orthopedics is practiced both in the clinic and operating room, particularly in shoulder surgery. Such cutting-edge technology will optimize value-based payment models, risk stratification of patients, and patient outcomes through patient-specific optimization, and evidence-based shared decision-making models. As highlighted previously, there has been success in applying AI to shoulder arthroplasty, rotator cuff patients, and other areas for many different applications, including predicting medical and surgical complications, predicting patient outcomes, and identifying implants. Nonetheless, there remains room for caution prior to clinical translation of AI to shoulder surgery, with the two greatest limitations being the lack of external validation, limiting generalizability, lack of well-performing models, and limiting accuracy.

Not all studies need ML for analysis, particularly in the setting of nongranular data. In the study by Gowd et al[13], for example, logistic regression represents the most rudimentary form of ML and outperformed advanced AI-based models in predicting adverse events, particularly surgical site infections. When comparing the distance traveled between an exotic sports car (ie, AI) and an economical sedan (ie, logistic regression), the advantage of the sports car in terms of aerodynamics, horsepower, and other engineering considerations means nothing when both cars are stuck in an unfavorable environment, like a highway traffic jam (ie, poor quality data and limited volume of data). Because of the lack of robust, high quality datasets, the performance of these AI-based algorithms in the shoulder arthroplasty literature often failed to demonstrate a good to excellent performance level, as 9 of the 13 evaluated studies reporting AUCs failed to exceed an AUC of 0.90 (excellent performance). It is likely that a model that is only performing at good or fair during internal validation will perform poorly when externally validated.[54,59] To improve model performance, there are several areas to consider. First, large accurate datasets from national and international regions with tens to hundreds of thousands of patients are needed. Despite the use of large databases, current sample sizes in shoulder surgery research are mostly limited within the realm of AI. Second, high-quality data with multiple input metrics, such as shoulder-specific physical examination findings, functional outcomes, patient-reported outcomes, medical complications, and value-based metrics are needed. Third, it is important to choose the optimal algorithm for each AI-based application, which should be carefully selected, based on the type of dataset and anticipated outputs. Currently, there appears to be an impetus to report studies utilizing AI without careful consideration of the quality of prediction from the predictive models, or consideration of non-AI-based models which may represent an acceptable alternative in some cases.

Beyond the lacking efficacy of many of the AI-based shoulder arthroplasty models, 0 of the 16 studies were externally validated. Likewise, only 2 of the 13 rotator cuff studies and 0 of the other studies were externally validated. Several conclusions in the studies suggest clinical application is imminent but applying these models to external populations and institutions for algorithm refinement is needed first. This finding illustrates that currently published shoulder arthroplasty studies and most rotator cuff studies using AI to date are exploratory, proof-of-concept reports. The lack of external validation is a critical limiting factor of current studies with respect to generalizability. This is an important barrier to clinical translation of AI-based technologies, as many prediction models perform worse during external validation.[54,59] For example, patient data or datasets from geographically varied sites are necessary, as there may be practice variations that result in differential models. Furthermore, future studies should externally validate any models in a prospective setting to better understand how the model performs in the current health care climate. Surgeons must be aware that current publicly available websites with risk calculators in shoulder surgery are based off studies that have only been internally validated and are not safe nor designed for immediate clinical use. In the case of shoulder arthroplasty, for example, new surgical techniques, redesigned prostheses and implants, varying patient populations, and improved medical management for patients may lead to differences in model performance.

As with other emerging technologies, there are several important limitations of AI that need to be addressed and understood. One major limitation involves the "black box" phenomenon, in which the users are able to see the inputs and final outputs of the AI algorithm but are unable to see or understand how the algorithm reached that final decision. For example, an algorithm may recommend using a certain prosthesis for rTSA, given the patient's medical comorbidities and other preoperative data, but the surgeon may not be able to determine how it reached this conclusion. This phenomenon creates both clinical applicability and ethical dilemmas, such as whether these AI-derived clinical tools should be trusted and if physicians should rely on AI-derived decisions.[10] This concern is slowly being dispelled with analyses that can determine the weights of risk factors on outcome analyses (ie, SHapley Additive exPlanations analysis) or the locations of interest in implant identification (ie, heatmapping). Another important limitation is data bias. Any form of bias in the data used for building an AI algorithm will lead to bias in the outcomes and conclusions it reaches. For example, certain algorithms may be developed using institutional data that has patients that are predominantly from a certain socioeconomic group or practice that may not be generalizable to the general population. Likewise, surgeons may have certain practice preferences that influence their care and decision-making, leading to algorithms that poorly classify or reach erroneous conclusions in other populations that cannot be generalized. Another important limitation is that algorithms are unable to take into account contextual information, which requires surgeons to critically evaluate AI-based decisions and apply them in the appropriate clinical context, recognizing when to choose differently than the model suggests. Additionally, there is a current lack of relevant, AI-specific guidelines for improving the quality and transparency of data being reported. The transparent reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines involve a 22-item checklist for improving the quality of reporting for any study focused on prediction models.[45] However, the original TRIPOD guidelines do not apply to AI research, as these guidelines were designed for traditional multivariate models such as the logistic regression and Cox regression. The development of TRIPOD-AI that is specific for prediction model studies based on AI is in progress.[9]

Conclusion

Applications of AI in the field of shoulder surgery are expanding rapidly and offer patient-specific risk stratification for shared decision-making and process automation for resource preservation. However, model performance is modest and external validation remains to be demonstrated, suggesting increased scientific rigor is warranted prior to deploying AI-based applications to the clinical setting in shoulder surgery. It is important for shoulder surgeons to be aware of AI-related advancements but to also be wary before broadly applying algorithms without model efficacy and external validation.

Disclaimers:

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
Conflicts of interest: William N. Levine: This author, or their immediate family, has received financial payments from Zimmer. Riley J. Williams: This author, or their immediate family, has received financial payments from Arthrex, BICMD, Cymedica, Engage Surgical, Gramercy Extremity Orthopedics, Histogenics, JRF Ortho, Lipogems, Pristine Surgical, and RecoverX. Prem N Ramkumar: This author, or their immediate family, has received financial payments or other benefits from ConforMIS, FocusMotion, Globus Medical, Johnson & Johnson, and Stryker. He is on the board of the American Association of Hip and Knee Surgeons and the Journal of Arthroplasty. The other authors, their immediate families, and any research foundation with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

References

1. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care 2019;8:2328-31. https://doi.org/10.4103/jfmpc.jfmpc_440_19.
2. Arvind V, London DA, Cirino C, Keswani A, Cagle PJ. Comparison of machine learning techniques to predict unplanned readmission following total shoulder arthroplasty. J Shoulder Elbow Surg 2021;30:e50-9. https://doi.org/10.1016/j.jse.2020.05.013.
3. Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? J Arthroplasty 2018;33:2358-61. https://doi.org/10.1016/j.arth.2018.02.067.
4. Biron DR, Sinha I, Kleiner JE, Aluthge DP, Goodman AD, Sarkar IN, et al. A novel machine learning model developed to assist in patient selection for outpatient total shoulder arthroplasty. J Am Acad Orthop Surg 2020;28:e580-5. https://doi.org/10.5435/JAAOS-D-19-00395.
5. Bullock GS, Thigpen CA, Collins GS, Arden NK, Noonan TK, Kissenberth MJ, et al. Machine learning does not improve humeral torsion prediction compared to regression in baseball pitchers. Int J Sports Phys Ther 2022;17:390-9. https://doi.org/10.26603/001c.32380.
6. Burns DM, Leung N, Hardisty M, Whyne CM, Henry P, McLachlin S. Shoulder physiotherapy exercise recognition: machine learning the inertial signals from a smartwatch. Physiol Meas 2018;39:75007. https://doi.org/10.1088/1361-6579/aacfd9.
7. Cantarelli Rodrigues T, Deniz CM, Alaia EF, Gorelik N, Babb JS, Dublin J, et al. Three-dimensional MRI bone models of the glenohumeral joint using deep learning: evaluation of normal anatomy and glenoid bone loss. Radiol Artif Intell 2020;2:e190116. https://doi.org/10.1148/ryai.2020190116.
8. Chung SW, Han SS, Lee JW, Oh K-S, Kim NR, Yoon JP, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop 2018;89:468-73. https://doi.org/10.1080/17453674.2018.1453714.
9. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021;11:e048008. https://doi.org/10.1136/bmjopen-2020-048008.
10. Durán JM, Jongsma KR. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. J Med Ethics 2021;47:329-35. https://doi.org/10.1136/medethics-2020-106820.
11. Exactech, Predict+. Patient-Specific Outcome Predictor. https://www.exac.com/extremities/predict-plus/; 2021. Accessed December 1, 2022.
12. Farooq H, Deckard ER, Ziemba-Davis M, Madsen A, Meneghini RM. Predictors of patient satisfaction following primary total knee arthroplasty: results from a traditional statistical model and a machine learning algorithm. J Arthroplasty 2020;35:3123-30. https://doi.org/10.1016/j.arth.2020.05.077.
13. Gowd AK, Agarwalla A, Amin NH, Romeo AA, Nicholson GP, Verma NN, et al. Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. J Shoulder Elbow Surg 2019;28:e410-21. https://doi.org/10.1016/j.jse.2019.05.017.
14. Goyal A, Ngufor C, Kerezoudis P, McCutcheon B, Storlie C, Bydon M. Can machine learning algorithms accurately predict discharge to nonhome facility and early unplanned readmissions following spinal fusion? Analysis of a national surgical registry. J Neurosurg Spine 2019:1-11. https://doi.org/10.3171/2019.3.SPINE181367.
15. Grauhan NF, Niehues SM, Gaudin RA, Keller S, Vahldiek JL, Adams LC, et al. Deep learning for accurately recognizing common causes of shoulder pain on radiographs. Skeletal Radiol 2022;51:355-62. https://doi.org/10.1007/s00256-021-03740-9.
16. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient dissatisfaction following total knee arthroplasty: a systematic review of the literature. J Arthroplasty 2017;32:3854-60. https://doi.org/10.1016/j.arth.2017.07.021.
17. Hahn S, Yi J, Lee H-J, Lee Y, Lim Y-J, Bang J-Y, et al. Image quality and diagnostic performance of accelerated shoulder MRI with deep learning-based reconstruction. AJR Am J Roentgenol 2022;218:506-16. https://doi.org/10.2214/AJR.21.26577.
18. Helm JM, Swiergosz AM, Haeberle HS, Karnuta JM, Schaffer JL, Krebs VE, et al. Machine learning and artificial intelligence: definitions, applications, and future directions. Curr Rev Musculoskelet Med 2020;13:69-76. https://doi.org/10.1007/s12178-020-09600-8.
19. Ho TT, Kim G-T, Kim T, Choi S, Park E-K. Classification of rotator cuff tears in ultrasound images using deep learning models. Med Biol Eng Comput 2022;60:1269-78. https://doi.org/10.1007/s11517-022-02502-6.
20. Hopkins BS, Yamaguchi JT, Garcia R, Kesavabhotla K, Weiss H, Hsu WK, et al. Using machine learning to predict 30-day readmissions after posterior lumbar fusion: an NSQIP study involving 23,264 patients. J Neurosurg Spine 2019:1-8. https://doi.org/10.3171/2019.9.SPINE19860.
21. Innovations GS. Shoulder Innovations Announces Exclusive License Agreement for Genesis Software Innovations PreView Shoulder^{TM} Software. 2021. https://genesissoftwareinnovations.com/shoulder-innovations-exclusive-license-agreement/#software.

22. Jiang H, Chen L, Zhao Y-J, Lin Z-Y, Yang H. Machine learning-based ultrasomics for predicting subacromial impingement syndrome stages. J Ultrasound Med 2022;41:2279-85. https://doi.org/10.1002/jum.15914.
23. Kang Y-J, Yoo J-I, Cha Y-H, Park CH, Kim J-T. Machine learning-based identification of hip arthroplasty designs. J Orthop Transl 2020;21:13-7. https://doi.org/10.1016/j.jot.2019.11.004.
24. Karnuta JM, Churchill JL, Haeberle HS, Nwachukwu BU, Taylor SA, Ricchetti ET, et al. The value of artificial neural networks for predicting length of stay, discharge disposition, and inpatient costs after anatomic and reverse shoulder arthroplasty. J Shoulder Elbow Surg 2020;29:2385-94. https://doi.org/10.1016/j.jse.2020.04.009.
25. Karnuta JM, Golubovsky JL, Haeberle HS, Rajan PV, Navarro SM, Kamath AF, et al. Can a machine learning model accurately predict patient resource utilization following lumbar spinal fusion? Spine J 2020;20:329-36. https://doi.org/10.1016/j.spinee.2019.10.007.
26. Karnuta JM, Haeberle HS, Luu BC, Roth AL, Molloy RM, Nystrom LM, et al. Artificial intelligence to identify arthroplasty implants from radiographs of the hip. J Arthroplasty 2021;36:S290-4. https://doi.org/10.1016/j.arth.2020.11.015. e1.
27. Karnuta JM, Navarro SM, Haeberle HS, Helm JM, Kamath AF, Schaffer JL, et al. Predicting inpatient payments prior to lower extremity arthroplasty using deep learning: which model architecture is best? J Arthroplasty 2019;34:2235-41. https://doi.org/10.1016/j.arth.2019.05.048. e1.
28. Kim JY, Ro K, You S, Nam BR, Yook S, Park HS, et al. Development of an automatic muscle atrophy measuring algorithm to calculate the ratio of supraspinatus in supraspinous fossa using deep learning. Comput Methods Programs Biomed 2019;182, 105063. https://doi.org/10.1016/j.cmpb.2019.105063.
29. Kim Y, Choi D, Lee KJ, Kang Y, Ahn JM, Lee E, et al. Ruling out rotator cuff tear in shoulder radiograph series using deep learning: redefining the role of conventional radiograph. Eur Radiol 2020;30:2843-52. https://doi.org/10.1007/s00330-019-06639-1.
30. 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;30:e225-36. https://doi.org/10.1016/j.jse.2020.07.042.
31. Kumar V, Roche C, Overman S, Simovitch R, Flurin P-H, Wright T, et al. What is the accuracy of three different machine learning techniques to predict clinical outcomes after shoulder arthroplasty? Clin Orthop Relat Res 2020;478:2351-63. https://doi.org/10.1097/CORR.0000000000001263.
32. Kunze KN, Polce EM, Sadauskas AJ, Levine BR. Development of machine learning algorithms to predict patient dissatisfaction after primary total knee arthroplasty. J Arthroplasty 2020;35:3117-22. https://doi.org/10.1016/j.arth.2020.05.061.
33. Lee K, Kim JY, Lee MH, Choi C-H, Hwang JY. Imbalanced loss-integrated deep-learning-based ultrasound image analysis for diagnosis of rotator-cuff tear. Sensors (Basel) 2021;21:2214. https://doi.org/10.3390/s21062214.
34. Lin C-C, Wang C-N, Ou Y-K, Fu J. Combined image enhancement, feature extraction, and classification protocol to improve detection and diagnosis of rotator-cuff tears on MR imaging. Magn Reson Med Sci 2014;13:155-66. https://doi.org/10.2463/mrms.2013-0079.
35. Liu J, Yang Y, Wan S, Yao Z, Zhang Y, Zhang Y, et al. A new prediction model for patient satisfaction after total knee arthroplasty and the roles of different scoring systems: a retrospective cohort study. J Orthop Surg Res 2021;16:329. https://doi.org/10.1186/s13018-021-02469-4.
36. Liu N, Kumara S, Reich E. Gaining insights into patient satisfaction through interpretable machine learning. IEEE J Biomed Health Inform 2021;25:2215-26. https://doi.org/10.1109/JBHI.2020.3038194.
37. Lopez CD, Constant M, Anderson MJJ, Confino JE, Heffernan JT, Jobin CM. Using machine learning methods to predict nonhome discharge after elective total shoulder arthroplasty. JSES Int 2021;5:692-8. https://doi.org/10.1016/j.jseint.2021.02.011.
38. Lopez CD, Gazgalis A, Boddapati V, Shah RP, Cooper HJ, Geller JA. Artificial learning and machine learning decision guidance applications in total hip and knee arthroplasty: a systematic review. Arthroplast Today 2021;11:103-12. https://doi.org/10.1016/j.artd.2021.07.012.
39. Lu Y, Labott JR, Salmons HI, Gross BD, Barlow JD, Sanchez-Sotelo J, et al. Identifying modifiable and non-modifiable cost drivers of ambulatory rotator cuff repair: a machine learning analysis. J Shoulder Elbow Surg 2022;31:2262-73. https://doi.org/10.1016/j.jse.2022.04.008.
40. Lu Y, Pareek A, Wilbur RR, Leland DP, Krych AJ, Camp CL. Understanding anterior shoulder instability through machine learning: new models that predict recurrence, progression to surgery, and development of arthritis. Orthop J Sports Med 2021;9:23259671211053330. https://doi.org/10.1177/23259671211053326.
41. Makhni EC, Makhni S, Ramkumar PN. Artificial intelligence for the orthopaedic surgeon: an overview of potential benefits, limitations, and clinical applications. J Am Acad Orthop Surg 2021;29:235-43. https://doi.org/10.5435/JAAOS-D-20-00846.
42. McLendon PB, Christmas KN, Simon P, Plummer OR, Hunt A, Ahmed AS, et al. Machine learning can predict level of improvement in shoulder arthroplasty. JB JS Open Access 2021;6:e20.00128. https://doi.org/10.2106/JBJS.OA.20.00128.
43. Menendez ME, Shaker J, Lawler SM, Ring D, Jawa A. Negative patient-experience comments after total shoulder arthroplasty. J Bone Joint Surg Am 2019;101:330-7. https://doi.org/10.2106/JBJS.18.00695.
44. Minelli M, Cina A, Galbusera F, Castagna A, Savevski V, Sconfienza LM. Measuring the critical shoulder angle on radiographs: an accurate and repeatable deep learning model. Skeletal Radiol 2022;51:1873-8. https://doi.org/10.1007/s00256-022-04041-5.
45. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1-73. https://doi.org/10.7326/M14-0698.
46. Mu X, Cui Y, Bian R, Long L, Zhang D, Wang H, et al. In-depth learning of automatic segmentation of shoulder joint magnetic resonance images based on convolutional neural networks. Comput Methods Programs Biomed 2021;211:106325. https://doi.org/10.1016/j.cmpb.2021.106325.
47. Murphy M, Killen C, Burnham R, Sarvari F, Wu K, Brown N. Artificial intelligence accurately identifies total hip arthroplasty implants: a tool for revision surgery. Hip Int 2021;32:766-70. https://doi.org/10.1177/1120700020987526.
48. Myers TG, Ramkumar PN, Ricciardi BF, Urish KL, Kipper J, Ketonis C. Artificial intelligence and orthopaedics: an introduction for clinicians. J Bone Joint Surg Am 2020;102:830-40. https://doi.org/10.2106/JBJS.19.01128.
49. Nicholson KF, Collins GS, Waterman BR, Bullock GS. Machine learning and statistical prediction of pitching arm kinetics. Am J Sports Med 2022;50:238-47. https://doi.org/10.1177/03635465211054506.
50. 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;30:e290-9. https://doi.org/10.1016/j.jse.2020.09.007.
51. Ramkumar PN, Haeberle HS, Navarro SM, Sultan AA, Mont MA, Ricchetti ET, et al. Mobile technology and telemedicine for shoulder range of motion: validation of a motion-based machine-learning software development kit. J Shoulder Elbow Surg 2018;27:1198-204. https://doi.org/10.1016/j.jse.2018.01.013.
52. Ramkumar PN, Karnuta JM, Navarro SM, Haeberle HS, Scuderi GR, Mont MA, et al. Deep learning preoperatively predicts value metrics for primary total knee arthroplasty: development and validation of an artificial neural network model. J Arthroplasty 2019;34:2220-7. https://doi.org/10.1016/j.arth.2019.05.034. e1.
53. Ramkumar PN, Kunze KN, Haeberle HS, Karnuta JM, Luu BC, Nwachukwu BU, et al. Clinical and research medical applications of artificial intelligence. Arthroscopy 2021;37:1694-7. https://doi.org/10.1016/j.arthro.2020.08.009.
54. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J 2021;14:49-58. https://doi.org/10.1093/ckj/sfaa188.
55. Ro K, Kim JY, Park H, Cho BH, Kim IY, Shim SB, et al. Deep-learning framework and computer assisted fatty infiltration analysis for the supraspinatus muscle in MRI. Sci Rep 2021;11:15065. https://doi.org/10.1038/s41598-021-93026-w.
56. Roche C, Kumar V, Overman S, Simovitch R, Flurin P-H, Wright T, et al. Validation of a machine learning-derived clinical metric to quantify outcomes after total shoulder arthroplasty. J Shoulder Elbow Surg 2021;30:2211-24. https://doi.org/10.1016/j.jse.2021.01.021.
57. Shah AA, Devana SK, Lee C, Bugarin A, Lord EL, Shamie AN, et al. Prediction of major complications and readmission after lumbar spinal fusion: a machine learning-driven approach. World Neurosurg 2021;152:e227-34. https://doi.org/10.1016/j.wneu.2021.05.080.
58. Shim E, Kim JY, Yoon JP, Ki S-Y, Lho T, Kim Y, et al. Automated rotator cuff tear classification using 3D convolutional neural network. Sci Rep 2020;10:15632. https://doi.org/10.1038/s41598-020-72357-0.
59. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol 2015;68:25-34. https://doi.org/10.1016/j.jclinepi.2014.09.007.
60. Sultan H, Owais M, Park C, Mahmood T, Haider A, Park KR. Artificial intelligence-based recognition of different types of shoulder implants in X-ray scans based on dense residual ensemble-network for personalized medicine. J Pers Med 2021;11:482. https://doi.org/10.3390/jpm11060482.
61. Taghizadeh E, Truffer O, Becce F, Eminian S, Gidoin S, Terrier A, et al. Deep learning for the rapid automatic quantification and characterization of rotator cuff muscle degeneration from shoulder CT datasets. Eur Radiol 2021;31:181-90. https://doi.org/10.1007/s00330-020-07070-7.
62. Tschannen M, Vlachopoulos L, Gerber C, Székely G, Fürnstahl P. Regression forest-based automatic estimation of the articular margin plane for shoulder prosthesis planning. Med Image Anal 2016;31:88-97. https://doi.org/10.1016/j.media.2016.02.008.
63. Vassalou EE, Klontzas ME, Marias K, Karantanas AH. Predicting long-term outcomes of ultrasound-guided percutaneous irrigation of calcific tendinopathy with the use of machine learning. Skeletal Radiol 2022;51:417-22. https://doi.org/10.1007/s00256-021-03893-7.
64. Wang G, Han Y. Convolutional neural network for automatically segmenting magnetic resonance images of the shoulder joint. Comput Methods Programs Biomed 2021;200:105862. https://doi.org/10.1016/j.cmpb.2020.105862.
65. Wang T-F, Chen D-S, Zhu J-W, Zhu B, Wang Z-L, Cao J-G, et al. Unsupervised machine learning-based analysis of clinical features, bone mineral density features and medical care costs of rotator cuff tears. Risk Manag Healthc Policy 2021;14:3977-86. https://doi.org/10.2147/RMHP.S330555.
66. Wang VM, Cheung CA, Kozar AJ, Huang B. Machine learning applications in orthopaedic imaging. J Am Acad Orthop Surg 2020;28:e415-7. https://doi.org/10.5435/JAAOS-D-19-00688.
67. Yi PH, Kim TK, Wei J, Li X, Hager GD, Sair HI, et al. Automated detection and classification of shoulder arthroplasty models using deep learning. Skeletal Radiol 2020;49:1623-32. https://doi.org/10.1007/s00256-020-03463-3.

68. Yi PH, Wei J, Kim TK, Sair HI, Hui FK, Hager GD, et al. Automated detection & classification of knee arthroplasty using deep learning. Knee 2020;27:535-42. https://doi.org/10.1016/j.knee.2019.11.020.
69. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index. Acad Radiol 2004;11:178-89. https://doi.org/10.1016/S1076-6332(03)00671-8.
70. Urban G, Porhemmat S, Stark M, Feeley B, Okada K, Baldi P. Classifying shoulder implants in X-ray images using deep learning. Comput Struct Biotechnol J 2020;18:967-72.
71. Medina G, Buckless CG, Thomasson E, Oh LS, Torriani M. Deep learning method for segmentation of rotator cuff muscles on MR images. Skeletal Radiol 2021;50:683-92.