Al-Azhar International Medical Journal

Volume 5 | Issue 5 Article 40

5-31-2024

Section: Orthopedics

The Role of Artificial Intelligence in Shoulder Arthroplasty; A systematic Review and Meta-Analysis

Faisal Hasan Zayed
Orthopedic Surgery, Faculty of Medicine for Boys, Al-Azhar University, Cairo, Egypt

Samir Ahmed Nematallah
Orthopedic Surgery, Faculty of Medicine for Boys, Al-Azhar University, Cairo, Egypt

Alaa Eldin Mahmoud Mohamed Elsaed
Orthopedic Surgery, Faculty of Medicine for Boys, Al-Azhar University, Cairo, Egypt, ala425826@gmail.com

How to Cite This Article

Zayed, Faisal Hasan; Nematallah, Samir Ahmed; and Elsaed, Alaa Eldin Mahmoud Mohamed (2024) "The Role of Artificial Intelligence in Shoulder Arthroplasty; A systematic Review and Meta-Analysis," *Al-Azhar International Medical Journal*: Vol. 5: Iss. 5, Article 40.
DOI: https://doi.org/10.58675/2682-339X.2442

META-ANALYSIS

The Role of Artificial Intelligence in Shoulder Arthroplasty; A systematic Review and Meta-Analysis

Faisal H. Zayed , Samir A. Nematallah, Alaa Eldin M. M. Elsaed *

Department of Orthopedic Surgery, Faculty of Medicine for Boys, Al-Azhar University, Cairo, Egypt

Abstract

Background: ***Artificial intelligence (AI) and machine learning (ML) have significant potential for improving clinical decision-making by analyzing complex healthcare data. However, their use in predicting outcomes after shoulder arthroplasty still needs to be explored.***

Aim: ***This study aims to evaluate the impact of AI in predicting the success rates of total and reverse total shoulder arthroplasty.***

Methods: ***We conducted a comprehensive literature search in PubMed, Web of Science, and SCOPUS, focusing on randomized controlled trials and observational studies. Our analysis included various outcomes such as Area Under the Precision-Recall Curve (AUPRC) scores and Mean Absolute Errors (MAE) in predicting various shoulder function scores.***

Results: ***Our meta-analysis included data from 154,988 patients with an average age of 69.63 years. We found the average AUPRC score to be 0.839, indicating robust model performance. The MAEs for various shoulder function scores were as follows: Global Shoulder Function score showed an MAE of 1.025, indicating a high level of prediction accuracy. The VAS pain score prediction had an MAE of 1.00, demonstrating the model's efficacy in pain assessment. The ASES score prediction yielded an MAE of 11.61, while active forward elevation had an MAE of 17.663. Active external rotation was associated with an MAE of 12.771, and the constant score prediction showed an MAE of 9.095.***

Conclusion: ***AI has the potential to revolutionize the field of shoulder arthroplasty, enhancing surgical decision-making and patient outcomes. Despite challenges, AI offers promising avenues for improved orthopedic care.***

Keywords: Artificial intelligence; Machine Learning; Total Shoulder Arthroplasty; Reverse Total Shoulder Arthroplasty

1. Introduction

Glenohumeral arthritis can be successfully treated with total shoulder arthroplasty (TSA), a routine orthopedic operation. TSA is typically connected to an inpatient hospital stay.[1]

Outpatient TSA has drawn more interest as a way to cut healthcare expenditures without sacrificing quality. Due to costs associated with the procedure itself, medications, rehabilitation, and nursing, inpatient TSA can cost three times as much as outpatient TSA.[2]

(AI) represents a diverse discipline focused on activities and automation that now require human intelligence. It has emerged as an innovative technology, transforming many aspects of life, although it exhibits limited public awareness.[3]

Supervised (ML) represents an AI type, allowing computers to learn the intricate relationships and complicated structures of huge datasets in order to build prediction models using labeled features.[4]

Innovations in medicine can be defined as any enhancements in the individual's quality of life and the overall quality of services.[5]

Recent innovations permitted the accessibility of information, diagnostic, and management services for most people, including those in low-income countries.[6]

A large field of computer science called artificial intelligence (AI) explores theories, techniques, tools, and software to replicate, enhance, and grow ML.[7,8]

ML is an AI branch that creates intelligent systems using statistical methods. Without being specifically designed, it is capable of learning as well as enhancing its functionality automatically, in terms of precision, using either an unsupervised or supervised approach.[9,10]

Accepted 21 May 2024.
Available online 31 May 2024

* Corresponding author at: Orthopedic Surgery, Faculty of Medicine for Boys, Al-Azhar University, Cairo, Egypt.
E-mail address: ala425826@gmail.com (A. M. M. Elsaed).

https://doi.org/10.58675/2682-339X.2442

Deep learning (DL) is a method for ML that takes advantage of cutting-edge approaches and has been very successful regarding operations related to computer vision as well as (NLP).[10] Such a success could be attributed to its superior capabilities of pattern recognition as well as extraction, accomplished by utilizing many processing layers, known as artificial neurons, to learn input representations at various abstraction levels.

2. Patients and methods

We performed this systematic review and meta-analysis based on the (PRISMA) guidelines.

Literature search: A literature review utilizing MEDLINE, Cochrane Library, Web of Science, and Scopus. Our search strategy was as follows: ("shoulder" OR "upper arm bone" OR "glenohumeral ") AND ("arthroplasty" OR "hemiarthroplasty" OR "shoulder replacement" OR "reversed shoulder") AND ("artificial intelligence" OR "machine learning" OR "deep learning" OR "AI").

Inclusion and exclusion criteria. This study includes Prospective or retrospective studies possessing follow-up periods, males and females of various ages were involved, Factors' evaluation linked to demographics, characteristics, injuries, surgeries' timeframe, sports or activities, biomechanical characteristics, muscular maturity, articular geometry, participant-reported findings as well as function at baseline. English articles and studies, Living humans, Cases of shoulder osteoarthritis need arthroplasty,

This study excludes published studies in other languages rather than English, animal or cadaveric studies, Review articles or non-prognostic studies, Missing data, Acute cases, and shoulder arthroplasty without artificial intelligence.

Screening of search results: The studies resulting from the literature search were imported to Excel software by EndNote X8.0.1, screening the imported records according to the eligibility criteria in two phases: the title /abstract phase as well as the full-text screening phase. Any conflict about the final decision on a specific study was managed by discussion.

Data extraction

Data went through extraction utilizing a standardized electronic form.

Prior to the analysis, any disagreements would be settled through discussion.

Extracting data was as follows:

Authors, Publication date, country, Number of participants, Patient demographics, Operative techniques used, Outcome measures, Follow-up period, and Complications

Data synthesis and analysis

This study analyzed the mean of our outcomes after treatment compared to pre-treatment findings. This used an Open Meta-analysis to conduct the analysis process .[11] All outcomes were continuous, so they went through analysis utilizing (MD) along with 95% (CI). We judged the heterogeneity in the outcomes according to I2 as well as the p-value of the Chi-square tests. Outcomes of I^2 more than fifty percent or P<0.1 were deemed to be heterogeneous, while those of I^2 less than fifty percent or P>0.1 were deemed to be homogenous.

3. Results

Results of the literature search

About 673 articles were included for abstract as well as title screening. Then forty-six articles were involved in the full-text screening. Finally, This study involved sixteen articles that met our criteria from the different databases as demonstrated on Figure 1 .

Our meta-analysis involved 154988 patients who underwent aTSA or rTSA. The oarticipants' mean age exhibited 69.63 years Figure 2. The included patients were 70731 men and 99842 women Figure 3.

The average follow-up period was 36 months. Table 1-2 demonstrates the involved studies' baseline characteristics.

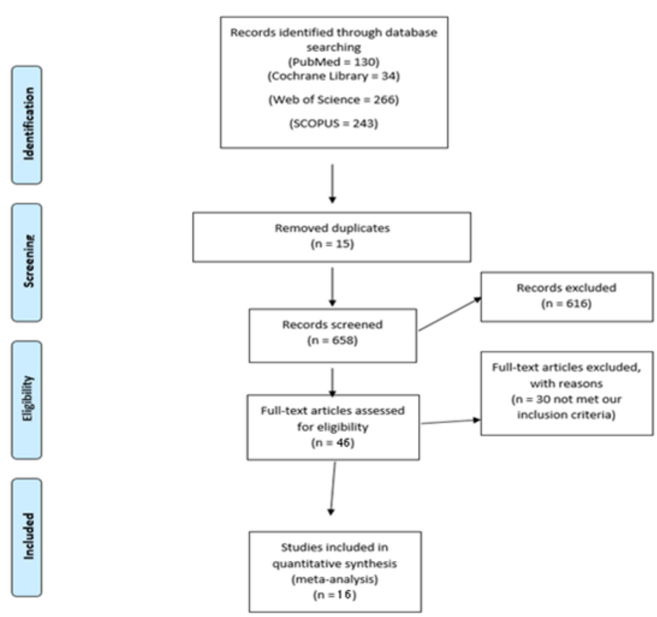

Figure 1. PRISMA flow chart.

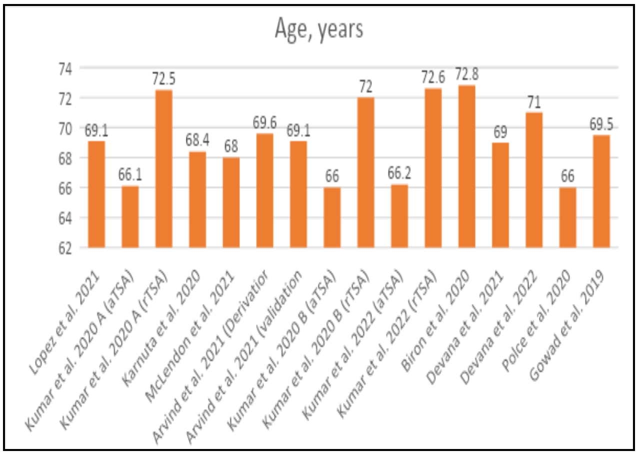

Figure 2. shows the participants' mean age within each study.

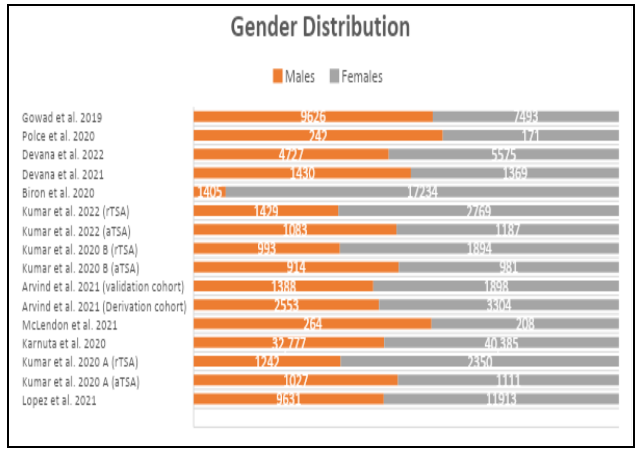

Figure 3. Gender distribution in the included study.

Table 1. demonstrates the involved studies' baseline characteristics.

Study ID	Country	study design	sample size	age (years)	Male	female
Lopez et al. 2021	USA	retrospective cohort study	21544	69.1	9631	11913
Kumar et al. 2020 (aTSA)	USA	Open-label clinical trial	2135	66.1 ± 9.2	1027	1111
Kumar et al. 2020 (rTSA)	USA	Open-label clinical trial	3621	72.5 ± 7.8	1242	2350
Karnuta et al. 2020	USA	retrospective cohort study	73,162	68.4	32,777	40,385
McLendon et al. 2021	USA	retrospective cohort study	472	68	264	208
Arvind et al. 2021 (Derivation cohort)	USA	Prospective study	5,857	69.6	2553	3304
Arvind et al. 2021 (validation cohort)	USA	Prospective study	3186	69.1	1388	1898
Kumar et al. 2021 (aTSA)	USA	Prospective study	1895	66 ± 9	914	981
Kumar et al. 2021 (rTSA)	USA	Prospective study	2887	72 ± 8	993	1894
Kumar et al. 2022 (aTSA)	USA	retrospective cohort study	2270	66.2 ± 9.1	1083	1187
Kumar et al. 2022 (rTSA)	USA	retrospective cohort study	4198	72.6 ± 7.9	1429	2769
Biron et al. 2020	USA	Prospective study	3128	72.8	1405	17234
Devana et al. 2021	USA	retrospective cohort study	2799	69	1430	1369
Devana et al. 2022	USA	retrospective cohort study	10302	71	4727	5575
Polce et al. 2020	USA	retrospective cohort study	413	66 ± 3.5	242	171
Gowad et al. 2019	USA	Prospective study	17,119	69.5 ± 9.6	9626	7493

Table 2: demonstrates the baseline characteristics of the included studies.

Study ID	Follow up (months)	BMI	Osteoarthritis	History of smoking	Diabetes	Hypertension
Lopez et al. 2021	NR	31.5	NR	4114	8897	1410
Kumar et al. 2020 (aTSA)	46.4 ± 35.6	29.9 ± 6.3	1980	210	260	1011
Kumar et al. 2020 (rTSA)	31± 25.8	28.7 ± 6.0	660	260	496	1930
Karnuta et al. 2020	NR	NR	NR	NR	13681	NR
McLendon et al. 2021	36	NR	472	NR	NR	NR
Arvind et al. 2021 (Derivation cohort)	NR	NR	4054	579	983	3942
Arvind et al. 2021 (validation cohort)	NR	NR	1439	354	575	2126
Kumar et al. 2021 (aTSA)	40 ± 30	NR	NR	NR	NR	NR
Kumar et al. 2021 (rTSA)	31 ± 22	NR	NR	NR	NR	NR
Kumar et al. 2022 (aTSA)	36	NR	NR	NR	NR	NR
Kumar et al. 2022 (rTSA)	36	NR	NR	NR	NR	NR
Biron et al. 2020	NR	NR	NR	352	666	2135
Devana et al. 2021	NR	NR	213	NR	404	NR
Devana et al. 2022	NR	NR	**737**	NR	1401	NR
Polce et al. 2020	NR	28.6	319	11	38	214
Gowad et al. 2019	NR	31.1 ± 6.8	13,725	1786	14,110	11,615

Outcomes:

AUPRC score for the best model

The results of pooled data from eleven studies (2-12) showed that the average AUPRC score was 0.839 (0.786, 0.893), and significant heterogeneity among pooled studies was observed (I2 =99.8%, P< 0.001) Figure 4

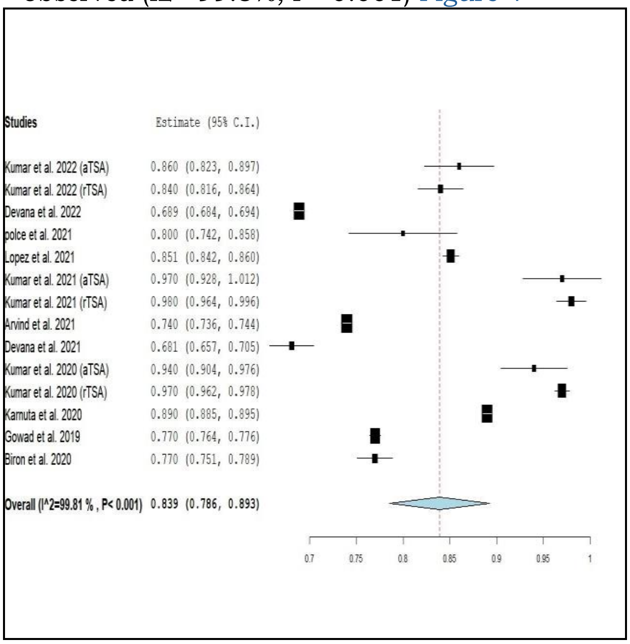

Figure 4. Forst plot shows the analysis of AUPRC score.

MAE linked to the Global Shoulder Function score ML prediction.

Kumar et al. (1)conducted a comparison of mean absolute error (MAE) among the full as well as abbreviated AI models, revealing comparable predictive accuracy for Global Shoulder Function in each model (±1.4 vs. ±1.5). Notably, over all examined postoperative time points, the average MAE variation in predictions among the full as well as abbreviated models exhibited consistently ±0.1 for Global Shoulder Function (±0.1 for aTSA and ±0.1 for rTSA). In our pooled analysis, results showed that the mean absolute error linked to the Global Shoulder Function score was 1.025 (0.980, 1.070). A significant heterogeneity was observed among pooled studies (I2 =93.02%, P = 0.000) Figure 5.

The features associated with improvement of Global shoulder function were age less than or equal to 58 (p<0.001), non diabetic (p<0.001), ASA class less than 2 (p<0.001), male sex (p<0.001), white race (p<0.001), as well as surgical procedure done within a more recent year (p=0.001)

MAE linked to the VAS pain score ML prediction

Throughout the analysis of postoperative time points, Kumar et al. (1) observed that both the full as well as abbreviated models exhibited comparable Mean Absolute Error (MAE) for the (VAS) pain (±1.3 vs. ±1.4). Our combined data for this metric indicated that the MAE linked to the VAS pain score was 1.00 (.997, 1.003). Pooled studies were homogeneous (I2 = 0%, P = 1.00), Figure 5

The features associated with decrease VAS score are age less than or equal to 55 (p<0.001),not diabetic (p<0.001), not hypertensive (p=0.008), ASA class less than 3 (p<0.001), female sex (p<0.001), black race (p<0.001),

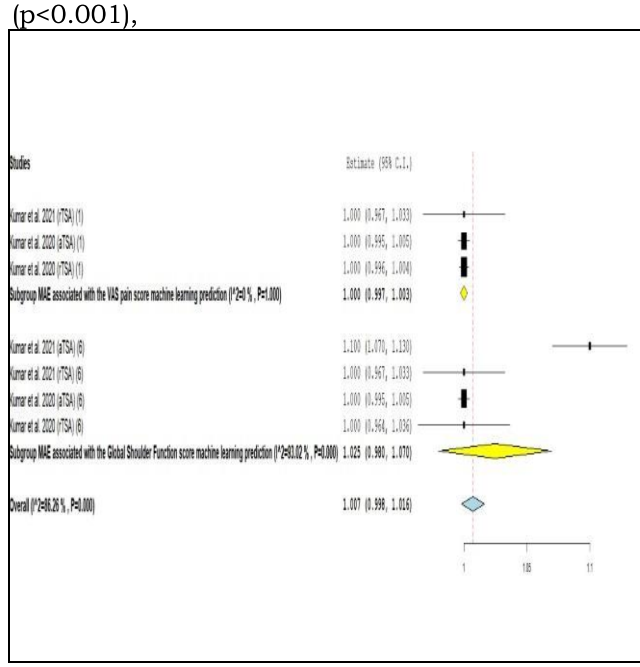

Figure 5. Forst plot shows MAE analysis linked to the VAS pain score and the Global Shoulder Function score.

MAE associated with ASES score machine learning prediction

Throughout the examination of various postoperative time points by Kumar et al.(1), they observed that the full as well as abbreviated models exhibited comparable Mean Absolute Error (MAE) for external rotation (±12.2° vs. ±12.6°). Our combined analysis indicated an overall mean absolute error for active external rotation at 12.771 (11.866, 13.675). Notably, significant heterogeneity was observed (I2 = 98.44%, P = 0.000), Figure 6

The result improved with a patient with patient preoperative comfortable sleep on the affected side, patient's usual activity, patient comb hair, washing back, lifting ten ibs above the shoulder

MAE is associated with the active abduction score machine learning prediction.

Kumar et al.[1] demonstrated that at all evaluated postoperative time intervals, both the full as well as abbreviated models exhibited comparable Mean Absolute Error (MAE) for active abduction (±20.4° vs. ±21.8°). Our combined analysis revealed that the mean MAE linked to the active abduction score was 20.998 (19.727, 22.269). The pooled data displayed heterogeneity (I2 = 96.9%, P = 0.000), Figure 6.

The features linked to improved active abduction involved ages of fifty or younger ($p<0.001$), not diabetics ($p<0.001$), ASA class less than 3 ($p<0.001$), male sex ($p<0.001$), and black race ($p<0.001$).

MAE linked to Active forward elevation, machine learning prediction

Kumar et al. (1)noted that throughout the analysis of various postoperative time points, both the full as well as abbreviated models demonstrated comparable Mean Absolute Error (MAE) for forward elevation (±17.6° vs. ±19.2°). The combined analysis revealed an overall MAE associated with active forward elevation at 17.663 (16.985, 18.341). Significant heterogeneity was observed (I2 = 92.88%, P = 0.000), Figure 6

The features associated with improved Active forward elevation involved ages 57 or younger ($p<0.001$), not diabetes ($p<0.001$), nonsmoker ($p=0.008$), not hypertensive ($p=0.008$), ASA class less than 4 ($p<0.001$), male sex ($p<0.001$), white race ($p<0.001$).

MAE associated with the Active external rotation machine learning prediction

Throughout the examination of various postoperative time points by Kumar et al.(1), they observed that the full as well as abbreviated models exhibited comparable Mean Absolute Error (MAE) for external rotation (±12.2° vs. ±12.6°). Our combined analysis indicated an overall mean absolute error for active external rotation at 12.771 (11.866, 13.675). Notably, significant heterogeneity was observed (I2 = 98.44%, P = 0.000), Figure 6.

The features linked to improving Active external rotation involved ages 53 or younger ($p<0.001$), not diabetics ($p<0.001$), nonsmoker ($p<0.001$), not hypertensive ($p=0.008$), ASA class less than 3($p<0.001$), female sex ($p<0.001$), white race ($p<0.001$).

MAE associated with the constant score machine learning prediction

Throughout the analysis of various postoperative time points, Kumar et al. (1) observed that both the full as well as abbreviated models demonstrated comparable Mean Absolute Error (MAE) for the Constant score (±8.9 vs. ±9.8). Our pooled analysis showed that the overall MAE associated with the constant score was 9.095 (8.916, 9.275). We faced a significant heterogeneity (I2 =81.5%, P = 0.001) Figure 6.

The result improved with preoperative raise of the arm to the top of the head, pre-operative move of the arm to the lumbosacral junction, preoperative move of dorsum of the hand to buttocks, preoperative move of arm behind the head with elbow held forward, preoperative do a usual activity, preoperative move arm to xiphoid, preoperative comfort of sleep, preoperative more dorsum of the hand to lateral thigh.

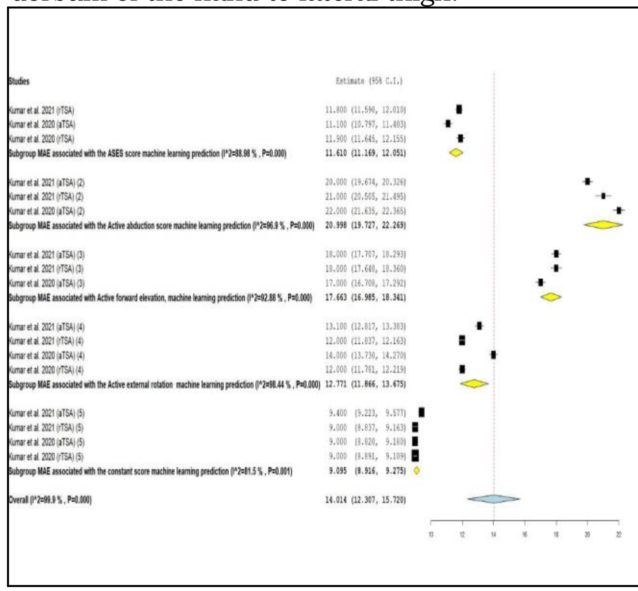

Figure 6. Forst plot shows the analysis of the remaining outcomes.

4. Discussion

(AI) is capable of revolutionizing orthopedics practice, particularly in shoulder surgery, influencing both clinical settings and operating rooms. This advanced technology is poised to enhance value-based payment models, assist in patients' categorization based on their risks, and enhance overall outcomes, utilizing personalized optimization and empirically supported collaborative decision-making models. Past achievements while utilizing (AI) for shoulder arthroplasty and rotator cuff patients, as well as other applications involving medical as well as surgical issues' prediction, overall outcomes, and implant identification, have been remarkable.[12]

These results were equivalent to results done by Kumar et al., who discovered that three distinct machine learning techniques, commercially available, effectively predicted therapeutic outcomes across various postoperative time points following aTSA and TSA. Our success extended to the accurate risk stratification of patients, predicting those likely to exhibit clinical enhancement surpassing the MCID and substantial clinical benefit patient satisfaction thresholds for each outcome measure. While further refinement of the model is required for efficient application in clinical procedures, the current research underscores the imminent capabilities of ML algorithms to address unresolved inquiries as well as enhance the decision process for improved overall results in shoulder arthroplasty.[13]

In Gowd's study, logistic regression, representing the predominant ML basic type, proved to be more effective than cutting-edge AI systems regarding negative events' prediction, specifically surgical site infections; the AI-based algorithms utilized for shoulder arthroplasty literature often failed to achieve a satisfactory performance, ranging from acceptable to exceptional. Of the sixteen studies assessed, only 11 (area under the curve) were reported to have failed to exceed 0.90, which is considered excellent performance. Likely, a model performing only at a satisfactory or fair level when internally validated will exhibit poor performance during external validation.[14,15].

Lopez et al. observed that Both the boosted decision tree as well as artificial neural network demonstrated effective performance regarding non-home discharge prediction, exhibiting comparable precision. However, the artificial neural network exhibited superior classification accuracy. The study's results suggest that machine learning can precisely forecast facility-based discharge after elective TSA. Clinicians could leverage these methods to inform patients about potential outcomes and enhance preoperative discharge planning, ultimately aiming to reduce hospitalization duration as well as enhance affordability.[16]

4. Conclusion

As technology continues to advance, the use of AI in shoulder arthroplasty is becoming increasingly common. AI can be used to assist surgeons in planning and executing the procedure, as well as in predicting outcomes and identifying potential complications.

(AI) is capable of revolutionizing shoulder arthroplasty. After huge patient data analysis, AI algorithms could detect sequences as well as predictors of complications and unplanned readmissions, allowing surgeons to make more informed decisions and enhance overall results. AI could enhance preoperative planning, implant selection, and surgical navigation, leading to more accurate and efficient procedures. However, some challenges could exist involving information confidentiality as well as algorithm validation; utilizing AI in shoulder arthroplasty holds great promise for improving patient care and advancing the field of orthopedic surgery.

The use of AI in shoulder surgery is rapidly growing, providing personalized risk categorization for collaborative decision processes along with automation to conserve resources. Nevertheless, the performance of AI models is moderate, and the need for external validation is yet to be established. This implies the necessity for heightened scientific rigor before integrating models based on AI into the clinical practice for shoulder surgical procedures. Shoulder surgeons should stay informed about AI advancements but exercise caution in applying algorithms widely until their efficacy and external validation are confirmed.

Disclosure

The authors have no financial interest to declare in relation to the content of this article.

Authorship

All authors have a substantial contribution to the article

Funding

No Funds : Yes

Conflicts of interest

There are no conflicts of interest.

References

1. Familiari F, Rojas J, Doral MN, Huri G, McFarland EG. Reverse total shoulder arthroplasty. EFORT Open Rev [Internet]. 2018;3(2):58–69.
2. Frank JK, Siegert P, Plachel F, Heuberer PR, Huber S, Schanda JE. The Evolution of Reverse Total Shoulder Arthroplasty—From the First Steps to Novel Implant Designs and Surgical Techniques. J Clin Med [Internet]. 2022;11(6) 1512.
3. Singh AM, Haju WB. Project Management in the Era of Artificial Intelligence. European Journal of Theoretical and Applied Sciences,2022; 1(3), 349-359
4. Viscaino M, Maass JC, Delano PH, Torrente M, Stott C, Auat Cheein F. Computer-aided diagnosis of external and middle ear conditions: A machine learning approach. PLoS One [Internet]. 2020;15(3):229226.
5. Flessa S, Huebner C. Innovations in health care—a conceptual framework. Int J Environ Res Public Health [Internet]. 2021;18(19):10026.
6. Sandberg CEJ, Knight SR, Qureshi AU, Pathak S. Using telemedicine to diagnose surgical site infections in low-And middle-income countries: Systematic review. JMIR mHealth uHealth [Internet]. 2019;7(8):13309.
7. Ju S, Chen C, Zhang J, Xu L, Zhang X, Li Z, Chen Y, Zhou J, Ji F, Wang L. Detection of circulating tumor cells: opportunities and challenges. Biomark Res. 2022 Aug 13;10(1):58
8. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence. AI Mag [Internet]. 2006;27(4):12-14
9. Samuel AL. Some studies in machine learning using the game of checkers. In: IBM Journal of Research and Development. Springer New York; 2000; 44.1.2: 206-226.
10. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature [Internet]. 2015;521(7553):436–444.
11. Wallace BC, Lajeunesse MJ, Dietz G, Dahabreh IJ, Trikalinos TA, Schmid CH, et al. OpenMEE: Intuitive, open-source software for meta-analysis in ecology and evolutionary biology. Methods Ecol Evol [Internet]. 2017;8(8):941–947.
12. Tian C, Chen X, Zhu H, Qin S, Shi L, Rui Y. [Application and prospect of machine learning in orthopaedic trauma]. Journal of Reparative and Reconstructive Surgery, 2023; 37.12: 1562.
13. Kumar V, Roche C, Overman S, 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(10):2351-2363.
14. 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. 2020;14(1):49-58.
15. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol. 2015;68(1):25-34.
16. 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(4):692-698.