Arthroplasty Today 27 (2024) 101396

Contents lists available at ScienceDirect

Arthroplasty Today

journal homepage: http://www.arthroplastytoday.org/

Special Issue: Highlights from the Young Arthroplasty Group (YAG) of AAHKS. Edited by Anna A.R.Cohen-Rosenblum

Overview of Artificial Intelligence Research Within Hip and Knee Arthroplasty

John P. Mickley, MD[a], Elizabeth S. Kaji, BA[a], Bardia Khosravi, MD, MPH, MHPE[a, b], Kellen L. Mulford, PhD[a], Michael J. Taunton, MD[a, c], Cody C. Wyles, MD[a, c, d, *]

[a] Orthopedic Surgery Artificial Intelligence Laboratory (OSAIL), Department of Orthopedic Surgery, Mayo Clinic, Rochester, MN, USA
[b] Radiology Informatics Lab (RIL), Department of Radiology, Mayo Clinic, Rochester, MN, USA
[c] Department of Orthopedic Surgery, Mayo Clinic, Rochester, MN, USA
[d] Department of Clinical Anatomy, Mayo Clinic, Rochester, MN, USA

ARTICLE INFO

Article history:
Received 8 September 2023
Received in revised form
14 March 2024
Accepted 2 April 2024
Available online 23 April 2024

Keywords:
Arthroplasty
Artificial intelligence
Deep learning
Registry construction
Risk modeling
Uncertainty quantification

ABSTRACT

Hip and knee arthroplasty are high-volume procedures undergoing rapid growth. The large volume of procedures generates a vast amount of data available for next-generation analytics. Techniques in the field of artificial intelligence (AI) can assist in large-scale pattern recognition and lead to clinical insights. AI methodologies have become more prevalent in orthopaedic research. This review will first describe an overview of AI in the medical field, followed by a description of the 3 arthroplasty research areas in which AI is commonly used (risk modeling, automated radiographic measurements, arthroplasty registry construction). Finally, we will discuss the next frontier of AI research focusing on model deployment and uncertainty quantification.

Introduction

Total hip arthroplasty (THA) and total knee arthroplasty (TKA), which we will refer to as “arthroplasty” for the rest of this paper, are high-volume procedures that have witnessed remarkable advancements and growth. In the U.S. alone, there are over 7 million individuals with total hip or total knee replacements [1,2]. The national arthroplasty volume increases annually, and this trajectory is expected to continue for the foreseeable future [3,4]. In addition to the growing case volume, there has also been an increase in the availability of arthroplasty data due to the widespread adoption of electronic medical record systems and clinical orthopaedic registries, which provides an opportunity for population-scale research [5,6]. Analyzing the large volume of data produced annually poses a significant hurdle to traditional methods of data analysis that require extensive human manual effort. Fortunately, the gradual adoption of techniques from the field of artificial intelligence (AI) in medicine provides a new opportunity to leverage these data resources.

AI is a broad subfield of computer science that deals with technologies that are capable of mimicking human cognitive functions. While AI is not a new field, its use in orthopaedics has grown significantly during the last decade [7]. AI algorithms can be trained to efficiently extract meaningful insights from large datasets, which may enable more personalized treatment decisions, optimized surgical techniques, enhanced postoperative care, and improved patient outcomes in total joint arthroplasty. This article briefly describes AI in the context of arthroplasty, followed by a summary of the major AI-related research topics within arthroplasty. We will end by discussing the future direction of AI in arthroplasty.

What is artificial intelligence?

AI encompasses a range of computer systems that learn to emulate human intelligence. Commonly used phrases “machine learning” (ML) and “deep learning” (DL) are actually subfields of AI (Fig 1). ML is the application of statistical models that learn patterns

* Corresponding author. Cody C. Wyles, MD, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. Tel.: +1 507 284 2511.
E-mail address: wyles.cody@mayo.edu

https://doi.org/10.1016/j.artd.2024.101396

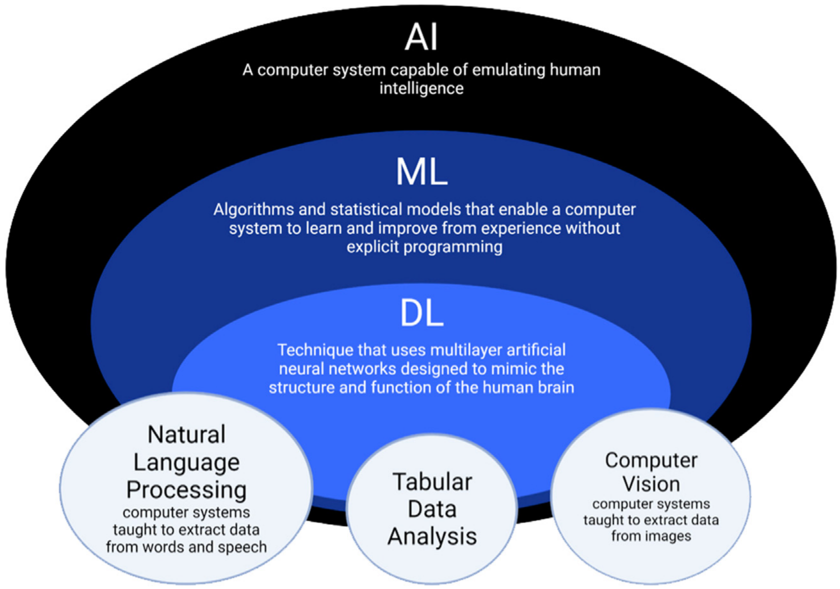

Figure 1. Subfields of artificial intelligence.

in datasets and can provide predictions on unseen data. Many popular algorithms involved in regression, classification, and clustering can be thought of as ML models. Within ML, DL is an increasingly popular technique that relies on the use of a specific ML algorithm and artificial neural networks with many layers [8]. Because DL algorithms can easily process unstructured data (like images and clinical report text) and because they learn for themselves what features of the data are important, they are ideal for addressing the common challenges encountered in the medical field. Which ML/DL solution is most appropriate is inextricably tied to the characteristics of the dataset being analyzed.

Data can be either structured or unstructured. One example of structured data is a spreadsheet, where each row represents an instance and each column is a variable or feature associated with the instances. For example, the rows in a dataset may represent patients, with lab test results occurring in the columns. While a table is structured data at a basic level, complex relations between instance types can be modeled using relational database technologies such as Structured Query Language. Many statistical and ML techniques naturally lend themselves to structured data.

The vast majority of medical data is unstructured. Unstructured data cannot be easily formatted into rows and columns. Clinical notes, operative reports, and medical images are all examples of data that are unstructured. Entire subfields of AI have grown out of the need to develop specialized ways of analyzing unstructured data (Fig 1). Natural language processing (NLP) is a subfield of AI that involves the development of algorithms that can interpret and generate human language in a meaningful way [9,10]. Likewise, computer vision refers to the field of computer science that focuses on enabling computers to interpret visual information from images or videos. In the context of arthroplasty research, computer vision plays a crucial role in analyzing medical images, such as radiographs, computed tomography scans, or magnetic resonance imaging scans, to assist in diagnosis and surgical planning. DL is especially powerful when used for the analysis of unstructured data because deep neural networks simultaneously learn a task and learn which features are most important to that task.

The following 3 sections will focus on major arthroplasty research topics.

Risk modeling

The use of AI in risk modeling refers to the analysis of patient data to make predictions that can then assist with clinical decision-making. While clinical prediction models have always been a fundamental part of biomedical research, the concomitant rise in computing power and large datasets has catalyzed the use of AI techniques in prediction models [11,12]. In this section, we will focus on complication and outcome prediction on the topic of risk modeling.

Using the enormous quantity of patient data now available, DL/ML models can predict the likelihood of complications such as infections, implant-related issues, and other adverse events with high accuracy. Yeo et al. used structured clinical data points to develop ML models that predict surgical site infections following a primary TKA [13]. While the most important features in that model were mostly previously identified factors for infection, the ML algorithm employed allowed for much higher accuracy than traditional analytic methods. Wyles et al. created an algorithm able to show how the risk of periprosthetic fracture changes in a specific patient based on surgical factors such as uncemented femoral fixation, collarless femoral implants, and surgical approach [14]. Likewise, Jo et al. developed a model that used patient demographics, labs, and history to predict transfusion requirements after a TKA. This model also demonstrated good predictive performance when applied to patient data from an external institution [15]. In many cases, ML models learn a data distribution from a single institution too well and fail to generalize to data elsewhere. There are many more examples of risk modeling for TKA and THA complications including prolonged opioid abuse [16], delirium [17], and acute kidney injury [18].

ML/DL models can also integrate data of multiple types. Wyles et al. created a model capable of incorporating the possible dislocation risk modification based on surgeon decisions such as the use of dual-mobility constructs and elevated liners [19]. In a subsequent publication, Khosravi et al. demonstrated that adding embedding data (the abstract features a DL model learns during training) from a radiograph to that model improved its performance [20]. Further, this algorithm was designed to show the patient-specific risk in addition to the degree of risk modification achievable with surgical decisions, thus yielding actionable tools for surgeons.

Similar to complication prediction, outcome prediction models forecast patient outcomes after surgery, including pain levels, functional improvements, and overall satisfaction with the procedure. In a study using Medicare data, the investigators developed an algorithm to predict postoperative outcomes, which was then compared to 3 of the most commonly used risk adjustment indices [21]. The novel Complexity Score had the highest accuracy in predicting perioperative morbidity. Harris et al. developed a model to predict the Activities of Daily Living, Pain, Symptoms, and Quality of Life subscales of the Knee Injury and Osteoarthritis Outcome Score following a TKA [22]. Fontana et al. used presurgical registry data to train 4 different models to predict which patients would not achieve 2-year postsurgical minimally clinically important differences following total joint arthroplasty with fair-to-good ability [23].

In comparison with traditional statistical methods, ML/DL models typically achieve higher predictive performance at the expense of explainability. This “Black Box Phenomenon” represents a key challenge that differentiates AI-based calculators from other calculation tools. Several explainable AI techniques, from feature importance metrics to saliency maps and uncertainty quantification, can help end users of an algorithm feel more confident about how a model is making its prediction. Which method is most appropriate is specific to the ML/DL model employed. Additionally, some may be concerned about the impact of online, publicly available calculators and how they could confuse rather than help patients. Clear explanations of how individual factors impact a risk score will help with popular usage as well. The advent of a trusted and accurate calculator for patient-specific postoperative complications and outcomes offers an additional tool for clinicians in order to continue to improve shared decision-making between surgeons and patients.

Automated radiographic analysis

Medical providers in orthopaedic surgery rely on imaging to diagnose pathology, plan treatment, and monitor outcomes. It is perhaps the most important data source in arthroplasty. Clinicians take measurements using imaging to assess either the degree of anatomic abnormality in a patient or to determine the position of implanted components, for example. In the context of revision surgery, accurate identification of existing implants' manufacturer and model is paramount. While this information plays a significant role in providing optimal patient care, the measurements themselves can be tedious, difficult to perform, and potentially subject to significant interrater variability. Computer vision algorithms can automate radiographic measurements and improve measurement robustness.

AI in radiology is a huge topic, spawning numerous journals and conferences in recent years. In arthroplasty research, AI studies have focused mostly on automating measurements and extracting semantic information (radiological findings) [24-27]. Rouzrokh et al. utilized computer vision techniques to measure femoral component subsidence between 2 serial anteroposterior radiographs; the median difference between the independent orthopaedic surgeon reviewer and automated measurements was 0.3 mm [28]. Another study by the same group presented an algorithm that calculates acetabular inclination and version with similarly high performance [27]. Evaluation of these algorithms is crucial; investigators must show that the performance of their algorithm meets or exceeds the performance of a human annotator.

AI techniques can also automate the extraction of semantic information from the image itself. Stotter et al. demonstrated that an AI algorithm ranked better than at least one manual reader for the majority of outcome measures when measuring radiological parameters that identify femoroacetabular impingement and hip dysplasia [29]. A particularly exciting use for DL/ML models is to extract information that would be difficult for an expert to ascertain. For example, while an experienced surgeon may be able to identify several models of hip arthroplasty implants, several recent published studies have trained models to identify a wide variety of implants with near-perfect accuracy [30].

Automated radiographic measurements can greatly increase the efficiency and generalizability of treatment planning for arthroplasty. Lambrechts et al. demonstrated a 39.7% reduction in the number of corrections the surgeon had to make from an AI-generated preoperative plan compared to the manufacturer's default plan [31]. Following the current process, THA surgeons often template based on personal experience resulting in different outcomes based on experience level [32]. A universally accepted algorithm for templating may eliminate some of this variability, and the ability to create preoperative plans within seconds would save surgeons time [29,33]. Generative AI may also help with visualization of postoperative hips [34].

Of course, employing AI in the analysis of radiographs presents several potential challenges. One primary concern lies in the diversity of image collection methodologies across various hospitals or even between different imaging personnel based in the same hospital. It is conceivable that differences in positioning can significantly impact measurements on planar images. However, uniform imaging methodologies are necessary even without the application of AI to radiograph analysis. Human readers will suffer from the same errors as AI when faced with radiographs taken using different techniques or of patients in different positions. Developers of AI algorithms can combat this variability with diverse data sources, best practices to avoid data leakage during training, and robust external validation. An algorithm is only as good as its training data, so proper oversight and training of the annotators curating the training data is also important. Finally, these AI algorithms allow for the extraction of radiographic information for further research on a volume of images not previously possible. In datasets of that size, it is not possible to validate each measurement or data point extracted by the algorithm. Again, robust validation can help improve the trustworthiness of these algorithms.

Moving on, we will now discuss the development of large registries, an extremely important task required for robust orthopaedic research, which can be expedited by automated annotation of radiographic images and other applications of AI.

Arthroplasty registry construction

Large-scale clinical registries have long been an important source of data for orthopaedic research, where clinical trials are especially expensive and difficult. Institutional and national registries provide an invaluable resource for researchers. Registries that rely on manual abstraction of data points are expensive, while registries that only use data coded into electronic health record (EHR) fields are usually shallow or incomplete. In the most recent American Joint Replacement Registry update, only 10% of procedures had a surgical approach reported. Automated methods of data curation from images and medical records could help bridge the gap from depth to completeness.

Most data in the EHR is unstructured text data, which requires specific analytic techniques in a field called NLP. NLP is a broad field that uses a wide range of techniques, from traditional statistical models adapted to analyze unstructured text to highly advanced DL models. When applied to medical research, NLP techniques can analyze text found in EHRs and subsequently use that for registry input [35-39]. Wyles et al. published a proof-of-concept of the NLP technology and utilized it to identify common elements described by surgeons in operative THA. The NLP algorithm extracted the operative approach, fixation technique, and bearing surface with accuracies of 99.2%, 90.7%, and 95.8%, respectively, mimicking the performance of human annotators with much higher efficiency [35]. NLP decreases the necessity of specialized, highly trained medical professionals to extract the data. By removing the labor-intensive part of the project, NLP allows the information to be collected expeditiously and cost-effectively.

Quite recently, the field of NLP has experienced a renaissance with the advent of increasingly sophisticated large language models (LLMs). LLMs are gigantic DL models (most have *billions* of parameters) that generate text after receiving some input text (or images) as a prompt. ChatGPT, GPT-4, LLaMa, and Gemini are all recent examples of LLMs and offer unique promise for the efficient extraction of free text data and also for the novel generation of data summaries [40]. While well-known for their human-like response capabilities, LLMs have shown remarkable success in completing medical exams [41], summarizing radiology reports [42], and plenty of other tasks that had previously employed NLP techniques. Their main function is to understand and generate natural language that can be applied to tasks such as summarization, translation, and question-answering [43]. LLMs may soon supplant NLP algorithms in automating registry curation.

As previously mentioned, AI has the potential to aid in radiographic annotation and measurements, which in turn can be leveraged for imaging data extraction for registry establishment. Rouzrokh et al. trained a DL model to efficiently annotate and categorize the view, laterality, and operative status of THA patient radiographs. The algorithm demonstrated impressive results, achieving 99.9% accuracy, 99.6% precision, 99.5% recall, and a 99.6% F1 score in assessing radiographic characteristics [44]. A fully automated registry could rely on NLP to extract valuable data from clinical text and computer vision to automate data extraction from medical images, collectively improving registry accuracy and efficiency.

What is next?

Over the next decade, the field of AI in arthroplasty will continue to expand and change. We believe the next wave of AI research will focus on 3 themes: 1) clinical implementation of algorithms; 2) AI trustworthiness; and 3) increased utilization of generative AI including LLMs [45-47].

Of the thousands of AI algorithms published in biomedical research each year, it is likely that very few will be integrated into the clinic and impact patient care. This is a complicated issue, but we see a few obvious reasons. The first, as we have discussed above, is that AI algorithms rarely generalize well outside of the data used to train it. This means each algorithm, unless trained and validated on a broad swath of data, only performs well at the institution that trained it. The second reason is that there is a lot of infrastructure necessary to transform a model on a researcher's computer into something that can be integrated into a clinical system and then monitored and updated. This process is the focus of a field called MLOps, well known to technology companies but still somewhat new to healthcare organizations. Any MLOps processes need to be closely paired with implementation studies. Successful implementation of the model is not the final step; it is essential to validate and intermittently improve the model by adding new training data [48]. A third reason that makes implementation difficult is the complex legal and regulatory issues related to the clinical use of an algorithm [49]. Ultimately, the final responsibility for the patient's health rests with the attending physician; for clinicians to regularly use AI algorithms, they need to trust them.

One of the current problems with AI approaches is the lack of explanation for the output of the models, commonly called the "black-box" phenomenon [50]. Without using the techniques of explainable AI, it is difficult to comprehend how AI arrives at its outputs or predictions, raising concerns about accountability and potential biases embedded within its operations [51]. One way to help providers appropriately use the output of models is to know how certain the model is about its prediction (uncertainty quantification) or to include information on what factors helped the algorithm reach its prediction (feature importance).

Explainable involves assessing i) the variability in scientific models and ii) the way the algorithm uses input features to make a prediction. There are a variety of techniques for adding explainability to an algorithm that analyze input parameters, model assumptions, and measurement errors (56-58). For example, Rouzrokh et al. added conformal prediction to an AI model trained to identify THA implants, thus providing the ability to quantify prediction uncertainty and flag outlier test data—both essential for clinical trustworthiness [30,34]. By comprehensively characterizing model uncertainties, researchers can enhance the accuracy and reliability of their models, ultimately leading to improved AI trustworthiness. Furthermore, quantifying uncertainty can identify areas that necessitate further research to reduce uncertainties and enhance our understanding of arthroplasty approaches.

Another rapidly evolving field in AI is generative AI. This branch of AI focuses on the novel synthesis of content rather than analyzing existing data. The newly created data can come in the form of images, text, audio, and other mediums. This area of computer science is rapidly advancing in medicine; Epic recently announced a collaboration with Microsoft Corp. to "develop and integrate generative AI into healthcare" [52]. LLMs, a type of DL previously discussed in this manuscript, are a type of generative AI. In the field of arthroplasty, generative AI could potentially be used for data augmentation and synthesis, custom implant design, and surgical simulation. LLMs hold promise to accurately summarize extensive patient data and research publications to aid physicians in informed decisions [53]. The translation function could be applied to not just language barriers but also the jargon-dense medical text within EHRs that sometimes challenges patients [54]. The question-answering function could relieve providers of the often tedious task of answering simple questions patients send via the online messaging function within most EHR systems. Additionally, a chatbox-like function could also be implemented in other areas of AI to attempt to add transparency to existing algorithms [50].

Conclusions

An increase in the volume of arthroplasty procedures and the data produced have opened the door to new research opportunities. AI techniques are a powerful way to analyze these new data streams. This article has surveyed several major research areas of AI within arthroplasty: risk modeling, automated radiographic analysis, and automated registry curation. These themes are both mechanistic and infrastructural. In the coming years, we expect some of the major themes of future AI research in medicine to include 1) implementation science, 2) explainable AI, and 3) generative AI. Despite already having a profound effect on the research landscape, we expect that the largest changes to the arthroplasty community will occur with the migration of AI technologies to the clinic.

CRediT authorship contribution statement

John P. Mickley: Writing – original draft, Methodology, Investigation, Conceptualization, Writing – review & editing. **Elizabeth S. Kaji:** Investigation, Writing – original draft, Writing – review & editing. **Bardia Khosravi:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Kellen L. Mulford:** Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Michael J. Taunton:** Supervision, Writing – original draft, Writing – review & editing. **Cody C. Wyles:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Acknowledgments

The authors express their sincere gratitude to the entire Orthopedic Surgery Artificial Intelligence Laboratory (OSAIL) at Mayo Clinic for their assistance, resources, and expertise in assisting in the preparation of this manuscript.

Conflicts of interest

M. Taunton receives IP royalties from Enovis, is a paid consultant for ONKOS and Enovis, is an editorial/governing board member of the Journal of Arthroplasty, and is a board or committee member of the American Association of Hip and Knee Surgeons, AAOS, and MAOA. All other authors declare no potential conflicts of interest.

For full disclosure statements refer to https://doi.org/10.1016/j.artd.2024.101396.

References

[1] Maradit Kremers H, Larson DR, Crowson CS, Kremers WK, Washington RE, Steiner CA, et al. Prevalence of total hip and knee replacement in the United States. J Bone Joint Surg Am 2015;97:1386–97.
[2] Siddiqi A, Levine BR, Springer BD. Highlights of the 2021 American joint replacement registry annual report. Arthroplast Today 2022;13:205–7.
[3] Kurtz SM, Lau E, Ong K, Zhao K, Kelly M, Bozic KJ. Future young patient demand for primary and revision joint replacement: national projections from 2010 to 2030. Clin Orthop Relat Res 2009;467:2606–12.
[4] Cram P, Lu X, Kates SL, Singh JA, Li Y, Wolf BR. Total knee arthroplasty volume, utilization, and outcomes among Medicare beneficiaries, 1991-2010. JAMA 2012;308:1227–36.
[5] Findlay S. Physician compare [Internet]. https://healthaffairs.org/do/10.1377/hpb20141211.552144/full. [Accessed 29 May 2023].
[6] Sharfstein JM. Using health care data to track and improve public health. JAMA 2015;313:2012–3.
[7] Farhadi F, Barnes MR, Sugito HR, Sin JM, Henderson ER, Levy JJ. Applications of artificial intelligence in orthopaedic surgery. Front Med Technol 2022;4:995526.
[8] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[9] Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol 2019;28:73–81.
[10] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015;115:211–52.
[11] Huang Z, Martin J, Huang Q, Ma J, Pei F, Huang C. Predicting postoperative transfusion in elective total HIP and knee arthroplasty: comparison of different machine learning models of a case-control study. Int J Surg 2021;96:106183.
[12] Rodríguez-Merchán EC. The current role of the virtual elements of artificial intelligence in total knee arthroplasty. EFORT Open Rev 2022;7:491–7.
[13] Yeo I, Klemt C, Robinson MG, Esposito JG, Uzosike AC, Kwon YM. The use of artificial neural networks for the prediction of surgical site infection following TKA. J Knee Surg 2023;36:637–43.
[14] Wyles CC, Maradit-Kremers H, Fruth KM, Larson DR, Khosravi B, Rouzrokh P, et al. Frank Stinchfield Award: creation of a patient-specific total hip arthroplasty periprosthetic fracture risk calculator. J Arthroplasty 2023;38:S2–10.
[15] Jo C, Ko S, Shin WC, Han HS, Lee MC, Ko T, et al. Transfusion after total knee arthroplasty can be predicted using the machine learning algorithm. Knee Surg Sports Traumatol Arthrosc 2020;28:1757–64.
[16] Karhade AV, Schwab JH, Bedair HS. Development of machine learning algorithms for prediction of sustained postoperative opioid prescriptions after total hip arthroplasty. J Arthroplasty 2019;34:2272–2277.e1.
[17] Jung JW, Hwang S, Ko S, Jo C, Park HY, Han HS, et al. A machine-learning model to predict postoperative delirium following knee arthroplasty using electronic health records. BMC Psychiatry 2022;22:436.
[18] Nikkinen O, Kolehmainen T, Aaltonen T, Jämsä E, Alahuhta S, Vakkala M. Developing a supervised machine learning model for predicting perioperative acute kidney injury in arthroplasty patients. Comput Biol Med 2022;144:105351.
[19] Wyles CC, Maradit-Kremers H, Larson DR, Lewallen DG, Taunton MJ, Trousdale RT, et al. Creation of a total hip arthroplasty patient-specific dislocation risk calculator [Internet]. J Bone Joint Surg Am 2022;104:1068–80. https://doi.org/10.2106/jbjs.21.01171.
[20] Khosravi B, Rouzrokh P, Maradit Kremers H, Larson DR, Johnson QJ, Faghani S, et al. Patient-specific hip arthroplasty dislocation risk calculator: an explainable multimodal machine learning–based approach. Radiol Artif Intell 2022;4:e220067.
[21] Hyer JM, White S, Cloyd J, Dillhoff M, Tsung A, Pawlik TM, et al. Can we improve prediction of adverse surgical outcomes? Development of a surgical complexity score using a novel machine learning technique. J Am Coll Surg 2020;230:43–52.e1.
[22] Harris AHS, Kuo AC, Bowe TR, Manfredi L, Lalani NF, Giori NJ. Can machine learning methods produce accurate and easy-to-use preoperative prediction models of one-year improvements in pain and functioning after knee arthroplasty? J Arthroplasty 2021;36:112–117.e6.
[23] Fontana MA, Lyman S, Sarker GK, Padgett DE, MacLean CH. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? Clin Orthop Relat Res 2019;477:1267–79.
[24] Jang SJ, Kunze KN, Brilliant ZR, Henson M, Mayman DJ, Jerabek SA, et al. Comparison of tibial alignment parameters based on clinically relevant anatomical landmarks: a deep learning radiological analysis. Bone Jt Open 2022;3:767–76.
[25] Mulford KL, Johnson QJ, Mujahed T, Khosravi B, Rouzrokh P, Mickley JP, et al. A deep learning tool for automated landmark annotation on hip and pelvis radiographs. J Arthroplasty 2023;38:2024–2031.e1. https://doi.org/10.1016/j.arth.2023.05.036.
[26] Yang W, Ye Q, Ming S, Hu X, Jiang Z, Shen Q, et al. Feasibility of automatic measurements of hip joints based on pelvic radiography and a deep learning algorithm. Eur J Radiol 2020;132:109303.
[27] Rouzrokh P, Wyles CC, Philbrick KA, Ramazanian T, Weston AD, Cai JC, et al. A deep learning tool for automated radiographic measurement of acetabular component inclination and version after total hip arthroplasty. J Arthroplasty 2021;36:2510–2517.e6.
[28] Rouzrokh P, Wyles CC, Kurian SJ, Ramazanian T, Cai JC, Huang Q, et al. Deep learning for radiographic measurement of femoral component subsidence following total hip arthroplasty. Radiol Artif Intell 2022;4:e210206.
[29] Stotter C, Klestil T, Röder C, Reuter P, Chen K, Emprechtinger R, et al. Deep learning for fully automated radiographic measurements of the pelvis and hip. Diagnostics (Basel) 2023;13:497. https://doi.org/10.3390/diagnostics13030497.
[30] Rouzrokh P, Mickley JP, Khosravi B, Faghani S, Moassefi M, Schulz WR, et al. THA-AID: deep learning tool for total hip arthroplasty automatic implant detection with uncertainty and outlier quantification. J Arthroplasty 2024;39:966–973.e17. https://doi.org/10.1016/j.arth.2023.09.025.
[31] Lambrechts A, Wirix-Speetjens R, Maes F, Van Huffel S. Artificial intelligence based patient-specific preoperative planning algorithm for total knee arthroplasty. Front Robot AI 2022;9:840282.
[32] Holzer LA, Scholler G, Wagner S, Friesenbichler J, Maurer-Ertl W, Leithner A. The accuracy of digital templating in uncemented total hip arthroplasty. Arch Orthop Trauma Surg 2019;139:263–8.
[33] Archer H, Reine S, Alshaikhsalama A, Wells J, Kohli A, Vazquez L, et al. Artificial intelligence-generated hip radiological measurements are fast and adequate for reliable assessment of hip dysplasia : an external validation study. Bone Jt Open 2022;3:877–84.
[34] Rouzrokh P, Khosravi B, Mickley JP, Erickson BJ, Taunton MJ, Wyles CC. THA-Net: a deep learning solution for next-generation templating and patient-specific surgical execution. J Arthroplasty 2023;39:727–733.e4. https://doi.org/10.1016/j.arth.2023.08.063.
[35] Wyles CC, Tibbo ME, Fu S, Wang Y, Sohn S, Kremers WK, et al. Use of natural language processing algorithms to identify common data elements in operative notes for total hip arthroplasty. J Bone Joint Surg Am 2019;101:1931–8.
[36] Sagheb E, Ramazanian T, Tafti AP, Fu S, Kremers WK, Berry DJ, et al. Use of natural language processing algorithms to identify common data elements in operative notes for knee arthroplasty. J Arthroplasty 2021;36:922–6.
[37] Wyles CC, Fu S, Odum SL, Rowe T, Habet NA, Berry DJ, et al. External validation of natural language processing algorithms to extract common data elements in THA operative notes. J Arthroplasty 2023;38:2081–4. https://doi.org/10.1016/j.arth.2022.10.031.
[38] Han P, Fu S, Kolis J, Hughes R, Hallstrom BR, Carvour M, et al. Multicenter validation of natural language processing algorithms for the detection of common data elements in operative notes for total hip arthroplasty: algorithm development and validation. JMIR Med Inform 2022;10:e38155.
[39] Shah RF, Bini S, Vail T. Data for registry and quality review can be retrospectively collected using natural language processing from unstructured charts of arthroplasty patients. Bone Joint J 2020;102-B:99–104.
[40] OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, et al. GPT-4 technical report [Internet] [cs.CL] arXiv 2023., http://arxiv.org/abs/2303.08774.

[41] Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2:e0000198.
[42] Adams LC, Truhn D, Busch F, Kader A, Niehues SM, Makowski MR, et al. Leveraging GPT-4 for post hoc transformation of free-text radiology reports into structured reporting: a multilingual feasibility study. Radiology 2023;307:e230725.
[43] Introducing ChatGPT [Internet]. https://openai.com/blog/chatgpt. [Accessed 28 August 2023].
[44] Rouzrokh P, Khosravi B, Johnson QJ, Faghani S, Vera Garcia DV, Erickson BJ, et al. Applying deep learning to establish a total hip arthroplasty radiography registry: a stepwise approach. J Bone Joint Surg Am 2022;104:1649–58.
[45] Zhang A, Xing L, Zou J, Wu JC. Shifting machine learning for healthcare from development to deployment and from models to data. Nat Biomed Eng 2022;6:1330–45.
[46] He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25:30–6.
[47] Erickson BJ. Chapter 2 - artificial intelligence in medicing: technical basis and clinical applications. In: Xing L, Giger ML, Min JK, editors. Artificial intelligence in medicine: technical basis and clinical applications. Cambridge, MA: Academic Press; 2020. p. 19–34.
[48] Oeding JF, Williams 3rd RJ, Camp CL, Sanchez-Sotelo J, Kelly BT, Nawabi DH, et al. A practical guide to the development and deployment of deep learning models for the orthopedic surgeon: part II. Knee Surg Sports Traumatol Arthrosc 2023;31:1635–43.
[49] Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. Nat Med 2021;27:582–4.
[50] Kumar V, Patel S, Baburaj V, Vardhan A, Singh PK, Vaishya R. Current understanding on artificial intelligence and machine learning in orthopaedics - a scoping review. J Orthop 2022;34:201–6.
[51] Zou J, Gichoya JW, Ho DE, Obermeyer Z. Implications of predicting race variables from medical images. Science 2023;381:149–50.
[52] Microsoft News Center. [No title] [Internet]. https://news.microsoft.com/2023/04/17/microsoft-and-epic-expand-strategic-collaboration-with-integration-of-azure-openai-service/. [Accessed 18 August 2023].
[53] van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. Nature Publishing Group UK. ChatGPT: five priorities for research, https://doi.org/10.1038/d41586-023-00288-7; 2023. [Accessed 18 August 2023].
[54] Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. Radiology 2023;307:e230163.