12-14-2020

Artificial Intelligence for the Orthopaedic Surgeon: An Overview of Potential Benefits, Limitations, and Clinical Applications

Eric C. Makhni
Henry Ford Health, emakhni1@hfhs.org

Sonya Makhni

Prem N. Ramkumar

Recommended Citation

Makhni EC, Makhni S, and Ramkumar PN. Artificial Intelligence for the Orthopaedic Surgeon: An Overview of Potential Benefits, Limitations, and Clinical Applications. J Am Acad Orthop Surg 2020.

Review Article

Artificial Intelligence for the Orthopaedic Surgeon: An Overview of Potential Benefits, Limitations, and Clinical Applications

Eric C. Makhni, MD, MBA
Sonya Makhni, MD, MBA
Prem N. Ramkumar, MD, MBA

ABSTRACT

Artificial intelligence (AI), along with its subset technology machine learning, has transformed numerous industries through newfound efficiencies and supportive decision-making. These technologies have similarly begun to find application within United States healthcare, particularly orthopaedics. Although these modalities have the potential to similarly transform health care, there exist limitations that must also be recognized and understood. Unfortunately, most clinicians do not have an understanding of the fundamentals of AI and therefore may have challenges in contextualizing its impact in modern healthcare. The purpose of this review was to provide an overview of the key concepts of AI and machine learning with the orthopaedic surgeon in mind. The review further highlights the potential benefits and limitations of AI, along with an overview of its applications, in orthopaedics.

Henry Ford Hospital (E. C. Makhni), Detroit, MI, Department of Medicine (S. Makhni), University of Chicago, Chicago, IL, and Orthopaedic Machine Learning Laboratory (Ramkumar), Cleveland Clinic, Cleveland, OH.

E. C. Makhni or an immediate family member is a member of a speakers' bureau or has made paid presentations on behalf of Xodus Medical; serves as a paid consultant to Smith & Nephew; and serves as a board member, owner, officer, or committee member of American Orthopaedic Society for Sports Medicine. Ramkumar or an immediate family member has stock or stock options held in Johnson & Johnson; has received research or institutional support from FocusMotion; and serves as a board member, owner, officer, or committee member of AAOS. Neither Dr. S. Makhni nor any immediate family member has received anything of value from or has stock or stock options held in a commercial company or institution related directly or indirectly to the subject of this article.

J Am Acad Orthop Surg 2020;00:1-9

DOI: 10.5435/JAAOS-D-20-00846

The application of artificial intelligence (AI) in the field of medicine has been widely forecasted since John McCarthy first coined the term over 60 years ago. In his proposal in 1955, McCarthy originally envisioned AI as "the science and engineering of making intelligent machines."[1] He predicted that these machines would one day be capable of doing feats previously thought to be exclusive to the domain of human intelligence, such as abstract thought, advanced problem-solving, and iterative self-improvement. In fact, in 1976, Jerrold S. Maxmen,[2] a professor of psychiatry at Columbia University, predicted that AI would bring about the "postphysician era" by the 21st century,[3] describing the change as "possible, inevitable, and desirable."

The true impact of AI on the future practice of medicine is still unknown. What is clear, however, is that AI has already begun to transform numerous industries in a variety of sectors. Some common examples include autonomous vehicles, online purchase recommendations, targeted advertising, and even forecasting of stock market fluctuations. However, the incorporation of AI in the healthcare sector has lagged behind that of other industries. Despite initial excitement over the possibilities of AI in the medical field, practical applications of AI have only recently begun to materialize.

Table 1. **Definitions of Relevant ML Terms and Concepts**

Term	Definition
Artificial intelligence (AI)	A broad term referring to the application of computational algorithms that can analyze large data sets (ie, "Big Data") to classify, predict, or gain useful inference
ML	A subset of AI that involves using real-world data sets to predict or estimate an outcome; these data sets encompass "training sets" that the machine is able to study and "learn" from using pattern recognition, which is then compared with a "test set" that quantifies the accuracies of the aforementioned inferences for further calibration
Deep learning	Sophisticated algorithms that require little to no human supervision to analyze, calibrate, and provide inferences; these include deep neural network models
Variance inflation factor (VIF)	A measure of multicollinearity in a regression analysis. A higher VIF indicates that predictors are highly correlated with each other, generally indicating a less reliable result.
Python StatsModel package	A Python module that provides resources for conducting statistical analysis in Python.
K nearest neighbors	A pattern recognition algorithm used for both classification and regression. This algorithm classifies a case based on the classification of most its neighbors.
Naïve Bayes	An algorithm that classifies cases based on the application of Bayes' theorem with the assumption of conditional independence.
XGBoost	A ML algorithm that utilizes a gradient boosting framework to solve prediction problems.
Top three ensemble	An ensemble algorithm that incorporates multiple machine learning algorithms (top three) to augment predictive performance.
Broyden-Fletcher-Goldfarb-Shanno optimizer	An iterative algorithm that allows for the solving of unconstrained optimization problems.
Brier score loss	A calculation of the mean squared error between predicted and expected values. A low Brier score indicates better predictions.
Area under the curve (AUC) of the receiver operating characteristics curve	An aggregate measure of a model's classification performance. AUC ranges in value from 0 to 1.0, with an AUC 1 meaning that a model is capable of distinguishing between classes 100% of the time.
SHapley Additive exPlanations (SHAP) scores	A measure of feature importance in predictive modelling. A higher SHAP value indicates a factor that predicts higher injury probability, whereas a lower SHAP value indicates a factor that predicts a lower injury probability

ML = machine learning

Within the field of orthopaedic surgery, AI holds promise for several cutting-edge applications that can transform the quality of care rendered, accelerate the delivery of services, and improve the value rendered. The technology also has important limitations and vulnerabilities that must be understood to maintain (and improve on) the current healthcare quality standards. The purpose of this review was to provide a summary of AI and highlight its potential applications and limitations within the context of clinical orthopaedics. In particular, the review will summarize the key-related technical aspects and subtopics of AI. This framework may be beneficial to practitioners who wish to understand the role of this important technology in the context of healthcare quality and operations.

Definitions

Broadly speaking, the term AI (Table 1) refers to the "mimicking of human cognition by computers."[4] One important subset of AI is that of machine learning (ML), which involves the use of computational algorithms that can analyze large data sets to classify, predict, or gain useful inference.[5,6] In its most rudimentary form, ML models are given inputs and outputs of "training sets" using real-world data to analyze and determine relationships using various methods of pattern recognition. The models are then tasked with creating predictions, given inputs from a "testing set," and these predictions are compared with actual known outcomes to quantify and refine the accuracy of the algorithm with positive or negative reinforcement. These algorithms are comparable to the same experiential "learning" associated with human intelligence, having the capacity to continually assess, and improve the quality of its analyses, given an adequate amount of data inputs, with the potential to continue learning after implementation because new data are available.[7–9] Thus, the predictive power and accuracy of an AI algorithm is only as powerful as its training experience and volume, not unlike the expertise and judgment of a surgeon. Moreover, these algorithms can be seen as doing similar tasks as traditional regression analyses, which determine relationships between disparate variables.

Deep learning can be thought of as an additional subset of ML. Made possible with increasingly powerful computational processing capabilities, deep learning models are more sophisticated algorithms that require less human supervision for development. Also known as deep neural networks, these models mimic the structure and function of the neuron by receiving several inputs (ie, dendrites) than determining which signal meets the internal threshold to pass forward along the axon to the next neuron. Unlike traditional ML algorithms, which generally require human expertise and the predetermined transformation of raw data into a suitable format, deep learning models are a form of representation learning. They function autonomously, allowing the system to discover alternative representations with differing levels of abstractions. The neural network begins with an input tier that receives the raw data. The network then progresses to a number of "hidden tiers" that each respond to different features of the input.[10] Similarly, "backpropagation" of the neural network exists, in which the model continues to learn through the refinement of weighting regarding the known training sets. Through this process of developing multiple hidden layers, the model continues to develop more and more abstract representations of the data. Similar to the way the human brain functions, the machine is able to make "neuronal" connections from "dendrites" on multiple hierarchical data levels.[8] Eventually, the model learns to appreciate a concept on multiple layers and dimensions, building on itself to create a web of interconnected relationships.[7] However, there is room for bias because these models rely on arbitrary weightings that must be manually assigned.

Role of Artificial Intelligence in United States Healthcare: Potential and Limitations

Modern healthcare is primed for positive transformation by AI.[5] Despite having the highest healthcare expenditure per capita among developed countries, the United States consistently ranks poorly in key quality metrics such as average life expectancy, maternal and infant mortalities, and health equity.[11,12] Innovation in the form of AI offers exciting potential in both improving healthcare outcomes and reducing inefficiencies that currently plague modern medicine. A second contributing factor is the recent generation of tremendous data, known as "Big Data." From high-resolution medical imaging, granular electronic medical records (EMRs) data, genome sequencing, and numerous diagnostic testing capabilities, each patient encounter generates tremendous Big Data that cannot be effectively analyzed with human processing or standard statistical methods. One study of EMRs found that a single patient's health record was associated with an average of approximately 32,000 data elements.[13] As another example, the human genome of a single individual requires 125 gigabytes of data storage.[5] In an age of information overload and evidence-based decision-making, the physician is tasked with the integrating this overwhelming amount of data and synthesizing a clinical decision, a seemingly impossible task, given limited time and context.[5] Judicious use of AI and its predictive abilities represent a solution to delivering high value care in the setting of an overwhelming degree of Big Data.

It is important to note the inherent limitations of these technologies. As with all data analysis, the quality of the output and conclusion is heavily dependent on the quality of the input data. Therefore, just as in clinical research efforts, application of ML algorithms to databases that are of low quality is unlikely to yield meaningful and accurate results. Examples of low-quality input data include data sets with large amounts of missing information, poorly organized data sets (which can introduce error upon attempted analysis), low-volume databases that are not powered enough to draw meaningful conclusions, and inaccurate but accessible databases. Additional opportunities for external validation of predictive models exist, such as through Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) standards of reporting.[14]

A second concern regarding data inputs lies in the relevance of the data. Many large databases in orthopaedics draw on administrative and claims-based data that can be susceptible to discrepancies 25% of the time.[15] This concern affects traditional clinical research and AI-driven research. Moreover, although claims based data are important, ample evidence exists that they do not represent the most relevant and meaningful outcomes to patients,[16] particularly regarding patient satisfaction.[17] Instead, they represent data that are relatively easy to extract and aggregate from EMRs. Data inputs such as patient-reported outcome measures (PROMs) and social determinants of health may be far more relevant in predicting clinical outcomes when compared with this claims-based data, which was not intended to be primary inputs for such ML algorithms.

Finally, despite the relatively autonomous nature of analysis through ML algorithms, a potential for bias still exists. This bias may be a result of the algorithm that is used to analyze the data or with the data itself (eg, skewed datasets). For example, in one study by Obermeyer et al,[18] the authors identified racial bias in the "ground truth," which led to the conclusion that Black patients are more medically complex and costly to the system than White patients, when in reality the model failed to account for access. Similarly, when Amazon (Seattle, Washington, United States) attempted to build an AI-based tool to aid in recruiting new talent, the algorithm negatively selected against women because the training data were primarily rooted in male-dominated applications to accrue data.[5]

These guidelines, however, are not comprehensive and do not encompass the importance of model validation. As technically challenging as it may be to build and train an accurately predictive model, clinical application requires external validation on an outside cohort, as was done by Ramkumar et al with primary hip and knee replacement patients using both institutional and National Inpatient Sample data.[19,20] Without external validation using institutional data, the model from the National Inpatient Sample database could have created many false predictions if applied too soon, given the inherent weaknesses of this administrative database. Thus, external validation with multiple data sets is important beyond simply reporting performance metrics.

A commonly critiqued limitation is the "black box" nature of AI-based algorithms, which intimates that the inner workings of the model's decision-making or "rationale" behind particular inferences will never be known. However, there exist several vehicles to determine the weight or importance of the data inputted into the algorithm. As one example, Shapely Additive Explanation summary aggregate Plots are a method to show the relative importance and direction of each modeling variable used to generate a prediction across a data set. For image processing, as with classifying implants from plain radiographs, heatmaps (Figure 1) can be developed to identify aspects of the image that trigger specific classification. These images are created through layering techniques that, when retroactively analyzed, help make the "black box" nature more transparent.

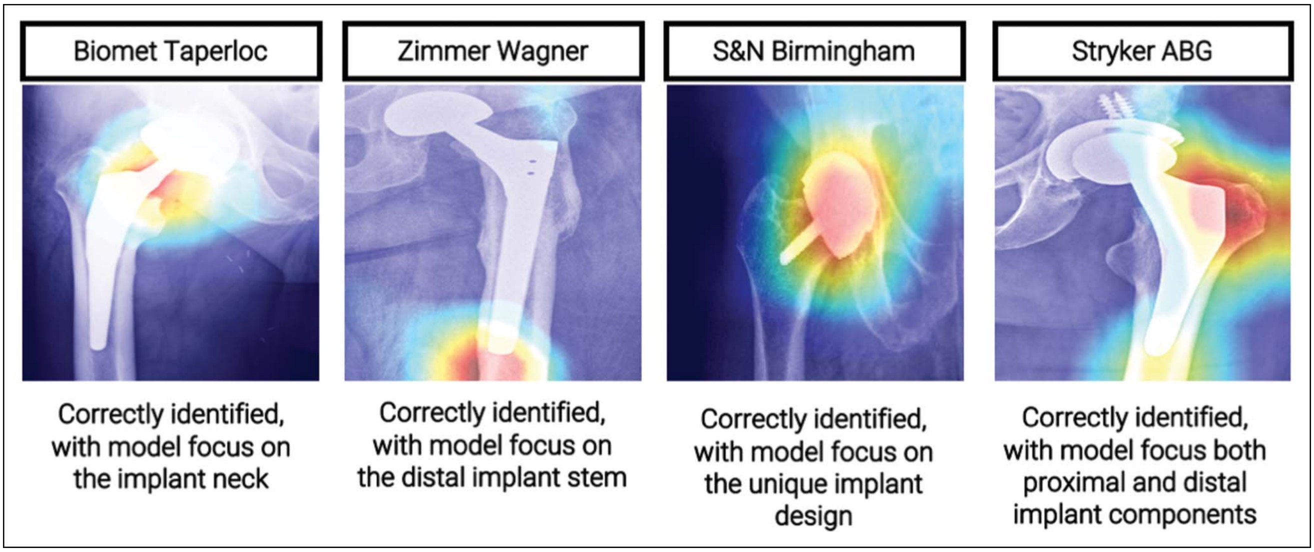
Applications in Orthopaedics

Remote Patient Monitoring

Remote patient monitoring systems (Table 2) represents an avenue that can increase the value of care during the perioperative period and has become increasingly important since the coronavirus disease 2019 (COVID-19) pandemic. Although many companies have developed software to monitor step counts and activity level, the application of ML with an open architecture system (eg, one that allows broad sharing and integration with other systems) allows patients and healthcare providers to track their participation in home exercise programs and general activity levels.[7,8] The surgical team can therefore track rehabilitation and intervene with calls or additional office visits if postoperative milestones are not being met.

Remote patient monitoring systems have been proven to be effective for patients undergoing primary total knee arthroplasty (TKA) for osteoarthritis. In one cohort study of 25 patients, patients who underwent this procedure, downloaded an AI-based, open architecture mobile application (FocusMotion) onto their personal iPhones (Apple), and recorded preoperative mobility and PROMs, beginning 2 to 4 weeks before surgery.[27] A knee sleeve was paired with the patient's iPhone via Bluetooth, and the application notified the patient to complete weekly exercises. Home exercise compliance and range of motion were detected by AI-based interpretation of the sensors on the knee sleeve that displayed

Figure 1



Heat map illustrates unique stem features from pixel processing and analysis that contributed to implant classification.

Table 2. Summary of Artificial Intelligence-based Studies in Orthopaedics

Genre/Author	Application
Remote patient monitoring: Ramkumar et al[20]	AI-driven wearable technology demonstrating benefit in monitoring of patient compliance with home exercise program, range of motion
Postoperative outcomes and cost: Karhade et al[21]	Machine learning algorithm for preoperatively predicting prolonged opioid prescription after total hip arthroplasty
Postoperative outcomes and cost: Navarro et al[22]	Postoperative detection of length of stay and cost according to factors such as age, race, sex, and medical comorbidity after total knee arthroplasty
Postoperative outcomes and cost: Ramkumar et al[19]	Neural network in predicting length of stay, charge, and disposition after total knee arthroplasty with application to value-based care delivery
Imaging and gait analysis: Urish et al[23]	Development of algorithms used for predicting the presence of osteoarthritis from radiograph images
Imaging and gait analysis: Kotti et al[24]	Correlation of body kinetics with likelihood of the presence of knee osteoarthritis
Imaging and gait analysis: Karnuta et al[25]	Classification of hip and knee arthroplasty implants from radiographs with manufacturer model
Implant design: Kozic et al[26]	Optimization of implant design according to implant shape and bone characteristics

range of motion and overall compliance with exercise form. This system was found to be reliable, low maintenance and well received during the process of recovery from TKA.[27]

Postoperative Outcomes and Cost

ML has been shown to be useful in using patient-specific factors to predict postoperative outcomes. This feature can be applied to further improve payment models by bringing greater, more nuanced specificity to tiered reimbursement. The Comprehensive Care for Joint Replacement model for bundled payments and quality measures was established to improve value and incentivize high-quality care at lower costs. However, hospitals that have demonstrated savings with bundled payments are more likely to be large, high volume, and associated with postacute care facilities.[28] Although bundling care has been shown to improve outcomes (readmissions decreased from 5% to 1.6%–2.7%, and patients are more likely to be discharged home), bundling care does not account for the specific factors each patient possesses.[29] Patient-level factors are essential in predicting the likely true cost and outcome of a procedure.[30,31] These specificities may shape or determine the course of their treatment. A single reimbursement fee for a single procedure may therefore fall short, failing to acknowledge or incorporate patient-level factors that influence cost. A comprehensive model that can identify patient-specific factors that influence cost may be able to help determine more appropriate reimbursements and reduce the phenomenon of "cherry-picking" or "lemon-dropping" patients.[32,33]

ML has also been applied to predict the necessity of prolonged opioid prescription after an operation. A 2019 study by Karhade et al[21] developed ML algorithms for preoperative prediction of prolonged opioid prescriptions after total hip arthroplasty. In addition, the algorithm's predictive power presents an opportunity to more accurately estimate the true cost of a procedure—a cost estimate that includes the likelihood of prolonged opioid prescription or dependence, in addition to the direct costs of the procedure.[21]

In this manner, ML technologies can be applied to determine a particular patients' likelihood of increased resource utilization—in this highlighted study, investigators predicted prescription utilization, but others have examined length of stay and inpatient charges. These specific outcomes help characterize a case's predicted complexity based on the patient's specific factors. Identifying and understanding the complexity of a case creates an opportunity to more accurately understand value. Although value has been understood as the relative benefit of the outcome to the cost, value is not standardized because individuals may require different resources, bring different goals, and achieve different outcomes.

Beyond simply identifying risk factors for increased cost, risk stratification for the purpose of improving cost may be a useful tool in improving equity. A 2018 study by Navarro et al[22] used a Bayesian model to forecast length of stay and cost, using factors such as age, race, sex, and comorbidity scores. A proposed risk-based patient-specific payment model was created based on the output. As patient complexity increased, cost add-ons then increased in tiers of 3%, 10%, and 15% for moderate, major, and extreme mortality risks. This proposition has the potential to encourage cost sharing, reduce patient selection, and even reinforce patient access by reimbursing in proportion to complexity. The ability to predict a specific patients' outcomes and resource utilization based on their preoperative variables has important implications in increasing the efficiency of payment models to improve cohort health. In the present era, risk is not distributed equally between payers and the treating team—surgeons are not incentivized to take on a large proportion of patients with increased comorbidities or case complexity. However, using ML to predict complexity offers an opportunity to fairly reward surgeons and institutions who take on greater risk. In the context of an aging cohort with increased comorbidities, a flat bundle reimbursement fee for patients with varying risk fails to match value with reimbursement.

Similarly, a 2019 study by Ramkumar et al[20] developed and validated an artificial neural network that was able use patient-specific factors and outcomes to "learn" and predict length of stay, inpatient charge, and discharge disposition in unfamiliar patients undergoing TKA. Furthermore, this predictive model was applied to propose a risk-based, patient-specific payment model. The neural network was created using 175,042 total knee arthroplasties and had an area under the curve of 0.748 for length of stay, 0.828 for charges, and 0.761 for discharge disposition. The model "learns" iteratively from training groups until it is able to predict value-based patient outcomes. This predictive capability has promise in application to patient-specific payment models and tiering reimbursement based on case complexity, in which patients may be preoperatively assigned to a tier based on their risk factors, with a reimbursement commensurate with their stratified risk.

With the advancement of data aggregation and deep learning algorithms, the field or orthopaedics is on the cusp of a transformation. The adoption of ML in orthopaedics has the power to improve patient care by estimating complexity of cases and supporting progress toward patient-specific payment models that are more capable of incorporating specificities of each case.[8]

Imaging and Gait Analysis

ML has important applications in diagnosis, using both imaging and gait analysis. It has been used to automatically detect osteoarthritis using imaging patterns and movement patterns, a feature that holds promise for efficient and objective automated diagnosis.[23]

For example, preprogrammed mathematical algorithms and measurements have been shown to accurately diagnose arthritis on a radiograph. Urish and Reznik[23] describe the use of medical imaging data in a technique that analyzes pixels in a radiograph image to recognize pertinent structures and specific features to create a pattern. When presented with an unknown image, the algorithm was shown to "decide" whether it was consistent with a known model for osteoarthritis or whether it did not match. This algorithm may be used in both clinical applications and research applications to confirm the presence of osteoarthritis. Furthermore, it could be expanded to predict which patients have more advanced pathology or would benefit most from surgical intervention. With an algorithm capable of processing images, a health system may be able to more efficiently triage a patient to the appropriate care provider—whether it be a specialist arthroplasty surgeon, sports medicine surgeon, or a nonsurgical physician. These clinical decisions can be made using data rather than reliance on nonclinical schedulers. In addition, they provide the benefit of increasing efficiency.

ML has utility in detecting knee osteoarthritis using gait analysis. A computer system developed by Kotti et al[24] took input body kinetics and produced as output an estimate of the likelihood of the presence of knee osteoarthritis. Furthermore, it identifies the discriminating parameters and set of rules that led to the decision. This explanation mimics "interpretation" and increases the value of the diagnosis. With an accuracy of 72.6%, this automatic detection of knee osteoarthritis provides a unique opportunity to create objective, sensitive diagnostic tools that can increase efficiency and quality of care delivered to patients.

Recently, Karnuta et al[34] trained, validated, and externally tested a deep-learning system to classify total hip arthroplasty and hip resurfacing arthroplasty femoral implants as one of 18 different manufacturer models from 1,972 retrospectively collected AP plain radiographs from four sites in one quaternary referral health system (Figure 1). After 1,000 training epochs by the deep-learning system, the system discriminated 18 implant models with an area under the curve of 0.999, accuracy of 99.6%, sensitivity of 94.3%, and specificity of 99.8% in the external-testing data set of 206 AP radiographs. Similarly, the same group[25] built a deep-learning system to identify TKA, unicompartmental knee arthroplasty, and distal femoral replacement images and found the model discriminated nine implant models with an area under the curve of 0.99, accuracy 99%, sensitivity of 95%, and specificity of 99% in the external-testing data set of 74 radiographs.

Implant Design

Optimization of implants and devices can increase the value that they provide both to the patient and the value of investments made by developers. Currently, implant design is not as efficient as possible because of constraints in testing fit.

Kozic et al[26] present a method to assess specific anatomical and morphological criteria that transcend shape variability in a cohort to optimize orthopaedic implant design. Although implants are mostly designed and tested through fitting on cadaver bones, which provide only a limited sample that does not represent the entire variability of the patient cohort, this technology provides an alternative. The framework allows an implant design to be virtually fit to samples drawn from a statistical model, determining which range of the cohort is most appropriate for a particular implant. Certain patterns of bone variability are more important for implant fitting, and this method allows for improvement of implant design such that a maximum target cohort can have a benefit.

This study demonstrated the optimization of implant design, using their proposed design and virtual validation method, of a proximal human tibia used for internal fracture fixation. Overall, implant design can benefit from these methods to improve fit for the patient, designer, and physician.

Summary

The day-to-day employment of ML in an independent orthopaedic practice is not yet widely used. However, individual orthopaedic services across the United States have pioneered the use of ML technology in recent years and may serve as examples of what future deployment will look like. In 2019, Goltz et al,[35] in affiliation with Duke University Medical Center, released a 90-day readmission risk calculator after primary unilateral total hip and knee arthroplasties. Using patient data from 10,155 primary unilateral total hip and knee arthroplasties done at a single institution, a multivariable regression model was created to "adequately predict" the likelihood of 90-day readmission, based on preoperative parameters, duration of surgery, postoperative laboratory results (hemoglobin and blood-urea-nitrogen level), and nine comorbidities. This tool is freely available online for any provider to use and serves as an example of how applied statistical techniques, in conjunction with large amounts of available patient data, can be harnessed to provide benefit in patient care.

Although many may not consider such regression models as "ML", it may be useful to consider ML as a natural extension of statistical techniques that have been used in the profession of healthcare for decades. Other online applications are similarly available, ranging from predicting risk of increased length-of-stay after joint arthroplasty to predicting inpatient payments.[21,36,37] It is likely that the use of such applications and online tools, powered by ML algorithms, will become more ubiquitous in the future and paid private consulting services.

In 2018, the Cleveland Clinic's Department of Orthopaedic Surgery established the ML Arthroplasty Laboratory, with the goal of exploring practical implementation of ML techniques in the practice of orthopaedic medicine.[38] The team has developed and validated several ML models in several areas of research interest, all with the ultimate focus of providing patient-specific, value-based care.[20,27,38,39] The team has recently developed an image classifier to read preoperative radiographs and identify arthroplasty implant class and manufacturer before revision. Such a tool would be valuable to any arthroplasty surgeon, avoiding the increase in costs associated with delays in care, and misidentifications leading to lack of appropriate equipment available during the operation.[8] Another area of study for the group has been the establishment of value-based payment models for hip and knee total joint arthroplasty.[22,39] Although the development of these theoretical payment models may not serve any utility for any single orthopaedic practice, these studies represent an initial foray exploring the utility of ML in better informing reimbursement.

Yet a third example of incorporation of AI and ML to routine orthopaedic care lies in with decision-support tools using PROM data to predict outcomes after hip and knee arthroplasties. Using AI, clinicians at the UT Health Austin Musculoskeletal Institute, Dell Medical School at the University of Texas at Austin discuss likelihood of post-operative success before scheduling surgery with patients.[40] These models have the potential to improve in accuracy as more data/inputs are incorporated into the models.

Although the true impact of AI and ML on clinical orthopaedics is still yet to be determined, ample evidence exists that these technologies may assist in generating healthcare value through improving outcomes or decreasing cost/inefficiencies. ML has the capability of automating redundant tasks, thereby allowing physicians to spend more time with patients. The technology should be viewed as a physician-aid—a tool that can better augment a physician's capabilities rather than replace their responsibilities. To maximize the benefit of these tools, however, clinicians, researchers, and policy makers must first understand the fundamentals of the technology, along with its potential benefits and limitations. Numerous applications in orthopaedics have already been demonstrated, and these applications will increase in quantity and impact as AI continues to grow as a key healthcare technology.

References

References printed in **bold type** are those published within the past 5 years.

1. McCarthy J, Minsky ML, Shannon CE: A proposal for the Dartmouth summer research project on artificial intelligence—August 31, 1955. *AI Mag* 2006;27:12-14.

2. Maxmen JS: Long-term trends in health care: The post-physician era reconsidered, in Schwefel D ed: *Indicators and Trends in Health and Health Care*. Berlin, Heidelberg, Springer Berlin Heidelberg, 1987, pp 109-115.

3. Maxmen JS: The post-physician era: Medicine in the 21st century. *JAMA* 1976;237:2336-2337.

4. **Jha S, Topol EJ: Adapting to artificial intelligence: Radiologists and pathologists as information specialists. *JAMA* 2016;316:2353-2354.**

5. **Topol EJ: High-performance medicine: The convergence of human and artificial intelligence. *Nat Med* 2019;25:44-56.**

6. **Naylor CD: On the prospects for a (deep) learning health care system. *JAMA* 2018;320:1099-1100.**

7. **Haeberle HS, Helm JM, Navarro SM, et al: Artificial intelligence and machine learning in lower extremity arthroplasty: A review. *J Arthroplasty* 2019;34:2201-2203.**

8. **Helm JM, Swiergosz AM, Haeberle HS, et al: Machine learning and artificial intelligence: Definitions, applications, and future directions. *Curr Rev Musculoskelet Med* 2020;13:69-76.**

9. **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;102:830-840.**

10. **Mintz Y, Brodie R: Introduction to artificial intelligence in medicine. *Minim Invasive Ther Allied Technol* 2019;28:73-81.**

11. **Papanicolas I, Woskie LR, Jha AK: Health care spending in the United States and other high-income countries. *JAMA* 2018;319:1024-1039.**

12. **MacDorman MF, Declercq E, Cabral H, Morton C: Recent increases in the U.S. maternal mortality rate: Disentangling trends from measurement issues. *Obstet Gynecol* 2016;128:447-455.**

13. **Milinovich A, Kattan MW: Extracting and utilizing electronic health data from epic for research. *Ann Transl Med* 2018;6:42.**

14. Available at: www.tripod-statement.org. Accessed October 8, 2020.

15. Saucedo J, Marecek GS, Lee J, Huminiak L, Stulberg SD, Puri L: How accurately are we coding readmission diagnoses after total joint arthroplasty? *J Arthroplasty* 2013;28:1076-1079.

16. Wu AW, Kharrazi H, Boulware LE, Snyder CF: Measure once, cut twice—adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. *J Clin Epidemiol* 2013;66:S12-S20.

17. **Sivaganesan A, Khan I, Pennings JS, et al: Why are patients dissatisfied after spine surgery when improvements in disability and pain are clinically meaningful? *Spine J* 2020;20:1535-1543.**

18. Obermeyer Z, Powers B, Vogeli C, Mullainathan S: Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447-453.

19. Ramkumar PN, Karnuta JM, Navarro SM, et al: Preoperative prediction of value metrics and a patient-specific payment model for primary total hip arthroplasty: Development and validation of a deep learning model. *J Arthroplasty* 2019;34: 2228-2234 e2221.

20. Ramkumar PN, Karnuta JM, Navarro SM, 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;34:2220-2227 e2221.

21. 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 e2271.

22. Navarro SM, Wang EY, Haeberle HS, et al: Machine learning and primary total knee arthroplasty: Patient forecasting for a patient-specific payment model. *J Arthroplasty* 2018;33:3617-3623.

23. Urish K, Reznik AM: How would a computer diagnose arthritis on a radiograph? *AAOS Now* 2018:32-33.

24. Kotti M, Duffell LD, Faisal AA, McGregor AH: Detecting knee osteoarthritis and its discriminating parameters using random forests. *Med Eng Phys* 2017;43:19-29.

25. Karnuta JMLB, Haeberle HS, Roth AL, et al: Automated detection of total knee arthroplasty, unicompartmental knee arthroplasty, and distal femoral replacement implants from plain radiographs: A deep learning application. *J Arthroplasty* In press.

26. Kozic N, Weber S, Büchler P, et al.: Optimisation of orthopaedic implant design using statistical shape space analysis based on level sets. *Med Image Anal* 2010;14:265-275.

27. Ramkumar PN, Haeberle HS, Ramanathan D, et al: Remote patient monitoring using mobile health for total knee arthroplasty: Validation of a wearable and machine learning-based surveillance platform. *J Arthroplasty* 2019;34:2253-2259.

28. Navathe AS, Liao JM, Shah Y, et al: Characteristics of hospitals earning savings in the first year of mandatory bundled payment for hip and knee surgery. *JAMA* 2018;319:930-932.

29. Mouille B, Higuera C, Woicehovich L, Deadwiler M: How to succeed in bundled payments for total joint replacement. *NEJM Catal* 2016;10:930-932.

30. Rondon AJ, Tan TL, Greenky MR, et al: Who goes to inpatient rehabilitation or skilled nursing facilities unexpectedly following total knee arthroplasty? *J Arthroplasty* 2018;33:1348-1351 e1341.

31. Courtney PM, Bohl DD, Lau EC, Ong KL, Jacobs JJ, Della Valle CJ: Risk adjustment is necessary in medicare bundled payment models for total hip and knee arthroplasty. *J Arthroplasty* 2018;33:2368-2375.

32. Clement RC, Derman PB, Kheir MM, et al: Risk adjustment for medicare total knee arthroplasty bundled payments. *Orthopedics* 2016;39:e911-e916.

33. Humbyrd CJ: The ethics of bundled payments in total joint replacement: "Cherry picking" and "lemon dropping." *J Clin Ethics* 2018;28:62-68.

34. Karnuta JMHH, Luu BC, Roth AL, et al: Artificial intelligence to identify arthroplasty implants from radiographs of the hip. *J Arthroplasty* [published online ahead of print November 16, 2020] doi: 10.1016/j.arth.2020.11.015.

35. Goltz DE, Ryan SP, Hopkins TJ, et al: A novel risk calculator predicts 90-day readmission following total joint arthroplasty. *J Bone Joint Surg Am* 2019;101:547-556.

36. Manning DW, Edelstein AI, Alvi HM: Risk prediction tools for hip and knee arthroplasty. *J Am Acad Orthop Surg* 2016;24:19-27.

37. Karnuta JM, Golubovsky JL, Haeberle HS, et al: Can a machine learning model accurately predict patient resource utilization following lumbar spinal fusion? *Spine J* 2020;20:329-336.

38. Ramkumar PN, Haeberle HS, Bloomfield MR, et al: Artificial intelligence and arthroplasty at a single institution: Real-world applications of machine learning to bg data, value-based care, mobile health, and remote patient monitoring. *J Arthroplasty* 2019;34:2204-2209.

39. Ramkumar PN, Navarro SM, Haeberle HS, et al: Development and validation of a machine learning algorithm after primary total hip arthroplasty: Applications to length of stay and payment models. *J Arthroplasty* 2019;34:632-637.

40. Jayakumar P, Bozic KJ: Advanced decision-making using patient-reported outcome measures in total joint replacement. *J Orthop Res* 2020;38:1414-1422.