Level V Evidence

Clinical and Research Medical Applications of Artificial Intelligence

Prem N. Ramkumar, M.D., M.B.A., Kyle N. Kunze, M.D., Heather S. Haeberle, M.D., Jaret M. Karnuta, M.S., Bryan C. Luu, B.S., Benedict U. Nwachukwu, M.D., M.B.A., and Riley J. Williams, M.D.

Abstract: Artificial intelligence (AI), including machine learning (ML), has transformed numerous industries through newfound efficiencies and supportive decision-making. With the exponential growth of computing power and large datasets, AI has transitioned from theory to reality in teaching machines to automate tasks without human supervision. AI-based computational algorithms analyze "training sets" using pattern recognition and learning from inputted data to classify and predict outputs that otherwise could not be effectively analyzed with human processing or standard statistical methods. Though widespread understanding of the fundamental principles and adoption of applications have yet to be achieved, recent applications and research efforts implementing AI have demonstrated great promise in predicting future injury risk, interpreting advanced imaging, evaluating patient-reported outcomes, reporting value-based metrics, and augmenting telehealth. With appreciation, caution, and experience applying AI, the potential to automate tasks and improve data-driven insights may be realized to fundamentally improve patient care. The purpose of this review is to discuss the pearls, pitfalls, and applications associated with AI.

The application of artificial intelligence (AI) in the field of medicine has been widely forecasted since the concept was first described by John McCarthy over 60 years ago.[1] Although the maturity of AI in the field of orthopaedics has lagged behind fields such as ophthalmology,[2] dermatology,[3] and cardiology,[4] interest has grown rapidly in the past 2 decades as techniques have become accessible to researchers and clinicians.

Broadly, AI is the science and engineering of creating intelligent machines that can achieve tasks that otherwise require human input.[5] Machine learning (ML) is a subset of AI that uses computational algorithms to analyze large data sets to classify and predict without explicit instructions.[5,6] In its most rudimentary form, ML models are given inputs and outputs of "training sets" using real-world data to determine relationships using pattern recognition.[6] As such, the model is dependent on the accuracy and biases of the given data set. The models are then tasked with creating predictions based on inputs from a "testing set," and these predictions are compared with actual known outcomes. As the data in the training sets grows and the number of testing repetitions increases, the machine's algorithm becomes more accurate and predictive, not unlike "experiential learning" in arthroscopy training. Thus, algorithms possess the capacity to "reflect" by continually assessing and improving the quality of its analyses, with the potential to continue incremental learning after addition of new data so as to not "reinvent the wheel," thereby permitting global data

From the Orthopaedic Machine Learning Lab (P.N.R., H.S.H., J.M.K., B.C.L.), Cleveland Clinic, Cleveland, Ohio, U.S.A.; Department of Orthopaedic Surgery, Brigham & Women's Hospital (P.N.R.), Boston, Massachusetts, U.S.A.; Department of Orthopaedic Surgery, Hospital for Special Surgery (K.N.K., H.S.H., B.U.N., R.J.W.), New York, New York, U.S.A.; and Department of Orthopaedic Surgery, Baylor College of Medicine (B.C.L.), Houston, Texas, U.S.A.

The authors report the following potential conflicts of interest or sources of funding: B.U.N. reports grants, personal fees, nonfinancial support, and other from Arthrex and Smith & Nephew, nonfinancial support from Stryker, personal fees from Medwest Associates and Medical Device Business Services, and a pending patent. P.N.R. reports a pending patent and serving as an advisory board member at Focus Motion. R.J.W. reports personal fees, nonfinancial support, and other from Arthrex and a pending patent. Full ICMJE author disclosure forms are available for this article online, as supplementary material.

Investigation performed at the Cleveland Clinic, Cleveland, Ohio.

Received August 4, 2020; accepted August 12, 2020.

Address correspondence to Prem N Ramkumar, M.D., M.B.A., 9500 Euclid Ave, Cleveland Clinic, Cleveland, OH 44118, U.S.A. E-mail: premramkumar@gmail.com

© 2020 by the Arthroscopy Association of North America

0749-8063/201366/$36.00

https://doi.org/10.1016/j.arthro.2020.08.009

sharing and interconnectivity.[5-7] Understanding appropriate use for AI models is required prior for application to clinical scenarios; further discussion of the various models is outside the scope of this review.

Three drivers have accelerated AI implementation in health care: (1) accessible computing power, (2) spiraling systemic costs, and (3) omnipresence of data. Although computing power and systemic costs reflect trends from their respective technology and health care sectors over the past few decades, the generation of digital data is unprecedented. From high-resolution medical imaging, continuously evolving electronic health records, and numerous diagnostic tests, each patient encounter produces tremendous discrete data points, generating Big Data that cannot be effectively analyzed with human processing or standard statistical methods. One study of electronic health records found that a single patient's health record was associated with an average of approximately 32,000 unique data elements.[8,9] With the fundamental understanding that AI should and could never replace the patient-facing tasks of the sports medicine surgeon, the most appropriate application involves removing time-intensive, administrative burdens that drive nearly 40% of health care costs.[6] Supportive AI tools may automate the redundant tasks from care coordination to routine documentation and orders in the electronic health record, in turn, allowing for increased focus on patient-facing activities.

Limitations of AI

AI carries limitations. ML techniques create a "black box" phenomenon, in which the user can only access the inputs and outputs of an algorithm, but not the inner workings or the specific relationships evaluated by the algorithm.[10] Some believe the "black box" phenomenon risks deskilling of physicians and other providers, the displacement of physician jobs, and the devaluation of human experience and clinical intuition. As an example for the sports surgeon, we may be able to produce a suitable algorithm that decides which biologic agents to use, but we will may not gain the reasoning behind the algorithm's prediction. However, having supportive decision-making derived from the global body of evidence would allow for improved insights from improve data sharing.

As with all data analysis, the quality of the output and conclusion is heavily dependent on the quality and relevance of input data. Therefore, just as with any other clinical research effort, application of algorithms to databases of low quality and relevance are unlikely to yield meaningful and accurate results. Examples of low-quality input data include datasets with large amounts of missing information, low-volume databases that are not powered enough to draw meaningful conclusions, and inaccurate databases. Therefore, ML efforts must draw upon outcomes that are clinically accurate and relevant to patients to be meaningful.

Despite the relatively autonomous nature of analysis through machine learning algorithms, there is still a potential for bias. This bias may be a result of the algorithm that is used to analyze the data or with the data itself (e.g., skewed datasets). For example, when Amazon (Seattle, WA, U.S.A.) attempted to build an AI-based tool to aid in recruiting new talent, the algorithm negatively selected against females because the training data consisted of male-dominated applications.[11] Moreover, although this bias may be unintended, it may be difficult to recognize because of the nature of the unsupervised learning techniques inherent to these algorithms.

Applications: AI in the Sports Medicine Literature

In this section, we review the use of AI-based techniques in sports medicine as a prelude for what the sports medicine surgeon may anticipate reading in future literature.

Athlete Injury Prediction

Professional sports represent a multibillion-dollar industry that depends on the maintenance of player health through coordinated efforts with the goal of optimizing player performance and availability. With the breadth of metrics surrounding professional sports, ML may hold value in injury prevention and prediction. Luu et al.[12] compiled publicly reported National Hockey League injury data, player-specific metrics, 85 different performance metrics, and injury history to demonstrate that the best machine learning algorithm predicted next season injury with an accuracy of 94.6% (standard deviation 0.5%). For Major League Baseball players, Karnuta et al.[13] evaluated data 1931 position players and 1245 pitchers and found that the best performing algorithm demonstrated an accuracy of 70% (standard deviation 2%) at predicting next season injury. In both studies, ML techniques were superior to logistic regression at predicting future player injury.

Imaging

The powerful pattern recognition capabilities of AI naturally lend to the automated interpretation of imaging. When presented with an unknown image, an algorithm can interpret the imaging to provide a supportive decision based on the query. With advanced computed tomography or magnetic resonance imaging (MRI) data, acute cartilage or ligamentous pathology could be immediately detected for triage to a sports specialist. Ramkumar et al. applied an ML model to discern, for 1735 patients undergoing arthroscopic correction of femoroacetabular impingement syndrome (FAIS), which preoperative radiographic indices from computed tomography scans of the hip predicted significant changes in 1- and 2-year patient-reported outcome measures (PROMs).[14] The study found that no specific radiographic index or combination of indices was found to be predictive of improvement in any of the 4 PROMs at either 1 or 2 years follow-up in the setting of strict surgical indications. Similarly, Fritz et al. applied deep learning to 100 MRIs and found the algorithm detected meniscus tears with similar specificity but lower sensitivity than musculoskeletal radiologists.[15] In a study of 260 patients with knee MRIs, Chang et al. reported anterior cruciate ligament tears were detected with 96% accuracy using a deep learning architecture.[16]

Patient-Reported Outcome Measures

PROMs have become increasingly valuable quality metrics in determining the success of an intervention. Nwachukwu et al. investigated the application of ML to predict changes in PROMs after arthroscopic FAIS surgery.[17] An ML model was built using the least absolute shrinkage and selection operator algorithm for feature selection, followed by logistic regression for the selected features. The model, trained on 898 FAIS patients, was able to identify across 3 separate hip-specific PROMs the following predictors for failure to achieve clinically meaningful outcomes: presence of anxiety/depression, symptom duration > 2 years, preoperative intra-articular injection, and high preoperative outcome scores.[17]

Value-Based Metrics

Another relevant application of AI is the promotion of value-conscious care. One key issue is the current inability to preoperatively communicate value of care rendered to a specific patient for elective surgery. Karnuta et al. assessed the capability of artificial neural networks to predict length of stay, discharge disposition, and inpatient charges for primary anatomic, reverse, and hemi-shoulder arthroplasty.[18] This model predicted inpatient costs with an accuracy ranging from 69% to 77%, as well as discharge disposition and length of stay with fair to good accuracy (72%-75% and 78%-92%, respectively). Future ML models may provide physicians with the ability to offer an evidence-based, patient-specific tool that preoperatively communicates value metrics for valuable discourse in terms of expectation management and reimbursement arbitration from payor preauthorization.

Telehealth

In the era of the COVID-19 pandemic, the patient experience is increasingly tied to access via remote patient monitoring and telemedicine. One particular system (FocusMotion, Santa Monica, CA, U.S.A.) implements AI to remotely monitor patients recovering from knee arthroplasty, arthroscopy, and anterior cruciate ligament reconstruction through the use of Bluetooth-enabled braces and mobile health data that relay data to an AI-based algorithm.[19,20] These data are instantaneously contextualized to highlight warning signs, as stipulated by the surgeon, and display mobility, range of motion, PROMs, opioid consumption, wound appearance, and rehabilitation compliance in a central dashboard shared with the patient and care team; this was found to increase patient engagement to rehabilitate postoperatively after knee surgery and remains under current prospective evaluation.[19,20]

Conclusion

AI is poised to transform medicine through automated task performance and has demonstrated potential in predicting future injury risk, interpreting imaging, evaluating patient-reported outcomes, reporting value-based metrics, and augmenting telemedicine visits. This technology should be viewed as a tool to augment the capabilities of physicians and researchers, rather than replace their responsibilities. Though unfamiliar and complex, we advocate for embracing the use of AI given its many potential applications as the future of medicine necessitates physicians and caregivers gain sufficient familiarity with AI-based concepts for responsible application.

References

1. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence. *AI Mag* 1955;27:12-14.
2. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316:2402-2410.
3. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-118.
4. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019;25:65-69.
5. Helm JM, Swiergosz AM, Haeberle HS, et al. Machine learning and artificial intelligence: Definitions, applications, and future directions. *Curr Rev Musculoskelet Med* 2020:69-76.
6. 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.
7. Myers TG, Ramkumar PN, Ricciardi BF, Urish KL, Kipper J, Ketonis C. Artificial intelligence and orthopaedics. *J Bone Joint Surg Am* 2020 Feb 6. Published online ahead of print.
8. Milinovich A, Kattan MW. Extracting and utilizing electronic health data from Epic for research. *Ann Transl Med* 2018;6:5.
9. Wyatt JC. Fifty million people use computerised self triage. *BMJ* 2015:351.
10. The "black box" metaphor in machine learning - Towards Data Science, https://towardsdatascience.com/the-black-box

-metaphor-in-machine-learning-4e57a3a1d2b0. Accessed June 28, 2020.
11. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44-56.
12. Luu B, Wright A, Haeberle H, et al. Machine learning outperforms logistic regression analysis to predict next season NHL player injury: An analysis of 2,322 players from 2007-2017. *Orthop J Sport Med*. (in press.)
13. Karnuta J, Luu B, Haeberle H, et al. Machine learning outperforms regression analysis to predict next season MLB player injury: Epidemiology and validation of 13,982 player-years from performance and injury profile trends between 2000-17. *Orthop J Sport Med*. (in press.)
14. Ramkumar PN, Karnuta JM, Haeberle HS, et al. Radiographic indices are not predictive of clinical outcome among 1,735 patients indicated for hip arthroscopy: A machine learning analysis. *Am J Sports Med* 2020 (in press.).
15. Fritz B, Marbach G, Civardi F, Fucentese SF, Pfirrmann CWA. Deep convolutional neural network-based detection of meniscus tears: Comparison with radiologists and surgery as standard of reference. *Skeletal Radiol* 2020;49:1207-1217.
16. Chang PD, Wong TT, Rasiej MJ. Deep learning for detection of complete anterior cruciate ligament tear. *J Digit Imaging* 2019;32:980-986.
17. Nwachukwu BU, Beck EC, Lee EK, et al. Application of machine learning for predicting clinically meaningful outcome after arthroscopic femoroacetabular impingement surgery. *Am J Sports Med* 2020;48:415-423.
18. Karnuta JM, Churchill JL, Haeberle HS, et al. The value of artificial neural networks for predicting length of stay, discharge disposition and inpatient costs after anatomic and reverse shoulder arthroplasty. J Shoulder Elb Surg *June* 2020;9. Published online ahead of print.
19. 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.
20. Ramkumar PN, Haeberle HS, Bloomfield MR, et al. Artificial intelligence and arthroplasty at a single institution: Real-world applications of machine learning to big data, value-based care, mobile health, and remote patient monitoring. *J Arthroplasty* 2019;34(10):2204-2209.