



■ ANNOTATION

Potential benefits, unintended consequences, and future roles of artificial intelligence in orthopaedic surgery research

A CALL TO EMPHASIZE DATA QUALITY AND INDICATIONS

**K. N. Kunze,
M. Orr,
V. Krebs,
M. Bhandari,
N. S. PiuZZi**

From Cleveland Clinic
Foundation, Cleveland,
Ohio, USA

Artificial intelligence and machine-learning analytics have gained extensive popularity in recent years due to their clinically relevant applications. A wide range of proof-of-concept studies have demonstrated the ability of these analyses to personalize risk prediction, detect implant specifics from imaging, and monitor and assess patient movement and recovery. Though these applications are exciting and could potentially influence practice, it is imperative to understand when these analyses are indicated and where the data are derived from, prior to investing resources and confidence into the results and conclusions. In this article, we review the current benefits and potential limitations of machine-learning for the orthopaedic surgeon with a specific emphasis on data quality.

Cite this article: *Bone Jt Open* 2022;3-1:93–97.

Keywords: Artificial intelligence, Machine-learning, Data management, Predictive modelling, Orthopedics

The race to be first

The application of artificial intelligence (AI) and machine-learning in orthopaedics is a concept that has elicited an aura of complexity, mystery, and excitement in recent years among researchers and clinicians. However, machine-learning and AI are rather well-established methodological techniques. Indeed, these have formed an extensively applied and well-accepted set of methodological processes among many industries that depend on data science for their success, and AI has been applied in other realms of medicine.¹ AI is a heterogeneous term, but may be defined as technologies or machines capable of performing tasks like problem-solving and learning, language interpretation, pattern recognition, and planning, similar to human cognitive function.² Machine-learning is a subset of AI that uses experiential learning through incremental adjustments in internal parameters to objectively strengthen or weaken select features (inputs) through mathematical functions to optimize model prediction

accuracy.² Orthopaedic research had simply not caught wind of the potential applications of AI and machine-learning, though recent years have seen it quick to adopt and laud these methods proposed to be superior. However, the adoption of these techniques and their potential for clinical impact should not proceed without understanding the limitations of these methods, and their potential to misinform readers and healthcare professionals. Therefore, it is imperative to understand the importance of data quality, appropriate reporting, and proper application of these models as it pertains to machine-learning in orthopaedic research.

Who has the data has the power

In 2021, Tim O'Reilly's maxim, "who has the data has the power" assumes a new meaning, and comes down to man versus the machine. Medical data were estimated to double every 73 days in 2020, and with the costs and resources used to collect so much patient data, it is imperative to use them in meaningful ways to benefit healthcare

Correspondence should be sent to
Nicolas S. PiuZZi; email:
nspiuzzi@gmail.com

doi: 10.1302/2633-1462.31.BJO-
2021-0123.R1

Bone Jt Open 2022;3-1:93–97.

Table 1. Recent applications and rise of machine-learning in arthroplasty literature since 2015.

Year	Studies, n	Study topic(s)
2015	1	Lower limb muscle activation patterns after TKA
2016	1	Gait analysis after TKA/UKA
2017	3	Classification of revision TKA cause, effect of femoral stem morphology on stress shielding, prediction of opioid use after THA
2018	5	Cost use after THA and TKA, patient activity monitoring after TKA, image-based rating of corrosion severity for THA implants, readmissions after TJR
2019	32	Clinical outcome prediction (adverse events and patient-reported outcomes), resource use and cost of episodes of care, patient activity monitoring (wearable sensors, gait analysis), automatic chart review using natural language processing, implant identification*
2020	56	Clinical outcome prediction (adverse events and patient-reported outcomes), resource use and cost of episodes of care, patient activity monitoring (wearable sensors, gait analysis), automatic chart review using natural language processing, implant identification*
2021 to date	45	Clinical outcome prediction (adverse events and patient-reported outcomes), resource use and cost of episodes of care, patient activity monitoring (wearable sensors, gait analysis), automatic chart review using natural language processing, implant identification*

*Applications categorized into four major domains for brevity.

THA, total hip arthroplasty; TJR, total joint replacement; TKA, total knee arthroplasty; UKA, unicompartmental knee arthroplasty.

centres and patients. For example, by applying AI to image recognition, it may be possible to save time and resources by allowing for the development of preoperative risk assessment or surveillance tools.³ It is undeniable that big data will transform medicine, and though high-quality data is imperative, ultimately all data must be appropriately analyzed, interpreted appropriately, and allow for change in the way we practice medicine. To this end, there currently exists much “hype” as to the potential of machine-learning; however, this currently remains disproportionate to the translation of such models in clinical practice.

What are the benefits?

Machine-learning models are critical to the advancement of research in orthopaedic surgery, given their ability to analyze and interpret vast numbers of predictor variables in non-linear and highly complex manners, “learn,” and make accurate predictions.² Furthermore, incremental inputs can improve the predictive ability of these models as they continue to learn from new data.⁴ The manner in which a model makes these predictions was appropriately named after our own brains: artificial neural networks. Similarly to how we as humans unconsciously recognize patterns and use these to inform decision-making, while sometimes not being able to articulate how we came to our conclusions (trusting our ‘gut instinct’), computer algorithms exist in a “black box.” While we can literally open this black box, view the code, and reapply the model to more data in a more consistent application than a human is capable of, the mapping of input to output is complex, and involves numerous connections between components, thus rendering it indescribable. While publications can cite how accurately their models predict outcomes in their dataset, these models often exist only in the paper and are not extended across institutions or implemented in practice.

However, machine-learning may serve as a solution to handling the plethora of medical data now available. In the 1950s, the amount of medical data was doubling at a rate of every 50 years, which increased to every seven years by 1980. By 2010, the amount of medical data was doubling every 3.5 years, and is currently estimated to double every 73 days.⁵ This trend necessitates rigorous and efficient data analytic methods, and machine-learning is already transforming medicine by advancing prognostication, interpreting digitized images, and improving diagnostic accuracy. Interestingly, early machine-learning analyses have already demonstrated these capabilities within orthopaedic surgery specifically.⁶⁻¹⁰ A recent systematic review of 11 studies performed by Kunze et al¹¹ investigated the ability of AI to identify anterior cruciate ligament (ACL) tears and meniscus lesions on imaging. The authors found that the accuracy of detecting ACL tears ranged between 90% to 98%, meniscus lesions 85% to 91%, and that the addition of AI models significantly increased the diagnostic performance of radiologists compared to their efforts without these models. Within total joint replacement (TJR), several recent studies have demonstrated clinically relevant capabilities of AI by using them to personalize risk prediction, monitor rehabilitation, and identify implant types (Table 1).^{3,6-10,12-18} Furthermore, these algorithms hold the ability to autonomously retrieve data from electronic medical charts, which may expedite registry creation and chart review.¹⁹ The future of AI in TJR and orthopaedics will likely leverage several recently established applications to automate data extraction and analysis, where a patient’s likelihood of experiencing a clinically meaningful outcome or complication, or the exact manufacturer and sizing of their hip arthroplasty, will be automatically calculated and presented to the clinician.

Table II. Transparent reporting of a multivariate prediction model for individual prognosis or diagnosis guideline checklist.

Topic	Question number	Checklist item
Methods		
Source of data	4a	Describe the study design or source of data (e.g. randomized trial, cohort, or registry data), separately for the development and validation data set, if applicable
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up
Participants	5a	Specify key elements of the study setting (e.g. primary care, secondary care, general population) including number and location of centres
	5b	Describe eligibility criteria for participants
	5c	Give details of treatments received, if relevant
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed
	6b	Report any actions to blind assessment of the outcome to be predicted
Predictors	7a	Clearly define all predictors used in developing the machine learning model, including how and when they were measured
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors
Missing data	9	Describe how missing data were handled (e.g. complete-case analysis, single imputation, multiple imputation) with details of any imputation method
Statistical analysis methods	10a	Describe how predictors were handled in the analyses
	10b*	Specify type of model, all model-building procedures (including any predictor selection or hyperparameter selection if applicable), and method for internal validation
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models
Risk groups	11	Provide details on how risk groups were created, if done
Results		
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome
Model development	14a	Specify the number of participants and outcome events in each analysis
	14b	If done, report the unadjusted association between each candidate predictor and outcome
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e. links to the final model online, code, and final parameters/coefficients), with the architecture described in full in the article
	15b	Explain how to use the prediction model
Model performance	16	Report performance measures (with CIs) for the prediction model

CI, confidence interval.

Unintended consequences

Despite numerous theoretical and realized benefits, we must use machine-learning appropriately and avoid the temptation of the “gold rush” to publish as doing so may lead to misleading conclusions.²⁰ Furthermore, several limitations of the current use of machine-learning in orthopaedic surgery research must be recognized. Though we must trust algorithms to explore and learn from data, we cannot simply assume they are accurate and valid.²¹ Algorithms frequently overfit predictions, responding to patterns specific to one dataset that do not accurately reflect the trends of an entire population, especially when multicollinear correlated predictors are present. This may lead to inflated accuracy and inaccurate conclusions about the potential benefits in practice, highlighting why it is essential to externally validate all algorithms on independent populations.

Data quality must also be considered. The pattern and degree of missing data, such as when not missing at random,^{22,23} or variables exceeding 40% missing data,²⁴⁻²⁶ may cause statistical performance bias and portray inaccurate relationships within the model. Furthermore, data

must not be biased, and included variables should be representative of the sample of patients.²¹ For example, though a feature on a preoperative CT scan of the knee might be an excellent predictor of long-term aseptic loosening after total knee arthroplasty (TKA), it is possible that preoperative CT scans are only obtained in a small and non-representative sample of TKA patients, and therefore limit the generalizability of the model. For these reasons, we recommend abiding by predictive modelling guidelines and expert consensus, such as the Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines,²⁷ and the Guidelines for Developing and Reporting Machine Learning Models in Biomedical Research,²⁸ to help ensure that such analyses are appropriately conducted (Table II). It is imperative that researchers and clinicians look for such adherence when evaluating machine-learning research. Furthermore, we recommend against adopting the use of predictive tools presented in these research articles until rigorous external validation has been performed to confirm their efficacy and reliability.

Where we are, and where we need to go

There remain many challenges as we continue to navigate the use of machine-learning, such as better explaining “black box” phenomenon, inadequate data, and model regulation, and integrating such models into clinical workflow. Publishing methods will be essential as more original research continues to be published, likely with differing results. Not only should open source code be published, but explanations as to how and why decisions were made in algorithmic development processes should also be shared among institutions. This increased clarity in reporting will allow researchers and clinicians to be more confident in the results and conclusions presented, and will be a step towards integration of automated models into electronic medical records for real-time use.

Additionally, we must ensure that we are investing in and using high-quality data, as machine-learning models are only as good as the data that are fed into them. High-quality data are unbiased, and it will be essential that electronic medical records and patient data are collected in a standardized and comprehensive manner to avoid erroneous errors and bias. In the private sector, companies spend substantial amounts of money and resources to ensure that algorithms are developed using high-quality and unbiased data. Researchers should not take conclusions at face value without a comprehensive understanding of where the data were derived from and how it was handled.

If we are to continue using machine-learning, we must thoroughly understand the benefits and limitations of these processes. We must be responsible and informed clinicians and surgeons, and practice integrity with our research; this requires concerted efforts to improve the standardization, integration, and availability of relevant high-quality data. Orthopaedic research is now at a crossroads in terms of man versus machine; however, it must be recognized that we are still in the era of needing man to coexist with machine, especially when defining “meaning.” Though machine-learning is great at working quickly and efficiently in defined roles, we still require humans to define these parameters, as the machine cannot learn what is meaningful and does not function autonomously. Therefore, physician input is another exceedingly important component in creating these models.

AI and machine-learning can find patterns in any large sets of data, but the larger question becomes do they mean anything and can we trust the data? We should continue to attempt to develop clinically meaningful applications, and embrace collaboration in order to ensure the creation of models trained on diverse sets of patients with the potential to be used in real-world settings. However, until there is increased reporting of methodological processes and data by authors, in addition to efforts by clinicians and researchers to become

familiar with machine-learning methods, these studies will likely remain limited to interesting analyses criticized by scepticism. Machine-learning has great potential to benefit patients, and as it continues to be applied in orthopaedic research, it will be up to us to be ahead of or behind the curve.



Take home message

- Artificial intelligence and machine-learning research have seen a substantial increase in recent years given several clinically relevant applications.
- Despite widespread optimism and excitement about the capabilities of artificial intelligence and machine-learning, researchers and clinicians must be critical as to whether these methods are indicated and the quality of data used.
- Increased methodological and data reporting clarity, and increased reader familiarity with machine-learning methods, will be two major steps towards transforming machine-learning studies from interesting studies into clinically meaningful and trustworthy applications.

Twitter

Follow K. N. Kunze @kylekunzemd

References

1. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med.* 2016;375(13):1216–1219.
2. 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-A(9):830–840.
3. Karnuta JM, Haeberle HS, Luu BC, et al. Artificial intelligence to identify arthroplasty implants from radiographs of the hip. *J Arthroplasty.* 2021;36(7S):S290–S294.
4. Shah RF, Bini SA, Martinez AM, Podoia V, Vail TP. Incremental inputs improve the automated detection of implant loosening using machine-learning algorithms. *Bone Joint J.* 2020;102-B(6_Supple_A):101–106.
5. Densen P. Challenges and opportunities facing medical education. *Trans Am Clin Climatol Assoc.* 2011;122:48–58.
6. Karnuta JM, Navarro SM, Haeberle HS, et al. Predicting inpatient payments prior to lower extremity arthroplasty using deep learning: Which model architecture is best? *J Arthroplasty.* 2019;34(10):2235–2241.
7. Hyer JM, Ejaz A, Tsilimigras DI, Paredes AZ, Mehta R, Pawlik TM. Novel machine learning approach to identify preoperative risk factors associated with super-utilization of medicare expenditure following surgery. *JAMA Surg.* 2019;154(11):1014–1021.
8. Shohat N, Goswami K, Tan TL, Yayac M, Soriano A, Sousa R. 2020 Frank Stinchfield Award: Identifying who will fail following irrigation and debridement for prosthetic joint infection. *Bone Joint J.* 2020;102-B(7_Supple_B):11–19.
9. Hyer JM, Paredes AZ, White S, Ejaz A, Pawlik TM. Assessment of utilization efficiency using machine learning techniques: A study of heterogeneity in preoperative healthcare utilization among super-utilizers. *Am J Surg.* 2020;220(3):714–720.
10. Ranti D, Warburton AJ, Hanss K, Katz D, Poeran J, Moucha C. K-means clustering to elucidate vulnerable subpopulations among Medicare patients undergoing total joint arthroplasty. *J Arthroplasty.* 2020;35(12):3488–3497.
11. Kunze KN, Rossi DM, White GM, et al. Diagnostic performance of artificial intelligence for detection of anterior cruciate ligament and meniscus tears: A systematic review. *Arthroscopy.* 2021;37(2):771–781.
12. Borjali A, Chen AF, Muratoglu OK, Morid MA, Varadarajan KM. Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural network. *J Orthop Res.* 2020;38(7):1465–1471.
13. Murphy M, Killen C, Burnham R, Sarvari F, Wu K, Brown N. Artificial intelligence accurately identifies total hip arthroplasty implants: a tool for revision surgery. *Hip Int.* 2021;2021:1120700020987526.
14. Yi PH, Wei J, Kim TK, Sair HI, Hui FK, Hager GD. Automated detection & classification of knee arthroplasty using deep learning. *Knee.* 2020;27(2):535–542.
15. Teuff W, Taetz B, Miezal M, Lorenz M, Pietschmann J, Jollenbeck T. Towards an inertial sensor-based wearable feedback system for patients after total hip arthroplasty: Validity and applicability for gait classification with gait kinematics-based features. *Sensors (Basel).* 2019;19(22):E5006.

16. **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(10):S0883-5403(19)30495-4;2253–2259.
17. **Hsieh CY, Huang HY, Liu KC, Chen KH, Hsu SJ, Chan CT.** Subtask segmentation of timed up and go test for mobility assessment of perioperative total knee arthroplasty. *Sensors*. 2020;20(21):E6302.
18. **Kunze KN, Polce EM, Clapp I, Nwachukwu BU, Chahla J, Nho SJ.** Machine learning algorithms predict functional improvement after hip arthroscopy for femoroacetabular impingement syndrome in athletes. *J Bone Joint Surg Am*. 2021;103-A(12):1055–1062.
19. **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(7_Supple_B):99–104.
20. **Gazendam A, Ekhtiari S, Wong E, Madden K, Naji L, Phillips M.** The “infodemic” of journal publication associated with the novel coronavirus disease. *J Bone Joint Surg Am*. 2020;102-A(13):e64.
21. **Cabitza F, Rasoini R, Gensini GF.** Unintended consequences of machine learning in medicine. *JAMA*. 2017;318(6):517–518.
22. **De Silva AP, Moreno-Betancur M, De Livera AM, Lee KJ, Simpson JA.** Multiple imputation methods for handling missing values in a longitudinal categorical variable with restrictions on transitions over time: a simulation study. *BMC Med Res Methodol*. 2019;19(1):14.
23. **Lee KJ, Carlin JB.** Multiple imputation in the presence of non-normal data. *Stat Med*. 2017;36(4):606–617.
24. **Hardt J, Herke M, Leonhart R.** Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. *BMC Med Res Methodol*. 2012;12:184.
25. **Karhade AV, Shah AA, Bono CM, Ferrone ML, Nelson SB, Schoenfeld AJ.** Development of machine learning algorithms for prediction of mortality in spinal epidural abscess. *Spine J*. 2019;19(12):1950–1959.
26. **Resche-Rigon M, White IR.** Multiple imputation by chained equations for systematically and sporadically missing multilevel data. *Stat Methods Med Res*. 2018;27(6):1634–1649.
27. **Collins GS, Reitsma JB, Altman DG, Moons KG.** Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. *Br J Surg*. 2015;102(3):148–158.
28. **Luo W, Phung D, Tran T, et al.** Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J Med Internet Res*. 2016;18(12):e323:12.

Author information:

- K. N. Kunze, MD, Orthopaedic Surgery Resident, Department of Orthopedic Surgery, Hospital for Special Surgery, New York, New York, USA.
- M. Orr, BS, Data Scientist
- V. Krebs, MD, Orthopaedic Surgeon
- N. S. Piuizzi, MD, Orthopedic Surgeon; Director of Adult Reconstructive Surgery Research Department of Orthopaedic Surgery, Cleveland Clinic, Cleveland, Ohio, USA.
- M. Bhandari, MD, Orthopaedic Surgeon; Associate Chair of Research; Chair of Orthopaedic Surgery, Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada; Department of Surgery, Division of Orthopaedic Surgery, McMaster University, Cleveland, Ohio, USA.

Author contributions:

- K. N. Kunze: Conceptualization, Writing – original draft.
- M. Orr: Writing – review & editing.
- V. Krebs: Writing – review & editing.
- M. Bhandari: Supervision, Writing – review & editing.
- N. S. Piuizzi: Supervision, Writing – review & editing.

Funding statement:

- The author(s) received no financial or material support for the research, authorship, and/or publication of this article.

Open access funding

- The authors confirm that the open access fee for this study was self-funded.

© 2022 Author(s) et al. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (CC BY-NC-ND 4.0) licence, which permits the copying and redistribution of the work only, and provided the original author and source are credited. See <https://creativecommons.org/licenses/by-nc-nd/4.0/>