Artificial intelligence and computer vision in orthopaedic trauma

Machine Learning Consortium; Prijs, Jasper; Liao, Zhibin; Ashkani-Esfahani, Soheil; Olczak, Jakub; Gordon, Max; Jayakumar, Prakash; Jutte, Paul C.; Jaarsma, Ruurd L.; IJpma, Frank F.A.

Published in:
The bone & joint journal

DOI:
10.1302/0301-620X.104B8.BJJ-2022-0119.R1

■ ANNOTATION

Artificial intelligence and computer vision in orthopaedic trauma

THE WHY, WHAT, AND HOW

J. Prijs,
Z. Liao,
S. Ashkani-Esfahani,
J. Olczak,
M. Gordon,
P. Jayakumar,
P. C. Jutte,
R. L. Jaarsma,
F. F. A. IJpma,
J. N. Doornberg,
on behalf of the Machine Learning Consortium

From University Medical Centre, Groningen, the Netherlands, and Flinders University/ Medical Centre, Adelaide, Australia

Artificial intelligence (AI) is, in essence, the concept of 'computer thinking', encompassing methods that train computers to perform and learn from executing certain tasks, called machine learning, and methods to build intricate computer models that both learn and adapt, called complex neural networks. Computer vision is a function of AI by which machine learning and complex neural networks can be applied to enable computers to capture, analyze, and interpret information from clinical images and visual inputs. This annotation summarizes key considerations and future perspectives concerning computer vision, questioning the need for this technology (the 'why'), the current applications (the 'what'), and the approach to unlocking its full potential (the 'how').

Cite this article: ***Bone Joint J*** **2022;104-B(8):911–914.**

Why AI and computer vision?

In orthopaedic surgery, we have been shown time and time again that "surgeons agree mostly with themselves, but not so much with each other".[1–8] Daniel Kahneman coined this form of human bias "WYSIATI": What You See Is All There Is.[9] Our field is rife with unsatisfactory levels of interobserver reliability in the recognition and classification of fractures among surgeons. The issue of reliability covers trauma from injuries involving the upper[2–4,7] and lower limb.[5,6,8] Advances in the power of hardware and computing, the development of more accurate imaging techniques, and improvements in the capabilities of software by using computer vision, promise to increase the speed and accuracy of diagnosis and overcome concerns about reliability for the evaluation of images in trauma.[10,11] The widely used complex neural networks have several characteristic features and merits. Compared with conventional machine-learning methods such as decision tree, random forest, boosting, and support vector machines, which are typically used to solve problems in machine-learning on top of structured data, the convolutional filtering operations in a complex neural network can respond to local patterns in features of input which are spatially and temporally correlated. These consume fewer computational resourses compared with a matrix multiplication process, and hence are predominantly used in the processing of images and videos. Complex neural networks also include a type of deep neural network in which the structure of the model can be easily modified by adding or removing a layer. There is a rich literature involving the structure of complex neural networks in the computer vision domain, providing a good methodological basis for the analysis of medical images. Finally, the number of learnable parameters in a complex neural networks is at a scale of millions to billions, and the optimization of the parameters of models is often favourable when dealing with a massive amount of data.

Rather than a replacement for human interpretation, we believe that the attraction of computer vision in the practice of trauma surgery lies in augmenting the diagnostic capabilities of surgeons and musculoskeletal radiologists, reducing bias and variation, minimizing error and mismanagement, and ultimately buying time to focus on our patients and delivering optimal care.[10,12,13]

How does computer vision work?

AI algorithms are now incorporated into many digital products, from smartphones to automated vehicles. The data generated through use of these devices serve as a perpetual source of information for further computer learning and improvement. In orthopaedic surgery, AI is being used in the development of advanced models of prediction as well as automated methods for the diagnosis and classification of different conditions. Models which predict the stratification of risk using machine-learning now go beyond conventional statistics identifying non-linear relationships between individual characteristics and outcomes.[14,15] For

Correspondence should be sent to J. Prijs; email: jasperprijs@icloud.com

doi:10.1302/0301-620X.104B8. BJJ-2022-0119.R1 $2.00

Bone Joint J 2022;104-B(8):911–914.

instance, models have been used to predict same-day discharge and assess balance and prosthetic alignment during total knee arthroplasty.[16,17] Computer vision has been evaluated in the detection and classification of fractures using radiographs and CT scans.[18,19] In other specialties, clinicians are using this technology to interpret images such as mammograms, fundoscopies for papilloedema, and CT scans for the identification of intracerebral haemorrhage.[20–22] There has been a considerable increase in the number of studies aiming to improve clinical decision-making through the analysis of large databases using AI and computer vision.[18,19,23] The next phase should focus on prospective clinical evaluation, the maturation of techniques, and expansion of work to gain external validity in geographical areas and populations, in order to consolidate accuracy, reliability, and transferability while minimizing bias.[19] Kunze et al[24] and others have emphasized these factors and the need for improvement in the regulations and standards for taxonomy, the quality of data, critical appraisal, and reporting.[25–28]

What are we doing with AI and computer vision?

Appreciating the fundamental differences in 'learning' – the process of absorbing information to increase knowledge, skills, and capabilities, and applying this intelligence across a variety of different contexts – between humans and AI-powered machines can help us improve our understanding of the technology behind computer vision. Humans use the brain's computational power, memory, and innate ability to learn from direct experience or to be trained by others. We are also taught to explain how and why we came to certain conclusions about the things we have learned and interpret, and write out the mathematics (or 'logic') so that it can be understood and validated by others. In contrast, machines driven by AI rely on the provision of data and the respective outcomes into the system to build current and future logic, and understand how outcomes might be inferred. A trained machine-learning model is highly complex, encapsulating millions of numerical parameters that collectively contribute to any decision it makes. Therefore, it is beyond our human capacity to fully explain why a model came to a certain conclusion, as the decision could be based on either a pattern that makes sense (clinically) or on a pattern with apparent association to the decision (i.e. a model may learn to recognize sheep by learning the texture of grass, as sheep are always found on grass).

Increasing the number of labels and observers is the most common way to deal with inadvertent human interobserver variation and mistakes. However, what we are teaching the computer is the majority-voted decision, which is usually the best available truth but unfortunately not error-free. If we want the computer to learn beyond what is given (i.e. information based on our understanding such as of classifications) it needs to act with the task (environment) and trial-and-error actions, where the process is in many ways similar to the evolutionary process. For example, the AlphaGo Zero chess player made by Google AI was created by allowing AI players to play against each other.[29] This was different from the original AlphaGo,[30] which learned from human moves. After a huge number of games, the AI players start to invent moves. As the computer can play so much quicker than a human, it may cover or surpass the entirety of games played throughout human history and thus generate a huge amount of data, which is key to an excellent model. In order to generate enormous datasets and create models that outperform us, it is essential that we collaborate, not only nationally but globally. However, it is also essential to consider ethical issues. For example, what if a dataset of 100,000 images is lost? Even though these images were anonymized, it would still lead to headlines and have an enormous effect on the future collection of these datasets. In addition to ethical considerations, laws between countries about sharing data between institutions, each with their own protocols and mandates, often significantly impair collaborations.

In computer vision-based analysis of orthopaedic images, the input can include any form of digital data, most often radiographs and CT scans. Medical images are usually stored in the Digital Imaging and Communications in Medicine (DICOM) format. As this contains substantial, often unnecessary, and sometimes incorrect, information about the patients and the study, the data are converted into more generic formats such as Portable Network Graphics (PNG, lossless) or Joint Photographic Experts Group (JPEG, lossy compression) files to minimize redundancy and increase efficiency. These data and converted formats are then split into training and test sets in a 60:40 or 80:20 ratio. Within the training set, a separate set of images is selected or stochastically sampled, often using n-fold cross-validation, to develop the validation set. This is then used to optimize the performance of the training set without compromising the objectivity of the test set, which is then finally used to evaluate performance. In other words, one is not directly training the model to fit the test set as a strategy to avoid overfitting the model. Thus, the computer model can effectively perform the designated task, not only on the images it has seen before, but on the images it has yet to see. This characteristic is termed 'generalization'.

The computer can reach human-level performance, or even outperform humans in certain tasks, but limitations in the ways of validating decisions can lower the reliability of medical AI systems, making the use of applied AI in medicine challenging.

Pitfalls and what to look out for when appraising manuscripts dealing with complex neural networks for fractures

There is a healthy reservation or resistance towards using AI in diagnostics and medical decision-making, and anyone who has had AI take the wheel can attest that the deviation from the normal situation is challenging. However, as we gain more experience with the applications of AI, it will become easier to understand and navigate through these situations. Even though computers, given the 'artificial' intelligence, might be able to perform certain tasks better than humans, they do not possess common sense and are therefore always 'stupid' or cold as robots. The main weakness of complex neural networks is the fact that their quality relies heavily on the database upon which they were trained. One cannot expect such a network to recognize fractures or pathology it has not seen before, even though they may be similar to what it already 'knows'. Therefore, the utmost care must be taken when choosing the data that are used for training, testing, and validation, either internally or externally. External validation is a crucial step in the validation of a model on new data from a different geographical location, as this could expose possible biases and performance weaknesses.[19] Many AI models in orthopaedic surgery have not undergone external validation.[19] However, assisting clinicians with AI-based solutions has some important strengths, including consistent predictions, no mental fatigue, no inherent bias, and analysis in just a few seconds. It can reach the level of an experienced clinician and is therefore able to provide continual top-level expertise effortlessly.[31–36]

Future perspectives

Regardless of the challenges in the past, present, and future, there has been a rapid development of AI and a surge of practical applications in day-to-day life. We enjoy the use of voice assistance to turn on the lights, dictate a message, or as reminders. We believe the future of medicine will enjoy similar quality-of-life improvements, with significant effects on the lives of our patients. Would it not bring comfort to patients and doctors to be able to make informed decisions together, based on the patient's specific medical characteristics, and to focus on the patients who require close monitoring, and spend one's time where it is the most efficient? We do not believe that AI will replace doctors, but will instead reduce the burdens on us and allow us to spend our time more efficiently with our patients.

In order to achieve these goals, we need to overcome one of the most difficult challenges yet: the relative shortage of quality data in a single hospital. We need to rise above isolated models that are developed, tested, and applied clinically in one centre, and thus are not applicable elsewhere. Only together can we create large enough databases to predict the conditions that matter, such as patient-specific outcomes based on individual characteristics, the risks of postoperative infection, hardware failure, morbidities, and mortality.

Take home message

- Artificial intelligence has seen a surge of applications; however, only together can the orthopaedic community create large databases so we can train models that are globally applicable and with a greater ability to predict the conditions that matter.

References

1. **Becker SJE**, **Bruinsma WE**, **Guitton TG**, **et al**. Interobserver agreement of the Eaton-Glickel classification for trapeziometacarpal and scaphotrapezial arthrosis. *J Hand Surg Am*. 2016;41(4):532–540.
2. **Beks RB**, **Drijkoningen T**, **Claessen F**, **Guitton TG**, **Ring D**, **Variation G**. Interobserver variability of the diagnosis of scaphoid proximal pole fractures. *J Wrist Surg*. 2018;7(4):350–354.
3. **Bruinsma WE**, **Guitton TG**, **Warner JJP**, **Ring D**, **Science of Variation Group**. Interobserver reliability of classification and characterization of proximal humeral fractures: a comparison of two and three-dimensional CT. *J Bone Joint Surg Am*. 2013;95-A(17):1600–1604.
4. **Doornberg JN**, **Guitton TG**, **Ring D**, **Variation G**. Diagnosis of elbow fracture patterns on radiographs: interobserver reliability and diagnostic accuracy. *Clin Orthop Relat Res*. 2013;471(4):1373–1378.
5. **Mellema JJ**, **Doornberg JN**, **Molenaars RJ**, **Ring D**, **Kloen P**, **Traumaplatform Study Collaborative & Science of Variation Group**. Tibial plateau fracture characteristics: reliability and diagnostic accuracy. *J Orthop Trauma*. 2016;30(5):e144-51.
6. **Mellema JJ**, **Doornberg JN**, **Molenaars RJ**, **Ring D**, **Kloen P**, **Traumaplatform Study Collaborative & Science of Variation Group**. Interobserver reliability of the Schatzker and Luo classification systems for tibial plateau fractures. *Injury*. 2016;47(4):944–949.
7. **Doornberg J**, **Lindenhovius A**, **Kloen P**, **van Dijk CN**, **Zurakowski D**, **Ring D**. Two and three-dimensional computed tomography for the classification and management of distal humeral fractures. Evaluation of reliability and diagnostic accuracy. *J Bone Joint Surg Am*. 2006;88-A(8):1795–1801.
8. **Doornberg JN**, **Rademakers MV**, **van den Bekerom MP**, **et al**. Two-dimensional and three-dimensional computed tomography for the classification and characterisation of tibial plateau fractures. *Injury*. 2011;42(12):1416–1425.
9. **Kahneman D**. *Thinking, fast and slow*. 1st ed. New York, New York, USA: Farrar, Straus and Giroux, 2011.
10. **Ring D**. How Artificial Intelligence May Improve Compassion in Orthopaedic Surgery. AAOS Now; November 1, 2019. https://www.aaos.org/aaosnow/2019/nov/commentary/commentary04/ (date last accessed 29 June 2022).
11. **Ashkani-Esfahani S**, **Mojahed Yazdi R**, **Bhimani R**, **et al**. Assessment of ankle fractures using deep learning algorithms and convolutional neural network. *Orthopedics*. Jul 282020.
12. **Oosterhoff JHF**, **Doornberg JN**, **Machine Learning Consortium**. Artificial intelligence in orthopaedics: false hope or not? A narrative review along the line of Gartner's hype cycle. *EFORT Open Rev*. 2020;5(10):593–603.
13. **Ashkani-Esfahani S**, **Mojahed-Yazdi R**. How artificial intelligence improves orthopaedic practice. *EC Orthop*. 2021;12(7):84–87.
14. **Hendrickx LAM**, **Sobol GL**, **Langerhuizen DWG**, **et al**. A machine learning algorithm to predict the probability of (occult) posterior malleolar fractures associated with tibial shaft fractures to guide "malleolus first" fixation. *J Orthop Trauma*. 2020;34(3):131–138.
15. **Oosterhoff JHF**, **Karhade AV**, **Oberai T**, **Franco-Garcia E**, **Doornberg JN**, **Schwab JH**. Prediction of postoperative delirium in geriatric hip fracture patients: a clinical prediction model using machine learning algorithms. *Geriatr Orthop Surg Rehabil*. 2021;12:21514593211062276.
16. **Wei C**, **Quan T**, **Wang KY**, **et al**. Artificial neural network prediction of same-day discharge following primary total knee arthroplasty based on preoperative and intraoperative variables. *Bone Joint J*. 2021;103-B(8):1358–1366.
17. **Verstraete MA**, **Moore RE**, **Roche M**, **Conditt MA**. The application of machine learning to balance a total knee arthroplasty. *Bone Jt Open*. 2020;1(6):236–244.
18. **Langerhuizen DWG**, **Janssen SJ**, **Mallee WH**, **et al**. What are the applications and limitations of artificial intelligence for fracture detection and classification in orthopaedic trauma imaging? A systematic review. *Clin Orthop Relat Res*. 2019;477(11):2482–2491.
19. **Oliveira E Carmo L**, **van den Merkhof A**, **Olczak J**, **et al**. An increasing number of convolutional neural networks for fracture recognition and classification in orthopaedics: are these externally validated and ready for clinical application? *Bone Jt Open*. 2021;2(10):879–885.
20. **Chilamkurthy S**, **Ghosh R**, **Tanamala S**, **et al**. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet*. 2018;392(10162):2388–2396.
21. **Milea D**, **Najjar RP**, **Zhubo J**, **et al**. Artificial intelligence to detect papilledema from ocular fundus photographs. *N Engl J Med*. 2020;382(18):1687–1695.
22. **Lotter W**, **Diab AR**, **Haslam B**, **et al**. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat Med*. 2021;27(2):244–249.
23. **McDonnell JM**, **Evans SR**, **McCarthy L**, **et al**. The diagnostic and prognostic value of artificial intelligence and artificial neural networks in spinal surgery: a narrative review. *Bone Joint J*. 2021;103-B(9):1442–1448.
24. **Kunze KN**, **Orr M**, **Krebs V**, **Bhandari M**, **Piuzzi NS**. Potential benefits, unintended consequences, and future roles of artificial intelligence in orthopaedic surgery research: a call to emphasize data quality and indications. *Bone Jt Open*. 2022;3(1):93–97.
25. **Farrow L**, **Zhong M**, **Ashcroft GP**, **Anderson L**, **Meek RMD**. Interpretation and reporting of predictive or diagnostic machine-learning research in trauma & orthopaedics. *Bone Joint J*. 2021;103-B(12):1754–1758.
26. **Olczak J**, **Pavlopoulos J**, **Prijs J**, **et al**. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. *Acta Orthop*. 2021;92(5):513–525.
27. **Bayliss L**, **Jones LD**. The role of artificial intelligence and machine learning in predicting orthopaedic outcomes. *Bone Joint J*. 2019;101-B(12):1476–1478.
28. **Jones LD**, **Golan D**, **Hanna SA**, **Ramachandran M**. Artificial intelligence, machine learning and the evolution of healthcare: A bright future or cause for concern? *Bone Joint Res*. 2018;7(3):223–225.
29. **Silver D**, **Schrittwieser J**, **Simonyan K**, **et al**. Mastering the game of Go without human knowledge. *Nature*. 2017;550(7676):354–359.

30. **Silver D**, **Huang A**, **Maddison CJ**, **et al**. Mastering the game of Go with deep neural networks and tree search. *Nature*. 2016;529(7587):484–489.
31. **Yamada Y**, **Maki S**, **Kishida S**, **et al**. Automated classification of hip fractures using deep convolutional neural networks with orthopedic surgeon-level accuracy: ensemble decision-making with antero-posterior and lateral radiographs. *Acta Orthop*. 2020;91(6):699–704.
32. **Chung SW**, **Han SS**, **Lee JW**, **et al**. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop*. 2018;89(4):468–473.
33. **Urakawa T**, **Tanaka Y**, **Goto S**, **Matsuzawa H**, **Watanabe K**, **Endo N**. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. *Skeletal Radiol*. 2019;48(2):239–244.
34. **Mawatari T**, **Hayashida Y**, **Katsuragawa S**, **et al**. The effect of deep convolutional neural networks on radiologists' performance in the detection of hip fractures on digital pelvic radiographs. *Eur J Radiol*. 2020;130:109188.
35. **Blüthgen C**, **Becker AS**, **Vittoria de Martini I**, **Meier A**, **Martini K**, **Frauenfelder T**. Detection and localization of distal radius fractures: Deep learning system versus radiologists. *Eur J Radiol*. 2020;126:108925.
36. **Olczak J**, **Emilson F**, **Razavian A**, **Antonsson T**, **Stark A**, **Gordon M**. Ankle fracture classification using deep learning: automating detailed AO Foundation/Orthopedic Trauma Association (AO/OTA) 2018 malleolar fracture identification reaches a high degree of correct classification. *Acta Orthop*. 2021;92(1):102–108.

Author information:
J. Prijs, BSc, PhD Candidate, Department of Orthopaedic Surgery, Groningen University Medical Centre, Groningen, the Netherlands; Department of Surgery, Groningen University Medical Centre, Groningen, the Netherlands; Department of Orthopaedic & Trauma Surgery, Flinders University, Flinders Medical Centre, Adelaide, Australia.

Z. Liao, PhD, Postdoctoral Computer Scientist, Australian Institute for Machine Learning, Adelaide, Australia.

S. Ashkani-Esfahani, MD, Physician, Department of Orthopaedic Surgery, Massachusetts General Hospital, Boston, USA.

J. Olczak, MD, Orthopaedic Surgery Resident
M. Gordon, MD, PhD, Orthopaedic Surgeon
Institute of Clinical Sciences, Danderyd University Hospital, Karolinska Institute, Stockholm, Sweden.

P. Jayakumar, MD, PhD, Assistant Professor in Surgery and Perioperative Care, The University of Texas at Austin, Dell Medical School, Austin, Texas, USA.

P. C. Jutte, MD, PhD, Professor
F. F. A. IJpma, MD, PhD, Trauma Surgeon
Department of Orthopaedic Surgery, Groningen University Medical Centre, Groningen, the Netherlands.

R. L. Jaarsma, MD, PhD, FRACS, Professor, Department of Orthopaedic & Trauma Surgery, Flinders University, Flinders Medical Centre, Adelaide, Australia.

J. N. Doornberg, MD, PhD, Professor of Orthopaedic Surgery, Department of Orthopaedic Surgery, Groningen University Medical Centre, Groningen, the Netherlands; Department of Orthopaedic & Trauma Surgery, Flinders University, Flinders Medical Centre, Adelaide, Australia.

Author contributions:
J Prijs: Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing
Z. Liao: Validation, Writing – original draft, Writing – review & editing.
S. Ashkani-Esfahani: Validation, Writing – original draft, Writing – review & editing.
J. Olczak: Validation, Writing – original draft, Writing – review & editing.
M. Gordon: Validation, Writing – original draft, Writing – review & editing.
P. Jayakumar: Validation, Writing – original draft, Writing – review & editing.
P. C. Jutte: Validation, Writing – original draft, Writing – review & editing.
R. L. Jaarsma: Validation, Writing – original draft, Writing – review & editing.
F. F. A. IJpma: Validation, Writing – original draft, Writing – review & editing.
J. N. Doornberg: Conceptualization, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding statement:
The authors received no financial or material support for the research, authorship, and/or publication of this article.

ICMJE COI statement:
M. Gordon declares a leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid, and stock or stock options as founder of DeepMed, all of which are unrelated to this article.

The Machine Learning Consortium:
Britt Barvelink, Joost Colaris, Chris DiGiovanni, Andrew Duckworth, Hamid Ghaednia, Daniel Guss, Merilyn Heng, Sanne Hoeksema, Mike Hogervorst, Gino Kerkhoffs, Charlotte Laane, Koen Oude Nijhuis, Peter van Ooijen, Jacobien Oosterhoff, David Ring, Jospeph Schwab, Sheila Sprague, Vincent Stirler, Matthieu Wijffels.

This article was primary edited by J. Scott.