



Artificial intelligence in musculoskeletal imaging: a perspective on value propositions, clinical use, and obstacles

Jan Fritz^{1,2} · Richard Kijowski¹ · Michael P. Recht¹

Received: 18 March 2021 / Revised: 25 April 2021 / Accepted: 25 April 2021 / Published online: 13 May 2021
© ISS 2021

Abstract

Artificial intelligence and deep learning (DL) offer musculoskeletal radiology exciting possibilities in multiple areas, including image reconstruction and transformation, tissue segmentation, workflow support, and disease detection. Novel DL-based image reconstruction algorithms correcting aliasing artifacts, signal loss, and noise amplification with previously unobtainable effectiveness are prime examples of how DL algorithms deliver promised value propositions in musculoskeletal radiology. The speed of DL-based tissue segmentation promises great efficiency gains that may permit the inclusion of tissue compositional-based information routinely into radiology reports. Similarly, DL algorithms give rise to a myriad of opportunities for workflow improvements, including intelligent and adaptive hanging protocols, speech recognition, report generation, scheduling, precertification, and billing. The value propositions of disease-detecting DL algorithms include reduced error rates and increased productivity. However, more studies using authentic clinical workflow settings are necessary to fully understand the value of DL algorithms for disease detection in clinical practice. Successful workflow integration and management of multiple algorithms are critical for translating the value propositions of DL algorithms into clinical practice but represent a major roadblock for which solutions are critically needed. While there is no consensus about the most sustainable business model, radiology departments will need to carefully weigh the benefits and disadvantages of each commercially available DL algorithm. Although more studies are needed to understand the value and impact of DL algorithms on clinical practice, DL technology will likely play an important role in the future of musculoskeletal imaging.

Keywords Artificial intelligence · Deep learning · Musculoskeletal · Radiology

Introduction

Deep learning (DL) is an advanced form of artificial intelligence that has revolutionized computer vision, leading to a

rather rapid expansion into a much broader range of science and engineering disciplines, including medical imaging [1].

New DL applications to musculoskeletal radiology have rapidly expanded over the past decade, including image reconstruction, image data transformation, tissue segmentation, body compositional analysis, opportunistic screening, workflow support, and disease detection [2–6].

We provide a perspective on the prospects and value propositions of DL applications to musculoskeletal radiology, adoption into clinical practice, and obstacles.

✉ Jan Fritz
jan.fritz@nyulangone.org

Richard Kijowski
richard.kijowski@nyulangone.org

Michael P. Recht
michael.recht@nyulangone.org

¹ New York University Grossman School of Medicine, New York University, New York, NY 10016, USA

² Department of Radiology, Musculoskeletal Radiology, New York University Grossman School of Medicine, 660 1st Ave, 3rd Floor, Rm #313, New York, NY 10016, USA

Image reconstruction and transformation

Continuous improvement of musculoskeletal magnetic resonance imaging (MRI) efficiency is an essential cornerstone for retaining and further increasing its value through expanding MRI availability and accessibility, improving tolerability, decreasing motion artifacts, decreasing needs for sedation and

anesthesia, and augmenting throughput [7, 8]. Novel DL-based image reconstruction algorithms are prime examples of how DL satisfies this need and delivers promised value propositions in musculoskeletal radiology in an exemplary fashion.

Novel DL reconstruction algorithms correct aliasing artifacts, signal loss, and noise amplification with previously unobtainable effectiveness, permitting reconstruction of 4-fold accelerated MRI acquisitions with better image quality than widely used 2-fold parallel imaging acceleration [9]. Scanner-based DL image reconstruction algorithms can further reduce the acquisition times of a broad range of accelerated pulse sequences by 50% or more, without delays, need for external software, or additional information technology infrastructure to enable clinical MRI exams under 5 min [10].

Similarly promising are DL algorithms for noise reduction of computed tomography (CT) images, which have great potential to redefine both the thresholds of reduced-dose CT and frontiers of ultra-high spatial resolution CT [11].

DL-based imaging data transformation is another exciting avenue that promises multiple benefits, including efficiency gains, reduced radiation exposure, and shorter acquisition times. DL algorithms can accurately and automatically transform MRI data into synthetic CT images with measurable Hounsfield unit data and show promise to synthetically add fat suppression to native MR images [12–15].

Tissue segmentation

Tissue segmentation is an essential component of articular cartilage mapping, compositional analysis of muscle bulk and fatty tissues, and often a requisite for advanced tasks, such as lesion detection [5, 16, 17]. Humans can perform segmentation tasks accurately, but most are time-consuming and tedious. The speed of DL-based tissue segmentation promises great efficiency gains that may permit the inclusion of tissue compositional-based information routinely into radiology reports, including opportunistic screening for sarcopenia, obesity, and osteoporosis, as well as volumes and fat fractions of muscle tissues [18].

Workflow enhancements

DL algorithms give rise to multiple opportunities for improving musculoskeletal workflows, including intelligent and adaptive hanging protocols, speech recognition, report generation, scheduling, precertification, and billing.

Prospective protocolling of MRI and CT studies is an important task that ensures every patient receives the most appropriate imaging exam. However, the synthesis of demographic information, surgical and medical history, allergies,

history of present illness, previous imaging studies, pathological diagnosis, microbiological analysis, and laboratory data is a time-consuming process, which can be greatly expedited with DL support [19].

The ability to customize hanging protocols is a much desired and utilized feature for the systematic evaluation of musculoskeletal imaging studies. However, the current hanging protocol capabilities of PACS systems are often unsatisfactory due to the myriad variations of imaging protocols and labels, which leads to a large amount of time spent by radiologists to manually rearrange image series. While many conventional hanging protocols depend on DICOM header information, DL algorithms can be trained to recognize countless image characteristics for accurately assigning viewports, display modes, and features, including imaging modality, anatomy, contrast weighting, reconstruction kernel, and spatial orientation [20].

Computerized speech recognition for report generation is now used almost ubiquitously in musculoskeletal radiology. While it has increased efficiency and is cost-effective, the phonetic error rate has increased. More than 20% of radiology reports may contain errors in laterality, wrong-word substitution, nonsense phrases, missing words, and spelling errors [21]. Advanced DL-based algorithms could provide important safeguards and reduce error rates by utilizing more information, including patients, studies, and idiosyncrasies [22].

DL-based scheduling support is another potentially high yield area where advanced algorithms could add tremendous value. By incorporating multiple layers of information and patterns based on patient demographics, past behavior, socioeconomic data, site-specific idiosyncrasies, referrers, planned procedures, as well as weather, traffic, holidays, public health, and seasonal information, DL algorithms could accelerate scheduling, reduce no-show rates, and permit sustainable overbooking to avoid empty imaging slots [23–25].

Disease detection

The value proposition of disease-detecting DL algorithms is primarily footed on improved diagnostic performance while reducing human subjectivity and errors due to distraction and fatigue. DL algorithms can also incorporate many complex and higher dimensional radiomics features for diagnosis [26, 27].

Fundamental questions that require answers before the widespread adoption of disease-detecting DL algorithms into clinical practice include whether DL algorithms measurably improve the already low (3–5%) error rate of radiologists (21) and what type of systematic errors DL algorithms make (e.g., subtle errors versus egregious errors). A subtle error may constitute a missed single bundle partial-thickness anterior cruciate ligament tear, which radiologists may or may not diagnose,

whereas an egregious error may constitute a missed full-thickness double-bundle anterior cruciate ligament tear, which most radiologists would diagnose [28].

A recent meta-analysis summarized the performance of DL algorithms to detect osseous fractures [29]. A total of 5914 radiographic examinations with 1343 fractures (prevalence, 22.7%) qualified for inclusion. The DL algorithms made 1086 true-positive (18.4%), 259 false-positive (4.3%), 257 false-negative (4.3%), and 4312 true-negative (73%) diagnoses, which suggests a pooled area under the curve of 87%, sensitivity of 80%, and specificity of 94%. While the diagnostic performance is promising, several of the included studies showed similar or higher diagnostic performances of human interpretations [30–34]. Similarly, studies evaluating the diagnostic performances of DL algorithms detecting and classifying internal derangement of the knee joint on MRI have also demonstrated that human performance is similar or better than DL performance [17, 35–38].

Although studies have demonstrated similar performances to humans, studies have also demonstrated that the combination of DL algorithms and radiologists performs better than either alone, suggesting that the most viable clinical practice model might be musculoskeletal radiologists using DL algorithms for interpretations rather than DL algorithms working independently [36, 39].

Clinical efficiency gains

DL algorithms may reduce the time needed to read musculoskeletal studies. A recent study demonstrated reduced reading time when using a deep DL algorithm to detect rib fractures on chest CT exams compared to interpreting CT exams alone [40]. However, promised efficiency gains by DL algorithms might require careful consideration if study designs did not include a clinically applicable workflow.

For example, a DL algorithm proposed for detecting a certain musculoskeletal condition in a certain anatomical region may be studied with a test set of 250 radiographs. The study may conclude that the radiologists read cases 30% faster with the DL algorithm while maintaining high diagnostic performance, which could equate to finishing a day's work earlier or interpreting more cases during the workday. However, the efficiency gains derived by such a study design may not translate into clinical practice because the 250 radiographs of a certain anatomical region will most likely not be read back-to-back but will be scattered within a mixed volume of cases, meaning the efficiency gains will span over a longer period of time than proposed.

Additional statistically significant studies comparing the impact of DL algorithms on diagnostic accuracy and interpretation speed are necessary to fully understand the value of DL algorithms for musculoskeletal disease detection.

Practice integration

In addition to the central focus on diagnostic performance and efficiency gains, an important consideration for delivering the value propositions of DL algorithms is the successful integration into the workflow of musculoskeletal radiologists. Despite a growing number of regulatory agency-approved DL algorithms and artificial intelligence products [41], widespread adoption by musculoskeletal radiologists has not yet occurred. While hesitancy may be due to uncertainties about diagnostic performance improvements and clinical efficacy gains, challenges in successfully integrating well-validated DL algorithms into the workflow of musculoskeletal radiologists also represent a major roadblock.

Given the current focus of DL algorithms on single tasks, comprehensive artificial intelligence support for musculoskeletal radiology reports will likely require the adoption of multiple algorithms. The most effective algorithms will likely be from different vendors and include detection, segmentation, classification, and prediction of multiple conditions, including fractures, masses, degeneration, deformities, malalignment, bone density, and soft tissue lesions.

While the practice integration of a single algorithm may already pose substantial challenges, the integration and management of multiple algorithms is difficult and may represent the largest roadblock for practical integration of DL algorithms.

Fundamental prerequisites for seamless integration of DL algorithms include routing imaging studies to the DL algorithm, fast image processing, presentation and formatting of the results, and interfacing with viewing software and dictation systems. Vendor-specific app stores featuring DL algorithms for their hospital and radiology information systems, PACS systems, and post-processing software may provide solutions that can utilize and integrate seamlessly into preexisting infrastructure. However, vendor-neutral infrastructures providing a universal plug-and-play environment may be needed ultimately for successful practice integration and management [42].

It is important to note that many DL algorithms may render a diagnosis regardless of what images are being provided as input. For example, a bone age-determining algorithm will return a bone age even if presented a chest radiograph or photo of a cat [43]. Therefore, the routing of accurately preselected images to the DL algorithms is of paramount importance, as many DL algorithms cannot automatically select appropriate and reject inappropriate images. For such DL algorithms, imaging studies will either have to be curated digitally, which may be achieved with other DL algorithms or manually routed [44].

An additional requirement for the successful adoption of DL into daily practice is the PACS integration of DL results to eliminate the need to open and switch between multiple applications while interpreting exams using algorithms from different vendors.

Costs versus benefits

The ultimate question regarding using DL algorithms in clinical practice may pertain to the tradeoffs between benefits and costs. The answer will need to factor in the true diagnostic value, efficiency gains, and problem-solving abilities of DL algorithms.

The cost of DL algorithms may be offset through efficiency gains. However, musculoskeletal radiologists and DL algorithms may detect the vast majority of musculoskeletal diseases and injuries with similar accuracy, and indeterminate musculoskeletal tumors may still need tissue sampling despite a DL-augmented differential diagnosis.

For DL algorithms that increase the diagnostic performance for disease detection, multiple business models are under debate, including radiology departments bearing the cost, managed healthcare bearing the cost by paying for DL with the hope of recouping the cost through improved patient outcomes, or shifting the cost to patients.

Each radiology department will need to prioritize which DL algorithms should be purchased based upon their individual needs. The needs of musculoskeletal radiologists are different from those of neuroradiologists or breast imagers. Therefore, the impact and value of DL need to be carefully evaluated for each use case before adoption.

DL algorithms that do not result in meaningful improvements in diagnostic performance and productivity may still be beneficial by reducing fatigue and stress, which could help combat the cautioning 80% prevalence of burnout symptoms that a study by the Society of Skeletal Radiology found among musculoskeletal radiologists [22].

Conclusion

DL offers musculoskeletal radiology exciting possibilities, including faster MRI, reduced-dose CT, image data transformation, automated tissue segmentation, workflow support, and disease detection. Novel DL-based image reconstruction algorithms are prime examples of how DL algorithms deliver promised value propositions in musculoskeletal radiology in an exemplary fashion. Although additional studies are needed to understand the value and impact of DL algorithms on clinical practice, DL technology will likely play an important role in the future of musculoskeletal imaging.

Declarations

Conflict of interest Jan Fritz received institutional research support from Siemens AG, BTG International, Zimmer Biomed, DePuy Synthes, QED, and SyntheticMR; is a scientific advisor for Siemens AG, SyntheticMR, GE Healthcare, QED, BTG, ImageBiopsy Lab, Boston Scientific, and Mirata Pharma; and has shared patents with Siemens Healthcare and Johns Hopkins University. Richard Kijowski: none. Michael Recht: none.

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–44.
2. Kijowski R, Liu F, Caliva F, Pedeoia V. Deep learning for lesion detection, progression, and prediction of musculoskeletal disease. *J Magn Reson Imaging*. 2020;52(6):1607–19.
3. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial intelligence in musculoskeletal imaging: current status and future directions. *AJR Am J Roentgenol*. 2019;213(3):506–13.
4. Chea P, Mandell JC. Current applications and future directions of deep learning in musculoskeletal radiology. *Skelet Radiol*. 2020;49(2):183–97.
5. Grauhan NF, Niehues SM, Gaudin RA, Keller S, Vahldiek JL, Adams LC, et al. Deep learning for accurately recognizing common causes of shoulder pain on radiographs. *Skeletal Radiol*. 2021.
6. Guan B, Liu F, Mizaian AH, Demehri S, Samsonov A, Guermazi A, et al. Deep learning approach to predict pain progression in knee osteoarthritis. *Skeletal Radiol*. 2021.
7. Del Grande F, Guggenberger R, Fritz J. Rapid musculoskeletal MRI in 2021: value and optimized use of widely accessible techniques. *AJR Am J Roentgenol*. 2021;216(3):704–17.
8. Fritz J, Guggenberger R, Del Grande F. Rapid musculoskeletal MRI in 2021: clinical application of advanced accelerated techniques. *AJR Am J Roentgenol*. 2021;216(3):718–33.
9. Recht MP, Zbontar J, Sodickson DK, Knoll F, Yakubova N, Sriram A, et al. Using deep learning to accelerate knee MRI at 3 T: results of an interchangeability study. *AJR Am J Roentgenol*. 2020;215(6):1421–9.
10. Del Grande F, Rashidi A, Luna R, Delcogliano M, Stern SE, Dalili D, et al. Five-minute five-sequence knee MRI using combined simultaneous multislice and parallel imaging acceleration: comparison with 10-minute parallel imaging knee MRI. *Radiology*. 2021;203655.
11. Nakamura Y, Higaki T, Tatsugami F, Honda Y, Narita K, Akagi M, et al. Possibility of deep learning in medical imaging focusing improvement of computed tomography image quality. *J Comput Assist Tomogr*. 2020;44(2):161–7.
12. Jans LBO, Chen M, Elewaut D, Van den Bosch F, Carron P, Jacques P, et al. MRI-based synthetic CT in the detection of structural lesions in patients with suspected sacroiliitis: comparison with MRI. *Radiology*. 2021;298(2):343–9.
13. Fayad LM, Parekh VS, de Castro Luna R, Ko CC, Tank D, Fritz J, et al. A deep learning system for synthetic knee magnetic resonance imaging: is artificial intelligence-based fat-suppressed imaging feasible? *Invest Radiol*. 2020; Publish Ahead of Print.
14. Breighner RE, Endo Y, Konin GP, Gulotta LV, Koff MF, Potter HG. Technical developments: zero echo time imaging of the shoulder: enhanced osseous detail by using MR imaging. *Radiology*. 2018;286(3):960–6.
15. Fritz J. Automated and radiation-free generation of synthetic CT from MRI data: does AI help to cross the finish line? *Radiology*. 2021;298(2):350–2.
16. Medina G, Buckless CG, Thomasson E, Oh LS, Torriani M. Deep learning method for segmentation of rotator cuff muscles on MR images. *Skelet Radiol*. 2021;50(4):683–92.
17. Liu F, Zhou Z, Samsonov A, Blankenbaker D, Larison W, Kanarek A, et al. Deep learning approach for evaluating knee MR images: achieving high diagnostic performance for cartilage lesion detection. *Radiology*. 2018;289(1):160–9.
18. Boutin RD, Lenchik L. Value-added opportunistic CT: insights into osteoporosis and sarcopenia. *AJR Am J Roentgenol*. 2020;215(3):582–94.

19. Kalra A, Chakraborty A, Fine B, Reicher J. Machine learning for automation of radiology protocols for quality and efficiency improvement. *J Am Coll Radiol*. 2020;17(9):1149–58.
20. Lee M, Kasmanoff N, Parente M, Razavian N, Lui Y. Using deep multi-task learning to classify brain MR images by sequence and orientation. *American Society of Neuroradiology Annual Meeting*. 2020.
21. Quint LE, Quint DJ, Myles JD. Frequency and spectrum of errors in final radiology reports generated with automatic speech recognition technology. *J Am Coll Radiol*. 2008;5(12):1196–9.
22. Chew FS, Mulcahy MJ, Porrino JA, Mulcahy H, Relyea-Chew A. Prevalence of burnout among musculoskeletal radiologists. *Skeletal Radiol*. 2017;46(4):497–506.
23. Chong LR, Tsai KT, Lee LL, Foo SG, Chang PC. Artificial intelligence predictive analytics in the management of outpatient MRI appointment no-shows. *AJR Am J Roentgenol*. 2020;215(5):1155–62.
24. Harvey HB, Liu C, Ai J, Jaworsky C, Guerrier CE, Flores E, et al. Predicting no-shows in radiology using regression modeling of data available in the electronic medical record. *J Am Coll Radiol*. 2017;14(10):1303–9.
25. Curtis C, Liu C, Bollerman TJ, Pianykh OS. Machine learning for predicting patient wait times and appointment delays. *J Am Coll Radiol*. 2018;15(9):1310–6.
26. Fritz B, Muller DA, Sutter R, Wumig MC, Wagner MW, Pfirmann CWA, et al. Magnetic resonance imaging-based grading of cartilaginous bone tumors: added value of quantitative texture analysis. *Investig Radiol*. 2018;53(11):663–72.
27. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563–77.
28. Harvey HB, Gowda V. Clinical applications of AI in MSK imaging: a liability perspective. *Skelet Radiol*. 2021.
29. Yang S, Yin B, Cao W, Feng C, Fan G, He S. Diagnostic accuracy of deep learning in orthopaedic fractures: a systematic review and meta-analysis. *Clin Radiol*. 2020;75(9):713 e717–28.
30. Chung SW, Han SS, Lee JW, Oh KS, Kim NR, Yoon JP, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop*. 2018;89(4):468–73.
31. 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. *Skelet Radiol*. 2019;48(2):239–44.
32. Choi JW, Cho YJ, Lee S, Lee J, Lee S, Choi YH, et al. Using a dual-input convolutional neural network for automated detection of pediatric supracondylar fracture on conventional radiography. *Investig Radiol*. 2020;55(2):101–10.
33. Gan K, Xu D, Lin Y, Shen Y, Zhang T, Hu K, et al. Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. *Acta Orthop*. 2019;90(4):394–400.
34. Yu JS, Yu SM, Erdal BS, Demirer M, Gupta V, Bigelow M, et al. Detection and localisation of hip fractures on anteroposterior radiographs with artificial intelligence: proof of concept. *Clin Radiol*. 2020;75(3):237 e231–9.
35. Fritz B, Marbach G, Civardi F, Fucntese SF, Pfirmann CWA. Deep convolutional neural network-based detection of meniscus tears: comparison with radiologists and surgery as standard of reference. *Skelet Radiol*. 2020;49(8):1207–17.
36. Bien N, Rajpurkar P, Ball RL, Irvin J, Park A, Jones E, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. *PLoS Med*. 2018;15(11):e1002699.
37. Liu F, Guan B, Zhou Z, Samsonov A, Rosas H, Lian K, et al. Fully automated diagnosis of anterior cruciate ligament tears on knee MR images by using deep learning. *Radiol Artif Intell*. 2019;1(3):180091.
38. Germann C, Marbach G, Civardi F, Fucntese SF, Fritz J, Sutter R, et al. Deep convolutional neural network-based diagnosis of anterior cruciate ligament tears: performance comparison of homogeneous versus heterogeneous knee MRI cohorts with different pulse sequence protocols and 1.5-T and 3-T magnetic field strengths. *Investig Radiol*. 2020;55(8):499–506.
39. Fritz J, Germann C, Sutter R, Fritz B. AI-augmented MRI diagnosis of ACL tears: which readers benefit? *SSR 2021 Annual Meeting*. 2021.
40. Zhang B, Jia C, Wu R, Lv B, Li B, Li F, et al. Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: a clinical evaluation. *Br J Radiol*. 2021;94(1118):20200870.
41. Benjamens S, Dhunoo P, Mesko B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med*. 2020;3:118.
42. Leiner T, Bennink E, Mol CP, Kuijff HJ, Veldhuis WB. Bringing AI to the clinic: blueprint for a vendor-neutral AI deployment infrastructure. *Insights Imaging*. 2021;12(1):11.
43. Yi PH, Arun A, Hafezi-Nejad N, Choy G, Sair HI, Hui FK, et al. Garbage In, Garbage out? Performance of state-of-the-art deep learning-based bone age algorithm on inappropriate data inputs. *SSR 2021 Annual Meeting* 2021.
44. Yi PH, Kim TK, Wei J, Shin J, Hui FK, Sair HI, et al. Automated semantic labeling of pediatric musculoskeletal radiographs using deep learning. *Pediatr Radiol*. 2019;49(8):1066–70.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.