The Use of Artificial Intelligence in the Evaluation of Knee Pathology

Elisabeth R. Garwood, MD[1]   Ryan Tai, MD[1]   Ganesh Joshi, MD[1]   George J. Watts V., MD[1]

[1] Division of Musculoskeletal Imaging and Intervention, Department of Radiology, University of Massachusetts Memorial Medical Center and University of Massachusetts Medical School, Worcester, Massachusetts

Address for correspondence Elisabeth R. Garwood, MD, Department of Radiology, University of Massachusetts Memorial Medical Center and University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655.

Abstract

Keywords
► artificial intelligence
► magnetic resonance imaging
► deep learning
► knee

Artificial intelligence (AI) holds the potential to revolutionize the field of radiology by increasing the efficiency and accuracy of both interpretive and noninterpretive tasks. We have only just begun to explore AI applications in the diagnostic evaluation of knee pathology. Experimental algorithms have already been developed that can assess the severity of knee osteoarthritis from radiographs, detect and classify cartilage lesions, meniscal tears, and ligament tears on magnetic resonance imaging, provide automatic quantitative assessment of tendon healing, detect fractures on radiographs, and predict those at highest risk for recurrent bone tumors. This article reviews and summarizes the most current literature.

Knee pain and injury are commonly encountered in clinical practice. Sports-related knee injuries alone account for > 2.5.million emergency department visits annually,[1] and increasingly, clinicians are relying on the results of musculoskeletal (MSK) imaging to guide diagnosis and management.[2] Interpretation of advanced MSK imaging is both labor intensive and subject to reader variability even when interpreted by subspecialty trained MSK radiologists, partly attributed to the large quantity of data presented by each study and the level of imaging detail.[3] The integration of artificial intelligence (AI) algorithms into the workflow of MSK radiology holds the potential to improve diagnostic accuracy, expedite cases with urgent findings, reduce reader fatigue, and provide decision support where radiology expertise is unavailable.[4]

In the diagnostic evaluation of knee pathology, most AI literature has focused on building convolutional neural networks (CNNs) that can perform a single interpretive task under the categories of pathology detection (ligament or meniscus tear, cartilage lesion), classification (assign osteoarthritis grading to knee radiographs, classify meniscus tears), and segmentation (cartilage and meniscus segmentation) (►**Fig. 1**). CNNs are a form of deep learning, a subcategory of machine learning (ML) that refers to algorithms with multiple interconnected layers reminiscent of the layered approach used by neurons in the brain.[5] CNNs are a specific class of deep-learning technique that use a mathematical operation called a convolution. This class of network is commonly used for image classification and analysis tasks. All of these techniques fall under the umbrella of "artificial intelligence" (►**Fig. 2**). Advances are being made using AI to accelerate magnetic resonance imaging (MRI) acquisition. This topic is addressed in a separate dedicated article in this journal. In terms of image-based tasks in the evaluation of knee pathology following image acquisition, AI currently holds the most potential for influencing lesion detection, characterization, and disease monitoring (►**Fig. 3**).

Cartilage/Osteoarthritis

Among the many AI applications being investigated in the setting of diagnostic imaging of the knee, one of the more established is the evaluation of cartilage disease. This is in part due to the worldwide disease prevalence of osteoarthritis (OA).[6] Knee OA is one of the most common forms of arthritis and a leading cause of chronic disability, projected to affect 59 million people in the United States in 2019.[7,8] Another key driver in ML algorithm development for cartilage evaluation is the public availability of large repositories of combined clinical and imaging data through the Osteoarthritis Initiative (OAI) and other large epidemiologic studies of OA. Building and training robust ML algorithms for cartilage evaluation depends
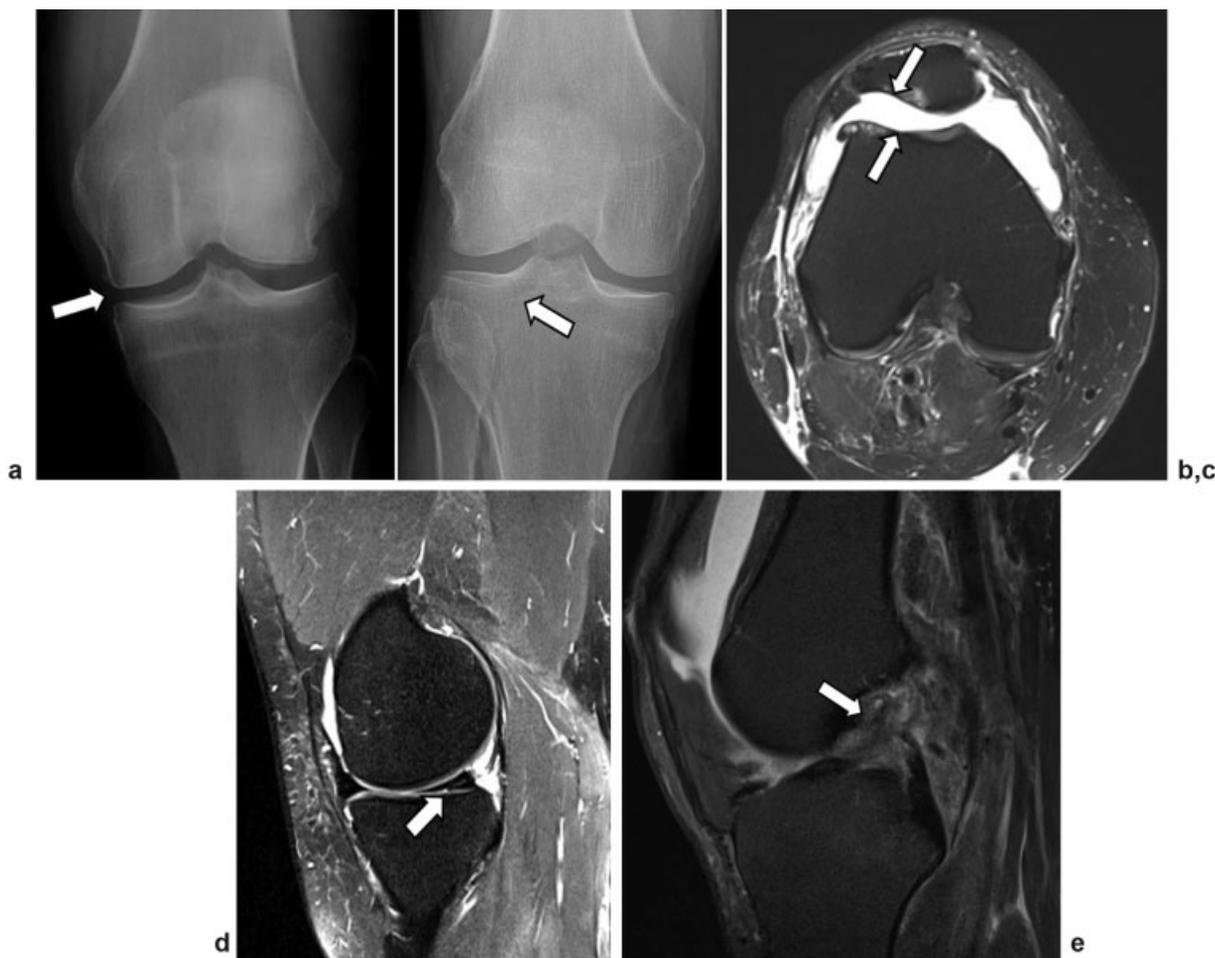
Fig. 1 Interpretive applications of artificial intelligence (AI) in the evaluation of knee pathology. AI algorithms have been built that can perform the following interpretive tasks: (**a**) Assign osteoarthritis severity grade to radiographs: Anteroposterior (AP) weight-bearing radiograph demonstrates mild medial tibiofemoral joint space narrowing and osteophytic spurring, Kellgren-Lawrence grade 2 (arrow). (**b**) Detect fractures on radiographs: AP radiograph demonstrates subtle lateral tibial plateau fracture (arrow). (**c**) Detect and classify cartilage lesions on MRI: Axial T2 fat-saturated sequence demonstrates broad-based full-thickness cartilage loss in the patellofemoral compartment (arrows). (**d**) Detect and classify meniscus tears on MRI: Sagittal proton-density (PD) fat-saturated sequence demonstrates a horizontal tear of the medial meniscus (arrow). (**e**) Detect anterior cruciate ligament (ACL) tears on MRI: Sagittal PD fat-saturated sequence demonstrates a complete midsubstance tear of the ACL (arrow).
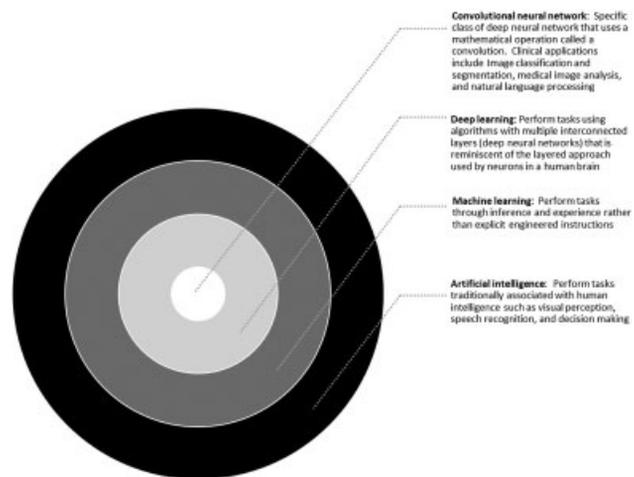
Convolutional neural network: Specific class of deep neural network that uses a mathematical operation called a convolution. Clinical applications include image classification and segmentation, medical image analysis, and natural language processing

Deep learning: Perform tasks using algorithms with multiple interconnected layers (deep neural networks) that is reminiscent of the layered approach used by neurons in a human brain

Machine learning: Perform tasks through inference and experience rather than explicit engineered instructions

Artificial intelligence: Perform tasks traditionally associated with human intelligence such as visual perception, speech recognition, and decision making

Fig. 2 Schematic of definitions. Convolutional neural networks (CNNS) are a specific class of deep neural network commonly used for image classification, analysis, and segmentation tasks. CNNs are a form of deep learning, a subcategory of machine learning. All of these techniques fall under the umbrella of artificial intelligence.

on access to large data sets containing this type of curated information.

The diagnosis of knee OA is currently made by clinical assessment of symptoms in combination with radiographic findings indicative of OA including joint space narrowing and osteophyte formation. Although radiographs are widely available, safe, and inexpensive, severity grading of knee OA by radiographs, the current imaging standard, suffers from the inherent insensitivity of radiographs to detect early changes indicative of OA and subjective variability in radiographic interpretation. The semiquantitative Kellgren-Lawrence (KL) grading scale[9] is the traditional method by which knee OA is assessed on radiographs, where a 0 through 4 ordinal scale is used (0 = normal; 4 = severe osteoarthritis). Interrater agreement when using this method ranges from 0.5 to 0.8, reflecting substantial levels of interobserver disagreement.[10,11] Driven by these challenges, interest in automating the task of knee OA quantification from radiographs has a long history, dating back to the late 1980s.[12]
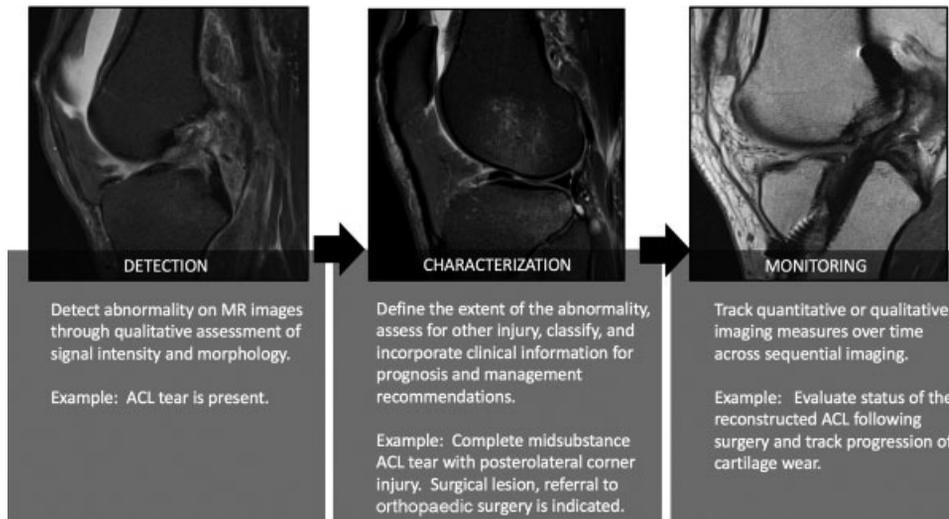
Fig. 3 How artificial intelligence (AI) may impact image-based tasks in the evaluation of knee pathology. This schematic outlines imaging-based tasks following image acquisition in musculoskeletal radiology where AI may have a potential impact, using an anterior cruciate ligament (ACL) tear as an example.

Cartilage/Osteoarthritis Evaluation: Radiographs

Several research groups have recently developed computer-assisted and ML models for predicting KL scores based on radiographic features.[7,13–18] The studies using AI methods for automation are summarized in ►Table 1.

Tiulpin et al sought to grade knee OA automatically by assigning a KL score based on a computer-aided diagnosis tool powered by a novel CNN based on Deep Siamese CNN architecture using image symmetry.[18] The algorithm was trained using the Multicenter Osteoarthritis Study (MOST) data set, a publicly available data set of manually KL-graded knee radiographs, and testing it using the OAI data set containing radiographs of 5,960 knee joints from 3,000 subjects. The algorithm had excellent agreement with manually graded radiographs, with a quadratic κ coefficient of 0.83. The study also used visual "attention maps" to aid in multiclass discrimination, and the algorithm performed with an average multiclass accuracy of 66.71%.

In a similar study conducted by Norman et al, a CNN utilizing DenseNet ensemble learning and direct demographic input was used to automatically produce KL grades

of OA on knee radiographs.[19] Sensitivity rates for the detection of no OA, mild, moderate, and severe OA were 83.7%, 70.2%, 68.9%, and 86.0%, respectively. The corresponding specificity rates were higher at 86.1%, 83.8%, 97.1%, and 99.1%. Visual depictions, or "saliency maps," were used to confirm that the neural networks were basing the assessment on osteoarthritic features rather than to regions that do not have relevant radiologic features.

Antony et al introduced the concept of an "end-to-end" AI model for the automatic radiographic grading of knee OA based on the KL scale.[13] By first using a convolutional neural network, the authors were able to localize the knee joint accurately on radiographic images. In the second stage, the classification network was trained using a regression model to treat the discrete 0 through 4 KL grades as a continuous scale, which the authors argue better approximates the continuous progression of OA in vivo. The study found that the jointly trained classification and regression CNN produced a higher multiclass classification accuracy (63.4%) compared with the classification-only CNN (60.3%). The same research group followed this report, in 2019, with a study using statistical

Table 1 Studies investigating artificial intelligence in diagnostic evaluation of knee cartilage[a]

Modality	Study	Study features	Specificity	Sensitivity	Multiclass classification accuracy	Reader agreement
Radiograph	Tiulpin et al[18]	Attention maps	NA	NA	66.7%	0.83 κ
Radiograph	Norman et al[19]	Saliency maps	83.8–99.1%	68.9–86.0%	NA	NA
Radiograph	Antony et al[17]	CNN plus regression model	NA	NA	63.4%	NA
MRI	Liu et al[23]	Tibiofemoral cartilage	85.2%, 87.9%	84.1%, 80.5%	NA	0.57–0.73 κ (humans) 0.76 κ (system)
MRI	Pedoia et al[24]	Patellar cartilage	80.3%	80.0%	NA	NA

Abbreviations: CNN, convolutional neural network; MRI, magnetic resonance imaging; NA, not applicable.
[a]Segmentation literature excluded.

models to predict the severity of knee OA based on patient data alone and compared that result with CNN-based analysis of knee radiographs alone, again using the OAI data set with the hypothesis that a good predictive model based on clinical data alone may obviate the need for radiographs in the diagnosis and quantification of OA severity.[20] They found that the statistical models based on patient data could predict OA severity with a good level of accuracy comparable with the CNN model using radiographs alone.

Cartilage/Osteoarthritis Evaluation: MRI

Although radiography remains the most common imaging modality by which OA is detected, the evolution of cartilage damage on radiographs must necessarily be inferred through secondary changes such as joint space narrowing because the cartilage itself is not directly visualized. But MRI offers direct visualization of both acute and degenerative cartilaginous lesions. Historically MRI was shown to be a highly specific and moderately sensitive tool in the assessment of knee cartilage lesions,[21] and specifically tailored morphologic cartilage MRI techniques have continued to advance over the years.[22] Recently, there has been growing interest in the development of AI applications aimed at enhancing MRI utility in the assessment of knee cartilage lesions (►Table 1). Many of these techniques rely on algorithms that automate cartilage segmentation, and this topic is covered in a dedicated article on segmentation in this issue.

Liu et al used a two-step process in the creation of a method based on deep learning to detect cartilage lesions by sequentially utilizing two different two-dimensional (2D) CNNs in the evaluation of 175 knee MRIs.[23] The first CNN was used for cartilage segmentation and the second for lesion detection. The authors trained and tested the system twice, treating them as separate data sets to assess intraobserver agreement. The results of the study showed improved sensitivity for lesion detection by the automated system (84.1% and 80.5%) compared with the radiologists (60.8–80.2%), but lower specificity for the automated system (85.2% and 87.9%) compared with the radiologists (92.2–96.5%). There was also higher intraobserver agreement (κ of 0.76) for the automated system compared with the interobserver agreement (κ of 0.57–0.73) between the human radiologists. The lower specificity of the automated system was attributed to the limitations of a 2D system, only using one sagittal fast spin-echo sequence (fat-saturated T2) for lesion detection, whereas the radiologists used three sagittal sequences for lesion detection. The authors contend that the increased sensitivity and intraobserver agreement demonstrated by the automated system may allow for more reliable detection of early cartilage damage because superficial cartilage lesion detection is currently a known limitation among radiologists.[21]

In a study conducted by Pedoia et al,[24] a deep-learning system was created to detect both meniscal injury and patellar cartilage defects in a radiologist-annotated data set of 302 patients (1478 total MRIs), composed of individuals with and without OA, after anterior cruciate ligament (ACL) injury and after ACL reconstruction. In this data set, the meniscus and the patellofemoral cartilage compartment were graded by radiol-

ogists using the modified Whole-organ Magnetic Resonance Imaging Score (WORMS). As in the automated study by Liu et al,[23] a two-step process was used. A 2D U-Net architecture was first used for automatic meniscal and cartilage segmentation, followed by a three-dimensional (3D) CNN used for lesion detection. The results of the study showed binary cartilage lesion detection (WORMS score 2–6 for lesion versus score 0–1 for no lesion) sensitivity of 80.0% and specificity of 80.27% using the radiologist annotation as the gold standard. A human group composed of three radiologists also reviewed a very small subset of cases (17 MRIs) to ascertain interrater variability, yielding average agreements of 89.56% for no cartilage lesion and 79.74% for the presence of a cartilage lesion.

With the continued advancement of clinical AI support, it may be possible to provide earlier and more reliable detection of knee cartilage disease and make outcome predictions based on those observations. AI support tools could potentially play a key role in the diagnosis and treatment of a globally debilitating and costly disease.

Ligaments

Anterior Cruciate Ligament Evaluation: MRI

ACL tears are common orthopaedic injuries, frequently warranting surgical management. Untreated or delayed treatment of ACL tears can impact quality of life, leading to premature knee OA, chronic instability, irreparable medial meniscal tears, and early chondral wear.[25–28] Although history and physical examination can raise the suspicion for ACL tear, MRI is often performed to confirm the diagnosis of tear given the high accuracy of MRI[29] and the ability to identify concomitant injuries. The findings of ACL tear on MRI include discontinuity or nonvisualization of the ligament fibers, abnormal course or contour of the ligament, and abnormal signal.[30] Given the frequency of injury and clinical importance, the ACL is the only ligament around the knee that has been targeted with ML approaches. Several articles address the use of AI to diagnose ACL tears on MRI (►Table 2).

Štajduhar et al demonstrated the feasibility of using a semiautomated model for the detection of ACL tears.[31] Reference standard was established through consensus between radiologists. From the original source data, the region of interest containing the ACL was manually extracted on sagittal proton-density (PD)-weighted fat-suppressed images by a radiologist. The models experimented were composed of a (1) feature extraction method with a (2) ML classification system. The highest performing model used a histogram of the oriented gradient feature extraction method coupled with a support vector ML classification system. This model achieved an area under the receiving operating characteristics curve (AUC) of 0.894 and 0.943 in identifying an injured and completely torn ACL, respectively.[31]

Subsequently, Bien et al developed MRNet, a CNN based on mapping a 3D MRI series to a probability, for the purposes of detecting ACL and meniscus tears. This method demonstrated similar specificity but lower sensitivity in the identification of a tear when compared with a cohort of general radiologists and orthopaedic surgeons.[3] The authors defined an ACL tear as a

Table 2 Studies investigating performance of ML algorithms in detection of ACL tear

Lesion	Study	Reference standard	Sequence	Results	Notes
Injury vs complete ACL tear	Štajduhar et al[31]	Radiology consensus read	Sagittal: PD FS	AUC 0.894 injured 0.943 complete tear	Feature extraction and ML classification
No tear vs partial or complete tear	Bien et al[3]	MSK radiology consensus read on a subset	Sagittal: T2 Coronal: T1 Axial: PD	AUC 0.965 for tear	Similar specificity but lower sensitivity compared with readers
No complete tear vs complete tear	Liu et al[33]	Arthroscopy	Sagittal: PD Sagittal: T2	Sensitivity: 0.96 Specificity: 0.96 AUC: 0.98	No statistically significant difference algorithm vs readers of various training levels
Normal vs complete tear	Chang et al[32]	MSK radiologist read	Coronal: PD	Sensitivity: 1.00 Specificity: 0.933	Cases selected with normal ACL versus complete tear. All others excluded (partial tear, mucoid degeneration)

Abbreviations: ACL, anterior cruciate ligament; AUC, area under the curve; FS, fat suppressed; ML, machine learning; MSK, musculoskeletal; PD, proton density.

low-grade partial tear, high-grade partial tear, or complete tear. Normal ACL and ACL with sprain, mucoid degeneration, or ganglion cysts were considered intact. The images used by the CNN included sagittal T2-weighted, coronal T1-weighted, and axial PD-weighted sequences. The reference standard was established through consensus between three fellowship-trained MSK radiologists.

The model achieved an AUC of 0.965, a sensitivity of 0.759, and a specificity of 0.968 for the identification of an ACL tear. The general radiologists achieved a sensitivity and specificity of 0.906 and 0.933. The model was statistically significantly less sensitive than general radiologists in the identification of an ACL tear. The authors also evaluated how using the model can affect the diagnostic performance of general radiologists with the diagnosis of ACL tears. When general radiologists used the algorithm, there was a 4.8% increased specificity for the identification of an ACL tear when compared with radiologist performance alone, which was statistically significant.[3]

Chang et al developed a CNN method that achieved a high level of accuracy for the diagnosis of complete ACL tears using a coronal PD-weighted sequence and a MSK radiologist's interpretation as the reference standard.[32] Cases demonstrating an ACL partial tear or mucoid degeneration were excluded. The performance of three CNNs were evaluated. The CNN with the highest performance used an initial localization network to crop the area of interest and also included dynamically sampled cropped patches of anatomy that did not include the ACL. The authors found that the diagnostic performance of the model improved with an increased number of input slices. The final model achieved a sensitivity and specificity of 1.00 and 0.933, respectively, for the identification of complete ACL tears.[32]

Liu et al developed a fully automated deep learning–based diagnosis system for the diagnosis of a complete ACL tear that achieved a similar level of specificity and sensitivity when compared with a cohort of radiologists with varying levels of training.[33] Their deep learning–based diagnosis system was composed of CNNs to (1) select the MR images containing the ACL, (2) isolate the intercondylar notch region containing the ACL, and (3) determine the presence of a tear. The images used by the ACL tear diagnosis system included sagittal PD-weighted and T2-weighted sequences. Using arthroscopic knee surgery reports as the reference standard, the ACL tear diagnosis system achieved a sensitivity and specificity of 0.96 and 0.96, and an AUC of 0.98. The clinical radiologists of varying levels of experience, ranging from radiology resident to fellowship-trained MSK radiologist, had a sensitivity and specificity of 0.96 to 0.98 and 0.90 to 0.98, respectively, in the diagnosis of a complete ACL tear. There was no statistically significant difference in the diagnostic performance between the ACL tear diagnosis system and radiologists in the diagnosis of a complete ACL tear.[33]

Meniscus

Meniscus Evaluation: MRI

The fibrocartilaginous meniscus is commonly injured, can lead to accelerated cartilage wear, and is frequently managed surgically with the rise in popularity of meniscus-preserving surgeries.[34,35] MRI remains the noninvasive modality of choice for the diagnosis of meniscal tears that are characterized by abnormal meniscal morphology and/or signal intensity.[36] Diagnostic performance of MRI when interpreted by radiologists in terms of sensitivity and specificity is 93% and 88% for medial meniscus tears and 79% and 96% for the lateral meniscus.[37] The literature reflects a long-standing interest in automatic segmentation and diagnosis of meniscus tears with several computer-assisted detection methods using texture analysis or supervised image classifiers published; however, no clinical applications have resulted to date.[38–43] Most recently, several novel CNNs for meniscus pathology detection and localization were developed and described in the literature (►**Table 3**).

Table 3 Studies investigating performance of CNNs in detection of meniscus tear[a]

Lesion	Study	Reference standard	Sequence	Results	Notes
WORMs score meniscus lesions	Pedoia et al[24]	MSK radiologist read	Three-dimensional FSE CUBE	Sens: 81.98% Spec: 89.81% AUC 0.89	WORMS categorizes intrasubstance degeneration as a lesion
No tear (degenerative signal, postoperative, normal) vs tear	Bien et al[3]	MSK radiology consensus read on a subset	Sagittal T2 Coronal T1 Axial PD	AUC: 0.847 Specificity: 0.741	Algorithm specificity for meniscal tear lower when compared with readers
Normal vs tear	Couteaux et al[44]	Annotated data set	Sagittal T2 single image	Weighted AUC: 0.906	Weighted AUC included presence/absence of tear, orientation, and location
Normal vs tear	Roblot et al[45]	Annotated data set	Sagittal T2 single image	AUC: 0.94	Presence/absence of tear, orientation, and location were assessed

Abbreviations: AUC, area under the curve; CNN, convolutional neural network; FSE, fast spin echo; MSK, musculoskeletal; PD, proton density; WORMS, Whole-organ Magnetic Resonance Imaging Score.
[a]Segmentation literature excluded.

Pedoia et al, in addition to evaluating for ACL tears, on the same data set, used their two-stage approach for binary detection of a meniscal "lesion" (present/absent) and then severity scoring of that lesion (mild/moderate versus severe) using the WORMs criteria.[24] Notably, this study included intrasubstance/degenerative meniscus signal abnormality as a "lesion" per the WORMs criteria. For binary meniscus lesion detection, the CNN achieved a sensitivity of 81.98% and a specificity of 89.81% with AUCs of 0.95, 0.84, and 0.89 on training, validation, and testing data sets, respectively.

Bien et al, in addition to using MRNet to detect ACL tears as described in the previous section, also classified menisci as intact (normal, degenerative, or postsurgical changes without tear) or torn (increased signal reaching the articular surface on at least two slices or morphology change), and the CNN performance was compared with that of radiologists. The reference standard was a radiologist consensus read on an internal validation set of 120 examinations from the 1,370-examination data set. The model achieved an AUC of 0.847 for meniscal tear, and the model's specificity for meniscal tear was lower than that of radiologists, 0.741 compared with 0.892.[3]

Couteaux et al used a CNN-based approach to classify menisci as "healthy" versus torn and to categorize the orientation and location of the meniscus tear if present. This was performed on an annotated data set consisting of sagittal-only single MR images manually cropped to include the meniscus. This approach yielded a weighted AUC of 0.906 for the three tasks (tear detection, orientation, and anatomical location).[44] Roblot et al also performed these three tasks but on a larger data set, also using a CNN-based approach, yielding a weighted AUC of 0.90.[45]

Tendons

Tendon injuries around the knee most commonly involve the extensor mechanism. Although no research currently addresses the use of ML in the diagnosis or management of tendon injuries around the knee, Kapiński et al reported the use of a CNN to assess the Achilles tendon. The CNN provided automatic quantitative assessment of Achilles tendon healing, classification of healthy versus injured tendon, and pathologic tissue localization through the analysis of MR images.[46]

Peripheral Nerves

Lower extremity neuropathies are common, and the diagnosis is frequently challenging, made through a combination of physical examination data as well as the results of electrodiagnostic testing and magnetic resonance neurography (MRN).[47] MRN analysis involves manual segmentation or manual 3D reconstruction and semiquantitative visual assessment of the peripheral nerve through the measurement of the cross-sectional area, and detection of morphological changes or signal intensity abnormalities that indicate nerve pathology. Balsiger et al developed a CNN to automatically segment the sciatic nerve through the tibial and common peroneal bifurcation at the knee using MRN images from healthy volunteers and those with diagnosed sciatic neuropathy.[48] This work represents an important initial step in automated peripheral nerve segmentation and quantitative analysis that potentially aid the radiologist in the diagnosis of peripheral neuropathies.

Musculoskeletal Ultrasound

CNNs have been developed to automate lesion classification, detection, and segmentation tasks with medical ultrasound. Early progress was demonstrated in thyroid nodule detection and classification, fetal biometry, breast lesion detection, and prostate cancer detection and grading, to name a few.[49] To date, no literature addresses the use of neural networks in MSK ultrasound. Potential applications around the knee might include the detection and classification of extensor mechanism injury, assessment of tendon healing, or quantitative analysis of knee joint effusions or synovitis.

Bone Tumors

Bone tumors commonly occur around the knee, particularly in the distal femur and proximal tibia, less frequently in the proximal fibula and patella.[50] Radiology has significant limitations in discrimination between malignant and benign tumors and predicting those at highest risk for recurrence. Thus automated or assisted diagnosis of bone tumors is attractive. The ability to integrate clinical information, pathology results, and risk factors could be very helpful in identifying patients at greatest risk for incidental bone tumors or for recurrence following treatment.

He et al developed a CNN to predict local recurrence of giant cell bone tumors of the proximal tibia or distal femur following curettage, using a combination of clinical data and presurgical noncontrast MR features. Their method outperformed radiologists, demonstrating greater accuracy and sensitivity in predicting tumors that recurred within 2 years of operative treatment in 56 patients.[51]

Fractures

Missed fractures in the emergent setting account for 41 to 80% of all diagnostic errors.[52] Lindsey et al targeted this shortcoming by developing a CNN that improved the accuracy of fracture detection when radiographs are interpreted by emergency medicine clinicians.[52] This study focused on wrist radiographs, and the reference standard was senior orthopaedic surgeon radiograph interpretation for binary fracture detection. On average, clinicians in this study demonstrated a relative reduction in misinterpretation rate of 47% when using the CNN. There are no current published studies using deep-learning methods for fracture detection around the knee. This is a potential area of research and may be particularly useful in detecting difficult to diagnose fractures or injury patterns including tibial plateau fractures, osteochondral fractures, stress fractures of the proximal tibia, and vertical patellar fractures. Other potential applications include identification of patients at highest risk for radiographically occult fractures based on patient characteristics, mechanism of injury, and bone mineralization.

Discussion

The current literature indicates AI performance similar to humans for the detection of cartilage lesions on MRI and less variability than humans in the grading of knee OA severity on radiographs. For ACL tears, AI performs well where humans do in determining full-thickness tear versus normal ACL, but it struggles in discriminating ACL anatomical variation, sprain, and mucoid degeneration from tear. In the evaluation of the meniscus, AI specificity for meniscal tear is lower than that of humans, higher specificity is reported when intrasubstance/degenerative meniscus signal abnormality, and tears are considered equivalent lesions.

Most of the published literature to date that explores interpretive applications of AI to the evaluation of knee pathology focuses on cartilage and OA. Investigations into the ability of AI systems to predict cartilage lesions (MRI) and stage of OA (radiographs) have shown early promise, with proof of concept established. This is not in small part due to the public availability of large annotated imaging data sets such as OAI and MOST. The potential benefits of AI in cartilage evaluation include increased speed of diagnosis, decreased costs associated with interpretation, and decrease in reader variability. The integration of fully automated OA severity grading on radiographs holds the potential to reduce reader fatigue by freeing up the radiologist for more complex or difficult to diagnose problems while interpreting knee radiographs such as the presence of subtle fractures, malignancy, or soft tissue abnormalities.

The incorporation of automatically applied objective grading systems in terms of cartilage wear on MRI and radiographic OA, rather than free-text verbal impressions, could reduce inter- and intrareader variability and potentially improve our process of tracking cartilage disease progression.[24] As neural network cartilage lesion detection becomes more streamlined and widely accessible, the opportunity to scale research and test multiple data sets will allow more robust studies to take place, specifically aimed at lesion detection and grading. Although important to initially establish AI system performance compared with human radiologists, it will be useful to make comparisons with a more robust reference standard with arthroscopic or surgical data and incorporate patient outcomes data. A combined AI system incorporating automated radiographic or MRI segmentation, detection, and staging eventually could be paired with an AI system incorporating clinical data[53,54] to provide more reliable outcome predictions for patients. Although we are not there yet, AI may soon play a useful role in the automated detection of cartilage lesions and in the distinction between early stages of OA, difficult tasks for radiologists in both MRI and radiography.

ACL tears are rarely a diagnostic challenge for the radiologist, and visual assessment of the ACL is not a particularly time-consuming task. In ACL evaluation, AI algorithms may be more helpful for prioritizing studies to be read or for assisting diagnosis of ACL tear when interpretation by an experienced radiologist is not immediately available.

Meniscus tears, depending on anatomical location, can be diagnostically challenging as reflected in the published numbers for MRI sensitivity and specificity. The few articles that have addressed AI applications for meniscal tear detection and classification have very disparate methods in terms of what is considered a meniscus tear, and the one study that directly compared the algorithms versus human performance found that radiologists outperformed the algorithm.

Current literature indicates the potential for AI algorithms to increase accuracy and efficiency in the evaluation of knee pathology. Many challenges remain, however, and more work on a larger scale needs to be done before a statement could be made on the practicality or reliability of such models in a modern clinical practice. The contents of the algorithms themselves as authored by individual research groups remain somewhat of a mystery, limiting the ability for the reproduction and validation of published results, a

process uncommon in the radiology literature. Additionally, the generalizability of these algorithms may be limited because the training occurs on very homogeneous data sets.

ML algorithms are ultra-specialized in the sense that each algorithm is designed for one very specific task, such as binary ACL classification as torn or intact. Diagnostic interpretation of a complete knee MRI for example, would require a litany of separate algorithms. These algorithms must be trained for the diagnosis of pathology that requires agreement on a reference standard because most MSK radiologic diagnoses lack a true gold standard. Some options include a consensus read of multiple radiologists to establish "truth" or a surgically proven lesion. All introduce an additional layer of complexity to labeling large data sets.

Finally, these algorithms currently require immense, ano-nymized, and usually annotated data sets of high-quality medical imaging. Only institutions/entities with the resources to build and manage these data sets will be able to achieve substantial forward progress unless there is a push to make data sets publicly available or multi-institutional collaboration is encouraged. Additional questions inherent to relying on the results of ML algorithms for medical decision making have yet to be fully addressed including issues surrounding medical liability, public perception, and trust in removing the human element from some aspects of medical image interpretation.[4]

Conclusions

We are at the epicenter of a research explosion in the arena of AI applications for medical image interpretation driven on the health care side by increased utilization of medical imaging and on the technology side by advances in AI algorithms and processing power. Although the potential is there, interpretive AI algorithms for the detection of knee pathology are currently single task oriented, have not yet delivered a clinical product, and significant limitations remain. Exciting future directions include AI-aided diagnostics, automated and standardized tracking of OA progression or injury healing, incorporation of clinical data into the image interpretation process, and the potential for AI to extract clinically important imaging features from MRI or radiographs that have yet to be defined.

Conflict of Interest
None declared.

References

1 Gage BE, McIlvain NM, Collins CL, Fields SK, Comstock RD. Epidemiology of 6.6 million knee injuries presenting to United States emergency departments from 1999 through 2008. Acad Emerg Med 2012;19(04):378–385

2 Gyftopoulos S, Harkey P, Hemingway J, Hughes DR, Rosenkrantz AB, Duszak R Jr. Changing musculoskeletal extremity imaging utilization from 1994 through 2013: a Medicare beneficiary perspective. AJR Am J Roentgenol 2017;209(05):1103–1109

3 Bien N, Rajpurkar P, Ball RL, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. PLoS Med 2018;15(11): e1002699

4 Hirschmann A, Cyriac J, Stieltjes B, Kober T, Richiardi J, Omoumi P. Artificial intelligence in musculoskeletal imaging: review of current literature, challenges, and trends. Semin Musculoskelet Radiol 2019;23(03):304–311

5 Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018;18(08): 500–510

6 Cross M, Smith E, Hoy D, et al. The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014;73(07):1323–1330

7 Norman B, Pedoia V, Majumdar S. Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. Radiology 2018;288(01):177–185

8 Lawrence RC, Felson DT, Helmick CG, et al; National Arthritis Data Workgroup. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum 2008;58(01):26–35

9 Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16(04):494–502

10 Gossec L, Jordan JM, Mazzuca SA, et al; OARSI-OMERACT Task Force Total Articular Replacement as Outcome Measure in OA. Comparative evaluation of three semi-quantitative radiographic grading techniques for knee osteoarthritis in terms of validity and reproducibility in 1759 X-rays: report of the OARSI-OMERACT task force. Osteoarthritis Cartilage 2008;16(07):742–748

11 Culvenor AG, Engen CN, Øiestad BE, Engebretsen L, Risberg MA. Defining the presence of radiographic knee osteoarthritis: a comparison between the Kellgren and Lawrence system and OARSI atlas criteria. Knee Surg Sports Traumatol Arthrosc 2015; 23(12):3532–3539

12 Dacre JE, Huskisson EC. The automatic assessment of knee radiographs in osteoarthritis using digital image analysis. Br J Rheumatol 1989;28(06):506–510

13 Antony J, McGuinness K, O'Connor NE, Moran K. Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. Abstract available at: https://arxiv-org/abs/1609.02469. Accessed July 7, 2019

14 Oka H, Muraki S, Akune T, et al. Fully automatic quantification of knee osteoarthritis severity on plain radiographs. Osteoarthritis Cartilage 2008;16(11):1300–1306

15 Shamir L, Ling SM, Scott W, Hochberg M, Ferrucci L, Goldberg IG. Early detection of radiographic knee osteoarthritis using computer-aided analysis. Osteoarthritis Cartilage 2009;17(10): 1307–1312

16 Shamir L, Ling SM, Scott WW Jr, et al. Knee x-ray image analysis method for automated detection of osteoarthritis. IEEE Trans Biomed Eng 2009;56(02):407–415

17 Antony J, McGuinness K, Moran K, O'Connor N. Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks. In: Perner P, ed. Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science. Vol 10358. Cham, Switzerland: Springer; 2017

18 Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. Sci Rep 2018;8(01):1727

19 Norman B, Pedoia V, Noworolski A, Link TM, Majumdar S. Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs. J Digit Imaging 2019;32(03):471–477

20 Abedin J, Antony J, McGuinness K, et al. Predicting knee osteoarthritis severity: comparative modeling based on patient's data and plain X-ray images. Sci Rep 2019;9(01):5761

21 Quatman CE, Hettrich CM, Schmitt LC, Spindler KP. The clinical utility and diagnostic performance of magnetic resonance imaging for identification of early and advanced knee osteoarthritis: a systematic review. Am J Sports Med 2011;39(07):1557–1568

22 Kijowski R, Blankenbaker DG, Munoz Del Rio A, Baer GS, Graf BK. Evaluation of the articular cartilage of the knee joint: value of adding a T2 mapping sequence to a routine MR imaging protocol. Radiology 2013;267(02):503–513

23 Liu F, Zhou Z, Samsonov A, et al. Deep learning approach for evaluating knee MR images: achieving high diagnostic performance for cartilage lesion detection. Radiology 2018;289(01):160–169

24 Pedoia V, Norman B, Mehany SN, Bucknor MD, Link TM, Majumdar S. 3D convolutional neural networks for detection and severity staging of meniscus and PFJ cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects. J Magn Reson Imaging 2019;49(02):400–410

25 Spindler KP, Wright RW. Clinical practice. Anterior cruciate ligament tear. N Engl J Med 2008;359(20):2135–2142

26 Lohmander LS, Englund PM, Dahl LL, Roos EM. The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. Am J Sports Med 2007;35(10):1756–1769

27 Fridén T, Zätterström R, Lindstrand A, Moritz U. Disability in anterior cruciate ligament insufficiency. An analysis of 19 untreated patients. Acta Orthop Scand 1990;61(02):131–135

28 Everhart JS, Kirven JC, Abouljoud MM, DiBartola AC, Kaeding CC, Flanigan DC. Effect of delayed primary anterior cruciate ligament reconstruction on medial compartment cartilage and meniscal health. Am J Sports Med 2019;47(08):1816–1824

29 Mink JH, Levy T, Crues JV III. Tears of the anterior cruciate ligament and menisci of the knee: MR imaging evaluation. Radiology 1988;167(03):769–774

30 Robertson PL, Schweitzer ME, Bartolozzi AR, Ugoni A. Anterior cruciate ligament tears: evaluation of multiple signs with MR imaging. Radiology 1994;193(03):829–834

31 Štajduhar I, Mamula M, Miletić D, Ünal G. Semi-automated detection of anterior cruciate ligament injury from MRI. Comput Methods Programs Biomed 2017;140:151–164

32 Chang PD, Wong TT, Rasiej MJ. Deep learning for detection of complete anterior cruciate ligament tear. J Digit Imaging 2019; March 11 (Epub ahead of print)

33 Liu F, Guan B, Zhou Z, et al. Fully automated diagnosis of anterior cruciate ligament tears on knee MR images by using deep learning. Radiol Artif Intell 2019;1(03). Doi: 10.1148/ryai. 2019180091

34 Zikria B, Hafezi-Nejad N, Roemer FW, Guermazi A, Demehri S. Meniscal surgery: risk of radiographic joint space narrowing progression and subsequent knee replacement-data from the Osteoarthritis Initiative. Radiology 2017;282(03):807–816

35 Recht MP, Kramer J. MR imaging of the postoperative knee: a pictorial essay. Radiographics 2002;22(04):765–774

36 De Smet AA. How I diagnose meniscal tears on knee MRI. AJR Am J Roentgenol 2012;199(03):481–499

37 Nguyen JC, De Smet AA, Graf BK, Rosas HG. MR imaging-based diagnosis and classification of meniscal tears. Radiographics 2014;34(04):981–999

38 Boniatis I, Panayiotakis G, Panagiotopoulos E. A computer-based system for the discrimination between normal and degenerated menisci from magnetic resonance images. Paper presented at: 2008 IEEE International Workshop on Imaging Systems and Techniques, September 10–12, 2008; Crete, Greece. Doi: 10.1109/IST.2008.4659996

39 Fu JC, Lin CC, Wang CN, Ou YK. Computer-aided diagnosis for knee meniscus tears in magnetic resonance imaging. J Ind Prod Eng 2013;30:67–77

40 Ramakrishna B, Liu W, Saiprasad G, et al. An automatic computer-aided detection system for meniscal tears on magnetic resonance images. IEEE Trans Med Imaging 2009;28(08):1308–1316

41 Köse C, Gençalioğlu O, Şevik U. An automatic diagnosis method for the knee meniscus tears in MR images. Expert Syst Appl 2009;36 (2, Part 1):1208–1216

42 Saygılı A, Albayrak S. An efficient and fast computer-aided method for fully automated diagnosis of meniscal tears from magnetic resonance images. Artif Intell Med 2019;97:118–130

43 Zarandi MHF, Khadangi A, Karimi F, Turksen IB. A computer-aided type-II fuzzy image processing for diagnosis of meniscus tear. J Digit Imaging 2016;29(06):677–695

44 Couteaux V, Si-Mohamed S, Nempont O, et al. Automatic knee meniscus tear detection and orientation classification with Mask-RCNN. Diagn Interv Imaging 2019;100(04):235–242

45 Roblot V, Giret Y, Bou Antoun M, et al. Artificial intelligence to diagnose meniscus tears on MRI. Diagn Interv Imaging 2019;100 (04):243–249

46 Kapiński N, Zieliński J, Borucki BA, et al. Monitoring of the Achilles tendon healing process: can artificial intelligence be helpful? Acta Bioeng Biomech 2019;21(01):103–111

47 Garwood ER, Duarte A, Bencardino JT. MR imaging of entrapment neuropathies of the lower extremity. Radiol Clin North Am 2018; 56(06):997–1012

48 Balsiger F, Steindel C, Arn M, et al. Segmentation of peripheral nerves from magnetic resonance neurography: a fully-automatic, deep learning-based approach. Front Neurol 2018;9:777

49 Liu S, Wang Y, Yang X, et al. Deep learning in medical ultrasound analysis: a review. Engineering 2019;5(02):261–275

50 Peleteiro-Pensado M, Barrientos-Ruiz I, Ortiz-Cruz EJ. Bone tumors around the knee. In: Rodríguez-Merchán EC, Liddle AD, eds. Joint Preservation in the Adult Knee. Cham, Switzerland;: Springer; 2017: 153–173

51 He Y, Guo J, Ding X, et al. Convolutional neural network to predict the local recurrence of giant cell tumor of bone after curettage based on pre-surgery magnetic resonance images. Eur Radiol 2019;29(10):5441–5451

52 Lindsey R, Daluiski A, Chopra S, et al. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A 2018;115(45):11591–11596

53 Joseph GB, McCulloch CE, Nevitt MC, et al. Tool for osteoarthritis risk prediction (TOARP) over 8 years using baseline clinical data, X-ray, and MRI: data from the Osteoarthritis Initiative. J Magn Reson Imaging 2018;47(06):1517–1526

54 Lim J, Kim J, Cheon S. A deep neural network-based method for early detection of osteoarthritis using statistical data. Int J Environ Res Public Health 2019;16(07):E1281