Advances in Clinical Radiology 1 (2019) 83–94
ADVANCES IN CLINICAL RADIOLOGY

Deep Learning in Musculoskeletal Imaging

Fang Liu, PhD*, Richard Kijowski, MD
Department of Radiology, University of Wisconsin School of Medicine and Public Health, 1111 Highland Avenue, Madison, WI 53705-2275, USA

KEYWORDS
• Deep learning • Machine learning • Medical imaging • Musculoskeletal imaging • Radiographs
• Magnetic resonance imaging • Computed tomography • Review

KEY POINTS
- Deep learning methods have been shown to be highly efficient and accurate for segmenting musculoskeletal tissues from medical images, which may eventually allow incorporation of quantitative image analysis into clinical practice.
- Deep learning methods have shown promising preliminary results for accelerating musculoskeletal magnetic resonance imaging.
- Deep learning methods have shown promising preliminary results for disease detection on medical images with diagnostic performance comparable to human readers for identifying a wide variety of musculoskeletal pathologic abnormalities.

INTRODUCTION

Deep learning is one form of machine learning that uses multiple levels of representation obtained by composing simple modules that each transforms the representation at one level into a representation at a higher, slightly more abstract level [1]. With the combination of enough such transformations, very complex functions can be learned. Conventional machine learning–based techniques have been used in medical imaging for decades [2–4]. However, applications of more advanced deep learning methods have recently become more feasible due to the vast increase in computational efficiency provided by newly developed computing hardware and advanced algorithm implementation.

One milestone event for deep learning was the 2012 ImageNet challenge [5], a benchmark of computer vision competition in which participants classified millions of natural images into discrete image categories (eg, cat, dog, car). A first deep learning–based entry called AlexNet achieved first place in the 2012 competition and exceeded the prior years' results that used conventional machine learning methods by a surprisingly large margin [6]. Ever since then, deep learning has revolutionized computer science, especially computer vision, and has quickly expanded into a much broader range of science and engineering disciplines, including medical imaging. In particular, recent surveys on deep learning in medical imaging have shown a wide variety of applications in different imaging subspecialties,

Disclosure Statement: The authors have no relationship with a commercial company that has a direct financial interest in subject matter or materials discussed in this article or with a company making a competing product.

*Corresponding author. Department of Radiology, Wisconsin Institutes for Medical Research, 1111 Highland Avenue, Madison, WI 53705-2275. *E-mail address:* fliu37@wisc.edu

https://doi.org/10.1016/j.yacr.2019.04.013
2589-8744/19/

including neuro, lung, abdomen, cancer, breast, and cardiac imaging [7,8]. However, applications in musculoskeletal imaging have remained relatively limited. This article reviews the current uses of deep learning in musculoskeletal imaging with a focus primarily on tissue segmentation, image reconstruction, and disease detection where great successes and breakthroughs have been recently achieved.

OVERVIEW OF DEEP LEARNING

Deep learning fits into a broader context of artificial intelligence. By definition, artificial intelligence can refer to any machine that performs tasks usually requiring human knowledge in a manner that resembles how humans solve problems. Machine learning, a subconcept under artificial intelligence, is more specific to techniques that enable a machine to learn features and functions from data without explicitly modeling the mechanism. Examples of conventional machine learning include regression, clustering, support vector machine, and random forest. Deep learning is one type of machine learning that uses a specific algorithm architecture, called a neural network, for processing information. The design of the neural network was initially inspired by the anatomy of neurons and how they function in the brain to perform complex cognitive tasks. Typically, multiple hidden layers with artificial neurons interconnecting with each other by "weights" in the neural network can represent the real interconnected neurons in the brain. Thus, deep learning usually refers to a deep structure with a large number of hidden layers. A unique feature of this architecture is that considerable flexibility can make it scale easily from a simple structure with just a few layers to an ultrasophisticated structure with hundreds of layers suitable for complex functions way beyond the capabilities of conventional machine learning techniques.

When this neural network is applied for computer vision tasks, such as image recognition, the input of image is usually treated as an array of voxels at different spatial locations and with different intensities. A process to characterize the intercorrelation among these voxels is called convolution, in which a set of the feature extracting "filters" is used to identify intensity variations, edges, and patterns at low-level layers. Next, new sets of filters at intermediate-level layers can then process the output from low-level features to construct composite information such as local shapes and patterns. This process repeats and continues to a high-level layer, in which multiple global features are combined to represent a unique set of patterns for one image. In image recognition, this set of patterns can then be used to characterize the image into different categories using predictors such as multiclass cross-entropy classifier. The convolutional neural network (CNN) with multiple convolutional layers is a widely used modern neural network architecture in computer vision and subsequently in medical imaging. One common component in the CNN design is the activation function, in which, similar to real neuron activation, a nonlinearity is added to the output of the convolutional layer to increase model complexity. Although a large variety of nonlinear functions was proposed, researchers found the rectified linear activation (ReLU) [9] is typically the easiest and most efficient to implement. A pooling (eg, Max-pooling) or unpooling process also can be performed in the network to downscale or upscale the size of the features to maintain only important image information while removing redundancy.

The CNN typically needs to be trained in a supervised manner in which a large number of paired image cases and reference labels are divided into several groups for training (with usually a fraction of 50%∼60%), validation (10%∼20%), and testing (20%∼30%). The training data are used to update the weights of the filters in the convolutional layers through an error back-propagation procedure (eg, stochastic gradient descent [10] after comparing the CNN estimates against the reference labels with a goal that the estimates should approximate the reference). Because the training process is performed iteratively through all images in the training group, the CNN estimation on the validation group determines how many total iterations (epochs) provide an optimal training performance to stop the training process. This estimation on data other than the training group ensures that the training can be generalized to other datasets without the model being overly adapted to the training data (ie, overfitting). Finally, once the training is complete, the model (eg, filter weights) of the CNN is fixed, and the testing group is used to assess the model accuracy in the real-world scenario. The CNN model can be trained from scratch from random numbers, or it can be initialized by training the same CNN with a set of data other than the task-specific one. This, referred to as transfer learning, is usually applied in deep learning studies using small training datasets, which is typical of the case in medical imaging. Because the networks could have millions of potential weights for some applications, transfer learning is able to reduce the demands for abundant training data and improve the model accuracy with high training efficiency [11,12]. In addition, unsupervised or weakly supervised training also can be applied for training CNNs

for applications in which collecting the paired data and reference labels is challenging, expensive, or not possible.

Deep learning reflects a generic image analysis pipeline using highly flexible, scalable, and efficient deep CNN architectures and training strategies to perform a wide variety of imaging applications. Due to the current availability of highly efficient and well-evaluated CNNs in computer vison and other domains, medical imaging researchers are less hindered by the challenges of algorithm design and tend to primarily focus on how to tailor specific imaging applications into the deep learning framework for optimal performance. For medical imaging applications, and in particular for musculoskeletal imaging, performance breakthrough has occurred in 3 major areas, including tissue segmentation, image reconstruction, and disease detection.

TISSUE SEGMENTATION

Segmentation of tissue structure on medical images is a fundamental step in many medical imaging–based applications. Traditionally, this is done by having experienced users scroll through many 2-dimensional (2D) or 3-dimensional (3D) image slices and manually identify specific tissues from adjacent structures. However, manual segmentation is exceptionally time-consuming, generally requiring several hours for complete segmentation, and its repeatability and accuracy are influenced by many factors, including the level of human expertise, the image quality, and the applied segmentation tools [13,14]. Thus, there has been much interest in the past decade for developing fully automated techniques for segmenting medical images [15]. The recent implementation of deep learning techniques in image processing has been shown to have significant impacts on medical image segmentation [7,8]. Deep CNN-based methods have achieved state-of-the-art performance in many medical image segmentation tasks for a variety of tissue structures throughout the body on computed tomography (CT) and magnetic resonance (MR) images. With the rapid advance of deep learning techniques, a few preliminary studies have described CNN applications for segmenting musculoskeletal tissues.

In a pilot study, Prasoon and colleagues [16] proposed to segment the tibial cartilage on knee MR image datasets using a deep learning–based voxel classification method. A voxel was identified into cartilage or background by combining the results of 3 simple CNNs, each of which was applied in a separate coronal, sagittal, and axial plane. These triplanar CNNs were trained on 25 knee image datasets and evaluated on 114 knee image datasets and achieved a high Dice coefficient of 0.82 for segmenting tibial cartilage. In another study, Liu and colleagues [17] implemented a more efficient CNN for performing cartilage and bone segmentation on multicontrast knee MR image datasets (Fig. 1). This network featured a pair of convolutional encoder and decoder networks, which was previously demonstrated to be highly accurate and efficient for several dense image segmentation tasks in computer vision [18,19]. Liu and colleagues [17] evaluated the network for segmenting femoral and tibial cartilage and bone on 3 clinical knee image datasets with different tissue contrasts, including sagittal T1-weighted 3D spoiled gradient recalled-echo images in 100 subjects, sagittal fat-suppressed proton density–weighted 3D fast spin-echo images in 60 subjects, and sagittal multiecho spin-echo T2 mapping images in 100 subjects. The results showed a significant improvement in segmentation accuracy compared with conventional state-of-the-art model or atlas-based segmentation methods, with the deep learning approach also providing a dramatically reduced segmentation time in the range of 20 seconds to 4 minutes depending on the image dataset. In another study performed by the same group [20], the convolutional encoder-decoder network was further combined with spatial image postprocessing algorithm to simultaneously segment 12 different knee tissue types on sagittal fat-suppressed proton density–weighted 3D fast spin-echo image datasets of 20 subjects. The deep learning method produced a high segmentation accuracy with mean Dice coefficient above 0.90 for femur, tibia, muscle, and other nonspecified tissues, and mean Dice coefficient ranging between 0.80 and 0.90 for femoral cartilage, tibial cartilage, patella, patellar cartilage, meniscus, quadriceps, and patellar tendons, and infrapatellar fat pad. Only joint effusion and Baker cyst had a Dice coefficient below 0.80 (Fig. 2). Besides the segmentation of knee joint structures, a recent study by Deniz and colleagues [21] evaluated a convolutional encoder-decoder for segmenting the proximal femur from hip MR image datasets and reported a high Dice coefficient of 0.95 in a cohort of 86 subjects.

A few studies have investigated the use of deep learning–based segmentation for the analysis of pathologic group differences in large clinical cohorts including subjects in the Osteoarthritis Initiative (OAI). Norman and colleagues [22] performed an analysis using a convolutional encoder-decoder network for segmenting cartilage and meniscus on knee MR image datasets from 2 large subject cohorts that included 3D

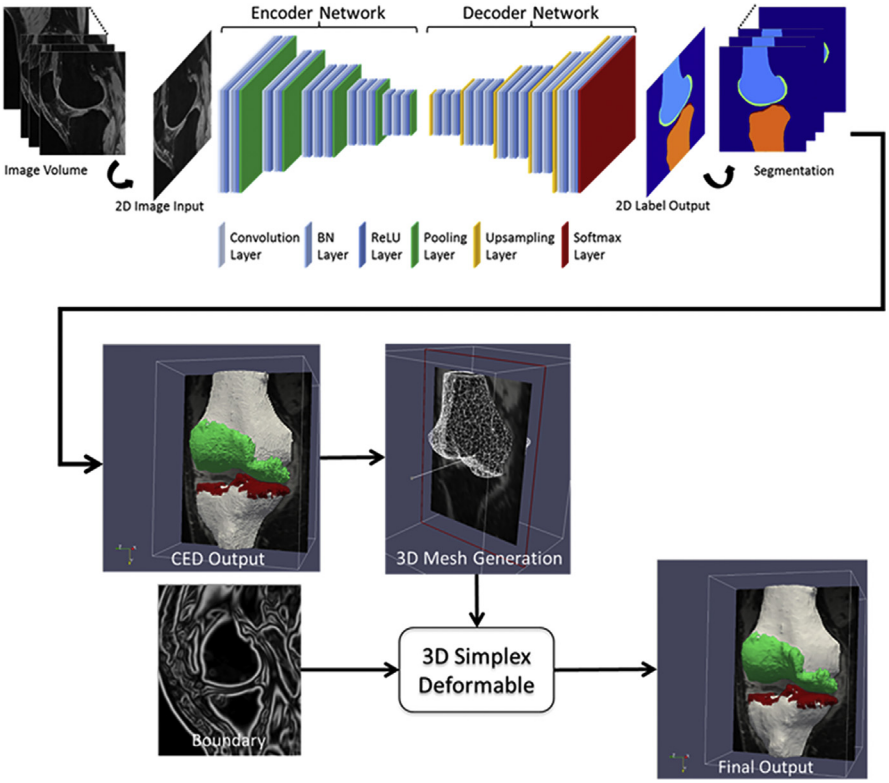

FIG. 1 An illustration of a deep learning–based segmentation framework combining a convolutional encoder-decoder network with a 3D simplex-mesh deformable process for bone and cartilage segmentation. The convolutional encoder-decoder contains an encoder network and a decoder network. The output of the network is high-resolution pixel-wise tissue classification, which is further processed with a 3D deformable process and regularized with image intensity–based tissue boundaries to produce an accurate and desirable tissue surface. BN, Batch Normalization; CED, Convolutional Encoder-Decoder. (*Adapted from* Liu F, Zhou Z, Jang H, et al. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. Magn Reson Med 2018;79(4):2379–2391; with permission.)

spoiled gradient recalled-echo images with T1ρ and T2 weighting and 3D dual-echo steady-state images. They found a strong Dice coefficient for 3D dual-echo steady-state images ranging between 0.770 and 0.878 for segmenting different cartilage compartments and 0.809 and 0.753 for segmenting the lateral and medial meniscus, respectively. They also reported a strong correlation between manual and deep learning–based cartilage relaxometry measurements with an average correlation coefficient of 0.823 and 0.860 for T1ρ and T2 values, respectively. In another study, Tack and colleagues [23] evaluated a deep learning method using a convolutional encoder-decoder in combination with statistical shape modeling for segmenting meniscus on 3D dual-echo steady-state image datasets of 88 subjects in the OAI. The deep learning method provided Dice coefficients of 0.838 and 0.889 for segmenting the medial and lateral meniscus at baseline, respectively, and 0.831 and 0.883 at 12-month follow-up, respectively. They also found that medial meniscal extrusion quantified at baseline using the deep learning method in 600 subjects in the OAI was significantly greater ($P<.05$) for subjects with osteoarthritis than subjects without osteoarthritis.

Unsupervised learning methods also have demonstrated initial promising results for musculoskeletal tissue segmentation. A recent study published by Liu [24] showed that a deep learning image-to-image translation method could segment multicontrast MR images using a single set of annotated training images with unsupervised learning. The results demonstrated that by training only with segmentation labels from

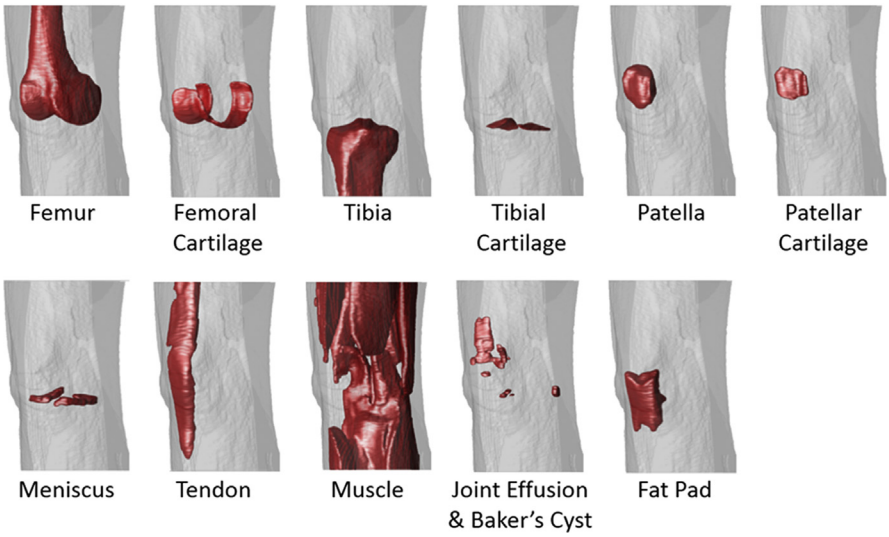

FIG. 2 Examples of 3D-rendered models for all knee joint structures in a subject with knee osteoarthritis created using segmented tissue masks from fat-suppressed proton density–weighted 3D fast spin-echo images. (*From* Zhou Z, Zhao G, Kijowski R, et al. Deep convolutional neural network for segmentation of knee joint anatomy. Magn Reson Med 2018;80(6):2759–2770; with permission.)

T1-weighted 3D spoiled gradient recall-echo images of 60 subjects, the unsupervised learning algorithm could achieve high segmentation performance for proton density–weighted fast spin-echo images with mean Dice coefficients of 0.95 and 0.65 for bone and cartilage, respectively and for fat-suppressed T2-weighted fast spin-echo images with mean Dice coefficients of 0.93 and 0.75 for bone and cartilage, respectively. The use of unsupervised learning would allow for efficient and accurate musculoskeletal tissue segmentation of MR images with any tissue contrast using a single annotated training dataset.

Recent studies focusing primarily on the knee joint have demonstrated the feasibility of using deep learning methods for efficient and accurate segmentation of musculoskeletal tissues. However, extension of these deep learning techniques to segment tissues in other anatomic regions such as the spine, hand, hip, and ankle is also of great interest. Furthermore, additional work is needed to create a user-friendly platform for fully automated deep learning segmentation methods, which may eventually allow incorporation of quantitative image analysis into clinical practice.

IMAGE RECONSTRUCTION

In recent years, there has been a rapid interest growth for applying deep learning in the reconstruction of medical imaging including dose reduction CT imaging [25], rapid MR imaging [26–28], and multimodality image synthesis [29–31]. In particular, many pioneer works in MR image reconstruction have shown that deep learning–based methods can yield better image quality with dramatically decreased acquisition time in comparison with conventional state-of-the-art reconstruction methods including parallel imaging [32–34] and compressed sensing [35].

Hammernik and colleagues [26] presented a deep learning method with a Variational Network to optimize the compressed sensing reconstruction framework. The Variation Network provided a scheme to learn reconstruction parameters of the compressed sensing formulation using training image datasets. Experiments were performed on 5 clinical knee MR image datasets acquired on 20 subjects in 3 acquisition planes using different signal-to-noise ratios and 2 image contrasts. The Variation Network was found to outperform standard reconstruction algorithms on both quantitative and qualitative image analysis at an acceleration factor up to 4. A similar study by Mardani and colleagues [27] also focused on re-implementing the compressed sensing framework using deep learning. The study incorporated a generative adversarial neural network [36] into the deep learning reconstruction method and evaluated the method on different body regions including 19 knee MR image

datasets acquired with an axial proton density–weighted fast spin-echo sequence. The deep learning method was found to outperform standard reconstruction methods with better reconstruction quality and sharper image appearance at an acceleration factor up to 5. Chaudhari and colleagues [28] proposed to accelerate musculoskeletal imaging using a deep learning–based super-resolution technique, which restored the missing slice information of low-resolution thick-slice images to produce visually plausible high-resolution thin-slice images. Using 3D dual-echo steady-state image datasets with 0.7-mm slice thickness of 124 subjects in the OAI for training, the super-resolution algorithm was able to convert down-sampled thick-slice images into a simulated image with 0.7-mm slice thickness (Fig. 3). The super-resolution algorithm was found to outperform a conventional interpolation-based method and a state-of-the-art sparse-coding super-resolution method on both quantitative and qualitative image analysis.

Although recent studies have demonstrated the feasibility of using deep learning methods for image reconstruction, further work using large image datasets is needed to investigate the clinical utility and diagnostic performance of these new techniques for different image contrasts and different anatomic regions. Future research also is needed to investigate the use of deep learning reconstruction methods for accelerating multidimensional and quantitative MR imaging, which may provide useful clinical information, but typically require substantially increased scan time.

DISEASE DETECTION

The use of deep learning methods for disease detection would be especially important for musculoskeletal radiologists, as it could potentially be used in clinical practice for identifying pathologic abnormalities on medical images. The use of machines to aid human readers in the interpretation of medical imaging studies could maximize diagnostic performance while reducing subjectivity and errors due to distraction and fatigue. Machines also could improve workflow by providing an initial interpretation of imaging studies and identify cases with abnormal findings requiring an immediate interpretation by human readers.

Deep learning methods have been used to evaluate hand and wrist radiographs. Larson and colleagues [37] used 14,036 anterior-posterior hand radiographs to train and evaluate a neural network to estimate pediatric bone age. The root mean square and mean absolute difference between the bone age estimates provided by the machine and the bone age estimates provided by experienced human readers was 0.63 years and 0.50 years, respectively. Kim and MacKinnon [38] used lateral wrist radiographs of 695 subjects with and without fracture to train and evaluate a neural network to detect distal radius fractures using the interpretation of an experienced human reader as the reference standard. The machine was found to have an area under the curve (AUC) of 0.954 for detecting distal radius fractures with a sensitivity and specificity of 90% and 88%, respectively.

Deep learning methods have been used to grade the severity of osteoarthritis on anterior-posterior knee

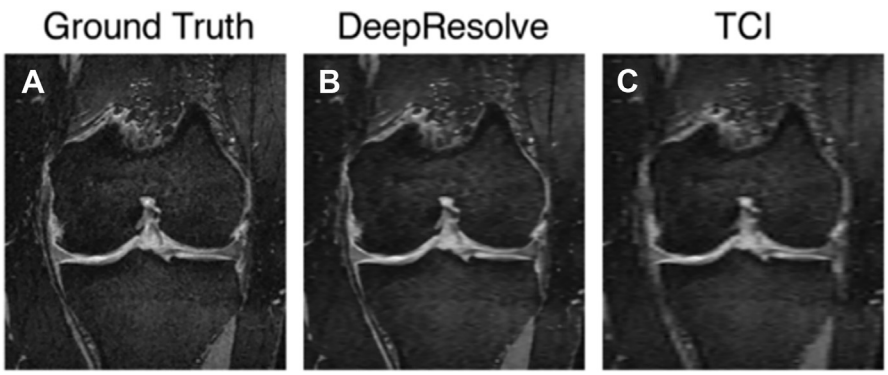

FIG. 3 (**A**) Coronal ground-truth image, (**B**) coronal DeepResolve (deep learning method) image, and (**C**) coronal TCI (conventional interpolation method) image of the knee in a subject with knee osteoarthritis. The deep learning image appears more similar to the high-resolution ground-truth image than the conventional interpolation image. TCI, Tricubic Interpolated. (*From* Chaudhari AS, Fang Z, Kogan F, et al. Super-resolution musculoskeletal MRI using deep learning. Magn Reson Med 2018;80(5):2139–2154; with permission.)

radiographs. Radiographs from more than 6000 subjects in the OAI were used to train and evaluate the neural networks using Kellgren-Lawrence (KL) grades provided by experienced human readers as the reference standard. Tiulpin and colleagues [39] found the machine to have an average multiclass accuracy of 66.7% for assigning a KL grade with an AUC of 0.930 for determining the presence or absence of radiographic osteoarthritis. Norman and colleagues [40] found the machine to have sensitivity values ranging between 68.9% and 86.0% for assigning KL grade 3 and grade 4, respectively, and specificity values ranging between 83.8% and 99.1% for assigning KL grades 2 and 4, respectively (Fig. 4). In a smaller study, Xue and colleagues [41] used 420 anterior-posterior pelvic radiographs to train and evaluate a neural network to detect hip osteoarthritis using the interpretation of an experienced human reader as the reference standard. The machine was found to have an AUC of 0.940 for determining the presence or absence of radiographic osteoarthritis with a sensitivity and specificity of 95.0% and 90.7%, respectively.

Deep learning methods have been used to detect spine fractures on CT images using the interpretations of experienced human readers as the reference standard. Using segmented spine images from CT scans of 100 subjects with fracture and 60 subjects without fracture to train and evaluate a neural network, Raghavendra and colleagues [42] found the machine to have 99% sensitivity and 100% specificity for detecting vertebral body fractures. Tomita and colleagues [43] used segmented spine images from CT scans of 713 subjects with fracture and 719 subjects without fracture to train and evaluate a neural network and found the machine to have 85.2% sensitivity and 95.8% specificity for detecting vertebral body fractures. However, both studies only classified the presence or absence of a fracture for the entire spine, which raises questions regarding the ability of the deep learning methods to localize the exact site of injury. Roth and colleagues [44] segmented the posterior elements of the spine from axial CT images of 18 subjects with 55 displaced posterior element fractures and 5 subjects without spine fracture to train and evaluate a neural network. The machine was found to have 71% sensitivity and 81% specificity for detecting posterior element fractures.

Deep learning methods have been used to localize vertebral body metastases on CT and MR images of the spine using metastatic lesions outlined by experienced human readers as the reference standard. Using segmented spine images from CT scans of 49 patients with 539 sclerotic lesions and 5 control subjects without metastatic disease to train and evaluate a neural network, Roth and colleagues [45] found the machine to have an AUC of 0.834 for localizing large metastatic lesions more than 3 cm in diameter with 9.5 false positives per subject at a sensitivity of 90.0%. Chmelik and colleagues [46] used segmented spine images from CT scans of 31 patients with 1046 lytic lesions and 1135 sclerotic lesions to train and evaluate a neural network.

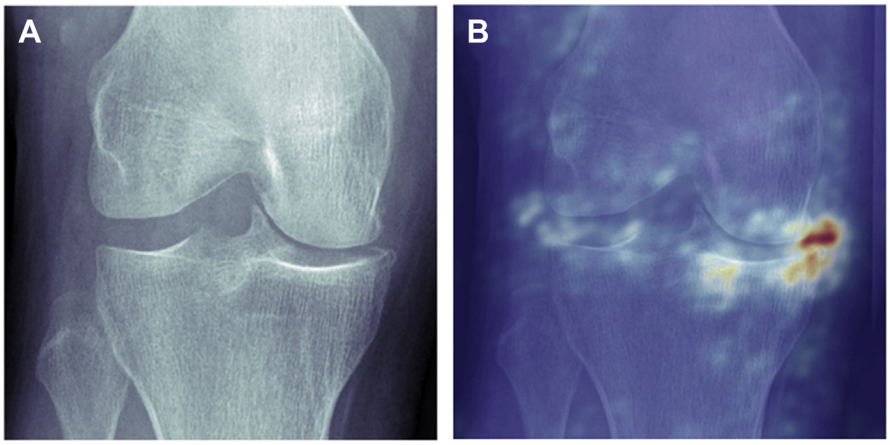

FIG. 4 (**A**) Anterior-posterior knee radiograph for which both the machine and the human reader assigned a KL grade 2 osteoarthritis. (**B**) Corresponding knee probability map showing the areas on the radiograph (*yellow* and *red color*) that the machine used to make its interpretation of the severity of knee osteoarthritis. (*From* Norman B, Pedoia V, Noworolski A, et al. Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs. J Digit Imaging 2018; with permission.)

The machine was found to have 45.6 false positives per subject at 92.0% sensitivity for localizing metastatic lesions smaller than 1.5 mm in diameter and 5.9 false positives per subject at 99.0% sensitivity for localizing large metastatic lesions more than 3 cm in diameter. Wang and colleagues [47] used nonsegmented sagittal fat-suppressed T2-weighted fast spin-echo MR images of the spine from 26 patients with metastatic disease to evaluate and train a neural network and found the machine to have 0.4 false positives per subject at 90.0% sensitivity for localizing metastatic lesions. However, the number and sizes of the metastatic lesions, which would influence diagnostic performance, was not specified in the study.

Deep learning methods have been used to detect a spectrum of pathologic abnormalities of the spine on MR images. Jamaludin and colleagues [48] used the segmented vertebral body and disk subunits on sagittal T2-weighted fast spin-echo images of the lumbar spine from 2009 subjects to train and evaluate a neural network to assign a Pfirrmann grade and disc height narrowing score and to determine the presence or absence of vertebral body endplate defects and bone marrow edema lesions. The interrater agreement between the machine and an experienced human reader was found to range between 70.4% for assigning a Pfirrmann grade and 92.5% for determining the presence or absence of vertebral body bone marrow edema lesions. Kim and colleagues [49] used nonsegmented axial T2-weighted fast spin-echo images in 80 patients with tuberculous spondylitis and 80 patients with pyogenic spondylitis to train and evaluate a neural network to distinguish between the different types of spinal infection. Using biopsy as the reference standard, the machine was found to have an AUC of 0.802 and sensitivity and specificity of 85.0% and 67.9%, respectively, for distinguishing between tuberculous and pyogenic spondylitis.

Deep learning methods have been used to detect internal derangement of the knee joint on MR images using the interpretation of experienced human readers as the reference standard. Liu and colleagues [50] used segmented cartilage on sagittal fat-suppressed T2-weighted fast spin-echo images in 175 subjects with 2642 cartilage lesions to train and evaluate a neural network to determine the presence or absence of cartilage lesions in approximately 100 regions of interest placed on the articular surface of the femur and tibia. The machine had a good intraobserver agreement between 2 individual evaluations with a kappa value of 0.76 in the kappa statistical analysis. The machine was found to have an AUC of 0.917 for detecting cartilage lesions with a sensitivity and specificity of 84.1% and 85.2%, respectively, for evaluation 1 and an AUC of 0.914 for detecting cartilage lesions with a sensitivity and specificity of 80.5% and 87.9%, respectively, for evaluation 2 (Figs. 5 and 6). Pedoia and colleagues [51] used segmented meniscus on sagittal fat-suppressed proton density–weighted 3D fast spin-echo images in 1478 subjects with 670 meniscus tears to train and evaluate a neural network to determine the presence or absence of meniscus tears (Fig. 7). The machine was found to have an AUC of 0.890 for detecting meniscus tears with sensitivity and specificity of 90% and 82%, respectively. Bien and colleagues [52] used nonsegmented axial fat-suppressed proton density–weighted fast spin-echo, coronal T1-weighted fast spin-echo, and sagittal fat-suppressed T2-weighted fast spin-echo images in 1370 subjects with 319 anterior cruciate ligament tears and 508 meniscus tears to

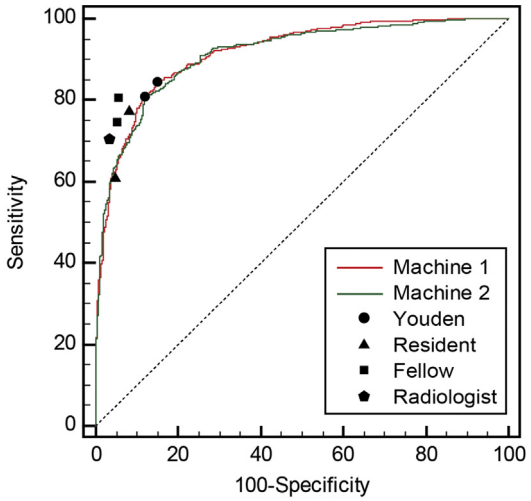

FIG. 5 Receiver operator characteristic curves showing the diagnostic performance of a deep learning method for detecting cartilage lesions within the knee joint on fat-suppressed T2-weighted fast spin-echo images using the interpretation of an experienced human reader as the reference standard. The AUC for the machine were 0.917 and 0.914 for evaluation 1 and 2, respectively, both indicating high overall diagnostic accuracy. Note that the sensitivity and specificity of the machine at the optimal threshold of the Youden index was comparable to the sensitivity and specificity of multiple human readers with varying levels of clinical experience. (*From* 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(1):160–169; with permission.)

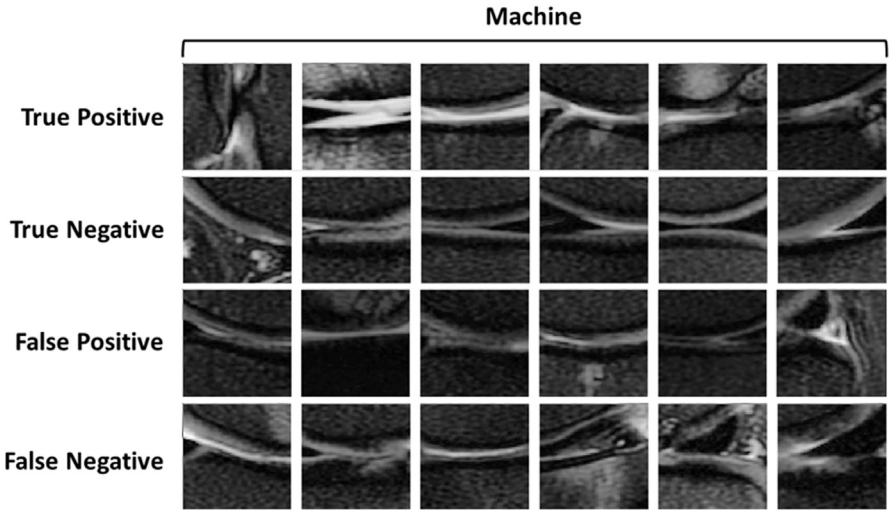

FIG. 6 Multiple examples of true positive, true negative, false positives, and false negative interpretations of the deep learning method for detecting cartilage lesions within the knee joint on fat-suppressed T2-weighted fast spin-echo images using the interpretation of an experienced human reader as the reference standard.

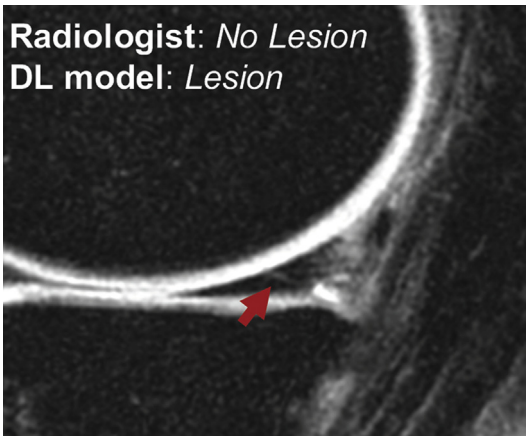

FIG. 7 Example of a false-positive interpretation of the deep learning method for detecting a meniscus tear within the knee joint on fat-suppressed proton density–weighted 3D fast spin-echo images using the interpretation of an experienced human reader as the reference standard. Note the intermediate signal in the meniscus (*red arrow*), which the machine interpreted as extending into the articular surface representing a tear and the human reading interpreted as not extending into the articular surface representing intrasubstance degeneration. (*From* Pedoia V, Norman B, Mehany SN, et al. 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(2):400–410; with permission.)

train and evaluate a neural network to determine the presence or absence of anterior cruciate ligament and meniscus tears. The machine was found to have a sensitivity and specificity of 75.9% and 96.8%, respectively, for detecting anterior cruciate ligament tears and a sensitivity and specificity of 71.0% and 74.1%, respectively, for detecting meniscus tears.

Recent studies have documented the feasibility of using deep learning methods for musculoskeletal disease detection with diagnostic performance comparable to human readers. However, additional technical development is needed to create fully automated deep learning techniques for reliable and repeatable interpretation of musculoskeletal imaging studies. Furthermore, the diagnostic performance of these deep learning methods must be further evaluated in prospective studies using large image datasets acquired at different institutions with different imaging parameters and different imaging hardware.

SUMMARY

There have been many recently published studies describing promising applications of deep learning in musculoskeletal imaging. Deep learning techniques have been shown to be highly efficient and accurate for segmenting musculoskeletal tissues and for accelerating musculoskeletal MR imaging. More importantly, multiple studies have documented the

feasibility of using deep learning methods for musculoskeletal disease detection with diagnostic performance comparable to human readers for identifying a wide variety of pathologic abnormalities. However, much additional work is needed in all applications of deep learning in musculoskeletal imaging before these techniques can be incorporated in clinical practice.

REFERENCES

[1] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44. Available at: https://www.ncbi.nlm.nih.gov/pubmed/26017442. Accessed August 18, 2017.

[2] Wernick MN, Yang Y, Brankov JG, et al. Machine learning in medical imaging. IEEE Signal Process Mag 2010;27(4):25–38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25382956. Accessed December 11, 2018.

[3] Giger ML. Machine learning in medical imaging. J Am Coll Radiol 2018;15(3):512–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29398494. Accessed December 11, 2018.

[4] Erickson BJ, Korfiatis P, Akkus Z, et al. Machine learning for medical imaging. Radiographics 2017;37(2):505–15. Available at: http://pubs.rsna.org/doi/10.1148/rg.2017160130. Accessed December 11, 2018.

[5] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015;115(3):211–52. Available at: http://arxiv.org/abs/1409.0575. Accessed April 13, 2018.

[6] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012;25:1097–105. Available at: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[7] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.

[8] Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017;19(1):221. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28301734. Accessed April 18, 2018.

[9] Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. Proc 27th Int Conf Mach Learn. Haifa, June 21–24, 2010. p. 807–814.

[10] Bottou L. Large-scale machine learning with stochastic gradient descent. 19th Int Conf Comput Stat Heidelberg. Paris, August 22–27, 2010. p. 177–186.

[11] Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging 2016;35(5):1299–312. Available at: http://ieeexplore.ieee.org/document/7426826/. Accessed March 7, 2018.

[12] Shin H-C, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: cnn architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016;35(5):1285–98. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26886976. Accessed December 14, 2018.

[13] McWalter EJ, Wirth W, Siebert M, et al. Use of novel interactive input devices for segmentation of articular cartilage from magnetic resonance images. Osteoarthr Cartil 2005;13(1):48–53. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1063458404002092. Accessed March 11, 2017.

[14] Shim H, Chang S, Tao C, et al. Knee cartilage: efficient and reproducible segmentation on high-spatial-resolution MR images with the semiautomated graph-cut algorithm method. Radiology 2009;251(2):548–56. Available at: http://pubs.rsna.org/doi/10.1148/radiol.2512081332. Accessed March 7, 2018.

[15] Sharma N, Aggarwal LM. Automated medical image segmentation techniques. J Med Phys 2010;35(1):3–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20177565. Accessed March 7, 2018.

[16] Prasoon A, Petersen K, Igel C, et al. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. Med Image Comput Comput Assist Interv 2013;16(Pt2):246–53, Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). Available at: http://www.ncbi.nlm.nih.gov/pubmed/24579147. Accessed September 17, 2017.

[17] Liu F, Zhou Z, Jang H, et al. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. Magn Reson Med 2017. https://doi.org/10.1002/mrm.26841. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28733975. Accessed July 29, 2017.

[18] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. ArXiv e-prints. arXiv preprint arXiv:1511.00561. 2015. Available at: http://arxiv.org/abs/1511.00561. Accessed February 21, 2017.

[19] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, et al, editors. Med image comput comput interv – MICCAI 2015 18th Int Conf Munich, ger Oct 5-9, 2015, proceedings, part III. Cham (Switzerland): Springer International Publishing; 2015. p. 234–41. https://doi.org/10.1007/978-3-319-24574-4_28.

[20] Zhou Z, Zhao G, Kijowski R, et al. Deep convolutional neural network for segmentation of knee joint anatomy. Magn Reson Med 2018. https://doi.org/10.1002/mrm.27229.

[21] Deniz CM, Xiang S, Hallyburton RS, et al. Segmentation of the proximal femur from MR images using deep convolutional neural networks. Sci Rep 2018;8(1):16485. Available at: http://www.nature.com/articles/s41598-018-34817-6. Accessed December 10, 2018.

[22] 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;172322. Available at: http://pubs.rsna.org/doi/10.1148/radiol.2018172322. Accessed April 12, 2018.

[23] Tack A, Mukhopadhyay A, Zachow S. Knee menisci segmentation using convolutional neural networks: data from the Osteoarthritis Initiative. Osteoarthr Cartilage 2018;26(5):680–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29526784. Accessed December 10, 2018.

[24] Liu F. SUSAN: segment unannotated image structure using adversarial network. Magn Reson Med 2018;81(5):3330–45. Available at: http://doi.wiley.com/10.1002/mrm.27627. Accessed December 10, 2018.

[25] Chen H, Zhang Y, Zhang W, et al. Low-dose CT via convolutional neural network. Biomed Opt Express 2017;8(2):679–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28270976. Accessed December 11, 2018.

[26] Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. Magn Reson Med 2017;79(6):3055–71, Wiley-Blackwell; Available at: http://arxiv.org/abs/1704.00447. Accessed April 29, 2017.

[27] Mardani M, Gong E, Cheng JY, et al. Deep generative adversarial neural networks for compressive sensing (GANCS) MRI. IEEE Trans Med Imaging 2018;38(1):167–79. Available at: https://ieeexplore.ieee.org/document/8417964/. Accessed July 31, 2018.

[28] Chaudhari AS, Fang Z, Kogan F, et al. Super-resolution musculoskeletal MRI using deep learning. Magn Reson Med 2018;80(5):2139–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29582464. Accessed July 2, 2018.

[29] Liu F, Jang H, Kijowski R, et al. Deep learning MR imaging–based attenuation correction for PET/MR imaging. Radiology 2017;286:676–84. Available at: http://pubs.rsna.org/doi/10.1148/radiol.2017170700. Accessed September 26, 2017.

[30] Leynes AP, Yang J, Wiesinger F, et al. Direct PseudoCT generation for Pelvis PET/MRI attenuation correction using deep convolutional neural networks with multiparametric MRI: zero echo-time and Dixon deep pseudoCT (ZeDD-CT). J Nucl Med 2017;59(5):852–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29084824. Accessed November 3, 2017.

[31] Liu F, Jang H, Kijowski R, et al. A deep learning approach for 18F-FDG PET attenuation correction. EJNMMI Phys 2018;5(1):24. Available at: https://ejnmmiphys.springeropen.com/articles/10.1186/s40658-018-0225-8. Accessed December 11, 2018.

[32] Pruessmann KP, Weiger M, Scheidegger MB, et al. SENSE: sensitivity encoding for fast MRI. Magn Reson Med 1999;42(5):952–62.

[33] Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 2002;47(6):1202–10.

[34] Sodickson DK, Manning WJ. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. Magn Reson Med 1997;38(4):591–603. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9324327. Accessed July 2, 2018.

[35] Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn Reson Med 2007;58(6):1182–95.

[36] Goodfellow IJI, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. arXiv Prepr arXiv …. 2014; 1–9. Available at: https://arxiv.org/pdf/1406.2661.pdf. Accessed July 4, 2017.

[37] Larson DB, Chen MC, Lungren MP, et al. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. Radiology 2018;287(1):313–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29095675. Accessed December 10, 2018.

[38] Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol 2018;73(5):439–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29269036. Accessed December 10, 2018.

[39] Tiulpin A, Thevenot J, Rahtu E, et al. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. Sci Rep 2018;8(1):1727. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29379060. Accessed December 10, 2018.

[40] Norman B, Pedoia V, Noworolski A, et al. Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs. J Digit Imaging 2019;32(3):471–7.

[41] Xue Y, Zhang R, Deng Y, et al. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. PLoS One 2017;12(6):e0178992. Available at: http://dx.plos.org/10.1371/journal.pone.0178992. Accessed December 12, 2017.

[42] Raghavendra U, Bhat NS, Gudigar A, et al. Automated system for the detection of thoracolumbar fractures using a CNN architecture. Future Gener Comput Syst 2018;85:184–9. Available at: https://www.sciencedirect.com/science/article/pii/S0167739X17321544. Accessed December 10, 2018.

[43] Tomita N, Cheung YY, Hassanpour S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. Comput Biol Med 2018;98:8–15. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29758455. Accessed December 10, 2018.

[44] Roth HR, Wang Y, Yao J, et al. Deep convolutional networks for automated detection of posterior-element fractures on spine CT. In: Tourassi GD, Armato SG, editors 2016;. 2016. p. 97850P. Available at: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2217146; 2016. Accessed December 10, 2018.

[45] Roth HR, Yao J, Lu L, et al. Detection of sclerotic spine metastases via random Aggregation of deep convolutional neural network classifications. Cham

(Switzerland): Springer; 2015. p. 3–12. Available at: http://link.springer.com/10.1007/978-3-319-14148-0_1. Accessed December 10, 2018.

[46] Chmelik J, Jakubicek R, Walek P, et al. Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. Med Image Anal 2018;49:76–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/30114549. Accessed December 10, 2018.

[47] Wang J, Fang Z, Lang N, et al. A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks. Comput Biol Med 2017;84:137–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28364643. Accessed September 15, 2018.

[48] Jamaludin A, Kadir T, Zisserman A. SpineNet: automated classification and evidence visualization in spinal MRIs. Med Image Anal 2017;41:63–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28756059. Accessed December 12, 2017.

[49] Kim K, Kim S, Lee YH, et al. Performance of the deep convolutional neural network based magnetic resonance image scoring algorithm for differentiating between tuberculous and pyogenic spondylitis. Sci Rep 2018;8(1):13124, Nat Publishing Group; Available at: http://www.nature.com/articles/s41598-018-31486-3. Accessed December 10, 2018.

[50] 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;172986. Available at: http://pubs.rsna.org/doi/10.1148/radiol.2018172986. Accessed July 31, 2018.

[51] Pedoia V, Norman B, Mehany SN, et al. 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 2018;49:400–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/30306701. Accessed December 10, 2018.

[52] Bien N, Rajpurkar P, Ball RL, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. Saria S, editor. PLoS Med 2018;15(11):e1002699. Available at: http://www.ncbi.nlm.nih.gov/pubmed/30481176. Accessed December 10, 2018.