Fully Automated Diagnosis of Anterior Cruciate Ligament Tears on Knee MR Images by Using Deep Learning

Fang Liu, PhD* • Bochen Guan, BS* • Zhaoye Zhou, PhD • Alexey Samsonov, PhD • Humberto Rosas, MD • Kevin Lian, MD • Ruchi Sharma, MD • Andrew Kanarek, MD • John Kim, MD • Ali Guermazi, MD • Richard Kijowski, MD

From the Department of Radiology, University of Wisconsin School of Medicine and Public Health, 600 Highland Avenue, Madison, WI 53705 (F.L., B.G., A.S., H.R., K.L., R.S., A.K., J.K., R.K.); Department of Electrical and Computer Engineering, University of Wisconsin School of Engineering, Madison, Wis (B.G.); Department of Biomedical Engineering, University of Minnesota, Minneapolis, Minn (Z.Z.); and Department of Radiology, Boston University School of Medicine, Boston, Mass (A.G.). Received December 16, 2018; revision requested February 14, 2019; final revision received March 26; accepted April 4. Address correspondence to F.L. (e-mail: flu37@wisc.edu).

Funding support for the research project was provided by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (R01-AR068373) and by the National Institutes of Health (R01EB027087).

* F.L. and B.G. contributed equally to this work.

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2019; 1(3):e180091 • <https://doi.org/10.1148/ryai.2019180091> • Content codes: MK MR IN

Purpose: To investigate the feasibility of using a deep learning–based approach to detect an anterior cruciate ligament (ACL) tear within the knee joint at MRI by using arthroscopy as the reference standard.

Materials and Methods: A fully automated deep learning–based diagnosis system was developed by using two deep convolutional neural networks (CNNs) to isolate the ACL on MR images followed by a classification CNN to detect structural abnormalities within the isolated ligament. With institutional review board approval, sagittal proton density–weighted and fat-suppressed T2-weighted fast spin-echo MR images of the knee in 175 subjects with a full-thickness ACL tear (98 male subjects and 77 female subjects; average age, 27.5 years) and 175 subjects with an intact ACL (100 male subjects and 75 female subjects; average age, 39.4 years) were retrospectively analyzed by using the deep learning approach. Sensitivity and specificity of the ACL tear detection system and five clinical radiologists for detecting an ACL tear were determined by using arthroscopic results as the reference standard. Receiver operating characteristic (ROC) analysis and two-sided exact binomial tests were used to further assess diagnostic performance.

Results: The sensitivity and specificity of the ACL tear detection system at the optimal threshold were 0.96 and 0.96, respectively. In comparison, the sensitivity of the clinical radiologists ranged between 0.96 and 0.98, while the specificity ranged between 0.90 and 0.98. There was no statistically significant difference in diagnostic performance between the ACL tear detection system and clinical radiologists at $P < .05$. The area under the ROC curve for the ACL tear detection system was 0.98, indicating high overall diagnostic accuracy.

Conclusion: There was no significant difference between the diagnostic performance of the ACL tear detection system and clinical radiologists for determining the presence or absence of an ACL tear at MRI.

© RSNA, 2019

Supplemental material is available for this article.

Deep learning has been successfully used for various medical imaging applications, including image reconstruction (1–3), tissue and lesion segmentation (4–7), and disease detection and characterization (8–12). The use of deep learning for disease detection would be especially important for clinical radiologists because it could maximize diagnostic performance while reducing subjectivity and errors due to distraction and fatigue. Most prior work on the use of deep learning for disease detection has focused on identifying abnormalities such as masses on breast mammograms (13,14); lung nodules, opacities, and cardiomegaly on chest radiographs (9,15); and lung nodules and interstitial lung disease at chest CT (16–18). The use of deep learning for disease detection at MRI has been less investigated and may be particularly challenging, as it often requires analyzing complex abnormalities on multiple sections of different image datasets (8,11,12).

Anterior cruciate ligament (ACL) tears are a common sports-related musculoskeletal injury (19). Detecting an ACL tear relies on evaluating an obliquely oriented structure on multiple image sections with different tissue contrasts using a combination of MRI findings including fiber discontinuity, changes in contour, and signal abnormality within the injured ligament (20,21). Investigation of the ability of a deep learning approach to detect an ACL tear would be useful to determine whether deep learning could aid in the diagnosis of complex musculoskeletal abnormalities at MRI. We have developed a fully automated system for detecting an ACL tear by utilizing two deep convolutional neural networks (CNNs) to isolate the ACL on MR images followed by a classification CNN to detect structural abnormalities within the isolated ligament. This study was performed to investigate the feasibility of using the deep learning–based approach to detect a full-thickness

Abbreviations

ACL = anterior cruciate ligament, AUC = area under the ROC curve, CI = confidence interval, CNN = convolutional neural network, DenseNet = Densely Connected Convolutional Networks, ROC = receiver operating characteristic, VGG16 = Very Deep Convolutional Networks for Large-Scale Image Recognition

Summary

There was no statistically significant difference between the anterior cruciate ligament (ACL) tear detection system and clinical radiologists with varying levels of experience for determining the presence or absence of a full-thickness ACL tear using sagittal proton density-weighted and fat-suppressed T2-weighted fast spin-echo MR images.

Key Points

- There was no significant difference between the diagnostic performance of a fully automated deep learning–based diagnosis system and clinical radiologists for detecting a full-thickness anterior cruciate ligament (ACL) tear at MRI.
- Sensitivity and specificity of the ACL tear detection system at the optimal threshold were 0.96 and 0.96, respectively; the sensitivity of the clinical radiologists ranged between 0.96 and 0.98 and specificity ranged between 0.90 and 0.98.

ACL tear within the knee joint at MRI by using arthroscopy as the reference standard.

Materials and Methods

Fully Automated ACL Tear Detection System

The proposed deep learning–based ACL tear detection system consisted of three separate CNNs. The first CNN selected the image sections that contained the ACL from the entire MR image dataset. The second CNN isolated the region of the intercondylar notch that contained the ACL on the selected image sections to narrow the range of information used for subsequent image recognition. The third classification CNN evaluated the isolated ACL on the selected image sections to determine the presence or absence of an ACL tear. The three CNNs were connected in a cascaded fashion to create a fully automated processing pipeline, as shown in Figure 1. The detailed structure of the CNNs is summarized in Table 1. The processing pipeline framework was implemented in a hybrid computing environment involving Python (version 2.7; Python Software Foundation, Wilmington, Del) and MATLAB (version 2016a; MathWorks, Natick, Mass). The CNNs were coded by using TensorFlow (version 1.04; Google, Mountain View, Calif).

The first section-selection CNN was adapted from LeNet-5, which was originally proposed for natural image analysis (23) but has recently been used for many medical imaging applications (24,25). This CNN, shown in the top of Figure 1, had two sets of convolutional layers, which were followed by two fully connected layers and then a SoftMax classifier. The CNN was modified to allow an input image size of 448×448 with the images normalized to a range of 0 and 1 with respect to the maximum MRI signal. Two output classes were defined: One class for image sections that

contained the ACL and the other for image sections that did not contain the ACL.

The second ligament isolation CNN was adapted from You Only Look Once, or YOLO, which has demonstrated top performance in many image object detection challenges (26). This CNN, shown in the middle of Figure 1, had 24 convolutional layers followed by two fully connected layers. The convolutional layers were used to extract useful image features from the selected image sections that contained the ACL. The fully connected layers then used the extracted image features to predict the coordinates of a rectangular box that defined the region of the intercondylar notch that contained the ACL. The cropped images of the predefined rectangle box that contained the ACL on the selected image sections were downsized to 112×112 and were used as input images to the classification CNN.

The classification CNN was adapted from Densely Connected Convolutional Networks (DenseNet) (27), which has demonstrated top performance for image recognition in the Large Scale Visual Recognition Challenge image dataset (28,29). This CNN, shown in the bottom of Figure 1, contained three dense blocks, each of which was connected by a convolutional layer and a maximum pooling layer. Dense connectivity was applied to enable efficient information sharing across the layers, which substantially reduced the network parameters needed for high-throughput training. For the output, a final probability score for the presence or absence of an ACL tear on the cropped images that contained the ACL was obtained from a global average pooling (GAP) layer that followed the dense blocks. The classification result was computed by the mean of the classification probability of an ACL tear on all image sections that contained the ACL. The GAP layer was further modified by using a gradient back-propagation method (30) to calculate a diagnosis probability map matching the input image size, which showed the pixel-by-pixel probability of the presence of an ACL tear. To compare with other popular CNNs used in previous image classification applications (8,9), two additional classification CNNs adapted from the commonly used image recognition networks Very Deep Convolutional Networks for Large-Scale Image Recognition (VGG16) (28) and AlexNet (31) were also evaluated and compared with the proposed CNN adapted from DenseNet.

MR Image Datasets

This retrospective study was performed in compliance with Health Insurance Portability and Accountability Act regulations, with approval from our institutional review board, and with a waiver of the need to obtain informed consent. MR image datasets were obtained in 175 subjects with a surgically confirmed ACL tear (98 male subjects and 77 female subjects; average age, 27.5 years; age range, 16–47 years) and 175 subjects with a surgically confirmed intact ACL (100 male subjects and 75 female subjects; average age, 39.4 years; age range, 17–51 years). All subjects underwent an MRI examination of the knee and subsequent arthroscopic knee surgery at our institution between December 15, 2010, and October 15, 2017. All MRI examinations were performed with the same 3.0-T imaging unit (Signa Excite

HDx; GE Healthcare, Wauke-
 sha, Wis) by using an eight-
 channel phased-array extremity
 coil (Invivo, Orlando, Fla) and
 consisted of a standard axial
 fat-suppressed T2-weighted
 fast spin-echo sequence, stan-
 dard sagittal proton density-
 weighted and fat-suppressed
 T2-weighted fast spin-echo se-
 quences, and standard coronal
 proton density-weighted and
 fat-suppressed proton density-
 weighted fast spin-echo se-
 quences. The imaging param-
 eters of all sequences in the MRI
 examination are summarized in
 Table 2. Subjects with and
 those without an ACL tear were
 selected by a fellowship-trained
 musculoskeletal radiologist
 with 15 years of clinical experi-
 ence (R.K.) using a medical re-
 cord archiving system (Hyper-
 space; Epic, Verona, Wis). The
 radiologist reviewed the sched-
 ule of arthroscopic knee surger-
 ies performed at our institution
 and determined whether the
 patient undergoing arthroscopy
 had undergone an MRI
 examination of the same knee
 performed on the same 3.0-T
 imaging unit within 4 months
 prior to surgery.

Arthroscopic Knee Surgery

All subjects underwent ar-
 throscopic knee surgery within
 4 months after their MRI ex-
 amination (mean time interval,
 38.5 days; range, 7–116 days).
 All arthroscopic knee surgeries
 were performed by one of three orthopedic sports medicine
 specialists at our institution who have between 10 and 25 years
 of clinical experience and who were aware of the findings of
 the MRI examination at the time of arthroscopy. The ortho-
 pedic surgeons at our institution routinely inspect the ACL
 both visually and with a surgical probe during arthroscopy to
 determine the presence or absence of an ACL tear and record
 the information in their surgical report. A fellowship-trained
 musculoskeletal radiologist with 15 years of clinical experience
 (R.K.) reviewed the surgical reports of all subjects using a med-
 ical record archiving system (Hyperspace; Epic) and recorded
 the presence or absence of an ACL tear. All subjects with an
 ACL tear were found to have full-thickness ligament injuries
 at arthroscopy.

Figure 1: Illustration of the convolutional neural network (CNN) architecture for the deep learning-based anterior cruciate ligament (ACL) tear detection system. The proposed method consisted of three separate CNNs connected in a cascaded fashion to create a fully automated image processing pipeline. BN = batch normalization, Conv2D = 2D convolution, ReLU = rectified linear activation, 2D = two-dimensional.

Training and Evaluation of the ACL Tear Detection System

Training and evaluation of the ACL tear detection system were performed on a desktop computer running a 64-bit Linux operating system (Ubuntu 16.04) with an Intel Core i7-7700 K CPU with 32 GB DDR4 RAM and two Nvidia GeForce GTX 1080 graphic cards (Nvidia driver 384.130) with 2560 CUDA cores and 8 GB GDDR5 RAM. Training of the system was performed for each CNN individually as described in Appendix E1 (supplement). However, once the training was completed for all three CNNs, the ACL tear detection system operated as a fully automated end-to-end network. To train the section-selection CNN and the ligament isolation CNN, R.K. manually identified the MR im-

Table 1: Detailed Network Structure for the Section-detection, ACL Isolation, and Classification CNNs of the Deep Learning-based ACL Tear Detection System

Section-detection CNN	Ligament Isolation CNN	Classification CNN
Input (gray-scale image)	Input (gray-scale image)	Input (gray-scale image)
Conv2D (6 7×7 filters), BN, ReLU	Conv2D (64 7×7 filters), BN, ReLU	Conv2D (32 7×7 filters)
MaxPool (window size 2×2)	MaxPool (window size 2×2)	MaxPool (window size 2×2)
Conv2D (16 5×5 filters), BN, ReLU	Conv2D (192 3×3 filters), BN, ReLU	Dense block: {Conv2D (1 $\times 1$ filters), BN, ReLU; Conv2D (3 $\times 3$ filters), BN, ReLU} $\times 6$
MaxPool (window size 2×2)	MaxPool (window size 2×2)	Conv2D (64 1×1 filters), BN, ReLU
FC (84 nodes)	Conv2D (128 1×1 filters), BN, ReLU Conv2D (256 3×3 filters), BN, ReLU Conv2D (256 1×1 filters), BN, ReLU Conv2D (512 3×3 filters), BN, ReLU	MaxPool (window size 2×2)
FC (2 nodes)	MaxPool (window size 2×2)	Dense block: {Conv2D (1 $\times 1$ filters), BN, ReLU; Conv2D (3 $\times 3$ filters), BN, ReLU} $\times 12$
SoftMax (2 classes)	{Conv2D (256 1×1 filters), BN, ReLU Conv2D (512 3×3 filters), BN, ReLU} $\times 4$ Conv2D (512 1×1 filters), BN, ReLU Conv2D (1024 3×3 filters), BN, ReLU	Conv2D (128 1×1 filters), BN, ReLU
	MaxPool (window size 2×2)	MaxPool (window size 2×2)
	{Conv2D (512 1×1 filters), BN, ReLU Conv2D (1024 3×3 filters), BN, ReLU} $\times 2$ {Conv2D (1024 3×3 filters), BN, ReLU} $\times 4$	Dense Block: {Conv2D (1 $\times 1$ filters), BN, ReLU; Conv2D (3 $\times 3$ filters), BN, ReLU} $\times 24$
	FC (4096 nodes)	GlobalAveragePool
	Tensor (7 $\times 7 \times 11$)	SoftMax (2 classes)

Note.—The convolutional neural networks (CNNs) were built by using the listed layers from the top to the bottom. Definitions of the different network structures and layers of a CNN have been described in a previously published review article (22). ACL = anterior cruciate ligament, BN = batch normalization, Conv2D = two-dimensional convolution, FC = fully connected layer, GlobalAveragePool = global average pooling, MaxPool = maximum pooling, ReLU = rectified linear activation.

age sections that contained the ACL and then manually outlined the region of the intercondylar notch that contained the ACL on the selected MR image sections. However, once the training was completed, the ACL tear detection system provided fully automated isolation of the ACL on all MR image sections to be used as inputs into the classification CNN.

Subjects with and those without a surgically confirmed ACL tear were randomly stratified into three non-overlapping groups for training, validation, and hold-out testing. The training dataset, which consisted of MR images from 100 subjects with an ACL tear and 100 subjects with an intact ACL, was used to iteratively optimize the CNN parameters to learn image features. The validation dataset, which consisted of MR images from 25 subjects with an ACL tear and 25 subjects with an intact ACL, was used to select the most optimal model during the training process. The hold-out testing dataset, which consisted of MR images from 50 subjects with an ACL tear and 50 subjects with an intact ACL, was used for final evaluation of the optimal model to avoid training overfitting and to ensure that the image features learned during the training process could be generalized to new image data.

Evaluation of Clinical Radiologists

So that we could compare the diagnostic performance of the ACL tear detection system with that of clinical radiologists, a fellowship-trained musculoskeletal radiologist with 13 years of clinical experience (H.R.), a musculoskeletal radiology fellow (K.L.), a 2nd-year radiology resident (J.K.), a 3rd-year radiology resident (A.K.), and a 4th-year radiology resident (R.S.) independently reviewed the MRI examinations of all 100 subjects in the hold-out test dataset by using a picture archiving and communication system workstation (McKesson, San Francisco, Calif). The clinical radiologists, who were blinded to the findings of the arthroscopic knee surgeries, used all five sequences in the MRI examination together to determine the presence or absence of an ACL tear. The radiologists received no formal training or calibration session prior to reviewing the MRI studies.

Statistical Analysis

Statistical analysis was performed by using MATLAB (version 2013a; MathWorks) and MedCalc (version 14.8; MedCalc Software, Ostend, Belgium), with $P < .05$ considered to indicate a significant difference. Contingency tables and sensitivity and specificity for determining the presence or absence of

Table 2: Parameters for the Imaging Sequences in the MRI Examination

Parameter	Axial Fat-Saturated T2-weighted FSE	Coronal Fat-Saturated Proton Density-weighted FSE	Coronal Proton Density-weighted FSE	Sagittal Proton Density-weighted FSE	Sagittal Fat-Saturated T2-weighted FSE
Repetition time (msec)	4300	1800	2000	2000	5300
Echo time (msec)	77	20	20	20	80
Flip angle (degrees)	90	90	90	90	90
Acquisition matrix size	448 × 224	384 × 224	384 × 224	384 × 224	384 × 224
Reconstruction matrix size	512 × 512	512 × 512	512 × 512	512 × 512	512 × 512
Field of view (cm)	18	14	14	14	14
Section thickness (mm)	3	2	2	2	3
Bandwidth (±kHz)	41.7	31.2	31.2	31.2	41.7
Echo train length	21	4	4	4	20
Signal averages	4	2	2	2	3
File type	DICOM	DICOM	DICOM	DICOM	DICOM
Bit depth (bit)	16	16	16	16	16
Imaging time (min)	3:30	3:25	3:26	3:26	3:16

Note.—DICOM = Digital Information and Communications in Medicine, FSE = fast spin echo.

Table 3: Sensitivity, Specificity, and AUCs for the Proposed ACL Tear Detection System with the Classification CNN Adapted from DenseNet and the Alternative ACL Tear Detection Systems

CNN	Sensitivity	Specificity	AUC
DenseNet	0.96 (0.89, 1.00)	0.96 (0.86, 1.00)	0.98 (0.93, 1.00)
VGG16	0.92 (0.82, 0.98)	0.92 (0.82, 0.98)	0.95 (0.90, 0.99)
AlexNet	0.89 (0.76, 0.96)	0.88 (0.76, 0.96)	0.90 (0.83, 0.96)

Note.—The alternative classification convolutional neural networks (CNNs) were adapted from Very Deep Convolutional Networks for Large-Scale Image Recognition (VGG16) and AlexNet for determining the presence or absence of an anterior cruciate ligament (ACL) tear in the hold-out test dataset. Data in parentheses are 95% confidence intervals. AUC = area under the receiver operating characteristic curve, DenseNet = Densely Connected Convolutional Networks.

an ACL tear using arthroscopy as the reference standard for the hold-out test dataset were determined for the musculoskeletal radiologist, the musculoskeletal radiology fellow, the radiology residents, the proposed ACL tear detection system with the classification CNN adapted from DenseNet, and alternative ACL tear detection systems with the classification CNNs adapted from VGG16 and AlexNet. Receiver operating characteristic (ROC) analysis was used to further evaluate the diagnostic performance of the proposed and alternative ACL tear detection systems, with areas under the ROC curves (AUCs) compared by using a nonparametric approach (32). The Youden index was used to determine the optimal sensitivity and specificity of the proposed and alternative ACL tear detection systems (33). Two-sided exact binomial tests were used to calculate 95% confidence intervals (CIs) for the sensitivity and specificity of the proposed and alternative ACL tear detection systems and the clinical radiologists and for the AUCs of the proposed and alternative ACL tear detection systems. Statistically significant differences between the sensitivity and specificity of the proposed ACL tear detection system and the clinical radiologists was defined as the sensitivity and specificity point estimates of the clinical radiologists located outside the 95% CIs of the AUC for the machine (10).

Results

The training times for the section selection, ligament isolation, and classification CNNs were 0.82, 5.11, and 5.70 hours, respectively, in the training datasets. However, the average time for the ACL tear detection system to provide an interpretation of the presence or absence of an ACL tear for one subject was only 9 seconds using the trained networks.

Table 3 compares the sensitivity, specificity, and AUC for determining the presence or absence of a surgically confirmed ACL tear for the proposed ACL tear detection system with the classification CNN adapted from DenseNet and the alternative ACL tear detection systems with the classification CNNs adapted from VGG16 and AlexNet. All classification CNNs performed well, with point estimates of the sensitivity and specificity ranging between 0.89 and 0.96 and between 0.88 and 0.96, respectively, and AUCs ranging between 0.90 and 0.98. However, the proposed ACL tear detection system with the classification CNN adapted from DenseNet had the highest overall diagnostic performance for detecting an ACL tear.

Tables 4 and 5 show the contingency tables and sensitivity and specificity, respectively, for the musculoskeletal radiologist, musculoskeletal radiology fellow, radiology residents,

and proposed ACL tear detection system for determining the presence or absence of a surgically confirmed ACL tear. The point estimates of the sensitivity and specificity of the proposed ACL tear detection system at the optimal threshold of the Youden index were 0.96 and 0.96, respectively. In comparison, the point estimates of the sensitivity of the clinical radiologists ranged between 0.96 and 0.98, while the point estimates of the specificity ranged between 0.90 and 0.98. In particular, the point estimate of both the sensitivity and specificity of the fellowship-trained musculoskeletal radiologist was 0.98. The proposed ACL tear detection system could create dense probability maps that showed the pixel-by-pixel probability of the presence of an ACL tear (Figs 2, 3). For all 100 subjects in the hold-out test dataset, the machine had only two false-negative interpretations of the absence of an ACL tear (Fig 4) and two false-positive interpretations of the presence of an ACL tear (Fig 5).

Figure 6 shows the ROC curve describing the diagnostic performance of the proposed ACL tear detection system for determining the presence or absence of a surgically confirmed ACL tear. The AUC for the ACL tear detection system was 0.98 (95% CI: 0.93, 1.00; $P < .001$). For comparison, the point estimates representing the sensitivity and specificity of the musculoskeletal radiologist, musculoskeletal radiology fellows, and radiology residents for determining the presence or absence of an ACL tear were plotted on the figure and were in close proximity to the ROC curve of the ACL tear detection system. All sensitivity and specificity point estimates of the clinical radiologists were located inside the 95% CIs of the AUC for the proposed ACL tear detection system, indicating no statistically significant differences in diagnostic performance between the clinical radiologists and the machine.

Discussion

Our study has demonstrated the feasibility of using a deep learning-based approach to detect a full-thickness ACL tear within the knee joint at MRI by using arthroscopy as the reference standard. The ACL tear detection system achieved high diagnostic performance for determining the presence or absence of an ACL tear, with an AUC of 0.98. Furthermore, there was no statistically significant difference between the sensitivity and specificity of the machine and those of clinical radiologists with varying levels of experience for detecting an ACL tear. The ACL tear detection system was also at a disadvantage because it evaluated only a limited number of sagittal proton density-weighted and fat-suppressed T2-weighted fast spin-echo images, while the clinical radiologists were able to evaluate all image sections from a total of five different MRI se-

quences. Furthermore, the clinical radiologists were able to use ancillary findings of an ACL tear, such as the presence of bone marrow edema, on MR image sections not evaluated by the machine, which could have increased their suspicion for an ACL tear (21,34).

Table 4: Contingency Tables for the Musculoskeletal Radiologist, Musculoskeletal Radiology Fellow, Radiology Residents, and Proposed ACL Tear Detection System for Determining the Presence or Absence of an ACL Tear on the Hold-out Test Dataset

Reader and Result	Disease Present	Disease Absent	Total
Radiologist			
Test positive	49	1	50
Test negative	1	49	50
Total	50	50	
Fellow			
Test positive	48	1	49
Test negative	2	49	51
Total	50	50	
Resident 1			
Test positive	48	5	53
Test negative	2	45	47
Total	50	50	
Resident 2			
Test positive	48	1	49
Test negative	2	49	51
Total	50	50	
Resident 3			
Test positive	49	1	50
Test negative	1	49	50
Total	50	50	
Machine			
Test positive	48	2	50
Test negative	2	48	50
Total	50	50	

Table 5: Sensitivity and Specificity for the Musculoskeletal Radiologist, Musculoskeletal Radiology Fellow, Radiology Residents, and Proposed ACL Tear Detection System for Determining the Presence or Absence of an ACL Tear on the Hold-out Test Dataset

Reader	Sensitivity	Specificity	AUC
Radiologist	0.98 (0.88, 1.00)	0.98 (0.88, 1.00)	0.98 (0.95, 1.00)
Fellow	0.96 (0.86, 1.00)	0.98 (0.88, 1.00)	0.98 (0.95, 1.00)
Resident 1	0.96 (0.86, 1.00)	0.90 (0.78, 0.97)	0.93 (0.88, 0.98)
Resident 2	0.96 (0.86, 1.00)	0.98 (0.88, 1.00)	0.97 (0.94, 1.00)
Resident 3	0.98 (0.88, 1.00)	0.98 (0.88, 1.00)	0.98 (0.95, 1.00)
Machine	0.96 (0.89, 1.00)	0.96 (0.86, 1.00)	0.98 (0.93, 1.00)

Note.—Data in parentheses are 95% confidence intervals. ACL = anterior cruciate ligament, AUC = area under the receiver operating characteristic curve.

To our knowledge, only one previous study has described the diagnostic performance of a deep learning–based approach to detect an ACL tear within the knee joint at MRI. Bien et al (12) 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 to train and evaluate a deep learning method to determine the presence or absence of an ACL tear, with the interpretations of experienced musculoskeletal radiologists as the reference standard. That deep learning–based approach had a sensitivity and specificity of 0.76 and 0.97, respectively, for detecting 319 ACL tears, with the sensitivity of the machine being significantly lower ($P = .002$) than the sensitivity of the clinical radiologists. Our ACL tear detection system achieved a higher sensitivity for detecting an ACL tear than the deep learning method described by Bien et al, despite using a much smaller training dataset. This was likely because our ACL tear detection system first isolated the ACL on the MR images, which narrowed the range of information used for subsequent image recognition and allowed the classification CNN to focus exclusively on the ACL during the training and evaluation processes.

Our study investigated three different types of classification CNNs, including DenseNet, VGG16, and AlexNet, and found that DenseNet provided the best diagnostic performance for detecting an ACL tear. AlexNet has a very simple network architecture, which was likely the cause of its lower diagnostic performance. VGG16 is a much more complex network and may require large training datasets for optimal performance. DenseNet provides deeper connectivity than AlexNet and VGG16 but with a unique dense connection, which allows direct propagation of information throughout different network layers and thereby reduces the number of parameters needed to create prediction models (27). The dense connection of DenseNet was the likely reason it provided the best diagnostic performance for detecting an ACL tear using the relatively small training dataset in our study.

Figure 2: Images in a 28-year-old man with a surgically confirmed anterior cruciate ligament (ACL) tear that the ACL tear detection system interpreted as present. **(a)** Cropped sagittal proton density-weighted fast spin-echo MR image and **(b)** cropped sagittal fat-suppressed T2-weighted fast spin-echo MR image of the knee analyzed by the classification convolutional neural network show disruption of fibers and increased signal within the ACL (arrow). **(c)** Probability map for the proton density-weighted fast spin-echo image and **(d)** probability map for the sagittal fat-suppressed T2-weighted fast spin-echo image show the high-probability areas in the ligament on which the machine based its interpretation of an ACL tear (arrow).

Our study had several limitations. First, our deep learning–based ACL tear detection system consisted of three individual CNNs connected in a cascaded fashion and not a single network from end to end. The three CNNs were trained individually, which may potentially increase the training burden. Another limitation was the relatively small training dataset used in our study. As a result, several data processing techniques, including transfer learning, were needed to maximize training efficiency. Use of larger training datasets in future studies may further improve the diagnostic performance of the ACL tear detection system. In addition, all subjects with and those without an ACL tear in our study were evaluated with the same MRI unit by using the same imaging protocol. Furthermore, the diagnostic performance of the combined use of clinical radiologist and machine interpretation of the MRI examinations was not evaluated. Another limitation was the presence of possible verification bias, because all

Figure 3: Images in a 26-year-old man with a surgically confirmed anterior cruciate ligament (ACL) tear that the ACL tear detection system interpreted as present. **(a)** Cropped sagittal proton density-weighted fast spin-echo MR image and **(b)** cropped sagittal fat-suppressed T2-weighted fast spin-echo MR image of the knee analyzed by the classification convolutional neural network show disruption of fibers and increased signal within the ACL (arrow). **(c)** Probability map for the proton density-weighted fast spin-echo image and **(d)** probability map for the sagittal fat-suppressed T2-weighted fast spin-echo image show the high-probability areas in the ligament on which the machine based its interpretation of an ACL tear (arrows).

subjects in our study underwent arthroscopic knee surgery. Such verification bias could have falsely increased the sensitivity and decreased the specificity estimates for the ACL tear detection system and for the clinical radiologists (35). A final limitation was the fact that all subjects with ACL tears in our study had full-thickness ligament injuries. The orthopedic surgeons at our institution typically do not surgically reconstruct partial-thickness tears and intrasubstance sprains of the ACL, which are more challenging to detect than full-thickness ACL tears at MRI because they show more subtle changes in the contour and signal intensity of the injured ligament (36,37).

In summary, our study has demonstrated the feasibility of using a deep learning-based approach to detect a full-thickness ACL tear within the knee joint at MRI, with arthroscopy as the reference standard. There was no statistically significant difference between the ACL tear detection system and clinical radiologists with varying levels of experience for determining the presence or absence of a full-thickness ACL tear using sagittal proton density-weighted and fat-suppressed T2-weighted

Figure 4: Cropped sagittal MR images in a 20-year-old man with a surgically confirmed anterior cruciate ligament (ACL) tear that the ACL tear detection system interpreted as an intact ACL. **(a)** Proton density-weighted fast spin-echo image and **(b)** fat-suppressed T2-weighted fast spin-echo image of the knee analyzed by the classification convolutional neural network show fiber disruption and increased signal within the torn ACL (arrow).

fast spin-echo images. However, future work is needed for further technical development and validation of the ACL tear detection system before it could be implemented in clinical

Figure 5: Images in a 21-year-old man with a surgically confirmed intact anterior cruciate ligament (ACL) that the ACL tear detection system interpreted as an ACL tear. **(a)** Cropped sagittal proton density-weighted fast spin-echo MR image and **(b)** cropped sagittal fat-suppressed T2-weighted fast spin-echo MR image of the knee analyzed by the classification convolutional neural network show continuous fibers and normal signal within the intact ACL (arrow). **(c)** Probability map for the proton density-weighted fast spin-echo image and **(d)** probability map for the sagittal fat-suppressed T2-weighted fast spin-echo image show the high-probability areas in the ligament on which the machine based its interpretation of an ACL tear (arrows).

Figure 6: Receiver operating characteristic (ROC) curves describing the diagnostic performance of the proposed anterior cruciate ligament (ACL) tear detection system for determining the presence or absence of a surgically confirmed ACL tear. The area under the ROC curve for the machine was 0.98, indicating high overall diagnostic accuracy. Sensitivity and specificity for the musculoskeletal radiologist, musculoskeletal radiology fellow, radiology residents, and machine at the optimal threshold of the Youden index are also plotted. Note that the sensitivity and specificity of the clinical radiologists are in close proximity to the ROC curve of the ACL tear detection system.

practice. In particular, the diagnostic performance of the ACL tear detection system must be further evaluated in large prospective studies that include subjects with both full-thickness and partial-thickness ACL tears who are evaluated at different institutions using different MRI units and different imaging protocols.

Disclosures of Conflicts of Interest: F.L. disclosed no relevant relationships. B.G. disclosed no relevant relationships. Z.Z. disclosed no relevant relationships. A.S.

disclosed no relevant relationships. H.R. disclosed no relevant relationships. K.L. disclosed no relevant relationships. R.S. disclosed no relevant relationships. A.K. disclosed no relevant relationships. J.K. disclosed no relevant relationships. A.G. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is a consultant for Pfizer, AstraZeneca, Roche, Galapagos, MerckSerono, and TissueGene; holds stock or stock options in BICL. Other relationships: disclosed no relevant relationships. R.K. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is a consultant for Boston Imaging Core Lab; institution has grants or grants pending with GE Healthcare. Other relationships: disclosed no relevant relationships.

Author contributions: Guarantors of integrity of entire study, F.L., B.G., A.K., R.K.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, F.L., B.G., Z.Z., H.R., R.K.; clinical studies, F.L., H.R., K.L., R.S., A.K., J.K., R.K.; experimental studies, F.L., B.G., A.S., J.K.; statistical analysis, F.L., B.G., Z.Z.; and manuscript editing, F.L., B.G., Z.Z., A.S., H.R., K.L., R.S., A.G., R.K.

References

1. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB. Deep learning MR imaging-based attenuation correction for PET/MR imaging. *Radiology* 2018;286(2):676–684.
2. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature* 2018;555(7697):487–492.

3. Chen F, Taviani V, Malkiel I, et al. Variable-density single-shot fast spin-echo MRI with deep learning reconstruction by using variational networks. *Radiology* 2018;289(2):366–373.
4. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. *Magn Reson Med* 2018;79(4):2379–2391.
5. Norman B, Podoia 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(1):177–185.
6. Zhou Z, Zhao G, Kijowski R, Liu F. Deep convolutional neural network for segmentation of knee joint anatomy. *Magn Reson Med* 2018;80(6):2759–2770.
7. Liu F. SUSAN: segment unannotated image structure using adversarial network. *Magn Reson Med* 2019;81(5):3330–3345.
8. 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.
9. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017;284(2):574–582.
10. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;24(9):1342–1350.
11. Podoia 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(2):400–410.
12. 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.
13. Arevalo J, González FA, Ramos-Pollán R, Oliveira JL, Guevara Lopez MA. Representation learning for mammography mass lesion classification with convolutional neural networks. *Comput Methods Programs Biomed* 2016;127:248–257.
14. Kooi T, Litjens G, van Ginneken B, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal* 2017;35:303–312.
15. Cicero M, Bilbily A, Colak E, et al. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Invest Radiol* 2017;52(5):281–287.
16. Ciompi F, de Hoop B, van Riel SJ, et al. Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. *Med Image Anal* 2015;26(1):195–202.
17. Setio AA, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans Med Imaging* 2016;35(5):1160–1169.
18. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging* 2016;35(5):1207–1216.
19. Gianotti SM, Marshall SW, Hume PA, Bunt L. Incidence of anterior cruciate ligament injury and other knee ligament injuries: a national population-based study. *J Sci Med Sport* 2009;12(6):622–627.
20. Lee JK, Yao L, Phelps CT, Wirth CR, Czajka J, Lozman J. Anterior cruciate ligament tears: MR imaging compared with arthroscopy and clinical tests. *Radiology* 1988;166(3):861–864.
21. Robertson PL, Schweitzer ME, Bartolozzi AR, Ugoni A. Anterior cruciate ligament tears: evaluation of multiple signs with MR imaging. *Radiology* 1994;193(3):829–834.
22. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Annu Rev Biomed Eng* 2017;19(1):221–248.
23. Lecun Y, Jackel LD, Bottou L, et al. Learning algorithms for classification: A comparison on handwritten digit recognition. In: Oh JH, Kwon C, Cho S, eds. *Neural networks: the statistical mechanics perspective*. World Scientific, 1995; 261–276. <https://nyuscholars.nyu.edu/en/publications/learning-algorithms-for-classification-a-comparison-on-handwrite>. Accessed September 15, 2018.
24. Malon CD, Cosatto E. Classification of mitotic figures with convolutional neural networks and seeded blob features. *J Pathol Inform* 2013;4(1):9.
25. Sarraf S, Tofghi G, for the Alzheimer's Disease Neuroimaging Initiative. DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *bioRxiv* 070441 v4 [preprint]. <https://www.biorxiv.org/content/10.1101/070441v4>. Posted August 21, 2016. Accessed January 12, 2018.
26. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. *ArXiv* 1506.02640 [preprint]. <http://arxiv.org/abs/1506.02640>. Posted 2015. Accessed September 15, 2018.
27. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. *ArXiv* 1608.06993 [preprint]. <http://arxiv.org/abs/1608.06993>. Posted 2016. Accessed September 15, 2018.
28. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv* 1409.1556 [preprint]. <http://arxiv.org/abs/1409.1556>. Posted 2014. Accessed March 27, 2018.
29. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *ArXiv* 1409.0575 [preprint]. <http://arxiv.org/abs/1409.0575>. Posted 2014. Accessed April 13, 2018.
30. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. *ArXiv* 1512.04150 [preprint]. <https://arxiv.org/abs/1512.04150>. Posted 2015. Accessed August 13, 2018.
31. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 2012; 25:1097–1105. <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>. Accessed October 20, 2017.
32. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44(3):837–845.
33. Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. *Biom J* 2005;47(4):458–472.
34. McCauley TR, Moses M, Kier R, Lynch JK, Barton JW, Joki P. MR diagnosis of tears of anterior cruciate ligament of the knee: importance of ancillary findings. *AJR Am J Roentgenol* 1994;162(1):115–119.
35. Richardson ML, Petsavage JM. Verification bias: an under-recognized source of error in assessing the efficacy of MRI of the meniscii. *Acad Radiol* 2011;18(11):1376–1381.
36. Vahey TN, Broome DR, Kayes KJ, Shelbourne KD. Acute and chronic tears of the anterior cruciate ligament: differential features at MR imaging. *Radiology* 1991;181(1):251–253.
37. Van Dyck P, Vanhoenacker FM, Gielen JL, et al. Three tesla magnetic resonance imaging of the anterior cruciate ligament of the knee: can we differentiate complete from partial tears? *Skeletal Radiol* 2011;40(6):701–707.