Skeletal Radiology (2022) 51:2269–2279
https://doi.org/10.1007/s00256-022-04081-x

SCIENTIFIC ARTICLE

Ensemble deep learning model for predicting anterior cruciate ligament tear from lateral knee radiograph

Dong Hyun Kim[1] · Jee Won Chai[1] · Ji Hee Kang[2] · Ji Hyun Lee[1] · Hyo Jin Kim[1] · Jiwoon Seo[1] · Jae Won Choi[3,4]

Received: 25 March 2022 / Revised: 24 May 2022 / Accepted: 24 May 2022 / Published online: 9 June 2022

Abstract

Objective To develop an ensemble deep learning model (DLM) predicting anterior cruciate ligament (ACL) tears from lateral knee radiographs and to evaluate its diagnostic performance.

Materials and methods In this study, 1433 lateral knee radiographs (661 with ACL tear confirmed on MRI, 772 normal) from two medical centers were split into training ($n = 1146$) and test sets ($n = 287$). Three single DLMs respectively classifying radiographs with ACL tears, abnormal lateral femoral notches, and joint effusion were developed. An ensemble DLM predicting ACL tears was developed by combining the three DLMs via stacking method. The sensitivities, specificities, and area under the receiver operating characteristic curves (AUCs) of the DLMs and three radiologists were compared using McNemar test and Delong test. Subgroup analysis was performed to identify the radiologic features associated with the sensitivity.

Results The sensitivity, specificity, and AUC of the ensemble DLM were 86.8% (95% confidence interval [CI], 79.9–92.0%), 89.4% (95% CI, 83.4–93.8%), and 0.927 (95% CI, 0.891–0.954), achieving diagnostic performance comparable with that of a musculoskeletal radiologist ($P = 0.193$, McNemar test; $P = 0.131$, Delong test). The AUC of the ensemble DLM was significantly higher than those of non-musculoskeletal radiologists ($P = 0.043$, $P < 0.001$). The sensitivity of the DLM was higher than that of the radiologists in the absence of an abnormal lateral femoral notch or joint effusion.

Conclusion The diagnostic performance of the ensemble DLM in predicting lateral knee radiographs with ACL tears was comparable to that of a musculoskeletal radiologist.

Keywords Deep learning · Ensemble deep learning · Knee · Anterior cruciate ligament · Radiograph

Abbreviations

ACL Anterior cruciate ligament
DLM Deep learning model
AUC Area under the receiver operating characteristic curve
CI Confidence interval

Dong Hyun Kim and Jee Won Chai contributed equally to this work.

✉ Ji Hee Kang
kangjihee1@gmail.com

1 Department of Radiology, SMG-SNU Boramae Medical Center, Seoul National University College of Medicine, Seoul, Republic of Korea

2 Department of Radiology, Konkuk University Medical Center, 120-1 Neungdong-ro, Gwangjin-gu, Seoul 05030, Republic of Korea

3 Armed Forces Yangju Hospital, Yangju, Republic of Korea

4 Department of Radiology, Seoul National University Hospital, Seoul, Republic of Korea

Introduction

An anterior cruciate ligament (ACL) tear is one of the common sports-related injuries of the knee. ACL tears can result in anteroposterior and rotatory instability of the knee, which increases the risk of cartilage or meniscal damage [1, 2]. Thus, prompt and accurate diagnosis of ACL tears with early surgical reconstruction is vital to prevent post-traumatic osteoarthritis. However, pain and muscle spasm may lower the accuracy of physical examination in patients with acute ACL tears [3]. Although MRI is the best imaging modality to confirm ACL tears, it is not widely used for initial evaluation because of the high cost. Therefore, conventional radiography, which is a readily available imaging tool, plays a role in screening ACL tears. If an ACL tear is suspected on radiography, follow-up physical examination and subsequent MRI examination would be needed even if ACL injury was not evident on the initial clinical examination.

There are several signs of ACL tears on conventional radiography. An abnormally deep lateral condylopatellar sulcus, which is called "lateral femoral notch sign," can be noted due to an impaction of the lateral femoral condyle by pivot shift injury mechanism [4]. Owing to the different impaction patterns, "abnormal lateral notches" including a deep notch, a long notch, and a double notch can also be observed [5, 6]. In addition, the amount of knee joint effusion usually increases in acute ACL tears [7]. These findings are best observed on the lateral radiograph. However, their sensitivity is quite low, and detection of these findings is significantly dependent on the interpreter's experience.

Recent studies have shown that deep learning model (DLM) is useful for disease detection and characterization on radiologic images [8–10]. Compared to single learning models, the performance of ensemble DLMs can be improved by combining several models [11]. Therefore, the purpose of this study was to develop three convolutional neural network-based single models that can respectively classify lateral knee radiographs with ACL tears, abnormal lateral notches, and increased amounts of knee joint effusion, and to finally develop an ensemble model by combining three single models and to validate the diagnostic performance of the models in predicting ACL tears.

Materials and methods

This retrospective study was approved by the institutional review boards of two medical centers (SMG-SNU Boramae Medical Center and Konkuk University Medical Center). The requirement for informed consent was waived due to its retrospective nature.

Dataset

Lateral knee radiographs from patients with MRI findings of ACL tear in SMG-SNU Boramae Medical Center (BRMC) and Konkuk University Medical Center (KUMC) between January 2010 and October 2020 were collected. Patients aged 45 years or below were included because more than 90% of ACL tears occur in this age group [12] and to exclude elderly patients with osteoarthritic changes of the knee. The following exclusion criteria were applied: (a) interval of longer than 15 days between lateral knee radiograph and knee MRI; (b) chronic ACL tear; (c) presence of fracture not related to ACL tear; (d) images with suboptimal image quality; (e) images with 60° or more of knee flexion; (f) previous knee operation; (g) presence of bone tumor; and (h) images with external immobilization or hardware (splint, screw, etc.). Chronic ACL tear was defined as fragmented appearance with low signal intensity or non-visualization on MRI [13, 14]. Finally, 661 lateral knee radiographs with ACL tears were included (Fig. 1). The time interval between the injury and lateral knee radiograph ranged from 0 to 168 days with a mean value of 9 days.

In addition, lateral knee radiographs obtained between October 2019 and October 2020 from the two musculoskeletal radiologists (J.H.K. and J.W.C., each with 7 and 18 years of experience). Normal lateral knee radiographs without any abnormal findings confirmed by two

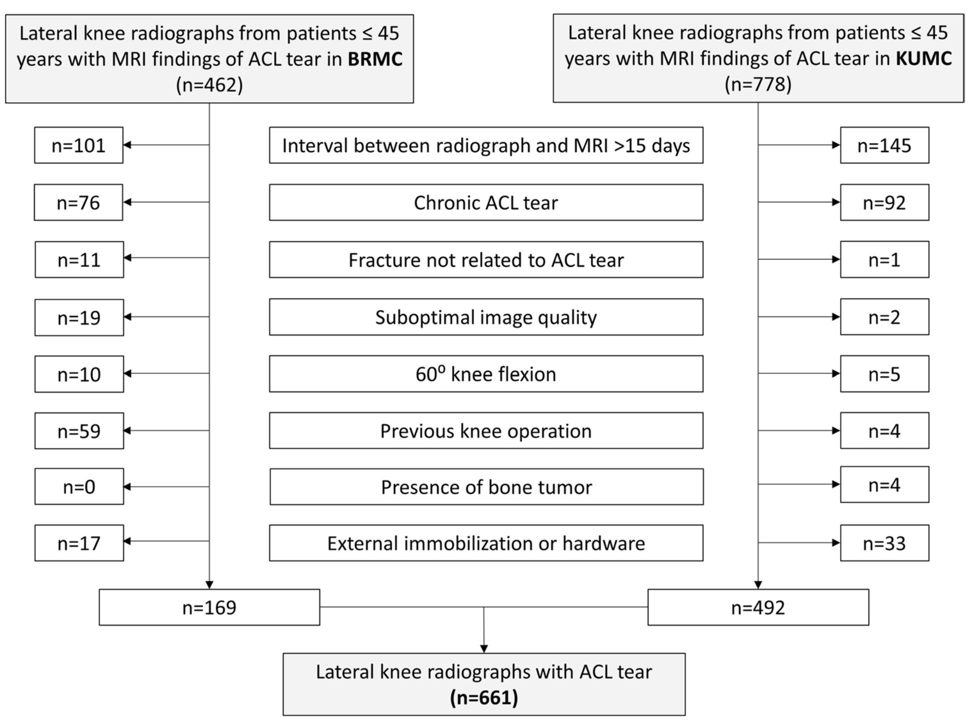

Fig. 1 Flowchart shows the process used to enroll lateral knee radiographs from ACL tear patients. ACL = anterior cruciate ligament, BRMC = SMG-SNU Boramae Medical Center, KUMC = Konkuk University Medical Center

musculoskeletal radiologists were included as controls. As a result, 772 normal lateral knee radiographs were enrolled.

Finally, a total of 1433 lateral knee radiographs from 1433 subjects were included in this study. The dataset consisted of 521 images from BRMC (169 with ACL tear, 352 normal) and 912 images from KUMC (492 with ACL tear, 420 normal). Among them, 252 radiographs (132 with ACL tear, 120 normal) were from pediatric subjects aged 18 years or below. Among 252 radiographs from pediatric patients, 38 radiographs (17 with ACL tear, 21 normal) showed skeletal immaturity, having an open physeal plate.

Image labeling and data split

All lateral knee radiographs were independently reviewed by two musculoskeletal radiologists (J.H.K. and J.W.C.). During the labeling process, they referred to MRI and contralateral lateral knee radiographs obtained at the same time as the ACL-injured knee (available in 72.5% [479/661] of the cases). They labeled lateral femoral notches as abnormal (deep, long, or double notches [5, 6]) or normal. The lateral femoral notch was designated abnormal when the notch was asymmetric by comparison of bilateral radiographs or was suspected abnormal in unilateral radiograph and there was corresponding bone lesion on MRI. The amount of knee joint effusion was also classified as increased (anteroposterior dimension of supra-patellar recess ≥ 5 mm [15]) or within the normal range. In cases of disagreement (10 radiographs for lateral femoral notches, and 15 radiographs for joint effusion), they re-evaluated the images in consensus.

Subsequently, 1433 lateral knee radiographs were split into training ($n = 1146$) and test sets ($n = 287$) at a ratio of 8:2. Owing to the imbalanced data distribution, a stratified split was performed. The detailed characteristics of the training and test sets are presented in Table 1.

Image preprocessing

The images were manually cropped to 1000 × 1000 pixels to include supra-patellar recess, patella, distal femur and tibial plateau. For the DLM detecting abnormal lateral femoral notch, the images were also cropped to 600 × 600 pixels centered on the lateral femoral notch. All images were resized to 512 × 512 pixels. Min–max normalization was used to standardize the images. Contrast limited adaptive histogram equalization was applied to enhance the contrast. To generalize the model, augmentation of the dataset was performed by rotation (within ± 20 degrees), horizontal flip, vertical flip, and center crop.

Table 1 Demographic information and characteristics of training and test sets

Characteristic	Training set ($n = 1146$)	Test set ($n = 287$)	Total ($n = 1433$)
Sex			
Male	799 (69.7)	201 (70.0)	1000 (69.8)
Female	347 (30.3)	86 (30.0)	433 (30.2)
Age (years)*	27.4 ± 8.7	27.0 ± 8.7	27.3 ± 8.7
ACL tear			
Absent	621 (54.2)	151 (52.6)	772 (53.9)
Present	525 (45.8)	136 (47.4)	661 (46.1)
Lateral femoral notch			
Normal	845 (73.7)	208 (72.5)	1053 (73.5)
Abnormal	301 (26.3)	79 (27.5)	380 (26.5)
Joint effusion			
Within normal range	727 (63.4)	176 (61.3)	903 (63.0)
Increased amount	419 (36.6)	111 (38.7)	530 (37.0)

Unless otherwise specified, data are reported as the number of patients or radiographs. Data in parentheses are percentages

*Data are reported as mean ± standard deviation

ACL anterior cruciate ligament

DLM

First, three single binary classification DLMs for ACL tear ("ACL tear model"), abnormal lateral femoral notch ("abnormal notch model"), and increased amount of knee joint effusion ("joint effusion model") were developed. Subsequently, an ensemble DLM detecting ACL tear, which was named "ensemble ACL tear model," was developed by combining the three single models by stacking method.

Figure 2 shows the architecture of the ensemble DLM. The orange box comprised multiple layers repeating 3 × 3 convolution, rectified linear unit, and batch normalization twice. The yellow box included max pooling with a 2 × 2 filter and a stride of 2, followed by a dropout layer with a dropout rate of 30%. Subsequently, the green box of a flatten layer converted the data into one dimension and sent it to the fully connected layer (blue circles). In the fully connected layer, the first dense layer included the rectified linear unit as the activation function and a dropout layer. In the second dense layer, a sigmoid activation function was used for the binary classification. The ensemble of the three single models was performed through the concatenate layer. In the last dense layer, a sigmoid function was used as the activation function. The model was trained for 100 epochs with a batch size and learning rate set to 20 and 0.0001, respectively. RMSProp was used as the optimizer, and the loss function was binary cross-entropy. To optimize the model

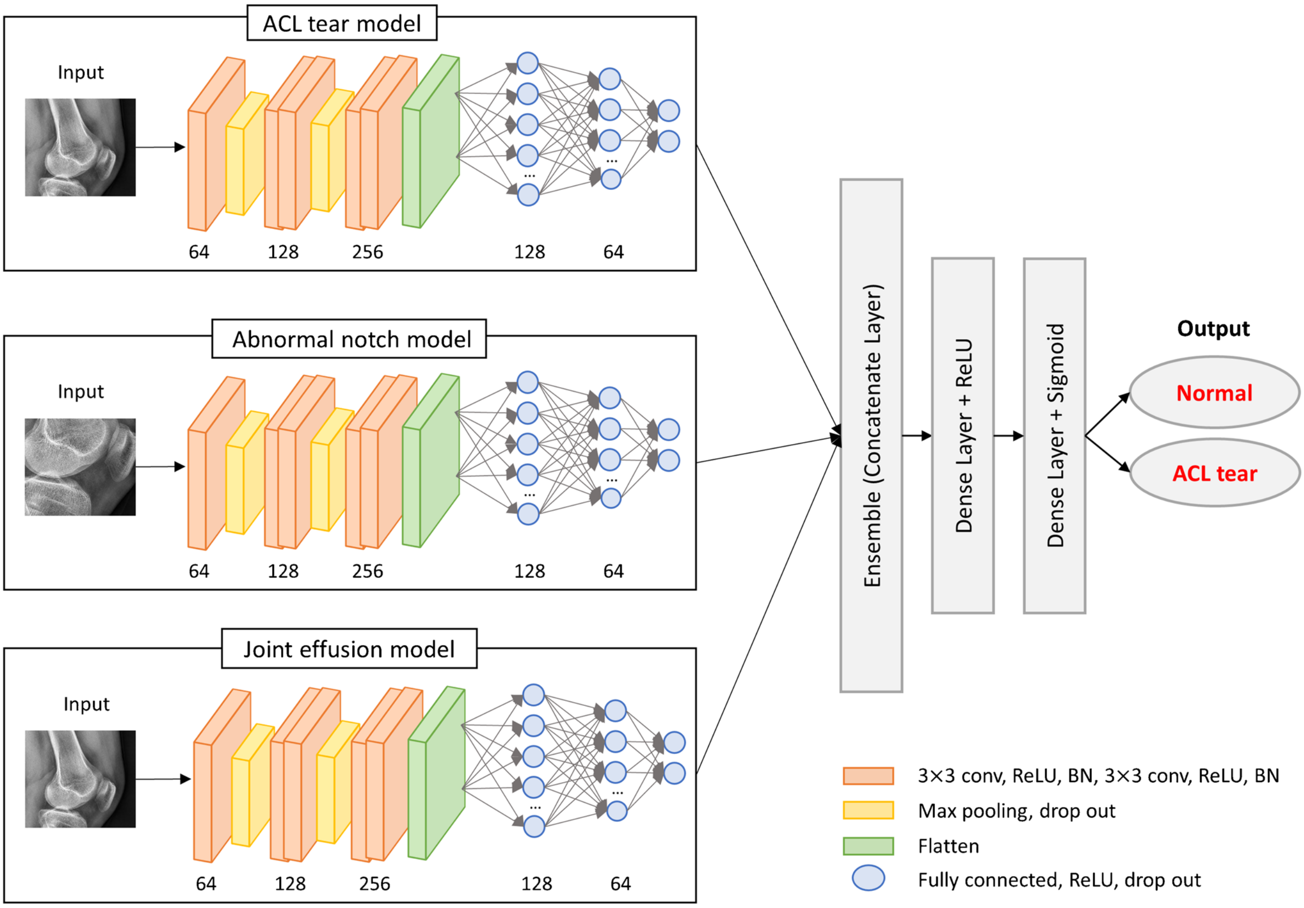

Fig. 2 Architecture of the ensemble ACL tear model. The ensemble ACL tear model was created by combining the ACL tear model, abnormal notch model, and joint effusion model using stacking method. The orange box comprised multiple layers repeating 3×3 convolution, rectified linear unit, and batch normalization twice. The yellow box included max pooling with a 2×2 filter and a stride of 2, followed by a dropout layer with a dropout rate of 30%. Subsequently, the green box of a flatten layer converted the data into one dimension and sent it to the fully connected layer (blue circles). In the fully connected layer, the first dense layer included the rectified linear unit as the activation function and a dropout layer. In the second dense layer, a sigmoid activation function was used for the binary classification. The ensemble of the three single models was performed through the concatenate layer. In the last dense layer, a sigmoid function was used as the activation function. The number below the box or circle represents the number of output units. ACL = anterior cruciate ligament, conv = convolution, ReLU = rectified linear unit, BN = batch normalization

and avoid overfitting, a stratified fourfold cross-validation was performed in the training set. Three single DLMs provided probability scores between 0 and 1 for the presence of an ACL tear, abnormal lateral femoral notch, and joint effusion, respectively. Finally, the ensemble DLM provided a probability score between 0 and 1 for the presence of an ACL tear.

The test set was evaluated using the trained DLMs. Heatmaps highlighting the most discriminative regions in images for model prediction were generated using gradient-weighted class activation mapping [16]. Training and evaluation of the DLMs were performed in Python 3.8 (Python Software Foundation, Wilmington, Del) using Tensorflow library version 2.2.0 (Google Brain, Mountain View, California) with four Nvidia GeForce GTX 1080 Ti GPUs. Finally, sensitivity, specificity, and accuracy of the DLMs were calculated. The cut-off probability values of the models were set to 0.5.

Review of radiologic features by radiologists

Three radiologists (reviewer 1, H.J.K., musculoskeletal radiologist; reviewer 2, J.W.C., non-musculoskeletal boardcertified radiologist; reviewer 3, K.L., radiology resident; each with 8, 5, and 1 year of experience) independently reviewed all images of the test set. They were asked to grade the likelihood of the presence of an ACL tear, abnormal lateral femoral notch, and increased amount of joint effusion on a 5-point scale as follows: 1, definitely absent, 0%; 2, probably absent, 25%; 3, indeterminate, 50%; 4, probably present, 75%; and

5, definitely present, 100%. The reviewers were blinded to the clinical information of each patient; however, they were aware that all images were either normal or with ACL tears. Anonymized DICOM image files without cropping were provided to the reviewers.

The sensitivity, specificity, and accuracy of the reviewers were calculated. Grades 1, 2, and 3 were regarded as normal and grades 4 and 5 were considered abnormal.

Subgroup analysis

For images with ACL tears in the test set, subgroup analysis was performed to identify the specific radiologic features associated with the diagnostic sensitivity of the DLMs and reviewers. Images were dichotomized according to the presence of an abnormal lateral femoral notch and a significant amount of joint effusion (anteroposterior dimension of supra-patellar recess ≥ 5 mm). The sensitivities of the DLMs and reviewers in each subgroup were calculated.

Statistical analysis

The sensitivity and specificity of the DLMs and reviewers were compared using McNemar test. The area under the receiver operating characteristic curves (AUCs) were calculated. Comparisons between the AUCs were performed using the Delong method. Statistical analyses were performed using MedCalc version 20.015 (MedCalc Software, Ostend, Belgium). A two-tailed *P* value less than 0.05 was considered statistically significant.

Results

Diagnostic performance of the DLMs

The diagnostic performance of the DLMs and human reviewers in detecting radiographs with ACL tears is summarized in Table 2. The sensitivity and specificity of the ensemble ACL tear model in detecting ACL tears were 86.8% (118/136; 95% confidence interval [CI], 79.9–92.0%) and 89.4% (135/151; 95% CI, 83.4–93.8%), while those of the ACL tear model were 84.6% (115/136; 95% CI, 77.4–90.2%) and 86.1% (130/151; 95% CI, 79.5–91.2%), respectively. The AUC of the ensemble ACL tear model and ACL tear model was 0.927 (95% CI, 0.891–0.954) and 0.856 (95% CI, 0.810–0.895) (Fig. 3). The AUC of the ensemble model was significantly higher than that of the ACL tear model ($P = 0.004$).

The sensitivity and specificity of the abnormal notch model in detecting abnormal lateral femoral notch were 79.8% (63/79; 95% CI, 69.2–88.0%) and 84.1% (175/208; 95% CI, 78.5–88.8%) (Supplemental Table 1). The joint effusion model achieved a sensitivity of 82.0% (91/111; 95% CI, 73.6–88.6%) and a specificity of 88.6% (156/176; 95% CI, 83.0–92.9%) in detecting an increased amount of knee joint effusion (Supplemental Table 2).

Heatmaps generated in the ACL tear model included supra-patella recess, lateral femoral notch, and Hoffa's fat pad within the highlighted regions. In the abnormal notch model, the highlighted regions included the lateral femoral notch. In the joint effusion model, supra-patellar recess was mostly highlighted (Figs. 4, 5, 6).

Comparison of the performance between DLMs and reviewers

The sensitivity and specificity of the human reviewers in detecting ACL tears ranged from 64.0 to 79.4% and 81.5 to 88.1%, respectively. The AUC of the three reviewers varied according to their level of experience, ranging from 0.785 to 0.898. The diagnostic accuracy of the ensemble ACL tear model was significantly higher than that of reviewer 3 (88.2% vs. 73.2%, $P = 0.042$, McNemar test). In addition, the AUC of the ensemble ACL tear model was significantly higher than that of reviewer 2 (0.927 vs. 0.886, $P = 0.043$)

Table 2 Diagnostic performance of ensemble ACL tear model, ACL tear model, and human reviewers in detecting ACL tear

Parameter	Ensemble ACL tear model	ACL tear model	Reviewer 1	Reviewer 2	Reviewer 3
Sensitivity (%)	86.8 (118/136) [79.9–92.0]	84.6 (115/136) [77.4–90.2]	77.9 (106/136) [70.0–84.6]	79.4 (108/136) [71.6–85.9]	64.0 (87/136) [55.3–72.0]
Specificity (%)	89.4 (135/151) [83.4–93.8]	86.1 (130/151) [79.5–91.2]	88.1 (133/151) [81.8–92.8]	87.4 (132/151) [81.1–92.3]	81.5 (123/151) [74.3–87.3]
Accuracy (%)	88.2 (253/287) [83.8–91.7]	85.4 (245/287) [80.7–89.3]	83.3 (239/287) [78.4–87.4]	83.6 (240/287) [78.8–87.7]	73.2 (210/287) [67.7–78.2]
AUC	0.927 [0.891–0.954]	0.856 [0.810–0.895]	0.898 [0.857–0.931]	0.886 [0.844–0.921]	0.785 [0.733–0.831]

Reviewer 1, musculoskeletal radiologist; reviewer 2, non-musculoskeletal board-certified radiologist; reviewer 3, radiology resident

Data in parentheses are the number of images. Data in brackets are 95% confidence intervals. *ACL* anterior cruciate ligament, *AUC* area under the receiver operating characteristic curve

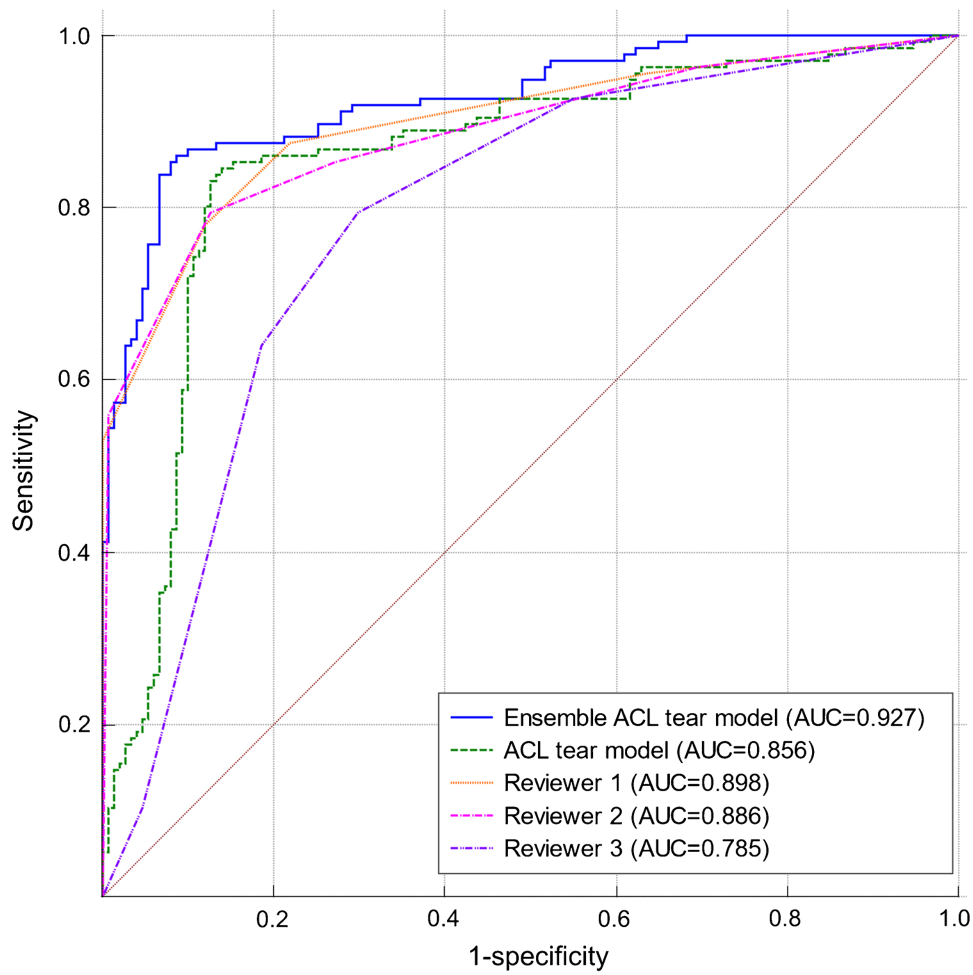

Fig. 3 Receiver operating characteristic curves of the models and human reviewers. AUC = area under the receiver operating characteristic curve

and reviewer 3 (0.927 vs. 0.785, $P < 0.001$). There was no significant difference in the diagnostic performance between the ensemble ACL tear model and reviewer 1 ($P = 0.193$, McNemar test; $P = 0.131$, Delong test).

Diagnostic sensitivity according to radiologic features

The results of the subgroup analyses are presented in Table 3. The sensitivity of the ensemble ACL tear model was higher than that of the reviewers in the absence of abnormal lateral femoral notch (75.4% vs. 52.6–66.7%; reviewer 1, $P = 0.049$; reviewer 2, $P = 0.302$; reviewer 3, $P = 0.011$) and in the cases of physiologic amount of joint effusion (56.0% vs. 20.0–36.0%; reviewer 1, $P = 0.008$; reviewer 2, $P = 0.227$; reviewer 3, $P = 0.049$). In contrast, in the presence of abnormal lateral femoral notch or joint effusion, sensitivities were not significantly different between the ensemble ACL tear model and reviewers except for reviewer 3 (in the presence of lateral femoral notch, 94.9% vs. 72.2%, $P < 0.001$; in the presence of joint effusion, 93.7% vs. 73.9%, $P < 0.001$).

Discussion

We developed an ensemble DLM that could screen lateral knee radiographs with ACL tears by combining DLMs classifying radiographs with indirect signs of ACL tears. The ensemble DLM provided the sensitivity, specificity, and AUC of 84.6%, 86.1%, and 0.927, respectively, achieving diagnostic performance comparable with that of a musculoskeletal radiologist. The AUC of the ensemble DLM was significantly higher than those of the single DLM and non-musculoskeletal radiologists. Especially in the absence of abnormal lateral femoral notch or joint effusion, the DLM provided better sensitivity for predicting ACL tears compared to radiologists.

In the current study, ensemble learning was used to improve overall performance of the DLM by simultaneously learning related tasks in the form of multi-task learning [17, 18]. In multi-task learning, information obtained during the classification task of "ACL tear" has a good influence on the other classification tasks of "abnormal lateral femoral notch" or "joint effusion", and vice versa. In addition, overfitting is reduced by learning a more generalized shared

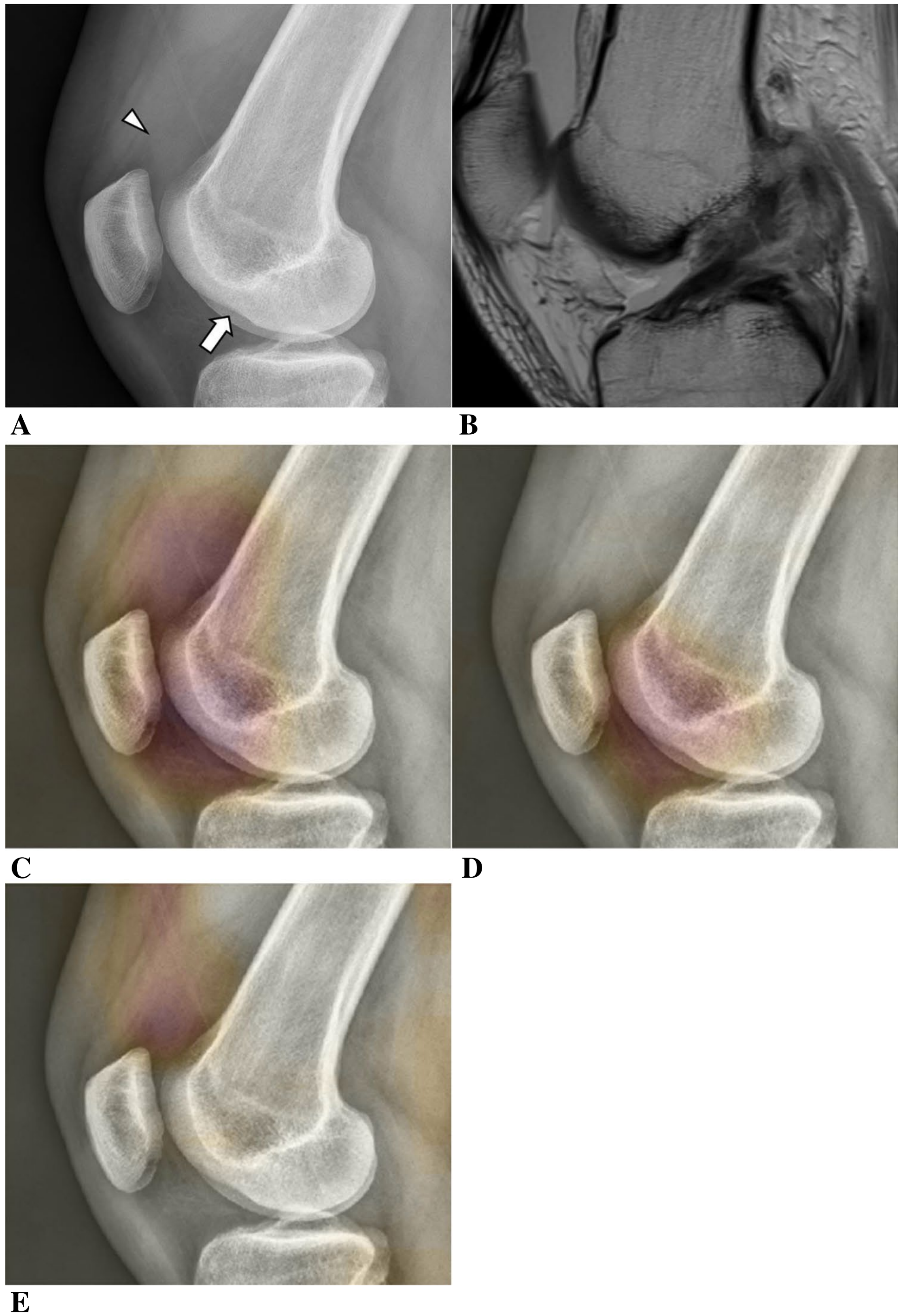

Fig. 4 Images in a 35-year-old man with anterior cruciate ligament (ACL) tear in which the ensemble ACL tear model and all human reviewers made correct predictions. **A** Cropped lateral knee radiograph shows deep lateral femoral notch (arrow) and joint effusion in supra-patellar recess (arrowhead). The probability score from the ensemble ACL tear model for the presence of ACL tear was 0.90. **B** Proton density weighted sagittal MRI shows complete tear of ACL. **C** Heatmap from the ACL tear model overlaid on the radiograph included supra-patella recess, lateral femoral notch, and Hoffa's fat pad within the highlighted region. **D** Heatmap from the abnormal notch model overlaid on the radiograph shows that an abnormal lateral femoral notch was successfully detected with a probability score of 0.80. **E** Heatmap from the joint effusion model overlaid on the radiograph shows that joint effusion was successfully detected with a probability score of 0.79.

representation to fit multiple tasks simultaneously. Previous works on cancer classifications also used multi-task learning with auxiliary classifiers to enhance the performance of a simple binary classification model [19–21].

In this study, the DLM showed better sensitivity compared to the radiologists. It is attributed to the higher sensitivity of the DLM in the absence of an abnormal lateral femoral notch or joint effusion. It has been reported that even if lateral femoral notch appears normal on radiograph, bone marrow edema is present on MRI in about half of the ACL tear patients [6, 22]. Therefore, we speculated that the DLM may recognize subtle bone marrow density changes at the grossly normal lateral femoral notch in images with ACL tears. In addition, heatmaps derived from the ACL tear model showed highlighted regions including Hoffa's fat pad as well as the supra-patellar recess and lateral

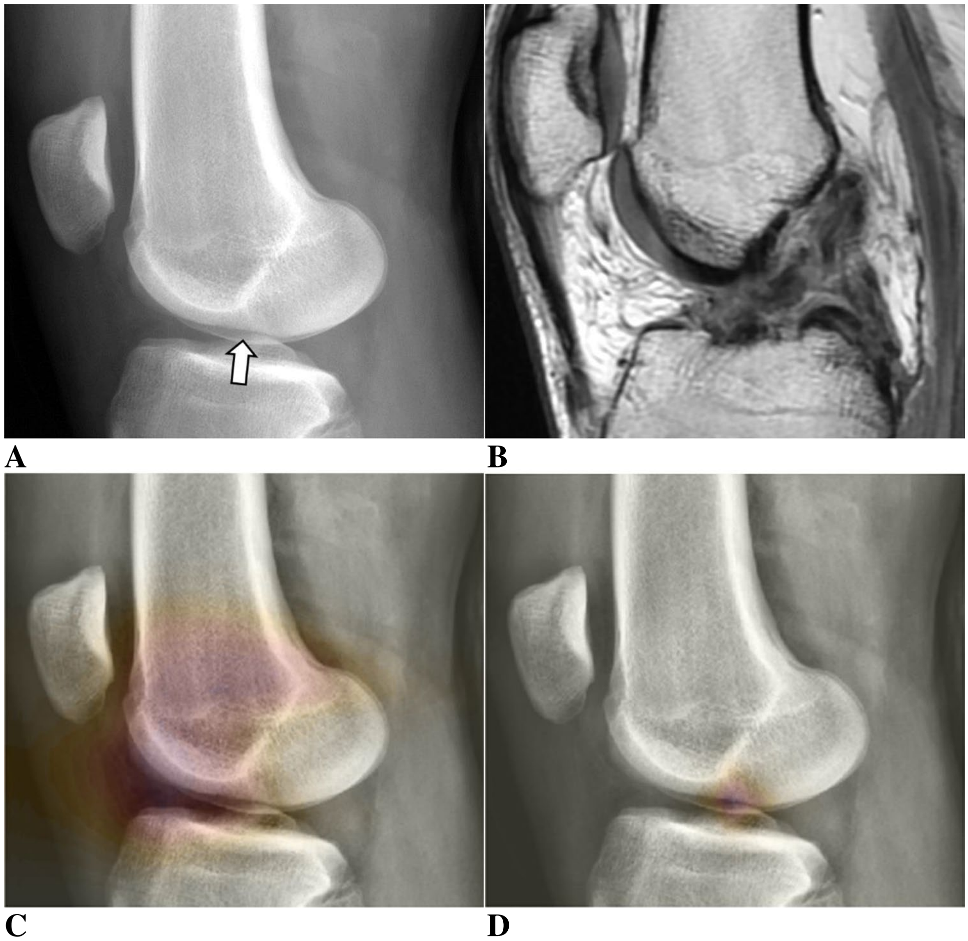

Fig. 5 Images in a 24-year-old man with anterior cruciate ligament (ACL) tear in which the ensemble ACL tear model made correct predictions but all human reviewers made incorrect predictions. **A** Cropped lateral knee radiograph shows a long lateral femoral notch (arrow). Joint effusion in supra-patellar recess is within a physiologic amount. Reviewers labeled this image as grade 3 (indeterminate), 3, and 2 (probably absent), respectively. The probability score from the ensemble ACL tear model for the presence of ACL tear was 0.91. **B** Proton density weighted sagittal MRI shows complete tear of ACL. **C** Heatmap from the ACL tear model overlaid on the radiograph included femoral condyle, lateral femoral notch, and Hoffa's fat pad within the highlighted region. The probability score of the ACL tear model for the presence of ACL tear was 0.61. **D** Heatmap from the abnormal notch model overlaid on the radiograph shows that an abnormal lateral femoral notch was successfully detected with a probability score of 0.64. Reviewers assessed the likelihood of abnormal lateral femoral notch as grade 4 (probably present), 0 (definitely absent), and 3 (indeterminate), respectively.

femoral condyle. This suggests that model used Hoffa's fat pad edema for the prediction of ACL tears, while the radiologists probably did not. Edema and synovial proliferation of Hoffa's fat pad on MRI is present in more than half of knees with ACL injury [23]. It can be inferred that Hoffa's fat pad edema should also be considered as an important radiographic sign in acute traumatic setting.

The clinical utility of the proposed DLM would be ancillary in the diagnosis of ACL tears. Admittedly, MRI is usually requested to confirm ACL tears based on trauma history and clinical knee instability since it provides excellent diagnostic performance and facilitates the detection of associated injuries [24]. However, physical exam in the emergency department is often overlooked or performed by inexperienced physicians. Since the diagnostic performance of the DLM was comparable to that of a musculoskeletal radiologist, the proposed DLM can be an adjunctive diagnostic tool for ACL tears especially in the emergency department where experienced radiologists and physicians are not always available. Considering the higher sensitivity of the DLM compared to the radiologists in the absence of indirect signs of ACL tears, the DLM can help prevent misinterpret radiographs with ACL tears as normal. In addition, an immobilizer can be applied to patients suspected of ACL tears by the DLM to prevent further incidents of injury.

There were 18 false negative cases (18/136, 13.2%) where the ensemble DLM classified the lateral knee radiographs as normal but ACL tears were present on MRI. In 11 cases,

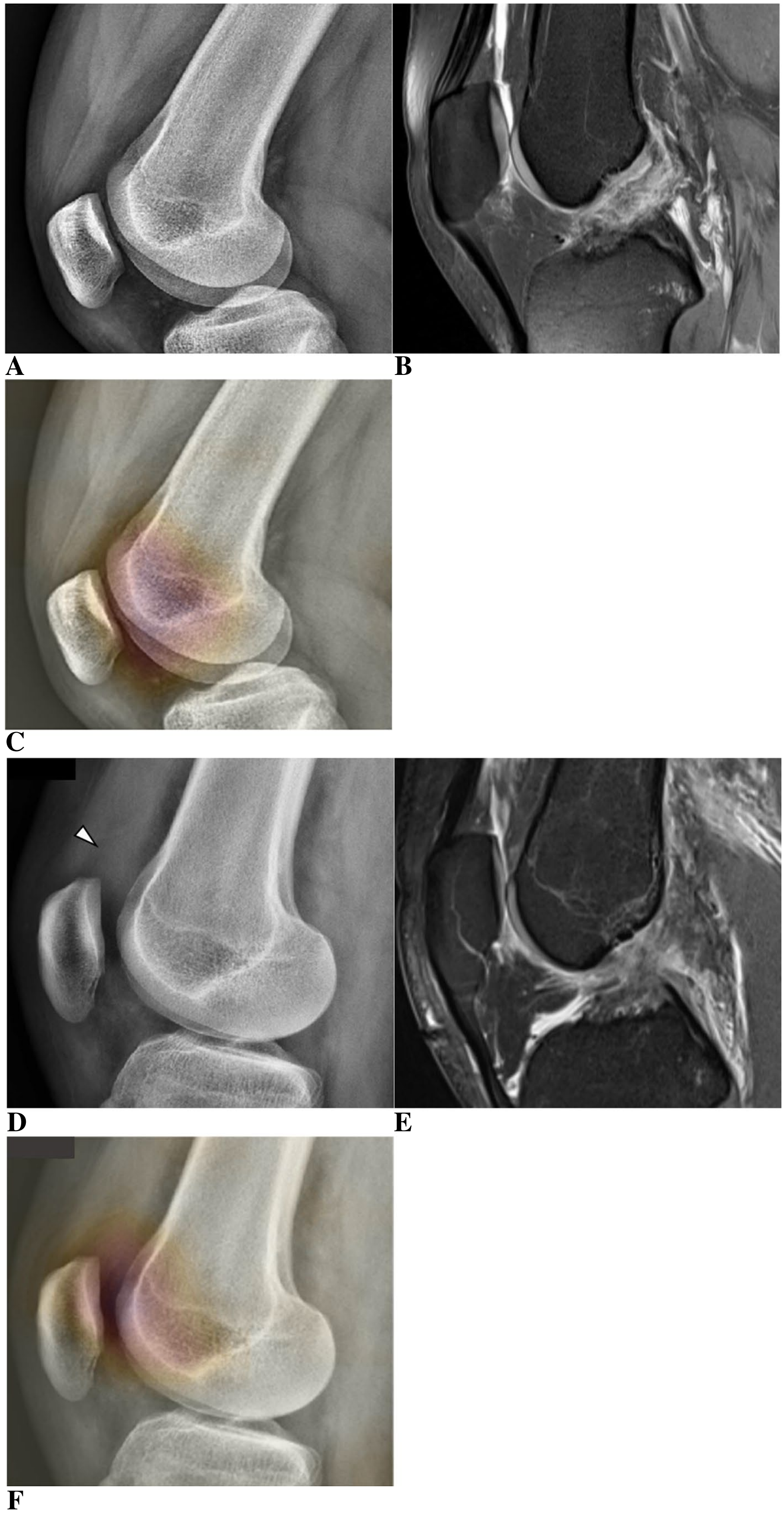

Fig. 6 Representative false negative cases. **A, B, C** Images in a 27-year-old man with anterior cruciate ligament (ACL) tear. **A** Cropped lateral knee radiograph shows normal configuration of lateral femoral condyle without joint effusion. All reviewers labeled this image as grade 0 (definitely absent). The probability score from the ensemble ACL tear model for the presence of ACL tear was 0.20. **B** Fat-suppressed proton density weighted sagittal MRI shows complete tear of ACL. **C** Heatmap from the ACL tear model overlaid on the radiograph included femoral condyle, lateral femoral notch, and Hoffa's fat pad within the highlighted region. **D, E, F** Images in a 21-year-old man with anterior cruciate ligament (ACL) tear. **D** Cropped lateral knee radiograph shows joint effusion in supra-patellar recess (arrowhead) with normal configuration of lateral femoral condyle. Reviewers labeled this image as grade 5 (definitely present), 5, and 4 (probably present), respectively. The probability score from the ACL tear model for the presence of ACL tear was 0.75. However, the probability score from the ensemble ACL tear model was 0.17. **E** Fat-suppressed proton density weighted sagittal MRI shows complete tear of ACL. **F** Heatmap from the ACL tear model overlaid on the radiograph included femoral condyle, supra-patella recess, and Hoffa's fat pad within the highlighted region.

Table 3 Diagnostic sensitivity of the models and human readers for images with or without radiologic features

Models and human readers	Abnormal lateral femoral notch		Joint effusion	
	Present ($n=79$)	Absent ($n=57$)	Increased amount ($n=111$)	Within normal range ($n=25$)
Ensemble ACL tear model	94.9% (75/79)	75.4% (43/57)	93.7% (104/111)	56.0% (14/25)
ACL tear model	92.4% (73/79)	73.7% (42/57)	92.8% (103/111)	48.0% (12/25)
Reviewer 1	91.1% (72/79)	59.6% (34/57)	90.1% (100/111)	24.0% (6/25)
Reviewer 2	88.6% (70/79)	66.7% (38/57)	89.2% (99/111)	36.0% (9/25)
Reviewer 3	72.2% (57/79)	52.6% (30/57)	73.9% (82/111)	20.0% (5/25)

Reviewer 1, musculoskeletal radiologist; reviewer 2, non-musculoskeletal board-certified radiologist; reviewer 3, radiology resident

Data in parentheses are presented as the number of true positives out of the total number of images. *ACL* anterior cruciate ligament

lateral femoral notch sign and joint effusion were absent or minimally present. On the contrary, indirect signs of ACL tears were present in the other 7 cases that reviewers and single DLM made correct predictions except for the ensemble DLM (Fig. 6). Although the overall sensitivity of the ensemble DLM was higher than that of the single DLM, there must be some trade-offs. Therefore, referring to the results of both ensemble and single DLMs would help make correct diagnoses.

In the current study, we developed a DLM classifying lateral knee radiographs with joint effusion. Classifying knee joint effusion in the setting of acute trauma would be helpful for screening injuries in various knee structures because it suggests a high likelihood of ligament or meniscus injury [25, 26]. Therefore, we expect that the joint effusion model in this study can be generally applied to patients with acute trauma. To turn this expectation into reality, the proposed joint effusion model needs to be further validated in a dataset composed of various conditions besides ACL tears.

This study has several limitations. First, among the various series of knee radiographs, only the lateral view was used. We intended to develop DLM classifying radiographs with abnormal lateral femoral notch and joint effusion, which are best seen on lateral knee radiograph. As a result, anteroposterior views or stress radiographs were not used. Second, the normal radiographs were not confirmed by MRI. There might have been hidden ACL tears or pathologies in images classified as normal. Third, our dataset consisted of only two groups of subjects: normal and ACL tears. Inclusion of other conditions would reduce the diagnostic performance of the DLMs because joint effusion or Hoffa's fat pad abnormality is not specific for ACL tears. To be applied in routine clinical settings, further validation of the model in datasets comprising various conditions is needed. Fourth, the DLM and human reviewers were blinded to the contralateral knee radiographs and trauma history of the subjects. In the real practice, comparison of the radiographs with the contralateral side and referring to trauma history would help diagnose ACL tears. Fifth, our dataset was limited to optimal images of young patients aged 45 years or below. In addition, images with interval of longer than 15 days from MRI or images with bone tumor were excluded. Exclusion of potentially confounding images might restrict the clinical utility of the DLM. Sixth, radiographs were manually cropped. Automated cropping process would be needed to improve workflow of the proposed DLM. Finally, an external validation of the models would be required to generalize our results.

In conclusion, the diagnostic performance of the ensemble DLM in predicting lateral knee radiographs with ACL tear was comparable to that of a musculoskeletal radiologist. By combining the models classifying radiographs with indirect signs of ACL tears, the ensemble model outperformed the single model in classifying images with ACL tears.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00256-022-04081-x.

Acknowledgements The authors thank Kangwhi Lee and Jung Oh Lee for data analysis.

Declarations

Ethics approval Approval from the Institutional Review Board was obtained, and in keeping with the policies for a retrospective review, informed consent was not required.

Conflict of interest The authors declare no competing interests.

References

1. Murrell GA, Maddali S, Horovitz L, Oakley SP, Warren RF. The effects of time course after anterior cruciate ligament injury in correlation with meniscal and cartilage loss. Am J Sports Med. 2001;29(1):9–14.
2. Granan L-P, Bahr R, Lie SA, Engebretsen L. Timing of anterior cruciate ligament reconstructive surgery and risk of cartilage lesions and meniscal tears: a cohort study based on the Norwegian National Knee Ligament Registry. Am J Sports Med. 2009;37(5):955–61.

3. Lubowitz JH, Bernardini BJ, Reid JB III. Current concepts review: comprehensive physical examination for instability of the knee. Am J Sports Med. 2008;36(3):577–94.
4. Pao DG. The lateral femoral notch sign. Radiology. 2001;219(3):800–1.
5. Grimberg A, Shirazian H, Torshizy H, Smitaman E, Chang EY, Resnick D. Deep lateral notch sign and double notch sign in complete tears of the anterior cruciate ligament: MR imaging evaluation. Skeletal Radiol. 2015;44(3):385–91.
6. Nakauchi M, Kurosawa H, Kawakami AJ. Abnormal lateral notch in knees with anterior cruciate ligament injury. J Orthop Sci. 2000;5(2):92–5.
7. Frobell R, Le Graverand M-P, Buck R, et al. The acutely ACL injured knee assessed by MRI: changes in joint fluid, bone marrow lesions, and cartilage during the first year. Osteoarthr Cartil. 2009;17(2):161–7.
8. Yoon AP, Lee Y-L, Kane RL, Kuo C-F, Lin C, Chung KC. Development and Validation of a Deep Learning Model Using Convolutional Neural Networks to Identify Scaphoid Fractures in Radiographs. JAMA Netw Open. 2021;4(5):e216096–e216096.
9. Park HS, Jeon K, Cho YJ, et al. Diagnostic Performance of a New Convolutional Neural Network Algorithm for Detecting Developmental Dysplasia of the Hip on Anteroposterior Radiographs. Korean J Radiol. 2021;22(4):612.
10. Nam JG, Park S, Hwang EJ, et al. Development and validation of deep learning–based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. Radiology. 2019;290(1):218–28.
11. Ganaie MA, Hu M, Tanveer M, Suganthan PN. Ensemble deep learning: A review. arXiv. 2021; arXiv:2104.02395. https://arxiv.org/abs/2104.02395. Accessed 4 Oct 2021.
12. Sanders TL, Maradit Kremers H, Bryan AJ, et al. Incidence of anterior cruciate ligament tears and reconstruction: a 21-year population-based study. Am J Sports Med. 2016;44(6):1502–7.
13. Yoon JP, Chang CB, Yoo JH, et al. Correlation of magnetic resonance imaging findings with the chronicity of an anterior cruciate ligament tear. J Bone Joint Surg Am. 2010;92(2):353–60.
14. Yoon JP, Yoo JH, Chang CB, et al. Prediction of chronicity of anterior cruciate ligament tear using MRI findings. Clin Orthop Surg. 2013;5(1):19–25.
15. Hall FM. Radiographic diagnosis and accuracy in knee joint effusions. Radiology. 1975;115(1):49–54.
16. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: 2017 IEEE Int Conf Comput Vis. IEEE; 2017: 618–26. https://doi.org/10.1109/ICCV.2017.74
17. Crawshaw M. Multi-task learning with deep neural networks: A survey. arXiv. 2020; arXiv:2009.09796. https://arxiv.org/abs/2009.09796. Accessed 11 May 2022.
18. Zhou Y, Chen H, Li Y, et al. Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. Med Image Anal. 2021;70:101918.
19. Wu B, Zhou Z, Wang J, Wang Y. Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction. IEEE 15th Int Symp Biomed Imaging. 2018; 1109–13, https://doi.org/10.1109/ISBI.2018.8363765.
20. Hussein S, Cao K, Song Q, Bagci U. Risk stratification of lung nodules using 3D CNN-based multi-task learning. In: 2017 Int Conf Inf Process Med Imaging. Springer Verlag; 2017: 249–60. https://doi.org/10.1007/978-3-319-59050-9_20
21. Xing J, Chen C, Lu Q, et al. Using BI-RADS stratifications as auxiliary information for breast masses classification in ultrasound images. IEEE J Biomed Health Inform. 2020;25(6):2058–70.
22. Herbst E, Hoser C, Tecklenburg K, et al. The lateral femoral notch sign following ACL injury: frequency, morphology and relation to meniscal injury and sports activity. Knee Surg Sports Traumatol Arthrosc. 2015;23(8):2250–8.
23. Abreu MR, Chung CB, Trudell D, Resnick D. Hoffa's fat pad injuries and their relationship with anterior cruciate ligament tears: new observations based on MR imaging in patients and MR imaging and anatomic correlation in cadavers. Skeletal Radiol. 2008;37(4):301–6.
24. Guenoun D, Le Corroller T, Amous Z, et al. The contribution of MRI to the diagnosis of traumatic tears of the anterior cruciate ligament. Diagn Interv Imaging. 2012;93(5):331–41.
25. Calmbach WL, Hutchens M. Evaluation of patients presenting with knee pain: Part I. History, physical examination, radiographs, and laboratory tests. Am Fam Physician. 2003;68(5):907–12.
26. Sternbach GL. Evaluation of the knee. J Emerg Med. 1986;4(2):133–43.