

# One-stop detection of anterior cruciate ligament injuries on magnetic resonance imaging using deep learning with multicenter validation

Mei Wang<sup>1,2#</sup>, Congjing Yu<sup>3#</sup>, Mianwen Li<sup>1</sup>, Xinru Zhang<sup>1</sup>, Kexin Jiang<sup>1</sup>, Zhiyong Zhang<sup>3</sup>, Xiaodong Zhang<sup>1</sup>

<sup>1</sup>Department of Medical Imaging, The Third Affiliated Hospital of Southern Medical University (Academy of Orthopedics Guangdong Province), Guangzhou, China; <sup>2</sup>Department of Radiology, Guangzhou First People's Hospital, Guangzhou, China; <sup>3</sup>School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen, China

*Contributions:* (I) Conception and design: M Wang, C Yu, Z Zhang, X Zhang; (II) Administrative support: X Zhang; (III) Provision of study materials or patients: M Wang, C Yu; (IV) Collection and assembly of data: M Wang, C Yu, M Li, X Zhang; (V) Data analysis and interpretation: M Wang, C Yu, K Jiang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work.

*Correspondence to:* Zhiyong Zhang, PhD. School of Electronics and Communication Engineering, Sun Yat-sen University, 66 Gongchang Road, Guangming District, Shenzhen 518107, China. Email: zhangzhy99@mail.sysu.edu.cn; Xiaodong Zhang, MD, PhD. Department of Medical Imaging, The Third Affiliated Hospital of Southern Medical University (Academy of Orthopedics Guangdong Province), 183 Zhongshan Ave W, Guangzhou 510630, China. Email: ddautumn@126.com.

**Background:** Anterior cruciate ligament (ACL) injuries are closely associated with knee osteoarthritis (OA). However, diagnosing ACL injuries based on knee magnetic resonance imaging (MRI) has been subjective and time-consuming for clinical doctors. Therefore, we aimed to devise a deep learning (DL) model leveraging MRI to enable a comprehensive and automated approach for the detection of ACL injuries.

**Methods:** A retrospective study was performed extracting data from the Osteoarthritis Initiative (OAI). A total of 1,589 knees (comprising 1,443 intact, 90 with partial tears, and 56 with full tears) were enrolled to construct the classification model. This one-stop detection pipeline was developed using a tailored YOLOv5m architecture and a ResNet-18 convolutional neural network (CNN) to facilitate tasks based on sagittal 2-dimensional (2D) intermediate-weighted fast spin-echo sequence at 3.0T. To ensure the reliability and robustness of the classification system, it was subjected to external validation across 3 distinct datasets. The accuracy, sensitivity, specificity, and the mean average precision (mAP) were utilized as the evaluation metric for the model performance by employing a 5-fold cross-validation approach. The radiologist's interpretations were employed as the reference for conducting the evaluation.

**Results:** The localization model demonstrated an accuracy of 0.89 and a sensitivity of 0.93, achieving a mAP score of 0.96. The classification model demonstrated strong performance in detecting intact, partial tears, and full tears at the optimal threshold on the internal dataset, with sensitivities of 0.941, 0.833, and 0.929, specificities of 0.925, 0.947, and 0.991, and accuracies of 0.940, 0.941, and 0.989, respectively. In comparison, on a subset consisting of 171 randomly selected knees from the OAI, the radiologists demonstrated a sensitivity ranging between 0.660 and 1.000, specificity ranging between 0.691 and 1.000, and accuracy ranging between 0.689 and 1.000. On a subset consisting of 170 randomly selected knees from the Chinese dataset, the radiologists exhibited a sensitivity ranging between 0.711 and 0.948, specificity ranging between 0.768 and 0.977, and accuracy ranging between 0.683 and 0.917. After retraining, the model achieved sensitivities ranging between 0.630 and 0.961, specificities ranging between 0.860 and 0.961, and accuracies ranging between 0.832 and 0.951, respectively, on the external validation dataset.

**Conclusions:** The proposed model utilizing knee MRI showcases robust performance in the domains of ACL localization and classification.

**Keywords:** Deep learning (DL); magnetic resonance imaging (MRI); anterior cruciate ligament (ACL); convolutional neural network (CNN)

Submitted Oct 30, 2023. Accepted for publication Mar 14, 2024. Published online Apr 10, 2024.

doi: 10.21037/qims-23-1539

**View this article at:** <https://dx.doi.org/10.21037/qims-23-1539>

## Introduction

The anterior cruciate ligament (ACL) is the most commonly affected structure within the knee, accounting for over 50% of knee injuries (1). An ACL injury increases the susceptibility to post-traumatic knee osteoarthritis (OA) and leads to a higher demand for total knee replacement (TKR) (2-5). Thus, an accurate diagnosis of ACL injuries is essential to initiate optimal treatments that reduce the occurrence of knee instability and improve the quality of life. Although arthroscopy is a standard diagnosis for ACL injury, it is an invasive procedure (6). Magnetic resonance imaging (MRI) has ascended as the preferred modality for evaluating ACL injuries, owing to its exceptional soft tissue resolution, absence of ionizing radiation, and the capability for multiparametric imaging. Usually, radiologists rely on visual assessment of the morphology and signal attributes of MRI scans to diagnose ACL injuries. According to recent literature (7,8), MRI has a sensitivity of 87%, specificity of 90%, and the area under the curve (AUC) of 0.93 for the diagnosis of ACL injury. Nevertheless, this outstanding diagnostic accuracy heavily hinges on the extensive expertise of radiologists. In clinical practice, the growing volume of medical images has resulted in increased workloads for diagnostic radiologists, longer patient waiting times, and reduced efficiency. Meanwhile, a lack of diagnostic experience and considerable anatomical variation increases the likelihood of missed diagnoses or misdiagnoses, particularly in situations of high workload or when radiologists are fatigued.

Deep learning (DL) employs automatic feature learning and modeling of complex relationships between medical images and their interpretations (9-11). In recent years, DL has shown promising applications in medical imaging (12), particularly in the field of musculoskeletal (MSK) radiology (13), such as tissue segmentation, disease detection, classification, and prediction (14). Researchers are leveraging DL techniques, employing the convolutional neural network (CNN) models

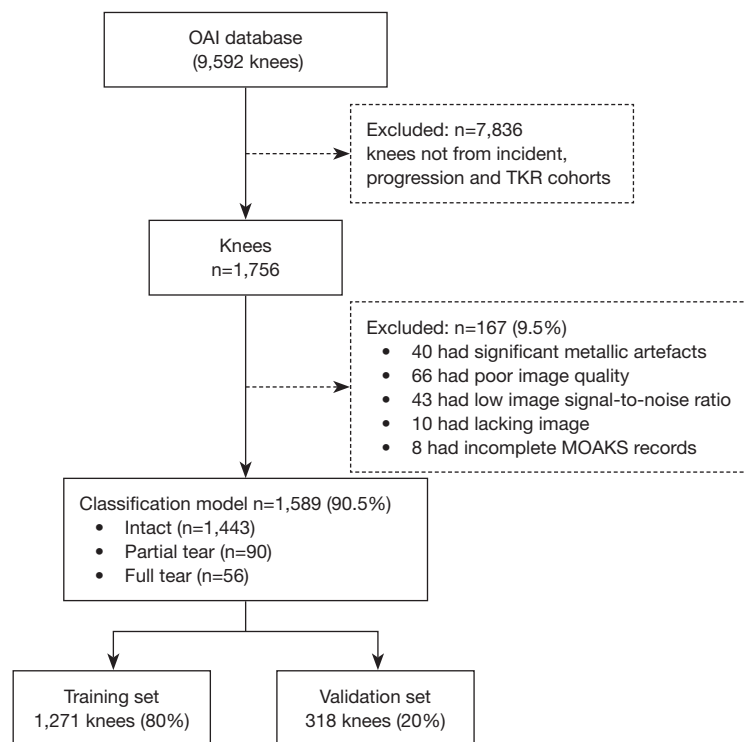
and their architectures across diverse applications. These CNN architectures typically encompass an input layer, an output layer, and multiple convolutional layers, pooling layers, rectified linear unit (ReLU) layers, dense layers, as well as dropout layers (15,16). The CNN shows huge success in the analysis of multi-classification diagnosis of ACL injuries. However, the sensitivity of the CNN model, particularly in distinguishing between partial and full tears, was reported to be low in studies conducted by Astuto (17), Namiri (18), and others. Furthermore, these studies did not validate the applicability of the model in external datasets and were often conducted using the images of a single manufacturer.

YOLOv5m (19) is a widely recognized object detection network known for its effectiveness on comprehensive datasets. ResNet-18 (20) is a currently superior model in image classification and recognition, and we hypothesize that more accurate ACL injury classification can be achieved based on YOLOv5m and ResNet-18. Therefore, in this study, a new fully automated DL model, modified from YOLOv5m and ResNet-18 network based on MRI, was devised for “one-stop” multi-tasking of ACL localization and accurate injury classification to enable clinical application. The model was validated in a multicenter dataset. We present this article in accordance with the TRIPOD reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-1539/rc>).

## Methods

### *Study design*

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). All of the training image data was obtained from the Osteoarthritis Initiative (OAI; <https://oai.ucsf.edu/>), a comprehensive multicenter, longitudinal, 10-year prospective cohort study. This study encompassed 4,796 patients aged between 45



**Figure 1** Study flowchart and inclusion and exclusion criteria for the classification model. OAI, Osteoarthritis Initiative; TKR, total knee replacement; MOAKS, MRI Osteoarthritis Knee Score; MRI, magnetic resonance imaging.

and 79 years, all of whom exhibited symptoms of OA, or were at least at risk of its development in 1 knee. The participant inclusion and exclusion criteria, imaging procedures, and assessments for the OAI study have been meticulously documented in prior works (21).

### Internal dataset

The internal dataset included 1,756 knees from the incident cohort, the TKR cohort, and the progression cohort within the OAI dataset. The exclusion criteria were significant metallic artefacts, poor image quality, low image signal-to-noise ratio, lacking image and incomplete MRI Osteoarthritis Knee Score (MOAKS) records. As a result, 167 knees were further excluded. Finally, a total of 1,443 intact ligaments, 90 partial tears, and 56 full tears met the inclusion criteria. The ground truth data used for labeling were derived from MOAKS image evaluations conducted by 2 radiologists at the Boston Imaging Core Lab (22,23). Labels for ACL status can be accessed for download from the following website: <https://oai.epi-ucsf.org/datarelease/>. A full tear was defined as a condition in which there was a total disruption of ACL

fibers, resulting in ligament discontinuity. A partial tear was defined as the presence of residual straight or taut ACL fibers in at least 1 pulse sequence. Signal alterations indicative of mucoid degeneration were excluded from the definitions. Intraligamentous hyperintense signal changes, not accompanied by apparent thinning or discontinuity of the ligament, which are characteristic of mucoid degeneration of the ACL, were not individually scored as per the MOAKS scoring system. Therefore, mucoid degeneration was considered within the “normal” spectrum, as the primary focus was on morphological abnormalities of the ligament consistent with partial or complete fiber disruption. *Figure 1* visually illustrates the study flow and the specific inclusion/exclusion criteria applied for the construction of our classification model.

### External dataset

For external validation purposes, we leveraged 3 distinct datasets: MRNet (USA) (24), KneeMRI (Croatia) (25), and a Chinese dataset. The MRNet (USA) dataset, comprising 988 knees with intact ACLs and 262 knees with ACL injuries, was employed to validate our binary classification

model. The KneeMRI dataset from Croatia consists of 917 knee samples, comprising 690 knees with intact ACLs, 172 knees with partial tears, and 55 knees with full tears. All these MRI examinations were labeled based on both report and additional reading by a radiologist. For a more detailed description of these 2 external validation datasets, please refer to [Appendix 1](#) (available online), which provides a comprehensive elaboration.

Knee images from patients who underwent MRI examinations between October 2011 and March 2022 were retrospectively collected by querying the picture archiving and communication system (PACS) of The Third Affiliated Hospital of Southern Medical University (Guangzhou, China). The collection of this dataset was approved by the Ethics Board of The Third Affiliated Hospital of Southern Medical University (No. 2013012), and individual consent for this retrospective analysis was waived. All participants were 18–80 years of age. This dataset, referred to as “Chinese dataset” initially queried a total of 489 cases. After excluding 153 knees based on the exclusion criteria, a total of 336 knees were ultimately included. The dataset consists of 207 knees with intact ACLs (approximately 62%), 65 with partial tears (about 21%), and 64 with complete tears (roughly 19%). The ground truth for the dataset was established by 2 radiologists, each possessing more than 2 decades of experience in diagnostic MSK imaging. In cases where ambiguity arose, a consensus was reached through discussions to establish the definitive diagnostic opinion. The diagnostic criteria employed by the radiologists were consistent with those used in the OAI dataset.

### *MRI data acquisition*

MR images from the OAI database were obtained using 4 identical 3.0 Tesla scanners (Magnetom Trio; Siemens, Erlangen, Germany) located in Columbus, Ohio; Baltimore, Maryland; Pittsburgh, Pennsylvania; and Pawtucket, Rhode Island. The following sequences were acquired: sagittal 2-dimensional (2D) intermediate-weighted fast spin-echo sequence with a repetition time/echo time (TR/TE) of 3,200/30 ms, spatial resolution of 0.357 mm × 0.511 mm, and slice thickness of 3.0 mm. Further details of the MRI protocol and evaluation of 3 external datasets are described in [Appendix 2](#) (online).

### *Deep CNN*

The developed DL model consists of 2 essential

components: the first component is responsible for localizing the ACL, whereas the second component classifies the ACL as intact, partial tears, or full tears. For a visual representation of our model, please refer to [Figure 2](#).

### *Image labeling*

M.W., a radiology resident with 3 years of clinical experience manually outlined the intercondylar notch region containing the ACL on selected MR image sections. Following this, Junjie Guo, a radiologist with 7 years of clinical experience, reviewed the annotated images.

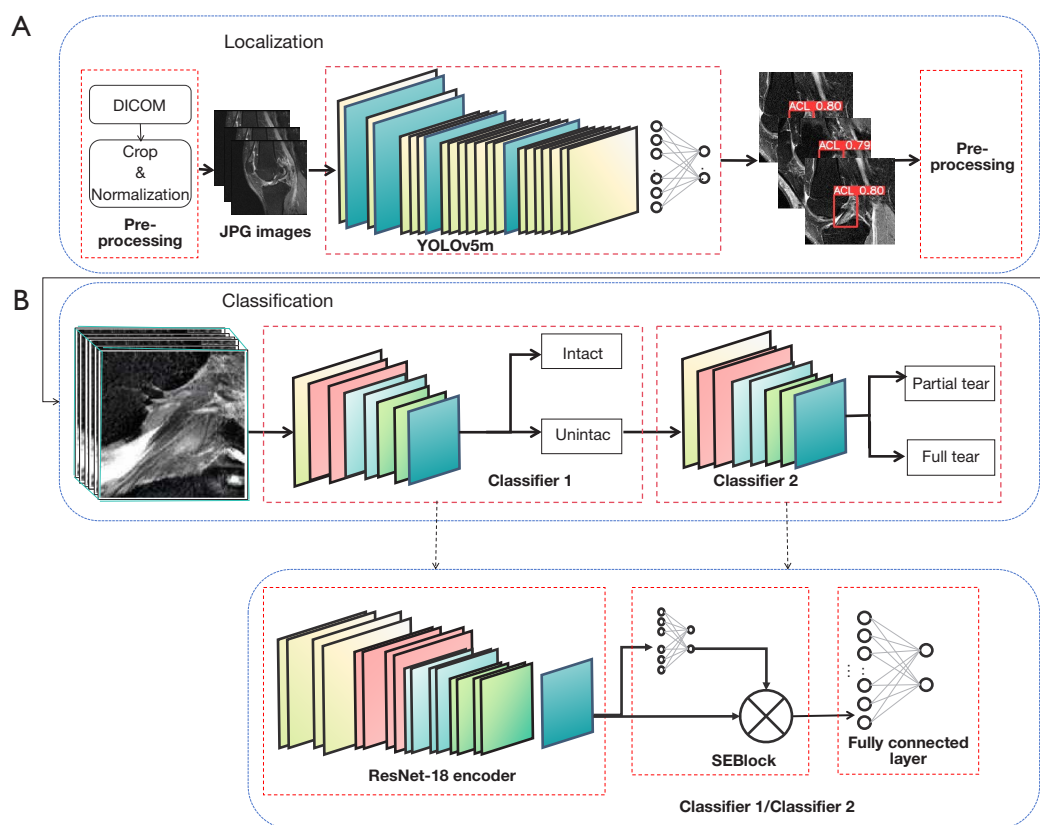
### *Pre-processing*

Initially, we limited our dataset to include only slices 15 through 23 of each case, focusing on the middle portion with a size of 256×256 pixels to narrow down the input scope. Following this, we removed the top and bottom 1% of the grayscale values from each slice and conducted grayscale min-max normalization. Subsequently, we converted the images into a single-channel JPG format, which was used as the input. These steps are all automated.

### *ACL localization*

We employed the YOLOv5m (19) architecture to design our ACL localization model, which was implemented using the PyTorch V1.11.0 package (<https://pytorch.org/>). We enhanced the accuracy of ACL detection across multiple slices within an MRI series by introducing a post-processing technique. The specific algorithm for post-processing involves retaining only the prediction boxes with the highest confidence in each slice as a preliminary result, and then setting 2 confidence thresholds: `conf_LB = 0.35` and `conf_thred = 0.5`. We considered slices with confidence between `conf_LB = 0.35` and `conf_thred = 0.5` as candidate predictions. If the number of prediction boxes with confidence greater than `conf_thred = 0.5` in a case was fewer than 2, we included these candidate prediction boxes. Finally, since the ACL must appear on continuous MRI slices, we included any dropped slices, regardless of the confidence level predicted by YOLOv5m, to ensure that the selected slices are continuous.

We seamlessly integrated the post-processing methodology into the YOLO detection code, subsequently applying the modified model to compute the region of interest (ROI) across the complete dataset. After determining



**Figure 2** Overview of the one-stop detection model pipeline. The proposed method consisted of two separate systems connected in a cascaded fashion to create a fully automated image processing pipeline. (A) ACL localization model. (B) ACL injury classification. DICOM, Digital Imaging and Communications in Medicine; SE, Squeeze-and-Excitation; ACL, anterior cruciate ligament.

the ROI, we cropped the images, resized them to dimensions of 128×96 pixels, and then applied z-score normalization before inputting them into the classification network.

### ACL classification

The classification network was an upgraded iteration of the MRNet (24). In this improvement, we replaced its backbone network AlexNet (26) with ResNet-18 (20). Once the backbone network had extracted feature maps from each 2D image, a global average pooling layer was applied to reduce these feature maps into feature vectors.

To effectively utilize information from multiple slices, we replaced the subsequent max-pooling layer in MRNet (24) with Squeeze-and-Excitation (SE) blocks (27). In our SE block, the feature vector of each slice was first mapped to a value using a fully connected layer. Then, these values were normalized through the softmax function to obtain weights for the corresponding slice. Finally, the feature vectors were

weighted summed according to these weights to obtain a comprehensive vector. By incorporating the SE block, the network can assign varying levels of attention to different slices, effectively integrating information from multiple slices into the final comprehensive vector.

The final layer of the network featured a fully connected layer employing a sigmoid activation function, facilitating the prediction of probabilities within the 0 to 1 range. In order to enhance the model's performance, data augmentation techniques were employed as part of this study. Within the training dataset, images underwent augmentation through random image translations ( $\pm 15$  pixels) both horizontally and vertically, as well as random rotations ( $\pm 15^\circ$ ). In the training of each fold, the partial tears and full tears in the training set were augmented by a factor of 5 and 9, respectively. These augmented images were uniformly cropped using the ACL localization model and subsequently incorporated into the classification network's training set, augmenting its robustness. However,

**Table 1** Comparison of diagnostic performance of different models

| Severity     | Study                       | Sensitivity (95% CI) | Specificity (95% CI) | Accuracy (95% CI)   |
|--------------|-----------------------------|----------------------|----------------------|---------------------|
| Intact       | Proposed model              | 0.941 (0.926–0.956)  | 0.925 (0.903–0.947)  | 0.940 (0.927–0.952) |
|              | Namiri <i>et al.</i> (2020) | 0.835 (0.722–0.948)  | 0.744 (0.595–0.894)  | 0.839 (0.739–0.939) |
|              | Astuto <i>et al.</i> (2021) | 0.835 (0.722–0.948)  | 0.744 (0.595–0.894)  | 0.839 (0.739–0.939) |
| Partial tear | Proposed model              | 0.833 (0.752–0.915)  | 0.947 (0.931–0.964)  | 0.941 (0.928–0.954) |
|              | Namiri <i>et al.</i> (2020) | 0.744 (0.594–0.894)  | 0.864 (0.747–0.981)  | 0.857 (0.754–0.961) |
|              | Astuto <i>et al.</i> (2021) | 0.744 (0.594–0.894)  | 0.864 (0.747–0.981)  | 0.857 (0.754–0.961) |
| Full tear    | Proposed model              | 0.929 (0.836–1.000)  | 0.991 (0.986–0.996)  | 0.989 (0.985–0.992) |
|              | Namiri <i>et al.</i> (2020) | 0.928 (0.836–1.000)  | 0.970 (0.957–0.983)  | 0.969 (0.958–0.979) |
|              | Astuto <i>et al.</i> (2021) | 0.928 (0.836–1.000)  | 0.970 (0.957–0.983)  | 0.969 (0.958–0.979) |

The sensitivity, specificity, and accuracy of the models of Namiri *et al.* and Astuto *et al.* were obtained using 5-fold cross-validation based on the OAI dataset we constructed. CI, confidence interval; OAI, Osteoarthritis Initiative.

our model did not utilize augmented data during validation and metric computation.

### Evaluation of radiologists

In this study, we conducted a comparative analysis of the diagnostic performance of a classification model for ACL injuries with that of MSK radiologists. An experienced radiologist with 10 years of experience (Reader 1) and a first-year radiology resident (Reader 2) independently reviewed original whole MRI of 171 randomly selected knees from the OAI dataset and 170 knees from the Chinese dataset to determine the type of ACL injuries. For honest comparison, 30 cases of whole knee image with labels from the OAI dataset and Chinese dataset were provided for the MSK radiologists to train to perform the diagnostics before the actual evaluation. The radiologists also did not utilize augmented datasets during the assessment of the images. As a comparison, our classification model was tested on the abovementioned cases.

### Statistical analysis

The statistical analysis was conducted using the software PyCharm 2021.2.3 (<https://www.jetbrains.com/pycharm/download/?section=windows>). To evaluate the accuracy of the ACL localization model, we used the intersection over union (IoU) index, with the radiology resident's labeling results as the gold standard. Model performance was further assessed using metrics such as average precision

(AP), sensitivity, and the mean average precision (mAP). In evaluating the classification model for diagnosing ACL injury, we calculated sensitivity, specificity, and accuracy based on the threshold value corresponding to the Youden index as the performance metrics. To assess the reliability of the diagnoses between radiologists, the intraclass correlation coefficient (ICC) was used. This study employed 5-fold cross-validation, and all reported results represent the average of 5 evaluations.

## Results

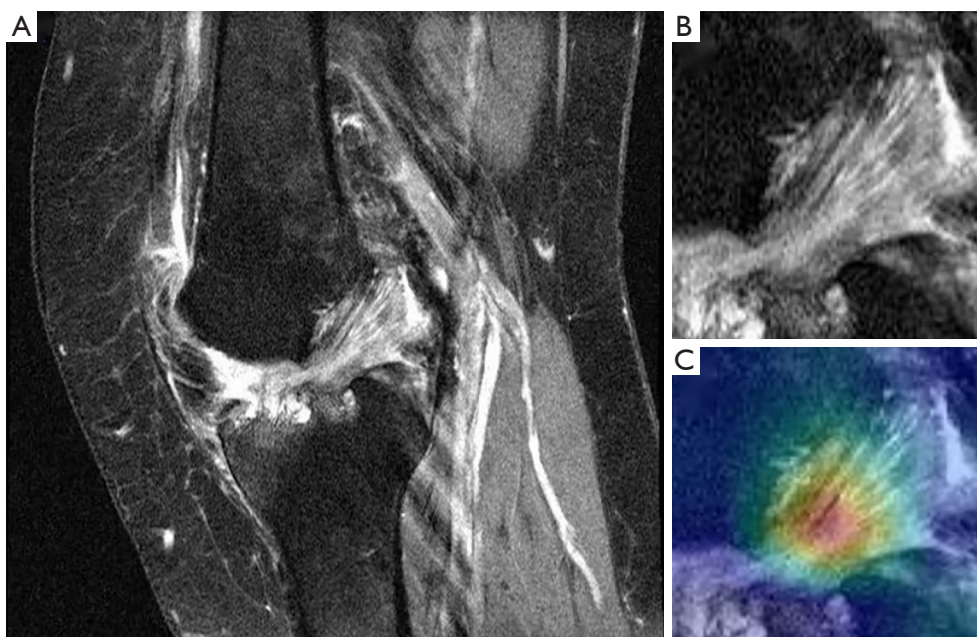
### ACL localization model performance

The precision of the object detection network in the validation set was 0.89 and the sensitivity was 0.93. The mAP scores of the model (IoU threshold  $\geq 0.5$ ) was 0.96.

### ACL classification model performance

#### Performance on the OAI dataset

Table 1 presents a comparative analysis of our model's performance alongside the studies conducted by Namiri *et al.* (18) and Astuto *et al.* (17). Our model achieved noteworthy levels of sensitivity, with values of 0.941, 0.833, and 0.929 for intact ligaments, partial tears, and full tears, respectively. Additionally, our model exhibited high specificity values of 0.925, 0.947, and 0.991, and accuracy values of 0.940, 0.941, and 0.989, respectively, for the corresponding ACL conditions. In order to objectively



**Figure 3** Sagittal MRI views of (A) a correctly classified knee with a partial tear ACL and (B) its ACL localization with, (C) probability plot showing the high probability region in the ACL on which our model is based to explain ACL tears. MRI, magnetic resonance imaging; ACL, anterior cruciate ligament.

compare the results of the models, we obtained new sensitivities, specificities, and accuracies for the studies of Namiri *et al.* and Astuto *et al.* based on our internal dataset using 5-fold cross-validation. To visually illustrate the performance of our model, we provide a case study in *Figure 3*, depicting a knee with a correctly identified partial ACL tear. This correct case demonstrates how our pipeline adeptly localizes the ACL and generates a saliency map that effectively emphasizes the high-intensity features of the ligament.

### Diagnosis by radiologists

The ICC value between reader 1 and reader 2 were 0.79 and 0.81 on the OAI subset (n=171) and the subset of Chinese dataset (n=170), respectively. *Table 2* shows the results of the comparison of the reader 1, reader 2 and ACL injury classification models. The model performed superiorly in detecting ACL tears compared to the readers diagnostic results.

### Performance on the external validation datasets

*Table 3* presents the performance of the model on external

datasets, both before and after retraining. After retraining, the model demonstrated impressive sensitivity (0.935) and specificity (0.961) in effectively identifying untear ACLs within the MRNet (USA) dataset. On the KneeMRI (Croatia) multi-class dataset, the model exhibited accuracies of 0.866, 0.832, and 0.928 for intact, partial tear, and full tear ACLs, respectively. On the Chinese dataset, the model achieved accuracies of 0.860, 0.866, and 0.898 for intact, partial tears, and full tears, respectively. These results indicate the model's robust performance across a range of external datasets after retraining.

### Discussion

In this study, we present a comprehensive detection system designed for accurate ACL localization and the efficient multi-classification of ACL injuries, accompanied by an in-depth exploration of relevant literature. This is achieved through the implementation of the YOLOv5m and ResNet-18 network architecture, complemented by a rigorous multi-center external data validation approach.

Prior to classifying ACL injuries, the extraction of the ROI can effectively mitigate interference from adjacent regions, such as bone, thereby leading to a significant

**Table 2** Comparison of the results of MRI diagnosis of ACL tears by 2 readers and the proposed model

| Dataset                           | Models and human readers | Class name   | Evaluation metrics   |                      |                     |
|-----------------------------------|--------------------------|--------------|----------------------|----------------------|---------------------|
|                                   |                          |              | Sensitivity (95% CI) | Specificity (95% CI) | Accuracy (95% CI)   |
| OAI subset (n=17)                 | Reader 1                 | Intact       | 0.672 (0.596–0.748)  | 1.000 (1.000–1.000)  | 0.701 (0.630–0.772) |
|                                   |                          | Partial tear | 0.664 (0.330–0.998)  | 0.691 (0.618–0.764)  | 0.689 (0.618–0.761) |
|                                   |                          | Full tear    | 0.999 (0.937–1.061)  | 0.976 (0.953–0.999)  | 0.977 (0.955–0.999) |
|                                   | Reader 2                 | Intact       | 0.942 (0.907–0.978)  | 0.934 (0.803–1.000)  | 0.942 (0.907–0.976) |
|                                   |                          | Partial tear | 0.660 (0.320–0.999)  | 0.944 (0.910–0.978)  | 0.930 (0.893–0.967) |
|                                   |                          | Full tear    | 1.000 (1.000–1.000)  | 0.988 (0.971–1.000)  | 0.988 (0.972–1.000) |
|                                   | Proposed model           | Intact       | 0.936 (0.897–0.976)  | 1.000 (1.000–1.000)  | 0.942 (0.905–0.978) |
|                                   |                          | Partial tear | 1.000 (1.000–1.000)  | 0.939 (0.900–0.977)  | 0.942 (0.905–0.978) |
|                                   |                          | Full tear    | 1.000 (1.000–1.000)  | 1.000 (1.000–1.000)  | 1.000 (1.000–1.000) |
| Subset of Chinese dataset (n=170) | Reader 1                 | Intact       | 0.711 (0.654–0.799)  | 0.977 (0.824–1.000)  | 0.683 (0.611–0.794) |
|                                   |                          | Partial tear | 0.752 (0.714–0.845)  | 0.793 (0.696–0.895)  | 0.801 (0.768–0.916) |
|                                   |                          | Full tear    | 0.946 (0.858–0.971)  | 0.959 (0.893–0.994)  | 0.917 (0.819–0.947) |
|                                   | Reader 2                 | Intact       | 0.948 (0.876–0.969)  | 0.899 (0.796–0.932)  | 0.914 (0.845–0.977) |
|                                   |                          | Partial tear | 0.748 (0.713–0.901)  | 0.789 (0.645–0.857)  | 0.875 (0.819–0.973) |
|                                   |                          | Full tear    | 0.916 (0.877–0.975)  | 0.768 (0.778–0.839)  | 0.877 (0.811–0.965) |
|                                   | Proposed model           | Intact       | 0.874 (0.810–0.922)  | 0.818 (0.746–0.895)  | 0.859 (0.796–0.919) |
|                                   |                          | Partial tear | 0.731 (0.717–0.855)  | 0.882 (0.794–0.933)  | 0.852 (0.816–0.939) |
|                                   |                          | Full tear    | 0.714 (0.655–0.836)  | 0.975 (0.886–0.997)  | 0.961 (0.913–0.998) |

MRI, magnetic resonance imaging; ACL, anterior cruciate ligament; CI, confidence interval. OAI, Osteoarthritis Initiative.

improvement in accuracy (28). A study by Liu *et al.* (29) developed 2 separate convolutional networks for ACL detection (level selection) and localization (determination of bounding box coordinates). However, this approach increased the training burden. Germann *et al.* (30) proposed a convolutional network that identified the layers containing ACLs, but the subsequent cropping step was still manual and imprecise, resulting in uniform-sized regions of 123×320 pixels along the ACL. Namiri *et al.* (18) introduced a 3-dimensional (3D) segmentation network that divided the entire knee into 11 small regions, using the background region as the ROI for ACL. However, this localization method was considered indirect and inaccurate. Therefore, the existing studies were insufficient in accurately locating the ACL.

The YOLO network is a well-established framework in the field of object detection for natural images, and its effectiveness extends to various diagnostic modalities.

Notably, it has been successfully applied to tasks such as the automated identification of lung nodules in computed tomography images (31) and the detection of nodules in X-ray mammography. Expanding on the YOLO network, we have introduced several enhancements to address the limitations observed in earlier studies. Firstly, we have developed a method for selecting a rectangular bounding box for segmentation, with a specific focus on the ACL. This approach not only reduces segmentation loss but also accurately identifies torn ACLs, even when their original morphology is altered. Secondly, we have incorporated post-processing techniques to enhance the automation, maturity, and efficiency of the YOLO localization network. This refinement allows for the simultaneous selection of the appropriate level of detail and precise localization of the ACL. Additionally, our model integrates adaptive resizing of the rectangular box, dynamically adjusting to the precise position of the ACL within each image. This adaptive



**Table 3** Model performance on external datasets

| Dataset          | Verification method | Class name   | Sensitivity (95% CI) | Specificity (95% CI) | Accuracy (95% CI)   |
|------------------|---------------------|--------------|----------------------|----------------------|---------------------|
| MRNet, USA       | Without retraining  | Intact       | 0.808 (0.698–0.918)  | 0.750 (0.597–0.903)  | 0.786 (0.698–0.874) |
|                  |                     | Unintact     | 0.750 (0.597–0.903)  | 0.808 (0.698–0.918)  |                     |
|                  | After retraining    | Intact       | 0.961 (0.908–1.000)  | 0.935 (0.849–1.000)  | 0.951 (0.906–0.997) |
|                  |                     | Unintact     | 0.935 (0.849–1.000)  | 0.961 (0.908–1.000)  |                     |
| KneeMRI, Croatia | Without retraining  | Intact       | 0.914 (0.835–0.994)  | 0.673 (0.558–0.788)  | 0.852 (0.809–0.895) |
|                  |                     | Partial tear | 0.351 (0.229–0.473)  | 0.924 (0.868–0.981)  | 0.810 (0.759–0.862) |
|                  |                     | Full tear    | 0.546 (0.191–0.902)  | 0.921 (0.868–0.974)  | 0.899 (0.853–0.944) |
|                  | After retraining    | Intact       | 0.868 (0.830–0.906)  | 0.860 (0.803–0.918)  | 0.866 (0.851–0.881) |
|                  |                     | Partial tear | 0.630 (0.504–0.757)  | 0.881 (0.849–0.914)  | 0.832 (0.817–0.847) |
|                  |                     | Full tear    | 0.721 (0.602–0.840)  | 0.940 (0.912–0.969)  | 0.928 (0.906–0.951) |
| Chinese dataset  | Without retraining  | Intact       | 0.773 (0.685–0.861)  | 0.790 (0.620–0.961)  | 0.779 (0.685–0.874) |
|                  |                     | Partial tear | 0.569 (0.290–0.849)  | 0.671 (0.611–0.732)  | 0.652 (0.563–0.740) |
|                  |                     | Full tear    | 0.188 (0.000–0.501)  | 0.959 (0.917–1.000)  | 0.812 (0.731–0.893) |
|                  | After retraining    | Intact       | 0.801 (0.777–0.827)  | 0.953 (0.927–0.979)  | 0.860 (0.840–0.880) |
|                  |                     | Partial tear | 0.800 (0.731–0.869)  | 0.882 (0.846–0.918)  | 0.866 (0.848–0.884) |
|                  |                     | Full tear    | 0.860 (0.795–0.925)  | 0.908 (0.885–0.931)  | 0.898 (0.889–0.909) |

CI, confidence interval; MRI, magnetic resonance imaging.

resizing significantly reduces the impact of irrelevant data, streamlining the processing workflow and reducing redundant operations. Collectively, these enhancements effectively address the limitations present in previous studies, resulting in a model that is not only more accurate and efficient but also more user-friendly for the detection and localization of the ACL.

Numerous studies (24,28,30) have delved into ACL injury classification tasks using DL frameworks. Although most of these studies (32,33) focused on lesion detection using binary classifiers, a small number of studies (34,35) explored multi-classification. Namiri *et al.* (18) developed a model based on 3D CNN and 2D CNN, achieving accurate multi-classification of ACL with accuracies of 0.89 and 0.92. However, their study only included 18 injuries. Astuto *et al.* (17) examined the sensitivity of a 3D DL model in predicting multiple classifications of ACL injuries, reporting a sensitivity of 0.75 for partial tears and 0.77 for full tears. In their pertinent multi-classification study, a central challenge was the underwhelming performance of the classification model, particularly with sensitivities of 0.75 for cases involving partial tears and full tears. Moreover,

their study solely relied on an internal dataset for model validation, lacking external validation, thus warranting improvements in generalization and model transferability.

In contrast, our model achieved remarkable sensitivities of 0.83 and 0.93 for the aforementioned categories. This improvement is attributed to the substitution of the classification network's maximum pooling layer with a SE block within our model's architecture. The SE block intelligently recalibrates feature responses by explicitly capturing interdependencies between channels. This approach enables the network to prioritize the most influential aspects for classification outcomes while amalgamating information from all image layers cohesively. Furthermore, our study distinguished itself by embracing a multi-center validation strategy, encompassing diverse databases originating from various global populations, distinct scanning equipment models, and sequences involving disparate parameter settings. Worth noting is our model's commendable validation performance on the Chinese dataset, affirming its robustness. Our investigation unveiled an initial dip in the model's performance during external validation, followed by a marked improvement

after retraining. This observation aligns with the findings of Germann *et al.* (30) and Tran *et al.* (36). The variations evident in scanning instruments, sequences, and imaging parameters across diverse datasets contributed to these outcomes. These disparities underscore the substantial diversity inherent in multi-center MRI data. Furthermore, it is common for models trained exclusively on single-center datasets to exhibit suboptimal performance on other datasets due to overfitting. Consequently, retraining the model, particularly with regards to detecting partial tears, becomes an essential step for optimal performance.

Furthermore, the system devised in our study can be seamlessly integrated with the MR scanning system. Digital Images and Communications in Medicine (DICOM) images from all sequences can serve as direct input data for the system, facilitating a step-by-step automatic analysis to yield a diagnostic result. The radiologist is only required to supervise the output diagnostic results, no additional operations are necessary. The system has the potential to be seamlessly connected with the scanning system, offering enhanced practicality and convenience in future applications.

There are several limitations in our study that should be considered. Firstly, the unbalanced data and the relatively limited number of injury cases could potentially lead to overfitting of the classification model. Secondly, our classification scheme focused solely on distinguishing partial and full tears, highlighting the need for future investigations to delve into a more nuanced and precise classification of partial injuries. Thirdly, it is important to acknowledge that our injury classification reference standard relied solely on diagnostic radiologists' assessments, lacking validation through arthroscopic evaluation. Lastly, we did not compare the diagnostic performance of radiologists to the model on all the datasets we constructed, which may have led to slightly biased results.

## Conclusions

This study introduces a fully automatic pipeline using DL based on MR images for the localization and classification of ACL injuries. The proposed system aims to enhance the efficiency of aiding radiologist readings and reduce intergrader variability.

## Acknowledgments

The authors would like to acknowledge participants in the

OAI study for providing this unique open-access database. The authors would like to thank Junjie Guo for reviewing the annotated images in this manuscript.

*Funding:* This project was supported by the Guang Dong Basic and Applied Basic Research Foundation (No. A1515010352) and President Foundation of The Third Affiliated Hospital of Southern Medical University (No. YM2021012).

## Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-23-1539/rc>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-1539/coif>). All authors report that this work was supported by the President Foundation of The Third Affiliated Hospital of Southern Medical University (No. YM2021012). The authors have no other conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The collection of this dataset was approved by the Ethics Board of The Third Affiliated Hospital of Southern Medical University (No. 2013012), and individual consent for this retrospective analysis was waived.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Al-Antari MA, Al-Masni MA, Kim TS. Deep Learning Computer-Aided Diagnosis for Breast Lesion in Digital Mammogram. *Adv Exp Med Biol* 2020;1213:59-72.

2. Prodromos CC, Han Y, Rogowski J, Joyce B, Shi K. A meta-analysis of the incidence of anterior cruciate ligament tears as a function of gender, sport, and a knee injury-reduction regimen. *Arthroscopy* 2007;23:1320-1325.e6.
3. Hunter DJ, Lohmander LS, Makovey J, Tamez-Peña J, Totterman S, Schreyer E, Frobell RB. The effect of anterior cruciate ligament injury on bone curvature: exploratory analysis in the KANON trial. *Osteoarthritis Cartilage* 2014;22:959-68.
4. Brophy RH, Gill CS, Lyman S, Barnes RP, Rodeo SA, Warren RF. Effect of anterior cruciate ligament reconstruction and meniscectomy on length of career in National Football League athletes: a case control study. *Am J Sports Med* 2009;37:2102-7.
5. Suter LG, Smith SR, Katz JN, Englund M, Hunter DJ, Frobell R, Losina E. Projecting Lifetime Risk of Symptomatic Knee Osteoarthritis and Total Knee Replacement in Individuals Sustaining a Complete Anterior Cruciate Ligament Tear in Early Adulthood. *Arthritis Care Res (Hoboken)* 2017;69:201-8.
6. Bari AA, Kashikar SV, Lakhkar BN, Ahsan MS. Evaluation of MRI versus arthroscopy in anterior cruciate ligament and meniscal injuries. *J Clin Diagn Res* 2014;8:RC14-8.
7. Li K, Du J, Huang LX, Ni L, Liu T, Yang HL. The diagnostic accuracy of magnetic resonance imaging for anterior cruciate ligament injury in comparison to arthroscopy: a meta-analysis. *Sci Rep* 2017;7:7583.
8. Challen J, Tang Y, Hazratwala K, Stuckey S. Accuracy of MRI diagnosis of internal derangement of the knee in a non-specialized tertiary level referral teaching hospital. *Australas Radiol* 2007;51:426-31.
9. Zhan H, Teng F, Liu Z, Yi Z, He J, Chen Y, Geng B, Xia Y, Wu M, Jiang J. Artificial Intelligence Aids Detection of Rotator Cuff Pathology: A Systematic Review. *Arthroscopy* 2024;40:567-78.
10. Xu M, Chen Z, Zheng J, Zhao Q, Yuan Z. Artificial intelligence-aided optical imaging for cancer theranostics. *Semin Cancer Biol* 2023;94:62-80.
11. Mirikharaji Z, Abhishek K, Bissoto A, Barata C, Avila S, Valle E, Celebi ME, Hamarneh G. A survey on deep learning for skin lesion segmentation. *Med Image Anal* 2023;88:102863.
12. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pinykh OS, Geis JR, Pandharipande PV, Brink JA, Dreyer KJ. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology* 2018;288:318-28.
13. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial Intelligence in Musculoskeletal Imaging: Current Status and Future Directions. *AJR Am J Roentgenol* 2019;213:506-13.
14. Ding J, Zhao R, Qiu Q, Chen J, Duan J, Cao X, Yin Y. Developing and validating a deep learning and radiomic model for glioma grading using multiplanar reconstructed magnetic resonance contrast-enhanced T1-weighted imaging: a robust, multi-institutional study. *Quant Imaging Med Surg* 2022;12:1517-28.
15. Al-Waisy AS, Al-Fahdawi S, Mohammed MA, Abdulkareem KH, Mostafa SA, Maashi MS, Arif M, Garcia-Zapirain B. COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images. *Soft comput* 2023;27:2657-72.
16. Mohammed MA, Abdulkareem KH, Mostafa SA, Khanapi Abd Ghani M, Maashi MS, Garcia-Zapirain B, Oleagordia I, Alhakami H, Al-Dhief FT. Voice pathology detection and classification using convolutional neural network model. *Appl Sci* 2020;10:3723.
17. Astuto B, Flament I, Namiri NK, Shah R, Bharadwaj U, Link TM, Bucknor MD, Pedoia V, Majumdar S. Erratum: Automatic Deep Learning-assisted Detection and Grading of Abnormalities in Knee MRI Studies. *Radiol Artif Intell* 2021;3:e219001.
18. Namiri NK, Flament I, Astuto B, Shah R, Tibrewala R, Caliva F, Link TM, Pedoia V, Majumdar S. Deep Learning for Hierarchical Severity Staging of Anterior Cruciate Ligament Injuries from MRI. *Radiol Artif Intell* 2020;2:e190207.
19. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016:779-88.
20. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:770-8.
21. Peterfy CG, Schneider E, Nevitt M. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthritis Cartilage* 2008;16:1433-41.
22. Roemer FW, Kwok CK, Hannon MJ, Hunter DJ, Eckstein F, Fujii T, Boudreau RM, Guermazi A. What comes first? Multitissue involvement leading to radiographic osteoarthritis: magnetic resonance imaging-based trajectory analysis over four years in the osteoarthritis initiative. *Arthritis Rheumatol* 2015;67:2085-96.
23. Hunter DJ, Guermazi A, Lo GH, Grainger AJ, Conaghan PG, Boudreau RM, Roemer FW. Evolution of semi-

- quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). *Osteoarthritis Cartilage* 2011;19:990-1002.
24. Bien N, Rajpurkar P, Ball RL, Irvin J, Park A, Jones E, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. *PLoS Med* 2018;15:e1002699.
  25. Štajduhar I, Mamula M, Miletić D, Ūnal G. Semi-automated detection of anterior cruciate ligament injury from MRI. *Comput Methods Programs Biomed* 2017;140:151-64.
  26. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. *Commun ACM* 2017;60:84-90.
  27. Hu J, Shen L, Sun G, editors. Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018:7132-41.
  28. Zhang L, Li M, Zhou Y, Lu G, Zhou Q. Deep Learning Approach for Anterior Cruciate Ligament Lesion Detection: Evaluation of Diagnostic Performance Using Arthroscopy as the Reference Standard. *J Magn Reson Imaging* 2020;52:1745-52.
  29. Liu F, Guan B, Zhou Z, Samsonov A, Rosas H, Lian K, Sharma R, Kanarek A, Kim J, Guermazi A, Kijowski R. Fully Automated Diagnosis of Anterior Cruciate Ligament Tears on Knee MR Images by Using Deep Learning. *Radiol Artif Intell* 2019;1:180091.
  30. Germann C, Marbach G, Civardi F, Fucentese SF, Fritz J, Sutter R, Pfirrmann CWA, Fritz B. Deep Convolutional Neural Network-Based Diagnosis of Anterior Cruciate Ligament Tears: Performance Comparison of Homogenous Versus Heterogeneous Knee MRI Cohorts With Different Pulse Sequence Protocols and 1.5-T and 3-T Magnetic Field Strengths. *Invest Radiol* 2020;55:499-506.
  31. Liu C, Hu SC, Wang C, Lafata K, Yin FF. Automatic detection of pulmonary nodules on CT images with YOLOv3: development and evaluation using simulated and patient data. *Quant Imaging Med Surg* 2020;10:1917-29.
  32. Chang PD, Wong TT, Rasiej MJ. Deep Learning for Detection of Complete Anterior Cruciate Ligament Tear. *J Digit Imaging* 2019;32:980-6.
  33. Jeon Y, Yoshino K, Hagiwara S, Watanabe A, Quek ST, Yoshioka H, Feng M. Interpretable and Lightweight 3-D Deep Learning Model for Automated ACL Diagnosis. *IEEE J Biomed Health Inform* 2021;25:2388-97.
  34. Awan MJ, Rahim MSM, Salim N, Mohammed MA, Garcia-Zapirain B, Abdulkareem KH. Efficient Detection of Knee Anterior Cruciate Ligament from Magnetic Resonance Imaging Using Deep Learning Approach. *Diagnostics (Basel)* 2021;11:105.
  35. Awan MJ, Rahim MSM, Salim N, Rehman A, Nobanee H, Shabir H. Improved Deep Convolutional Neural Network to Classify Osteoarthritis from Anterior Cruciate Ligament Tear Using Magnetic Resonance Imaging. *J Pers Med* 2021;11:1163.
  36. Tran A, Lassalle L, Zille P, Guillin R, Pluot E, Adam C, Charachon M, Brat H, Wallaert M, d'Assignies G, Rizk B. Deep learning to detect anterior cruciate ligament tear on knee MRI: multi-continental external validation. *Eur Radiol* 2022;32:8394-403.

**Cite this article as:** Wang M, Yu C, Li M, Zhang X, Jiang K, Zhang Z, Zhang X. One-stop detection of anterior cruciate ligament injuries on magnetic resonance imaging using deep learning with multicenter validation. *Quant Imaging Med Surg* 2024. doi: 10.21037/qims-23-1539

## Appendix 1

Details of KneeMRI dataset (Croatia) and MRNet dataset (USA)

KneeMRI dataset, Croatia

The KneeMRI dataset was obtained from the Clinical Hospital Centre Rijeka, Croatia and was acquired using a 1.5T Siemens Avanto MRI scanner. It contains 917 sagittal PD-weighted examinations from 2007 until 2014.

MRNet dataset, USA

The MRNet dataset comprises 1370 examinations between 2001 and 2012 by a research team at Stanford University Medical Center. MRNet dataset was divided into 3 sets: a training set with 1,130 patient samples, a validation set with 120 patient samples, and a testing set with 120 patient samples. The training and validation sets are publicly available, but the testing set is not. It also includes other knee conditions, such as normal knees, meniscal tears, and ACL tears, but we restricted our analysis to cases with intact and ACL tears. Each examination contained the following sequences: coronal T1-weighted, coronal T2 with fat saturation, sagittal proton density (PD)-weighted, sagittal T2 with fat saturation, and axial PD-weighted with fat saturation. Examinations were performed using GE scanners, 775 (56%) with a 3T magnetic field, the remainder with a 1.5T magnetic field.

## Appendix 2

MRI imaging protocol and specific details of the all datasets in this study

MRNet dataset, USA

Imaging parameters for the 3.0T sagittal T2-FS sequence included: field-of-view (FOV) = 15cm, echo time (TE) = 54 ms, relaxation time (TR) = 5599 ms, slice thickness = 2.5 mm, gap = 0 mm, flip angle = 142°, number of sections=42, acquisition matrix size=384×192.

Imaging parameters for the 1.5T sagittal T2-FS sequence included: field-of-view (FOV) = 16 cm, echo time (TE) = 54 ms, relaxation time (TR) = 3266 ms, slice thickness = 3.5 mm, gap = 0.5 mm, flip angle = 90°, number of sections=24, acquisition matrix size= 448×192.

Chinese dataset

All image acquisition is done on a 1.5T Achieva (221 participants) or 3T Ingenia (115 participants). Imaging parameters for the 3.0T sagittal PDW-SPAIR sequence included the following: field-of-view (FOV) = 18 cm, echo time (TE) = 30 ms, relaxation time (TR) = 2844 ms, slice thickness = 3.5 mm, and gap = 0.35 mm.

Imaging parameters for the 1.5T sagittal PDW-SPAIR sequence include: FOV = 16 cm, TE = 30 ms, TR = 3000 ms, slice thickness = 3.5 mm, and gap = 0.35 mm.