# POLISH
# JOURNAL *of* RADIOLOGY

Original paper

# Feasibility of the fat-suppression image-subtraction method using deep learning for abnormality detection on knee MRI

Shusuke Kasuya[1,A,B,E,F], Tsutomu Inaoka[1,A,B,C,D,G], Akihiko Wada[2,A,B,C,D], Tomoya Nakatsuka[1,C,D,E], Koichi Nakagawa[3,E], Hitoshi Terada[1,E]

[1]Department of Radiology, Toho University Sakura Medical Center, Sakura, Japan
[2]Department of Radiology, Juntendo University, Tokyo, Japan
[3]Department of Orthopaedic Surgery, Toho University Sakura Medical Center, Sakura, Japan

## Abstract

**Purpose**: To evaluate the feasibility of using a deep learning (DL) model to generate fat-suppression images and detect abnormalities on knee magnetic resonance imaging (MRI) through the fat-suppression image-subtraction method.

**Material and methods**: A total of 45 knee MRI studies in patients with knee disorders and 12 knee MRI studies in healthy volunteers were enrolled. The DL model was developed using 2-dimensional convolutional neural networks for generating fat-suppression images and subtracting generated fat-suppression images without any abnormal findings from those with normal/abnormal findings and detecting/classifying abnormalities on knee MRI. The image qualities of the generated fat-suppression images and subtraction-images were assessed. The accuracy, average precision, average recall, F-measure, sensitivity, and area under the receiver operator characteristic curve (AUROC) of DL for each abnormality were calculated.

**Results**: A total of 2472 image datasets, each consisting of one slice of original T1WI, original intermediate-weighted images, generated fat-suppression (FS)-intermediate-weighted images without any abnormal findings, generated FS-intermediate-weighted images with normal/abnormal findings, and subtraction images between the generated FS-intermediate-weighted images at the same cross-section, were created. The generated fat-suppression images were of adequate image quality. Of the 2472 subtraction-images, 2203 (89.1%) were judged to be of adequate image quality. The accuracies for overall abnormalities, anterior cruciate ligament, bone marrow, cartilage, meniscus, and others were 89.5-95.1%. The average precision, average recall, and F-measure were 73.4-90.6%, 77.5-89.4%, and 78.4-89.4%, respectively. The sensitivity was 57.4-90.5%. The AUROCs were 0.910-0.979.

**Conclusions**: The DL model was able to generate fat-suppression images of sufficient quality to detect abnormalities on knee MRI through the fat-suppression image-subtraction method.

**Key words**: deep learning, 2-dimensional convolutional neural network, fat suppression, subtraction image, knee, MRI.

## Introduction

The development of deep learning (DL), an emerging field of artificial intelligence (AI), has greatly facilitated clinical decision support for interpreting medical images such as echocardiograms, chest radiographs, and magnetic resonance images (MRI) [1,2]. The use of DL algorithms to diagnose internal joint derangement through MRI analysis presents numerous possibilities. Investigational DL algorithms for internal joint derangement have been developed to detect tears in the anterior cruciate ligament (ACL), meniscus, and articular cartilage in the knee, rotator cuff tears

**Correspondence address:**
Tsutomu Inaoka, Department of Radiology, Toho University Sakura Medical Center, Sakura, Japan, e-mail: tsutomu.inaoka@med.toho-u.ac.jp

**Authors' contribution:**
**A** Study design · **B** Data collection · **C** Statistical analysis · **D** Data interpretation · **E** Manuscript preparation · **F** Literature search · **G** Funds collection

in the shoulder, Achilles tendon tears in the ankle, as well as to identify nerves, bones, and muscles [1,2]. Previous DL-based knee MRI studies have mostly focused on a single area such as the ACL, meniscus, or articular cartilage [3-14]. Recent studies have used 2-dimensional (2D) and 3-dimensional (3D) convolutional neural network (CNN) DL models and MRI data to developed automated algorithms that detect and grade abnormalities of multiple joint tissues [15,16]. Despite these advances, inputting image datasets remains a challenge for complex DL algorithms.

Fat suppression technique is crucial for visualizing oedema, cartilage structures, and bone marrow lesions. Sagittal fat-suppressed proton-density or T2-weighted images have shown higher detection rates for abnormalities in the ligaments, bone marrow, articular cartilage, meniscus, and soft tissues [17,18]. In this study, we hypothesized that abnormal findings could be more easily detected by subtracting knee fat-suppression images without abnormal findings from those with abnormal findings using DL. To test our hypothesis, we developed a DL model with 2D CNNs to accurately generate fat-suppression images from original non-fat-suppression images acquired with 2 different sequences, T1-weighted imaging (T1WI), and intermediate-weighted imaging, as well as to generate fat-suppression images in which abnormal findings were completely removed (referred to as knee normal fat-suppression images). We then used the generated images to detect and classify abnormalities using the fat-suppression image-subtraction method, which involved subtracting abnormal minus normal images. The purpose of the study is to develop a DL model that can accurately generate fat-suppression images and easily detect and classify abnormalities on knee MRI.

## Material and methods

This study was conducted with the approval of the Ethics Committee in our institution (S20064). The use of clinical data for this research was disclosed on the institutional website, and the potential participants were given the opportunity to decline to be further enrolled in the study.

### MRI acquisition

All images were obtained in our institution using a 3 T MR scanner (Magnetom Skyra, Siemens Healthcare, Erlangen, Germany) with an 8-channel knee coil. All studies consisted of 2D-FSE T1-weighted (T1WI) and intermediate-weighted images, with and without fat suppression in the sagittal plane. The parameters are shown in Table 1. All the images were extracted in Digital Imaging and Communications in Medicine (DICOM) file format, converted to 8-bit greyscale Portable Network Graphics (PNG) format, and resized to $128 \times 128$ pixels.

### Sample selection

Forty-five knee studies in 45 consecutive symptomatic patients (mean age 54.6 ± 20.3 years; 16 males/29 females; 21 right/24 left) performed at 3 T in our institution between April 2020 and July 2020 were included. Cases after ligament reconstruction were excluded. The final diagnoses were osteoarthritis ($n = 18$), meniscal tear ($n = 38$), ligament tear ($n = 5$), post-resection of benign tumour ($n = 2$), Osgood-Schlatter disease ($n = 1$), and muscle injury ($n = 1$). In addition, 12 knee MR studies in 6 healthy volunteers who had neither symptoms nor history of trauma in the knee (mean age 34.2 ± 9.5 years; 4 males/2 females; 6 right/6 left) were included.

### Deep learning (DL) model

The DL model uses 2D CNNs on the open-source Neural Network Console ver.2.1 deep learning library (A. Hayakawa *et al.*, unpublished data, 2021), which was commercially developed (Sony Network Communications, Tokyo, Japan, https://dl.sony.com) and was based on the Python programming language (version 3.6.3; Python Software Foundation, Wilmington, DE, USA), running on a computer (eX.computer, Windows 10 operating system) with an AMD Ryzen 9 3950X 3.5 GHz processor, 64 GB RAM, and an NVIDIA GeForce RTX 2080Ti 11 GB graphics processing unit (NVIDIA, Santa Clara CA, USA).

Our DL algorithm consisted of two consecutive processes: generation of fat-suppression images and detection and classification of abnormalities (Figure 1).

### Generation of fat-suppression images

In this process, we developed 2 types of DL models using 2D CNNs: one for generating fat-suppression images without

Table 1. Imaging parameters of knee MRI at 3 Tesla

| Sequence | Plane | TR/TE (ms) | FOV (mm) | Slice thickness (mm) | Spacing (%) | Matrix |
| --- | --- | --- | --- | --- | --- | --- |
| T1WI | Sagittal | 520/10 | 160-160 | 3 | 10 | $448 \times 291$ |
| Intermediate-weighted image | Sagittal | 4500/31 | 160-160 | 3 | 10 | $448 \times 291$ |
| FS-intermediate-weighted image | Sagittal | 4500/31 | 160-160 | 3 | 10 | $448 \times 291$ |

T1WI – T1-weighted image, FS – fat suppression, TR – repetition time, TE – echo time, FOV – field of view
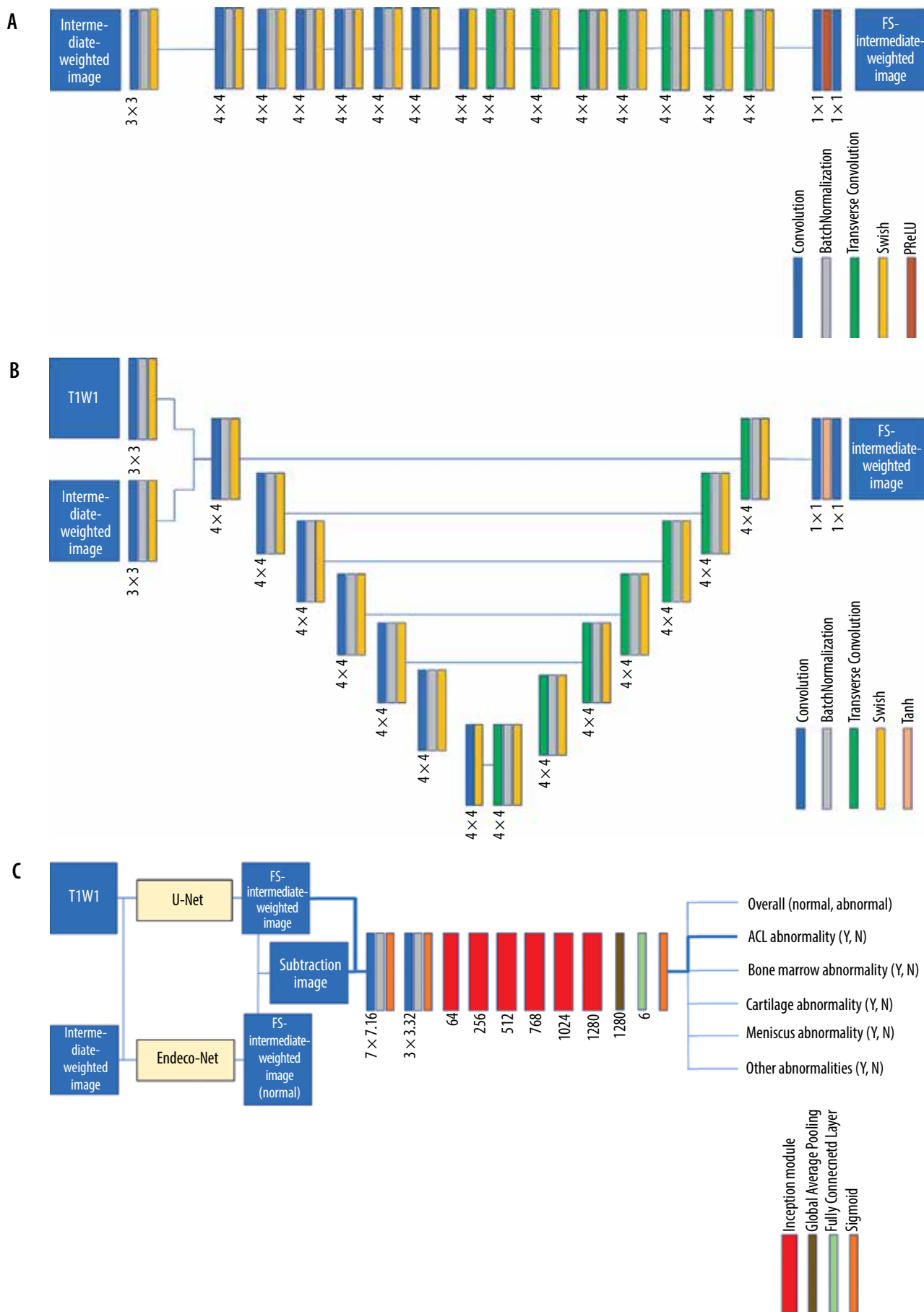
**Figure 1.** Structure of deep learning model. **A**) Encoder-decoder network (Endeco-Net) for normal fat-suppression images. **B**) U-Net for fat-suppression images incorporating both normal and abnormal findings. **C**) Convolutional neural network for subtracting the generated normal fat-suppression images (Figure 1A) from the generated fat-suppression images with abnormal findings (Figure 1B) and detecting and classifying abnormal findings
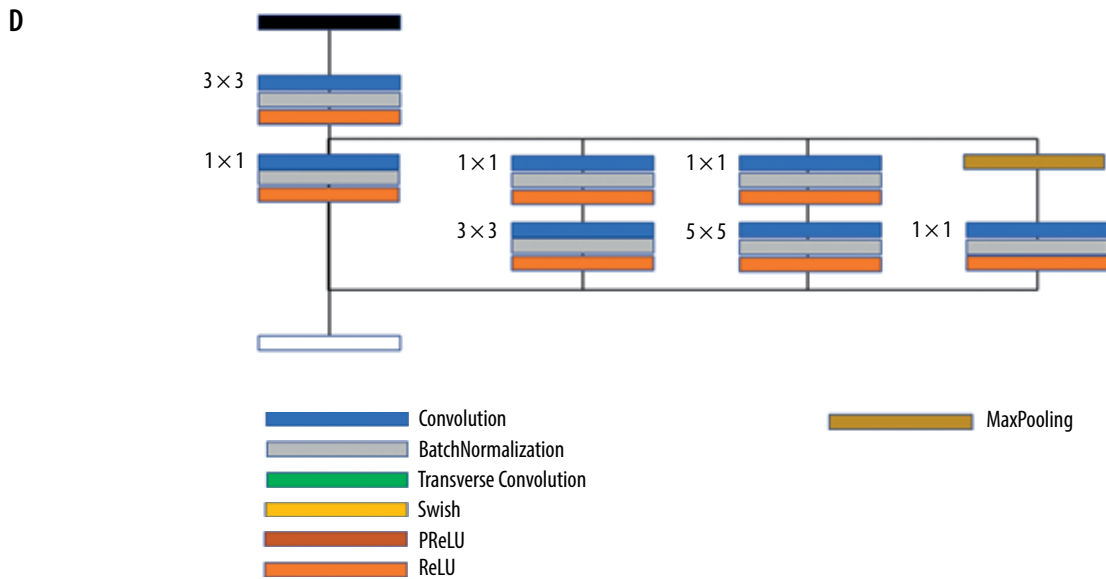
**D**



Convolution

BatchNormalization

Transverse Convolution

Swish

PReLU

ReLU

MaxPooling

**Figure 1.** Cont. Structure of deep learning model. **D**) Inception module in the convolutional neural network (Figure 1C)

any abnormal findings (containing only normal findings), and another for generating fat-suppression images with normal and/or abnormal findings. The first DL model was specifically designed to synthesize fat-suppression images without any abnormal findings (referred to as normal fat-suppression images) using an in-house convolutional encoder-decoder network (Endeco-Net) (Figure 1A). For this model, as a control group, we exclusively used normal MR images from 12 knee studies involving healthy volunteers. Consequently, the Endeco-Net was trained solely on normal findings and was capable of generating only normal findings. Thus, we created 348 image datasets, each consisting of one slice of original T1WI, original intermediate-weighted images, original fat-suppressed (FS)-intermediate-weighted images, and generated FS-intermediate-weighted images at the same cross-section.

Next, we developed a DL model utilizing a U-Net network for faithfully synthesizing fat-suppression images incorporating both normal and/or abnormal findings (Figure 1B). By using the 45 abnormal knee MR studies from 45 patients, we created 1263 image datasets, each consisting of one slice of original T1WI, original intermediate-weighted images, original FS-intermediate-weighted images, and generated FS-intermediate-weighted images at the same cross-section.

### Detection and classification of abnormalities on knee MRI

In this process, apart from the previously mentioned Endeco-Net and U-Net, we developed a dedicated DL model for subtracting the generated normal fat-suppression images (FS-intermediate-weighted images) (fat-suppression images without any abnormal findings) from the generated fat-suppression images (FS-intermediate-weighted im-

ages) with abnormal findings in order to effectively detect abnormalities. Furthermore, we developed an additional DL model specifically designed for the detection and classification of these abnormalities (Figures 1C and D).

Based on the luminance of the subtraction images, a threshold of 128 out of 256 colour tones between free fluid and oedema was determined. Thus, the colour blue was displayed when a certain area was free fluid ($\geq$ 128), while the colour red was displayed when it was judged to be oedema ($\leq$ 128). The value of 128 was employed to effectively exclude free fluid. Prior to setting the value of 128, a radiologist assessed whether it was possible to distinguish between free fluid and oedema in several cases. Therefore, abnormalities of the ACL, bone marrow, articular cartilage, and menisci are depicted in the colour red. Attenuation map images were generated by superimposing the subtraction images on acquired intermediate-weighted images.

We augmented the image data by randomly zooming in and out, rotating within a range of –0.15 and +0.15 radians, and flipping it left and right [19]. We primarily focused on fine structures in the knee joints, so we applied such data augmentation.

From the 1263 image datasets, each consisting of one slice of original T1WI, original intermediate-weighted images, and original FS-intermediate-weighted images, a total of 2472 image datasets, each consisting of one slice of original T1WI, original intermediate-weighted images, FS-intermediate-weighted images synthesized using Endeco-Net, FS-intermediate-weighted images synthesized using U-Net, and subtraction images between the FS-intermediate-weighted images synthesized using Endeco-Net and synthesized using U-Net at the same cross-section, were created.

### Image assessments

Two board-certified radiologists independently assessed the image quality of the 2472 FS-intermediate-weighted images synthesized by Endeco-Net, 2472 FS-intermediate-weighted images synthesized by U-Net, and 2472 subtraction images between these fat-suppression images. One of the radiologists conducted a second evaluation after a one-month interval. The image assessments were performed on a liquid crystal display monitor (diagonal 80 cm [31.5"], screen ratio 16:9; resolution 2560 × 1440, 3.7 megapixels). Regarding inter-reader agreement by 2 radiologists and intra-reader agreement by one radiologist, kappa values were calculated. The strength of agreement quantified by the kappa statistic was graded as follows: < 0, poor; 0.01-0.20, slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; and 0.81–1, almost perfect [20].

In regard to the labelling of image sets, we labelled each image set. The presence or absence of overall abnormalities on the image sets, which consisted of one slice each of the original sagittal T1WI, original sagittal intermediate-weighted images, and original sagittal FS-intermediate-weighted images at the same cross-sections, was determined by the radiologist. In addition, the presence or absence of abnormalities involving the ACL, bone marrow, articular cartilage, and menisci, joint effusion with capsular distention, soft-tissue oedema, and other fluid collections was also assessed. Thus, determinations of abnormal findings were made on a per-slice basis. In cases where equivocal findings were encountered, the adjacent 1- or 2-slice images were reviewed to reach a definitive conclusion. Finally, a cross-check was conducted with clinical diagnoses based on the medical information.

"ACL abnormality" was defined as a complete tear, partial tear, or mucoid degeneration of the ACL. A partial tear of the ACL was defined as any fibre discontinuity of up to 80%. Mucoid degeneration of the ACL was a thickened, ill-defined ACL, with increased signal intensity on all MR sequences but without fibre discontinuity [8]. Because these categories are difficult to distinguish even on the radiologist's readings of MRI, we did not differentiate among them. "Bone marrow abnormality" was defined as any bone marrow oedema pattern. "Cartilage abnormality" was defined as articular cartilage irregularity, focal defect, or diffuse thinning due to cartilage degeneration at the medial and lateral femorotibial joints and patellofemoral joints [6]. "Meniscus abnormality" was defined as a vertical tear, horizontal tear, complex tear, irregular-shaped meniscus, or disappeared or displaced meniscus.

### Statistical analysis

We divided the image datasets, including the FS-intermediate-weighted images synthesized by U-Net, and the subtraction images, into 3 groups for training, validation, and testing. To evaluate our DL model, 5 metrics of predictive power (accuracy, average precision, average recall, F-measure, and sensitivity) were calculated on a Neural Network Console ver. 2.1 deep learning library (Sony). In addition, area under the receiver operator characteristic curve (AUROC) values were also calculated with commercial software (SPSS for Windows ver. 28.0, IBM, Armonk, NY, USA).

## Results

### Generation of fat-suppression images

Out of the 2472 image datasets (each consisting of one slice of original T1WI, original intermediate-weighted images, FS-intermediate-weighted images synthesized using Endeco-Net, FS-intermediate-weighted images synthesized using U-Net, and subtraction images between the FS-intermediate-weighted images synthesized using Endeco-Net and synthesized using U-Net at the same cross-section), 2203 (89.1%) were deemed to have adequate image quality, while 269 (10.9%) were determined to have inadequate image quality. In terms of the breakdown of judging of inadequate image quality, the fat suppression images synthesized through the Endeco-Net and U-Net were considered satisfactory. However, 10.9% of the subtraction images synthesized by Endeco-Net and U-Net were judged to have inadequate image quality. One radiologist determined that the 10.9% of the subtraction images did not affect the diagnosis of internal derangements in the knee joint.

The inter-reader agreement was substantial (κ value: 0.77), and the intra-reader agreement was almost perfect (κ value: 0.88).

### Detection and classification of abnormalities on knee MRI

The image datasets for training, validation, and testing were 1799 (81.7%), 99 (4.5%), and 305 (13.8%), respectively. A summary of the training, validation, and testing image datasets is shown in Table 2.

Of the 2203 image datasets, 976 (44.3%) were interpreted by the radiologist as "normal" and 1227 (55.7%) were determined to be "abnormal". Of the "abnormal" images, an ACL abnormality was detected in 11.9%, a bone marrow abnormality in 27.3%, a cartilage abnormality in 55.0%, a meniscus abnormality in 64.9%, and other diagnoses in 14.4%.

Accuracy, average precision, average recall, F-measure, and sensitivity of our DL model for determining whether presence or absence of overall abnormalities on knee MRI were 89.5%, 89.4%, 89.4%, 89.4%, and 90.5%, respectively. The AUROC (95% confidence interval [CI]) was 0.931 (0.899–0.963). Accuracies, average precisions, average recalls, F-measures, sensitivities, and AUROCs to detect each abnormality are shown in Table 3.

Representative cases are shown in Figures 2 and 3.

**Table 2.** A summary of datasets for training, validation, and testing

| | Total | Training | Validation | Testing |
|---|---|---|---|---|
| Number of image sets (%) | 2203 | 1799 (81.7) | 99 (4.5) | 305 (13.8) |
| Number of normal (%) | 976 (44.3) | 788 (43.8) | 51 (51.5) | 137 (44.9) |
| Number of abnormal (%) | 1227(55.7) | 1011(56.2) | 48 (48.5) | 168 (55.1) |
| Number of ACL abnormalities (%) | 146 (11.9) | 115 (6.4) | 3 (3.0) | 28 (9.2) |
| Number of bone marrow abnormalities (%) | 335 (27.3) | 262 (14.6) | 12 (12.1) | 61 (20.0) |
| Number of cartilage abnormalities (%) | 675 (55.0) | 542 (30.1) | 24 (24.2) | 169 (35.7) |
| Number of meniscus abnormalities (%) | 796 (64.9) | 653 (36.3) | 32 (32.3) | 111 (36.4) |
| Number of other abnormalities (%) | 177 (14.4) | 157 (8.7) | 6 (6.1) | 14 (4.6) |

ACL – anterior cruciate ligament

**Table 3.** Results of metrics to detect abnormalities in knee MR imaging by DL model with 2D CNNs

| Abnormality | Accuracy | Average precision | Average recall | F-measure | Sensitivity | AUROC (95% CI) |
|---|---|---|---|---|---|---|
| ACL | 95.1% | 84.9% | 86.1% | 85.5% | 75% | 0.979 (0.965-0.993) |
| Bone marrow | 89.5% | 87.8% | 77.5% | 81.2% | 57.4% | 0.910 (0.863-0.956) |
| Cartilage | 89.8% | 90.6% | 87.2% | 88.5% | 80% | 0.947 (0.920-0.974) |
| Meniscus | 89.5% | 89.8% | 87.3% | 88.3% | 79.4% | 0.943 (0.916-0.970) |
| Others | 95.1% | 73.4% | 87.2% | 78.4% | 78.6% | 0.921 (0.819-1.000) |
| Overall* | 89.5% | 89.4% | 89.4% | 89.4% | 90.5% | 0.931 (0.899-0.963) |

DL – deep learning, 2D CNN – 2-dimensional convolutional neural network, AUROC – area under receiver operating characteristic curve, 95% CI – 95% confidence interval, ACL – anterior cruciate ligament. Others include joint effusion with capsular distention, soft-tissue oedema, and soft-tissue fluid collection. *Overall means presence or absence of any abnormalities on knee MR images



**Figure 2.** A 53-year-old male. (**A**) Original sagittal T1-weighted image, (**B**) original sagittal intermediate-weighted image, (**C**) original sagittal fat-suppressed intermediate-weighted image, (**D**) fat-suppressed intermediate-weighted image synthesized by U-Net, (**E**) fat-suppressed intermediate-weighted image synthesized by Endeco-Net (normal fat-suppression image), (**F**) subtraction image between fat-suppression images synthesized by U-Net and synthesized by Endeco-Net, (**G**) attenuation map image. (**D, E**) Fat-suppression intermediate-weighted images synthesized by U-Net and Endeco-Net both are of adequate image quality. (**D, F, G**) Joint effusion (arrow), bone marrow edema pattern and cartilage loss (thin arrow), and meniscus tear (curved arrow) are shown on the fat-suppressed intermediate-weighted image synthesized by U-Net, subtraction image, and attenuation map image. Joint effusion is represented in blue color, while bone marrow edema pattern and cartilage loss and meniscus tear are shown in red color on the attenuation map. Our deep learning model detected bone marrow abnormality, cartilage abnormality, and meniscus abnormality

## Discussion

We developed a DL model with 2D CNNs for the synthesis of fat-suppression images from 2 different non-fat-suppressed 2D-FSE sequences (T1WI and intermediate-weighted-image) and presented a fat-suppression subtraction-image method using 2D CNN DL algorithms for the detection and classification of abnormal findings
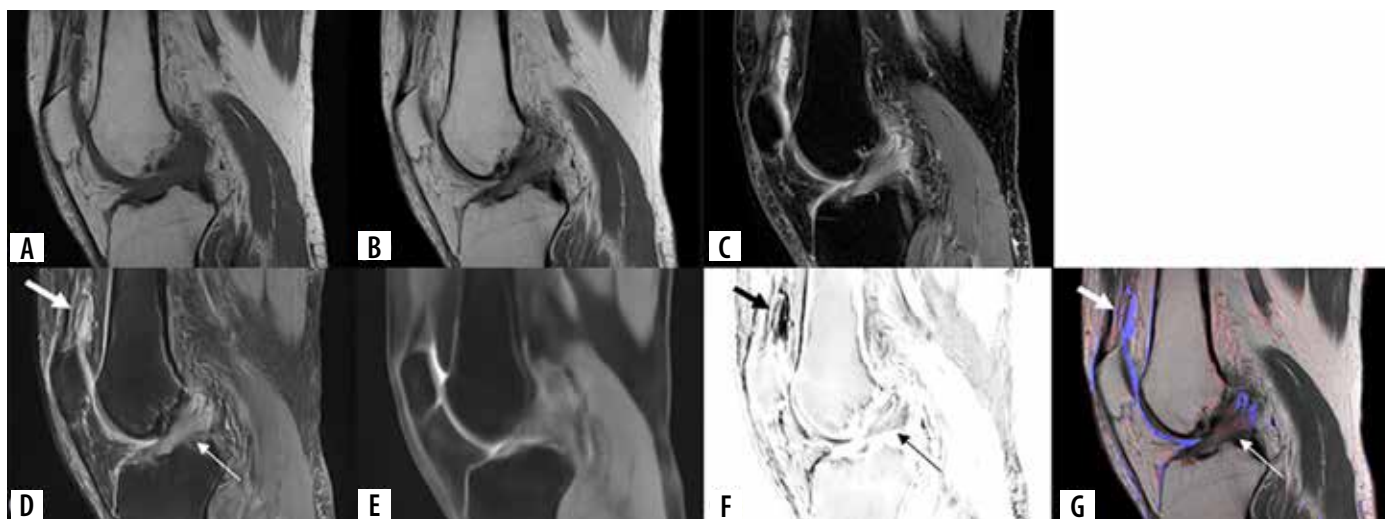
**Figure 3.** A 63-year-old male. (**A**) Original sagittal T1-weighted image, (**B**) original sagittal intermediate-weighted image, (**C**) original sagittal fat-suppressed intermediate-weighted image, (**D**) fat-suppressed intermediate-weighted image synthesized by U-Net, (**E**) fat-suppressed intermediate-weighted image synthesized by Endeco-Net (normal fat-suppression image), (**F**) subtraction image between fat-suppression images synthesized by U-Net and by Endeco-Net, (**G**) attenuation map image. (**D, E**) Fat-suppression intermediate-weighted images synthesized by U-Net and Endeco-Net both are of adequate quality. (**D, F, G**) Joint effusion (arrow) and anterior cruciate ligament abnormality (thin arrow) are shown on the fat-suppressed intermediate-weighted image synthesized by U-Net, subtraction image, and attenuation map image. Joint effusion is represented in blue color. A red color overlay is observed on the ACL, which indicate an ACL abnormality on the attenuation map. Our deep learning model detected anterior cruciate ligament abnormality

on knee MRI. The accuracy, average precision, average recall, F-measure, and sensitivity of our DL model for determining the presence or absence of overall abnormalities on knee MRI were 89.5%, 89.4%, 89.4%, 89.4%, and 90.5%, respectively. The AUROC was 0.931. For the specific abnormalities, the sensitivity was 57.4% for bone marrow oedema, while for ACL, cartilage, meniscus, and others it ranged from 75.0% to 80.0%. The accuracy was between 89.5% and 95.1% for all abnormalities. The AUROC was between 0.910 and 0.979. These results indicate that our DL model with 2D CNNs can generate fat-suppression images of sufficient quality from 2 different non-fat-suppressed 2D-FSE sequences and detect and classify abnormalities on knee MRI through the fat-suppression image-subtraction method.

Fayad *et al.* [21], who developed a DL model utilizing 2D CNNs for generating FS-intermediate-weighted-images from non-FS-intermediate-weighted-images with 3D-FSE, demonstrated the feasibility of DL-based synthesis of high-quality fat suppression images. In our study, 10.9% of the image datasets were determined to have inadequate image quality. Because there are no previous studies for comparison, it is challenging to determine whether this percentage is high or low. This 10.9% specifically pertained to issues observed in the subtraction images between the fat-suppression images synthesized through Endeco-Net and U-Net. Indeed, the fat-suppression images synthesized by Endeco-Net and U-Net were found to be satisfactory. The reasons for inadequate image quality were identified as blurring and misregistration at the anatomical edges, particularly in the medial and lateral aspects of the knee, observed in the subtraction images. This may be due to the

process of the synthesis of fat-suppression images without any abnormality. For the removal of any abnormal findings, we used an encoder-decoder to perform data reduction. In addition, only one sequence series of an original 2D-FSE intermediate-weighted imaging dataset was used as an input at the training. Initially, we considered including 2 sequence series as an input for Endeco-Net; however, we found that a lower data volume of input was more effective in removing abnormal findings and in generating fat-suppression images without any abnormal findings (referred to as normal fat-suppression images). This aspect represents one of the key principles of our DL models. While, for the synthesis of fat-suppression images with normal and/or abnormal findings, a U-Net network was used. In addition, original 2D-FSE T1WI and intermediate-weighted images were used as an input. This decision was made based on the belief that at least 2 different imaging sequences were necessary to accurately synthesize fat-suppression images with abnormalities. In any case, the sagittal plane images of the medial and lateral aspects of the knee joint, comprising only skin and subcutaneous fat, generally have minimal impact on the diagnosis of knee joint derangements. One radiologist determined that the 10.9% of subtraction images did not affect the diagnosis of internal derangements in the knee joint.

Previously, Bien *et al.* [15] developed a DL model using 2D CNNs for detecting abnormalities in the knee (MRNet) using a dataset of 1370 knee MRIs, including all 3 orthogonal planes. The reference standards were a majority vote of 3 musculoskeletal radiologists. Their DL model achieved accuracy, sensitivity, and AUROC of 85%, 87.9%, and 0.937 for detecting overall abnormal-

ity on knee MRI, respectively. In the following 2 studies, there are ambiguous aspects in the definition of general/overall abnormality. Irmakci *et al.* [22], who attempted to evaluate several DL models for abnormality detection in the knee, reported varying accuracy ranging from 82.5% to 85.8%, sensitivity ranging from 96.8% to 97.9%, and AUROC ranging from 0.811 to 0.909. Tsai *et al.* [23] reported accuracy of 91.7%, sensitivity of 96.8%, and AUROC of 0.941. In our study, the accuracy, sensitivity, and AUROC of our DL algorithms for the detection of overall abnormalities were 89.5%, 90.5%, and 0.931, respectively. The AUROC in our series was nearly on par with their results, while the accuracy and sensitivity were superior to those of Bien's study.

For the specific abnormalities, our results showed high accuracy (95.1%) and AUROC (0.979) but a low sensitivity of 75% for ACL tears. Bien *et al.* [15] reported a sensitivity of 76% for ACL tears, although the specificity was 97%, while Chang *et al.* [6] demonstrated a sensitivity of 100% and an accuracy of > 96%. Liu *et al.* [24] used a 2D CNN DL model to investigate another ACL assessment approach for binary ACL-tear classification and reported 96% for both sensitivity and specificity and 0.980 for an AUROC. Germann *et al.* [7] reported a DL model using 3D CNNs for detecting ACL tears with a sensitivity of 96.1%, a specificity of 93.1%, and an AUROC of 0.935. Irmarci *et al.* [22] reported sensitivity of 77.8%, specificity of 93.9%, accuracy of 86.7%, and AUROC of 0.954. Zhang *et al.*, [11] using 3D CNN model reported sensitivity of 97.6%, specificity of 94.4%, accuracy of 95.7%, and AUROC of 0.960. Astuto *et al.* [16], using a 3D CNN DL model and 3D-FSE MRI datasets, reported sensitivity of 88% and specificity of 89% for the detection of ACL abnormalities. In our study, we designed a DL model using 2D CNNs. The accuracy for the detection of ACL abnormalities in our series was comparable to these previous studies. The sensitivity for the detection of ACL abnormalities was similar to that in Bien's study [15], but both sensitivities were lower than the other previous studies. Namiri *et al.* [3] previously reported that 2D and 3D CNN DL models performed similarly in classifying ACL abnormalities. They reported sensitivity of 76.4% and 82.4% and specificity of 93.7% and 99.6%, respectively. Recently, Shin *et al.* [25], who used a 2D CNN DL model and one oblique-sagittal image along the ACL on which the largest ACL area was observed, reported accuracy of 94.1% and AUROC of 0.941. Our series included only a small number of ACL abnormalities. Although it is well known that most ACL tears are visible on sagittal images, [17] the entire ACL was not shown on a single-slice sagittal image. Therefore, for the detection of ACL abnormalities, cropping the images to the ACL, multi-slice input, or multi-plane input might increase the sensitivity.

Regarding the cartilage abnormalities, our results showed that the accuracy, sensitivity, and AUROC were 89.8%, 80%, and 0.947, respectively. Liu *et al.* [5] reported the sensitivity ranging from 80.5% to 84.1%, speci-

ficity from 85.2% to 87.9%, and AUROC from 0.914 to 0.917 using 2D CNN and 2D FSE FS-T2WI in the sagittal plane. Astuto *et al.* [16], using 3D CNN and 3D FSE images, reported sensitivity of 85%, specificity of 89%, and AUROC of 0.930. Our results were quite similar to theirs. In meniscus abnormalities, a relatively large number of studies using DL have previously been reported [4,9, 12,15,16,24-27]. Multi-slice or multi-plane input might increase the sensitivity in the detection of meniscus tears. Liu *et al.* [5] also found that DL-based detection had substantial intraobserver agreement, but clinical radiologists had moderate to substantial inter-observer agreement. DL was less prone to errors due to inexperience, distraction, or fatigue but had a high false-positive rate and was not effective in evaluating images with disparate parameters.

To date, only a limited number of studies have employed DL for the detection of bone marrow abnormalities. Astuto *et al.* [16], using DL for the detection through 3D CNN and 3D FSE sequences, reported lower sensitivity (70%) for the detection of bone marrow oedema compared to other tissues. Fayed *et al.* [21], who developed a DL model utilizing 2D CNNs for generating FS-intermediate-weighted-images from non-FS-intermediate-weighted-images with 3D-FSE, reported sensitivity of 76% and specificity of 90% when human readers interpreted the generated FS-intermediate-weighted-images. Kijowski *et al.* [28] reported the same trend in their evaluation for bone marrow oedema on 3D FSE FS-intermediate-weighted images by 2 musculoskeletal radiologists, with lower sensitivity (85.3%). In our study, although using DL detection based on 2D sequences, the accuracy and AUROC for the detection of bone marrow abnormalities were 89.5% and 0.910, respectively. However, the sensitivity was relatively low at 57.4%. This suggests that it is difficult to determine if bone marrow findings are abnormal or not due to significant variations in the size, location, and signal intensity of bone marrow abnormalities in the femur, tibia, and patella, even though our DL models were able to identify changes in bone marrow signals. We think that establishing a gold standard for bone marrow abnormality is also challenging. As a future step, to improve the diagnostic accuracy of bone marrow abnormality, it may be necessary to augment the number of cases and improve training data for the DL models. Additionally, incorporating the bone marrow finding with other findings could promote a comprehensive assessment.

A summary of DL studies for abnormal detection on knee MRI is presented in Table 4 [3-7,9,11-13,15,16,22-27,29].

There are several limitations to this study. Firstly, the number of image data included in this study, acquired from both patients and healthy volunteers, was small, and the knee MRI protocols and parameters were fixed. External cross-validation is necessary to confirm our preliminary observations [30]. Larger studies in an uncontrolled environment are also needed to assess the clinical usefulness of this method. Secondly, we did not use a surgical standard

**Table 4.** A summary of DL studies for abnormality detection on knee MRI

| Author | Year | Analysed pulse sequence | Field strengths [T] | DL model | Sensitivity | Specificity | Accuracy | AUROC | Comments |
|---|---|---|---|---|---|---|---|---|---|
| **Overall abnormality** | | | | | | | | | |
| Bien et al. [15] | 2018 | 2D*, sag FST2, cor T1, ax PD | 1.5, 3 | 2D CNN | 87.9% | 71.4% | 85% | 0.937 | |
| Irmakci et al. [22] | 2020 | 2D*, sag T2, cor T1, ax PD | 1.5, 3 | 2D CNN | 96.8-97.9% | 28-40% | 82.5-85.8% | 0.811-0.909 | The definition of general abnormality is unclear. |
| Tsai et al. [23] | 2020 | | 1.5, 3 | 2D CNN | 96.8% | 72% | 91.7% | 0.941 | Analysed pulse sequences are unclear. The definition of general abnormality is unclear. |
| Range | | | | | 87.9-97.9% | 28-72% | 82.5-91.7% | 0.811-0.941 | |
| The present study | | 2D*, sag T1, IWI | 3 | 2D CNN | 90.5% | | 89.5% | 0.931 | |
| **Anterior cruciate ligament abnormality** | | | | | | | | | |
| Bien et al. [15] | 2018 | 2D*, sag FST2, cor T1, ax PD | 1.5, 3 | 2D CNN | 75.9% | 96.8% | 86.7% | 0.965 | |
| Chang et al. [6] | 2019 | 2D*, cor PD | 1.5, 3 | 2D CNN | 100% | 93.3% | 96.7% | | Mucoid degeneration and partial tear are excluded. |
| Liu et al. [24] | 2019 | 2D*, sag FST2, sag PD | 3 | 2D CNN | 96% | 96% | | 0.980 | |
| Germann et al. [7] | 2020 | 2D*, cor STIR, sag FST2 | 1.5, 3 | 3D CNN | 96.1% | 93.1% | | 0.935 | |
| Irmakci et al. [22] | 2020 | 2D*, sag T2, cor T1, ax PD | 1.5, 3 | 2D CNN | 77.8% | 93.9% | 86.7% | 0.954 | |
| Zhang et al. [11] | 2020 | 2D*, sag FSPD | 1.5, 3 | 3D CNN | 97.6% | 94.4% | 95.7% | 0.960 | |
| Namiri et al. [3] | 2020 | 3D*, sag FSPD | 3 | 2D CNN | 82.4% | 93.7% | | | Full-thickness tears are assessed. |
| Namiri et al. [3] | 2020 | 3D*, sag FSPD | 3 | 3D CNN | 76.4% | 99.6% | | | Full-thickness tears are assessed. |
| Tsai et al. [23] | 2020 | | 1.5, 3 | 2D CNN | 92.3% | 89.1% | 90.4% | 0.960 | Analysed pulse sequences are unclear. |
| Awan et al. [13] | 2021 | 2D*, sag FSPD | 1.5 | 2D CNN | 91.7% | 94.7% | | 0.980 | |
| Astuto et al. [16] | 2021 | 3D*, sag FSPD | 3 | 3D CNN | 88% | 89% | | 0.900 | |
| Shin et al. [25] | 2022 | 2D*, sag FST2 | 1.5 | 2D CNN | | | 94.1% | 0.941 | Single oblique-sagittal images along the anterior cruciate ligament are assessed. |
| Tran et al. [29] | 2022 | 2D*, 3D*; sag, cor FSPD/ FST2 | 1, 1.5, 3 | 2D CNN | 87% | 91% | 90.2% | 0.941 | MRI studies from 12 imaging centres are included. |
| Range | | | | | 75.9-100% | 89-99.6% | 86.7-96.7% | 0.900-0.980 | |
| The present study | | 2D*, sag T1, IWI | 3 | 2D CNN | 75% | | 95.1% | 0.979 | |

of reference for correlation. Thirdly, in this study, we employed a value of 128 within a range of 256 colour tones to differentiate between free fluid and oedema, especially aiming to enhance the specificity of bone marrow abnormality detection. Before setting the value of 128, a radiologist deliberated if it was possible to distinguish between free fluid and oedema in several cases. Because there might be overlap between free fluid and oedema, reassessment of the threshold might be required. Fourthly, due to being based on the accuracy on a per-slice basis, there appears to be a tendency for the diagnostic accuracy to be relatively low. Moreover, in cases where the ACL, articular cartilage, and

**Table 4.** Cont.

| Author | Year | Analysed pulse sequence | Field strengths [T] | DL model | Sensitivity | Specificity | Accuracy | AUROC | Comments |
|---|---|---|---|---|---|---|---|---|---|
| **Bone marrow abnormality** | | | | | | | | | |
| Astuto *et al.* [16] | 2021 | 3D*, sag FSPD | 3 | 3D CNN | 70% | 88% | | 0.830 | |
| Range | | | | | 70% | 88% | | 0.830 | |
| The present study | | 2D*, sag T1, IWI | 3 | 2D CNN | 57.4% | | 89.5% | 0.910 | |
| **Cartilage abnormality**** | | | | | | | | | |
| Liu *et al.* [5] | 2018 | 2D*, sag FST2 | 3 | 2D CNN | 80.5-84.1% | 85.2-87.9% | | 0.914-0.917 | |
| Pedoia *et al.* [4] | 2019 | 3D*, sag FSPD | 3 | 3D CNN | 80% | 80.3% | | 0.880 | Patellar cartilages are assessed. |
| Astuto *et al.* [16] | 2021 | 3D*, sag FSPD | 3 | 3D CNN | 85% | 89% | | 0.930 | |
| Range | | | | | 80-85% | 80.3-89% | | 0.880-0.930 | |
| The present study | | 2D*, sag T1, IWI | 3 | 2D CNN | 80% | | 89.8% | 0.947 | |
| **Meniscus abnormality** | | | | | | | | | |
| Bien *et al.* [15] | 2018 | 2D*, sag FST2, cor T1, ax PD | 1.5, 3 | 2D CNN | 71% | 74.1% | 72.5% | 0.847 | |
| Pedoia *et al.* [4] | 2019 | 3D*, sag FSPD | 3 | 3D CNN | 82% | 89.8% | | 0.890 | |
| Fritz *et al.* [12] | 2020 | 2D*, sag FSIWI, cor STIR | 1.5, 3 | 3D CNN | 91.2% | 87.1% | 90% | 0.961 | |
| Irmakci *et al.* [22] | 2020 | 2D*, sag T2, cor T1, ax PD | 1.5, 3 | 2D CNN | 61.5-69.2% | 76.5-85.3% | 70.0-75.8% | 0.779-0.808 | |
| Tsai *et al.* [23] | 2020 | | 1.5, 3 | 2D CNN | 86% | 89% | 88% | 0.904 | Analysed pulse sequences are unclear. |
| Rizk *et al.* [9] | 2021 | Sag, cor FSPD | 1, 1.5, 3 | 3D CNN | 67-89% | 84-88% | 82-87% | 0.840-0.930 | Analysed pulse sequences are unclear. Medial and lateral meniscus are separately assessed. |
| Astuto *et al.* [16] | 2021 | 3D*, sag FSPD | 3 | 3D CNN | 85% | 85% | | 0.930 | |
| Li *et al.* [26] | 2022 | 2D*, sag FSPD | 3 | 3D CNN | 94.1% | 78.5% | 92.4% | 0.907 | |
| Shin *et al.* [27] | 2022 | 2D*. sag, cor FST2 | 1.5 | 2D CNN | 78.6% | 93.3% | 92.0% | 0.924 | |
| Range | | | | | 61.5-94.1% | 74.1-93.3% | 70-92.4% | 0.779-0.961 | |
| The present study | | 2D*, sag T1, IWI | 3 | 2D CNN | 79.4% | | 89.5% | 0.943 | |

DL – deep learning, AUROC – area under the receiver operator characteristic curve, 2D* – 2-dimensional pulse sequence, 3D* – 3-dimensional pulse sequence, sag – sagittal image, cor – coronal image, axi – axial image, FS – fat-suppressed, T2 – T2-weighted image, T1 – T1-weighted image, PD – proton density-weighted image, IWI – intermediated-weighted image, 2D – two-dimensional, 3D – three-dimensional, CNN – convolutional neural network
**From the category of cartilage abnormality, osteoarthritis studies are excluded.

meniscus are entirely absent within a slice, free fluid may be detected. However, typically, even if oedema is not detected at the site where these tissues are absent, it is common for findings of oedema to be present in the neighbouring tissues. Alternatively, in consecutive adjacent slices, a finding of oedema is detected, which is expected to be present in the surrounding tissues. Hence, 3-dimensional analysis may be considered necessary to improve the diagnostic accuracies. Finally, we did not evaluate the diagnostic performance of human readers when assisted by our DL model. Although our initial results are promising, further technical development and correlation with surgical findings as the gold standard will be required before this method can be implemented fully in clinical practice.

## Conclusions

We present a DL model that can generate fat suppression images of sufficient quality from 2 different non-fat-suppression images and detect and classify abnormalities on knee MRI. This method could be useful even in cases of poor-quality fat suppression images or when fat suppression images are unavailable. Furthermore, our results suggest that the use of a DL model with 2D CNNs for fat-suppression subtraction-image method could be a useful tool in detecting and classifying abnormalities on knee MRI.

## Acknowledgments

## Conflicts of interest

The authors declare that they have no conflicts of interest.

## References

1. Siouras A, Moustakidis S, Giannakidis A, et al. Knee injury detection using deep learning on MRI studies: a systematic review. Diagnostics (Basel) 2022; 12: 537. doi: 10.3390/diagnostics12020537.
2. Fritz B, Fritz J. Artificial intelligence for MRI diagnosis of joints: a scoping review of the current state-of-the art of deep learning-based approaches. Skeletal Radiol 2022; 51: 315-329.
3. Namiri NK, Flament I, Astuto B, et al. Deep learning for hierarchical severity staging of anterior cruciate ligament injuries from MRI. Radiol Artif Intell 2020; 2: e190207. doi: 10.1148/ryai.2020190207.
4. Pedoia V, Norman B, Mehany SN, et al. 3D convolutional neural networks for detection and severity staging of meniscus and PFJ cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects. J Magn Reson Imaging 2019; 49: 400-410.
5. 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: 160-169.
6. Chang PD, Wong TT, Rasiej MJ. Deep learning for detection of complete anterior cruciate ligament tear. J Digit Imaging 2019; 32: 980-986.
7. Germann C, Marbach G, Civardi F, et al. Deep convolutional neural network-based diagnosis of anterior cruciate ligament tears. Invest Radiol 2020; 55: 499-506.
8. Couteaux V, Si-Mohamed S, Nempont O, et al. Automatic knee meniscus tear detection and orientation classification with Mask-RCNN. Diag Interv Imaging 2019; 100: 235-242.
9. Rizk B, Brat H, Zille P, et al. Meniscal lesion detection and characterization in adult knee MRI: a deep learning model approach with external validation. Physica Medica 2021; 83: 64-71.
10. Jeon YS, Yoshino K, Hagiwara S, et al. Interpretable and lightweight 3-D deep learning model for automated ACL diagnosis. IEEE J Biomed Health Inform 2021; 25: 2388-2397.
11. Zhang L, Li M, Zhou Y, et al. 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: 1745-1752.
12. Fritz B, Marbach G, Civardi F, et al. Deep convolutional neural network-based detection of meniscus tears: comparison with radiologists and surgery as standard of reference. Skeletal Radiol 2020; 49: 1207-1217.

13. Awan MJ, Rahim MSM, Salim N, et al. Efficient detection of knee anterior cruciate ligament from magnetic resonance imaging using deep learning approach. Diagnostics 2021; 11: 105. doi: 10.3390/diagnostics11010105.
14. Tack A, Shestakov A, Ludke D, Zachow S. A multi-task deep learning method for detection of meniscal tears in MRI data from the osteoarthritis initiative database. Front Bioeng Biotechnol 2021; 9: 747217. doi: 10.3389/fbioe.2021.747217.
15. 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: e1002699. doi: 10.1371/journal.pmed.1002699.
16. Astuto B, Flament I, Namiri NK, et al. Automatic deep learning-assisted detection and grading of abnormalities in knee MRI studies. Radiol Artif Intell 2021; 3: e200165. doi: 10.1148/ryai.2021200165.
17. Schaefer FKW, Schaefer PJ, Brossmann J, et al. Value of fat-suppressed PD-weighted TSE-sequences for detection of anterior and posterior cruciate ligament lesions – comparison to arthroscopy. Eur J Radiol 2006; 58: 411-415.
18. Khoury NJ, Mahfoud Z, Masrouha KZ, et al. Value of sagittal fat-suppressed proton-density fast-spin-echo of the knee joint as a limited protocol in evaluating internal knee derangements. J Comput Assist Tomogr 2011; 35: 653-661.
19. Molnar D, Enqvist O, Ulen J, et al. Artificial intelligence based on automatic quantification of epicardial adipose tissue suitable for large scale population studies. Sci Rep 2021; 11: 23905. doi: 10.1038/s41598-021-03150-w.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-174.
21. Fayad LM, Parekh VS, Luna RDC, et al. A deep learning system for synthetic knee magnetic resonance imaging. Is artificial intelligence-based fat-suppressed imaging feasible? Invest Radiol 2021; 56: 357-368.
22. Irmakci I, Anwar SM, Torigian DA, Bagci U. Deep learning for musculoskeletal image analysis. arXiv:2003.00541v1; 2020. doi: https://doi.org/10.48550/arXiv.2003.00541
23. Tsai CH, Kiryati N, Konen E, et al. Knee injury detection using MRI with efficiently-layered network (ELNet). In: Tal A, Ismail Ben A, Marleen de B, Maxime D, Herve L, Christopher P (eds.). Proceedings of the Third Conference on Medical Imaging with Deep

Learning. Proceedings of Machine Learning Research: PMLR; 2020, p. 784-794.

24. Liu F, Guan B, Zhou Z, et al. Fully automated diagnosis of anterior cruciate ligament tears on knee MR images by using deep learning. Radiol Artif Intell 2019; 1: e180091. doi: 10.1148/ryai.2019180091.

25. Shin H, Choi GS, Chang MC. Development of convolutional neural network model for diagnosing teat of anterior cruciate ligament using only one knee magnetic resonance image. Medicine (Baltimore) 2022; 101: e31510. doi: 10.1097/MD.0000000000031510.

26. Li YZ, Wang Y, Fang KB, et al. Automated meniscus segmentation and detection of knee MRI with a 3D mask-RCNN. Eur J Med Res 2022; 27: 247. doi: 10.1186/s40001-022-00883-w.

27. Shin H, Choi GS, Shon OJ, et al. Development of convolutional neural network model for diagnosing meniscus tear using magnetic reso-

nance image. BMC Musculoskelet Disord 2022; 23: 510. doi: 10.1186/s12891-022-05468-6.

28. Kijowski R, Davis KW, Woods MA, et al. Knee joint: comprehensive assessment with 3D isotropic resolution fast spin-echo MR imaging – diagnostic performance compared with that of conventional MR imaging at 3.0T. Radiology 2009; 252: 486-495.

29. Tran A, Lassalle L, Zille P, at al. Deep learning to detect anterior cruciate ligament tear on knee MRI: multi-continental external validation. Eur Radiol 2022; 32: 8394-8403.

30. Si L, Zhong J, Huo J, et al. Deep learning in knee imaging: a systematic review utilizing a Checklist for Artificial Intelligence in Medical Imaging (CLAIM). Eur Radiol 2022; 32: 1353-1361.