



Prediction model for knee osteoarthritis using magnetic resonance–based radiomic features from the infrapatellar fat pad: data from the osteoarthritis initiative

Keyan Yu^{1,2#}, Jia Ying^{3#}, Tianyun Zhao³, Lan Lei^{4,5}, Lijie Zhong¹, Jiaping Hu¹, Juin W. Zhou³, Chuan Huang^{3,6,7}, Xiaodong Zhang¹

¹Department of Medical Imaging, The Third Affiliated Hospital of Southern Medical University (Academy of Orthopedics Guangdong Province), Guangzhou, China; ²Department of Radiology, Peking University Shenzhen Hospital, Shenzhen, China; ³Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY, USA; ⁴Program in Public Health, Stony Brook Medicine, Stony Brook, NY, USA; ⁵Department of Medicine, Northside Hospital Gwinnett, Lawrenceville, GA, USA; ⁶Department of Radiology, Stony Brook Medicine, Stony Brook, NY, USA; ⁷Department of Psychiatry, Stony Brook Medicine, Stony Brook, NY, USA

Contributions: (I) Conception and design: K Yu, J Ying, C Huang, X Zhang; (II) Administrative support: C Huang, X Zhang; (III) Provision of study materials or patients: K Yu, X Zhang, L Zhong, J Hu; (IV) Collection and assembly of data: K Yu, J Ying, X Zhang; (V) Data analysis and interpretation: K Yu, J Ying, C Huang, X Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work.

Correspondence to: Chuan Huang, PhD. Department of Radiology, Stony Brook Medicine, 101 Nicolls Rd, Stony Brook, NY 11794, USA. Email: chuan.huang@stonybrookmedicine.edu; Xiaodong Zhang, MD. Department of Medical Imaging, The Third Affiliated Hospital of Southern Medical University, 183 Zhongshan Ave W, Guangzhou, 510630, China. Email: ddautumn@126.com.

Background: The infrapatellar fat pad (IPFP) plays an important role in the incidence of knee osteoarthritis (OA). Magnetic resonance (MR) signal heterogeneity of the IPFP is related to pathologic changes. In this study, we aimed to investigate whether the IPFP radiomic features have predictive value for incident radiographic knee OA (iROA) 1 year prior to iROA diagnosis.

Methods: Data used in this work were obtained from the osteoarthritis initiative (OAI). In this study, iROA was defined as a knee with a baseline Kellgren-Lawrence grade (KLG) of 0 or 1 that further progressed to KLG ≥ 2 during the follow-up visit. Intermediate-weighted turbo spin-echo knee MR images at the time of iROA diagnosis and 1 year prior were obtained. Five clinical characteristics—age, sex, body mass index, knee injury history, and knee surgery history—were obtained. A total of 604 knees were selected and matched (302 cases and 302 controls). A U-Net segmentation model was independently trained to automatically segment the IPFP. The prediction models were established in the training set (60%). Three main models were generated using (I) clinical characteristics; (II) radiomic features; (III) combined (clinical plus radiomic) features. Model performance was evaluated in an independent testing set (remaining 40%) using the area under the curve (AUC). Two secondary models were also generated using Hoffa-synovitis scores and clinical characteristics.

Results: The comparison between the automated and manual segmentations of the IPFP achieved a Dice coefficient of 0.900 (95% CI: 0.891–0.908), which was comparable to that of experienced radiologists. The radiomic features model and the combined model yielded superior AUCs of 0.700 (95% CI: 0.630–0.763) and 0.702 (95% CI: 0.635–0.763), respectively. The DeLong test found no statistically significant difference between the receiver operating curves of the radiomic and combined models ($P=0.831$); however, both models outperformed the clinical model ($P=0.014$ and 0.004 , respectively).

Conclusions: Our results demonstrated that radiomic features of the IPFP are predictive of iROA 1 year prior to the diagnosis, suggesting that IPFP radiomic features can serve as an early quantitative prediction biomarker of iROA.

Keywords: Osteoarthritis (OA); infrapatellar fat pad (IPFP); radiomics; prediction model

Submitted Apr 16, 2022. Accepted for publication Oct 31, 2022. Published online Nov 17, 2022.

doi: 10.21037/qims-22-368

View this article at: <https://dx.doi.org/10.21037/qims-22-368>

Introduction

Knee osteoarthritis (OA) is a degenerative joint disease in which mechanical, biochemical, and biological processes can deteriorate articular cartilage. Knee OA initially manifests as the interaction between the cartilage matrix degradation products and proinflammatory mediators and the synovium tissue, which causes increasing proinflammatory reaction and further leads to cartilage degradation, bone remodeling, osteophyte formation, joint inflammation, and a loss of normal joint function (1). This disorder is one of the leading causes of chronic pain and disability (2). A validation study from Sweden predicted that clinically diagnosed knee OA in adults over 45 years old will increase from 13.8% to 15.7% by 2032 (3). Further, a 12.5% increase in disease-related claims for knee OA was observed among Medicare beneficiaries across the United States in 2014 compared to that in 2009 (4). Current trends in prevalence and healthcare expense escalation require medical and technical innovation that promotes knee OA prediction and prevention at an early stage.

The infrapatellar fat pad (IPFP) is an intraarticular adipose tissue located inferior to the patella. It actively interacts with periarticular tissues, such as cartilage, subchondral bone, and synovial membrane (5). Recent publications have shown that the IPFP changes the articular microenvironment and contributes to knee OA by secreting cytokines, adipokines, and lipid mediators (5,6). Currently, magnetic resonance (MR) imaging is the state-of-the-art modality used to evaluate IPFP abnormalities, such as edema, inflammation, synovial proliferation, and fibrosis (7,8). Research findings have revealed that alterations of MR signal intensity, shape, and volume in the IPFP are related to the incidence, development, and prognosis of knee OA (9-13). In current clinical practice, MR signal intensity changes in the IPFP are most commonly assessed semiquantitatively with the MR Imaging Osteoarthritis Knee Score (MOAKS) system using Hoffa-synovitis score, which grades the size of the diffuse hyperintense signal from 0 to 3 (0= normal/no hyperintensity, 1= mild, 2= moderate, and 3= severe) (14). By using radiography and MR imaging, the association between IPFP semiquantitative assessments (e.g.,

Hoffa-synovitis) and incident radiographic knee OA (iROA) has been established (15,16). However, this semiquantitative scoring system can suffer from inter-rater variability, and there is very limited evidence to prove that the IPFP can serve as an independent predictor for iROA.

Radiomics is a quantitative approach to medical image analysis that converts digital medical images into mineable, high-dimensional features, including signal intensity, geometric shape, and texture features. These features can provide a comprehensive description of the image and could subsequently be used in prediction model development to investigate disease diagnosis and prognosis (17-19). Using MR imaging, Hirvasniemi *et al.* (20) demonstrated that the radiomic features of the tibia bone can discriminate between knees with and without iROA. However, iROA development has yet to be predicted by a radiomics method. Moreover, accurate and effective segmentation is a practical challenge in model development. Recent studies have shown that deep learning methods, such as convolutional neural network (CNN), can automatically delineate regions of interest (ROIs) and achieve reliable and accurate segmentation of large amounts of data (21-24).

In this study, we aimed to investigate the predictive value of MR-based radiomic features of the IPFP in iROA with the aid of a fully automated segmentation method and to develop a prediction model using data obtained 1 year prior to clinical diagnosis. We would like to emphasize that the goal of this study was not to demonstrate a superior prediction performance of using IPFP compared to previously published models using other knee regions. Instead, the purpose is to determine whether IPFP radiomic features can be used as independent imaging markers for iROA, as this has not been well explored in the existing literature. We present the following article in accordance with the TRIPOD reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-368/rc>).

Methods

The osteoarthritis initiative (OAI)

The study was conducted in accordance with the

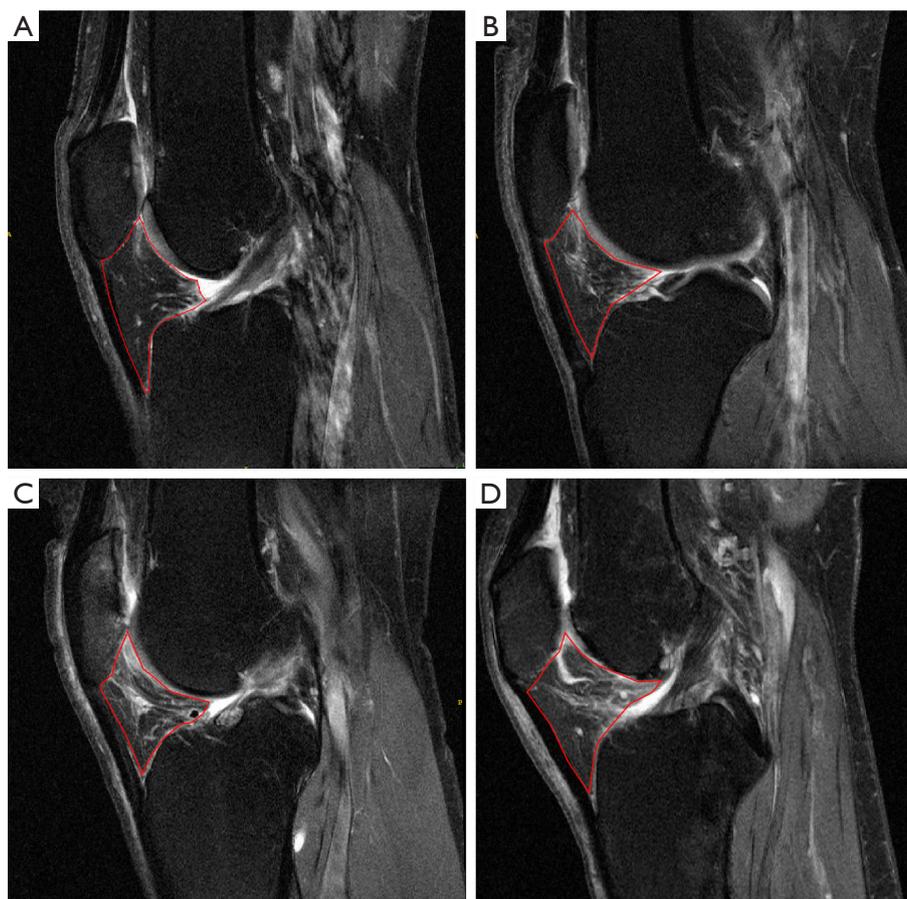


Figure 1 IPFP MR images obtained on sagittal intermediate-weighted turbo spin-echo from the iROA dataset. The red line delineates the IPFP for each case. The Hoffa-synovitis score was determined to be 0 = normal/no hyperintensity, 1= mild, 2= moderate, and 3= severe for (A) to (D), respectively. IPFP, infrapatellar fat pad; MR, magnetic resonance; iROA, incident radiographic knee osteoarthritis.

Declaration of Helsinki (as revised in 2013). All data used in this study were retrieved from the OAI (<https://nda.nih.gov/oai/>), which is an open-access public-domain research resource supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases. The OAI is a multicenter, longitudinal, prospective, and observational cohort study that aims to identify risk factors associated with the incidence and progression of knee OA. A total of 4,796 individuals aged 45–79 years were enrolled in this initiative, with detailed inclusion and exclusion criteria described on the designated website. This database contains (I) clinical characteristics; (II) biospecimen information; and (III) bilateral fixed-flexion radiographs and 3 T MR images at enrollment (baseline), and at 12, 24, 36, and 48 months. Radiographs of enrolled participants were read by central readers using standard protocols, including Kellgren-Lawrence grade (KLG) and the Osteoarthritis Research

Society International classification systems. Hoffa-synovitis scores assessed using the MOAKS were obtained from the OAI. *Figure 1* shows the representative IPFP MR images with different severities of Hoffa-synovitis.

MR imaging

As part of the OAI image acquisitions, 3 T MR scanners (Siemens Magnetom Trio, Siemens, Erlangen, Germany) were used to acquire MR data. This study only used images acquired with the intermediate-weighted 2D turbo spin-echo sequence as suggested in (13) with the following parameters (25): field of view =160 mm, slice thickness =3 mm, interslice gap =0 mm, slice number =35–47 (varied by knee size), flip angle =180°, echo time =30 ms, recovery time =3,200 ms, matrix =313×448, and in-plane spatial resolution =0.357 mm × 0.511 mm.

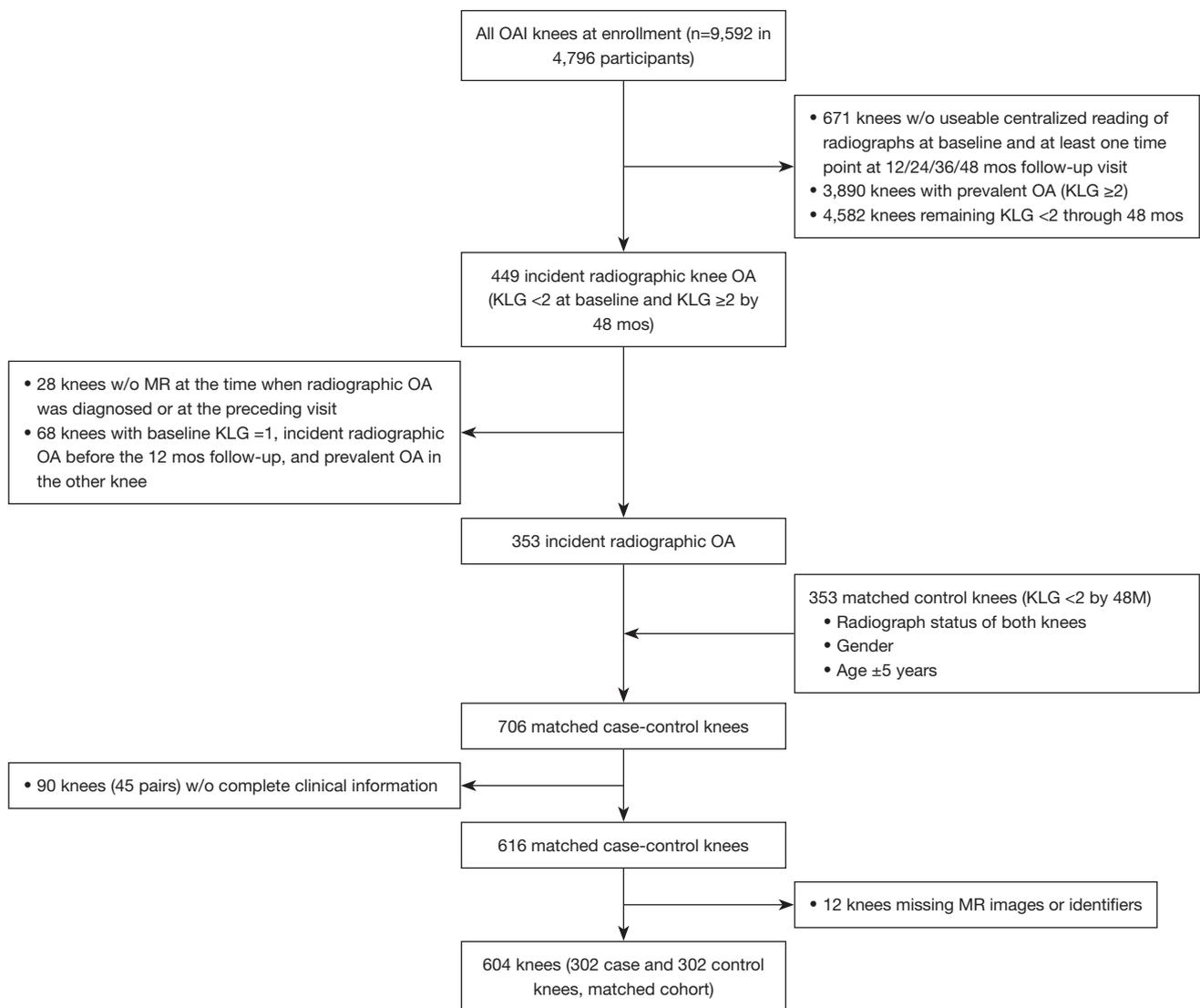


Figure 2 A flowchart showing participant inclusion and exclusion. OAI, Osteoarthritis Initiative; KLG, Kellgren-Lawrence grade; OA, osteoarthritis; w/o, without; MR, magnetic resonance.

iROA dataset

The *iROA* dataset was created for radiomic analysis using eligible participants from the OAI. Details of participant inclusion and exclusion criteria are shown in *Figure 2*. *iROA* was considered to be the first occurrence of radiographic findings compatible with OA and a knee with a baseline KLG of 0 or 1 that further progressed to KLG ≥ 2 during the designated 48-month follow-up visit. The time point for the diagnosis of *iROA* was defined as P0, and the time point for 1 year before *iROA* was defined as P-1. Cases included knees diagnosed with *iROA*

during the 48-month period in the absence of *iROA* at baseline. Controls were knees that maintained a KLG of 0 or 1 during the same period. Case and control knees were matched by well-established risk factors of knee OA including age (within 5 years), sex, and contralateral knee status (i.e., KLG status in the other knee) (26,27). As a result, 113 of 302 case knees and 113 of 302 control knees had contralateral knees with KLG ≥ 2 . Knees with missing radiographic or MR images from baseline to P0 or with missing clinical characteristics were excluded. Finally, 302 case knees and 302 control knees were selected, resulting in a total of 604 knees.

Clinical characteristics

Five clinical characteristics—age, sex, body mass index (BMI), knee injury history, and knee surgery history—were collected from the OAI. As mentioned previously, age and sex were used as the matching parameters, while the latter 3 characteristics were chosen as the clinical predictors for model optimization based on previous studies (28,29). Injury history was defined as a positive response to the question: “Have you had knee injury severe enough to limit the ability to walk for at least 2 days?” at baseline, or “Have you had knee injury severe enough to limit the ability to walk for at least 2 days since the last visit about 12 months ago?” at the follow-up visit on or before P-1. Surgical history was defined as a positive response to the question: “Have you ever had knee surgery or arthroscopy” at baseline or at the follow-up visit on or before P-1. Knee pain was not included in this study due to the following considerations: (I) given that the maximum score was 20, all enrolled participants had very low knee pain scores with a median of 0–1 at P-1 and baseline; (II) a previous study indicated that knee pain scores showed no predictive value to iROA (28); and (III) the knee pain rating scale is cumbersome and highly subjective.

Automated prediction model

IPFP segmentation: 2-and-a-half-dimensional U-Net model

Before initiation of the prediction model, a deep learning segmentation method, 2-and-a-half-dimensional (2.5D) U-Net, was implemented based on the U-Net architecture (30) to automatically segment the region of interest (ROI; i.e., the entire IPFP volume). The 2.5D U-Net was developed using 196 knee MR scans from the OAI. These scans were excluded from the dataset used for radiomic analysis (iROA dataset). Data were randomly split into training, validation, and testing (testing set 1) subsets at a ratio of 70:15:15. ROIs were manually drawn by radiologist 1 with 3 years of experience. An additional dataset (testing set 2) containing 20 participants from the iROA dataset was used to evaluate the results further.

Preprocessing

All right knees were flipped horizontally to match the left knees. Images were normalized using min-max normalization. To remove excessive and unimportant

slices, a median filter was first applied to the normalized image to mitigate noise. Next, each slice’s Shannon entropy representing the complicity of the image (31) was calculated using the following formula:

$$E(x) = -\sum_{k=0}^L p(k) * \log_2 p(k) \quad [1]$$

where x is the original image slice, $p(k)$ is the count of occurrence of the value k in slices x , and L represents the total number of bins. The normalized histogram count with 256 bins was calculated for each slice. Slices at both ends of the image have lower entropy as they contain less structural information compared to middle slices. To remove excessive image background, slices with entropy less than 5 were removed where the threshold of 5 was conservatively chosen based on the training set. As a result, at least 2 slices were kept around the ROI. The resultant slices for each patient ranged from 22 to 31 (mean=39), depending on the knee size of the patient.

U-Net development

The network structure is shown in *Figure 3*. In the 2.5D technique, the neighboring 4 slices of each target slice were used as the input. The manual ROI of the target slice was used as ground truth for the training. The model was trained using Adam optimizer (32) with a learning rate of 0.001, a first-moment exponential decay rate of 0.900, and a second-moment exponential decay rate of 0.999. The weights of the kernels were initialized using the Glorot uniform method (33), and the initial biases were set to 0. The network was trained in batches that were randomly selected from the training subset. The sparse categorical cross-entropy was used as the loss function, which was computed using the following equation:

$$J(w) = -\sum_{i=1}^N y_i \log(\hat{y}_i) \quad [2]$$

where N is the number of neurons in a layer, w refers to the model parameters, y_i is the true label, and \hat{y}_i is the predicted label. The model with the best validation sparse categorical accuracy was saved.

Postprocessing and evaluation

The output of the CNN layers ranged from 0 to 1 and indicated the possibility that the pixel belongs to the ROI. The optimal threshold was chosen using the training set,

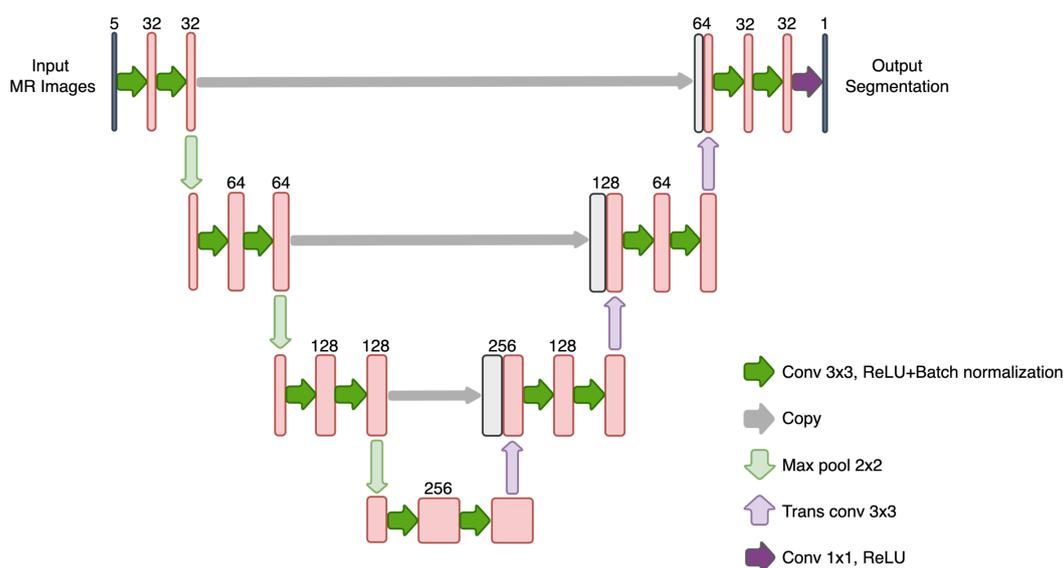


Figure 3 Architecture of 2.5D U-Net for IPFP segmentation. IPFP, infrapatellar fat pad; MR, magnetic resonance.

which was the one that achieved the best average Dice coefficient (DC) at the subject level rather than at the slice level. The model performance was assessed using the DC by comparing the automated ROIs with the manual ROIs drawn by radiologist 1 for both testing sets. To assess the inter-rater reliability, with testing set 2, another 2 versions of ROIs were manually drawn by radiologist 2 (12 years' experience) and radiologist 3 (1 year's experience) independently, without prior knowledge from radiologist 1. The DC was used to evaluate the inter-rater reliability by comparing different versions of the manual ROIs at the subject level.

Feature extraction

After obtaining the ROIs using the above-described U-Net model, an intensity normalization step was performed to remove outliers and improve reproducibility (34,35). This step was needed because the intensity of the MR image changes from scan to scan, even within the same MR imaging protocol. The following formula (36) was used:

$$\frac{x - \min(x)}{99 \text{ percentile of } x - \min(x)} \times C \quad [3]$$

where x is the image data, and C is a constant. In this study, C was empirically set to 5,000, which covered the intensity range of the original scan. Wavelet transformations (i.e., HHH, HHL, HLH, HLL, LLL, LLH, LHL, and LHH) were implemented using the Python package

PyWavelets (<https://github.com/PyWavelets/pywt>) (23,37). Laws filters were applied to the normalized data. This transformation was used to extract energy, skewness, and kurtosis features only. See Supplementary Table S1 for a detailed list of the extracted features. Three-dimensional (3D) shape features were extracted from the ROI using MATLAB R2020a (MathWorks, Natick, MA, USA). All other features were extracted from both normalized data and transformed data using LIFEx v6.44 (<https://www.lifexsoft.org>) (38). For histogram-based features, the maximum bound for gray-level resampling was set to the 99th percentile of the ROI, and the minimum bound was set to the minimum value of the data. A total of 699 features, including first-order, shape, and texture features, were extracted from each IPFP.

Dataset splitting

The iROA dataset was randomly divided into a training set (~60%, 362 knees with case:control =181:181) and a testing set (~40%, 242 knees with case:control =121:121), as performed similarly in previous studies (39-41). The training set was used for feature selection and prediction model generation, while the testing set was reserved for model assessment.

Feature selection and prediction model development (training set)

The Mann-Whitney U test was performed to identify

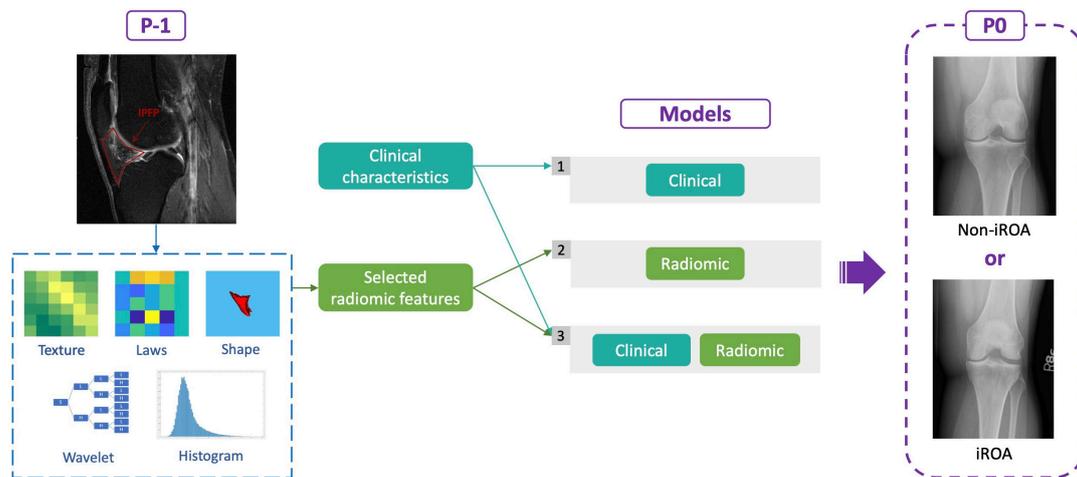


Figure 4 The prediction model development flowchart. Three models were generated to predict iROA diagnosis (P0) using features collected 1 year before the diagnosis (P-1). The following features were included in the models: (I) clinical characteristics, (II) radiomic features, and (III) clinical plus radiomic features. iROA, incident radiographic knee osteoarthritis. IPFP, infrapatellar fat pad; iROA, incident radiographic knee osteoarthritis.

features with a significant difference between the cases and controls in the training set ($P < 0.05$). The identified features that exhibited a high Spearman correlation coefficient ($|\rho| \geq 0.8$) were replaced by the one with the highest area under the receiver operating characteristic (ROC) curve (AUC). The final candidate radiomic features were then combined with the 3 clinical characteristics as input of the feature selection step. A least absolute shrinkage selection operator (LASSO) regression model with 3-fold cross-validation was applied to select the most important predictors (both clinical and radiomic). The cross-validation was performed using the training set. To avoid overfitting, the LASSO was restricted to a maximum of 10 features. These features were then used to generate logistic regression models to predict OA 1 year before iROA. The optimal threshold of the ROC prediction analysis was determined by maximizing the Youden index (sensitivity + specificity - 1). Additionally, AUC, sensitivity, specificity, and accuracy were compared to evaluate the performance of the model. The model that achieved the highest accuracy in the training set was selected as the prediction model for further assessment. Three main models were generated using (I) clinical characteristics, (II) radiomic features, and (III) combined (clinical plus radiomic) features, respectively, as illustrated in *Figure 4*. Moreover, we constructed 2 secondary models using (I) Hoffa-synovitis alone and (II) Hoffa-synovitis plus clinical characteristics for comparison purposes. All the analyses in this section were conducted in MATLAB R2020a.

Prediction performance evaluation (testing set)

The prediction performance of the generated models was further assessed in the independent testing set using the same threshold determined in the training set. The corresponding AUC, sensitivity, specificity, and accuracy were evaluated in MATLAB R2020a.

Statistical analysis

Continuous variables were assessed with a *t* test and 1-way analysis of variance, as appropriate. Categorical variables were compared with either the chi-squared test or Fisher exact test. All features that were included in the final prediction models were evaluated using univariate analysis. The analyses used for prediction model assessment were described above. The results were further visualized using the *t*-distributed stochastic neighbor embedding (*t*-SNE) algorithm (42). Finally, the DeLong test was applied to compare the AUCs across different models. The confidence interval (CI) was set at 95%.

Results

Demographic

In the iROA dataset, the average age was 61.54 ± 8.50 years, predominantly female (68%), and obese (average BMI = 28.30 ± 4.64). Other population characteristics are presented

Table 1 Populational characteristics of the iROA dataset

Characteristics	Total (n=604)	Case (n=302)	Control (n=302)	P value
Age, year	61.54±8.50	61.68±8.61	61.40±8.40	0.691
BMI, kg/m ²	28.30±4.64	29.14±4.59	27.46±4.54	<0.001
Normal	150 (24.8)	52 (17.2)	98 (32.5)	<0.001
Overweight	244 (40.4)	126 (41.7)	118 (39.1)	
Obesity	210 (34.8)	124 (41.1)	86 (28.5)	
Sex				
Male	194 (32.1)	99 (32.8)	95 (31.5)	0.727
Female	410 (67.9)	203 (67.2)	207 (68.5)	
Baseline KLG status [†]				
0/0	111 (18.4)	55 (18.2)	56 (18.5)	1.000
0/1	135 (22.4)	67 (22.2)	68 (22.5)	
1/1	131 (21.7)	66 (21.9)	65 (21.5)	
0/≥2	101 (16.7)	51 (16.9)	50 (16.6)	
1/≥2	126 (20.9)	63 (20.9)	63 (20.9)	
Injury history [§]				
0	470 (77.8)	218 (72.2)	252 (83.4)	0.006
1	106 (17.5)	64 (21.2)	42 (13.9)	
2	19 (3.1)	14 (4.6)	5 (1.7)	
3	9 (1.5)	6 (2.0)	3 (1.0)	
Total injured	134 (22.2)	84 (27.8)	50 (16.6)	0.001
Surgical history	38 (6.3)	21 (7.0)	17 (5.6)	0.503
Hoffa-synovitis score				
0	302 (50.0)	120 (39.7)	182 (60.3)	<0.001
1	264 (43.7)	154 (51.0)	110 (36.4)	
2	36 (6.0)	26 (8.6)	10 (3.3)	
3	2 (0.3)	2 (0.7)	0 (0.0)	

Data were presented as mean ± SD and n (%). BMI <25 kg/m²; overweight, 25 kg/m² ≤ BMI <30 kg/m²; obese, BMI ≥30 kg/m²; †, KLG status: “-/-” represents “ipsilateral knee KLG/contralateral knee KLG”; §, injury history: “0, 1, 2, or 3” refers to the amount of knee injury in the past medical history that resulted in limited movement for at least 2 days. iROA, incident radiographic knee osteoarthritis; BMI, body mass index; KLG, Kellgren-Lawrence grade; SD, standard deviation.

in *Table 1*. No statistical discrepancies in age, sex, or surgical history were observed between cases and controls. BMI ($P<0.001$) and the presence of knee injury history ($P=0.001$) were significantly higher in the case group.

Segmentation performance

The DC between the manual ROIs drawn by the

2 experienced radiologists (radiologist 1 *vs.* 2) was 0.900 (95% CI: 0.874–0.939), while the DCs between the more experienced and less experienced radiologists were 0.798 (95% CI: 0.782–0.815; radiologist 1 *vs.* 3) and 0.728 (95% CI: 0.638–0.820; radiologist 2 *vs.* 3). The DC between the automated and manual ROIs (radiologist 1) achieved 0.900 (95% CI: 0.891–0.908). More details of the deep learning segmentation performance are shown in *Table 2*.

Table 2 Performance of the 2.5D U-Net segmentation

	Training set	Validation set	Testing set 1	Testing set 2
DC	0.966	0.902	0.898	0.900
95% CI	0.965–0.969	0.895–0.909	0.890–0.906	0.891–0.908

DC, dice coefficient; CI, confidence interval.

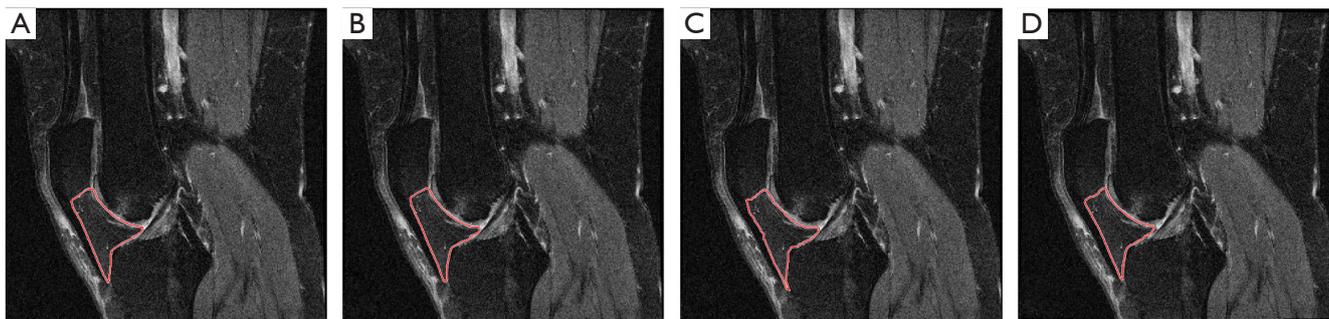


Figure 5 Comparison of manual and automated IPFP segmentations of the same knee. The red line delineates (A) manual segmentation by radiologist 1, (B) manual segmentation by radiologist 2, (C) manual segmentation by radiologist 3, and (D) 2.5D U-Net model. IPFP, infrapatellar fat pad.



Figure 6 2.5D U-Net model generated IPFP segmentations of 4 representative participants. The red line indicates the segmentation results. IPFP, infrapatellar fat pad.

A comparison between the ROIs obtained manually and automatically is shown in *Figure 5*. The ROIs of 4 representative knees generated using the deep learning model are shown in *Figure 6*.

Prediction models performance

Using the training set, 256 out of 699 radiomic features survived after the Mann-Whitney *U* test and redundant removal. A total of 3 models were optimized afterward. The prediction performances of each model were further

validated in the testing set. The outcomes are summarized in *Table 3* and are visualized in *Figure 7* and *Figure 8*. Among the 3 prediction models, the radiomic and combined models yielded superior AUCs of 0.700 (95% CI: 0.630–0.763) and 0.702 (95% CI: 0.635–0.763), respectively, when compared to the clinical model (AUC =0.591; 95% CI: 0.519–0.663). The selected predictive features of each model are shown in *Table 4*. The DeLong test showed no statistical difference between the AUCs of the radiomic and combined models ($P=0.831$). However, a significant difference was observed when comparing the radiomic ($P=0.014$) and combined

Table 3 Prediction performance in the training and testing sets for the main models

Model	Training set				Testing set			
	Accuracy	Sensitivity	Specificity	AUC (95% CI)	Accuracy	Sensitivity	Specificity	AUC (95% CI)
Clinical	0.621	0.801	0.442	0.650 (0.596–0.717)	0.531	0.639	0.421	0.591 (0.519–0.663)
Radiomic	0.668	0.580	0.790	0.744 (0.684–0.788)	0.663	0.525	0.802	0.700 (0.630–0.763)
Combined	0.682	0.652	0.746	0.747 (0.692–0.792)	0.671	0.639	0.702	0.702 (0.635–0.763)

AUC, area under the receiver operating characteristic curve; CI, confidence interval.

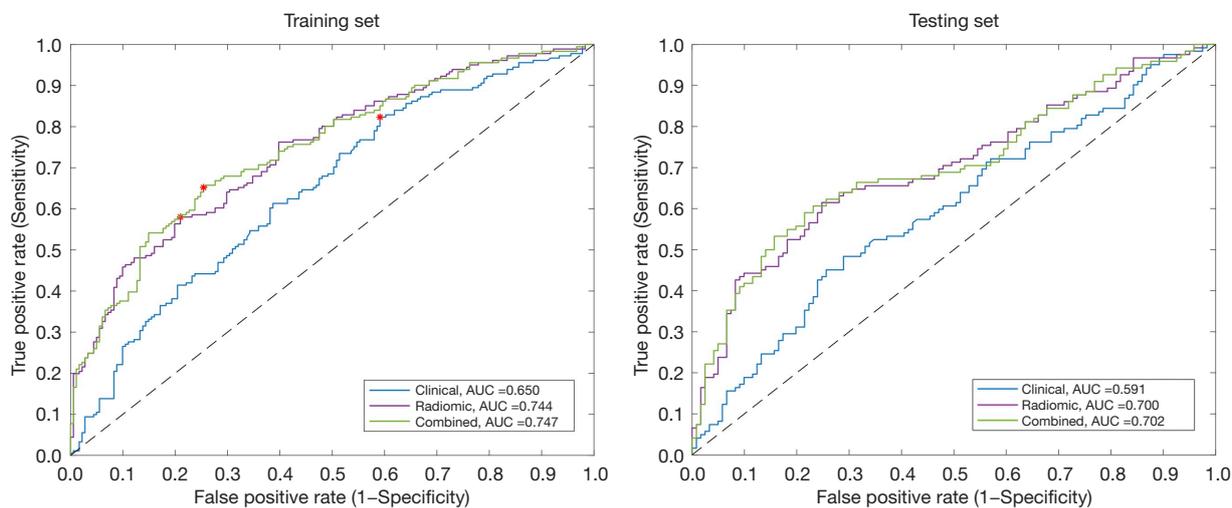


Figure 7 ROC curves of the training and testing sets for the 3 different models. The red points on the ROC curves of the training set are the optimal cutoff points (threshold) determined by maximizing the Youden index using the training set. ROC, receiver operating characteristic; AUC, area under the curve.

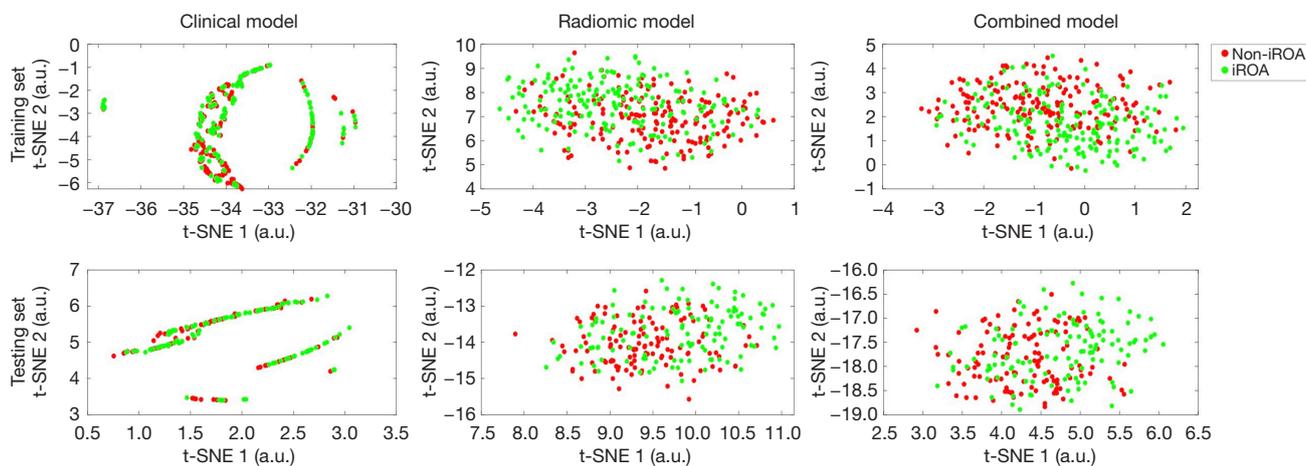


Figure 8 t-SNE visualization for the training and testing sets of all 3 models. t-SNE, t-distributed stochastic neighbor embedding; iROA, incident radiographic knee osteoarthritis.

Table 4 Selected predictive features for the 3 main prediction models

Model	Selected predictive features
Clinical	BMI
	Knee injury history
	Knee surgery history
Radiomic	Entropy_log10_ER_RE
	DISCRETIZED_peakSphere1mL_HHH
	CONVENTIONAL_max_HLH
	CONVENTIONAL_std_LHL
	DISCRETIZED_peakSphere0.5mL_LLH
	GLCM_Correlation
	GLCM_Energy[=AngularSecondMoment]_LLL GLCM_Contrast[=Variance]_LLL
	GLZLM_SZE_LHL
	GLZLM_SZHGE_LLL
Combined	BMI
	Entropy_log10_ER_RE
	DISCRETIZED_peakSphere1mL_HHH
	CONVENTIONAL_max_HLH
	CONVENTIONAL_std_LHL
	GLCM_Correlation
	GLCM_Energy[=AngularSecondMoment]_LLL
	GLCM_Contrast[=Variance]_LLL
GLZLM_SZHGE_LLL	

BMI, body mass index; GLCM, gray-level co-occurrence matrix; GLZLM, gray-level zone length matrix.

Table 5 Prediction performance in the training and testing sets for the secondary models

Model	Training set				Testing set			
	Accuracy	Sensitivity	Specificity	AUC (95% CI)	Accuracy	Sensitivity	Specificity	AUC (95% CI)
Hoffa-synovitis alone	0.611	0.591	0.630	0.618 (0.570–0.668)	0.588	0.615	0.562	0.598 (0.535–0.663)
Hoffa-synovitis plus clinical	0.652	0.635	0.696	0.697 (0.638–0.747)	0.621	0.574	0.669	0.651 (0.587–0.719)

AUC, area under the receiver operating characteristic curve; CI, confidence interval.

($P=0.004$) models with the clinical characteristics model. Moreover, for quick comparison, the Hoffa-synovitis alone model and Hoffa-synovitis plus clinical characteristics model showed AUCs of 0.598 (95% CI: 0.535–0.663) and 0.651 (95% CI: 0.587–0.719), respectively (*Table 5*). Both of them were inferior to the radiomic model.

Discussion

This study establishes iROA prediction models using MR-based radiomic features from the IPFP and corresponding clinical characteristics. Our work demonstrates that radiomic features of IPFP are indeed predictive of iROA

1 year in advance of the radiographic diagnosis.

The model combining the clinical and radiomic features yielded the best performance (AUC =0.702, sensitivity =0.639, and specificity =0.702). The model derived only from radiomic features was also promising. Its performance (AUC =0.700, sensitivity =0.525, and specificity =0.802) was very similar to the combined model, and the seemingly slightly lower prediction performance difference was not found to be statistically significant ($P=0.831$). Moreover, the radiomic model was superior to both the Hoffa-synovitis alone model (AUC =0.598, sensitivity =0.615, and specificity =0.562) and the Hoffa-synovitis plus clinical characteristics model (AUC =0.651, sensitivity =0.574, and specificity =0.669) in iROA prediction. The above evidence demonstrates that MR-based radiomics of the IPFP is a meaningful tool in prediction model development.

The IPFP is an adipose tissue in the knee that can secrete numerous inflammatory mediators associated with the pathological process of knee OA (43). The induced inflammation can interact with other structures in the knee. The association between knee OA and IPFP morphology has long been controversial due to the observed discrepancies between studies and the difficulty in evaluating IPFP. Han *et al.* (43) suggested that increased IPFP volume plays a protective role in knee OA by reducing cartilage loss and defect. However, Masaki *et al.* (10) reported that IPFP volume is associated with cartilage degeneration. Cowan *et al.* (44) further demonstrated that the enlarged IPFP volume correlates with patellofemoral joint OA and contributes to knee pain. Nonetheless, these discrepancies could be a consequence of inaccurate IPFP measurement. For signal intensity assessment, semiquantitative tools, such as effusion-synovitis and Hoffa-synovitis scores, are subjective and only reflect information in hyperintense regions of the IPFP (15,16). Quantitative assessments have demonstrated that the intensity alterations agree with the semiquantitative scores (9). However, the intensity measures applied here were mostly basic statistics, such as mean, standard deviation, and median. To reflect more detailed and comprehensive information of the IPFP, additional methods for calculating the statistical inter-relationships between neighboring voxels (e.g., textural information) are needed. Therefore, we applied radiomics analysis to capture the total signal change and precise morphology of the IPFP, which improved the reproducibility of our study. Moreover, different from the association studies, we took a step further and explored the predictive validity of IPFP radiomic features in iROA.

In our study, 10 out of the 699 radiomic features, including 5 first-order features and 5 texture features, were selected in radiomic model construction. These features are difficult to perceive by the naked eye; however, they may contain important information about the IPFP. Interestingly, all radiomic predictors selected by the regression model are signal intensity heterogeneity features rather than shape features. For instance, several gray-level co-occurrence matrix (GLCM) features were selected: GLCM_correlation measures the joint probability occurrence of the specified pixel pairs, GLCM_energy provides the sum of squared elements in the GLCM which is also known as uniformity or the angular second moment, and GLCM_contrast indicates a measure of local variations in the GLCM. As shown in *Figure 9*, the patterns of GLCM texture maps of case and control knees differed. An increased bone marrow heterogeneity can be observed in the case (iROA) patients. GLCM analysis has been found to be a powerful tool for investigating symptomatic OA (45,46), cartilage lesions (46,47), and those at high risk for OA (47,48). Moreover, a few gray-level zone length matrix (GLZLM)-related features (i.e., GLZLM_SZHGE) have shown significant value and were selected in the model. These hyperintense signal alterations may be associated with inflammation, edema, fibrosis, and acute hemorrhage (7,49). Even though prior research has revealed that IPFP shape alteration is associated with structural change and pain in knee OA (10,44), more commonly, patients are usually asymptomatic before being diagnosed with iROA because only minor deterioration has developed in the knee joints. Thus, shape features, at that time, would not be as prominent in the image and would carry less weight than signal intensity in the predictive model. Radiomics can provide readily mineable features, and our results indicate that a radiomic model can perform promisingly in this prediction task.

Previous studies have presented predictive models that used clinical and medical imaging information, as well as other regions of the knee. Janvier *et al.* (50) proposed a model that combined subchondral tibial bone texture, joint space narrowing, and clinical covariates. These features were found to be useful in predicting knee OA progression (AUC =0.77). Morales Martinez *et al.* (51) implemented a deep learning model to predict iROA from 1 to 8 years using the last healthy 3D MR images of the femur, tibia, and patella prior to the patients' diagnosis of OA. Garriga *et al.* (52) developed a model to predict iROA after 4 years based on clinical and radiographic risk factors obtained

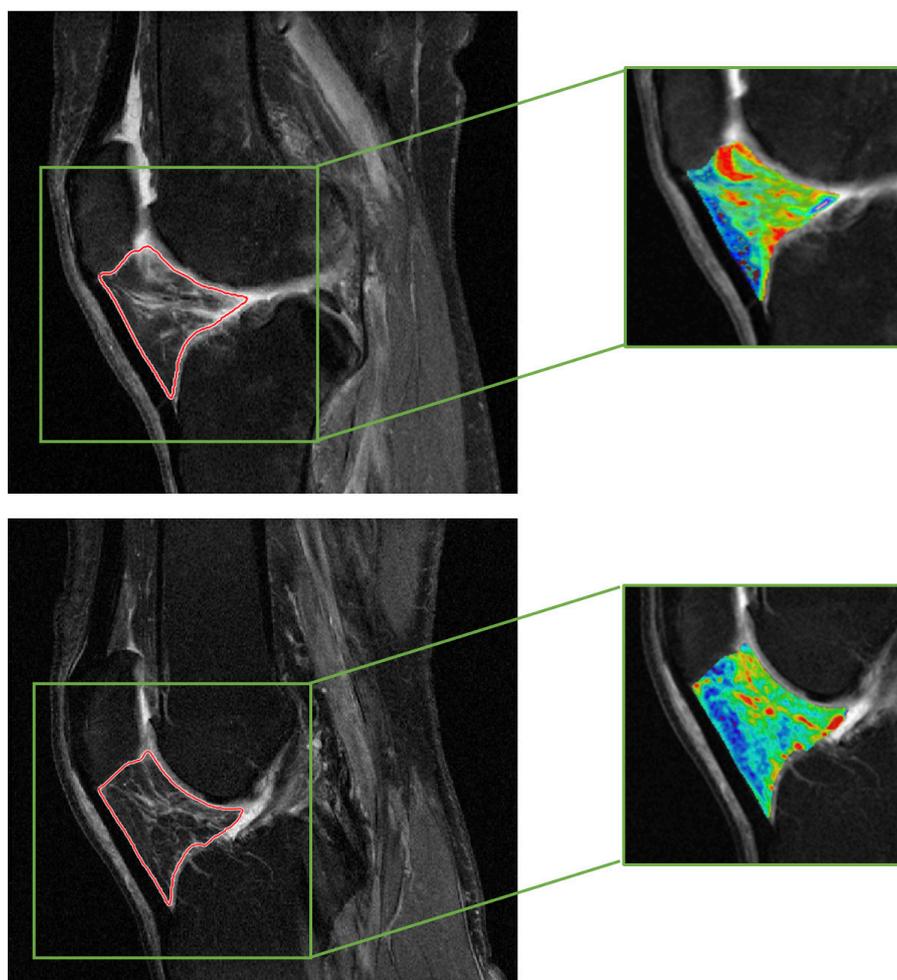


Figure 9 GLCM texture maps of representative participants. The top and bottom are from a case and a control, respectively. GLCM, gray-level cooccurrence-matrix.

from 649 middle-aged women (AUC =0.797). Sharma *et al.* (28) proposed an iROA prediction model using the MOAKS of cartilage, meniscus, and bone marrow lesions and preselected clinical predictors from high-risk individuals (AUC =0.84). However, there are a limited number of publications investigating the utility of IPFP as an independent biomarker for iROA. There is increasing interest in using IPFP-targeted interventions, particularly for the determination of IPFP preservation or resection during total knee arthroplasty (53,54). The current clinical standard using semiquantitative assessment can introduce intra- and inter-rater variability, which is not reliable to support such interventions. Very few studies are conducting research on fully automated, quantitative analyses of IPFP signal heterogeneity regarding the predictive value of iROA.

Given the fact that knee OA is a multifactorial inflammatory disease of the whole joint with a complex pathomechanism involving interactions across multiple joint tissues, we did not aim to develop a superior iROA prediction model using IPFP; instead, the main goal of our study was to explore its predictive validity in iROA using radiomic analysis on both genders across a relatively wide age range.

Additionally, to accurately and efficiently segment the IPFP, a deep learning automated model was developed. The DCs between the automated segmentation and the manual segmentations obtained from experienced radiologists achieved 0.900 (CNN *vs.* radiologist 1). For the inter-rater reliability analysis, the DC calculated using the ROIs drawn by the 2 experienced radiologists achieved 0.900 (radiologist 1 *vs.* 2), while a lower DC was shown

between the ROIs drawn by the more experienced and less experienced radiologist (0.798; radiologist 1 *vs.* 3). This indicates that the CNN generated ROIs with an accuracy comparable to that of an experienced radiologist while taking much less time compared to the manual ROI segmentation (75 s of a desktop computer *vs.* 15 min of a radiologist per case). A study conducted by Zhou *et al.* (55) implemented an approach to segment 12 tissue types of the knee joint (e.g., IPFP) using deep CNN, 3D fully connected conditional random field, and 3D simplex deformable modeling; 3D fast spin-echo image data from 20 participants were used to train the model and a leave-one-out cross-validation was performed. Their study used a smaller sample size. For the IPFP segmentation, the DC achieved 0.882, which was slightly lower than our result. More importantly, our deep learning approach is specifically designed for IPFP segmentation, which further supports our radiomic analysis.

The aforementioned studies also suggested that including clinical characteristics, such as age, gender, BMI, injury history, and surgical history, could have a notable impact on model simulation since knee OA is a multifactorial disease. These characteristics are well-established risk factors for iROA (26) and can be obtained during clinical visits; thus, we considered including these in the model development, along with the radiomic features. Our combined model only selected BMI, and this is consistent with previous publications which present evidence that weight gain increases the risk of knee OA by aggravating mechanical load, resulting in knee medial cartilage volume loss (56,57). Other demographic characteristics, such as age, gender, and knee surgery history, were not selected. This could be due to the risk factors being well-balanced between the case and control groups. In addition, knee injury, as a known risk factor of OA (58), was selected in the clinical-only model but not in the combined (clinical plus radiomic) model in this study. To further investigate this finding, we added knee injury into the combined model and reperformed model regression. The result showed that the AUC slightly increased to 0.705 (*vs.* 0.702) in the testing set; however, the DeLong test showed no statistically significant difference between the new model and the combined model ($P=0.316$). This result indicates that knee injury history provides little additional information to the radiomic model, likely because the radiomic features can capture knee injuries. Similarly, when all the unselected clinical characteristics (i.e., age, sex, knee injury history, and knee surgery history) were added manually to the combined model and the regression

analysis was reperformed, no statistically significant difference in the predictive accuracy of the model was noted (AUC =0.701; $P=0.835$, DeLong test). Moreover, Felson *et al.* (15) suggested that synovitis is an independent risk factor of iROA and that the association between cartilage and meniscal damage and iROA is mediated by synovitis. Therefore, our study also constructed 2 secondary models that included Hoffa-synovitis. The results showed that the model using IPFP radiomic features (AUC =0.700) was superior to both the Hoffa-synovitis alone model (AUC =0.598) and the Hoffa-synovitis plus clinical characteristics model (AUC =0.651) in iROA prediction. Further, during the revision of this manuscript, another study using 20 IPFP texture features also indicated that the MR-based texture analysis of the IPFP could help predict iROA 1 year before iROA diagnosis (59). This study and our own both found that IPFP texture features were predictive in iROA and outperformed clinical characteristics and/or traditional semiquantitative assessment (i.e., Hoffa-synovitis scoring system). In addition to the texture features, we also extracted and included shape features (e.g., volume and sphericity) of the IPFP when building the model. Even though the shape features were not selected in the final model, this revealed that the morphological characteristics of the IPFP were not meaningful for iROA prediction, suggesting that the inflammatory responses and metabolic functions of the IPFP play a more important role in iROA than does the structural information. Furthermore, in their study, the ROIs (i.e., IPFPs) were manually segmented, which is burdensome and clinically impractical. In contrast, we developed and used a deep learning-based automated segmentation method that enabled our model to be useful in clinical practice.

This study has some limitations. First, although the prediction models demonstrated the association between knee OA and the IPFP, the prediction accuracy using IPFP alone is not sufficient to be clinically useful. This is expected due to the fact that the poly-pathogenic process of cartilage, bone, meniscus, and synovium can contribute to the development of knee OA (16,28). A more predictive model could possibly be developed by combining other prediction factors. Second, our models have not yet been validated outside the OAI dataset, but given that OAI has a good representative population, the generalizability of this work is expected to be high. In addition, power analysis was not performed before the data analysis for our study. Moreover, post hoc power analysis was not carried out, as it has been well-documented that post hoc power

calculations are not informative and should be avoided (60–62). Moreover, BMI and the presence of knee injury history were significantly higher in the case group than in the control group, which might have influenced the prediction outcome; however, 2 simple regression models conducted using BMI only and knee injury only yielded AUCs of less than 0.575 (0.574 for the BMI-only model and 0.564 for the knee injury-only model). Thus, there was no material difference in terms of the interpretation of our results. Lastly, the time frame between P-1 and confirmed iROA was relatively short for effective lifestyle and clinical interventions. Although it would be interesting for a future study, this work did not use data with a longer follow-up period because these data are limited and are more difficult to match due to lack of imaging data, loss to follow-up, and other factors, which could affect the effectiveness and accuracy of the study. Whether radiomic analysis of the IPFP is useful in identifying high-risk patients earlier than 1 year before iROA certainly is worth further investigation. Examining the predictive value of radiographic features in predicting frequent knee pain may also be another fruitful avenue of future research.

Conclusions

This study presents a radiomic knee iROA prediction model 1 year before diagnosis using IPFP radiomic features on MR images, with fully automated deep learning segmentation. Our findings indicate that IPFP radiomic features may potentially serve as an independent imaging biomarker for iROA development.

Acknowledgments

The authors would like to acknowledge participants in the OAI study for providing this unique open-access database.

Funding: This work was in part supported by the President Foundation of the Third Affiliated Hospital of Southern Medical University (No. YM2021012 to K Yu, L Zhong, J Hu, and X Zhang). The other authors received no financial support from any source other than their affiliated institutes for the research, authorship, and publication of this work.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-22-368/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-368/coif>). KY, LZ, JH, and XZ report that this work was supported by the President Foundation of the Third Affiliated Hospital of Southern Medical University (No. YM2021012). The other authors have no 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).

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. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. *Lancet* 2019;393:1745–59.
2. GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018;392:1859–922.
3. Turkiewicz A, Petersson IF, Björk J, Hawker G, Dahlberg LE, Lohmander LS, Englund M. Current and future impact of osteoarthritis on health care: a population-based study with projections to year 2032. *Osteoarthritis Cartilage* 2014;22:1826–32.
4. Chen F, Su W, Bedenbaugh AV, Oruc A. Health care resource utilization and burden of disease in a U.S. Medicare population with a principal diagnosis of osteoarthritis of the knee. *J Med Econ* 2020;23:1151–8.
5. Zeng N, Yan ZP, Chen XY, Ni GX. Infrapatellar Fat Pad and Knee Osteoarthritis. *Aging Dis* 2020;11:1317–28.
6. Favero M, El-Hadi H, Belluzzi E, Granzotto M, Porzionato A, Sarasin G, Rambaldo A, Iacobellis C,

- Cigolotti A, Fontanella CG, Natali A, Ramonda R, Ruggieri P, De Caro R, Vettor R, Rossato M, Macchi V. Infrapatellar fat pad features in osteoarthritis: a histopathological and molecular study. *Rheumatology (Oxford)* 2017;56:1784-93.
7. Eymard F, Chevalier X. Inflammation of the infrapatellar fat pad. *Joint Bone Spine* 2016;83:389-93.
 8. van der Heijden RA, de Vries BA, Poot DHJ, van Middelkoop M, Bierma-Zeinstra SMA, Krestin GP, Oei EHG. Quantitative volume and dynamic contrast-enhanced MRI derived perfusion of the infrapatellar fat pad in patellofemoral pain. *Quant Imaging Med Surg* 2021;11:133-42.
 9. Wang K, Ding C, Hannon MJ, Chen Z, Kwok CK, Hunter DJ. Quantitative Signal Intensity Alteration in Infrapatellar Fat Pad Predicts Incident Radiographic Osteoarthritis: The Osteoarthritis Initiative. *Arthritis Care Res (Hoboken)* 2019;71:30-8.
 10. Masaki T, Takahashi K, Hashimoto S, Ikuta F, Watanabe A, Kiuchi S, Okuaki T, Mochizuki Y, Takai S. Volume change in infrapatellar fat pad is associated not with obesity but with cartilage degeneration. *J Orthop Res* 2019;37:593-600.
 11. Roemer FW, Jarraya M, Felson DT, Hayashi D, Crema MD, Loeuille D, Guermazi A. Magnetic resonance imaging of Hoffa's fat pad and relevance for osteoarthritis research: a narrative review. *Osteoarthritis Cartilage* 2016;24:383-97.
 12. Han W, Aitken D, Zhu Z, Halliday A, Wang X, Antony B, Cicuttini F, Jones G, Ding C. Signal intensity alteration in the infrapatellar fat pad at baseline for the prediction of knee symptoms and structure in older adults: a cohort study. *Ann Rheum Dis* 2016;75:1783-8.
 13. Cai J, Xu J, Wang K, Zheng S, He F, Huan S, Xu S, Zhang H, Laslett L, Ding C. Association Between Infrapatellar Fat Pad Volume and Knee Structural Changes in Patients with Knee Osteoarthritis. *J Rheumatol* 2015;42:1878-84.
 14. 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.
 15. Felson DT, Niu J, Neogi T, Goggins J, Nevitt MC, Roemer F, Torner J, Lewis CE, Guermazi A; MOST Investigators Group. Synovitis and the risk of knee osteoarthritis: the MOST Study. *Osteoarthritis Cartilage* 2016;24:458-64.
 16. Atukorala I, Kwok CK, Guermazi A, Roemer FW, Boudreau RM, Hannon MJ, Hunter DJ. Synovitis in knee osteoarthritis: a precursor of disease? *Ann Rheum Dis* 2016;75:390-5.
 17. Alvarez-Jimenez C, Sandino AA, Prasanna P, Gupta A, Viswanath SE, Romero E. Identifying Cross-Scale Associations between Radiomic and Pathomic Signatures of Non-Small Cell Lung Cancer Subtypes: Preliminary Results. *Cancers (Basel)* 2020;12:3663.
 18. Woźnicki P, Westhoff N, Huber T, Riffel P, Froelich MF, Gresser E, von Hardenberg J, Mühlberg A, Michel MS, Schoenberg SO, Nörenberg D. Multiparametric MRI for Prostate Cancer Characterization: Combined Use of Radiomics Model with PI-RADS and Clinical Parameters. *Cancers (Basel)* 2020;12:1767.
 19. Ou J, Wu L, Li R, Wu CQ, Liu J, Chen TW, Zhang XM, Tang S, Wu YP, Yang LQ, Tan BG, Lu FL. CT radiomics features to predict lymph node metastasis in advanced esophageal squamous cell carcinoma and to discriminate between regional and non-regional lymph node metastasis: a case control study. *Quant Imaging Med Surg* 2021;11:628-40.
 20. Hirvasniemi J, Klein S, Bierma-Zeinstra S, Vernooij MW, Schiphof D, Oei EHG. A machine learning approach to distinguish between knees without and with osteoarthritis using MRI-based radiomic features from tibial bone. *Eur Radiol* 2021;31:8513-21.
 21. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. *Magn Reson Med* 2018;79:2379-91.
 22. Byra M, Wu M, Zhang X, Jang H, Ma YJ, Chang EY, Shah S, Du J. Knee menisci segmentation and relaxometry of 3D ultrashort echo time cones MR imaging using attention U-Net with transfer learning. *Magn Reson Med* 2020;83:1109-22.
 23. Spuhler K, Ding J, Altbach M, Galons J, Thompson P, Stopeck A, et al. Deep learning-based whole breast segmentation to support automated breast density measurements from fat-water decomposition MRI [Abstract #0600]. ISMRM 27th Annual Meeting and Exhibition 2019. Montreal, Quebec, Canada.
 24. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in Auto-Segmentation. *Semin Radiat Oncol* 2019;29:185-97.
 25. 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.
26. Heidari B. Knee osteoarthritis prevalence, risk factors, pathogenesis and features: Part I. *Caspian J Intern Med* 2011;2:205-12.
 27. Eckstein F, Maschek S, Roemer FW, Duda GN, Sharma L, Wirth W. Cartilage loss in radiographically normal knees depends on radiographic status of the contralateral knee - data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage* 2019;27:273-7.
 28. Sharma L, Hochberg M, Nevitt M, Guermazi A, Roemer F, Crema MD, Eaton C, Jackson R, Kwok K, Cauley J, Almagor O, Chmiel JS. Knee tissue lesions and prediction of incident knee osteoarthritis over 7 years in a cohort of persons at higher risk. *Osteoarthritis Cartilage* 2017;25:1068-75.
 29. Kerkhof HJ, Bierma-Zeinstra SM, Arden NK, Metrustry S, Castano-Betancourt M, Hart DJ, Hofman A, Rivadeneira F, Oei EH, Spector TD, Uitterlinden AG, Janssens AC, Valdes AM, van Meurs JB. Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors. *Ann Rheum Dis* 2014;73:2116-21.
 30. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *Proceedings of Medical Image Computing and Computer-Assisted Intervention 2015*. Springer; 2015;234-41.
 31. Mello Román JC, Vázquez Noguera JL, Legal-Ayala H, Pinto-Roa DP, Gomez-Guerrero S, García Torres M. Entropy and Contrast Enhancement of Infrared Thermal Images Using the Multiscale Top-Hat Transform. *Entropy (Basel)* 2019;21:244.
 32. P KD, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv 2014:1421.6980*.
 33. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics 2010*;249-56. *JMLR Workshop and Conference Proceedings*.
 34. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging-"how-to" guide and critical reflection. *Insights Imaging* 2020;11:91.
 35. Traverso A, Kazmierski M, Welch ML, Weiss J, Fiset S, Foltz WD, Gladwish A, Dekker A, Jaffray D, Wee L, Han K. Sensitivity of radiomic features to inter-observer variability and image pre-processing in Apparent Diffusion Coefficient (ADC) maps of cervix cancer patients. *Radiother Oncol* 2020;143:88-94.
 36. Cusumano D, Dinapoli N, Boldrini L, Chiloiro G, Gatta R, Masciocchi C, Lenkowicz J, Casà C, Damiani A, Azario L, Van Soest J, Dekker A, Lambin P, De Spirito M, Valentini V. Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer. *Radiol Med* 2018;123:286-95.
 37. Lee GR, Gommers R, Wasilewski F, Wohlfahrt K, O'Leary A. PyWavelets: A Python package for wavelet analysis. *J Open Source Softw* 2019;4:1237.
 38. Nioche C, Orhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, Pellot-Barakat C, Soussan M, Frouin F, Buvat I. LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res* 2018;78:4786-9.
 39. Araujo-Filho JAB, Mayoral M, Zheng J, Tan KS, Gibbs P, Shepherd AF, Rimmer A, Simone CB 2nd, Riely G, Huang J, Ginsberg MS. CT Radiomic Features for Predicting Resectability and TNM Staging in Thymic Epithelial Tumors. *Ann Thorac Surg* 2022;113:957-65.
 40. Franzese C, Cozzi L, Badalamenti M, Baldaccini D, D'Agostino G, Fogliata A, Navarra P, Franceschini D, Comito T, Clerici E, Reggiori G, Tomatis S, Scorsetti M. Radiomics-based prognosis classification for high-risk prostate cancer treated with radiotherapy. *Strahlenther Onkol* 2022;198:710-8.
 41. Iyer S, Ismail M, Tamrazi B, Margol A, Verma R, Correa R, Prasanna P, Beig N, Bera K, Statsevych V, Judkins A, Madabhushi A, Tiwari P. Deformation heterogeneity radiomics to predict molecular subtypes of pediatric Medulloblastoma on routine MRI. *Medical Imaging 2019: Computer-Aided Diagnosis 2019*;10950:359-64.
 42. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;9:2579-605.
 43. Han W, Cai S, Liu Z, Jin X, Wang X, Antony B, Cao Y, Aitken D, Cicuttini F, Jones G, Ding C. Infrapatellar fat pad in the knee: is local fat good or bad for knee osteoarthritis? *Arthritis Res Ther* 2014;16:R145.
 44. Cowan SM, Hart HF, Warden SJ, Crossley KM. Infrapatellar fat pad volume is greater in individuals with patellofemoral joint osteoarthritis and associated with pain. *Rheumatol Int* 2015;35:1439-42.
 45. Peuna A, Hekkala J, Haapea M, Podlipská J, Guermazi A, Saarakkala S, Nieminen MT, Lammintausta E. Variable angle gray level co-occurrence matrix analysis of T2 relaxation time maps reveals degenerative changes of cartilage in knee osteoarthritis: Oulu knee osteoarthritis study. *J Magn Reson Imaging* 2018;47:1316-27.
 46. Schooler J, Kumar D, Nardo L, McCulloch C, Li X, Link

- TM, Majumdar S. Longitudinal evaluation of T1 ρ and T2 spatial distribution in osteoarthritic and healthy medial knee cartilage. *Osteoarthritis Cartilage* 2014;22:51-62.
47. Joseph GB, Baum T, Alizai H, Carballido-Gamio J, Nardo L, Virayavanich W, Lynch JA, Nevitt MC, McCulloch CE, Majumdar S, Link TM. Baseline mean and heterogeneity of MR cartilage T2 are associated with morphologic degeneration of cartilage, meniscus, and bone marrow over 3 years--data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage* 2012;20:727-35.
 48. Baum T, Joseph GB, Nardo L, Virayavanich W, Arulanandan A, Alizai H, Carballido-Gamio J, Nevitt MC, Lynch J, McCulloch CE, Link TM. Correlation of magnetic resonance imaging-based knee cartilage T2 measurements and focal knee lesions with body mass index: thirty-six-month followup data from a longitudinal, observational multicenter study. *Arthritis Care Res (Hoboken)* 2013;65:23-33.
 49. Dragoo JL, Johnson C, McConnell J. Evaluation and treatment of disorders of the infrapatellar fat pad. *Sports Med* 2012;42:51-67.
 50. Janvier T, Jennane R, Toumi H, Lespessailles E. Subchondral tibial bone texture predicts the incidence of radiographic knee osteoarthritis: data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage* 2017;25:2047-54.
 51. Morales Martinez A, Caliva F, Flament I, Liu F, Lee J, Cao P, Shah R, Majumdar S, Pedoia V. Learning osteoarthritis imaging biomarkers from bone surface spherical encoding. *Magn Reson Med* 2020;84:2190-203.
 52. Garriga C, Sánchez-Santos MT, Judge A, Hart D, Spector T, Cooper C, Arden NK. Predicting Incident Radiographic Knee Osteoarthritis in Middle-Aged Women Within Four Years: The Importance of Knee-Level Prognostic Factors. *Arthritis Care Res (Hoboken)* 2020;72:88-97.
 53. Asai K, Nakase J, Oshima T, Shimozaki K, Yoshimizu R, Tsuchiya H. Partial resection of the infrapatellar fat pad during anterior cruciate ligament reconstruction has no effect on clinical outcomes including anterior knee pain. *Arch Orthop Trauma Surg* 2020;140:1751-7.
 54. Zhu Z, Han W, Lu M, Lin J, Yin Z, Shang X, Weng X, Zha Z, Tian J, Lei G, Hunter DJ, Ding C. Effects of infrapatellar fat pad preservation versus resection on clinical outcomes after total knee arthroplasty in patients with knee osteoarthritis (IPAKA): study protocol for a multicentre, randomised, controlled clinical trial. *BMJ Open* 2020;10:e043088.
 55. Zhou Z, Zhao G, Kijowski R, Liu F. Deep convolutional neural network for segmentation of knee joint anatomy. *Magn Reson Med* 2018;80:2759-70.
 56. Teichtahl AJ, Wluka AE, Tanamas SK, Wang Y, Strauss BJ, Proietto J, Dixon JB, Jones G, Forbes A, Cicuttini FM. Weight change and change in tibial cartilage volume and symptoms in obese adults. *Ann Rheum Dis* 2015;74:1024-9.
 57. Felson DT, Zhang Y, Anthony JM, Naimark A, Anderson JJ. Weight loss reduces the risk for symptomatic knee osteoarthritis in women. The Framingham Study. *Ann Intern Med* 1992;116:535-9.
 58. Muthuri SG, McWilliams DF, Doherty M, Zhang W. History of knee injuries and knee osteoarthritis: a meta-analysis of observational studies. *Osteoarthritis Cartilage* 2011;19:1286-93.
 59. Li J, Fu S, Gong Z, Zhu Z, Zeng D, Cao P, Lin T, Chen T, Wang X, Lartey R, Kwok CK, Guermazi A, Roemer FW, Hunter DJ, Ma J, Ding C. MRI-based Texture Analysis of Infrapatellar Fat Pad to Predict Knee Osteoarthritis Incidence. *Radiology* 2022;304:611-21.
 60. Plate JDJ, Borggreve AS, van Hillegersberg R, Peelen LM. Post Hoc Power Calculation: Observing the Expected. *Ann Surg* 2019;269:e11.
 61. Hoenig JM, Heisey DM. The abuse of power: the pervasive fallacy of power calculations for data analysis. *Am Stat* 2001;55:19-24.
 62. Althouse AD, Chow ZR. Comment on "Post-hoc Power: If You Must, At Least Try to Understand". *Ann Surg* 2019;270:e78-9.

Cite this article as: Yu K, Ying J, Zhao T, Lei L, Zhong L, Hu J, Zhou JW, Huang C, Zhang X. Prediction model for knee osteoarthritis using magnetic resonance-based radiomic features from the infrapatellar fat pad: data from the osteoarthritis initiative. *Quant Imaging Med Surg* 2023;13(1):352-369. doi: 10.21037/qims-22-368

Table S1 List of the extracted features

Category (Settings)	Features
First-order	
Intensity-based statistics	Mean, variance, skewness, kurtosis, median, minimum grey level, maximum grey level, energy, kurtosis, excess kurtosis, maximum intensity, minimum intensity, mean intensity, intensity quartiles, intensity skewness, intensity standard deviation, and peak intensity
Intensity histogram	Discretized kurtosis, discretized excess kurtosis, discretized maximum intensity, discretized minimum intensity, discretized mean intensity, discretized intensity quartiles, discretized intensity skewness, discretized intensity standard deviation, discretized peak intensity, discretized intensity entropy, and discretized intensity uniformity (energy)
Shape	
2D	Mesh surface, pixel area, perimeter, perimeter to surface ratio, sphericity, spherical disproportion, maximum 2D diameter, major axis length, minor axis length, and elongation
3D	Volume, solidity, eccentricity, equivalent diameter, extent, and surface area
Texture	
GLCM	Autocorrelation, cluster prominence, cluster shade, cluster tendency, contrast, correlation, difference entropy, dissimilarity, energy, joint entropy, inverse difference, homogeneity, informational measure of correlation 1, informational measure of correlation 2, inverse difference moment normalized, inverse difference normalized, inverse variance, joint maximum, sum entropy, sum variance, and joint variance
GLRLM	Short-run emphasis, long-run emphasis, low gray-level run emphasis, high gray-level run emphasis, short-run low gray-level emphasis, short-run high gray-level emphasis, long-run low gray-level emphasis, long-run high gray-level emphasis, gray-level non-uniformity for run, run length non-uniformity, and run percentage
NGLDM	Coarseness, contrast, and busyness
GLZLM	Short-zone emphasis, long-zone emphasis, low gray-level zone emphasis, high gray-level zone emphasis, short-zone low gray-level emphasis, short-zone high gray-level emphasis, long-zone low gray-level emphasis, long-zone high gray-level emphasis, gray-level non-uniformity for zone, zone length non-uniformity, and zone percentage

GLCM, grey-level co-occurrence matrix; GLRLM, grey-level run length matrix; NGLDM, neighborhood grey-level difference matrix; GLZLM, grey-level zone length matrix.