



OPEN

## Subchondral tibial bone texture of conventional X-rays predicts total knee arthroplasty

Ahmad Almhdie-Imjabbar<sup>1,2</sup>, Hechmi Toumi<sup>1,2,3</sup>, Khaled Harrar<sup>4</sup>, Antonio Pinti<sup>1,5</sup> & Eric Lespessailles<sup>1,2,3</sup>✉

Lacking disease-modifying osteoarthritis drugs (DMOADs) for knee osteoarthritis (KOA), Total Knee Arthroplasty (TKA) is often considered an important clinical outcome. Thus, it is important to determine the most relevant factors that are associated with the risk of TKA. The present study aims to develop a model based on a combination of X-ray trabecular bone texture (TBT) analysis, and clinical and radiological information to predict TKA risk in patients with or at risk of developing KOA. This study involved 4382 radiographs, obtained from the OsteoArthritis Initiative (OAI) cohort. Cases were defined as patients with TKA on at least one knee prior to the 108-month follow-up time point and controls were defined as patients who had never undergone TKA. The proposed TKA-risk prediction model, combining TBT parameters and Kellgren–Lawrence (KL) grades, was performed using logistic regression. The proposed model achieved an AUC of 0.92 (95% Confidence Interval [CI] 0.90, 0.93), while the KL model achieved an AUC of 0.86 (95% CI 0.84, 0.86;  $p < 0.001$ ). This study presents a new TKA prediction model with a good performance permitting the identification of at risk patient with a good sensitivity and specificity, with a 60% increase in TKA case prediction as reflected by the recall values.

Osteoarthritis (OA), the most common form of arthritis, affects a substantial and increasing proportion of the population due to the combined effect of ageing and increasing overweight and obesity in the population. It is estimated that 250 million people are already affected by this leading cause of disability in older adults<sup>1,2</sup>. Although both non weight-bearing (hand) and weight-bearing joints such as the spine, hip, and knee are commonly affected, knee OA constitutes 85% of the global burden of OA in general<sup>3</sup>, and is the most frequently studied location in OA imaging research<sup>4</sup>. Bone metabolism and particularly subchondral bone plays a crucial role in the pathophysiology of knee osteoarthritis (KOA)<sup>5</sup>. Although several promising therapies have recently emerged that could reshape the overview of OA management in the near future<sup>6</sup>, there are still no approved agents that can be considered as validated disease-modifying osteoarthritis drugs (DMOADs)<sup>7,8</sup>, and consequently joint replacement surgery is often considered to be the sole effective treatment option for KOA when non-surgical management of KOA has failed. Knee replacement is considered an important clinical outcome of KOA<sup>9,10</sup> and thus it is of the utmost importance to determine the most relevant factors that are associated with the risk of total knee arthroplasty (TKA)<sup>11</sup>.

Bone metabolism and particularly subchondral bone plays a key role in the pathophysiology of KOA<sup>5,12</sup>. Recently, trabecular bone texture (TBT) analysis of subchondral bone on conventional radiographs of the knee was shown to be a promising method for identifying patients at-risk of KOA progression<sup>13–18</sup>. An overview of the interest of TBT analysis in the assessment of KOA was recently published<sup>19</sup>. Applying a combined approach of new machine learning techniques and trabecular texture analysis has already been successfully used to improve the prediction of KOA progression in the Multicenter Osteoarthritis Study (MOST) and the OsteoArthritis Initiative (OAI) cohorts<sup>20</sup>.

The most important unresolved hurdle is identifying and classifying patients who are at risk for TKA based on bone texture analysis. In the present work, we aim to demonstrate that applying trabecular bone texture analysis to baseline knee radiographs can predict the risk of TKA in the OAI cohort. To our knowledge, no prior study

<sup>1</sup>EA 4708- I3MTO Laboratory, University of Orleans, Orleans, France. <sup>2</sup>Translational Medicine Research Platform, PRIMMO, Regional Hospital of Orleans, Orleans, France. <sup>3</sup>Department of Rheumatology, Regional Hospital of Orleans, 14 Avenue de l'Hôpital, 45067 Orleans Cedex 2, France. <sup>4</sup>LIST Laboratory, University M'Hamed Bougara of Boumerdes, Boumerdes, Algeria. <sup>5</sup>DeVisu-Design, Visuel, Urbain, EA 2445, UPHF, Valenciennes, France. ✉email: eric.lespessailles@chr-orleans.fr

has evaluated the utility of radiographic TBT analysis for predicting which individuals from a cohort of knee OA patients will undergo TKA.

## Methods

The data used in this study were obtained from the OAI cohort in patients with or at risk of developing symptomatic KOA in at least one knee. The OAI study protocol was approved by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) and is registered on ClinicalTrials.gov as “Osteoarthritis Initiative (OAI): A Knee Health Study”, NCT#00080171. The OAI is a public access research database. The National Institute of Mental Health (NIMH) Data Archive (NDA) website (<https://nda.nih.gov/oai/query-download>) provides useful information on how to obtain permission to access to the OAI data. The access to the raw data used in our study was granted by the NDA Osteoarthritis Initiative permission group. The OAI study was carried out in accordance with all pertinent guidelines and regulations, and written and informed consent was obtained from participants prior to each clinical visit in the study. Details about the acquisition and grading protocols are available online at (<https://nda.nih.gov/oai/study-details>). The OAI is a longitudinal cohort study designed to identify biomarkers of the incidence and/or progression of KOA. Both knees of 4796 participants were studied using bilateral posteroanterior fixed-flexion knee radiography at baseline, and annually over 8 years of follow-up. The TKA outcome data were collected during 9 years. OAI participants were 45 to 79 years old at baseline, with or at risk of developing symptomatic KOA in at least one knee. At each yearly follow-up time point, OAI participants were interviewed and asked about TKA in the preceding 12 months<sup>9</sup>. The OAI study was approved by the institutional review boards at each OAI clinical site and the coordinating center (University of California, San Francisco) and informed consent was obtained from the participants. All research was performed in accordance with relevant guidelines/regulations and with the Declaration of Helsinki.

**Data selection.** Cases were defined as patients underwent TKA on at least one knee from 12-months to 108-months follow-up time points from baseline. Controls were defined as patients who had never undergone a TKA prior to the 108-month follow-up time point. Exclusion criteria included patients about whom information on clinical covariates (CCov), namely age, gender and body mass index (BMI), and radiological Kellgren–Lawrence (KL) grades, was lacking, at baseline. Previously published studies have shown only moderate performance for predicting KOA progression when using pain and/or history of knee injury<sup>21,22</sup>. However, since data for WOMAC pain and history of previous injury were available for the OAI cohort, the CCov also included these additional predictors. Patients with any type of knee joint replacement reported on at least one knee at baseline were also excluded. At follow-up time points, patients with single-compartment prosthesis or non-confirmed TKA were also excluded. KL scoring was performed in the OAI using two expert readers who independently assessed each radiograph; differences were adjudicated by a group including a more senior reader<sup>23</sup>.

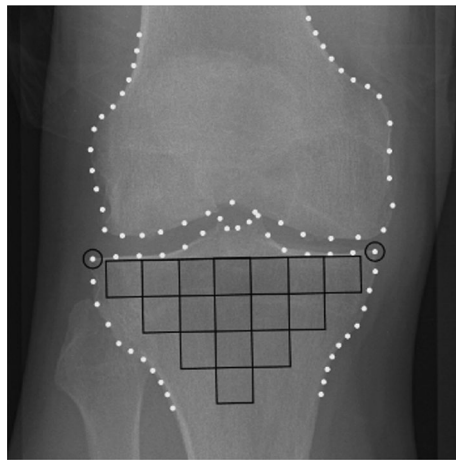
In this study, the inclusion/exclusion approach was set at baseline, and then knees were followed over time to see which ones underwent TKA. To evaluate the performance of the prediction models with respect to KOA severity, we considered two scenarios. In Scenario I, the knees of all patients (KL ≤ 4) at baseline were considered except those having previously undergone knee surgery, in accordance with the study of Leung et al.<sup>24</sup>. At follow-up time points, for a patient with TKA on one knee, only the knee that had undergone surgery was included. Furthermore, if both knees underwent TKA, only the first one was included. For each control patient, the knee with the most severe radiographic OA at baseline was excluded. Finally, if both knees of control patients had the same KL score, then one of them was randomly selected and the other one was excluded.

In accordance with the study of Podsiadlo et al.<sup>25</sup>, in Scenario II, knees with KL = 4 at baseline were excluded. In this scenario, for a patient who had undergone TKA on one knee, only the knee that had undergone TKA was included. If the included knee had a KL = 4 at baseline, both knees were excluded. If both knees underwent TKA, only the first one was included, unless the included knee had a KL = 4 at baseline, in which case this knee was excluded, and the other one was included. Both knees were excluded if they had a KL = 4 at baseline. For each control patient, the knee with the most severe radiographic OA at baseline was excluded. Here also, both knees were excluded if they had a KL = 4 at baseline. As a result, all included knees had a KL < 4 at baseline.

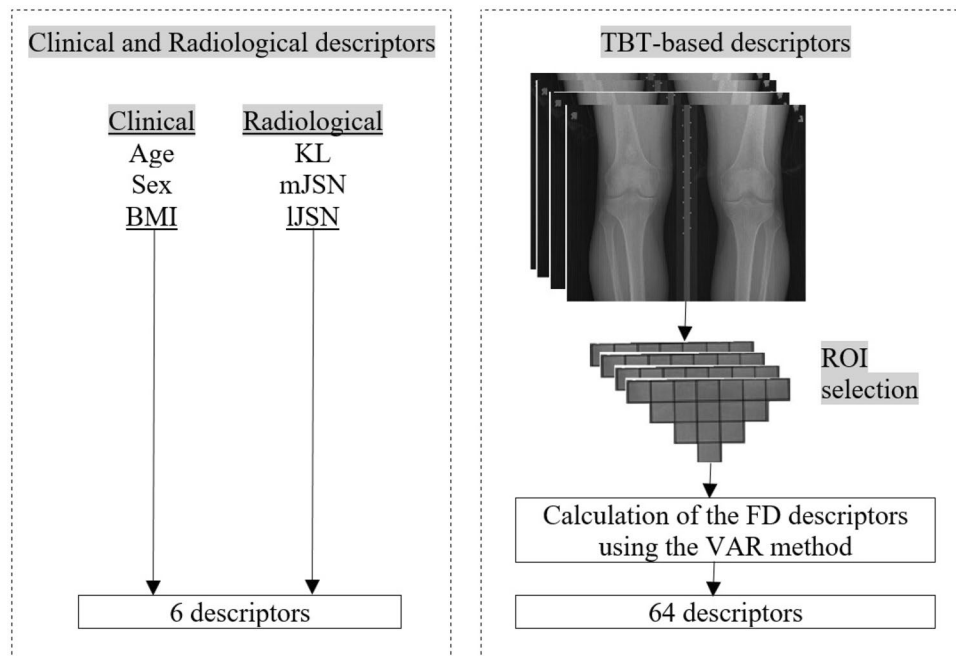
**Calculation of TBT parameters.** A total of 16 TBT regions of interest (ROIs) were selected for each knee X-ray using an automated method<sup>13,26</sup>. The 16 ROIs cover the entire tibial subchondral bone structure, immediately under the medial and lateral cortical plates of the tibia. The choice of such ROIs was inspired from the work of Janvier et al.<sup>13,27</sup>.

For each knee X-ray, the ROI selection method consisted in identifying several anatomical points of the tibia, using the BoneFinder software (<http://bone-finder.com/>)<sup>28,29</sup>. The segmentation provided by BoneFinder is based on a machine-learning approach that has been trained on a large set of manually-annotated radiographs. The selected radiographs were not preprocessed prior to the segmentation process. Applied on different datasets, BoneFinder was able to detect the contour points of skeletal structures from 2D radiographs, in a fully automatic manner with high accuracy<sup>30,31</sup>. In our study, BoneFinder was used to localize the contour points of both femoral and tibial parts of the knee. In addition, BoneFinder does not need dedicated personnel or special training. In the current study, BoneFinder successfully provided correct contours of the femoral and tibial parts of all considered knees. Using BoneFinder, the identification process of the ROIs for each knee took only a few seconds.

The tibial subchondral baseline was then defined as the line going between the lateral and medial extremities of the proximal tibia. To prevent periarticular osteophytes and fibular head overlay, an offset equal to 10% of the tibial subchondral baseline was applied to the horizontal positioning. Both the orientation and size of the 16-ROIs patchwork were calculated based on this line (Fig. 1). The dimension of the squared-shape ROIs was proportional to the knee width defined as the distance between the outer tibial margins.



**Figure 1.** ROIs automatically selected in the tibial subchondral bone. Dots represent the femoral and tibial bone edges, delimited by BoneFinder software.

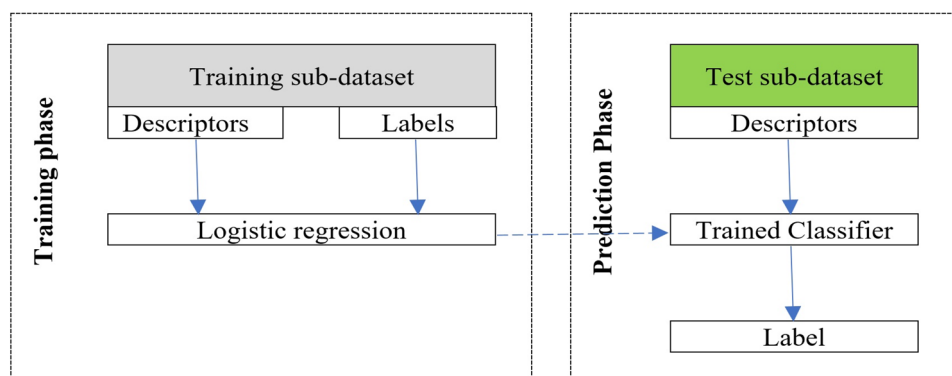


**Figure 2.** Descriptors used in the proposed model.

In the present study, TBT was characterized using fractal analysis<sup>25,32</sup> and the fractal parameters (H) were computed using the Variogram (VAR)<sup>27,33</sup>. Each ROI was quantified by means of 4 TBT variograms corresponding to the H values computed along with the two main loading directions: horizontal (*h*-direction, image lines) and vertical (*v*-direction, image columns), and for the two scales of observation: the  $\mu$ -scale (under 400 mm) and the *m*-scale (above 600 mm). As a result, TBTp represents a vector of four TBT descriptors  $\{H_{\mu v}; H_{\mu h}; H_{m v}; H_{m h}\}$  computed for each of the 16 ROIs, resulting in 64 descriptors for each image, where  $H_{\mu v}; H_{\mu h}; H_{m v}; H_{m h}$  are the microscopic horizontal, microscopic vertical, macroscopic horizontal and macroscopic vertical fractal parameters, respectively.

**Statistical analysis.** In this study, logistic regression was used to predict TKA risk. Logistic regression is widely used for modelling KOA progression prediction<sup>13,16,34</sup> and TKA risk prediction<sup>24</sup>. Several statistical models were developed involving not only clinical covariates and radiological scores but also TBT-based parameters (Fig. 2):

- Model 1: CCov.
- Model 2: KL.



**Figure 3.** Schematic diagram of one round of machine learning prediction, repeated  $300 \times 10$  times and then averaged.

- Model 3: TBTp.
- Model 4: TBTp + CCov.
- Model 5: TBTp + JSN (JSN: OARSI Joint Space Narrowing in medial and JSNL compartments).
- Model 6: TBTp + KL.
- Model 7: TBTp + KL + CCov + JSNM (JSNM: medial OARSI Joint Space Narrowing).
- Model 8: TBTp + KL + CCov + JSNL (JSNL: lateral OARSI Joint Space Narrowing).

To avoid any bias in the distribution of training and validation (testing) datasets, each model was evaluated using a tenfold cross-validation repeated 300 times. The tenfold cross-validation involved randomly partitioning the given dataset into 10 equally-sized subdatasets. At each iteration, a single subdataset was used as the validation data for testing the model, and the remaining 9 subdatasets were used as training data. The whole cross-validation process was then repeated 300 times. The validation results were averaged to give a single estimate of the model's predictive performance.

Note that all observations were used for both training and validation, and each observation was used only once for validation. Each observation involves the descriptors of one given set of patients. The flowchart of the proposed machine learning prediction model is illustrated in Fig. 3.

The AUC was used to evaluate the global performance of the models. The model classification accuracy (ACC), i.e. the probability that a random sample is correctly classified, was also computed to investigate the relevance of the different models. An ACC is defined as the ratio of the number of correct predictions to the total number of predictions. It is expressed by:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

and the balanced accuracy (BA), which is a metric that is usually used to evaluate the performance of a binary classifier, is expressed by:

$$\text{BA} = \frac{1}{2} \times \left( \frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}} \right) \quad (2)$$

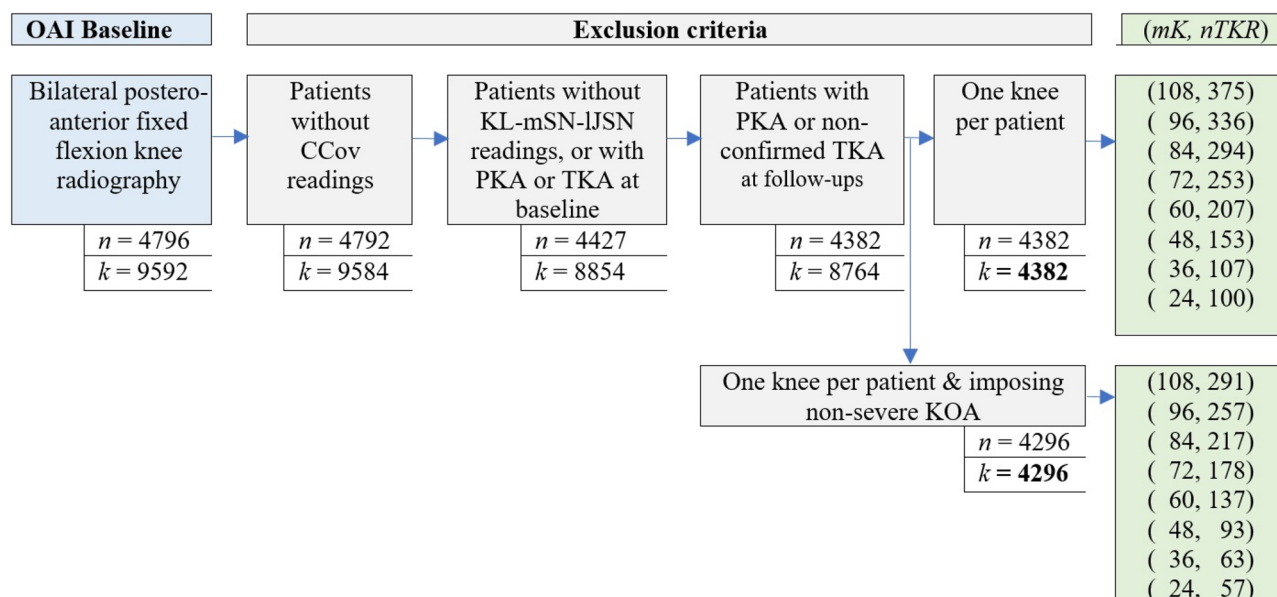
where TP, FP, TN, and FN mean True Positive, False Positive, True Negative, and False Negative, respectively. Moreover, the F1 score was calculated as another performance measure of the tested models<sup>35</sup>. The F1 score provides a combined measure of the precision and recall of the model and defined as the harmonic mean of precision and recall. The precision refers to the positive predictive value, the percentage of correctly predicted progressors to the number of actual progressors, and the recall provides the ability of a model to predict all progressors and refers to the true positive rate or the sensitivity.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

All statistical analyses were performed using the R Statistical tool<sup>36</sup> (version 4.0.2, 2020-06-22) including the MASS package (Modern Applied Statistics with S, version\_7.3-51.6, 2020-04-26) for stepwise Akaike Information Criterion (AIC)<sup>37</sup> optimization, the Caret package (Classification And REGression Training) for the



**Figure 4.** TKR-based data selection with and without imposing severe knees exclusion.  $n$  and  $k$  denote the number of patients and knees, respectively. PKA denotes partial knee arthroplasty.  $nTKR$  denotes the number of TKA knees prior to  $mK$  months' follow-up (closest contact after TKA).

	Baseline (scenario I)			Baseline (scenario II)		
	Controls	Cases	Total	Controls	Cases	Total
N° of knees	4007	375	4382	4005	291	4296
Age (years)	62.3(±9.0)	62.4(±8.1)	62.3(±8.8)	61.0(±9.20)	63.5(±8.0)	61.0(±9.2)
BMI (kg/m <sup>2</sup> )	29.4(±4.6)	30.6(±4.7)	29.6(±4.7)	28.4(±4.8)	30.0(±4.9)	28.5(±4.8)
<b>Gender</b>						
F	57.9%	60.3%	58.1%	57.9%	65.6%	58.4%
M	42.1%	39.7%	41.9%	42.1%	34.4%	41.6%
<b>KL grade</b>						
0	2113	12	2125	2113	14	2127
1	820	22	842	820	22	842
2	834	79	913	834	91	925
3	238	153	391	238	164	402
4	2	109	111	0	0	0

**Table 1.** Characteristics of the datasets included in this study. Values for age and BMI are represented as mean ( $\pm$  standard deviation).

cross-validation training and the pROC for Receiver Operating Characteristic (ROC) curve comparisons, in particular partial Areas Under the Curve (pAUC). The comparison of the models' performances was carried out with the ROC curves using the Delong method<sup>38</sup>. To reduce the number of parameters before training the prediction models, a backward selection of the parameters was automatically performed using the AIC as an iterative criterion. At each iteration, the AIC removes one parameter and preserves the most efficient parameter(s) to limit overfitting effects.

## Results

The dataset obtained for Scenario I included 4382 knees (58% women, 75% left knee) that met the inclusion criteria and were judged as eligible for this study, in which 375 knees underwent TKA on at least one knee from 12-months to 108-months follow-up time points. The dataset obtained for Scenario II included 4296 knees (58% women, 76% left knee) that were judged as eligible for this study, in which 291 knees underwent TKA prior to the 108-month follow-up. Figure 4 shows more details about our method of case-control selection, and Table 1 describes the characteristics of the two datasets.

Figures S1 and S2, presented in supplementary files, illustrate the ROIs which offered the best predictive association with TKA based on the TBT parameters selected by the AIC algorithm. In total, 38 and 30 TBT parameters were selected in Scenario I and Scenario II, respectively.

Out of the total number of knees that underwent TKA prior to the 108-month follow-up, more than 87% had a  $KL \geq 2$  at baseline, while more than 73% of the control knees had a  $KL < 2$  at baseline.

Figure 5 show the ROC curves obtained by the 8 prediction models listed in the previous section, using the datasets of Scenario I and Scenario II, respectively. The model based only on CCov was not strongly predictive ( $AUC < 0.7$ ). Combining TBT parameters and KL grades improved the performance of the TKA-risk prediction compared to the classical reference model (KL). The TBTP-KL model achieved an AUC of 0.92 (95% Confidence Interval [CI] 0.90, 0.93), using the dataset of Scenario I and an AUC of 0.89 (95% CI 0.87, 0.91), using the dataset of Scenario II, while the KL model achieved an AUC of 0.86 (95% CI 0.84, 0.86;  $P < 0.001$ ) using the dataset of Scenario I and an AUC of 0.81 (95% CI 0.78, 0.83;  $P < 0.001$ ) using the dataset of Scenario II. The sensitivity of Model 1 (using only cov parameters, Scenarios I and II) and of Model 2 (using only KL scores, Scenario I) is zero, as predicting all subjects as controls. Table 2 reports the obtained F1 scores for the different tested models for both scenarios. The prediction model Model 6, combining KL and TBT, provided a better performance than the models Model 1, Model 2, and Model 3 where only one descriptor was employed as summarized in Table 2. The performance of the TBTP-KL model was not improved when including CCov, JSNM, or JSNL parameters in addition to the TBTP and KL descriptors. More statistical results are summarized in Table 2. For example, while the KL model (Model 2) provided a recall value of 0.29, the proposed model (Model 6) provided a recall value of 0.47, resulting in a 60% increase in TKA case prediction in Scenario I.

## Discussion

In the current study, we explored a model based on the fractal analysis of radiographic TBT to identify a population at higher risk of TKA, even at less advanced stages of OA. Our prediction model showed that a combination of radiography-based KL scoring and TBTA provided a high level of performance as reflected by either an AUC of 0.92 (Scenario I) or an AUC of 0.89 (Scenario II).

The associations between knee arthroplasty and radiographic or clinical features of KOA have already been published<sup>24,25,39</sup>. The association between trabecular bone texture and knee arthroplasty was firstly studied in a limited dataset ( $n = 114$  patients, 28 TKA cases)<sup>25</sup>. In the latter study, TBT parameters were evaluated using the Variance Orientation Transform (VOT) method. The mean fractal dimension (Texture parameter) analyzed in the medial subchondral tibial region was found to be lower in the TKA cases than in control patients. However, the study did not specify whether total or partial knee replacement, or both, was considered. Moreover, no prediction model was proposed. In the present study, we used a different TBT analysis method, (VAR), which was previously validated for the prediction of KOA progression<sup>27</sup>. Furthermore, we included not only the patients with severe OA ( $KL = 4$ , Scenario I), in agreement with the VOT-based study, but also the patients with non-severe OA ( $KL < 4$ , scenario II). When considering patients with ( $KL = 4$ ) at baseline, the number of cases (109) was dramatically higher than the number of controls (2), contrarily to the other situations (patients with  $KL < 4$  at baseline), as shown in Table 1. As expected, including patients with  $KL = 4$  at baseline (Scenario I) in the training procedure provided a better prediction performance with an AUC of 0.92 than that of Scenario II (excluding patients with  $KL < 4$  at baseline) with an AUC of 0.89.

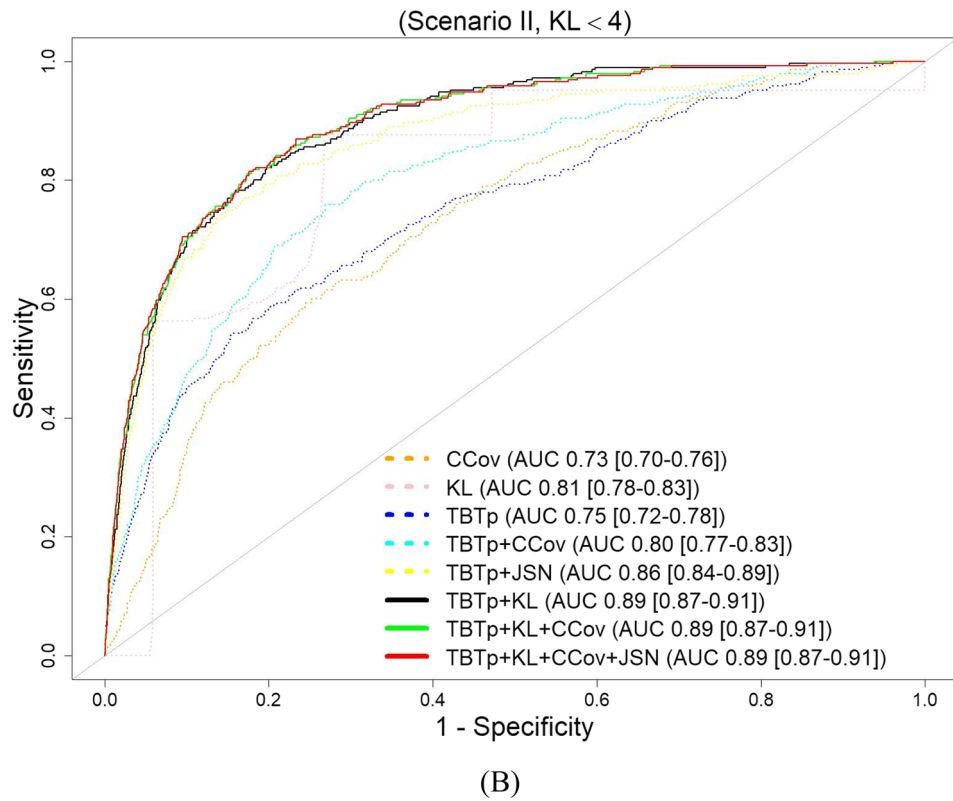
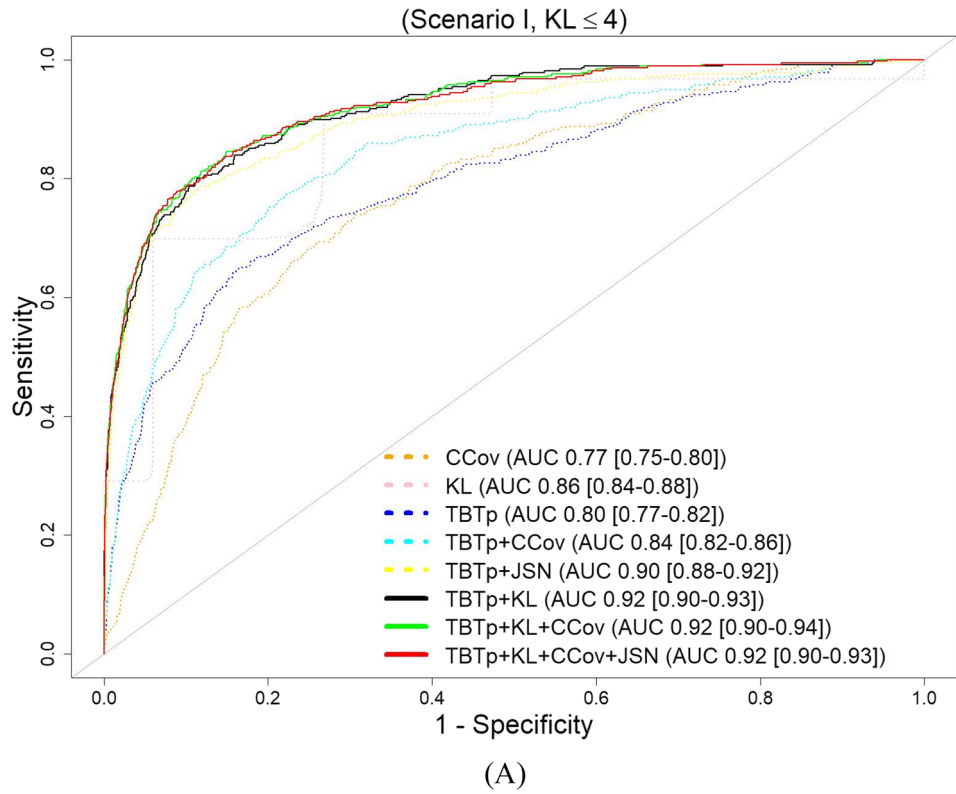
Recently, a deep learning (DL) based approach was used for the prediction of TKA risk<sup>24</sup> from the OAI cohort. In this study, the TKA outcome was defined, in line with our work, as whether a subject had undergone a TKA within 9 years or not. This binary definition of TKA does not reflect the TKA incident time. Considering time-dependent outcomes would involve the use of multinomial logistic regression instead of the binary logistic regression used in our study and it could be the subject of future investigation.

Several points, however, distinguish the present study from the previously reported one<sup>24</sup>. First, in the latter, radiographs were analyzed using a multitask DL approach, while in our study, a fractal-based TBT approach was used. The TBT analysis method, VAR, used in our study, has been shown to be invariant to both the quality and type of the images obtained in routine radiography<sup>20</sup>, whereas in DL-based methods, more attention should be paid to the variation in image quality during the training part<sup>24</sup>. Second, our case-control selection criteria did not impose the one-to-one case-control matching employed in their study<sup>24</sup>. Finally, in the current study, results related to the OARSI grades were omitted since better results of prediction were obtained with KL grades. Table 3 summarizes the main differences and similarities between the present study and the DL-based study<sup>24</sup>.

The AUC in Scenario II was 0.89. Although significantly different from 0.92 in Scenario I ( $p = 0.022$ ), they are both higher compared to the AUC of 0.87 (95% CI 0.85, 0.90) found in a DL-based prediction model<sup>24</sup>, although a direct comparison is not applicable due to the different knee-selection criteria used in both methods. The presented results highlight the potential of our prediction model for patients even with non-severe osteoarthritis ( $KL < 4$ ).

Based on a dataset from the OAI cohort<sup>18</sup>, it was shown that both the baseline values of TBT parameters and 12–18 month changes in MRI subchondral bone texture score were significantly associated with radiographic progression at 36 months. Combining such MRI-based parameters with 2D radiographic TBT-based parameters might help to better predict TKA.

In another study<sup>40</sup>, more than half of the knees with no/mild radiographic OA at baseline progressed to severe radiographic OA before knee replacement during 4 years of follow-up. Patellofemoral-femoral bone marrow lesions (BMLs) and worsening pain status were often associated with knees that had no/mild radiographic OA. MRI-defined joint inflammation defined as effusion- or Hoffa-synovitis, reflecting whole joint inflammation, developed over time in a majority of knees, indicating the presence of active disease at time points close to knee replacement. It would also be interesting to evaluate the prediction capacity of a model that combines DL-based and TBT-based descriptors. Such an idea has already been investigated in the context of the prediction of KOA progression and encouraging results were found<sup>20</sup>. Several DL approaches have recently been proposed for the



**Figure 5.** ROC curves obtained by the different TKA prediction models, using tenfold cross-validation: **(A)** using the dataset of Scenario I ( $0 \leq KL \leq 4$  at baseline), **(B)** using the dataset of Scenario II ( $0 \leq KL < 4$  at baseline).

Metrics	Model 1	Model 2	Model 3	Model 6	Model 7	Model 8
<b>Scenario I</b>						
Recall	0.04	0.29	0.15	0.47	0.47	0.46
Precision	0.38	0.98	0.60	0.75	0.75	0.75
Balanced accuracy	0.52	0.65	0.57	0.73	0.73	0.72
F1	0.07	0.45	0.24	0.58	0.58	0.57
AUC	0.77	0.86	0.80	0.92	0.92	0.92
<b>Scenario II</b>						
Recall	0.01	0.00	0.06	0.18	0.17	0.18
Precision	0.52	NA	0.56	0.55	0.54	0.53
Balanced accuracy	0.50	0.50	0.53	0.58	0.58	0.58
F1	0.01	NA	0.11	0.27	0.26	0.27
AUC	0.73	0.81	0.75	0.89	0.89	0.89

**Table 2.** Performance results of tested models in scenario I and scenario II. Where Model 1 included CCov parameters, Model 2 included KL parameters, Model 3 included TBTp parameters, Model 6 included TBTp and KL parameters, Model 7 included TBTp, KL, CCov and JSNM parameters, and Model 8 included TBTp, KL, CCov and JSNL parameters. NA refers to non-applicable values when sensitivity is zero.

	Leung et al. (2020) <sup>24</sup>	Our study	
Method	Deep learning: TL & MT approaches	Fractal texture analysis: VAR method	
Cohort	Public (OAI)	Public (OAI)	
Exclusion criteria	TKA at baseline PKA Incomplete radiographic data Not one-to-one case-control matching	TKA at baseline PKA Incomplete clinical and radiographic data	
Dataset selected	728 (364 cases)	Scenario I: 4382 (375 cases) Scenario II: 4296 (291 cases)	
Prediction models	KL OARSI DL-TL-MT	KL TBT TBTp-KL	
		<b>Scenario I</b>	<b>Scenario II</b>
KL inclusion	$0 \leq KL \leq 4$	$0 \leq KL \leq 4$	$0 \leq KL \leq 3$
AUC Results	KL = 0.74 OARSI = 0.75 DL-TL-MT = 0.87	KL = 0.86 TBT = 0.80 TBTp-KL = 0.92	KL = 0.81 TBT = 0.75 TBTp-KL = 0.89

**Table 3.** Prediction of radiographic TKA: comparison with a recent study. TBT Trabecular bone texture, VOT Variance Orientation Transform, VAR Variogram, TKA Total knee arthroplasty, PKA Partial knee arthroplasty, AUC Area under the ROC curve, ROC Receiver Operating Characteristic, DL Deep Learning, TL Transfer Learning, MT Multi-task, OARSI Osteoarthritis Research Society International, KL Kellgren-Lawrence.

prediction of TKA<sup>24,39,41</sup>. The association of such approaches with radiographic TBT analysis would certainly deserve to be investigated.

Changes in subchondral bone are constitutive of the OA disease process<sup>42</sup>. It has been demonstrated that the subchondral bone is intimately related to the structure and function of the cartilage<sup>43</sup>. Subchondral bone changes in its structure and thus in its biomechanical properties interfere with the interaction between the cartilage and the subchondral bone during mechanical loading<sup>44,45</sup>. Distribution of local stress in the subchondral bone is part of the parameters leading to the occurrence and development of the OA process as it has been demonstrated after changes in the joint's biomechanical environment following anterior cruciate ligament injury<sup>46</sup> or associated with abnormal knee alignment<sup>47</sup>. Subchondral bone structure and texture analyses can thus be considered as relevant tools to reflect bone remodeling changes induced by the osteoarthritis disease process<sup>19</sup>.

We have considered the TKA as a relevant outcome of KOA progression, as it has been proposed by the researchers interested in KOA related studies<sup>24,25,39,41,48,49</sup>. However, since the occurrence of TKA may vary by patient groups, within a country or different countries, it could be better in the next future to investigate another potential outcome measure as the "End Stage Knee Osteoarthritis" score proposed by Tufts Medical Center, and accepted to be included into the Center for Drug Evaluation and Research (CDER) Biomarker Qualification Program (BQP) in 2019<sup>50,51</sup>, which relies on a composite of radiographic disease severity, knee pain, knee function, knee mobility, and knee instability. Despite the inherent variability in the decision to undergo TKA, the development of a predictive model is a relevant task in the aim of aiming at improving our the current strategies to reduce the significant burden of TKA. As we lack in the absence of approved disease modifying therapy, TKA has been considered as a relevant proxy of the end stage of KOA and an important clinical outcome in KOA<sup>9</sup>,



even in a recently published work based on data from the OAI<sup>52</sup>. However, we also there is a lack of consensual medical indications for TKA, and with the arrival of potential new DMOADs, that target the structural course of OA disease, an earlier identification of KOA progressors will permit targeted interventions in high risk-high-risk groups.

The limitations of this study include the non-inclusion of several patient-reported outcomes such as stiffness, quality of life, and disability in the proposed model. Clinical decision-making in TKA is a complex process, thus considering undergoing a TKA within 9 years as a single clinical outcome measure can also be a limitation of the present study. Including decision criteria of TKA eligibility (i.e., radiographic severity<sup>53,54</sup>, functional limitations<sup>53</sup>, age<sup>54</sup>, pain<sup>53-55</sup>, history of arthroscopy<sup>55</sup>) would help to improve the implementation and performance of proposed TKA prediction models.

Our study has still some other limitations. Only clinical covariates and radiological scores collected at baseline were included in the proposed model. The variations in these readings during the subsequent follow-ups might improve the prediction performance of our proposed model. We investigated the TBT changes in the tibial subchondral bone of the knee. More studies including the femoral zone might also be conducted to investigate the potential additional information present in the femoral subchondral bone. The comparison with the work of Leung et al.<sup>24</sup> should ideally be conducted using the same set of radiographs. However, we did not have access to the reference of their selected dataset, thereby, leaving a more systematic comparison to future studies. The association between TKA and both 3D MRI bone texture<sup>56</sup> or shape<sup>57</sup> has recently been evaluated using data from the OAI<sup>41,56,57</sup>, in which features from the femoral region were also included, in addition to those from the tibial region.

The radiographs of selected patients were all acquired with a fixed flexion protocol. Therefore, our results may not applied to a study that uses a different imaging protocol. A further limitation of the dataset is that, despite the fairly large sample size, there are a very limited number of TKA cases. Only 100 such cases existed in the entire OAI for the time interval of 2 years. Consequently time-to-event experiments were not carried out in this study. Finally, since the occurrence of TKA may vary by patient groups, with a country or different countries, it might be better to investigate another potential outcome measure as the “End Stage Knee Osteoarthritis” (esKOA) Score proposed by Tufts Medical Center, and accepted by the FDA in 2019<sup>50,51</sup>, which relies on a composite of radiographic disease severity, knee pain, knee function, knee mobility, and knee instability.

Our study had several strengths. The data used in our study was obtained from a validated and well-structured research database (OAI cohort) that represents a large well-phenotyped US population. In this paper, we present the largest study, in terms of number of predicted events ( $n = 375$  in Scenario I and 291 in Scenario II) that investigated the role of TBT parameters as TKA risk predictors. Furthermore, a number of clinical predictors were included in our tested models such as history of knee injury, WOMAC pain, age, gender and BMI, in addition to radiographic disease stage (KL and two OARSI parameters: JSNM and JSNL). Another significant strength of our study is the performance analysis of several prediction models with strong validated performance metrics such as precision, recall (sensitivity), accuracy and F1 metrics<sup>35</sup>. In addition, technically, the ROIs used for our TBT analysis were segmented in a fully automated manner.

The clinical relevance of this study includes the feasibility of integrating the proposed model in a standard clinical routine. The used image processing tools are simple, fully automatic and based on radiography which is widely available and relatively not expensive. Furthermore, while the traditional KL model was capable to predict correctly less than 30% of total number of cases in Scenario I, and was completely incapable to predict any cases in Scenario II, the proposed KL + TBT model was capable to predict correctly 47% and 18% of total number of cases in Scenario I and Scenario II, respectively, Table 2. Suggesting radiographic TBT analysis may have a strong role in TKA risk screening, considering the lower cost and more readily implementable of radiography in primary care practice than MRI-based analysis.

Targeting high-risk individuals to prevent the progression of structural KOA might be possible with the advent of new algorithms based on multiple clinical, biologic and imaging biomarkers such as radiographic-based TBT parameters and MRI-derived features.

Our proposed prediction model performed well in the OAI cohort, in which a large number of participants were included. This type of modeling is often criticized for its inability to reproduce the same effectiveness in another sample. Although not with the same clinical endpoint (progression of the KOA disease), we demonstrated that the TBT-based models performed well not only when trained and tested on the same cohort, but also when trained on one cohort (OAI or MOST) and tested on the other one (MOST or OAI)<sup>20</sup>. As previously mentioned, the number of TBT parameters was optimized using the AIC.

## Conclusions

The model presented in this study, involving TBT analysis of conventional knee radiographs associated with radiological scoring (KL scores), improves the long-term prediction of TKA risk in patients with KOA.

The clinical decision-making process in TKA is complex<sup>53</sup>. Early detection and assessment of KOA prognostic factors are crucial for developing management and treatments that aim at preventing irreversible damage to the knee joint leading to arthroplasty. Identifying the most at-risk patients who may undergo total knee arthroplasty (TKA) could be the basis of a more aggressive therapeutic approach to prevent KOA progression.

Received: 13 October 2021; Accepted: 27 April 2022

Published online: 18 May 2022

## References

- Hunter, D. J. & Bierma-Zeinstra, S. Osteoarthritis. *The Lancet* **393**, 1745–1759 (2019).
- Musumeci, G. *et al.* Osteoarthritis in the XXIst century: Risk factors and behaviours that influence disease onset and progression. *Int. J. Mol. Sci.* **16**, 6093–6112 (2015).
- GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1545–1602 (2016).
- Hafezi-Nejad, N., Demehri, S., Guermazi, A. & Carrino, J. A. Osteoarthritis year in review 2017: Updates on imaging advancements. *Osteoarthr. Cartil.* **26**, 341–349 (2018).
- Brandt, K. D., Dieppe, P. & Radin, E. Etiopathogenesis of osteoarthritis. *Med. Clin. N. Am.* **93**, 1–24 (2009).
- Latourte, A., Kloppenburg, M. & Richette, P. Emerging pharmaceutical therapies for osteoarthritis. *Nat. Rev. Rheumatol.* <https://doi.org/10.1038/s41584-020-00518-6> (2020).
- Kolasinski, S. L. *et al.* 2019 American college of rheumatology/arthritis foundation guideline for the management of osteoarthritis of the hand, hip, and knee. *Arthritis Care Res.* **72**, 149–162 (2020).
- Bannuru, R. R. *et al.* OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. *Osteoarthr. Cartil.* **27**, 1578–1589 (2019).
- Eckstein, F., Wirth, W. & Nevitt, M. C. Recent advances in osteoarthritis imaging: The Osteoarthritis Initiative. *Nat. Rev. Rheumatol.* **8**, 622–630 (2012).
- Bowes, M. A. *et al.* Machine-learning, MRI bone shape and important clinical outcomes in osteoarthritis: Data from the Osteoarthritis Initiative. *Ann. Rheum. Dis.* <https://doi.org/10.1136/annrheumdis-2020-217160> (2020).
- Huynh, C. *et al.* Factors associated with the orthopaedic surgeon's decision to recommend total joint replacement in hip and knee osteoarthritis: An international cross-sectional study of 1905 patients. *Osteoarthr. Cartil.* **26**, 1311–1318 (2018).
- Felson, D. T. Osteoarthritis as a disease of mechanics. *Osteoarthr. Cartil.* **21**, 10–15 (2013).
- Janvier, T. *et al.* Subchondral tibial bone texture analysis predicts knee osteoarthritis progression: data from the osteoarthritis initiative: Tibial bone texture & knee OA progression. *Osteoarthr. Cartil.* **25**, 259–266 (2017).
- Kraus, V. B. *et al.* Predictive Validity of Radiographic Trabecular Bone Texture in Knee Osteoarthritis: The Osteoarthritis Research Society International/Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. *Arthritis & Rheumatology (Hoboken, N.J.)* **70**, 80–87 (2018).
- Kraus, V. B. *et al.* Subchondral bone trabecular integrity predicts and changes concurrently with radiographic and magnetic resonance imaging-determined knee osteoarthritis progression. *Arthritis Rheum.* **65**, 1812–1821 (2013).
- Woloszynski, T. *et al.* Prediction of progression of radiographic knee osteoarthritis using tibial trabecular bone texture. *Arthritis Rheum.* **64**, 688–695 (2012).
- Woloszynski, T., Podsiadlo, P., Stachowiak, G. W. & Kurzynski, M. A dissimilarity-based multiple classifier system for trabecular bone texture in detection and prediction of progression of knee osteoarthritis. *Proc. Inst. Mech. Eng. H* **226**, 887–894 (2012).
- MacKay, J. W. *et al.* Association of subchondral bone texture on magnetic resonance imaging with radiographic knee osteoarthritis progression: Data from the osteoarthritis initiative bone ancillary study. *Eur. Radiol.* **28**, 4687–4695 (2018).
- Almhdie-Imjabbar, A. *et al.* Trabecular bone texture analysis of conventional radiographs in the assessment of knee osteoarthritis: Review and viewpoint. *Arthritis Res. Ther.* **23**, 208 (2021).
- Nguyen, K. L. *et al.* Interest of texture analysis and neural networks for the characterization of knee osteoarthritis radiographic progression in OAI and MOST cohorts. *Bone Rep.* **13**, 100696 (2020).
- Hunter, D. J. *et al.* OARSI clinical trials recommendations: Knee imaging in clinical trials in osteoarthritis. *Osteoarthr. Cartil.* **23**, 698–715 (2015).
- LaValley, M. P. *et al.* Development of a clinical prediction algorithm for knee osteoarthritis structural progression in a cohort study: Value of adding measurement of subchondral bone density. *Arthritis Res. Ther.* **19**, 95 (2017).
- Ratzlaff, C. *et al.* A quantitative metric for knee osteoarthritis: Reference values of joint space loss. *Osteoarthr. Cartil.* **26**, 1215–1224 (2018).
- Leung, K. *et al.* Prediction of total knee replacement and diagnosis of osteoarthritis by using deep learning on knee radiographs: Data from the osteoarthritis initiative. *Radiology* **296**, 584–593 (2020).
- Podsiadlo, P., Cicuttini, F. M., Wolski, M., Stachowiak, G. W. & Wluka, A. E. Trabecular bone texture detected by plain radiography is associated with an increased risk of knee replacement in patients with osteoarthritis: A 6 year prospective follow up study. *Osteoarthr. Cartil.* **22**, 71–75 (2014).
- Nguyen, H. H., Saarakkala, S. & Tiulpin, A. Automatic knee osteoarthritis severity assessment from plain radiographs with limited data. *Osteoarthr. Cartil.* **28**, S314–S315 (2020).
- Janvier, T., Jennane, R., Toumi, H. & Lespessailles, E. Subchondral tibial bone texture predicts the incidence of radiographic knee osteoarthritis: Data from the osteoarthritis initiative. *Osteoarthr. Cartil.* **25**, 2047–2054 (2017).
- Lindner, C., Thiagarajah, S., Wilkinson, J. M., Wallis, G. A. & Coates, T. F. Accurate bone segmentation in 2D radiographs using fully automatic shape model matching based on regression-voting. In *Medical Image Computing and Computer-Assisted Intervention: MICCAI 2013* (eds Mori, K. *et al.*) 181–189 (Springer, 2013).
- Lindner, C. *et al.* Fully automatic segmentation of the proximal femur using random forest regression voting. *IEEE Trans. Med. Imaging* **32**, 1462–1472 (2013).
- Tiulpin, A., Melekhov, I. & Saarakkala, S. KNEEL: Knee anatomical landmark localization using hourglass networks. *IEEE Trans. Med. Imaging* <https://doi.org/10.1109/ICCVW.2019.00046> (2019).
- Krishnakumar Raja, V. B. *et al.* A-new-innovative-software-to-automatically-outline-condyles-in-orthopantomography. *Int. J. Med. Sci. Innov. Res. (IJMSIR)* **4**, 123–134 (2019).
- Messent, E. A., Buckland-Wright, J. C. & Blake, G. M. Fractal analysis of trabecular bone in knee osteoarthritis (OA) is a more sensitive marker of disease status than bone mineral density (BMD). *Calcif. Tissue Int.* **76**, 419–425 (2005).
- Bousson, V. *et al.* Trabecular bone score (TBS): Available knowledge, clinical relevance, and future prospects. *Osteoporos Int.* **23**, 1489–1501 (2012).
- Tiulpin, A. *et al.* Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. *Sci. Rep.* **9**, 20038 (2019).
- Widera, P. *et al.* Multi-classifier prediction of knee osteoarthritis progression from incomplete imbalanced longitudinal data. *Sci. Rep.* **10**, 8427 (2020).
- R Core Team. *R: The R Project for Statistical Computing*. (R Foundation for Statistical Computing, 2020) <https://www.r-project.org/>.
- Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
- DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**, 837–845 (1988).
- Heisinger, S., Hitzl, W., Hobusch, G. M., Windhager, R. & Cotofana, S. Predicting total knee replacement from symptomology and radiographic structural change using artificial neural networks—data from the osteoarthritis initiative (OAI). *J. Clin. Med.* **9**, 1–10 (2020).

40. Roemer, F. W. *et al.* From early radiographic knee osteoarthritis to joint arthroplasty: Determinants of structural progression and symptoms. *Arthritis Care Res.* **70**, 1778–1786 (2018).
41. Tolpadi, A. A., Lee, J. J., Pedoia, V. & Majumdar, S. Deep learning predicts total knee replacement from magnetic resonance images. *Sci. Rep.* **10**, 3395 (2020).
42. Loeser, R. F., Goldring, S. R., Scanzello, C. R. & Goldring, M. B. Osteoarthritis: A disease of the joint as an organ. *Arthritis Rheum.* **64**, 1697–1707 (2012).
43. Yuan, X. L. *et al.* Bone-cartilage interface crosstalk in osteoarthritis: Potential pathways and future therapeutic strategies. *Osteoarthr. Cartil.* **22**, 1077–1089 (2014).
44. Goldring, S. R. & Goldring, M. B. Changes in the osteochondral unit during osteoarthritis: Structure, function and cartilage-bone crosstalk. *Nat. Rev. Rheumatol.* **12**, 632–644 (2016).
45. Watkins, L. *et al.* Assessment of quantitative [<sup>18</sup>F]Sodium fluoride PET measures of knee subchondral bone perfusion and mineralization in osteoarthritic and healthy subjects. *Osteoarthr. Cartil.* **29**, 849–858 (2021).
46. Bhatla, J. L., Kroker, A., Manske, S. L., Emery, C. A. & Boyd, S. K. Differences in subchondral bone plate and cartilage thickness between women with anterior cruciate ligament reconstructions and uninjured controls. *Osteoarthr. Cartil.* **26**, 929–939 (2018).
47. Han, X. *et al.* Abnormal subchondral trabecular bone remodeling in knee osteoarthritis under the influence of knee alignment. *Osteoarthr. Cartil.* **30**, 100–109 (2022).
48. Wang, T., Leung, K., Cho, K., Chang, G. & Deniz, C. M. *Total Knee Replacement prediction using Structural MRIs and 3D Convolutional Neural Networks.* (2019).
49. Pelletier, J.-P. *et al.* What is the predictive value of MRI for the occurrence of knee replacement surgery in knee osteoarthritis?. *Ann. Rheum. Dis.* **72**, 1594–1604 (2013).
50. McAlindon, T. *Biomarker Qualification Letter of Intent (LOI) Template by Tufts Medical Center.* (Food and Drug Administration (FDA), 2019). <https://www.fda.gov/media/135989/download>.
51. Leptak, C. & Seymour, S. *Letter of Intent: Determination Letter.* (Food and Drug Administration (FDA), 2019). <https://www.fda.gov/media/135988/download>.
52. Brown, M. J. C., Vella-Baldacchino, M., O’Flaherty, E. & Jenkins, P. J. Predicting patient reported outcome in total knee arthroplasty using body mass index and limb measurements. *Knee* **25**, 915–922 (2018).
53. Skou, S. T. *et al.* Criteria used when deciding on eligibility for total knee arthroplasty: Between thinking and doing. *Knee* **23**, 300–305 (2016).
54. Verra, W. C. *et al.* The reason why orthopaedic surgeons perform total knee replacement: Results of a randomised study using case vignettes. *Knee Surg Sports Traumatol. Arthrosc* **24**, 2697–2703 (2016).
55. Zikria, B., HafeziNejad, N., Wilckens, J., Ficke, J. & Demehri, S. Determinants of knee replacement in subjects with a history of arthroscopy: Data from the osteoarthritis initiative. *Eur. J. Orthop. Surg. Traumatol.* **26**, 1–10 (2016).
56. Heilmeyer, U. *et al.* Baseline knee joint effusion and medial femoral bone marrow edema, in addition to MRI-based T2 relaxation time and texture measurements of knee cartilage, can help predict incident total knee arthroplasty 4–7 years later: Data from the osteoarthritis initiative. *Skeletal Radiol.* **48**, 89–101 (2019).
57. Barr, A. J. *et al.* The relationship between three-dimensional knee MRI bone shape and total knee replacement: A case control study: data from the osteoarthritis initiative. *Rheumatology* **55**, 1585–1593 (2016).

## Acknowledgements

All authors would like to thank the participants and staff of the OAI study. They gratefully acknowledge the European Regional Development Fund (ERDF) for financial support. All authors also wish to thank Mrs. Elizabeth Rowley-Jolivet for proofreading the article.

## Author contributions

A.A.I. and E.L. wrote the first draft of the manuscript. All authors contributed to the conception and design of the study. Statistical analyses were conducted by A.A.I. All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. E.L. takes responsibility for the integrity of the work as a whole, from inception to the finished article.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-12083-x>.

**Correspondence** and requests for materials should be addressed to E.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022