

Charting Aging Trajectories of Knee Cartilage Thickness for Early Osteoarthritis Risk Prediction: An MRI Study from the Osteoarthritis Initiative Cohort

Tengfei Li^{1,2†}, Tianyou Luo^{3†}, Boqi Chen⁴, Chao Huang⁵, Zhengyang Shen⁴, Zhenlin Xu⁴, Daniel Nissman¹, Yvonne M. Golightly^{6,7,8}, Amanda E. Nelson^{6,7}, Marc Niethammer^{2,4*}, and Hongtu Zhu^{1,2,3,4*}

¹Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

²Biomedical Research Imaging Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

³Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

⁴Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

⁵Department of Statistics, Florida State University, Tallahassee, FL, USA

⁶Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

⁷Thurston Arthritis Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

⁸College of Allied Health Professions, University of Nebraska Medical Center, Omaha, NE, USA

† *These authors contributed equally*

* *Corresponding authors*

mn@cs.unc.edu

htzhu@email.unc.edu

Abstract

Knee osteoarthritis (OA), a prevalent joint disease in the U.S., poses challenges in terms of predicting of its early progression. Although high-resolution knee magnetic resonance imaging (MRI) facilitates more precise OA diagnosis, the heterogeneous and multifactorial aspects of OA pathology remain significant obstacles for prognosis. MRI-based scoring systems, while standardizing OA assessment, are both time-consuming and labor-intensive. Current AI technologies facilitate knee OA risk scoring and progression prediction, but these often focus on the symptomatic phase of OA, bypassing initial-stage OA prediction. Moreover, their reliance on complex algorithms can hinder clinical interpretation. To this end, we make this effort to construct a computationally efficient, easily-interpretable, and state-of-the-art approach aiding in the radiographic OA (rOA) auto-classification and prediction of the incidence and progression, by contrasting an individual's cartilage thickness with a similar demographic in the rOA-free cohort. To better visualize, we have developed the toolset for both prediction and local visualization. A movie demonstrating different subtypes of dynamic changes in local centile scores during rOA progression is available at <https://tli3.github.io/KneeOA/>. Specifically, we constructed age-BMI-dependent reference charts for knee OA cartilage thickness, based on MRI scans from 957 radiographic OA (rOA)-free individuals from the Osteoarthritis Initiative cohort. Then we extracted local and global centiles by contrasting an individual's cartilage thickness to the rOA-free cohort with a similar age and BMI. Using traditional boosting approaches with our centile-based features, we obtain rOA classification of KLG ≤ 1 versus KLG = 2 (AUC = 0.95, F1 = 0.89), KLG ≤ 1 versus KLG ≥ 2 (AUC = 0.90, F1 = 0.82) and prediction of KLG2 progression (AUC = 0.98, F1 = 0.94), rOA incidence (KLG increasing from < 2 to ≥ 2 ; AUC = 0.81, F1 = 0.69) and rOA initial transition (KLG from 0 to 1; AUC = 0.64, F1 = 0.65) within a future 48-month period. Such performance in classifying KLG ≥ 2 matches that of deep learning methods in recent literature. Furthermore, its clinical interpretation suggests that cartilage changes, such as thickening in lateral femoral and anterior femoral regions and thinning in lateral tibial regions, may serve as indicators for prediction of rOA incidence and early progression. Meanwhile, cartilage thickening in the posterior medial and posterior lateral femoral regions, coupled with a reduction in the central medial femoral region, may signify initial phases of rOA transition.

Keywords: Knee osteoarthritis, KLG, knee MRI, local cartilage thickness, aging trajectory, early detection, prognosis

Introduction

Knee osteoarthritis (OA), a common chronic disease characterized in part by cartilage degeneration, inflammation, and pain (1, 2), presents a substantial global health and economic burden. Its global prevalence reached 364.6 million in 2019 (3), with 86.7 million new adult cases reported in 2020 (4). Nearly 30% of individuals above 45 demonstrate signs of knee radiographic OA (rOA), with around half reporting symptoms, and significantly more severe among women (5-7). Other health implications of knee OA include frequent comorbidities (8). Economically, OA and related disorders entail substantial total costs of approximately \$486.4 billion, with \$136.8 billion directly attributable to OA (9), and a lifetime cost of \$15,000 per patient (6, 10). Given the aging population and rising obesity rates, these statistics are likely to be further amplified in the future, underscoring the importance of knee OA management.

While knee OA is prevalent, its complex etiology involving various joint structures like cartilage, bone, synovium, ligaments, periarticular fat, meniscus, muscle, biomechanical influences, immune response variations (6, 11, 12), and their intricate interplay (13), complicates the prognosis and treatment of knee OA. Existing management strategies range from patient education (14), exercise (15), and weight management (16), to pharmacological interventions using nonsteroidal anti-inflammatories (17) and pain alleviation drugs (18). However, in advanced cases, total knee replacement (TKR) surgeries, although invasive and costly, become necessary. Timely diagnosis can decelerate OA progression, mitigate the need for TKR and curtail long-term medication use and associated adverse events, underscoring the need for effective early interventions for knee OA.

Diagnosing early OA, when typical signs and symptoms of advanced OA are not apparent, is challenging. Traditional radiographs primarily detect changes in bone structure, such as joint space narrowing or osteophyte formation, common in later stages (19). In early OA, tissue-related changes such as cartilage or synovial damage, which may not be visible on radiographs, can occur (20, 21). On the other hand, magnetic resonance imaging (MRI), which visualizes all joint tissues, can detect a diverse array of pathological changes, such as early cartilage damage, bone marrow lesions, synovitis, or changes in the menisci, ligaments, or tendons, that cannot be captured by radiographs (22). Through various modalities, OA can be assessed and classified by radiographic OA (rOA), MRI-identified OA (MOA) and histopathological OA (hOA) (20). The widely accepted Kellgren-Lawrence Grading (KLG) (23) system has long served as the rOA standard across clinical practice and research. This system classifies OA severity on a scale from 0 (rOA-free) to 4 (severe rOA) using radiographs. Despite its efficacy in detecting moderate (grade 3) to severe rOA, the KLG system's sensitivity is limited in accurately identifying early rOA stages, such as the grade 2 (24), and grade 1 (25). Enhancements in early OA detection have been driven by the development of MRI scoring systems such as the Magnetic Resonance Imaging Osteoarthritis Knee Score (MOAKS) (19), the Knee Osteoarthritis Scoring System (KOSS) (26), the Boston Leeds Osteoarthritis Knee Score (BLOKS) (27), and the Whole-Organ Magnetic Resonance Imaging Score (WORMS) (28). These systems capture early tissue changes more sensitively and are crucial for defining MOA. For instance, the European Society of Sports Traumatology, Knee Surgery and Arthroscopy (ESSKA) describes (29) early MOA using WORMS grades 3–6 for cartilage morphology Scores, 2-3 for bone marrow lesions, ≥ 2 for cartilage regional loss, and ≥ 3 for meniscal tears. From a histopathological perspective, early OA is determined (20) when $\leq 50\%$ of the cartilage thickness is involved, as per the macroscopic International Cartilage Repair Society (ICRS) classification. Considering these standards, early knee OA is classified (29) based on, a KLG of 0, 1, or 2 in conjunction with knee pain, with additional evidence such as arthroscopic findings of cartilage lesions, MRI evidence of articular cartilage or meniscal degeneration, or subchondral bone marrow lesions.

Over recent decades, a plethora of literature has been dedicated to extracting MRI-based scores or features to predict the future incidence of rOA. This includes the use of MOAKS (30, 31) and WORMS (32), and MRI bone shape features (33), among others, to predict incident rOA 12 months to 48 months prior. These studies propose that MRI changes in bone can presage the early development of rOA and knee pain. However, these methods have notable limitations. Foremost, they are both labor-intensive and time-consuming, necessitating a considerable investment of time and requiring significant clinical expertise for manual scoring or landmark labeling. Moreover, existing MRI scoring systems primarily focus on cartilage loss and lesions, while potentially overlooking the early stage swelling or thickening of the cartilage. It has been suggested (34) that the pathogenesis of OA might not be unidirectional, involving solely cartilage loss. On the contrary, some studies

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/) .

have reported that cartilage can also thicken in the early stages of OA, a phenomenon observed in both animal studies (35-39) and humans (34, 40-44) with early OA. Cartilage thickening in early OA stages, possibly resulting from water diffusion, inflammation, or an early reparative response (34), is not yet fully understood. These findings highlight the need for refinement and improvement of these scoring systems for better early detection of rOA.

On the other hand, despite the advancements in machine learning (ML) and deep learning (DL) for automating OA classification and prediction, significant challenges in this area remain. These technologies have been deployed for a broad range of applications (45, 46), including OA classification (47-57) and prediction (58-65), joint space narrowing (JSN) prediction (66-68), pain symptom classification (69-71) and prediction (72), cartilage lesion detection (73) and TKR incidence prediction (74-79) using MRI or radiograph data. However, the complex inner workings of these artificial intelligence (AI) algorithms often remain opaque, leading to limited interpretability (45). Efforts towards improving the transparency of DL-based models persist (80, 81), yet it remains a major challenge. Additionally, the reliance on the KLG system to classify the severity of rOA presents another limitation. The KLG system, while widely accepted, may not be well suited to early-stage disease detection. These models often focus on detecting and predicting OA progression in its later, symptomatic phases, ignoring OA at the early stages. Models capable of detecting and predicting the earliest stages of OA are needed.

In this study, we aim to predict the early rOA progression, including the potential transition from normal to rOA initial onset and from the initial stage to mild stage. To accomplish this, we first utilized MRI data to establish reference charts for rOA-free knee thickness. Our training data for these reference charts comprised 5,045 scans from 957 participants without rOA (KLG = 0), aged between 45 and 79, who were enrolled in the publicly available Osteoarthritis Initiative (OAI) dataset between 2004 and 2006 (82) (<https://nda.nih.gov/oai/>). We extracted local cartilage thickness (LCT) maps from knee MRIs using automatic segmentation and registration techniques (83, 84). Next, we constructed the aging charts for LCT among the rOA-free participants, using a nonlinear generalized additive modeling framework with mixed effects (GAMM), accounting for confounding factors such as age, sex, race, baseline height, and body mass index (BMI). We then evaluated centile scores for each individual based on their ranking among peers with similar demographics who were free from rOA. Using these centile score features, we conducted three predictive analyses. (i) Generated a MRI-based rOA auto-classification system with high accuracy, in classifying the early rOA phases into KLG0-1, and KLG2 with and without progression in 48 months. (ii) Predicted the early rOA transition (from KLG = 1 to 2). (iii) Predicted the rOA initial transition (from KLG = 0 to 1). In particular, we may have the following contributions to the field.

- 1) We developed comprehensive knee cartilage thickness growth reference charts, considering complex nonlinear relationships among age, BMI, and their interactions. Although the effects of sex, age, and BMI have been widely studied in previous literature (6), their complex nonlinear relationships have rarely been investigated. By hypothesizing that distinct cartilage thickness growth curves exist across different knee-OA groups, this approach could distinguish rOA phases effectively.
- 2) We introduced an automated, easily interpretable and local MRI-based scoring system, based on rankings of cartilage thickness growth trajectories among rOA-free participants. This system could facilitate $KLG \leq 1$ versus $KLG = 2$ classification (AUC = 0.95, F1 = 0.89), $KLG \leq 1$ versus $KLG \geq 2$ classification (AUC = 0.90, F1 = 0.82) and prediction of KLG2 progression (AUC = 0.98, F1 = 0.94), incidence rOA transition (AUC = 0.81, F1 = 0.69) and rOA initial transition (AUC = 0.64, F1 = 0.65).
- 3) Our findings supported the notion that cartilage thickening is an important biomarker in early rOA transition (34). We reported key biomarkers for the rOA initial transition, incidence-rOA transition and rOA early progression, such as anterior femoral cartilage thickening and lateral tibial cartilage thinning during the progression from KLG2 to KLG3.

We outlined our workflow and analysis process in **Figure 1**. Our work may offer valuable insights into OA by establishing a standard reference for cartilage aging comparison, and the early detection of OA transition.

Results

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

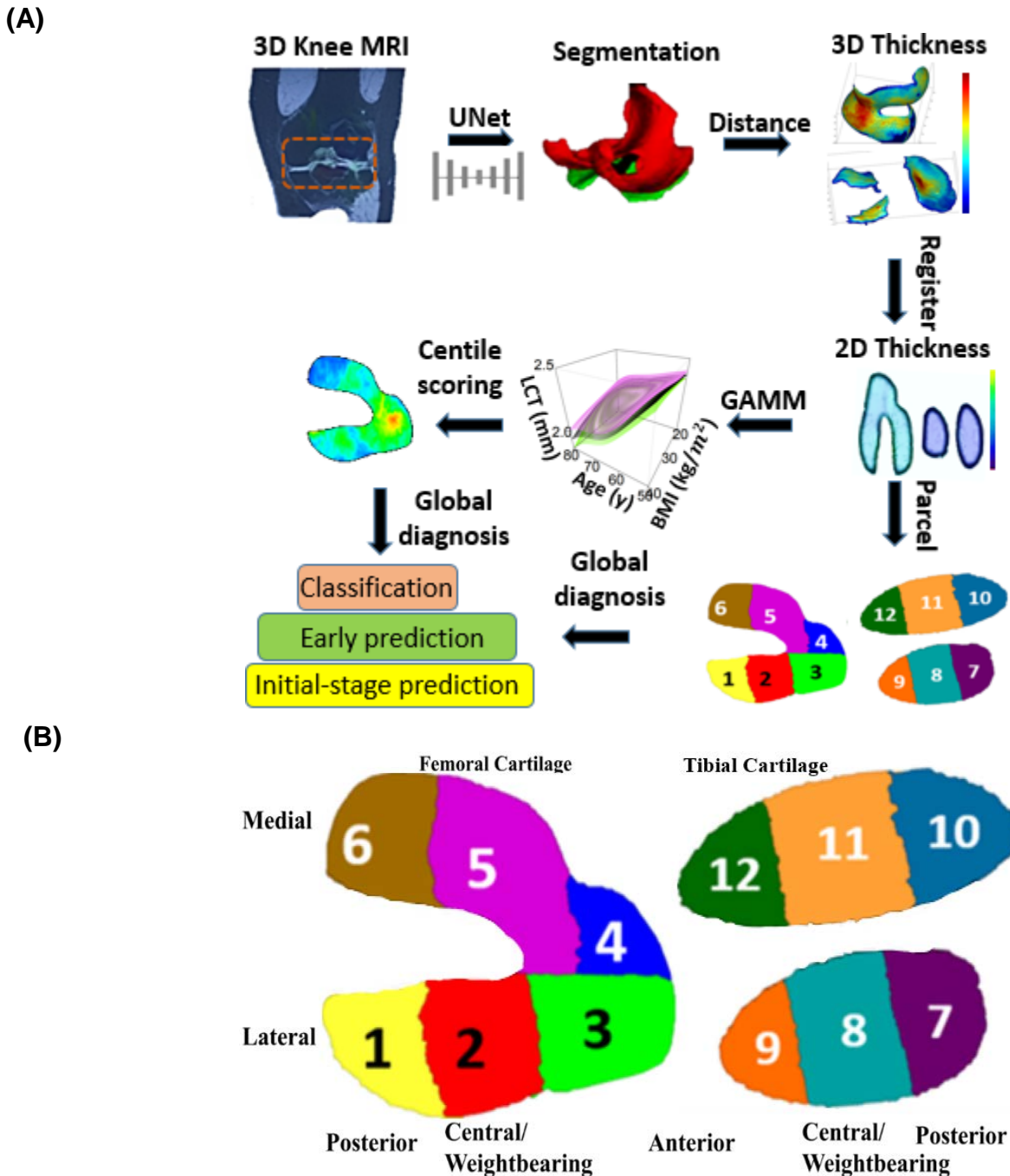


Fig. 1. Overview of the study design. (A) Overall workflow of the study. We took the 3D MRI from the OAI dataset as an input and generated the 2D (femoral and tibial) cartilage thickness map on the standard space. We fit aging trajectories at each spatial location and build the MRI centile scoring system. The centile scores were used to predict scans scans with future rOA progression. **(B)** Parcellations of the knee cartilage. Regions of interest (ROIs) 1-6 correspond to femoral cartilage; ROIs 7-12 correspond to tibial cartilage. 1: Lateral posterior femur (LPF); 2: Lateral central femur (LCF); 3: Lateral anterior femur (LAF); 4: Medial anterior femur (MAF); 5: Medial central femur (MCF); 6: Medial posterior femur (MPF); 7: Lateral posterior tibia (LPT); 8: Lateral central tibia (LCeT); 9: Lateral anterior tibia (LAT); 10: Medial posterior tibia (MPT); 11: Medial central tibia (MCT); 12: Medial anterior tibia (MAT).

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

Following prior literature (85-87), we extended specific terminology based on the KLG grades.

(1) Knees. Knees were classified into KLG0, KLG1, KLG2, KLG3 and KLG4 according to $KLG = 0, 1, 2, 3, 4$, respectively. Scans with $KLG \geq 2$ were designated as definite-rOA knees.

(2) Participants. We separated participants into two groups, the **KLG0-1-stable** ($KLG \leq 1$ across all visits within 96 months for both knees) and **rOA-prevalence** cases (KLG attains 2 within 96 months).

The KLG0-1-stable group included three subgroups,

- (i) the **rOA-free** ($KLG = 0$ across all visits for both knees),
- (ii) the **KLG1-stable** ($KLG = 1$ across all visits)
- (iii) the **KLG0-1-transition** (baseline $KLG = 0$ which transits to $KLG = 1$).

The rOA-prevalence group included two subgroups,

- (i) the **rOA-incidence** (baseline $KLG \leq 1$ which transits into $KLG \geq 2$), and
 - (ii) the **KLG2-4-progressive** ($KLG \geq 2$ across all visits),
- where the rOA-incidence subgroup includes the following subsets,
- (iii) the **KLG1-progressive** (baseline $KLG = 1$ and then to $KLG \geq 2$),
 - (iv) the **KLG0-progressive** (baseline $KLG = 0$ and then to $KLG \geq 2$),
- and both the rOA-incidence and KLG2-4-progressive subgroups included,
- (v) the **KLG2-3-transition** ($KLG = 2$ at one visit which transit into $KLG = 3$ in later phases),
 - (vi) the **KLG2-stable** ($KLG = 2$ at one visit and ≤ 2 until the last visit).

Classifications for other subpopulations were applied to left and right knees, separately. Furthermore, we referred to the rOA initial transition as the KLG transition from 0 to 1, and termed the incidence rOA transition as the KLG shift from less than 2 to 2. In **Table 1**, we offered descriptive statistics for 9 primary subpopulations, which included sample size, age, sex, BMI, race, and height distributions from the participants' initial visits.

Reference cartilage thickness charts highlight thickness growth patterns

We modeled the sex-specific trajectories of cartilage thickness with respect to age and BMI, adjusted for race and height (detailed methods are available in the *Methods* Section). **Figure 2A** provides reference charts depicting the non-linear interplay among the BMI, age, and average cartilage thickness in the tibial and femoral regions of the right knee, stratified by sex. These charts elucidate complex nonlinear associations through 3D perspective and 2D contour plots, with the corresponding confidence interval surfaces. Detailed reference charts for 12 regions of interest (ROIs) in both left and right knees are presented in the **Supplementary Figures S1** and **S2**. **Figure 2B** presents the effect size and statistical significance of sex, race, height, BMI, age-sex interaction, and sex-BMI interaction across all the regions.

In this analysis, we observed that cartilage thickness in all ROIs decreases significantly with age, with males typically possessing larger baseline thickness in most ROIs (except right posterior medial tibial region). While an increase in BMI negatively affected cartilage thickness in a majority of knee regions in males, the negative influence in females was restricted to the left and right posterior tibial and posterior lateral femoral regions, and right anterior and central medial femoral regions. Notably, all regions affected by BMI demonstrated a consistent negative age-BMI interaction, suggesting an increased negative effect of BMI on cartilage thickness as individuals age. The cartilage thickness reduction affected by both age and BMI was more evident in men compared to women among rOA-free participants. Regarding other factors, race typically negatively affected all ROIs with the exception of the anterior femoral region, and height typically exhibits positive effects, except for the anterior medial femoral, posterior lateral tibial, and left anterior medial tibial regions. In addition, femoral regions generally exhibited greater cartilage thickness than tibial ones.

Our results generally demonstrated bilateral symmetry between left and right knees, and confirmed a negative association between age and knee cartilage thickness at all sites (88). Previous research yielded inconsistent conclusions regarding age-sex interactions and cartilage loss. Cai, et al. (2019) (89) reported increased age-related lateral tibial cartilage loss in females, while Si, et al. (2021) (90) observed faster thinning in the lateral anterior and medial anterior femur in males. Our findings aligned more with Si, et al. (2021), likely due to our focus on rOA-free participants, in contrast to the higher proportion of OA-symptomatic participants in Cai, et al. (2019). Our data supported the notion of larger cartilage thickness in males (90-92). This might explain the lower OA prevalence rate in males (93), as thicker cartilage might offer better protection against shear stress (94).

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

Table 1. Descriptive statistics for 9 subpopulations in OAI. Those includes sample size of participants, number of visits, sample size for females and whites, and BMI, height and age at the baseline visits, of the KLG0-1-stable (KLG ≤ 1 across all visits), rOA-prevalence (all KLG >1 cases), rOA-free (KLG = 0 across all visits), rOA-incidence, KLG0-1-transition, KLG1-progressive, KLG2-3 transition and KLG2-stable participants. Statistics are based on the right knees.

	KLG0-1-stable		rOA-prevalence	
	Female	Male	Female	Male
Number of Participants (visits)	769 (4131)	593 (3339)	1252 (5602)	792 (3716)
Sex %	56.5%	43.5%	61.3%	38.7%
Race for white (%)	665 (86.5%)	525 (88.5%)	917 (73.2%)	670 (84.6%)
Baseline BMI \pm sd (kg/m ²)	26.2 \pm 4.6	27.9 \pm 4.1	29.9 \pm 5.3	29.7 \pm 4.1
Baseline height \pm sd (cm)	162.2 \pm 5.9	176.2 \pm 6.4	162.8 \pm 6.2	176.3 \pm 6.5
Baseline age \pm sd (year old)	59.8 \pm 9.1	58.5 \pm 9.2	62.5 \pm 8.8	62.4 \pm 9.2
	rOA-free		rOA-incidence	
	Female	Male	Female	Male
Number of Participants (visits)	542 (2799)	415 (2246)	134 (763)	79 (451)
Sex %	56.6%	43.4%	62.9%	37.1%
Race for white (%)	467 (86.2%)	361(87.0%)	124 (80.5%)	82 (92.1%)
Baseline BMI \pmsd (kg/m ²)	25.8 \pm 4.5	27.7 \pm 4.1	29.6 \pm 5.2	29.4 \pm 3.9
Baseline height \pmsd (cm)	162.3 \pm 5.9	176.3 \pm 6.5	163.1 \pm 6.6	176.5 \pm 7.1
Baseline age \pmsd (year old)	59.3 \pm 9.2	57.4 \pm 8.9	60.3 \pm 8.8	61.4 \pm 8.7
	KLG0-1-transition		KLG1-progressive	
	Female	Male	Female	Male
Number of Participants (visits)	27 (155)	19 (99)	94 (537)	49 (271)
Sex %	58.7%	41.3%	65.7%	34.3%
Race for white (%)	25 (92.6%)	17 (89.5%)	77 (81.9%)	46 (93.9%)
Baseline BMI \pmsd (kg/m ²)	24.7 \pm 3.7	28.5 \pm 5.8	29.9 \pm 4.5	29.7 \pm 4.2
Baseline height \pmsd (cm)	163.6 \pm 6.9	176.1 \pm 5.6	163.2 \pm 6.8	177.5 \pm 7.5
Baseline age \pmsd (year old)	58.2 \pm 7.7	53.7 \pm 7.7	60.6 \pm 8.9	60.7 \pm 8.6
	KLG2-4-progressive		KLG2-3-transition	
	Female	Male	Female	Male
Number of Participants (visits)	1118 (5313)	713 (3546)	126 (571)	53 (251)
Sex %	61.1%	39.9%	70.4%	29.6%
Race for white (%)	807 (72.2%)	597 (83.7%)	92 (73.0%)	44 (83.0%)
Baseline BMI \pmsd (kg/m ²)	30.0 \pm 5.4	29.7 \pm 4.1	30.7 \pm 5.1	29.7 \pm 4.3
Baseline height \pmsd (cm)	162.8 \pm 6.1	176.3 \pm 6.5	163.1 \pm 7.0	176.5 \pm 7.0
Baseline age \pmsd (year old)	62.9 \pm 8.8	62.4 \pm 9.3	61.8 \pm 7.9	60.9 \pm 9.3
	KLG2-stable			
	Female	Male		
Number of Participants (visits)	495 (2670)	315 (1719)		
Sex %	61.1%	38.9%		
Race for white (%)	368 (74.3%)	270 (85.7%)		
Baseline BMI \pmsd (kg/m ²)	29.3 \pm 5.2	29.4 \pm 3.9		
Baseline height \pmsd (cm)	163.1 \pm 5.8	176.0 \pm 6.4		
Baseline age \pmsd (year old)	61.2 \pm 8.7	61.0 \pm 9.4		

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

For the effects of BMI, though numerous studies have examined the impact of BMI on cartilage loss in both sexes among OA patients (89), findings on the effects of BMI on cartilage thickness in healthy individuals are inconsistent. Some studies suggested an association between increased BMI and thicker cartilage (95, 96), while Keng, et al. (2017) (97) reported the opposite. Our investigation unveiled region-specific effects of BMI on cartilage thickness in rOA-free participants of both sexes. Moreover, the interaction between age and BMI remains largely unexplored. We found that age and BMI interact negatively in all femoral and tibial regions adversely affected by BMI. This might indicate an amplified detrimental impact of BMI on cartilage thickness with age. Such an effect could result from cumulative joint damage, increased load from higher body weight, as well as age-related metabolic and inflammatory changes. Additionally, we observed thicker knee cartilage among white and taller individuals, consistent with Connolly, et al. (2008) (98).

Reference cartilage thickness charts highlight rOA progression patterns

We examined cartilage thickness differences among six KLG groups of knees, adjusting for confounding covariates. These groups included KLG2-4, KLG0, and KLG1. The KLG1 was further divided based on progression status to KLG > 1 as "KLG1progress" and "KLG1Stable". The boxplots in **Figure 2C** and heatmap in **Figure 2D** display these differences across 28 regions. For the transition from KLG0 to KLG1, thickness increased in the posterior femoral (ROIs 1 and 6) and posterior medial tibial (ROI10) regions. Compared to the KLG1 stable group, the KLG1progress group showed additional increases in the posterior femoral, central lateral femoral (ROI2), and anterior medial tibial (ROI12) regions. In KLG1-2, we noted reductions in the posterior and central tibial (ROIs 7 and 8) and the central medial tibial and femoral regions. However, there was an increase in the left posterior lateral femoral thickness (ROI 1). For KLG2-4, cartilage thinning was observed in all ROIs except the left and right anterior femoral (ROIs 3 and 4) and left lateral femoral (ROIs 1 and 2). The most significant thinning occurred in the central medial tibial and femoral regions (ROIs 11 and 5). Besides, there is an increase in anterior lateral femoral region (ROI 3) from KLG2 to KLG3.

In summary, one of the earliest indications of progression to rOA was an increased thickness in the posterior femoral and posterior medial tibial regions, when transitioning from KLG = 0 to 1. As progression continued from KLG = 1 to 2, cartilage thinning occurs in posterior and central lateral tibial and central medial tibial regions with possible cartilage increase in the posterior lateral femoral region. During the transition from the KLG = 2 to 4, a significant decline in cartilage thickness across the majority of femoral and tibial regions was observed, while at the earlier transition (KLG = 2 to 3), cartilage thickening may occur in the anterior lateral femoral region. This progression trajectory outlined a population-level course of rOA progression.

In previous literature, OA pathogenesis may not strictly follow a linear path. Instead, a more complex picture, suggesting a bi-directional cartilage change during OA's early stages, has been proposed (34). Cotofana, et al. (2012) (43) reported knees with early rOA displayed thicker cartilage specifically in the external femoral subregions of compartments with marginal osteophytes; Calvo, et al. (2001) (36) reported a progressive articular cartilage thickness increase in the femur at weeks 4, 6 and 8 but only at week 6 at the tibial region; findings from Hellio Le Graverland, et al. (2009) (44) suggested that, on average, female participants with KLG2 and KLG3 showed thicker and thinner cartilage than healthy controls, respectively, corroborating our observed thickness rise from KLG1 to KLG2 and a subsequent decline from KLG2 to KLG3. However, few studies have systematically examined detailed regional thickness changes across different phases and regions based on larger cohorts.

While our study identified several distinct patterns among different rOA categories, their predictive power needed to be examined. In the subsequent sections, we further explored the predictability of these patterns, including predicting rOA classification, incidence rOA and the initial-stage rOA transition.

Reference cartilage thickness charts facilitate early rOA classification

Using 22 machine learning methods (more details in *Method* Section), we classified the knees in early-rOA phases (KLG ≤ 2) into three stages, including KLG ≤ 1, and KLG2 with and without progression into KLG3. This was carried out through two binary classifications, including classifying KLG ≤ 1 versus KLG = 2, and classifying KLG2 with versus without progression, whose performances were shown in **Figure 3** and **Figure 4**, respectively. The comparison of the area under the receiver operating characteristic (AUC), macro-F1 score and prediction

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

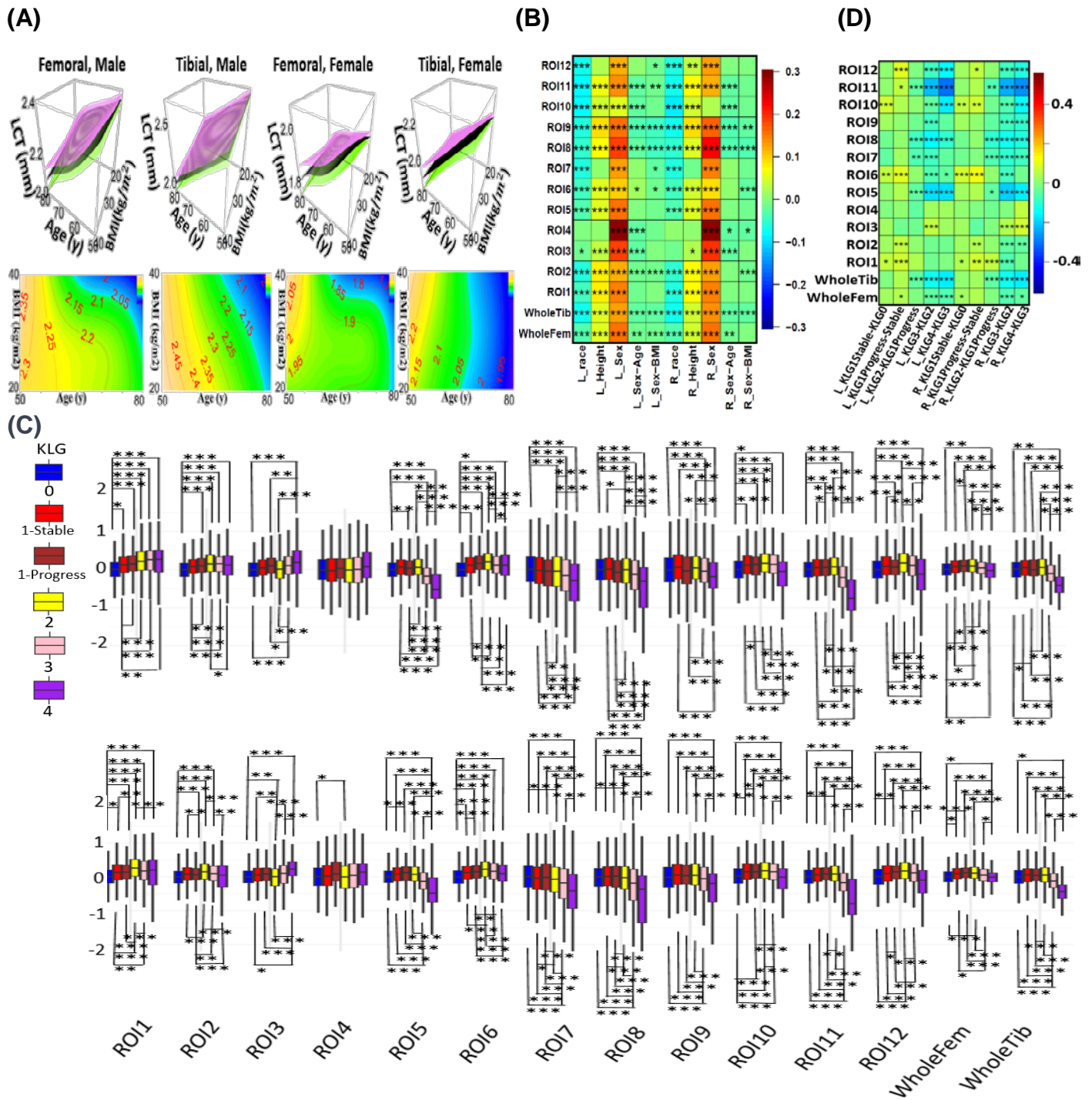


Fig. 2. Results of association analyses. (A) Overview of the trajectories of cartilage thicknesses in terms of BMI and age on the right knee, with both the 3D perspective view (row 1) and 2D contour plots (row 2) for the overall tibial and femoral regions among males and females, respectively. 95% confidence surfaces are shown in purple (upper) and green (lower) colors. (B) Heatmap to show the effect sizes (standardized coefficients) of race, height, sex, sex*age, sex*BMI on cartilage thicknesses in tibial, femoral and 12 regions of interest (ROIs) for left (column 1-5) and right (column 6-10) knees. (C) Boxplots depict cartilage thickness (adjusted for confounders) for left (upper) and right (lower) knees. Different colors represent various KLGs. KLG1 is split into 'KLG1progress' (progressed to KLG2 in 96 months) and 'KLG1Stable'. (D) Heatmap to show group mean differences in (C). We use *, ** and *** to denote significance levels 0.05, 0.01 and 0.001, respectively, after Bonferroni correction for 28 knee regions in (B-D). Significance levels in (C-D) are shown based on Tukey adjusted p-values.

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

accuracy (ACC) among any subset of demographic factors, raw regional cartilage thicknesses, and centile features depicted in **Figures 3B** and **4B**, and **Supplementary Tables S3-S5** and **S6-S8** using the combined demographic factors, raw regional cartilage thicknesses and centile features (refer to **Supplementary Table S2** for feature details). In both binary classifications, the combined set of predictors of demographic factors, raw regional cartilage thicknesses and centile features was optimal, which outperformed those reduced models without centile features, in terms of increased AUC, F1 score and ACC. Additionally, we conducted the classification to distinguish between ($KLG \leq 1$) and ($KLG \geq 2$) to evaluate our performance against existing literature. Results were shown in **Supplementary Figure S3** and **Supplementary Tables S9-S11**.

For the classification of $KLG \leq 1$ against $KLG = 2$, the LightGBM method, with random forest (RF) permutation importance feature selection, proved most effective (**Figure 3C**), achieving both the highest AUC of 0.953 ($\sigma = 0.001$) and F1 of 0.887 ($\sigma = 0.003$). Various plots, including the association circular plot (**Figure 3A**), SHapley Additive exPlanations (SHAP) plot (**Figure 3D**), pairwise comparison boxplots (**Figure 3E**), and nomogram (**Figure 3F**), elucidated the predictors' significance, effect sizes, prediction contributions, and joint effects. In particular, cartilage thickening in the posterior medial (ROI 6) and anterior lateral (ROI 3) femoral regions, cartilage thinning in the posterior and central lateral tibial (ROIs 7 and 8) and central medial femoral (ROI 5) regions, coupled with an increased BMI, contributed to the classification of $KLG = 2$.

In differentiating $KLG2$ based on progression, the RF and LightGBM (with random forest feature selection) achieved the highest F1 score and AUC, respectively (**Figure 4C**). In particular, the RF achieved an AUC of 0.979 ($\sigma = 0.002$) and F1 score of 0.941 ($\sigma = 0.006$). The corresponding circular plot, SHAP plot, boxplots, and nomograms were displayed in **Figure 4A**, **4D**, **4E** and **4F**, respectively. A decrease in cartilage thickness in the lateral tibial regions (ROIs 7, 8 and 9), an increase in the anterior femoral regions (ROIs 3 and 4), and a heightened BMI increased the risk of rOA progressing to $KLG3$.

When distinguishing between $KLG \leq 1$ and $KLG \geq 2$, the LightGBM method, combined with GLMBoost importance-based feature selection, yielded the best performance with an F1 score of 0.818 ($\sigma = 0.001$) and an AUC of 0.900 ($\sigma = 0.003$) (**Supplementary Figure S3C**). Factors contributing to a classification of $KLG \geq 2$ include cartilage thickening in posterior femoral (ROIs 6 and 1) region, thinning in central tibial (ROIs 11 and 8), central medial femoral (ROI 5) and the overall femoral and tibial regions, increased BMI and age, and identifying as female, and black (**Supplementary Figure S3A**, **S3D**, **S3E** and **S3F**).

In existing literature, there were studies (53) on the auto-classification of 5 KLG scores. However, few have specifically focused on differentiating between $KLG \leq 1$ and $KLG = 2$ or between $KLG2$ with and without progression. In addition, our performance in classifying $KLG \leq 1$ against $KLG \geq 2$ was comparable to the 3D CNN model proposed by Guida et al. (2021) (99) which achieved an F1-score of 0.831 and an AUC of 0.911. Our approach, however, offered enhanced simplicity and interpretability, and was validated on a larger cohort.

Reference cartilage thickness charts predict the incidence and initial-stage rOA

In this section, we distinguished between incidence rOA (KLG increasing from < 2 to ≥ 2) and initial-stage rOA (KLG from 0 to 1) transition. We evaluated the performance of 22 machine learning models, using a combination of demographic factors, raw regional cartilage thicknesses, and centile features. For the incident rOA prediction, the optimal models (highest F1 score) incorporated all these predictor variables, while for initial-stage predictions, the optimal model incorporated predictor variables without demographic variables. In both cases, they outperformed reduced models that did not include centile features. Detailed model performances for incidence and initial-stage rOA predictions can be found in **Supplementary Figures S4B** and **S5B**, and **Supplementary Tables S12-S14** and **S15-S17**. While ROC curves of the top models were similar (**Supplementary Figure 4C**) for incident rOA prediction, there were differences for initial-stage prediction (**Supplementary Figure 5C**). The linear discriminant analysis (LDA) model, coupled with Glmboost for feature pre-screening (AUC = 0.813, $\sigma = 0.002$; F1 = 0.685, $\sigma = 0.004$), and the logistic regression (Glmnet) method, coupled with random forest feature pre-screening (AUC = 0.653, $\sigma = 0.021$; F1 = 0.639, $\sigma = 0.036$) were optimal models for the incidence and initial-stage rOA transition prediction, respectively. The marginal associations including AUC, p-values and Bonferroni-corrected p-values of top predictors (AUC > 0.7) with rOA are illustrated in **Supplementary Figures S4A** and **S5A**, the importance of top predictors on the test set are shown in the SHAP plots (**Supplementary Figures**

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

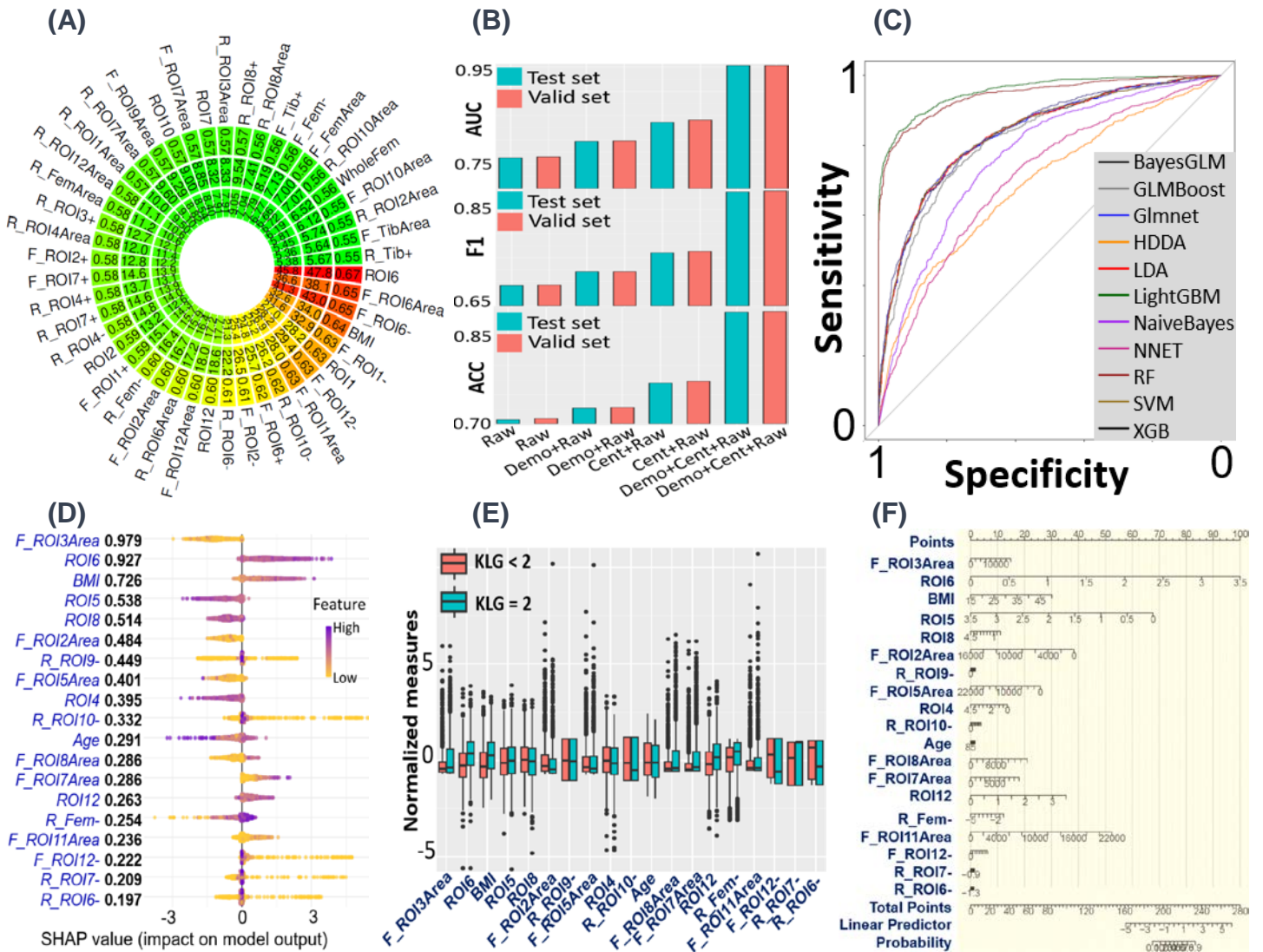


Fig. 3. Results of rOA classification between KLG = 2 and KLG < 2. (A) Circular plot showing associations between radiographic OA (rOA) status KLG = 2 and top predictors. Outer, middle, and inner circles signify Area Under the Receiver Operating Characteristic Curve (AUC), $-\log_{10}p$, and $-\log_{10}(\text{Bonferroni-adjusted } p)$, respectively. Only results with $\text{AUC} > 0.55$ are shown. (B) Barplots illustrating the median performance metrics (AUC, F1 score, Accuracy) across 22 machine learning methods using different predictor combinations ("Demo" for demographic factors, "Raw" for raw regional cartilage thicknesses and "Cent" for centile features). Performance for test (blue) and validation (red) datasets is indicated. (C) Receiver Operating Characteristic (ROC) curves of 11 models (each using random forest (RF) for variable selection) on the test set. The LightGBM model achieves the highest AUC and F1. (D) SHAP Additive exPlanations (SHAP) plot revealing impacts of top contributing 19 predictors (larger than 80% contribution in total) on the test set based on the LightGBM model. Predictors are sorted by their importance; for each predictor, each dot represents a patient's SHAP value plotted horizontally, colored by the value of the feature from low (yellow) to high (purple). If purple and yellow dots are plotted at the lower and higher side, respectively, the KLG = 2 risk becomes higher as the feature increases. (E) Boxplots of the top predictors (normalized to mean zero and unit variance) for KLG = 2 (red) and KLG < 2 (blue). (F) Nomogram (100) depicting coefficients of (not standardized) top predictors in a joint logistic model. The x-axis ("points") refers to the scores assigned to predictors based on their influence or weight in the predictive model.

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

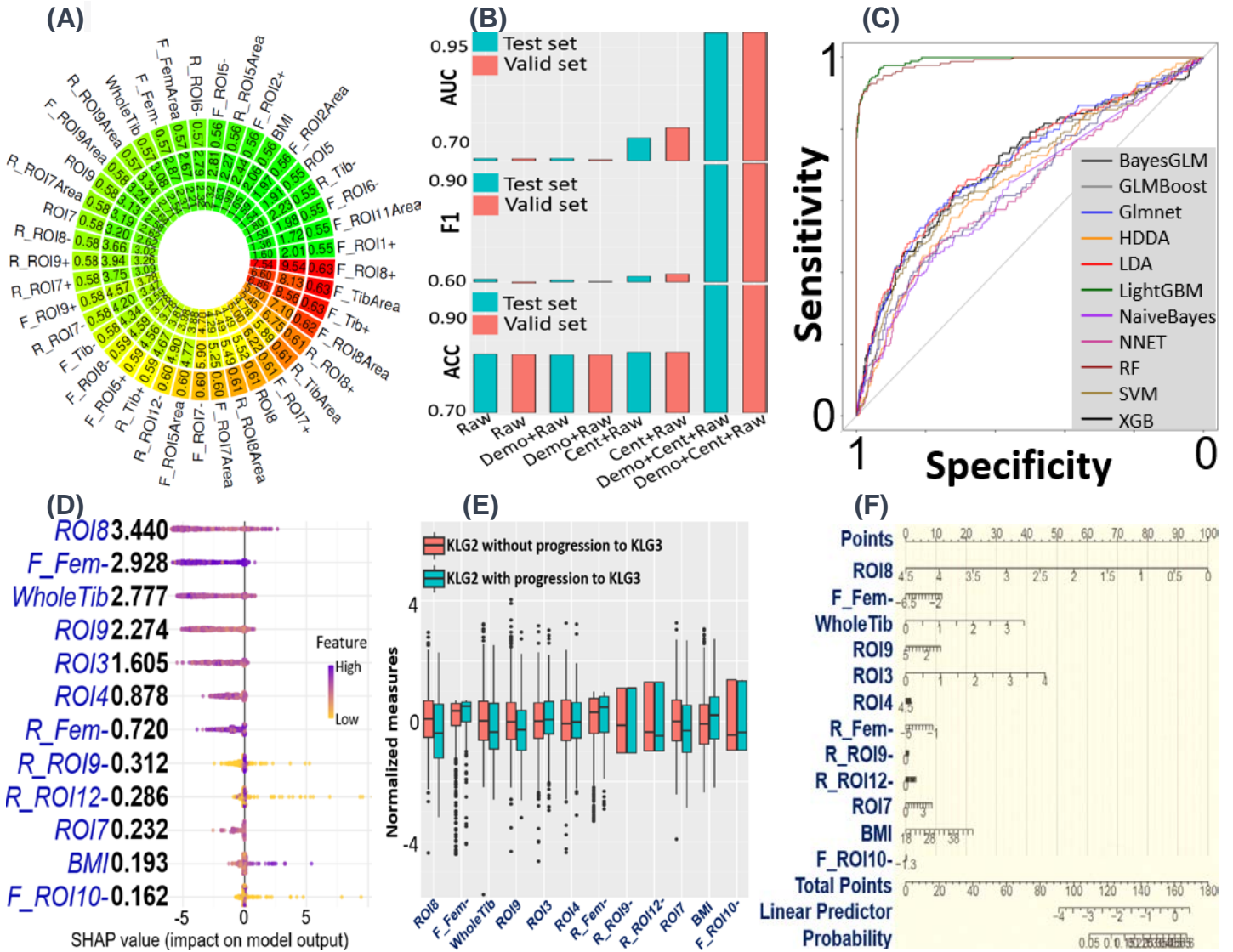


Fig. 4. Results of KLG2-KLG3 progression prediction. (A) Circular plot showing associations between radiographic progression status and top predictors. Outer, middle, and inner circles signify Area Under the Receiver Operating Characteristic Curve (AUC), $-\log_{10}p$, and $-\log_{10}(\text{Bonferroni-adjusted } p)$, respectively. Only results with $\text{AUC} > 0.55$ are shown. (B) Barplots illustrating the median performance metrics (AUC, F1 score, Accuracy) across 22 machine learning methods using different predictor combinations ("Demo" for demographic factors, "Raw" for raw regional cartilage thicknesses and "Cent" for centile features). Performance for test (blue) and validation (red) datasets is indicated. (C) Receiver Operating Characteristic (ROC) curves of the 11 models (each using random forest (RF) for variable selection) on the test set. The LightGBM and RF models achieve the highest AUC and F1, respectively. (D) SHapley Additive exPlanations (SHAP) plot revealing impacts of top 12 predictors (> 90% contribution) on the test set based on the RF model. Predictors are sorted by their importance; for each predictor, each dot represents a patient's SHAP value plotted horizontally, colored by the value of the feature from low (yellow) to high (purple). If purple and yellow dots are plotted at the lower and higher side, respectively, the progression risk becomes higher as the feature increases. (E) Boxplots of the top predictors (normalized to mean zero and unit variance) for progressive KLG2 (red) and stable KLG2 (blue). (F) Nomogram (100) depicting coefficients of top predictors in a joint logistic model. The x-axis ("points") refers to the scores assigned to predictors based on their influence or weight in the predictive model.

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/) .

S4D and **S5D**), and boxplots of group differences of normalized (zero mean and unit variance) top predictors were shown in **Supplementary Figures S4E** and **S5E**. These predictors' joint effects were shown in nomograms (**Supplementary Figures S4F** and **S5F**). From those results, we observed that the larger BMI, along with increased cartilage thickness at the posterior and central lateral femoral regions and reduced cartilage thickness at the overall tibial and femoral regions served as top contributors of rOA early-stage progression, while increased cartilage thickness at posterior lateral, posterior medial femoral regions, and reduced cartilage thickness in the central medial femoral region were initial-stage rOA progression indicators.

These findings were in line with previous research demonstrating the relationship between BMI and OA development (101, 102). The swelling of lateral femoral cartilage thickness in predicting early-stage rOA progression paralleled findings from studies such as Erhart, et al. (2008) (103). Furthermore, our study revealed the increase in the posterior medial and the reduction in the central medial femoral regions as potential biomarkers of initial rOA. This underlined the importance of further studies to explore underlying mechanisms.

Reference cartilage thickness charts aid in local disease mapping for rOA progression

We analyzed the difference of transformed local centile scores representing cartilage thinning and thickening (details in *Method* Section) between participants with and without rOA progression, delineated across five KLG groups in the femoral and tibial regions (**Figure 5A**). **Figure 5B** illustrates differences and significance of comparisons shown in **Figure 5A**, indicating locations where cartilage thickness differed between participants with and without rOA progression at each radiographic stage. In particular, we observed increased cartilage thickness in the posterior medial and posterior lateral femoral regions and a reduction in the central medial femoral region in the dashed black boxes in **Figure 5A**. Furthermore, we discerned (**Figure 5B**) a substantial reduction in the lateral tibial, increase in posterior medial femoral regions and minor reduction in the anterior femoral regions in participants with rOA progression at KLG = 2, which could potentially signify a progression towards KLG = 3. Similarly, a larger reduction (**Figure 5B**) in the medial tibial region among KLG3 participants might suggest further progression into KLG = 4. We prepared a movie available at <https://tli3.github.io/KneeOA/> which showcased dynamic changes in local centile scores for four subtypes of rOA progression during all visits: KLG = 0 → KLG = 1; KLG = 0 → KLG = 1 → KLG = 2; KLG = 0 → KLG = 1 → KLG = 2 → KLG = 3; KLG = 0 → KLG = 1 → KLG = 2 → KLG = 3 → KLG = 4. This demonstrated how local centile scores of the four rOA progression patterns changes across time. While there is extensive literature on the topic of rOA progression prediction, few studies specifically address an overall local progression of rOA.

Discussion

An integrative toolset for rOA predictive analyses

Traditional knee rOA association and prediction analyses (30-33) have primarily relied on manual readings of MR images, which often limit the study sample size and statistical power, potentially leading to conflicting results. Modern deep-learning models (45, 46), while offering accurate rOA prediction, pose challenges for clinical interpretability. In response to these challenges, we have developed an integrative toolset that combines the benefits of both traditional and modern approaches. This toolset includes DL-based LCT maps extracted from knee MR DESS image data, rOA-free nonlinear growth charts of knee cartilage thickness, enhanced statistical methods for extracting centile features and the implementation of ML-predictive modeling. These components work in harmony to allow more comprehensive OA diagnosis and prognosis. All codes are written in R/4.2.2 which we will make publicly accessible at www.github.com. To further supplement our toolset, we have proposed a variety of visualization methods, including circular plots, SHAP plots, boxplots, and nomogram plots. These are integrated with local diagnosis visualizations, providing a means to capture and visualize regional abnormalities during disease progression. Compared to previous research which has often examined cartilage thickness changes during rOA progression in a broader scope, our developed toolset advances precision by identifying localized alterations in cartilage thickness among participants at various radiographic stages with or without future progression. This enhanced our prior work (64), which although successfully detected local dynamic abnormal regions and rOA subclassifications, did not predict future rOA progression.

On the other hand, the effective contribution of our methodology is reflected in its robust performance and high prediction accuracy, with AUC values of 0.953, 0.900, 0.979 and 0.813 and F1 scores of 0.887, 0.818, 0.941,

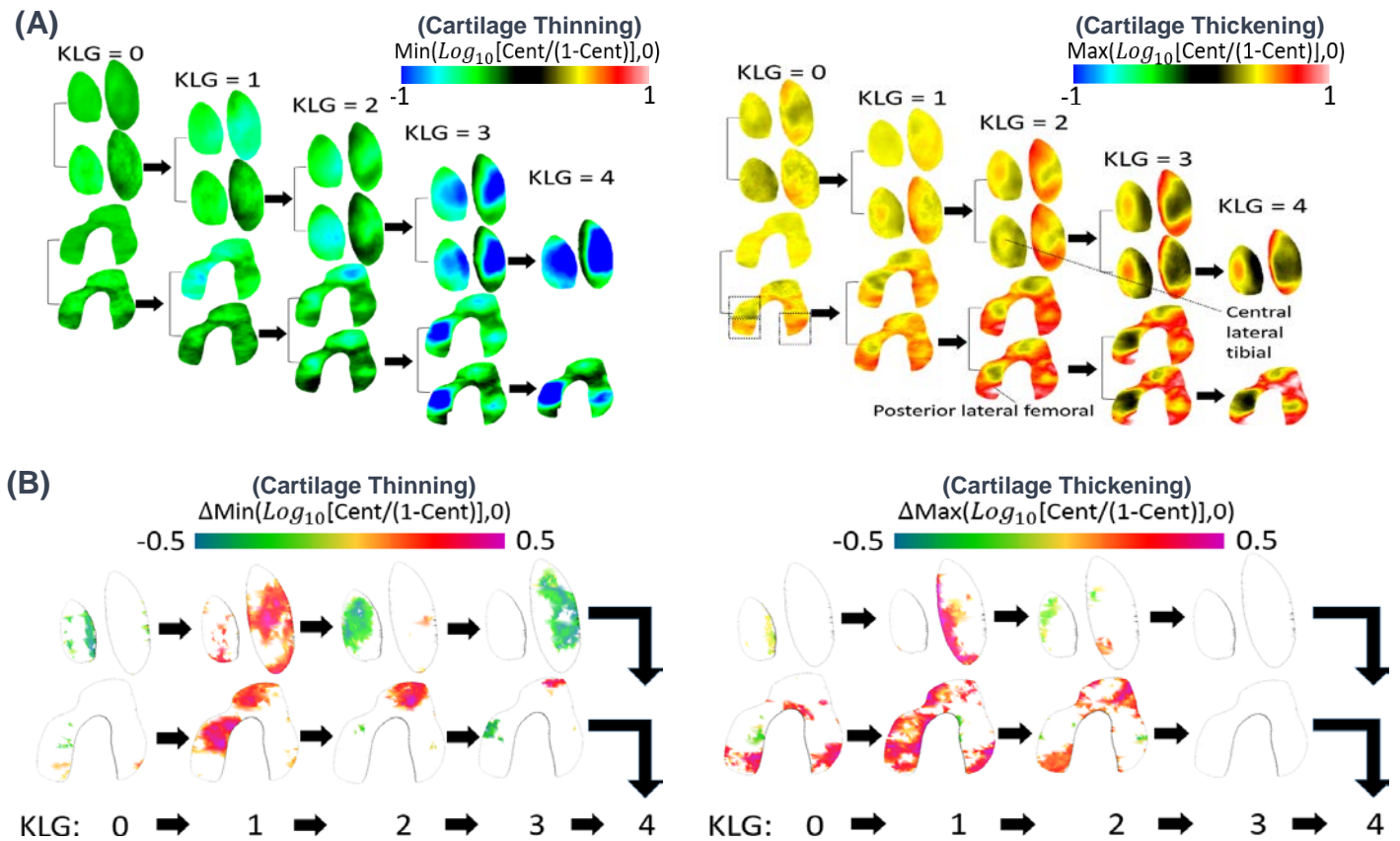


Fig. 5. Visualization of the local centile score map for different KLG progression patterns. (A) A comparison of mean local centile scores, which were logit10-transformed ($\log_{10}(x/(1-x))$), between participants with and without rOA progression across five KLG groups in the femoral and tibial regions. These scores were separated into positive and negative components, representing bidirectional thickness changes (reduction and increase). **(B)** Differences of each comparison in panel (A), showing locations with FDR-adjusted $p < 0.2$. $\Delta \text{Max}(\text{Log}_{10}[\text{Cent}/(1-\text{Cent})], 0) = \text{Max}(\text{Log}_{10}[\text{Cent}_2/(1-\text{Cent}_2)], 0) - \text{Max}(\text{Log}_{10}[\text{Cent}_1/(1-\text{Cent}_1)], 0)$, where Cent_2 and Cent_1 are local centile scores with and without progression, respectively (ΔMin defined similarly). A pink (or green) color in cartilage thinning indicates a decrease (or increase) in OA-induced cartilage reduction compared to non-progression; a pink (or green) color in cartilage thickening indicates an increase (or decrease) in OA-induced cartilage thickening compared to non-progression.

and 0.685, for rOA classification of $\text{KLG} \leq 1$ versus $\text{KLG} = 2$ and $\text{KLG} \leq 1$ versus $\text{KLG} \geq 2$, and prediction of KLG2-3 transition and incidence rOA transition, respectively. We here include the macro-F1 score into our evaluation metrics, which effectively balances precision and recall, providing a more conservative evaluation in scenarios with imbalanced data (104). Despite being frequently used in prior studies (48, 49, 52, 55, 60, 63), metrics like accuracy and AUC can be sensitive to the class distribution and may result in overly-optimistic results in imbalanced scenarios, as the model may perform well on the majority class but poorly on the minority class. In previous literature, Thomas, et al. (2020) (53) and Guida et al. (2021) (99) conducted the same classification task as ours in classifying $\text{KLG} \leq 1$ against $\text{KLG} \geq 2$, whose reported mean F1-score were 0.80 and 0.831 based on deep learning methods (our result was 0.818). Tiulpin, et al. (2020) (57) reported a macro-F1 of less than 0.60 and 0.57 for rOA multi-class classification. In terms of predicting OA progression, Widera, et al. (2020) (59) reported a weighted F1 of 0.69 based on the OAI data, while Chan, et al. (2021) (61) achieved F1 scores of 0.837 and 0.810 for predicting OA incidence and deterioration, respectively, indicating higher prediction accuracy than our model. However, it is worth noting that their model used additional symptomatic data, history of knee

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/) .

injury and surgery, metabolic syndrome and living habits, for prediction of OA progression characterized by JSW and WOMAC pain score, while our work focused on easily-measurable factors such as age, sex, BMI, race, and height to predict rOA progression.

In addition, previous literature on early prediction of OA generally focuses on the definite-OA phase. Studies predicting the rOA initial transition remain less explored. Shamir, et al. (2009) (105) reported the automatic detection of doubtful rOA (KLG = 1) with a relatively low accuracy of 57%; our analyses predicted the transition to KLG = 1 from KLG = 0 with a higher accuracy of 65.9% ($\sigma = 0.027$). Predicting initial-phase rOA onset would provide insights of the underlying mechanisms of early OA. Furthermore, in our study, the consistency of the identified regions and signals observed across diverse analyses, including regional association analyses, future predictive analyses, and local disease region analyses lends robustness to our findings.

Insights and findings from a clinical perspective

In knee OA patients, prevalent symptoms of pain and joint stiffness often originate from innervated tissues such as the synovium and bone (106, 107), instead of the typically avascular and aneural articular cartilage. Yet, cartilage degradation is closely linked to synovitis, bone marrow lesions, and osteophyte formation. The degradation of cartilage can release inflammatory mediators that trigger synovitis, which in turn can accelerate cartilage degradation, thus perpetuating a destructive cycle (108, 109). Moreover, as cartilage degrades, the resulting increased load and altered joint mechanics can lead to bone marrow lesions, detectable through imaging. As this process unfolds, cartilage thinning often occurs as one of the early detectable signs of OA. Thus, assessing changes in cartilage thickness using MRI serves as a valuable tool for diagnosing and tracking OA progression.

While considerable research has delineated the association between BMI, age, and cartilage thickness (89, 90, 95-97), studies on the age-BMI interplay remain scarce. We construct an age-BMI growth chart for knee cartilage thickness among rOA-free participants, incorporating complex age-BMI interactions. The observed interactions between age and BMI in our study support the notion that high BMI values may accelerate aging for rOA-free knee cartilage.

Previous literature (34, 40-44) has observed that cartilage thickening may act as important indicators in early OA. For instance, it was observed (34) that the frequency of knees classified with medial femorotibial cartilage thickening was equal to that of medial thinning in knees with KLG = 2, whereas cartilage thinning was 2.6 times more frequent than thickening in knees with KLG = 3. Despite these observations, understanding of cartilage thickening in early stages of OA is limited, necessitating further validation through larger-scale studies. Commonly used OA imaging scores like MOAKS and WOMBS primarily focus on cartilage loss, which may lead to underrepresentation of cartilage thickening in early stages. Furthermore, the potential of cartilage thickening to serve as a predictive biomarker for rOA progression remains unclear. Our results provide evidence to suggest that cartilage thickening may occur at the initial and earlier radiographic grades and may present a biomarker for rOA progression.

In summary, our study has shed light on the potential of combined statistical-ML approaches to provide valuable insights into rOA progression and its early indicators. Further studies are warranted to confirm our findings and explore the clinical implications for early detection and prognosis of rOA.

Limitations and future work

Certain limitations exist within our study that should be highlighted. Within our scope of study, our predictor variables were focused on easily-measurable factors such as age, sex, BMI, race, and height, which can be obtained without additional clinical examinations. Nevertheless, other influential factors, such as knee injury, knee alignment, laxity, quadriceps muscle strength, physical activity, genetic predisposition, and overall health conditions, could be integrated into future models to enhance disease incidence and progression predictions. Second, while we have chosen to predict rOA progression, a broader scope could include the prediction of JSN, symptoms, cartilage lesion incidence, and TKR. As rOA only forms a component of OA, pain scores could also be considered for the OA prediction (29) instead of rOA. Third, we have constructed cartilage thickness growth charts for rOA-free participants in the OAI, which are not the same as healthy controls. There lies potential for

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

future research to build these growth charts for healthy control participants without knee pain, which could be beneficial for comparative studies. Finally, while our models utilize scan features from one single scan to predict future rOA progression, exploring advanced statistical models that incorporate features from all historical scans, like autoregressive models, could offer deeper insights.

Method

Study population

The Osteoarthritis Initiative (OAI; <https://nda.nih.gov/oai/>) is a prospective, multi-center, longitudinal study initiated in 2004. It recruited 4,796 participants aged 45 to 79 years, who were either at high risk of or already had symptomatic knee OA (82). Participants were followed for up to 96 months, with imaging and clinic visits scheduled at baseline, 12 months, 24 months, 36 months, 48 months, 72 months and 96 months (110). We have observed significant difference in knee cartilage thickness between KLG = 0 and KLG = 1 participants, supporting the use of rOA-free participants for the creation of reference group charts. Generally, we used the 957 rOA-free participants as the training data for the construction of the rOA-free reference cartilage thickness growth charts. Participants included in all the other participants were used to test the prediction performance of rOA progression based on the reference charts. Given the potential bilateral symmetry in rOA progression (111), it is not recommended to consider a knee with KLG0 as rOA normal if its contralateral knee exhibits a KLG > 1. Based on this, we excluded participants if they met any of the following criteria: (i) one knee consistently presents with $KLG \leq 1$ across all visits while the contralateral knee has $KLG > 1$ at any visit, (ii) one knee consistently shows $KLG = 0$ across all visits while the contralateral knee displays $KLG = 1$ at any visit, (iii) knee MRIs are missing or are of poor quality across all visits, or (iv) participants of minority races, those identifying as hispanic, or those not identifying as either white or black. Sample size and other demographic information for participants at their baseline visits in different subpopulations were summarized in **Table 1** (based on the KLG of right knees).

Magnetic Resonance Images and preprocessing

Longitudinal assessments and measurements were performed on 4,796 participants at up to seven time points by using MR imaging data. The imaging data was collected using four 3.0 Tesla Siemens Trio MRI scanners over a four-year follow-up period. The use of 3T-MRI scanners and standardized protocols, as opposed to 1.5T-MRI, offered greater signal-to-noise and contrast-to-noise ratios due to its larger field strength. The OAI applied various imaging acquisition protocols. We specifically used those MRIs employing the sagittal 3D Dual Echo Steady State (DESS) technique with selective water excitation (WE), which had been demonstrated superior for cartilage discrimination (110). Although there were a few exceptions, the majority of these images had dimensions of 384 x 384 x 160 and a voxel resolution of 0.36 x 0.36 x 0.7 mm³. In prior research, we developed the Dynamic Abnormality Detection and Progression (DADP) pipeline (64) and an open-source knee-OAI analysis toolkit https://github.com/uncbiag/OAI_analysis_2 for the creation and analysis of knee thickness maps. In this study, the OAI MR images were preprocessed (**Figure 1A**) through the image analysis module including the following four steps: segmentation and meshing; 3D thickness map computation; registration; and thickness map projection. The output of this step was a two-dimensional (2D) thickness map in a common 2D atlas coordinate system for each OA MR image.

KLG measures for radiographic severity

The KLG system is a scoring system used to evaluate the severity of knee OA on radiographic images. The scale ranges from 0 to 4, with 0 indicating the absence of changes and 4 indicating the presence of substantial osteophytes, marked by joint space narrowing, severe sclerosis, and evident deformity of the bone contour. The KLG scores were obtained from radiographs taken in the standing postero-anterior position using the fixed-flexion protocol. This protocol involves acquiring bilateral X-rays while the patient stands with knees flexed between 20-30 degrees and feet internally rotated 10 degrees. To ensure consistent patient positioning, the degree of knee flexion and foot rotation was fixed using a plexiglass positioning frame (SynaFlexer, Synarc, Inc, San Francisco, CA, USA) (112).

Construction of the FKTR charts

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

After obtaining 2D maps of cartilage thickness for all individuals in the common coordinate system, we used statistical models to construct the rOA-Free Knee Thickness Reference (FKTR) charts. The charts aimed to characterize the aging trajectories of LCT maps using data from 957 rOA-free individuals. To do this, an experienced musculoskeletal radiologist (DN) manually annotated 12 ROIs with 6 for the femoral and 6 for the tibial cartilages on the standard 3D atlas space according to their physical locations. These ROIs were then mapped to the 2D atlas space (**Figure 1B**). Next, we fitted both regional-average and local cartilage thickness regression models, taking into account the demographic covariates including age, sex, BMI, height, and race of the individuals as covariates. Note that in this step, clinical covariates were considered contributing factors rather than confounding factors. We used a Generalized Additive Mixed Model (GAMM) (113), which allowed us to model the nonlinear and dynamic nature of cartilage thickness development during aging. Let $y_{i,v}(t, b)$ denote the cartilage thickness for the i -th subject at location v and age t , and with BMI b , with $i = 1, \dots, n$. Specifically, the GAMM model can be formulated as the model $M1$:

$$M1: y_{i,v}(t_{i,j}, b_{i,j}) = f_v(t_{i,j}, b_{i,j}) + \beta_{0,v}g_i + \beta_{1,v}b_{i,j}g_i + \beta_{2,v}g_it_{i,j} + \beta_{3,v}R_i + \beta_{4,v}h_{i,j} + \alpha_{i,v} + e_{i,v}(t_{i,j}).$$

Here g_i and r_i represent the sex and race for subject i , with $g_i = 1$ for male and 0 for female and $R_i = 1$ for black and 0 for white groups since most of the participants are white, $b_{i,j}$ and $h_{i,j}$ represent the BMI and height for the subject i at age $t_{i,j}$, respectively, and $\alpha_{i,v}$ and $e_{i,v}(t_{i,j})$ denote the random intercept effect and random Gaussian noise for subject i at location v and age t . $\beta_{0,v}$, $\beta_{1,v}$, $\beta_{2,v}$, $\beta_{3,v}$, and $\beta_{4,v}$ are coefficients of the sex, BMI-sex and sex-age interactions, race and height, respectively. Our rationale for examining the complex interactions between age and BMI is that both serve as important continuous indicators of aging and knee OA. Taking the variables described above, the two-dimensional functions $f_v(t, b)$ were estimated by full tensor product smoothing using cubic splines (114) to model the nonlinear age-BMI interactions nonparametrically. Smoothing hyperparameters in $f_v(t, b)$ were chosen by maximizing the marginal likelihood. ANOVA F-statistics were calculated to determine the significance of the main effects and BMI-sex, age-sex interactions. Two alternative models were compared with model $M1$ in the **Supplementary Text**. All the above estimations and 95% simultaneous confidence intervals were calculated through the `mgcv` package (115) with R version 4.2.2. This analysis was solely based on the rOA-free participants.

A comparison of cartilage thickness progression between different rOA groups

Participants with KLG = 0 and KLG = 1 were often grouped together as rOA-free due to the subtle distinctions between these grades (116). In our study, we separated them to enable detailed comparisons across OA progression phases. First, we adjusted the cartilage thickness for all confounding factors, by removing all fixed effects of rOA-free participants from model $M1$. Afterwards we fit the following linear mixed model $R1$

$$R1: \tilde{y}_{i,v}(t_{i,j}, b_{i,j}) = KLG_Group + \alpha_{i,v} + e_{i,v}(t_{i,j}),$$

where \tilde{y} is the residualized cartilage thickness, and KLG_Group is a categorical variable which includes six KLG group, the KLG0, KLG2, KLG3, KLG4, "KLG1progress" (progression to KLG > 1) and "KLG1Stable", and $\alpha_{i,v}$ and $e_{i,v}(t_{i,j})$ are Gaussian random effects and random noise, respectively. One-way ANOVA F tests statistics were used to test the effect of KLG on the residualized cartilage thickness, accounting for mixed effect models via Satterthwaite's degrees of freedom method (117). These tests were performed across 28 regions, including 12 ROIs and the overall femoral and tibial regions in both the left and right knees. P-values were adjusted for 28 independent tests using the Bonferroni correction. For regions with significant KLG group effect, posthoc analyses were performed to test pairwise differences using the Tukey test to correct for multiple comparison of pairs. This analysis was performed through the `lme4` and `emmeans` packages with R version 4.2.2, and was based on all scans from rOA-prevalence and KLG0-1-stable participants.

Local centile score extraction

In the preceding subsection, we constructed reference developmental trajectory charts to depict knee growth patterns in individuals without rOA. Using these charts as a baseline, we compared and evaluated the developmental trajectories of knee cartilage thicknesses from participants in different groups. We introduced the **Local centile Scores (LCS)** to assess the extent of local abnormalities/extremities for a given knee. Specifically, we defined the LCS at each location v as the centile of cartilage thickness among all rOA-free knees, adjusted for demographic confounding covariates. After accounting for these confounders, we ranked the residualized cartilage thickness of the knee relative to the residuals of all rOA-free knees at each v , and then convert these

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

ranks to centiles. The LCS could be calculated in two ways, which account for fixed effects only from confounding covariates (Method E) or both the fixed and participant-level random effects (Method R), respectively. More details and formulations were shown in the **Supplementary Text**. Similar methods can be found in the centile scores of Bethlehem, et al. (2022) (118).

Association analyses between rOA progression and LCS

First, the LCS underwent a logit transformation, a procedure that extended the values near the extremes of the centile (closer to 0 or 1),

$$\text{Logit}_{10}[LCS(v)] = \text{Log}_{10}[LCS(v)/(1 - LCS(v))].$$

The median is naturally mapped to $\text{Logit}_{10}0.5 = 0$. This transformation essentially increases the sensitivity of our model to extreme centiles, allowing us to assess the degree of departure from the median cartilage thickness in both directions. We then bifurcated these transformed scores into negative and positive components. This segregation represents the bidirectional changes in cartilage thickness: the **cartilage thinning LCS** and **thickening LCS**,

$$\begin{aligned} \text{Thinning LCS: } \text{Logit}_{10}[LCS(v)]_- &= \min\{\text{Logit}_{10}[LCS(v)], 0\}; \\ \text{Thickening LCS: } \text{Logit}_{10}[LCS(v)]_+ &= \max\{\text{Logit}_{10}[LCS(v)], 0\}. \end{aligned}$$

Specifically, the positive components reflect instances where cartilage thickness is greater than the median, while the negative components capture instances where cartilage thickness falls below the median. Next, we tested the difference of thinning and thickening LCS between KLG0 knees with and without KLG1 progression, between KLG1 knees with and without KLG2 progression, between KLG2 knees with and without KLG3 progression, and between KLG3 knees with and without KLG4 progression, using two-sided two-sample t-tests. P-values were FDR-corrected for all pixels v .

Centile feature extraction

We proposed a novel concept of the **global centile scores (GCS)** to assess the global abnormality/extremity of a given knee across all locations. Specifically, the GCS is defined as the mean centile of **the LCS spatial distribution** among rOA-free knees, where the LCS spatial distribution is characterized by the cumulative distribution function (CDF). Denote the CDF of $LCS(v)$ for $v \in V$ as $F(\cdot)$, which can be formulated as

$$F(x) \triangleq \int (v \in V)[LCS(v) \leq x] dv.$$

The centile function $GCS(x)$ (in region V) was defined as the centile of $F(x)$ among that of all rOA-free knees,

$$GCS(x) = \text{Pr}(F(x) \leq F_0(x)), F_0(x) \text{ is the CDF of } LCS(v) \text{ among rOA free subjects,}$$

and the final GCS is defined as the mean of $GCS(x)$ function (mean centile function),

$$GCS = \int GCS(x) \pi(x) dx$$

where $\pi(x)$ was a Bayesian prior on the interval $[0,1]$, such as a uniform distribution on $[0,1]$, which is used to specify which percentile of the LCS distribution is more important for global diagnosis.

This method was motivated by the cluster-extent based thresholding method (119-121). The CDF can also be interpreted as the area where LCS is less than the threshold x . For example, when $\pi(x) = 1$ at $x = 0.05$ and 0 otherwise, the CDF is the "disease region" size determined by the "p-value" < 0.05 . In this sense, the GCS is an extension to cluster-extent p-value-based inference, with a fix significance level. In our setting, $\pi(\cdot)$ was data-driven (**Supplementary Text**). Similar to the LCS, we calculated the **cartilage thinning GCS** and **thickening GCS** by

$$\begin{aligned} \text{Thickening GCS: } \text{Logit}_{10}[GCS]_- &= \min\{\text{Logit}_{10}[GCS], 0\}; \\ \text{Thinning GCS: } \text{Logit}_{10}[GCS]_+ &= \max\{\text{Logit}_{10}[GCS], 0\}. \end{aligned}$$

Besides the GCS, we extracted other centile features, the integrated CDF (ICDF), based on the CDF of the LCS. The ICDF was defined as

$$\text{ICDF} \triangleq \int F(x) \pi(x) dx.$$

This integration reflects the mean cartilage thinning area across all threshold x . Both a larger ICDF and GCS imply a reduced cartilage thickness. The thickening GCS, thinning GCS and ICDF features were extracted

It is made available under a [CC-BY-ND 4.0 International license](#) .

across 14 regions (12 ROIs and the overall femoral and tibial regions). In total, for each knee we extracted 84 centile features, including 14 Thinning GCS, 14 Thickening GCS, and 14 ICDF features, with each based on the LCS calculated from Method F or Method R ($3 \times 2 \times 14 = 84$). Detailed feature interpretations were given in the **Supplementary Table S2**.

Classification and prediction tasks

Given the extensive analyses undertaken in this work, we have chosen to focus on the right knees for the subsequent classification and prediction tasks. Nevertheless, the established predictive framework is equally applicable to the left knees.

In the first classification, we categorized knees in the early-rOA phases ($KLG \leq 2$) into three distinct stages: $KLG \leq 1$, $KLG2$ progressing to $KLG3$ in 48 months, and $KLG2$ without such progression. For $KLG \leq 1$ (referred to as Class 1-1), we chose the knee from 48 months prior to the last visit in $KLG0$ -1-stable participants, ensuring the KLG remained ≤ 1 in the subsequent 48 months. This resulted in 1,362 knees from 1,362 participants. For $KLG2$ without progression (Class 1-2), we selected the knee from 48 months prior to the last visit in $KLG2$ -stable participants, yielding 804 knees from 804 participants. For participants undergoing a $KLG2$ -3 transition (Class 1-3), we chose the knee from the final visit before its transition to $KLG3$, yielding 179 knees from 179 participants.

In the second classification, we aimed to distinguish between knees with $KLG \leq 1$ (Class 2-1) and those with $KLG \geq 2$ (Class 2-2). This task is a typical rOA classification problem, which was performed to validate our methods against existing literature. In particular, we incorporated all 449 $KLG4$ knees, and randomly selected 450 knees each from $KLG0$, $KLG1$, $KLG2$, and $KLG3$ categories to balance the sample size. For each participant, only one knee was randomly chosen, with the exception of those in the $KLG4$ category due to limited knees with $KLG4$. This configuration aligns well with the settings in Guida et al. (2021) (99).

In the rOA-incidence prediction, we predicted the transition into $KLG2$ in 48 months based on its current knee with $KLG < 2$. For knees with no such transition (Class 3-1), we selected the knee from 48 months prior to the last visit in $KLG0$ -1-stable participants. This resulted in 1,362 knees from 1,362 participants. For knees with transition (Class 3-2) we chose the knee from the final visit before its transition to $KLG2$, accounting for 213 knees from 213 participants.

In our rOA initial transition prediction, we aimed to predict the transition to $KLG1$ within 48 months for knees currently at $KLG = 0$. For knees with no transition (Class 4-1), we selected the knees from 48 months prior to the last visit in $KLG0$ -1-stable participants, yielding 957 knees from 957 individuals. On the other hand, for knees with transition (Class 4-2), we used data from the last scan before their transition to $KLG1$, resulting in 46 knees from 46 participants. Given the significant sample size imbalance between the two classes, we randomly downsized the Class 4-1 group, by randomly selecting 46 knees for a balanced 46:46 classification task.

Predictive modeling and hyperparameter selection

For better interpretation, we broke down the above three-stage classification of $KLG \leq 2$ into two binary classification problems: differentiating Class 1-1 from Classes 1-2 and 1-3, and distinguishing Class 1-2 from Class 1-3. In all the classification and prediction tasks, we extracted global centile features within each of the 14 ROIs, which were combined with the 14 raw mean cartilage thickness, age, sex, BMI, height, and race, which summed up to 103 potential biomarkers (**Supplementary Table S2**). The KLG of current knee was used as an additional predictor only in the rOA-incidence prediction. For each prediction task, we employed a 10-fold nested cross-validation to randomly divided knees into training, validation, and test sets.

We adopted a standard two-step predictive modeling approach. Initially, we computed the random forest permutation importance (122) and the GLMBoost importance (123) of each feature using the training dataset. Subsequently, features were ranked based on one of these importance scores. Afterwards, top S features were selected, which were fed into 11 baseline models, including the elastic net (Glmnet) (124), Glmboost (125), single-hidden-layer neural networks (NNET) (126), Naive Bayes (NaiveBayes) (127), Bayesian Generalized Linear Model (BayesGLM) (128), random forest (RF) (129), XGBoost (XGBTree) (130), Kernel-based Support

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/) .

Vector Machine (SVM) (131), Linear Discriminant Analysis (LDA) (132), High-Dimensional Discriminant Analysis (HDDA) (133) and LightGBM (134). The optimal hyperparameter S was determined by maximizing the AUC on the validation dataset. All other hyperparameters, such as learning rate, maximum depth and regularization parameters, were optimized using an internal 10-fold cross validation within the training data. Pairing the two feature selection methods with the 11 baseline models yielded 22 distinct predictive approaches. The method achieving the highest F1-score on the validation data was adopted as our final model.

Nested cross validation and model evaluation

To prevent overfitting, we used a three-layer-nested 10-fold cross-validation (CV) method to assess various predictive models. This method comprised three nested loops, all implementing a 10-fold cross-validation. Specifically, we partitioned all participants into 10 folds in the outer loop. Reserving one fold as an independent test set, the remaining folds were further divided into 10 subfolds in the middle loop. Choosing one subfold as the validation, the remaining folds were further divided into 10 internal subfolds in the inner loop. The outer loop was designed to give an unbiased performance evaluation of the model, the middle loop was used for determining the optimal number of selected features, and the inner loop was used for tuning all other hyperparameters. Because the test set was cycled through the 10 folds in the outer loop, each subject was evaluated once as an independent test subject. The prediction ACC, precision, recall, AUC, and F1-score for each machine learning model were reported. Furthermore, the three-layer-nested 10-fold CV was repeated five times to assess the standard deviation of these evaluation metrics.

Visualizing results for enhanced clinical interpretation

Circular plots were shown to include the association results between each feature and the response (class label), which specifically, included the p-values based on two-sided two-sample t-test, FDR-adjusted p-values, and the univariate AUC. We then performed the prediction tasks based on various feature combinations, such as raw cartilage thickness features only, the combination of raw and centile features and the combination of raw, centile and demographic features. Barplots were displayed to compare the AUC, F1 and ACC of different combinations, based on their corresponding final incorporated models. ROC curves was used to compare the specificity and sensitivity among 22 predictive approaches.

The SHAP summary plot were used to visualize both the SHAP importance (135) score of each feature and the feature's effect on the predictions. The y-axis represented the features sorted by their SHAP importance scores, while the x-axis reflects the Shapley value. Each dot on the plot represents a Shapley value for a specific feature and instance. An increase in the Shapley value suggests that, on average, including the feature would increase the predicted odds ratio for the instance. The color gradient indicates the feature's value, varying from low to high. Boxplots were used to display the group differences for features with top SHAP importance (contributions > 90% or 80%).

Finally, a nomogram chart was displayed to illustrate the combined effects of the top features, based on the assumption of a linear relationship between the log odds ratio and the features in a logistic regression model. In particular, the x-axis represent the points (100). Each predictor was assessed by moving vertically on the chart to find the corresponding points based on the patient's score, yielding the individual score. The total score was determined by summing up all the individual scores. Then, one moved vertically down from the "Total Points" scale to ascertain the probability of being identified as cases. Although the linear relationship did not typically provide the optimal fit, this nomogram chart roughly suggested coefficients and directions.

Acknowledgements

Data and/or research tools used in the preparation of this manuscript were obtained and analyzed from the controlled access datasets distributed from the Osteoarthritis Initiative (OAI), a data repository housed within the NIMH Data Archive (NDA). OAI is a collaborative informatics system created by the National Institute of Mental Health and the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS) to provide a worldwide

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

resource to quicken the pace of biomarker identification, scientific investigation and OA drug development. Dataset identifier(s): NIMH Data Archive Collection ID: 2343.

Author Contributions

TF. L., M. N., and H. Z. designed the study. M. N. supervised the imaging analyses and H. Z. supervised the statistical analyses. TF. L. and TY. L. analyzed the data. B. C., C. H., Z. S. and Z. X. processed imaging data. TF. L., D. N., YM. G, and AE. N. provided clinical interpretations of the analyses and reviewed the paper. TF. L. and TY. L. wrote the manuscript with feedback from all authors.

Funding Resources

Research reported in this publication was supported by the National Institutes of Health (NIH) under award number AR072013 (TY. L., B. C. and M. N.) and R01MH116527 (TF. L. and H.Z.). We also acknowledge support from NIH/NIAMS P30AR072580 (AE. N and YM. G). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interests

The authors declare no conflict of financial interests.

References

1. J. C. Sherwood, J. Bertrand, S. E. Eldridge, F. Dell'Accio, Cellular and molecular mechanisms of cartilage damage and repair. *Drug Discov Today* **19**, 1172-1177 (2014).
2. T. Neogi, The epidemiology and impact of pain in osteoarthritis. *Osteoarthritis Cartilage* **21**, 1145-1153 (2013).
3. G. Yang *et al.*, Burden of knee osteoarthritis in 204 countries and territories, 1990-2019: Results from the Global Burden of Disease Study 2019. *Arthritis Care Res (Hoboken)* 10.1002/acr.25158 (2023).
4. A. Cui *et al.*, Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. *EClinicalMedicine* **29-30**, 100587 (2020).
5. J. M. Jordan *et al.*, Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project. *J Rheumatol* **34**, 172-180 (2007).
6. J. N. Katz, K. R. Arant, R. F. Loeser, Diagnosis and Treatment of Hip and Knee Osteoarthritis: A Review. *JAMA* **325**, 568-578 (2021).
7. V. K. Srikanth *et al.*, A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. *Osteoarthritis Cartilage* **13**, 769-781 (2005).
8. R. B. Kreider *et al.*, International Society of Sports Nutrition position stand: safety and efficacy of creatine supplementation in exercise, sport, and medicine. *J Int Soc Sports Nutr* **14**, 18 (2017).
9. J. Lo, L. Chan, S. Flynn, A Systematic Review of the Incidence, Prevalence, Costs, and Activity and Work Limitations of Amputation, Osteoarthritis, Rheumatoid Arthritis, Back Pain, Multiple Sclerosis, Spinal Cord Injury, Stroke, and Traumatic Brain Injury in the United States: A 2019 Update. *Arch Phys Med Rehabil* **102**, 115-131 (2021).
10. E. Losina *et al.*, Lifetime medical costs of knee osteoarthritis management in the United States: impact of extending indications for total knee arthroplasty. *Arthritis Care Res (Hoboken)* **67**, 203-215 (2015).
11. J. Martel-Pelletier *et al.*, Osteoarthritis. *Nat Rev Dis Primers* **2**, 16072 (2016).
12. J. Sellam, F. Berenbaum, The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. *Nat Rev Rheumatol* **6**, 625-635 (2010).

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/) .

13. F. Berenbaum, Osteoarthritis as an inflammatory disease (osteoarthritis is not osteoarthrosis!). *Osteoarthritis Cartilage* **21**, 16-21 (2013).
14. F. P. Kroon *et al.*, Self-management education programmes for osteoarthritis. *Cochrane Database Syst Rev* 10.1002/14651858.CD008963.pub2, CD008963 (2014).
15. O. A. Uthman *et al.*, Exercise for lower limb osteoarthritis: systematic review incorporating trial sequential analysis and network meta-analysis. *BMJ* **347**, f5555 (2013).
16. R. Christensen, E. M. Bartels, A. Astrup, H. Bliddal, Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis. *Ann Rheum Dis* **66**, 433-439 (2007).
17. B. R. da Costa *et al.*, Effectiveness and safety of non-steroidal anti-inflammatory drugs and opioid treatment for knee and hip osteoarthritis: network meta-analysis. *BMJ* **375**, n2321 (2021).
18. G. C. Machado *et al.*, Efficacy and safety of paracetamol for spinal pain and osteoarthritis: systematic review and meta-analysis of randomised placebo controlled trials. *BMJ* **350**, h1225 (2015).
19. D. J. Hunter *et al.*, Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). *Osteoarthritis Cartilage* **19**, 990-1002 (2011).
20. H. Madry *et al.*, Early osteoarthritis of the knee. *Knee Surg Sports Traumatol Arthrosc* **24**, 1753-1762 (2016).
21. D. T. Felson *et al.*, A new approach yields high rates of radiographic progression in knee osteoarthritis. *J Rheumatol* **35**, 2047-2054 (2008).
22. A. Guermazi *et al.*, Prevalence of abnormalities in knees detected by MRI in adults without knee osteoarthritis: population based observational study (Framingham Osteoarthritis Study). *BMJ* **345**, e5339 (2012).
23. J. H. Kellgren, J. S. Lawrence, Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* **16**, 494-502 (1957).
24. A. G. Culvenor, C. N. Engen, B. E. Oiestad, L. Engebretsen, M. A. Risberg, Defining the presence of radiographic knee osteoarthritis: a comparison between the Kellgren and Lawrence system and OARSI atlas criteria. *Knee Surg Sports Traumatol Arthrosc* **23**, 3532-3539 (2015).
25. N. J. Collins, E. H. G. Oei, J. L. de Kanter, B. Vicenzino, K. M. Crossley, Prevalence of Radiographic and Magnetic Resonance Imaging Features of Patellofemoral Osteoarthritis in Young and Middle-Aged Adults With Persistent Patellofemoral Pain. *Arthritis Care Res (Hoboken)* **71**, 1068-1073 (2019).
26. P. R. Kornaat *et al.*, MRI assessment of knee osteoarthritis: Knee Osteoarthritis Scoring System (KOSS)--inter-observer and intra-observer reproducibility of a compartment-based scoring system. *Skeletal Radiol* **34**, 95-102 (2005).
27. D. J. Hunter *et al.*, The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). *Ann Rheum Dis* **67**, 206-211 (2008).
28. C. G. Peterfy *et al.*, Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. *Osteoarthritis Cartilage* **12**, 177-190 (2004).
29. F. P. Luyten, M. Denti, G. Filardo, E. Kon, L. Engebretsen, Definition and classification of early osteoarthritis of the knee. *Knee Surg Sports Traumatol Arthrosc* **20**, 401-406 (2012).
30. T. Neogi *et al.*, Magnetic resonance imaging-based three-dimensional bone shape of the knee predicts onset of knee osteoarthritis: data from the osteoarthritis initiative. *Arthritis Rheum* **65**, 2048-2058 (2013).
31. N. K. Namiri *et al.*, Deep learning for large scale MRI-based morphological phenotyping of osteoarthritis. *Sci Rep* **11**, 10915 (2021).
32. M. K. Javaid *et al.*, Pre-radiographic MRI findings are associated with onset of knee symptoms: the most study. *Osteoarthritis Cartilage* **18**, 323-328 (2010).

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/) .

33. F. W. Roemer *et al.*, What comes first? Multitissue involvement leading to radiographic osteoarthritis: magnetic resonance imaging-based trajectory analysis over four years in the osteoarthritis initiative. *Arthritis Rheumatol* **67**, 2085-2096 (2015).
34. R. J. Buck *et al.*, Osteoarthritis may not be a one-way-road of cartilage loss--comparison of spatial patterns of cartilage change between osteoarthritic and healthy knees. *Osteoarthritis Cartilage* **18**, 329-335 (2010).
35. M. E. Adams, K. D. Brandt, Hypertrophic repair of canine articular cartilage in osteoarthritis after anterior cruciate ligament transection. *J Rheumatol* **18**, 428-435 (1991).
36. E. Calvo *et al.*, High-resolution MRI detects cartilage swelling at the early stages of experimental osteoarthritis. *Osteoarthritis Cartilage* **9**, 463-472 (2001).
37. E. Calvo *et al.*, Histopathological correlation of cartilage swelling detected by magnetic resonance imaging in early experimental osteoarthritis. *Osteoarthritis Cartilage* **12**, 878-886 (2004).
38. E. Vignon, M. Arlot, D. Hartmann, B. Moyon, G. Ville, Hypertrophic repair of articular cartilage in experimental osteoarthrosis. *Ann Rheum Dis* **42**, 82-88 (1983).
39. P. J. Watson, T. A. Carpenter, L. D. Hall, J. A. Tyler, Cartilage swelling and loss in a spontaneous model of osteoarthritis visualized by magnetic resonance imaging. *Osteoarthritis Cartilage* **4**, 197-207 (1996).
40. R. B. Frobell *et al.*, The acutely ACL injured knee assessed by MRI: changes in joint fluid, bone marrow lesions, and cartilage during the first year. *Osteoarthritis Cartilage* **17**, 161-167 (2009).
41. T. G. Williams *et al.* (2006) Regional Quantitative Analysis of Knee Cartilage in a Population Study Using MRI and Model Based Correspondences. in *3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano, 2006.*, pp 311-314.
42. M. A. Bowes *et al.*, 422 Statistical Shape Modelling Reveals Focal Pattern of Cartilage Loss in Oai Progression Cohort. *Osteoarthritis and Cartilage* **16**, S183-S184 (2008).
43. S. Cotofana *et al.*, Cartilage thickening in early radiographic knee osteoarthritis: a within-person, between-knee comparison. *Arthritis Care Res (Hoboken)* **64**, 1681-1690 (2012).
44. M. P. Hellio Le Graverand *et al.*, Subregional femorotibial cartilage morphology in women--comparison between healthy controls and participants with different grades of radiographic knee osteoarthritis. *Osteoarthritis Cartilage* **17**, 1177-1185 (2009).
45. F. Caliva *et al.*, Studying osteoarthritis with artificial intelligence applied to magnetic resonance imaging. *Nat Rev Rheumatol* **18**, 112-121 (2022).
46. Y. X. Teoh *et al.*, Discovering Knee Osteoarthritis Imaging Features for Diagnosis and Prognosis: Review of Manual Imaging Grading and Machine Learning Approaches. *J Healthc Eng* **2022**, 4138666 (2022).
47. V. Pedoia, J. Lee, B. Norman, T. M. Link, S. Majumdar, Diagnosing osteoarthritis from T(2) maps using deep learning: an analysis of the entire Osteoarthritis Initiative baseline cohort. *Osteoarthritis Cartilage* **27**, 1002-1010 (2019).
48. A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, S. Saarakkala, Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. *Sci Rep* **8**, 1727 (2018).
49. R. Mahum *et al.*, A Novel Hybrid Approach Based on Deep CNN Features to Detect Knee Osteoarthritis. *Sensors (Basel)* **21** (2021).
50. P. Chen, L. Gao, X. Shi, K. Allen, L. Yang, Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. *Comput Med Imaging Graph* **75**, 84-92 (2019).
51. C. W. Yong *et al.*, Knee osteoarthritis severity classification with ordinal regression module. *Multimedia Tools and Applications* **81**, 41497-41509 (2021).
52. B. Norman, V. Pedoia, A. Noworolski, T. M. Link, S. Majumdar, Applying Densely Connected Convolutional Neural Networks for Staging Osteoarthritis Severity from Plain Radiographs. *J Digit Imaging* **32**, 471-477 (2019).

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/) .

53. K. A. Thomas *et al.*, Automated Classification of Radiographic Knee Osteoarthritis Severity Using Deep Neural Networks. *Radiol Artif Intell* **2**, e190065 (2020).
54. J. Abedin *et al.*, Predicting knee osteoarthritis severity: comparative modeling based on patient's data and plain X-ray images. *Sci Rep* **9**, 5761 (2019).
55. J. Hirvasniemi *et al.*, A machine learning approach to distinguish between knees without and with osteoarthritis using MRI-based radiomic features from tibial bone. *Eur Radiol* **31**, 8513-8521 (2021).
56. S. S. Gornale, P. U. Patravali, P. S. Hiremath, Automatic Detection and Classification of Knee Osteoarthritis Using Hu's Invariant Moments. *Front Robot AI* **7**, 591827 (2020).
57. A. Tiulpin, S. Saarakkala, Automatic Grading of Individual Knee Osteoarthritis Features in Plain Radiographs Using Deep Convolutional Neural Networks. *Diagnostics (Basel)* **10** (2020).
58. N. Lazzarini *et al.*, A machine learning approach for the identification of new biomarkers for knee osteoarthritis development in overweight and obese women. *Osteoarthritis Cartilage* **25**, 2014-2021 (2017).
59. P. Widera *et al.*, Multi-classifier prediction of knee osteoarthritis progression from incomplete imbalanced longitudinal data. *Sci Rep* **10**, 8427 (2020).
60. A. Tiulpin *et al.*, Multimodal Machine Learning-based Knee Osteoarthritis Progression Prediction from Plain Radiographs and Clinical Data. *Sci Rep* **9**, 20038 (2019).
61. L. C. Chan, H. H. T. Li, P. K. Chan, C. Wen, A machine learning-based approach to decipher multi-etiology of knee osteoarthritis onset and deterioration. *Osteoarthr Cartil Open* **3**, 100135 (2021).
62. J. C. Cheung, A. Y. Tam, L. C. Chan, P. K. Chan, C. Wen, Superiority of Multiple-Joint Space Width over Minimum-Joint Space Width Approach in the Machine Learning for Radiographic Severity and Knee Osteoarthritis Progression. *Biology (Basel)* **10** (2021).
63. E. Halilaj, Y. Le, J. L. Hicks, T. J. Hastie, S. L. Delp, Modeling and predicting osteoarthritis progression: data from the osteoarthritis initiative. *Osteoarthritis Cartilage* **26**, 1643-1650 (2018).
64. C. Huang *et al.*, DADP: Dynamic abnormality detection and progression for longitudinal knee magnetic resonance images from the Osteoarthritis Initiative. *Med Image Anal* **77**, 102343 (2022).
65. G. B. Joseph *et al.*, Tool for osteoarthritis risk prediction (TOARP) over 8 years using baseline clinical data, X-ray, and MRI: Data from the osteoarthritis initiative. *Journal of Magnetic Resonance Imaging* **47**, 1517-1526 (2018).
66. B. Guan *et al.*, Deep learning risk assessment models for predicting progression of radiographic medial joint space loss over a 48-MONTH follow-up period. *Osteoarthritis Cartilage* **28**, 428-437 (2020).
67. C. Ntakolia, C. Kokkosis, S. Moustakidis, D. Tsaopoulos, Prediction of Joint Space Narrowing Progression in Knee Osteoarthritis Patients. *Diagnostics (Basel)* **11** (2021).
68. N. Hafezi-Nejad *et al.*, Prediction of medial tibiofemoral compartment joint space loss progression using volumetric cartilage measurements: Data from the FNIH OA biomarkers consortium. *Eur Radiol* **27**, 464-473 (2017).
69. A. G. Morales *et al.*, Uncovering associations between data-driven learned qMRI biomarkers and chronic pain. *Sci Rep* **11**, 21989 (2021).
70. E. Pierson, D. M. Cutler, J. Leskovec, S. Mullainathan, Z. Obermeyer, An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nat Med* **27**, 136-140 (2021).
71. B. Guan *et al.*, Deep learning approach to predict pain progression in knee osteoarthritis. *Skeletal Radiol* **51**, 363-373 (2022).
72. J. J. Lee, F. Liu, S. Majumdar, V. Pedoia, An ensemble clinical and MR-image deep learning model predicts 8-year knee pain trajectory: Data from the osteoarthritis initiative. *Osteoarthritis Imaging* **1** (2021).

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/) .

73. V. Padoia *et al.*, 3D convolutional neural networks for detection and severity staging of meniscus and PFJ cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects. *J Magn Reson Imaging* **49**, 400-410 (2019).
74. A. A. Tolpadi, J. J. Lee, V. Padoia, S. Majumdar, Deep Learning Predicts Total Knee Replacement from Magnetic Resonance Images. *Sci Rep* **10**, 6371 (2020).
75. A. Jamshidi *et al.*, Machine Learning-Based Individualized Survival Prediction Model for Total Knee Replacement in Osteoarthritis: Data From the Osteoarthritis Initiative. *Arthritis Care Res (Hoboken)* **73**, 1518-1527 (2021).
76. H. R. Rajamohan *et al.*, Prediction of total knee replacement using deep learning analysis of knee MRI. *Scientific Reports* **13** (2023).
77. Q. Liu *et al.*, Prediction models for the risk of total knee replacement: development and validation using data from multicentre cohort studies. *The Lancet Rheumatology* **4**, e125-e134 (2022).
78. S. Heisinger, W. Hitzl, G. M. Hobusch, R. Windhager, S. Cotofana, Predicting Total Knee Replacement from Symptomology and Radiographic Structural Change Using Artificial Neural Networks—Data from the Osteoarthritis Initiative (OAI). *Journal of Clinical Medicine* **9** (2020).
79. K. Leung *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).
80. R. R. Selvaraju *et al.* (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp 618-626.
81. G. Montavon, S. Lapuschkin, A. Binder, W. Samek, K.-R. Müller, Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognition* **65**, 211-222 (2017).
82. F. Eckstein, W. Wirth, M. C. Nevitt, Recent advances in osteoarthritis imaging--the osteoarthritis initiative. *Nat Rev Rheumatol* **8**, 622-630 (2012).
83. M. J. Cardoso *et al.*, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Lecture Notes in Computer Science (2017), 10.1007/978-3-319-67558-9.
84. Z. Shen, X. Han, Z. Xu, M. Niethammer (2019) Networks for Joint Affine and Non-Parametric Image Registration. in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 4219-4228.
85. I. K. Haugen *et al.*, Comparison of features by MRI and radiographs of the interphalangeal finger joints in patients with hand osteoarthritis. *Ann Rheum Dis* **71**, 345-350 (2012).
86. M. D. Kohn, A. A. Sassoon, N. D. Fernando, Classifications in Brief: Kellgren-Lawrence Classification of Osteoarthritis. *Clin Orthop Relat Res* **474**, 1886-1893 (2016).
87. L. Wang *et al.*, T1rho MRI at 3T of menisci in patients with acute anterior cruciate ligament (ACL) injury. *J Magn Reson Imaging* **41**, 544-549 (2015).
88. C. Ding, F. Cicuttini, F. Scott, H. Cooley, G. Jones, Association between age and knee structural change: a cross sectional MRI based study. *Ann Rheum Dis* **64**, 549-555 (2005).
89. G. Cai, M. Jiang, F. Cicuttini, G. Jones, Association of age, sex and BMI with the rate of change in tibial cartilage volume: a 10.7-year longitudinal cohort study. *Arthritis Research & Therapy* **21** (2019).
90. L. Si *et al.*, Knee Cartilage Thickness Differs Alongside Ages: A 3-T Magnetic Resonance Research Upon 2,481 Subjects via Deep Learning. *Frontiers in Medicine* **7** (2021).
91. I. G. Otterness, F. Eckstein, Women have thinner cartilage and smaller joint surfaces than men after adjustment for body height and weight. *Osteoarthritis and Cartilage* **15**, 666-672 (2007).
92. M. T. Schneider, N. Rooks, T. Besier, Cartilage thickness and bone shape variations as a function of sex, height, body mass, and age in young adult knees. *Sci Rep* **12**, 11707 (2022).
93. M. Tschon, D. Contartese, S. Pagani, V. Borsari, M. Fini, Gender and Sex Are Key Determinants in Osteoarthritis Not Only Confounding Variables. A Systematic Review of Clinical Data. *J Clin Med* **10** (2021).

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/) .

94. W. Wilson *et al.*, Causes of mechanically induced collagen damage in articular cartilage. *J Orthop Res* **24**, 220-228 (2006).
95. K. Blazek, J. Favre, J. Asay, J. Erhart-Hledik, T. Andriacchi, Age and obesity alter the relationship between femoral articular cartilage thickness and ambulatory loads in individuals without osteoarthritis. *J Orthop Res* **32**, 394-402 (2014).
96. R. F. Shah *et al.*, Variation in the Thickness of Knee Cartilage. The Use of a Novel Machine Learning Algorithm for Cartilage Segmentation of Magnetic Resonance Images. *J Arthroplasty* **34**, 2210-2215 (2019).
97. A. Keng *et al.*, Association of body mass index with knee cartilage damage in an asymptomatic population-based study. *BMC Musculoskelet Disord* **18**, 517 (2017).
98. A. Connolly, D. FitzPatrick, J. Moulton, J. Lee, A. Lerner, Tibiofemoral cartilage thickness distribution and its correlation with anthropometric variables. *Proc Inst Mech Eng H* **222**, 29-39 (2008).
99. C. Guida, M. Zhang, J. Shan, Knee Osteoarthritis Classification Using 3D CNN and MRI. *Applied Sciences* **11** (2021).
100. M. W. Kattan, J. Marasco, What Is a Real Nomogram? *Seminars in Oncology* **37**, 23-26 (2010).
101. L. Jiang *et al.*, Body mass index and susceptibility to knee osteoarthritis: A systematic review and meta-analysis. *Joint Bone Spine* **79**, 291-297 (2012).
102. D. Coggon *et al.*, Knee osteoarthritis and obesity. *International Journal of Obesity* **25**, 622-627 (2001).
103. J. C. Erhart *et al.* (2008) Regional Cartilage Thinning Occurs First in the Walking Weight Bearing Regions of the Femur in Medial Compartment Knee Osteoarthritis. in *ASME 2008 Summer Bioengineering Conference, Parts A and B*, pp 835-836.
104. R. Soleymani, E. Granger, G. Fumera, F-measure curves: A tool to visualize classifier performance under imbalance. *Pattern Recognition* **100** (2020).
105. L. Shamir *et al.*, Knee X-Ray Image Analysis Method for Automated Detection of Osteoarthritis. *IEEE Transactions on Biomedical Engineering* **56**, 407-415 (2009).
106. C. R. Scanzello, S. R. Goldring, The role of synovitis in osteoarthritis pathogenesis. *Bone* **51**, 249-257 (2012).
107. S. Ashraf, D. A. Walsh, Angiogenesis in osteoarthritis. *Current Opinion in Rheumatology* **20**, 573-580 (2008).
108. M. B. Goldring, M. Otero, Inflammation in osteoarthritis. *Current Opinion in Rheumatology* **23**, 471-478 (2011).
109. W. H. Robinson *et al.*, Low-grade inflammation as a key mediator of the pathogenesis of osteoarthritis. *Nature Reviews Rheumatology* **12**, 580-592 (2016).
110. C. G. Peterfy, E. Schneider, M. Nevitt, The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthritis and Cartilage* **16**, 1433-1441 (2008).
111. A. J. Metcalfe, M. L. E. Andersson, R. Goodfellow, C. A. Thorstensson, Is knee osteoarthritis a symmetrical disease? Analysis of a 12 year prospective cohort study. *BMC Musculoskeletal Disorders* **13** (2012).
112. J. Kumm, A. Turkiewicz, F. Zhang, M. Englund, Structural abnormalities detected by knee magnetic resonance imaging are common in middle-aged subjects with and without risk factors for osteoarthritis. *Acta Orthopaedica* **89**, 535-540 (2018).
113. X. Lin, D. Zhang, Inference in Generalized Additive Mixed Models by Using Smoothing Splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **61**, 381-400 (1999).
114. D.-J. Lee, M. Durbán, P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling* **11**, 49-69 (2011).
115. S. N. Wood, *Generalized Additive Models* (2017), 10.1201/9781315370279.

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/) .

116. D. Chiba *et al.*, US detection of medial meniscus extrusion can predict the risk of developing radiographic knee osteoarthritis: a 5-year cohort study. *European Radiology* **30**, 3996-4004 (2020).
117. A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82** (2017).
118. R. A. I. Bethlehem *et al.*, Brain charts for the human lifespan. *Nature* **604**, 525-533 (2022).
119. K. J. Friston, A. Holmes, J. B. Poline, C. J. Price, C. D. Frith, Detecting Activations in PET and fMRI: Levels of Inference and Power. *NeuroImage* **4**, 223-235 (1996).
120. K. J. Friston, K. J. Worsley, R. S. J. Frackowiak, J. C. Mazziotta, A. C. Evans, Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping* **1**, 210-220 (1994).
121. C.-W. Woo, A. Krishnan, T. D. Wager, Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage* **91**, 412-419 (2014).
122. K. K. Nicodemus, J. D. Malley, C. Strobl, A. Ziegler, The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC Bioinformatics* **11** (2010).
123. M. E. Baxter, H. A. Miller, J. Chen, B. J. Williams, H. B. Frieboes, Metabolomic differentiation of tumor core versus edge in glioma. *Neurosurgical Focus* **54** (2023).
124. S. Engebretsen, J. Bohlin, Statistical predictions with glmnet. *Clinical Epigenetics* **11** (2019).
125. B. Hofner, A. Mayr, N. Robinzonov, M. Schmid, Model-based boosting in R: a hands-on tutorial using the R package mboost. *Computational Statistics* **29**, 3-35 (2012).
126. Z. Zhang, A gentle introduction to artificial neural networks. *Annals of Translational Medicine* **4**, 370-370 (2016).
127. K. Yadav, R. Thareja, Comparing the Performance of Naive Bayes And Decision Tree Classification Using R. *International Journal of Intelligent Systems and Applications* **11**, 11-19 (2019).
128. A. Gelman, A. Jakulin, M. G. Pittau, Y.-S. Su, A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* **2** (2008).
129. J. L. Speiser, M. E. Miller, J. Tooze, E. Ip, A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications* **134**, 93-101 (2019).
130. T. Chen, C. Guestrin (2016) XGBoost. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp 785-794.
131. R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, C.-J. Lin, LIBLINEAR: A library for large linear classification. *the Journal of machine Learning research* **9**, 1871-1874 (2008).
132. M. Kuhn, K. Johnson, "Discriminant Analysis and Other Linear Classification Models" in *Applied Predictive Modeling*. (2013), 10.1007/978-1-4614-6849-3_12 chap. Chapter 12, pp. 275-328.
133. C. Bouveyron, S. Girard, C. Schmid, High-Dimensional Discriminant Analysis. *Communications in Statistics - Theory and Methods* **36**, 2607-2623 (2007).
134. G. Ke *et al.*, Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* **30** (2017).
135. Y. Kim, Y. Kim, Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. *Sustainable Cities and Society* **79** (2022).