European Journal of Radiology Open 14 (2025) 100649

Contents lists available at ScienceDirect

European Journal of Radiology Open

journal homepage: www.elsevier.com/locate/ejro

Ultrasound-based radiomics and machine learning for enhanced diagnosis of knee osteoarthritis: Evaluation of diagnostic accuracy, sensitivity, specificity, and predictive value

Takeharu Kiso [a,b,*], Yukinori Okada [b,c], Satoru Kawata [d,e], Kouta Shichiji [a], Eiichiro Okumura [d], Noritaka Hatsumi [a], Ryohei Matsuura [a], Masaki Kaminaga [a], Hikaru Kuwano [a], Erika Okumura [b,f]

[a] Department of Radiology, Medical Corporation Seireikai Tachikawa Memorial Hospital, 2-12-14 Yakumo, Kasama, Ibaraki 309-1611, Japan
[b] Graduate School of Medical Sciences, Suzuka University, 1001-1, Kishioka-cho, Suzuka-shi, Mie 510-0293, Japan
[c] Tokyo Medical University Hospital, Department of Clinical Medicine, Division of Radiation Oncology, 6-7-1 Nishi-Shinjuku, Shinjuku-ku, Tokyo 160-0023, Japan
[d] Department of Radiology, Faculty of Medical and Health Sciences, Tsukuba International University, 6-20-1 Manabe, Tsuchiura-shi, Ibaraki 300-0051, Japan
[e] Postdoctoral Program, Graduate School of Health Sciences, Kyorin University, 5-4-1 Shimorenjaku, Mitaka-shi, Tokyo 181-8612, Japan
[f] Department of Radiology, Tsukuba Medical Center Hospital, 1-3-1 Amakubo, Tsukuba City, Ibaraki Prefecture 305-8558, Japan

HIGHLIGHTS

- Created a method for knee OA diagnosis and severity prediction using US radiomics.
- The statistical model achieved high specificity in excluding non-OA cases.
- Used a machine learning model with radiomics age, sex, and BMI for OA diagnosis.
- Classified mild vs. severe deformation with radiomics features post-OA diagnosis.

ARTICLE INFO

Keywords:
Knee joint
Ultrasonography
Machine learning
Osteoarthritis
Radiomics

ABSTRACT

Purpose: To evaluate the usefulness of radiomics features extracted from ultrasonographic images in diagnosing and predicting the severity of knee osteoarthritis (OA).
Methods: In this single-center, prospective, observational study, radiomics features were extracted from standing radiographs and ultrasonographic images of knees of patients aged 40–85 years with primary medial OA and without OA. Analysis was conducted using LIFEx software (version 7.2.n), ANOVA, and LASSO regression. The diagnostic accuracy of three different models, including a statistical model incorporating background factors and machine learning models, was evaluated.
Results: Among 491 limbs analyzed, 318 were OA and 173 were non-OA cases. The mean age was 72.7 (±8.7) and 62.6 (±11.3) years in the OA and non-OA groups, respectively. The OA group included 81 (25.5 %) men and 237 (74.5 %) women, whereas the non-OA group included 73 men (42.2 %) and 100 (57.8 %) women. A statistical model using the cutoff value of MORPHOLOGICAL_SurfaceToVolumeRatio (IBSI:2PR5) achieved a specificity of 0.98 and sensitivity of 0.47. Machine learning diagnostic models (Model 2) demonstrated areas under the curve (AUCs) of 0.88 (discriminant analysis) and 0.87 (logistic regression), with sensitivities of 0.80 and 0.81 and specificities of 0.82 and 0.80, respectively. For severity prediction, the statistical model using MORPHOLOGICAL_SurfaceToVolumeRatio (IBSI:2PR5) showed sensitivity and specificity values of 0.78 and 0.86, respectively, whereas machine learning models achieved an AUC of 0.92, sensitivity of 0.81, and specificity of 0.85 for severity prediction.
Conclusion: The use of radiomics features in diagnosing knee OA shows potential as a supportive tool for enhancing clinicians' decision-making.

Clinical Trial Registration: This study was registered with the UMIN Clinical Trials Registration System (Study ID: 000049395).

* Corresponding author at: Department of Radiology, Medical Corporation Seireikai Tachikawa Memorial Hospital, 2-12-14 Yakumo, Kasama, Ibaraki 309-1611, Japan.
E-mail address: rms5641@yahoo.co.jp (T. Kiso).

https://doi.org/10.1016/j.ejro.2025.100649
Received 9 December 2024; Received in revised form 21 March 2025; Accepted 26 March 2025

1. Introduction

Osteoarthritis of the knee (OA) is a chronic condition involving joint damage and deformity, involving a wide variety of factors, including age, sex, weight, and genetic and environmental factors [1–4] of the knee: a chronic disease involving joint damage and deformity [1–4]. It is a leading cause of disability in older adults, posing significant social and economic burdens and necessitating early intervention [5–7].

Although upright knee radiography is the gold standard for diagnosing knee OA [8–10], it primarily focuses on bone changes and has limited utility in assessing soft tissue. Magnetic resonance imaging (MRI) is the preferred modality for knee OA evaluation, offering detailed visualization of tissues such as cartilage, subchondral bone, and meniscus [11]. However, MRI is unsuitable for screening due to its supine positioning, which limits the assessment of functional changes, its lengthy examination times, and its high cost. Ultrasonography (US), in contrast, is short, noninvasive, and radiation-free, making it a viable option for knee OA diagnosis.

US effectively assesses soft tissue, cartilage, and bone deformities which are critical in knee OA. Saarakkala et al. [12] used ultrasonography (US) to assess knee cartilage degeneration and found a sensitivity of 0.83 for the medial condyle. Shimozaki et al. [13] reported a sensitivity and specificity of 0.88 and 0.85, respectively, for the diagnosis of medial meniscus tear, highlighting its utility in evaluating early OA dynamically. Majidi et al. [14] reported that US had a higher sensitivity of 0.74 and specificity of 0.91 than radiography in detecting osteophytes in knee OA.

In recent years, radiomics has emerged as a promising tool in diagnostic imaging. It involves extracting quantitative features from images to assist in disease diagnosis and prognosis [15–19]. In a study of radiomics analysis in knee OA, Xue et al. [20] developed a model to identify OA by radiomics analysis using MRI of tibial and femoral subchondral bone. Li et al. [21] created a nomogram model combining radiographic features and age, reporting its diagnostic accuracy. Villagran et al. [22] reported that radiomic features of the medial meniscus, extracted from MRI, could predict the development of future meniscus injuries.

However, no study has yet applied US radiomics to knee OA, and its usefulness has not been clarified. The purpose of this study was to investigate the utility of radiomics features extracted from ultrasonographic images to diagnose and predict the severity of knee OA.

2. Subjects and methods

2.1. Ethical considerations

This prospective, observational study was conducted at a single institution. Informed consent was obtained from all patients. The study was approved by the Ethical Review Committee of Tsukuba International University (No. R05-12) and was conducted in accordance with the Declaration of the Helsinki Code of Ethics. The study was registered in the UMIN Clinical Trials Registration System (UMIN-CTR) (Study ID: UMIN000049395) and was conducted in accordance with the Standards for Reporting of Diagnostic Accuracy (STARD) guidelines provided by the EQUATOR Network.

2.2. Case selection

Patients with and without OA who visited the Department of Orthopaedic Surgery at Tachikawa Memorial Hospital with knee pain between December 2022 and April 2024 and who underwent their first knee radiography and knee US during the study period were included. The following patients were excluded: a) those who underwent supine knee radiography, b) those after knee arthroplasty, c) those after anterior cruciate ligament (ACL) reconstruction, d) those who did not agree to participate in the study, e) those for whom data could not be collected due to the absence of a US examiner, f) those with external knee OA, and g) those with secondary knee joint OA (e.g., rheumatoid arthritis, gout/pseudogout, crystal-induced arthritis, idiopathic medial femoral condylar osteonecrosis). Re-measurements by the same examiner were also excluded.

Since the statistical analysis required verification of new cases, cases after November 2023 were excluded in the diagnosis using cutoff values and stored separately. On the other hand, case data from the entire period were used to build the machine learning model, and the full dataset was utilized to improve accuracy. This study includes some cases from the previous research conducted by the first author [23].

2.3. Research design

Radiomics features were utilized to develop statistical and machine learning models for evaluating the presence and severity (mild and severe) of knee OA. The study was conducted in two phases: the first phase employed statistical methods for analysis, while the second phase utilized machine learning models for evaluation. In each phase, three different models were constructed, and their diagnostic accuracy was compared. In this study, we initially planned to evaluate the diagnostic performance of a model using only radiomics features. However, our analysis revealed significant differences in age, sex, and body mass index (BMI) between the OA and non-OA groups ($p < 0.001$). Therefore, to explore the potential impact of these background factors, we conducted an additional sub-analysis by constructing a model that included these factors.

In Model 1, only radiomics features were used to assess the diagnostic performance of knee OA. In Model 2, radiomics features were combined with background factors such as age, sex, and BMI to enhance diagnostic accuracy. Finally, Model 3 was constructed using only background factors, and its diagnostic performance was evaluated. Fig. 1 provides a flow chart of the study procedure.

2.4. Criteria and methods in diagnosing knee OA

The diagnosis of knee OA was made in accordance with the European League Against Rheumatism criteria [24], integrating the patient's chief complaint, the physician's physical examination, and imaging findings. Radiographs of the knee were assessed using the Kellgren–Lawrence (KL) classification system [25,26]. Three orthopedic surgeons, with 29, 15, and 8 years of experience, respectively, diagnosed knee OA based on KL classification (grade 2 or higher). Primary medial-type knee OA, excluding secondary causes, was defined as clinical knee OA (hereafter referred to as clinical OA).

Patients with KL grade 3 or higher were categorized as the severely deformed group, whereas those with grades below 3 were classified as the mildly deformed group. Patients without evident knee OA were considered to have normal knee joints and were placed in the non-OA group. All knee OA diagnoses were determined by consensus among the three surgeons.

2.5. Standing knee radiographic and ultrasonographic examination

2.5.1. Standing knee radiography

Standing knee radiography was conducted under fluoroscopic guidance using the SONIALVISION G4 radiography system (D150BC-40S, Shimadzu Corporation, Kyoto, Japan). Patients were positioned with a footrest on a vertical fluoroscopic table, standing on one leg on the affected side. Fluoroscopy settings included pulse mode: N, pulse rate: 7.5 fps, and initial tube voltage: 50 kV. Imaging parameters were as follows: tube voltage, 60 kV; current-time product, 12.5 mAs; imaging time, 25 ms; and source-to-image distance: 150 cm.

2.5.2. Standing knee ultrasonographic examination

The standing knee US was performed on the same day as the standing

knee radiography, using an Aplio400 (TUS-A400/W1, Canon Medical Systems Inc., Tochigi, Japan) with a 12 MHz linear probe (PLT-1204AT). The imaging settings included a dynamic range of 65, mechanical index of 1.2, tissue harmonic imaging type Diff 18M, Precision 5, ApliPure 8, Tissue Specific Optimization 4, Time Smooth 4, and gamma 5. Gain, diagnostic depth, and receiver focus were adjusted for each patient. To standardize image acquisition, previous studies were referenced [23].

For the standing knee ultrasonographic examination, a long-axis image of the medial joint cleft was obtained, with the longitudinal section placed parallel to the medial collateral ligament of the knee. The medial femoral epicondyle was used as a landmark, and this position was reproducible by palpating the medial epicondyle before the examination and positioning the proximal end of the probe accordingly. Special care was taken to adjust the angle and position of the probe to ensure the medial collateral ligament was clearly delineated.

2.6. Ultrasonographic image acquirer experience and diagnostic competence in clinical OA diagnosis and severity prediction

US images of the standing knee joint were obtained by two sonographers certified by the Japanese Society of Ultrasonography with 17 and 10 years of ultrasonographic examination experience. The ultrasonographic images obtained by these examiners were compared with the diagnostic results obtained by the three orthopedic surgeons. Diagnostic accuracy was evaluated based on diagnostic rate, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

2.7. Segmentation methods in radiomics feature extraction

Image segmentation was performed using LIFEx software (version 7.2.n), an open-source platform known for its reliability and validity, widely used in radiomics analysis and specialized for medical imaging data [27–30]. Fig. 2 illustrates the image segmentation procedure for radiomics feature extraction. Segmentation was carried out manually on the high-echo lines of the medial meniscus tissue, osteophytes, and cortical portions of the femur and tibia.

The meniscus was segmented first, ensuring that the medial collateral ligament was not included. Next, the high-echoic lines of the femoral and tibial cortex were segmented, ensuring the inclusion of the proximal and distal boundaries where the high-echoic lines met. If osteophytes were present, they were included in the segmentation as well. The segmentation of the region of interest was performed manually by two examiners who conducted the standing knee ultrasonographic examination after consultation. Both examiners were blinded to the knee OA diagnostic results and independently evaluated the segmentation.

2.8. Evaluation of reproducibility of image segmentation methods

To assess the reliability of radiomics feature extraction, inter- and intra-examiner agreement was evaluated in 30 randomly selected cases. Two examiners independently extracted features, and each examiner repeated the extraction on the same dataset. The intraclass correlation coefficient (ICC) was calculated to evaluate both inter- and intra-examiner agreement for each feature, using the (1, 1) and (2, 1) models.

2.9. Selection of radiomics features used and analysis methods

The features listed in the Image Biomarker Standardisation Initiative (IBSI) Reference Manual were selected.

2.10. Feature selection

2.10.1. Feature selection and evaluation in statistical methods

Feature selection was performed using one-factor analysis of variance (ANOVA), a filter-type feature selection algorithm.

2.10.2. Feature selection and accuracy evaluation by least absolute shrinkage and selection operator (LASSO) in machine learning models

LASSO was used for feature selection. The optimal penalty parameter (λ) was determined through 10-part cross-validation ($k = 10$). Based on the optimal λ, features with non-zero regression coefficients were selected, and logistic regression models were constructed using these features. To evaluate the prediction accuracy, residual plots were created to check the fit of the model. Based on the results, the model fit and error trends were evaluated.

2.11. Building machine learning models

Two models were constructed: one for classifying the presence or absence of clinical OA and another for classifying the severity of clinical OA as either severe or mild. The dataset was randomly divided into two groups and cross-validated five times for each model. Seventy percent of

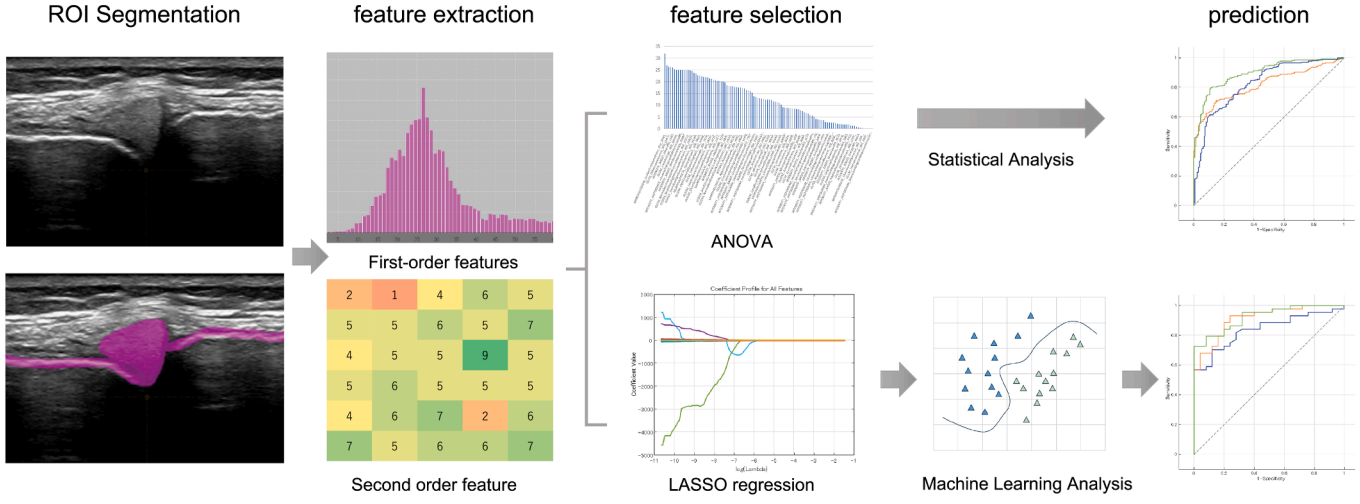

Fig. 1. Flowchart of the research procedure. ROI Segmentation: Image segmentation is performed on knee US images to extract radiomics features. Feature Extraction: Extracted radiomics features include shape features, statistical pixel value features, and histogram features. Feature Selection: ANOVA is used for statistical analysis, and LASSO is used for machine learning models. Prediction: diagnostic accuracy is evaluated and compared using area under the curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F value as indices.

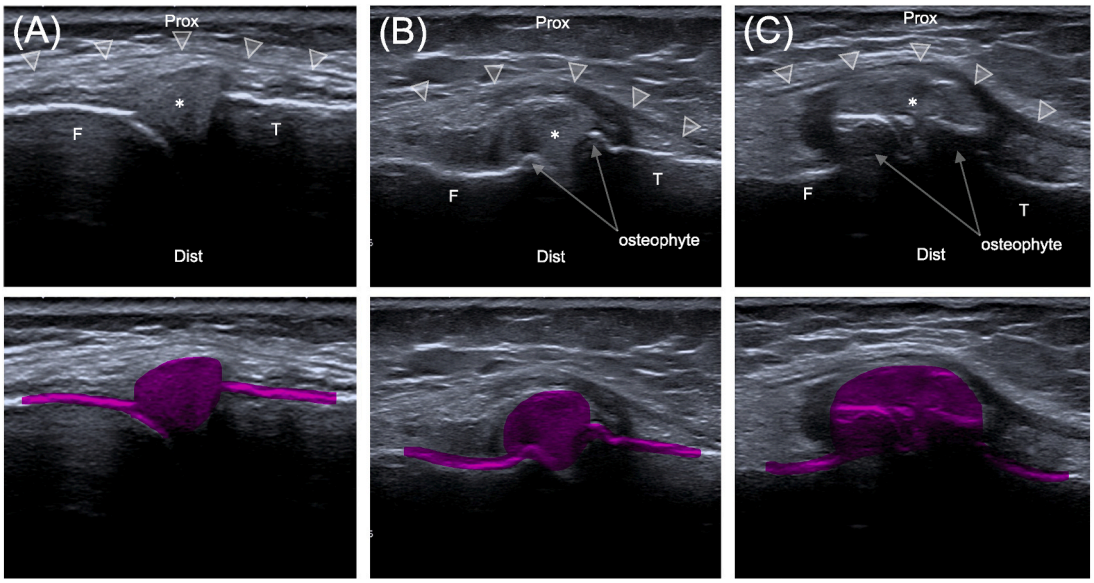

Fig. 2. Specific examples of segmentation performed for radiomics feature extraction. A: non-OA (KL1), B: clinical OA (KL2), C: severe OA (KL4); F: femur, T: tibia, Prox: proximal, Dist: distal, *: medial meniscus, △: medial collateral ligament.

the data was used as the training set, and the remaining 30 % served as the test set. Additionally, 20 % of the training set was further split for validation to optimize the models. In this study, we constructed a knee OA diagnostic model using classification algorithms within supervised learning. Specifically, we applied the following classification methods: Random Forest (RF), an ensemble learning method that combines multiple decision trees for classification; Discriminant Analysis (DA), which models class distributions and classifies data using linear or quadratic boundaries; Logistic Regression (LR), which estimates output probabilities using a sigmoid function for classification; Naive Bayes Classifier (NBC), a probabilistic model based on prior probabilities and conditional probabilities of features; Support Vector Machine (SVM), which learns a hyperplane that maximizes the margin between classes for classification; and K-Nearest Neighbors (KNN), which classifies data based on a majority vote of the k-nearest neighbors based on similarity. Machine learning was performed using MATLAB R2023a (MathWorks Inc., Sherborn, MA, USA). Details of the parameters for each model used are provided in Appendix A.

2.12. Statistical analysis

The machine learning models were evaluated using the following metrics: area under the curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F value. Receiver operating characteristic (ROC) curves for the validation and test data are shown, and the AUC and ROC curve p-values were used to compare the classification performance of each model. Significant differences in ROC curves between models were assessed using the DeLong test. The sample size was determined through power analysis based on the ROC analysis, with an expected AUC of 0.85, a null hypothesis AUC of 0.5, a significance level (α) of 0.05, and a power of 0.80. Consequently, the minimum required sample size was calculated to be 64 cases. However, in machine learning models, a small dataset increases the risk of overfitting, emphasizing the need for an adequate sample size. Therefore, we references previous studies [18,29,31–33] to determine an appropriate dataset size for the knee OA diagnostic model.

In the statistical analysis, Kolmogorov–Smirnov and F tests were performed to check for normality and variance of each distribution. If the distributions were normal and equally distributed, the Student's *t*-test was used to test for group differences. For categorical data analysis, the chi-square test or Fisher's exact test was applied based on the expected frequency. Youden's Index was used to determine cutoff values. Statistical analysis was performed using EZR (R version 2.4–0), with statistical significance defined as $p < 0.05$.

3. Results

3.1. Cases

Fig. 3 illustrates the flow chart of the study methodology: 872 limbs from 689 cases were initially selected. From these, patients who underwent supine knee radiography (162 cases, 191 limbs), those who had knee replacement (64 cases, 85 limbs), patients post-reconstructive surgery (6 cases, 6 limbs), patients who did not consent to participate (17 cases, 22 limbs), and patients whose data could not be collected due to the unavailability of a ultrasonographic examiner (51 cases, 69 limbs) were excluded prior to the ultrasonographic examination. The remaining 389 patients with 499 limbs underwent ultrasonographic examination. Additionally, patients with lateral-type knee OA (7 patients, 8 limbs) were excluded after the ultrasonographic examination. In the end, 491 limbs from 382 patients were included. Of these, 318 limbs were included in the clinical OA group and 173 limbs in the non-OA group. Statistical analysis confirmed normal distribution and equal variance for age and BMI; therefore, the Student's *t*-test was used for comparisons between groups. The chi-square test was applied for sex distribution as the expected frequency criteria were met. The mean age in the clinical OA group was 72.7 ± 8.7 years and that in the non-OA group was 62.6 ± 11.3 years, showing a statistically significant age difference ($p < 0.001$). In the clinical OA group, 81 limbs (25.5 %) were of male patients and 237 limbs (74.5 %) were of female patients, whereas in the non-OA group, 73 limbs (42.2 %) were of male patients and 100 limbs (57.8 %) were of female patients, reflecting a significant difference in sex distribution ($p < 0.001$). The mean BMI for the clinical OA group was 24.84 ± 3.89 kg/m^2, compared to 23.47 ± 3.33 kg/m^2 in the non-OA group, with a statistically significant difference ($p < 0.001$). Table 1 presents a comparison of patient backgrounds.

3.2. Intra- and inter-inspector agreement assessment (specific ICC results are shown in Appendix B)

3.2.1. Intra-inspector (1, 1) agreement (inspectors A and B)

Examiner A showed an ICC of 0.90 or higher for 93 features and an

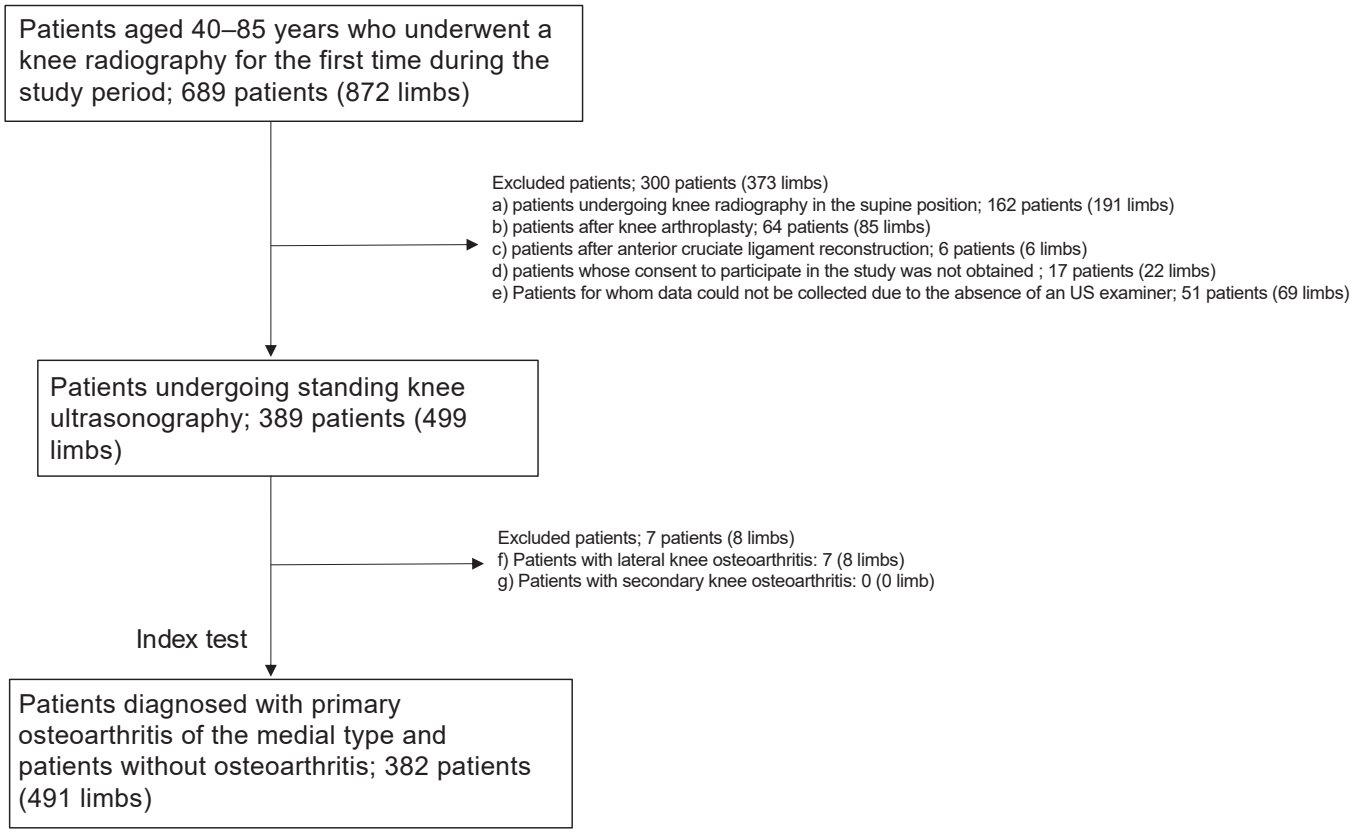

Fig. 3. Flowchart of case selection. ACL, anterior cruciate ligament; OA, knee osteoarthritis; US, ultrasonography.

ICC between 0.75 and 0.89 for 15 features. No features had an ICC between 0.50 and 0.74 or < 0.50. Examiner B showed an ICC of ≥ 0.90 for 73 features, an ICC between 0.75 and 0.89 for 33 features, and an ICC between 0.50 and 0.74 for 2 features. No features had an ICC less than 0.50.

3.2.2. *Inter-inspector (2, 1) agreement*

Eighty-nine features showed an ICC of ≥ 0.90, ten features showed an ICC of 0.75–0.89, eight features had an ICC of 0.50–0.74, and one feature had an ICC of < 0.50.

3.3. *Comparison of diagnostic performance of ultrasound image acquirers in clinical OA diagnosis and severity prediction*

In diagnosing the presence or absence of clinical OA, examiner A had a diagnostic rate of 0.86, sensitivity of 0.93, specificity of 0.77, PPV of 0.82, and NPV of 0.86, whereas examiner B had a diagnostic rate of 0.89, sensitivity of 0.87, specificity of 0.92, PPV of 0.93, and NPV of 0.86. In predicting the severity of clinical OA, examiner A had a diagnostic rate of 0.86, sensitivity of 0.93, specificity of 0.77, PPV of 0.82, and NPV of 0.91, whereas examiner B had a diagnostic rate of 0.89, sensitivity of 0.87, specificity of 0.92, PPV of 0.93, and NPV of 0.86.

3.4. *Radiomics features selected by LIFEx software*

A total of 108 radiomics features were extracted. The features consisted of 11 shape features, 19 statistical pixel-value features, 23 histogram features, and 55 texture features. The texture features include 23 gray-level co-occurrence matrices, 11 gray-level run-length matrices, 5 neighborhooding gray-level tone difference matrix, and 16 gray-level size zone matrices (GLSZM). A breakdown of these features is shown in Table 2. Details of the 108 radiomics features are provided in Appendix C.

3.5. *Comparison of diagnostic accuracy in predicting the presence and severity of clinical OA using statistical models*

3.5.1. *Diagnostic accuracy for identifying the presence or absence of clinical OA*

Comparison of patient background based on the presence or absence of clinical OA

Notably, 245 limbs were classified into the clinical OA group and 129 limbs were categorized into the non-OA group. The mean age in the clinical OA group was 73.3 ± 8.6 years and that in the non-OA group was 64.9 ± 10.9 years. Notably, 61 limbs (24.9 %) in the clinical OA

Table 1
Comparison of background of patients in the clinical OA and non-OA groups.

Variables	Clinical OA (n = 318)	Non-OA (n = 173)	p-value	All patients (n = 491)
Age	72.7 ± 8.7	62.6 ± 11.3	< 0.001	69.1 ± 8.3
Sex			< 0.001	
Male, n (%)	81 (25.5 %)	73 (42.2 %)		154(31.4 %)
Female, n (%)	237 (74.5 %)	100(57.8 %)		337(68.6 %)
BMI	24.84 ± 3.89	23.47 ± 3.33	< 0.001	24.36 ± 4.94

Data are expressed as mean ± standard deviation or value (%). BMI, body mass index; OA, osteoarthritis.

Table 2
Overview of selected features by category.

Category	Number of features	Description
Shape features	11 pieces	Features related to shape
Statistical features of pixel values	19 pieces	Statistical distribution of pixel values
Histogram features	23 pieces	Features of pixel intensity distribution
Texture features	55 pcs.	Texture-related features
GLCM	23 pieces	Measure co-occurrence between different gray levels
GLRLM	11 pieces	Length of consecutive identical gray levels
NGTDM	5 pieces	Tone difference between adjacent pixels
GLSZM	16 pieces	Size of the same gray level area

Abbreviations: GLCM, gray-level co-occurrence matrices; GLRLM, gray-level run-length matrices; NGTDM, neighborhooding gray-level tone difference matrix; GLSZM, gray-level size zone matrices.

group were of male patients and 184 limbs (75.1 %) were of female patients, whereas 55 limbs (42.6 %) in the non-OA group were of male patients and 74 limbs (57.4 %) were of female patients. The mean BMI of the clinical OA and non-OA groups was 24.96 ± 4.05 kg/m^2 and 23.73 ± 3.43 kg/m^2, respectively.

Selection results of important features by filter type feature selection algorithm

The ranking results showed that the shape feature MORPHOLOGICAL_SurfaceToVolumeRatio(IBSI:2PR5) had the lowest p-value ($p = 1.7 \times 10^{-26}$). The MORPHOLOGICAL_SurfaceToVolumeRatio (IBSI:2PR5) in the clinical OA group was 2.0442 ± 0.0126 and 2.0576 ± 0.0061 in the non-OA group, showing a statistically significant difference ($p < 0.001$). The results of the filter-type feature selection algorithm are described in detail in Supplementary Appendix D.

Predictive accuracy of clinical OA diagnostic models: comparison of radiomics features and background factors (shown in Fig. 4)

Diagnostic performance of clinical OA diagnostic Model 1 (radiomics features only)

Model 1 (using only radiomics features) showed an AUC of 0.81 (95 % CI: 0.77–0.86), sensitivity of 0.71 and specificity of 0.85.

Diagnostic performance of Clinical OA Diagnostic Model 2 (Radiomics + age, sex, BMI)

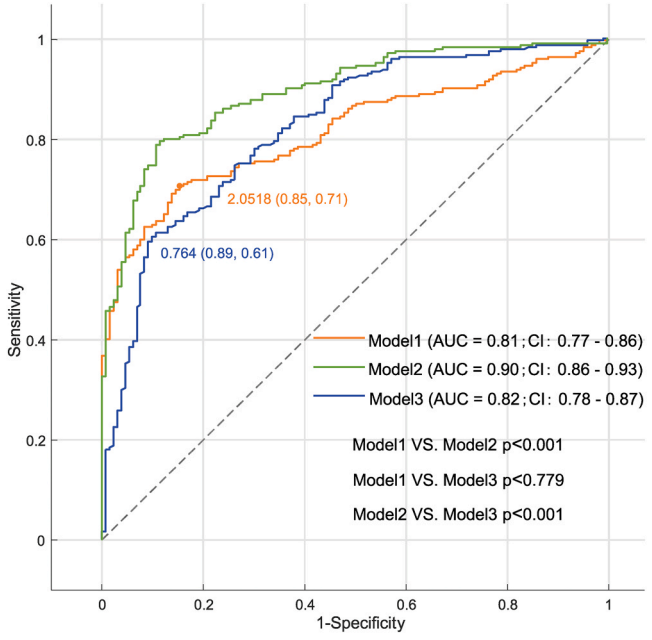

Fig. 4. Comparison of ROC curves of statistical models using radiomics features and background factors for clinical OA diagnosis. Model 1 (using only radiomics features): Uses radiomics features extracted from images. Model 2 (Radiomics + use of background factors): In addition to Radiomics, this model combines age, sex, and BMI. Model 3 (using only background factors): model using only age, sex, and BMI.

Model 2 (radiomics features + age, sex, BMI) showed an AUC of 0.90 (95 % CI: 0.86–0.93), sensitivity of 0.80 and specificity of 0.88. A statistically significant difference in diagnostic accuracy was observed when it compared with Model 1 ($p < 0.001$).

Diagnostic performance of Clinical OA Diagnostic Model 3 (age, sex, and BMI only)

Model 3 (age, sex, BMI) showed an AUC of 0.82 (95 % CI: 0.78–0.87), sensitivity 0.61 and specificity 0.89. No statistically significant difference was found in comparison with Model 1 ($p = 0.779$). On comparison with Model 2, it showed a statistically significant difference in diagnostic accuracy ($p < 0.001$).

Validation results for cutoff values based on MORPHOLOGICAL_SurfaceToVolumeRatio(IBSI:2PR5)

Applying the cutoff value "≤ 2.0518" resulted in a sensitivity of 0.47, specificity of 0.98, PPV of 0.97, NPV of 0.52, and diagnostic rate of 0.66. The 2 × 2 table showed 34 true positive (TP), 39 false negative (FN), 43 true negative (TN) and 1 false positive (FP) results.

3.5.2. Diagnostic accuracy for predicting the severity of clinical OA

Comparison of patient backgrounds based on severity of clinical OA

In the subset of patients with clinical OA (n = 245), 94 limbs were included in the severely deformed group and 151 limbs in the mildly deformed group. Statistical analysis confirmed normal distribution and equal variances for age and BMI; therefore, the Student's *t*-test was used to test for differences between groups. The chi-square test was applied for sex, as the criteria for expected frequencies were met. The mean age in the severely deformed group was 73.9 ± 7.8 years and in the mildly deformed group was 72.9 ± 9.0 years, with no statistically significant difference in age between the groups ($p = 0.380$). In the severely deformed group, 22 limbs (23.4 %) were of male patients and 72 limbs (76.6 %) were of female patients, whereas in the mildly deformed group, 39 limbs (25.8 %) were of male patients and 112 limbs (74.2 %) were of female patients, showing no statistically significant difference in sex distribution ($p = 0.789$). The mean BMI of the severely deformed group was 25.79 ± 4.25 kg/m^2, compared to 24.45 ± 3.85 kg/m^2 in the mildly deformed group, with a statistically significant difference ($p = 0.001$).

Selection results of important features by filter type feature selection algorithm

The ranking results showed that the shape feature MORPHOLOGICAL_SurfaceToVolumeRatio(IBSI:2PR5) had the lowest p-value ($p = 8.4 \times 10^{-27}$). The MORPHOLOGICAL_SurfaceToVolumeRatio (IBSI:2PR5) in the severely deformed group was 2.0345 ± 0.0093, and 2.0502 ± 0.0104 in the mildly deformed group, showing a statistically significant difference. ($p < 0.001$). The results of the filter-type feature selection algorithm are described in detail in Appendix E.

Predictive accuracy of the clinical OA severity prediction model: comparison of radiomics features and background factors (shown in Fig. 5)

Diagnostic performance of clinical OA severity prediction

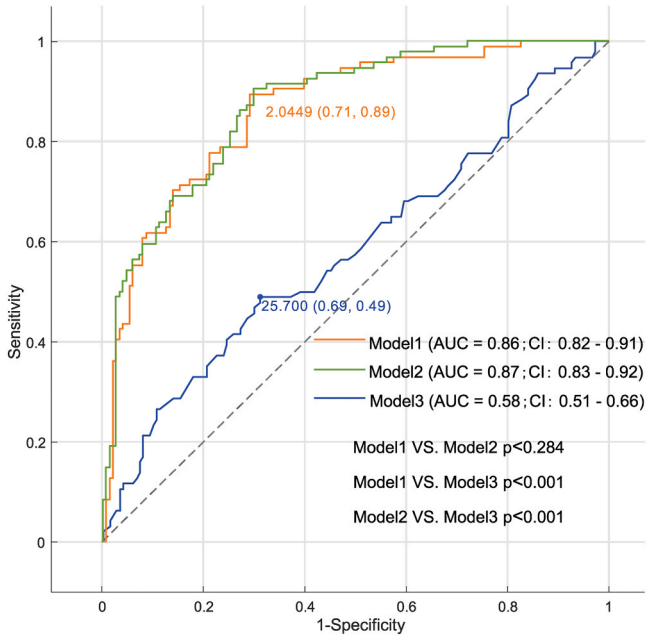

Fig. 5. Comparison of ROC curves of statistical models using radiomics features and background factors in predicting OA severity. Model 1 (using only radiomics features): Uses radiomics features extracted from images. Model 2 (Radiomics + background factor use): A model that combines BMI in addition to Radiomics. Model 3 (using only background factors): Model using only BMI.

model 1 (radiomics features only)

Model 1 (using only radiomics features) showed an AUC of 0.86 (95 % CI: 0.82–0.91), sensitivity of 0.89, and specificity of 0.71.

Diagnostic Performance of Clinical OA Severity Prediction Model 2 (Radiomics + BMI)

Model 2 (radiomics features + BMI) showed an AUC of 0.87 (95 % CI: 0.83–0.92), sensitivity of 0.90 and specificity of 0.70. Compared to Model 1, there was no statistically significant difference in diagnostic accuracy. ($p = 0.284$).

Diagnostic performance of clinical OA severity prediction model 3 (BMI only)

Model 3 (using only BMI) had an AUC of 0.58 (95 % CI: 0.51–0.66), sensitivity of 0.49 and specificity of 0.69. A statistically significant difference in diagnostic accuracy was observed when compared to Model 1 and Model 2 ($p < 0.001$).

Validation results for cutoff values based on MORPHOLOGICAL_SurfaceToVolumeRatio(IBSI:2PR5)

Using the cutoff value "≤2.0449," the sensitivity was 0.78, specificity 0.86, PPV 0.72, NPV 0.90, and diagnostic rate 0.84. The results in the 2 × 2 table were 18 true positives (TP), 5 false negatives (FN), 43 true negatives (TN), and 7 false positives (FP).

3.6. Comparison of diagnostic accuracy in predicting the presence and severity of clinical OA using machine learning models

3.6.1. LASSO regression results for the presence of clinical OA

Fig. 6A shows the change in mean squared error (MSE) based on the cross-validation results of the LASSO regression. Based on this result, 0.05706 was selected as the optimal λ value for the 1-SE criterion. Fig. 6B shows the change in the coefficient profile of each feature in the LASSO regression, where it is observed that as λ increases, the coefficients of many features converge to zero, and only important features retain their nonzero coefficients. Finally, three features were selected: MORPHOLOGICAL_SurfaceToVolumeRatio(IBSI:2PR5) (regression coefficient - 10.7651), INTENSITY-HISTOGRAM_IntensityHistogram10thPercentile (IBSI:GPMT) (regression coefficient - 0.0042), and GLSZM_HighGrayLevelZoneEmphasis(IBSI:5GN9) (regression coefficient - 0.0002).

Predictive accuracy of clinical OA diagnostic models: comparison of radiomics features and background factors (detailed in Table 3 and Fig. 7)

Diagnostic performance of clinical OA diagnostic model 1 (radiomics features only)

AUC 0.79, sensitivity 0.71, specificity 0.82, F value 0.79 for RF.
AUC 0.76, sensitivity 0.66, specificity 0.76, F value 0.74 for DA.
AUC 0.76, sensitivity 0.67, specificity 0.70, F value 0.73 for LR.
AUC 0.80, sensitivity 0.62, specificity 0.90, F value 0.74 for NBC.
AUC 0.74, sensitivity 0.55, specificity 0.92, F value 0.69 for SVM.
AUC 0.77, sensitivity 0.66, specificity 0.80, F value 0.75 for KNN.

Diagnostic performance of Clinical OA Diagnostic Model 2

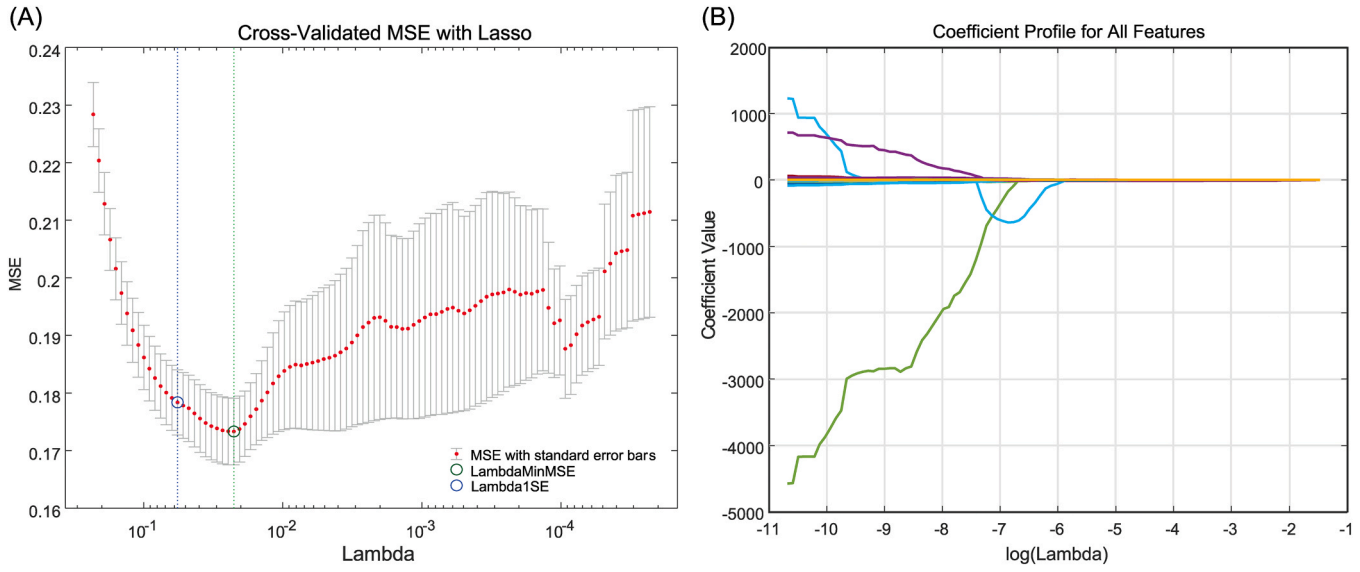

Fig. 6. Results of LASSO regression for clinical OA diagnosis. (A) shows the variation of the mean squared error (MSE) based on the cross-validation results of the LASSO regression. The horizontal axis represents the regularization parameter λ and the vertical axis represents the MSE obtained from cross-validation. The blue circle is the λ of the minimum MSE (LambdaMinMSE) and the green circle is the λ of the 1-SE criterion (Lambda1SE). (B) shows the variation of the coefficients of each feature in the Lasso regression. The horizontal axis represents the log scale of λ (log(λ)), and the vertical axis represents the coefficient value of each feature.

Table 3
Comparison of diagnostic accuracy for each clinical OA diagnostic model based on validation and test data of machine learning models.

Validation data (clinical OA Model 1)							Test data (clinical OA Model 1)						
Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F-value	Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F-value
RF	0.91	0.80	0.77	0.84	0.89	0.68	0.83	RF	0.79	0.75	0.71	0.82	0.88	0.59	0.79
DA	0.82	0.72	0.73	0.72	0.82	0.60	0.77	DA	0.76	0.69	0.66	0.76	0.84	0.54	0.74
LR	0.82	0.71	0.73	0.68	0.80	0.59	0.76	LR	0.76	0.68	0.67	0.70	0.81	0.52	0.73
NBC	0.86	0.74	0.64	0.92	0.93	0.59	0.76	NBC	0.80	0.71	0.62	0.90	0.92	0.55	0.74
SVM	0.76	0.67	0.55	0.88	0.89	0.52	0.68	SVM	0.74	0.67	0.55	0.92	0.93	0.51	0.69
KNN	0.83	0.70	0.64	0.80	0.85	0.56	0.73	KNN	0.77	0.71	0.66	0.80	0.86	0.55	0.75
Validation (Clinical OA Model 2)								Test data (clinical OA Model 2)							
Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F-value	Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F-value
RF	0.92	0.81	0.82	0.80	0.88	0.71	0.85	RF	0.86	0.79	0.77	0.82	0.89	0.65	0.83
DA	0.88	0.83	0.82	0.84	0.90	0.72	0.86	DA	0.88	0.81	0.80	0.82	0.90	0.68	0.85
LR	0.88	0.81	0.82	0.80	0.88	0.71	0.85	LR	0.87	0.81	0.81	0.80	0.89	0.69	0.85
NBC	0.89	0.80	0.73	0.92	0.94	0.66	0.82	NBC	0.89	0.73	0.64	0.90	0.93	0.56	0.76
SVM	0.87	0.81	0.80	0.84	0.90	0.70	0.85	SVM	0.86	0.79	0.77	0.82	0.89	0.65	0.83
KNN	0.88	0.78	0.89	0.60	0.80	0.75	0.84	KNN	0.89	0.80	0.90	0.62	0.82	0.76	0.86
Validation (Clinical OA Model 3)								Test data (clinical OA Model 3)							
Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F-value	Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F-value
RF	0.84	0.78	0.84	0.68	0.82	0.71	0.83	RF	0.82	0.77	0.86	0.60	0.81	0.68	0.83
DA	0.80	0.77	0.86	0.60	0.79	0.71	0.82	DA	0.83	0.78	0.88	0.60	0.81	0.71	0.84
LR	0.80	0.74	0.82	0.60	0.78	0.65	0.80	LR	0.83	0.78	0.90	0.56	0.80	0.74	0.85
NBC	0.78	0.68	0.86	0.36	0.70	0.60	0.77	NBC	0.83	0.80	0.97	0.48	0.78	0.89	0.86
SVM	0.80	0.78	0.84	0.68	0.82	0.71	0.83	SVM	0.80	0.76	0.85	0.60	0.80	0.67	0.82
KNN	0.77	0.71	0.95	0.28	0.70	0.78	0.81	KNN	0.81	0.71	0.97	0.20	0.70	0.77	0.81

(Radiomics + age, sex, BMI)

AUC 0.86, sensitivity 0.77, specificity 0.82, F value 0.83 for RF.
AUC 0.88, sensitivity 0.80, specificity 0.82, F value 0.85 for DA.
AUC 0.87, sensitivity 0.81, specificity 0.80, F value 0.85 for LR.
AUC 0.89, sensitivity 0.64, specificity 0.90, F value 0.76 for NBC.
AUC 0.86, sensitivity 0.77, specificity 0.82, F value 0.83 for SVM.
AUC 0.89, sensitivity 0.90, specificity 0.62, F value 0.86 for KNN.

Diagnostic performance of Clinical OA Diagnostic Model 3 (age, sex, and BMI only)

AUC 0.82, sensitivity 0.86, specificity 0.60, F value 0.83 for RF.
AUC 0.83, sensitivity 0.88, specificity 0.60, F value 0.84 for DA.
AUC 0.83, sensitivity 0.90, specificity 0.56, F value 0.85 for LR.
AUC 0.83, sensitivity 0.97, specificity 0.48, F value 0.86 for NBC.
AUC 0.80, sensitivity 0.85, specificity 0.60, F value 0.82 for SVM.
AUC 0.81, sensitivity 0.97, specificity 0.20, F value 0.81 for KNN.

3.6.2. Diagnostic accuracy for the severity of clinical OA

Comparison of patient backgrounds based on severity of clinical OA

In the subset of clinical OA patients (n = 318), the severely deformed group had 117 limbs, and the mildly deformed group had 201 limbs. Statistical analysis confirmed normal distribution and equal variances for age and BMI, so Student's *t*-test was used to test for differences between groups. The chi-square test was applied for sex, as the expected frequencies met the criteria. The mean age of the severely deformed group was 73.7 ± 8.5 years, and the mean age of the mildly deformed group was 72.1 ± 10.3 years, with no statistically significant difference in age (p = 0.121). In terms of sex distribution, 29 limbs (24.8 %) in the severely deformed group were of male patients, and 88 limbs (75.2 %) were of female patients, whereas 52 limbs (25.9 %) in the mildly deformed group were of male patients, and 149 limbs (74.1 %) were of female patients, with no statistically significant difference (p = 0.936). The mean BMI of the severely deformed group was 25.43 ± 4.09 kg/m^2, compared to 24.50 ± 3.74 kg/m^2 in the mildly deformed group, showing a statistically significant difference (p = 0.039).

Results of LASSO regression on clinical OA severity

Fig. 8A shows the change in MSE based on the cross-validation results of the LASSO regression. From this, 0.10079 was selected as the optimal λ value using the 1-SE criterion. Fig. 8B illustrates the change in the coefficient profile of each feature in the LASSO regression. As λ increases, the coefficients of many features converge to zero, with only the important features retaining nonzero coefficients. Ultimately, the following two features were selected: MORPHOLOGICAL_SurfaceToVolumeRatio (IBSI:2PR5) (regression coefficient −15.7344) and INTENSITY-HISTOGRAM_IntensityHistogramMean (IBSI:X6K6) (regression coefficient −0.0018).

Predictive accuracy of the clinical OA severity prediction model: comparison of radiomics features and background factors (detailed in Table 4 and Fig. 9)

Diagnostic performance of clinical OA severity prediction Model 1 (radiomics features only)

AUC 0.92, sensitivity 0.64, specificity 0.92, F value 0.82 for RF.
AUC 0.92, sensitivity 0.81, specificity 0.85, F value 0.78 for DA.
AUC 0.92, sensitivity 0.75, specificity 0.86, F value 0.76 for LR.
AUC 0.90, sensitivity 0.78, specificity 0.83, F value 0.76 for NBC.
AUC 0.91, sensitivity 0.56, specificity 0.97, F value 0.69 for SVM.
AUC 0.90, sensitivity 0.75, specificity 0.88, F value 0.77 for KNN.

Diagnostic performance of clinical OA severity prediction Model 2 (radiomics + BMI)

AUC 0.91, sensitivity 0.67, specificity 0.92, F value 0.74 for RF.
AUC 0.92, sensitivity 0.78, specificity 0.85, F value 0.77 for DA.
AUC 0.92, sensitivity 0.72, specificity 0.85, F value 0.73 for LR.
AUC 0.90, sensitivity 0.78, specificity 0.81, F value 0.75 for NBC.
AUC 0.89, sensitivity 0.53, specificity 0.93, F value 0.64 for SVM.
AUC 0.91, sensitivity 0.72, specificity 0.90, F value 0.76 for KNN.

Diagnostic performance of clinical OA severity prediction

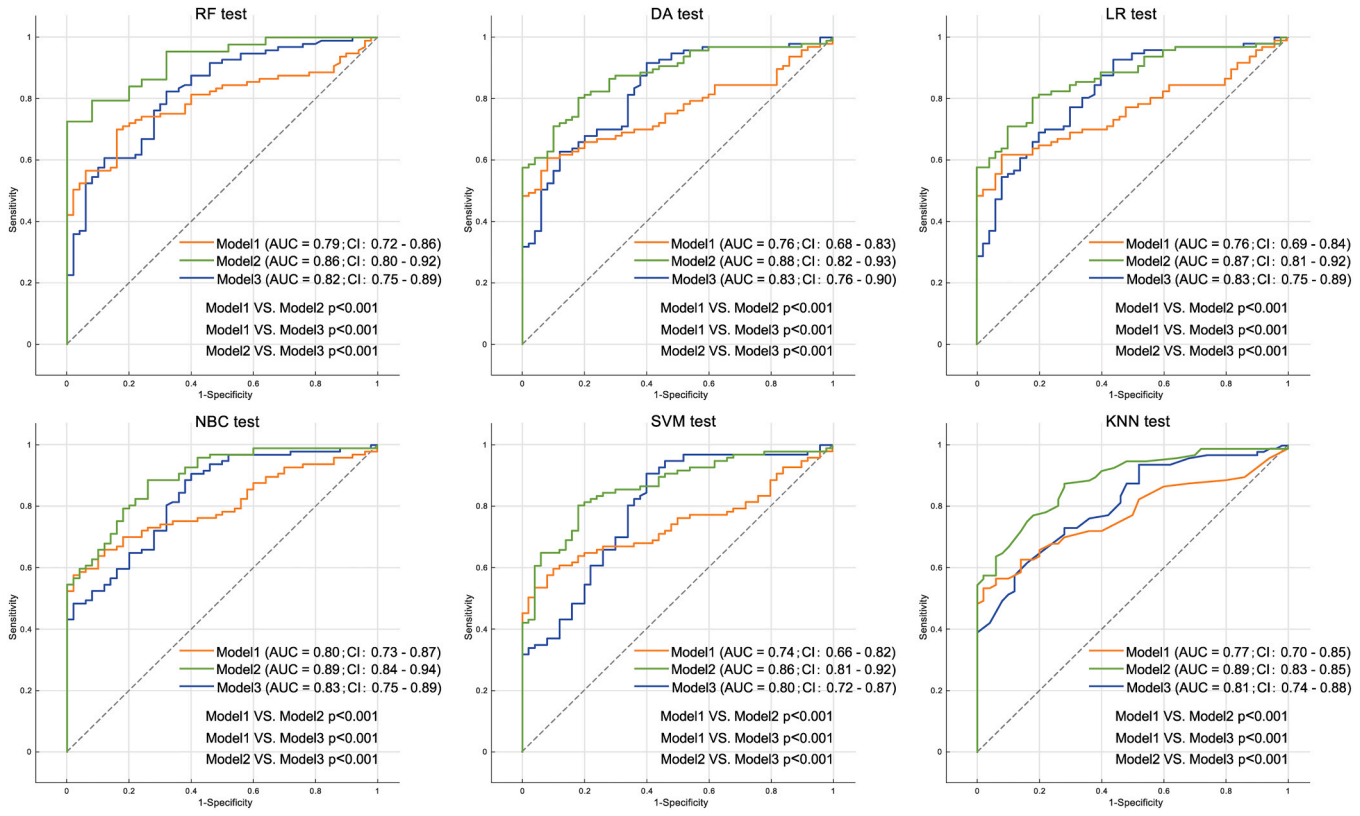

Fig. 7. ROC curve comparison of machine learning models with radiomics features and background factors for clinical OA diagnosis (test data). Model 1 (using only radiomics features): Uses radiomics features extracted from images. Model 2 (Radiomics + use of background factors): In addition to Radiomics, this model combines age, sex, and BMI. Model 3 (using only background factors): model using only age, sex, and BMI. RF, Random Forest; DA, Discriminant Analysis; LR, Logistic Regression; NBC, Naive Bayes Classifier; SVM, Support Vector Machine; KNN; K-Nearest Neighbors.

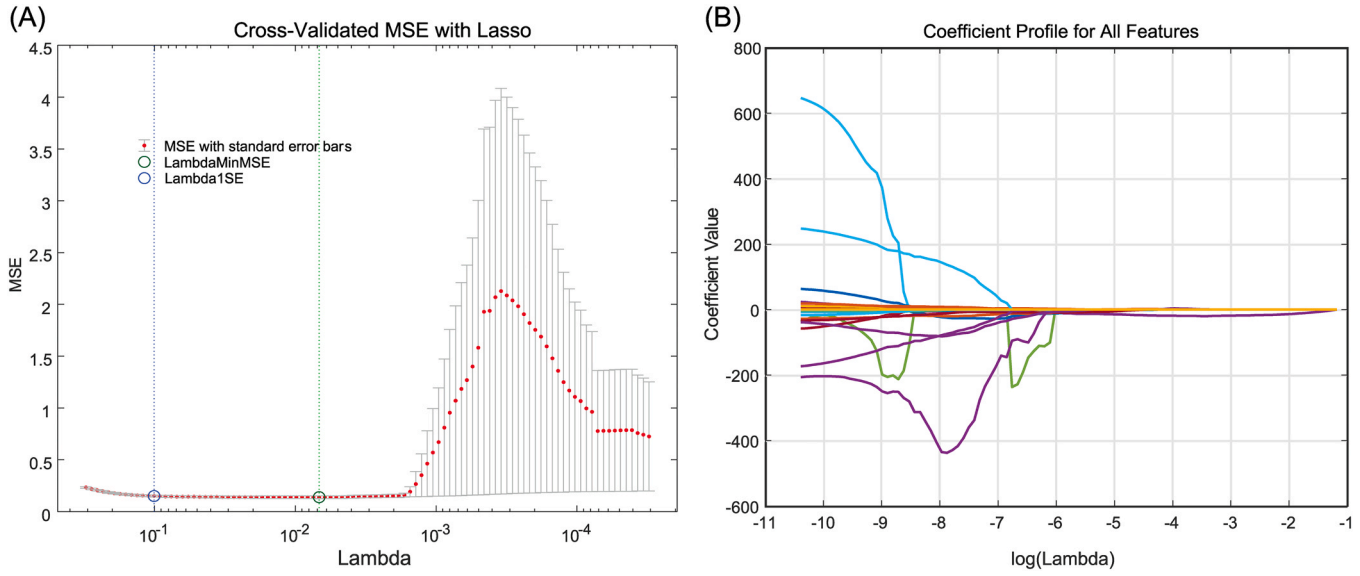

Fig. 8. Results of LASSO regression in predicting OA severity. (A) shows the variation of the mean squared error (MSE) based on the cross-validation results of the LASSO regression. The horizontal axis represents the regularization parameter λ and the vertical axis represents the MSE obtained from cross-validation. The blue circle is the λ of the minimum MSE (LambdaMinMSE) and the green circle is the λ of the 1-SE criterion (Lambda1SE). (B) shows the variation of the coefficients of each feature in the Lasso regression. The horizontal axis represents the log scale of λ (log(λ)), and the vertical axis represents the coefficient value of each feature.

Model 3 (BMI only)

AUC 0.51, sensitivity 0.11, specificity 0.92, F value 0.18 for RF.
AUC 0.53, sensitivity 0.06, specificity 1.00, F value 0.11 for DA.
AUC 0.53, sensitivity 0.06, specificity 1.00, F value 0.11 for LR.
AUC 0.52, sensitivity 0.00, specificity 1.00, F value NAN for NBC.
AUC 0.60, sensitivity 0.17, specificity 0.90, F value 0.25 for SVM.
AUC 0.56, sensitivity 0.00, specificity 1.00, F value NAN for KNN.

Table 4
Comparison of diagnostic accuracy for each clinical OA severity prediction model based on validation and test data of machine learning models.

Validation data (OA severity prediction Model 1)							Test data (OA severity prediction Model 1)						
Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F value	Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F value
RF	0.90	0.82	0.60	0.89	0.60	0.89	0.60	RF	0.92	0.81	0.64	0.92	0.82	0.81	0.72
DA	0.88	0.78	0.80	0.77	0.50	0.93	0.62	DA	0.92	0.83	0.81	0.85	0.76	0.88	0.78
LR	0.88	0.76	0.70	0.77	0.47	0.90	0.56	LR	0.92	0.82	0.75	0.86	0.77	0.85	0.76
NBC	0.92	0.84	1.00	0.80	0.59	1.00	0.74	NBC	0.90	0.81	0.78	0.83	0.74	0.86	0.76
SVM	0.87	0.84	0.60	0.91	0.67	0.89	0.63	SVM	0.91	0.81	0.56	0.97	0.91	0.78	0.69
KNN	0.91	0.87	0.90	0.86	0.64	0.97	0.75	KNN	0.90	0.83	0.75	0.88	0.79	0.85	0.77
Validation data (OA severity prediction Model 2)							Test data (OA severity prediction Model 2)						
Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F value	Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F value
RF	0.94	0.82	0.60	0.89	0.60	0.89	0.60	RF	0.91	0.82	0.67	0.92	0.83	0.82	0.74
DA	0.90	0.80	0.80	0.80	0.53	0.93	0.64	DA	0.92	0.82	0.78	0.85	0.76	0.86	0.77
LR	0.90	0.80	0.70	0.83	0.54	0.91	0.61	LR	0.92	0.80	0.72	0.85	0.74	0.83	0.73
NBC	0.93	0.84	1.00	0.80	0.59	1.00	0.74	NBC	0.90	0.80	0.78	0.81	0.72	0.86	0.75
SVM	0.84	0.82	0.60	0.89	0.60	0.89	0.60	SVM	0.89	0.78	0.53	0.93	0.83	0.76	0.64
KNN	0.93	0.84	0.70	0.89	0.64	0.91	0.67	KNN	0.91	0.83	0.72	0.90	0.81	0.84	0.76
Validation data (OA severity prediction model 3)							Test data (OA severity prediction model 3)						
Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F value	Model	AUC	Diagnosis rate	Sensitivity	Specificity	PPV	NPV	F value
RF	0.67	0.84	0.30	1.00	1.00	0.83	0.46	RF	0.51	0.61	0.11	0.92	0.44	0.63	0.18
DA	0.71	0.76	0.00	0.97	0.00	0.77	NAN	DA	0.53	0.64	0.06	1.00	1.00	0.63	0.11
LR	0.71	0.76	0.00	0.97	0.00	0.77	NAN	LR	0.53	0.64	0.06	1.00	1.00	0.63	0.11
NBC	0.65	0.78	0.00	1.00	NAN	0.78	NAN	NBC	0.52	0.62	0.00	1.00	NAN	0.62	NAN
SVM	0.59	0.76	0.20	0.91	0.40	0.80	0.27	SVM	0.60	0.62	0.17	0.90	0.50	0.64	0.25
KNN	0.38	0.78	0.00	1.00	NAN	0.78	NAN	KNN	0.56	0.62	0.00	1.00	NAN	0.62	NAN

4. Consideration

Radiomics-based models have shown their utility in various medical fields. For example, Warkentin et al. [34] developed a model to predict the risk of malignancy in lung nodules using low-dose CT, demonstrating the usefulness of radiomics. Zengjie et al. [35] demonstrated that radiological features of the spine from MRI images can be used to predict treatment response in patients with multiple myeloma, contributing to the advancement of personalized medicine. Zhu et al. [36] showed that radiomics models based on US images are useful for Ki-67 evaluation of breast cancer, whereas Adjei et al. [37] revealed that BRAFV600E mutation can be predicted before surgery in patients with papillary thyroid cancer.

Radiomics-based models overcome the subjective evaluation and inter-observer variability associated with conventional imaging methods (radiography and MRI), allowing for more objective and quantitative diagnosis [38–41]. Multiple studies have shown that radiomics can capture microscopic changes in tissues such as subchondral bone, cartilage, and fatty bodies below the knee cap, enabling the detection of early pathological changes and prediction of disease progression that are difficult with conventional visual assessment [21, 42,43].

In addition, the use of radiomics has been greatly improved with the introduction of machine learning and automated segmentation. The time and effort required for manual feature extraction have reduced, reproducibility and consistency have improved, and clinical utility has increased [41,44]. The use of radiofrequency features has also been shown to be useful in the clinical setting [45]. Furthermore, it is expected to become a tool to assist physicians in making treatment decisions [21,41] The following are a few examples of the use of the new technology.

In this study, a high specificity of 0.98 was achieved for diagnosing OA using the MORPHOLOGICAL_SurfaceToVolumeRatio (IBSI:2PR5). This near-100 % specificity is valuable, as it helps exclude individuals who do not have OA, making the primary discrimination effective. The MORPHOLOGICAL_SurfaceToVolumeRatio (IBSI:2PR5) remains relatively constant in normal joint structures, with changes indicating abnormalities [45,46], which likely aids in distinguishing patients with OA from those without OA. However, the low sensitivity of 0.47 suggests that not all patients OA exhibit a clear change in this ratio, indicating that this feature alone has limitations in detecting OA.

Conversely, combining background factors such as age, sex, and BMI with radiomics features improved the ability to differentiate OA from non-OA conditions (or simple aging). The sensitivity of the combined model increased to 0.80, compared to 0.71 for radiomics features alone and 0.61 for background factors alone. This 10 % improvement in sensitivity, without a significant drop in specificity, is critical for enhancing diagnostic accuracy. It is particularly noteworthy that both radiomics features and background factors alone help identify aspects that might otherwise be missed.

Next, a machine learning model was constructed by selecting features using Lasso regression. Lasso regression is a regularization method that selects only the most important features while controlling model complexity. It has been widely reported to improve performance [20,21, 44]. In this study, we applied the Lambda1SE criterion [47,48], which resulted in the selection of three features with nonzero regression coefficients. These features had been validated in previous studies [44, 49–51]. The model was built using six machine learning algorithms that have been proven effective. For each algorithm, models were constructed with three different combinations (radiomics features only, radiomics features + background factors, and background factors only), and their diagnostic performance was evaluated.

Among these models, the combination of radiomics features and background factors (Model 2) exhibited the highest diagnostic accuracy. In particular, DA and LR demonstrated the most well-balanced and stable diagnostic performance across multiple metrics, including AUC,

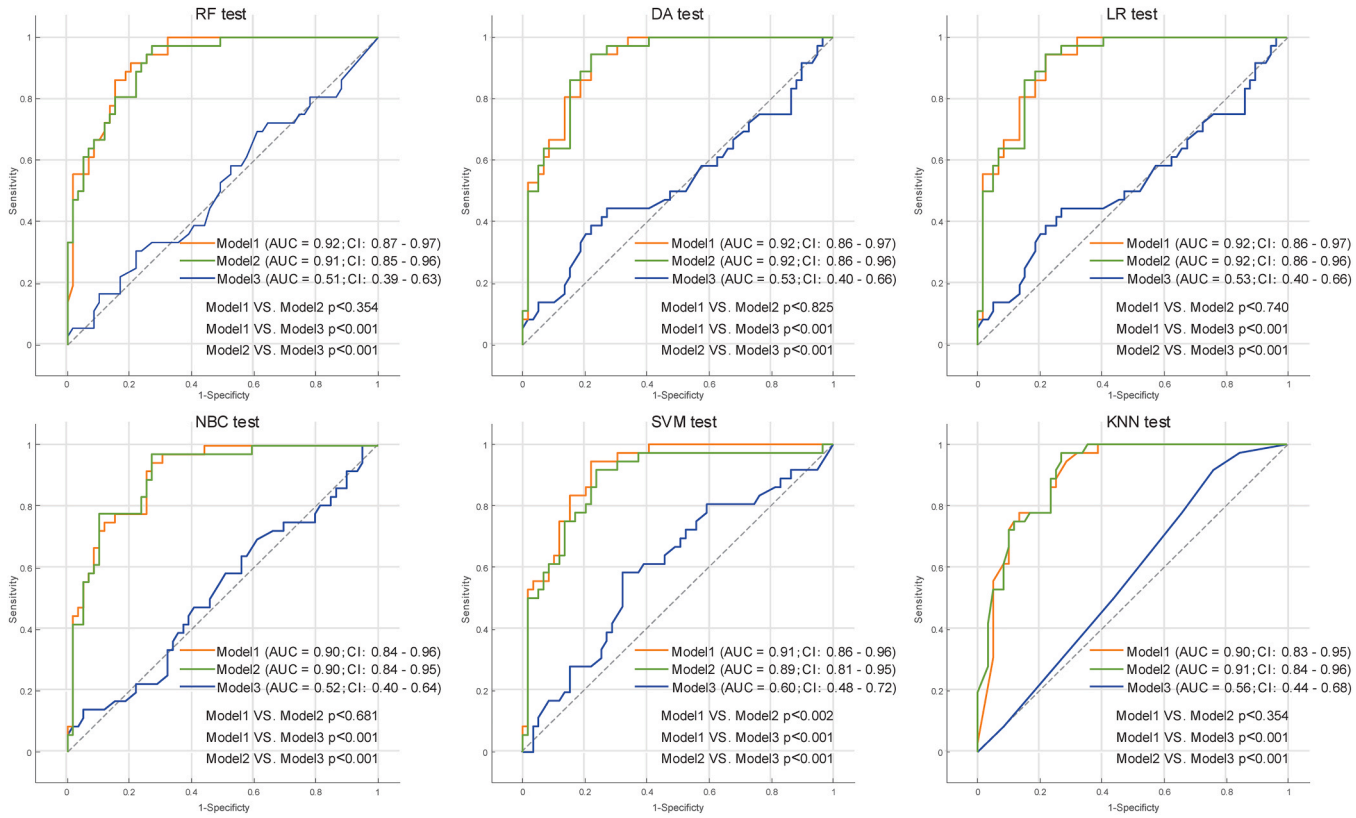

Fig. 9. ROC curve comparison of machine learning models with radiomics features and background factors in predicting clinical OA severity (test data). Model 1 (using only radiomics features): Uses radiomics features extracted from images. Model 2 (Radiomics + background factor use): A model that combines BMI in addition to Radiomics. Model 3 (using only background factors): Model using only BMI. RF, Random Forest; DA, Discriminant Analysis; LR, Logistic Regression; NBC, Naive Bayes Classifier; SVM, Support Vector Machine; KNN; K-Nearest Neighbors.

diagnostic accuracy, sensitivity, and specificity, making them promising tools for assisting in the diagnosis of knee OA.

Other machine learning models also exhibited distinct trends. Although RF and SVM achieved AUC and diagnostic accuracy comparable to those of DA and LR, their sensitivity was slightly lower than that of DA and LR, suggesting a potential risk of missing certain OA cases. Both RF and SVM exhibited a specificity of 0.82, indicating well-balanced diagnostic accuracy; however, their sensitivity was 0.77.

NBC had the highest AUC (0.89) but exhibited relatively low sensitivity (0.64) and NPV (0.56), indicating a tendency for a higher false-negative rate. However, its high specificity (0.90) and PPV (0.93) indicate that NBC could be useful in situations where minimizing false positives while maintaining a high PPV is a priority, though careful consideration is required for its use as a standalone model.

KNN exhibited the highest sensitivity (0.90) but had relatively low specificity (0.62), leading to a higher rate of false positives. Although its diagnostic accuracy (0.80) and F1-score (0.86) were high, its low specificity remains a challenge.

Based on these findings, DA and LR demonstrated the most balanced diagnostic performance among the models examined in this study, making them the most promising machine learning models for diagnosing knee OA. However, the diagnostic accuracy of other machine learning models can be enhanced considering the specific clinical needs. For example, NBC could be applied in scenarios where minimizing false positives is crucial, whereas KNN could be prioritized in situations where sensitivity is a primary concern. Thus, the models can be strategically selected for a tailored diagnostic support based on clinical objectives. Furthermore, combining clinical data, such as age, sex, and BMI, with radiomics features increased the sensitivity of the model to 80 %, suggesting that the model can detect deviations from the normal range due to age and sex and contribute to OA diagnosis. On the other hand, the model using only background factors (Model 3) showed a lower specificity of 0.20–0.60, indicating a higher risk of misdiagnosis. These results demonstrate that using radiomics features contributes to improved diagnostic accuracy.

Next, we focused on the severity assessment of knee OA, which is clinically critical in determining conservative treatment options or the indication for surgery, especially after KL grade 3, which increases the risk of surgery. The statistical method using the MORPHOLOGICAL_SurfaceToVolumeRatio (IBSI:2PR5) showed a sensitivity of 0.78, specificity of 0.86, positive predictive value (PPV) of 0.72, and negative predictive value (NPV) of 0.90. While the sensitivity is low and not suitable for screening severe cases, the high specificity reduces false positives and may serve as a useful objective indicator.

In machine learning, two features were selected using Lasso regression, and two models were evaluated: one using only radiomics features (Model 1) and the other combining radiomics and background factors (Model 2). In particular, DA (Model 1) demonstrated the most well-balanced diagnostic performance. With an AUC of 0.92, diagnostic accuracy of 0.83, sensitivity of 0.81, and specificity of 0.85, it facilitated accurate diagnosis while minimizing the risk of missing severe OA cases.

On the other hand, other machine learning models exhibited distinct trends. RF achieved an AUC of 0.92, comparable to that of DA, but had a lower sensitivity of 0.64, indicating a potential risk of missing some severe OA cases. However, its high specificity of 0.92 suggests its usefulness in reducing misdiagnoses.

NBC demonstrated an AUC of 0.90, sensitivity of 0.78, and specificity of 0.83, providing a relatively balanced diagnostic performance. However, compared to DA and LR, its specificity was slightly lower. With a PPV of 0.74 and an NPV of 0.86, it showed strengths in negative prediction, making it potentially useful in scenarios where avoiding false negatives is crucial.

SVM exhibited the highest specificity at 0.97, effectively minimizing false positives. However, its sensitivity was low at 0.56, posing a high risk of missing severe OA cases. Therefore, careful consideration is necessary when using it as a standalone diagnostic tool.

KNN demonstrated a relatively balanced performance with a sensitivity of 0.75 and specificity of 0.88, achieving an F1-score of 0.77—the second highest after DA. This suggests that it is a promising method from the perspective of diagnostic accuracy.

In contrast to the OA diagnostic model, no statistically significant difference in AUC was observed between Model 1 and Model 2 across all machine learning models, except for SVM in the severity prediction model, suggesting that incorporating the background factor (BMI) does not necessarily improve diagnostic accuracy.

These results confirm that machine learning models with radiomics features provide superior diagnostic accuracy in both the diagnosis and severity prediction of clinical OA. In particular, they suggest that a simple model using only radiomics features may increase diagnostic utility and efficiency, especially when the effect of additional background factors is limited.

In this study, BMI did not contribute sufficiently to the assessment of the severity of knee OA. However, previous studies have reported that increased BMI is associated with the risk and severity of knee OA [52, 53]. One possible reason for this discrepancy is that metabolic factors associated with obesity vary among racial groups. Differences in the association between BMI and serum leptin concentrations have been found between black and White women, and variations in metabolic factors among postmenopausal obese women depending on race have also been reported [53–56]. This suggests that the utility of BMI in assessing knee OA risk may vary by race.

The flowchart shown in Fig. 10 presents a novel approach for diagnosing and assessing the severity of knee OA. This method begins with primary discrimination using the cutoff value of the radiomics feature SurfaceToVolumeRatio (IBSI:2PR5) to distinguish between patients without OA and those with OA. OA diagnosis is then confirmed using a machine learning model that combines radiomics features with age, sex, and BMI. For OA cases, the machine learning model (using radiomics features only) can classify them into mildly deformed and severely deformed groups.

The results of this study, along with those from previous studies, are summarized in Table 5, providing a comparison of findings. Li et al. [21] used radiographs to develop a knee OA classification model, which achieved an AUC of 0.85, sensitivity of 0.78, and specificity of 0.75 for classifying OA (KL 2–4) and non-OA conditions (KL 0–1). In contrast, the DA model based on US radiomics analysis in this study achieved an AUC of 0.88, sensitivity of 0.80, and specificity of 0.82, demonstrating comparable diagnostic accuracy. Notably, the results suggest that US radiomics may offer an advantage in diagnosing knee OA, as US can also be used to evaluate soft tissue and meniscus degeneration, while radiographic imaging focuses primarily on bone deformities.

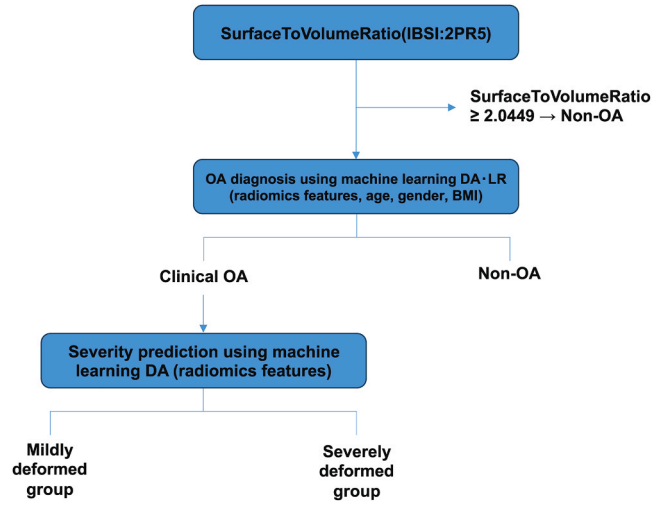

Fig. 10. Flowchart for diagnosis and severity classification of knee OA. DA, Discriminant Analysis; LR, Logistic Regression.

Additionally, Li et al. [40] conducted a study using anteroposterior and lateral radiographs, which showed lower sensitivity for OA classification (0.68 for mild OA, 0.57 for moderate OA, and 0.63 for severe OA). In contrast, our model demonstrated high performance in AUC and diagnostic accuracy, outperforming their results. Li et al. [40] classified each KL grade and discussed them in detail, but integrating the grades for a broader analysis may be clinically beneficial. For example, classifying knee OA as mild (KL 2) or severe (KL 3–4) may enable a simpler model more suitable for clinical use.

Xue et al. [20] developed a model based on MRI radiomics analysis, focusing on tibial and femoral bone marrow edema regions. Their model achieved an AUC of 0.96 for OA (KL 1–4) vs. non-OA (KL 0), with an AUC of 0.99 for severe (KL 3–4) vs. mild (KL 1–2) OA classification. The highest AUCs in this study were 0.89 for clinical OA (KL 2–4) vs. non-OA (KL 0–1) and 0.92 for the severely deformed group (KL 3–4) vs. the mildly deformed group (KL 2), which were slightly lower than the MRI-based models of Xue et al. [20] but still promising in terms of novelty.

Ye et al. [38] used MRI-based radiomics to assess the severity of knee OA, focusing on the subpatellar fat pad. They showed an AUC of 0.78, sensitivity of 0.73, and specificity of 0.70 in classifying OA (KL 2–4) vs. non-OA (KL 0–1). In comparison, all machine learning models in this study showed better diagnostic performance, especially in specificity. Li et al. [43] performed an MRI-based radiomics analysis focused on bone marrow edema and developed a nomogram combining clinical characteristics (age, KL grade, and WOMAC score), achieving an AUC of 0.84 for OA vs. non-OA, sensitivity of 0.95, specificity of 0.86, PPV of 0.69, and NPV of 0.63. In contrast, the US model in this study demonstrated comparable high performance [57]. MRI studies, such as those by Xue et al. [20], have limitations, such as the use of 3D BFFE (Balanced Fast-Field Echo) sequences, which are not widely used in clinical practice due to time constraints. In contrast, US has been used in several studies and, although slightly inferior to MRI in terms of diagnostic accuracy, offers advantages in practicality, cost, and accessibility. More importantly, previous studies by Li et al. [43] and Ye et al. [38] focused mainly on the diagnosis of knee OA, while this study also addresses OA severity. This study's model achieved high diagnostic accuracy (AUC 0.92) for classifying severely deformed (KL 3–4) vs. mildly deformed (KL 2) groups, demonstrating the model's potential in severity discrimination.

The radiomics analysis in this study differs from morphological diagnostic methods using cutoff values, such as those reported by Kiso et al. [23] and Yanagisawa et al. [57]. Yanagisawa et al. [58] reported a high diagnostic accuracy (0.91 sensitivity, 0.96 specificity, 0.98 PPV, and 0.84 NPV) using US, but their method depends on specific cutoff values and is prone to alignment changes due to evaluator experience and OA progression, which raises concerns about reproducibility. Kiso et al. [23] showed a sensitivity of 0.80, specificity of 0.76, PPV of 0.85, and NPV of 0.69 for clinical OA (KL 2–4) vs. non-OA (KL 0–1) groups. Their analysis also showed high reproducibility, as results were not affected by alignment changes in OA progression, though it did not evaluate the meniscus or articular cartilage, limiting its ability to assess early OA.

The US-based radiomics model proposed in this study is superior to the diagnostic methods of Yanagisawa et al. [57] and Kiso et al. [23] in terms of diagnostic accuracy and reproducibility. By eliminating subjective evaluation, this model is considered more reliable for assessing knee OA progression.

This study also evaluated intra- and inter-inspector reliability in extracting radiomics features. The features ultimately used in the analysis were selected based on consensus, which minimized measurement

Table 5
Comparison of diagnostic accuracy between statistical methods and machine learning models in this and previous studies.

Author	Modality	Method	Knee OA Classification Method	Number of cases	Sensitivity	Specificity	AUC	PPV	NPV	F1 value
Auto-research results	US	Radiomics + Statistical Methods	Clinical OA (KL 2–4) vs. Non-OA (KL 0–1)	OA: 318 cases, Non-OA: 173 limbs	0.47	0.98	-	0.97	0.52	-
			Highly deformed (KL 3–4) vs. mildly deformed (KL 2)		0.78	0.86	-	0.72	0.90	-
		Radiomics + Machine Learning	OA (KL 2–4) vs. Non-OA (KL 0–1)		0.81	0.80	0.89	0.89	0.69	0.85
			Clinical OA (KL 3–4) vs. OA (KL 2)		0.81	0.85	0.92	0.76	0.88	0.78
Li et al. (2023) [21]	Radiography	Radiomics + Machine Learning	(KL 2–4) vs. Non-OA (KL 0–1)	OA: 612 cases, Non-OA: 562 cases	0.78	0.75	0.85	0.91	0.92	0.84
Li et al. (2024) [40]	Radiography	Radiomics + Machine Learning	Normal (KL 0)	OA: 311 cases, Non-OA: 162 cases	0.85	0.99	0.99	0.98	0.85	0.91
			Minor (KL 1)		0.68	0.92	0.78	0.74	0.87	0.71
			Moderate OA (KL 2)		0.57	0.92	0.73	0.57	0.92	0.57
			Severe OA (KL 3)		0.75	0.92	0.91	0.63	0.94	0.68
			Very Severe OA (KL 4)		0.63	0.91	0.89	0.48	0.94	0.54
Xue et al. (2022) [20]	MRI	Radiomics + Machine Learning	OA (KL 1–4) vs. Non-OA (KL 0)	OA: 56 cases, Non-OA: 32 cases	0.98	0.81	0.96	-	-	0.92
			Severe OA (KL 3–4) vs. OA (KL 1–2)		0.97	0.96	0.99	-	-	0.96
Ye et al. (2023) [38]	MRI	Radiomics + Machine Learning	OA (KL 2–4) vs. Non-OA (KL 0–1)	OA: 130 cases, Non-OA: 34 cases	0.73	0.70	0.78	-	-	-
Li et al. (2024) [43]	MRI	Radiomics + Machine Learning	OA (KL 2–4) vs. Non-OA (KL 0–1)	OA: 203 cases, Non-OA: 99 cases	0.95	0.86	0.84	0.69	0.63	0.80
Kiso et al. (2024) [23]	US	TOH-DBB index (quantitative evaluation method)	Clinical OA (KL 2–4) vs. Non-OA (KL 0–1)	OA: 134 cases, Non-OA: 69 cases	0.80	0.76	-	0.85	0.69	-
			Highly deformed OA (KL 3–4) vs. mildly deformed (KL 2)		0.71	0.85	-	0.71	0.85	-
Yanagisawa et al. (2014) [57]	US	Three quantitative evaluation methods	OA (KL 2–4) vs. Non-OA (KL 0–1)	OA: 87 cases, Non-OA: 44 cases	0.91	0.96	-	0.98	0.84	-

variability. Additionally, high ICCs were found for the main features selected for ANOVA and Lasso regression, indicating consistent measurements.

Nevertheless, this study has some limitations that must be acknowledged. First, data collection was restricted to a single institution. Second, this study classified conditions with KL 0–1 into the non-OA group and those with KL 2–4 into the clinical OA group, while some previous studies [20,21,40] have classified conditions with KL 0 into the non-OA group and those with KL 1 into the OA group. The KL classification is a standard measure of knee OA progression and shows a degree of consistency across studies, making it valuable to compare the characteristics and relative strengths of each method.

Furthermore, the inter-inspector ICC was low for some radiomics features, suggesting that measurements for certain features may vary between examiners, possibly due to differences in calculation methods or image interpretation. Although the final feature selection was done collaboratively, reducing inter-inspector variability, it does not fully resolve discrepancies for all radiomics features. Future studies should focus on standardizing the feature extraction process and improving training to enhance inter-inspector agreement.

In this study, the analysis was conducted using a single device at a single institution; therefore, the results may vary when using different ultrasound devices. However, previous studies and meta-analyses on radiomics in the thyroid, breast, and liver have demonstrated its utility despite variations in devices and institutions [59–62].

5. Conclusion

In this study, we evaluated a diagnostic and severity prediction model for knee OA using radiomics features extracted from ultrasonographic images of the knee joint and validated its effectiveness. The machine learning model utilizing these radiomics features demonstrated high diagnostic accuracy and excellent results in predicting the severity of knee OA. Our approach has the potential to serve as an adjunctive tool to assist clinicians in decision-making. Future studies are expected to further enhance the model's accuracy and establish its clinical application through external validation at multiple centers and with larger datasets.

Ethics approval statement

This study was approved by the Ethical Review Committee of Tsukuba International University (No. R05-12) and conducted in accordance with the Declaration of Helsinki.

Funding

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Informed consent statement

All patients provided written informed consent.

Permission to reproduce material from other sources

Not applicable.

CRediT authorship contribution statement

Kawata Satoru: Formal analysis. **Okada Yukinori:** Methodology, Formal analysis, Conceptualization. **Kiso Takeharu:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Okumura Erika:** Investigation. **Kuwano**

Hikaru: Investigation. **Kaminaga Masaki:** Investigation. **Matsuura Ryohei:** Investigation. **Hatsumi Noritaka:** Investigation. **Okumura Eiichiro:** Investigation. **Shichiji Kouta:** Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Yuichiro Tachikawa, Makoto Kikuchi, and Hideki Ono of Tachikawa Memorial Hospital for their cooperation in the study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejro.2025.100649.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

[1] N. Taylor, Nonsurgical management of osteoarthritis knee pain in the older adult: an update, Rheum. Dis. Clin. N. Am. 44 (2018) 513–524, https://doi.org/10.1016/j.rdc.2018.03.009.
[2] R.F. Loeser, S.R. Goldring, C.R. Scanzello, M.B. Goldring, Osteoarthritis: a disease of the joint as an organ, Arthritis Rheum. 64 (2012) 1697–1707, https://doi.org/10.1002/art.34453.
[3] E. Sasaki, S. Ota, D. Chiba, Y. Kimura, S. Sasaki, Y. Yamamoto, E. Tsuda, S. Nakaji, Y. Ishibashi, Early knee osteoarthritis prevalence is highest among middle-aged adult females with obesity based on new set of diagnostic criteria from a large sample cohort study in the Japanese general population, Knee Surg. Sports Traumatol. Arthrosc. 28 (2020) 984–994, https://doi.org/10.1007/s00167-019-05614-z.
[4] M.L.A. Landsmeer, J. Runhaar, M. van Middelkoop, E.H.G. Oei, D. Schiphof, P.J. E. Bindels, S.M.A. Bierma-Zeinstra, Predicting knee pain and knee osteoarthritis among overweight women, J. Am. Board Fam. Med. 32 (2019) 575–584, https://doi.org/10.3122/jabfm.2019.04.180302.
[5] S.A. Hrnack, F.A. Barber, Managing the pain of knee osteoarthritis, Phys. Sport. 42 (2014) 63–70, https://doi.org/10.3810/psm.2014.09.2077.
[6] K.D. Allen, L.M. Thoma, Y.M. Golightly, Epidemiology of osteoarthritis, Osteoarthr. Cartil. 30 (2022) 184–195, https://doi.org/10.1016/j.joca.2021.04.020.
[7] W.B. Lim, O. Al-Dadah, Conservative treatment of knee osteoarthritis: a review of the literature, World J. Orthop. 13 (2022) 212–229, https://doi.org/10.5312/wjo.v13.i3.212.
[8] R.D. Altman, G.E. Gold, Atlas of individual radiographic features in osteoarthritis, revised, Osteoarthr. Cartil. 15 Suppl A (2007) A1–A56, https://doi.org/10.1016/j.joca.2006.11.009.
[9] S. Botha-Scheepers, M. Dougados, P. Ravaud, M.P. Hellio Le Graverand, I. Watt, F. C. Breedveld, M. Kloppenburg, Effect of medial tibial plateau alignment on serial radiographs on the capacity to predict progression of knee osteoarthritis, Osteoarthr. Cartil. 16 (2008) 272–276, https://doi.org/10.1016/j.joca.2007.10.020.
[10] A.C. Marijnissen, K.L. Vincken, P.A. Vos, D.B. Saris, M.A. Viergever, J.W. Bijlsma, L.W. Bartels, F.P. Lafeber, Knee Images Digital Analysis (KIDA): a novel method to quantify individual radiographic features of knee osteoarthritis in detail, Osteoarthr. Cartil. 16 (2008) 234–243, https://doi.org/10.1016/j.joca.2007.06.009.
[11] F.W. Roemer, A. Guermazi, S. Demehri, W. Wirth, R. Kijowski, Imaging in osteoarthritis, Osteoarthr. Cartil. 30 (2022) 913–934, https://doi.org/10.1016/j.joca.2021.04.018.
[12] S. Saarakkala, P. Waris, V. Waris, I. Tarkiainen, E. Karvanen, J. Aarnio, J.M. Koski, Diagnostic performance of knee ultrasonography for detecting degenerative changes of articular cartilage, Osteoarthr. Cartil. 20 (2012) 376–381, https://doi.org/10.1016/j.joca.2012.01.016.
[13] K. Shimozaki, J. Nakase, K. Asai, R. Yoshimizu, M. Kimura, T. Kanayama, T. Kitagawa, H. Tsuchiya, Usefulness of ultrasonography for dynamic evaluation of medial meniscus hoop function in early knee osteoarthritis, Sci. Rep. 11 (2021) 20091, https://doi.org/10.1038/s41598-021-99576-3.
[14] H. Majidi, F. Niksolat, K. Anbari, Comparing the accuracy of radiography and sonography in detection of knee osteoarthritis: a diagnostic study, Open Access Maced. J. Med. Sci. 7 (2019) 4015–4018, https://doi.org/10.3889/oamjms.2019.617.
[15] C. Liu, Y. Li, X. Xia, J. Wang, C. Hu, Application of radiomics feature captured from MRI for prediction of recurrence for glioma patients, J. Cancer 13 (2022) 965–974, https://doi.org/10.7150/jca.65366.
[16] G.Q. Hu, Y.Q. Ge, X.K. Hu, W. Wei, Predicting coronary artery calcified plaques using perivascular fat CT radiomics features and clinical risk factors, BMC Med. Imaging 22 (2022) 134, https://doi.org/10.1186/s12880-022-00858-7.
[17] F. Feng, P. Wang, K. Zhao, B. Zhou, H. Yao, Q. Meng, L. Wang, Z. Zhang, Y. Ding, N. An, X. Zhang, Y. Liu, Radiomic features of hippocampal subregions in Alzheimer's disease and amnestic mild cognitive impairment, Front. Aging Neurosci. 10 (2018) 290, https://doi.org/10.3389/fnagi.2018.00290.
[18] C.B. Huang, J.S. Hu, K. Tan, W. Zhang, T.H. Xu, L. Yang, Application of machine learning model to predict osteoporosis based on abdominal computed tomography images of the psoas muscle: a retrospective study, BMC Geriatr. 22 (2022) 796, https://doi.org/10.1186/s12877-022-03502-9.
[19] Y. Ji, H. Li, A.V. Edwards, J. Papaioannou, W. Ma, P. Liu, M.L. Giger, Independent validation of machine learning in diagnosing breast cancer on magnetic resonance imaging within a single institution, Cancer Imaging 19 (2019) 64, https://doi.org/10.1186/s40644-019-0252-2.
[20] Z. Xue, L. Wang, Q. Sun, J. Xu, Y. Liu, S. Ai, L. Zhang, C. Liu, Radiomics analysis using MR imaging of subchondral bone for identification of knee osteoarthritis, J. Orthop. Surg. Res. 17 (2022) 414, https://doi.org/10.1186/s13018-022-03314-y.
[21] W. Li, J. Feng, D. Zhu, Z. Xiao, J. Liu, Y. Fang, L. Yao, B. Qian, S. Li, Nomogram model based on radiomics signatures and age to assist in the diagnosis of knee osteoarthritis, Exp. Gerontol. 171 (2023) 112031, https://doi.org/10.1016/j.exger.2022.112031.
[22] M. Villagran, J.B. Driban, B. Lu, J.W. MacKay, T.E. McAlindon, M.S. Harkey, Radiomic features of the medial meniscus predicts incident destabilizing meniscal tears: data from the osteoarthritis initiative, J. Orthop. Res. 42 (2024) 2080–2087, https://doi.org/10.1002/jor.25851.
[23] T. Kiso, Y. Okada, S. Kawata, K. Shichiji, E. Okumura, N. Hatsumi, R. Matsuura, M. Kaminaga, H. Kuwano, Diagnostic accuracy of a novel ultrasound imaging index for knee osteoarthritis: evaluation of sensitivity, specificity, and predictive values, J. Clin. Ultrasound 52 (2024) 687–699, https://doi.org/10.1002/jcu.23691.
[24] J.S. Smolen, R.B.M. Landewé, J.W.J. Bijlsma, G.R. Burmester, M. Dougados, A. Kerschbaumer, I.B. McInnes, A. Sepriano, R.F. van Vollenhoven, M. de Wit, D. Aletaha, M. Aringer, J. Askling, A. Balsa, M. Boers, A.A. den Broeder, M.H. Buch, F. Buttgereit, R. Caporali, M.H. Cardiel, D. De Cock, C. Codreanu, M. Cutolo, C. J. Edwards, Y. van Eijk-Hustings, P. Emery, A. Finckh, L. Gossec, J.E. Gottenberg, M.L. Hetland, T.W.J. Huizinga, M. Koloumas, Z. Li, X. Mariette, U. Müller-Ladner, E.F. Mysler, J.A.P. da Silva, G. Poór, J.E. Pope, A. Rubbert-Roth, A. Ruyssen-Witrand, K.G. Saag, A. Strangfeld, T. Takeuchi, M. Voshaar, R. Westhovens, D. van der Heijde, EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update, Ann. Rheum. Dis. 79 (6) (2020) 685–699, https://doi.org/10.1136/annrheumdis-2019-216655.
[25] J.H. Kellgren, J.S. Lawrence, Radiological assessment of osteo-arthrosis, Ann. Rheum. Dis. 16 (1957) 494–502, https://doi.org/10.1136/ard.16.4.494.
[26] Y. Shimura, H. Kurosawa, Y. Sugawara, M. Tsuchiya, M. Sawa, H. Kaneko, I. Futami, L. Liu, R. Sadatsuki, S. Hada, Y. Iwase, K. Kaneko, M. Ishijima, The factors associated with pain severity in patients with knee osteoarthritis vary according to the radiographic disease severity: a cross-sectional study, Osteoarthr. Cartil. 21 (2013) 1179–1184, https://doi.org/10.1016/j.joca.2013.05.014.
[27] C. Nioche, F. Orlhac, S. Boughdad, S. Reuzé, J. Goya-Outi, C. Robert, C. Pellot-Barakat, M. Soussan, F. Frouin, I. Buvat, LIFEx: a Freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity, Cancer Res. 78 (2018) 4786–4789, https://doi.org/10.1158/0008-5472.CAN-18-0125.
[28] A. Zwanenburg, M. Vallières, M.A. Abdalah, H. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R.J. Beukinga, R. Boellaard, M. Bogowicz, L. Boldrini, I. Buvat, G.J.R. Cook, C. Davatzikos, A. Depeursinge, M.C. Desseroit, N. Dinapoli, C.V. Dinh, S. Echegaray, I. El Naqa, A.Y. Fedorov, R. Gatta, R.J. Gillies, V. Goh, M. Götz, M. Guckenberger, S.M. Ha, M. Hatt, F. Isensee, P. Lambin, S. Leger, R.T. H. Leijenaar, J. Lenkowicz, F. Lippert, A. Losnegård, K.H. Maier-Hein, O. Morin, H. Müller, S. Napel, C. Nioche, F. Orlhac, S. Pati, E.A.G. Pfaehler, A. Rahmim, A.U. K. Rao, J. Scherer, M.M. Siddique, N.M. Sijtsema, J. Socarras Fernandez, E. Spezi, R. Steenbakkers, S. Tanadini-Lang, D. Thorwarth, E.G.C. Troost, T. Upadhaya, V. Valentini, L.V. van Dijk, J. van Griethuysen, F.H.P. van Velden, P. Whybra, C. Richter, S. Löck, The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping, Radiology 295 (2020) 328–338, https://doi.org/10.1148/radiol.2020191145.
[29] J.H. Jo, H.W. Chung, Y. So, Y.B. Yoo, K.S. Park, S.E. Nam, E.J. Lee, W.C. Noh, FDG PET/CT to predict recurrence of early breast invasive ductal carcinoma, Diagnostics 12 (2022) 694, https://doi.org/10.3390/diagnostics12030694.
[30] A. Crombé, M. Lafon, S. Nougaret, M. Kind, S. Cousin, Ranking the most influential predictors of CT-based radiomics feature values in metastatic lung adenocarcinoma, Eur. J. Radiol. 155 (2022) 110472, https://doi.org/10.1016/j.ejrad.2022.110472.
[31] Y. Wu, J.H. Jiang, L. Chen, J.Y. Lu, J.J. Ge, F.T. Liu, J.T. Yu, W. Lin, C.T. Zuo, J. Wang, Use of radiomic features and support vector machine to distinguish Parkinson's disease cases from normal controls, Ann. Transl. Med. 7 (2019) 773, https://doi.org/10.21037/atm.2019.11.26.
[32] H. Pang, Z. Yu, R. Li, H. Yang, G. Fan, MRI-based radiomics of basal nuclei in differentiating idiopathic Parkinson's disease from Parkinsonian variants of multiple system atrophy: a susceptibility-weighted imaging study, Front. Aging Neurosci. 12 (2020) 587250, https://doi.org/10.3389/fnagi.2020.587250.

[33] Z. Xue, J. Huo, X. Sun, S.T. Ai, LichiZhang, C. Liu, Using radiomic features of lumbar spine CT images to differentiate osteoporosis from normal bone density, BMC Musculoskelet. Disord. 23 (2022) 336, https://doi.org/10.1186/s12891-022-05309-6.
[34] M.T. Warkentin, H. Al-Sawaihey, S. Lam, G. Liu, B. Diergaarde, J.M. Yuan, D. O. Wilson, S. Atkar-Khattra, B. Grant, Y. Brhane, E. Khodayari-Moez, K.R. Murison, M.C. Tammemagi, K.R. Campbell, R.J. Hung, Radiomics analysis to predict pulmonary nodule malignancy using machine learning approaches, Thorax 79 (2024) 307–315, https://doi.org/10.1136/thorax-2023-220226.
[35] Z. Wu, T. Bian, C. Dong, S. Duan, H. Fei, D. Hao, W. Xu, Spinal MRI-based radiomics analysis to predict treatment response in multiple myeloma, J. Comput. Assist. Tomogr. 46 (2022) 447–454, https://doi.org/10.1097/RCT.0000000000001298.
[36] Y. Zhu, Y. Dou, L. Qin, H. Wang, Z. Wen, Prediction of Ki-67 of invasive ductal breast cancer based on ultrasound radiomics nomogram, J. Ultrasound Med. 42 (3) (2023) 649–664, https://doi.org/10.1002/jum.16061.
[37] E.A. Agyekum, Y.G. Wang, F.J. Xu, D. Akortia, Y.Z. Ren, K.H. Chambers, X. Wang, J.O. Taupa, X.Q. Qian, Predicting BRAFV600E mutations in papillary thyroid carcinoma using six machine learning algorithms based on ultrasound elastography, Sci. Rep. 13 (2023) 12604, https://doi.org/10.1038/s41598-023-39747-6.
[38] Q. Ye, D. He, X. Ding, Y. Wang, Y. Wei, J. Liu, Quantitative evaluation of the infrapatellar fat pad in knee osteoarthritis: MRI-based radiomic signature, BMC Musculoskelet. Disord. 24 (2023) 326, https://doi.org/10.1186/s12891-023-06433-7.
[39] J. Hirvasniemi, S. Klein, S. Bierma-Zeinstra, M.W. Vernooij, D. Schiphof, E.H. G. Oei, A machine learning approach to distinguish between knees without and with osteoarthritis using MRI-based radiomic features from tibial bone, Eur. Radiol. 31 (2021) 8513–8521, https://doi.org/10.1007/s00330-021-07951-5.
[40] W. Li, J. Liu, Z. Xiao, D. Zhu, J. Liao, W. Yu, J. Feng, B. Qian, Y. Fang, S. Li, Automatic grading of knee osteoarthritis with a plain radiograph radiomics model: combining anteroposterior and lateral images, Insights Imaging 15 (2024) 143, https://doi.org/10.1186/s13244-024-01719-3.
[41] K. Yu, J. Ying, T. Zhao, L. Lei, L. Zhong, J. Hu, J.W. Zhou, C. Huang, X. Zhang, Prediction model for knee osteoarthritis using magnetic resonance-based radiomic features from the infrapatellar fat pad: data from the osteoarthritis initiative, Quant. Imaging Med. Surg. 13 (2023) 352–369, https://doi.org/10.21037/qims-22-368.
[42] J. Zhang, T. Jiang, L.C. Chan, S.H. Lau, W. Wang, X. Teng, P.K. Chan, J. Cai, C. Wen, Radiomics analysis of patellofemoral joint improves knee replacement risk prediction: data from the Multicenter Osteoarthritis Study (MOST), Osteoarthr. Cartil. Open 6 (2024) 100448, https://doi.org/10.1016/j.ocarto.2024.100448.
[43] X. Li, W. Chen, D. Liu, P. Chen, P. Li, F. Li, W. Yuan, S. Wang, C. Chen, Q. Chen, S. Guo, Z. Hu, Radiomics analysis using magnetic resonance imaging of bone marrow edema for diagnosing knee osteoarthritis, Front. Bioeng. Biotechnol. 12 (2024) 1368188, https://doi.org/10.3389/fbioe.2024.1368188.
[44] T. Cui, R. Liu, Y. Jing, J. Fu, J. Chen, Development of machine learning models aiming at knee osteoarthritis diagnosing: an MRI radiomics analysis, J. Orthop. Surg. Res. 18 (2023) 375, https://doi.org/10.1186/s13018-023-03837-y.
[45] G. Jones, C. Ding, F. Scott, M. Glisson, F. Cicuttini, Early radiographic osteoarthritis is associated with substantial changes in cartilage volume and tibial bone surface area in both males and females, Osteoarthr. Cartil. 12 (2004) 169–174, https://doi.org/10.1016/j.joca.2003.08.010.
[46] S. Tummala, A.C. Bay-Jensen, M.A. Karsdal, E.B. Dam, Diagnosis of osteoarthritis by cartilage surface smoothness quantified automatically from knee MRI, Cartilage 2 (2011) 50–59, https://doi.org/10.1177/1947603510381097.
[47] Y. Chen, Y. Yang, The one standard error rule for model selection: does it work? Stats 4 (2021) 868–892.
[48] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Softw. 33 (2010) 1–22.
[49] M. Foltyn-Dumitru, M. Schell, F. Sahm, T. Kessler, W. Wick, M. Bendszus, A. Rastogi, G. Brugnara, P. Vollmuth, Advancing noninvasive glioma classification with diffusion radiomics: exploring the impact of signal intensity normalization, Neurooncol. Adv. 6 (2024) vdae043, https://doi.org/10.1093/noajnl/vdae043.
[50] W. Wu, C. Parmar, P. Grossmann, J. Quackenbush, P. Lambin, J. Bussink, R. Mak, H.J. Aerts, Exploratory study to identify radiomics classifiers for lung cancer histology, Front. Oncol. 6 (2016) 71, https://doi.org/10.3389/fonc.2016.00071.
[51] J. Bin, M. Wu, M. Huang, Y. Liao, Y. Yang, X. Shi, S. Tao, Predicting invasion in early-stage ground-glass opacity pulmonary adenocarcinoma: a radiomics-based machine learning approach, BMC Med. Imaging 24 (2024) 240, https://doi.org/10.1186/s12880-024-01421-2.
[52] L. Jiang, W. Tian, Y. Wang, J. Rong, C. Bao, Y. Liu, Y. Zhao, C. Wang, Body mass index and susceptibility to knee osteoarthritis: a systematic review and meta-analysis, Jt. Bone Spine 79 (2012) 291–297, https://doi.org/10.1016/j.jbspin.2011.05.015.
[53] R. Shumnalieva, G. Kotov, S. Monov, Obesity-related knee osteoarthritis-current concepts, Life 13 (2023) 1650, https://doi.org/10.3390/life13081650.
[54] B.M. Dickson, A.J. Roelofs, J.J. Rochford, H.M. Wilson, C. De Bari, The burden of metabolic syndrome on osteoarthritic joints, Arthritis Res. Ther. 21 (2019) 289, https://doi.org/10.1186/s13075-019-2081-x.
[55] S.S. Cohen, J.H. Fowke, Q. Cai, M.S. Buchowski, L.B. Signorello, M.K. Hargreaves, W. Zheng, W.J. Blot, C.E. Matthews, Differences in the association between serum leptin levels and body mass index in black and white women: a report from the Southern Community Cohort Study, Ann. Nutr. Metab. 60 (2012) 90–97, https://doi.org/10.1159/000336180.
[56] B.J. Nicklas, M.J. Toth, A.P. Goldberg, E.T. Poehlman, Racial differences in plasma leptin concentrations in obese postmenopausal women, J. Clin. Endocrinol. Metab. 82 (1997) 315–317, https://doi.org/10.1210/jcem.82.1.3659.
[57] S. Yanagisawa, T. Ohsawa, K. Saito, T. Kobayashi, A. Yamamoto, K. Takagishi, Morphological evaluation and diagnosis of medial type osteoarthritis of the knee using ultrasound, J. Orthop. Sci. 19 (2014) 270–274, https://doi.org/10.1007/s00776-013-0524-9.
[58] K. Bevers, J.W. Bijlsma, J.E. Vriezekolk, C.H. van den Ende, A.A. den Broeder, The course of ultrasonographic abnormalities in knee osteoarthritis: 1 year follow up, Osteoarthr. Cartil. 22 (2014) 1651–1656, https://doi.org/10.1016/j.joca.2014.06.012.
[59] Z. Li, X. Liu, Y. Gao, X. Lu, J. Lei, Ultrasound-based radiomics for early predicting response to neoadjuvant chemotherapy in patients with breast cancer: a systematic review with meta-analysis, Radiol. Med. 129 (2024) 934–944, https://doi.org/10.1007/s11547-024-01783-1.
[60] S. Zhang, R. Liu, Y. Wang, Y. Zhang, M. Li, S. Wang, N. Ma, J. Ren, Ultrasound-base radiomics for discerning lymph node metastasis in thyroid cancer: a systematic review and meta-analysis, Acad. Radiol. 31 (2024) 3118–3130, https://doi.org/10.1016/j.acra.2024.03.012.
[61] J. Zhou, J. Wu, C. Zhu, Q. An, Ultrasound radiomics in the assessment of breast cancer molecular subtypes: a systematic review and meta-analysis, Med. Ultrason. (2024), https://doi.org/10.11152/mu-4449.
[62] Q. Xiao, W. Zhu, H. Tang, L. Zhou, Ultrasound radiomics in the prediction of microvascular invasion in hepatocellular carcinoma: a systematic review and meta-analysis, Heliyon 9 (2023) e16997, https://doi.org/10.1016/j.heliyon.2023.e16997.

Glossary

Radiomics: Radiomics is an image analysis technique that extracts high-dimensional quantitative features from medical images to support disease diagnosis and prognosis. It enables the identification of patterns not detectable through visual inspection.
Radiomic Features: Quantitative attributes derived from medical images are categorized as follows:
Shape features: Describe structural morphology (e.g., SurfaceToVolumeRatio).
First-order statistics: Represent intensity distribution (e.g., histogram-based features).
Texture features: Capture spatial variations in intensity (e.g., GLSZM, GLCM).
Higher-order features: Extracted via wavelet transforms or fractal analysis.
Machine Learning (ML): A subset of artificial intelligence in which algorithms learn patterns from data to make predictions. Common ML approaches in radiomics include:
Supervised learning: Uses labeled data for classification (e.g., logistic regression, random forest).
Unsupervised learning: Identifies patterns in unlabeled data (e.g., k-means clustering).
Reinforcement learning: Learns optimal actions through trial and error.
Least Absolute Shrinkage and Selection Operator (LASSO): A regression method that applies L1 regularization, reducing the coefficients of less important features to zero for automatic feature selection. It prevents overfitting and enhances model interpretability.
Filter-type Feature Selection Algorithm: An algorithm that selects the most relevant features using statistical methods (e.g., analysis of variance [ANOVA], correlation coefficients). It operates independently of machine learning models.