Received: 16 September 2024 | Accepted: 7 November 2024

DOI: 10.1002/jor.26018

RESEARCH ARTICLE

Journal of Orthopaedic Research®

Machine learning-based prediction of contralateral knee osteoarthritis development using the Osteoarthritis Initiative and the Multicenter Osteoarthritis Study dataset

Ji-Sahn Kim[1] | Byung Sun Choi[1] | Sung Eun Kim[1] | Yong Seuk Lee[2] | Do Weon Lee[3] | Du Hyun Ro[1,4,5,6]

[1]Department of Orthopaedic Surgery, Seoul National University Hospital, Seoul, South Korea

[2]Department of Orthopedic Surgery, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Gyeonggi-Do, South Korea

[3]Department of Orthopedic Surgery, Dongguk University Ilsan Hospital, Gyeonggi-do, South Korea

[4]Department of Orthopaedic Surgery, Seoul National University College of Medicine, Seoul, South Korea

[5]CONNECTEVE Co., Ltd, Seoul, South Korea

[6]Innovative Medical Technology Research Institute, Seoul National University Hospital, Seoul, South Korea

Correspondence

Do Weon Lee, Department of Orthopedic Surgery, Dongguk University Ilsan Hospital, 27 Dongguk-ro, Ilsandong-gu, Goyang-si, Gyeonggi-do, 10326, Republic of Korea.
Email: danthebest@naver.com

Du Hyun Ro, Department of Orthopedic Surgery, Seoul National University Hospital, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-Gu, Seoul, 03080, South Korea; CONNECTEVE Co., Ltd, Seoul, 06224, South Korea; Innovative Medical Technology Research Institute, Seoul National University Hospital, 101 Daehak-ro, Jongno-Gu, Seoul, 03080, South Korea.
Email: duhyunro@gmail.com

Abstract

Having osteoarthritis in one knee is reported as an independent risk factor for developing contralateral knee osteoarthritis (KOA). However, no study has been designed to predict the development of contralateral KOA (cKOA). The authors hypothesized that specific risk factors for cKOA development exist and that it could be accurately predicted with the assistance of machine learning. KOA was defined using the Kellgren-Lawrence grade (KLG) of 2 or higher. Data from 1353 unilateral KOA patients (900 from the Osteoarthritis Initiative [OAI] and 453 from the Multicenter Osteoarthritis Study [MOST]) over 4–5 years of follow-up were examined. The risk factors for cKOA development were analyzed, and a machine learning model was developed to predict cKOA using OAI as the development data set and MOST as the test data set. cKOA developed in 172 (19.1%) and 178 (39.3%) of the patients (OAI and MOST, respectively) over a period of 4–5 years. A machine learning model was developed using the Tree-based Pipeline Optimization Tool algorithm. This model utilized nine variables, including baseline lateral joint space narrowing grade of the contralateral knee (odds ratio 4.475). The area under the curve of the receiver operating characteristics curve, along with accuracy, precision, and F1-score, were recorded as 0.69, 0.60, 0.50, and 0.58, respectively, in the test data set. The development of cKOA could be effectively predicted using a limited number of variables through machine learning. Surgeons should consider the development of cKOA in patients with identified risk factors when managing KOA patients.

KEYWORDS

knee, machine learning, osteoarthritis, risk factor

1 | INTRODUCTION

Knee osteoarthritis (KOA) represents a predominant health issue, impeding the daily activities of over 650 million individuals globally.[1] Given its high prevalence, KOA demands comprehensive management strategies ranging from medication to surgery, imposing a significant socioeconomic impact on communities.[2] It has been demonstrated that the presence of KOA in one knee is an independent predictor for the development of contralateral KOA (cKOA), with accelerated joint space narrowing and a higher rate of arthroplasty in the contralateral knee.[3] Furthermore, individuals with unilateral KOA are more likely to sustain meniscal injuries in the contralateral knee.[4] Additionally, a prospective cohort study revealed that 80% of patients with unilateral KOA at baseline progressed to bilateral KOA within a span of 12 years.[5]

Several previous studies have explored the risk factors associated with developing cKOA in patients with unilateral KOA. Varus alignment of the unaffected knee has been identified as a significant risk factor for developing cKOA in obese females.[6] A high body mass index (BMI) has also been acknowledged as a potent risk factor for developing cKOA in patients with unilateral KOA.[7] Flexion contracture in KOA has been linked with higher Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) scores and flexion contracture in the contralateral knee.[8] For patients who have undergone unilateral total knee replacement arthroplasty (TKA), factors such as medial and lateral joint space narrowing,[9] baseline Kellgren-Lawrence grade (KLG),[9] and the tibiofemoral angle[9,10] of the contralateral knee were identified as risk factors for also requiring TKA on the contralateral knee. However, the majority of these studies focused on identifying risk factors rather than predicting the development of cKOA in individuals.

Machine learning has proven to be an effective tool for predicting the development of KOA.[11–13] One study achieved a prediction accuracy of 74.1% for KOA development using 55 risk factors and machine learning techniques.[11] Another study utilized 10 risk factors, achieving an area-under-the-curve (AUC) of 0.772 in the receiver operating characteristics (ROC) curve of the model.[12] Yet another model, which combined radiographic data with patient information, achieved an AUC of 0.75.[13] Despite these advances, few studies have been designed explicitly to predict cKOA development in KOA patients.

Identifying risk factors and predicting the development of cKOA could assist surgeons and patients with unilateral KOA, such as enlightening a need for early modification in lifestyle-provoking cKOA development. Therefore, the authors initially explored the risk factors for developing cKOA in individuals with unilateral KOA. Subsequently, a machine learning model was developed to predict the onset of cKOA in these patients. Two independent datasets were employed for model development and testing to minimize the risk of model overfitting. It was hypothesized that unilateral KOA patients might exhibit distinct risk factors for developing cKOA, apart from the general risk factors associated with KOA. Specifically, we proposed that the severity of unilateral KOA could act as a unique risk factor for cKOA development due to the increased mechanical burden on the contralateral knee in KOA patients.[14]

2 | METHODS

2.1 | Data acquisition

Data from the Osteoarthritis Initiative (OAI), a public, multi-center, 10-year observational study sponsored by the National Institutes of Health that provides clinical and radiological data on 4796 participants aged between 45 and 79 years,[15] were employed to validate hypotheses. In this study, the knees were classified according to the KLG[16] from 0 to 4 and KOA was defined by a KLG of 2 or higher. Among the initial 4796 participants, only those with KLG labels at baseline and at the 4-year follow-up for both knees were included, resulting in 3430 participants. Participants who underwent knee arthroplasty during the study were excluded. Additionally, 55 participants lacking basic demographic information such as age, sex, and race were excluded, leaving a total of 3375 participants. An independent data set was obtained from the Multicenter Osteoarthritis Study (MOST),[17] another large observational study that provides radiological and clinical data on individuals with KOA for model testing. The MOST study involved 3026 participants aged between 50 and 79 years. It was analyzed similarly, except that the follow-up period was 60 months due to a different study protocol. Among the 3026 participants, those with KLG labels at baseline and at the 60-month follow-up for both knees were included, totaling 1917 patients. Our retrospective cohort analysis is classified as Evidence Level III.

2.2 | Patient grouping

Participants with unilateral KOA at baseline were identified, resulting in 900 participants from the OAI and 453 from the MOST study. Participants with unilateral KOA at baseline who did not develop cKOA after the follow-up were categorized as the "cKOA (-)" group, and those who developed cKOA were categorized as the "cKOA (+)" group. A flow chart illustrating the patient selection process is depicted in Figure 1.

2.3 | Variables for cKOA development

Patient variables reported to be associated with KOA development in previous studies were selected for our study. Variables included age, sex (male and female), race (white and others), BMI,[7] presence of diabetes mellitus (DM),[18] alignment (degrees[°], with varus represented as negative),[10] total WOMAC score,[19] history of meniscectomy,[20] baseline KLG, and the grades for medial and lateral joint space narrowing according to the Osteoarthritis Research Society International (OARSI) atlas[9,21,22] ranging from 0 (no joint space narrowing) to 3 (severe narrowing) at baseline for all participants. All predictive variables were gathered from the initial clinic visit during the baseline period. These variables were collected from both the OA-affected knee and the contralateral knee of the participants and were analyzed as separate entities.

2.4 | Statistical analysis and machine learning

Demographic differences (sex, age, and race) between the cKOA (−) and cKOA (+) groups in the OAI and MOST datasets were examined using the chi-square test and independent *t*-test. Odd's ratio, its *p*-value, and 95% confidence intervals were determined for each variable using univariate logistic regression. A p-value < 0.05

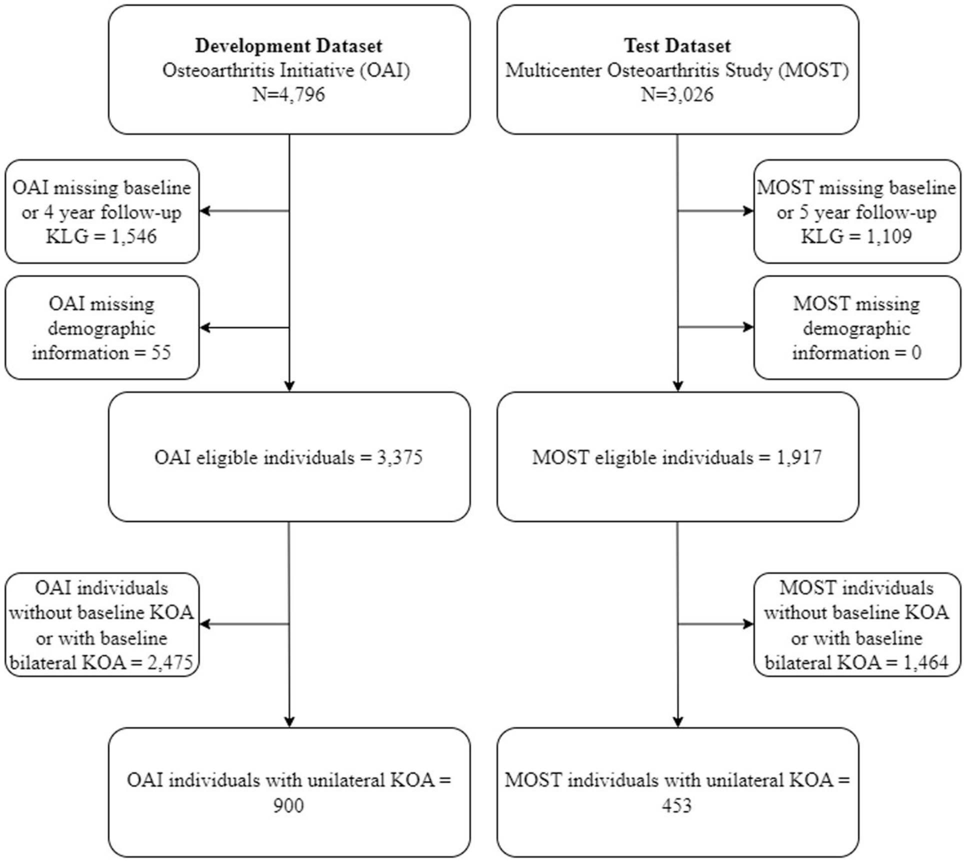


FIGURE 1 Flow chart detailing the selection of participants for the study. KLG refers to Kellgren-Lawrence grade, while KOA denotes knee osteoarthritis.

was deemed statistically significant. Variables were input into a machine learning model irrespective of their *p* values in the univariate analyses.

Using the predictor variables, a machine learning model was developed with the Tree-based Pipeline Optimization Tool (TPOT), an automated machine learning tool that explores the optimal combination of machine learning algorithms, feature selections, and hyperparameters.[23] All missing values were replaced with the median value for model development.[24] The OAI data set was divided randomly in a 3:1 ratio for training and validation. Various models with differing pipelines, selected features, and hyperparameter configurations were trained using the training set. The most effective model, as determined by TPOT, based on the highest validation F1 score, was then selected. The chosen model was evaluated using the MOST data set, and both the ROC curve and its AUC value were calculated. Measures such as the confusion matrix, F1 score, accuracy, precision, and recall for the test set were calculated. To assess the contributions of selected variables to the model and to visualize their importance, SHapley Additive exPlanations (SHAP) values.[25] were calculated and presented graphically. All statistical analyses, machine learning model development, and visualization were conducted using Python 3.9 (Python Software Foundation, http://python.org).

3 | RESULTS

3.1 | Demographic differences among the groups

cKOA developed in 172 (19.1%) and 178 (39.3%) of the patients (OAI and MOST, respectively). Demographic information and differences among cKOA (−) and cKOA (+) groups of OAI data set and MOST data set are each described in Table 1. In the OAI data set, sex (49.3% female in the cKOA (−) group; 62.8% female in the cKOA (+) group) and race (12.6% Nonwhite in the cKOA (−) group; 19.2% Nonwhite in the cKOA (+) group) were significantly different between the cKOA (−) and cKOA (+) groups ($p = 0.002$ and 0.035, respectively). For the MOST data set, sex (48.0% female in the cKOA (−) group; 65.7% female in the cKOA (+) group) was significantly different between the groups ($p < 0.001$).

3.2 | Variables of cKOA development

The odds ratio for the development of cKOA, 95% confidence interval, and *p*-value of the variables analyzed within the OAI and MOST datasets are detailed in Tables 2 and 3, respectively. In the OAI data set, variables such as sex, race, and BMI exhibited significant

TABLE 1 Demographic information and differences between the cKOA (−) and cKOA (+) groups in the OAI and MOST datasets.

OAI	cKOA (−) (*n* = 728)	cKOA (+) (*n* = 172)	*p*-Value
Age (years)	70.120 ± 9.204	69.035 ± 8.267	0.131
Sex (female)	359 (49.3%)	108 (62.8%)	0.002*
Race (non-white)	92 (12.6%)	33 (19.2%)	0.035*
MOST	**cKOA (−) (*n* = 275)**	**cKOA (+) (*n* = 178)**	***p*-Value**
Age (years)	62.007 ± 7.733	61.860 ± 7.587	0.841
Sex (female)	132 (48.0%)	117 (65.7%)	<0.001*
Race (non-white)	41 (14.9%)	29 (16.3%)	0.691

Note: Data are expressed as number (percentage %) for categorical variables and mean ± standard deviation for continuous variables; OAI, Osteoarthritis Initiative; MOST, Multicenter Osteoarthritis Study; cKOA (−), participants who did not develop contralateral knee osteoarthritis after follow-up; cKOA (+), participants who developed contralateral knee osteoarthritis after follow-up.

*p-Value below 0.05 indicates statistical significance.

differences between the cKOA (+) and cKOA (−) groups. Regarding the contralateral knee variables, baseline lateral joint space narrowing grade, prior meniscectomy history, baseline KLG, and total WOMAC score showed significant variations between the groups. Within the variables from the arthritic knee at baseline, the total WOMAC score demonstrated a significant difference between the groups.

In the MOST data set, significant differences were noted between groups in terms of sex and BMI. Variables of the contralateral knee, including baseline lateral joint space narrowing grade, baseline KLG, baseline medial joint space narrowing, and total WOMAC score, differed significantly between groups. Among the variables from the arthritic knee at baseline, both total WOMAC score and prior meniscectomy history showed significant differences between the groups.

3.3 | Machine learning model development and performance examination

The optimal model identified by TPOT was built using Linear Support Vector Classification, a type of Support Vector Machine typically employed for classifying linearly separable data.[26] The best predictor variables were selected through a feature selection process within TPOT, and they included the baseline lateral joint space narrowing grade of the contralateral knee, meniscectomy history of the contralateral knee, baseline KLG of the contralateral knee, sex, race, baseline KLG of the OA knee, BMI, total WOMAC score of the contralateral knee, and total WOMAC score of the OA knee. ROC curves for the validation data set and the test data set are shown in Figure 2. The AUC for the ROC curves of the validation and test datasets were 0.67 and 0.69, respectively. The confusion matrix is depicted in Figure 3, and the test accuracy, precision, recall, and F1-score were 0.60, 0.50, 0.70, and 0.58, respectively.

SHAP plot and mean absolute SHAP value of each variable are displayed in Figure 4. The variables identified as significant, in descending order of mean absolute SHAP value, include sex, baseline lateral joint space narrowing grade of the contralateral knee, history of meniscectomy in the contralateral knee, BMI, baseline KLG of the OA knee, race, baseline KLG of the contralateral knee, total WOMAC score of the contralateral knee, and total WOMAC score of the OA knee.

4 | DISCUSSION

The development of cKOA in a KOA patient could be well-predicted using a machine learning model with a select few variables. Importantly, this model was evaluated on an independent data set, unlike some other models[11,12] that predict KOA progression. Our study elucidated distinct risk factors for cKOA development in patients with unilateral KOA through both univariate logistic regression and machine learning analysis. Notably, significant risk factors included variables that represent the severity of unilateral KOA, such as the total WOMAC score of the OA knee.

Nine variables identified after the feature selection process suggest optimal performance. BMI, sex, total WOMAC score of the contralateral knee, previous meniscectomy of the contralateral knee, and baseline KLG of the contralateral knee were prominent risk factors in both machine learning and logistic regression analyses, aligning with prior research.[7,9,14,20,27,28] Figure 4 of SHAP values illustrates that following sex—which is already recognized as a strong risk factor for KOA[28]—the lateral joint space narrowing grade of the contralateral knee emerged as the second most influential feature in our model. This finding, coupled with its high odds ratio of 4.475 in univariate logistic regression, underscores the importance of early baseline lateral joint space narrowing grade in the contralateral knee as a predictor of cKOA development, a divergence from previous work.[9] It has been reported that the lateral meniscus endures higher axial loading than the medial meniscus,[29] and lateral meniscectomy has been identified as a more significant risk factor for KOA development than medial meniscectomy.[30] Additionally, meniscal alteration has been substantially linked to joint space narrowing.[31] Consequently, lateral meniscal alteration in the contralateral knee, which could manifest as lateral joint space narrowing, may increase the overall mechanical loading on the contralateral knee and lead to cKOA development. Future research focusing on gait mechanics and the lateral meniscus could further elucidate the role of baseline lateral joint space narrowing in cKOA development.

It is notable that both the KLG and the total WOMAC score of the arthritic knee at baseline contributed significantly to the prediction made by our model, corroborating the hypothesis that the severity of KOA can act as a significant risk factor for cKOA development, alongside other well-known risk factors associated with KOA progression. Due to pain, patients with severe KOA typically avoid placing weight on the affected knee, which can lead to increased stress on the contralateral knee as KOA progresses. Increases in joint contact forces

TABLE 2 Univariate logistic regression analysis in the OAI data set.

OAI	cKOA (−) (*n* = 728)	cKOA (+) (*n* = 172)	Missing value (group)	Odds ratio (OR)	95% confidence interval (CI)	*p*-Value
Sex (female)	359 (49.3%)	108 (62.8%)	0	1.733	1.236–2.451	0.002*
Race (non-white)	92 (12.6%)	33 (19.2%)	0	1.642	1.047–2.519	0.027*
Age (years)	70.120 ± 9.204	69.035 ± 8.267	0	1.002	0.985–1.019	0.854
BMI (kg/m^2)	28.299 ± 4.173	29.948 ± 4.698	1 (cKOA (−))	1.091	1.050–1.134	<0.001*
Diabetes mellitus	94 (12.9%)	24 (14.0%)	0	1.094	0.662–1.746	0.716
Knee alignment (cont) ([°], varus: negative)	−0.437 ± 3.446	−0.750 ± 4.128	0	0.976	0.931–1.022	0.303
Total WOMAC score (cont)	7.085 ± 11.163	10.344 ± 14.741	3 (cKOA (−)), 1 (cKOA (+))	1.020	1.007–1.032	0.002*
Prior meniscectomy History (cont)	15 (2.1%)	10 (5.8%)	0	2.934	1.255–6.581	0.010*
Baseline KLG (cont) (0, 1)	382 (52.5%), 346 (47.5%)	51 (29.7%), 121 (70.4%)	0	2.619	1.841–3.773	<0.001*
Baseline MJSN (cont) (0,1)	571 (78.4%), 157 (21.6%)	126 (73.3%), 46 (26.7%)	0	1.328	0.901–1.933	0.145
Baseline LJSN (cont) (0,1)	716 (98.4%), 12 (1.6%)	160 (93.0%), 12 (7.0%)	0	4.475	1.955–10.247	<0.001*
Knee alignment (OA) ([°], varus: negative)	−0.383 ± 3.688	−0.791 ± 4.442	0	0.973	0.931–1.016	0.211
Total WOMAC score (OA)	12.273 ± 14.275	14.812 ± 16.368	6 (cKOA (−)), 1 (cKOA (+))	1.011	1.000–1.022	0.044*
Prior meniscectomy History (OA)	22 (3.0%)	4 (2.3%)	0	0.764	0.221–2.028	0.625
Baseline KLG (OA) (2,3,4)	515 (70.7%), 164 (22.5%), 49 (6.7%)	107 (62.2%), 51 (29.7%), 14 (8.1%)	0	1.284	0.989–1.653	0.057
Baseline MJSN (OA) (0,1,2,3)	310 (42.6%), 267 (36.7%), 123 (16.9%), 28 (3.8%)	73 (42.4%), 50 (29.1%), 43 (25.0%), 6 (3.5%)	0	1.106	0.912–1.337	0.3
Baseline LJSN (OA) (0,1,2,3)	602 (82.7%), 60 (8.2%), 45 (6.2%), 21 (2.9%)	142 (82.6%), 13 (7.6%), 9 (5.2%), 8 (4.7%)	0	1.052	0.833–1.306	0.657

Note: Data are expressed as number (percentage %) for categorical variables and mean ± standard deviation for continuous variables; OAI, Osteoarthritis Initiative; cKOA (−), participants who did not develop contralateral knee osteoarthritis after follow-up; cKOA (+), participants who developed contralateral knee osteoarthritis after follow-up; BMI, body mass index; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; KLG, Kellgren-Lawrence grade; MJSN, medial joint space narrowing grade; LJSN, lateral joint space narrowing grade; cont, data from the contralateral knee at baseline; OA, data from the arthritic knee at baseline.

*p-Value below 0.05 indicates statistical significance.

TABLE 3 Univariate logistic regression analysis in the MOST data set.

MOST	cKOA (−) (*n* = 275)	cKOA (+) (*n* = 178)	Missing value (group)	Odds ratio (OR)	95% confidence interval (CI)	*p*-Value
Sex (female)	132 (48.0%)	117 (65.7%)	0	2.079	1.412–3.077	<0.001*
Race (non-white)	41 (14.9%)	29 (16.3%)	0	1.111	0.662–1.865	0.691
Age (years)	62.007 ± 7.733	61.860 ± 7.587	0	0.997	0.973–1.022	0.841
BMI (kg/m^2)	29.988 ± 5.724	31.509 ± 6.338	0	1.043	1.011–1.077	0.009*
Diabetes mellitus	28 (10.3%)	21 (12.1%)	4 (cKOA (−)), 4 (cKOA (+))	1.199	0.651–2.179	0.554
Knee alignment (cont) ([°], varus: negative)	−1.190 ± 2.533	−0.890 ± 2.635	6 (cKOA (−)), 6 (cKOA (+))	1.047	0.971–1.129	0.233
Total WOMAC score (cont)	14.054 ± 13.797	22.028 ± 16.765	3 (cKOA (−)), 1 (cKOA (+))	1.037	1.022–1.049	<0.001*
Prior meniscectomy History (cont)	6 (2.2%)	5 (2.8%)	2 (cKOA (−))	1.291	0.367–4.349	0.677
Baseline KLG (cont) (0, 1)	158 (57.5%), 117 (42.5%)	64 (36.0%), 114 (64.0%)	0	2.405	1.636–3.560	<0.001*
Baseline MJSN (cont) (0,1)	233 (84.7%), 42 (15.3%)	125 (70.2%), 53 (29.8%)	0	2.352	1.489–3.739	<0.001*
Baseline LJSN (cont) (0,1)	273 (99.3%), 2 (0.7%)	164 (92.1%), 14 (7.9%)	0	11.652	3.203–74.813	0.001*
Knee alignment (OA) ([°], varus: negative)	−1.776 ± 3.938	−1.786 ± 3.896	7 (cKOA (−)), 5 (cKOA (+))	0.999	0.952–1.049	0.979
Total WOMAC score (OA)	16.307 ± 15.410	24.113 ± 17.533	2 (cKOA (−)), 1 (cKOA (+))	1.029	1.017–1.042	<0.001*
Prior meniscectomy History (OA)	70 (25.7%)	27 (15.3%)	4 (cKOA (−)), 1 (cKOA (+))	0.519	0.313–0.841	0.009*
Baseline KLG (OA) (2,3,4)	179 (65.1%), 72 (26.2%), 24 (8.7%)	106 (59.6%), 62 (34.8%), 10 (5.6%)	0	1.063	0.787–1.430	0.688
Baseline MJSN (OA) (0,1,2,3)	99 (36.0%), 102 (37.1%), 58 (21.1%), 16 (5.8%)	49 (27.5%), 74 (41.6%), 48 (27.0%), 7 (3.9%)	0	1.144	0.921–1.422	0.223
Baseline LJSN (OA) (0,1,2,3)	216 (78.5%), 32 (11.6%), 17 (6.2%), 10 (3.6%)	138 (77.5%), 23 (12.9%), 13 (7.3%), 4 (2.2%)	0	0.992	0.763–1.279	0.948

Note: Data are expressed as number (percentage %) for categorical values and mean ± standard deviation for continuous variables; MOST, Multicenter Osteoarthritis Study; cKOA (−), participants who did not develop contralateral knee osteoarthritis after follow-up; cKOA (+), participants who developed contralateral knee osteoarthritis after follow-up; BMI, body mass index; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; KLG, Kellgren-Lawrence grade; MJSN, medial joint space narrowing grade; LJSN, lateral joint space narrowing grade; cont, data from the contralateral knee at baseline; OA, data from the arthritic knee at baseline.

*p-Value below 0.05 indicates statistical significance.

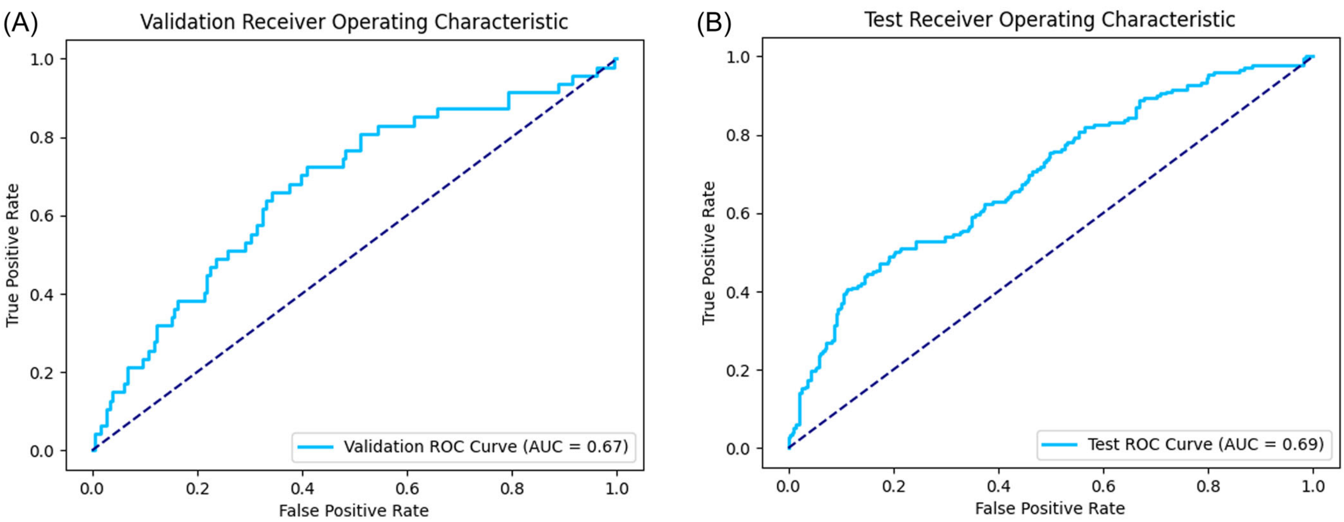
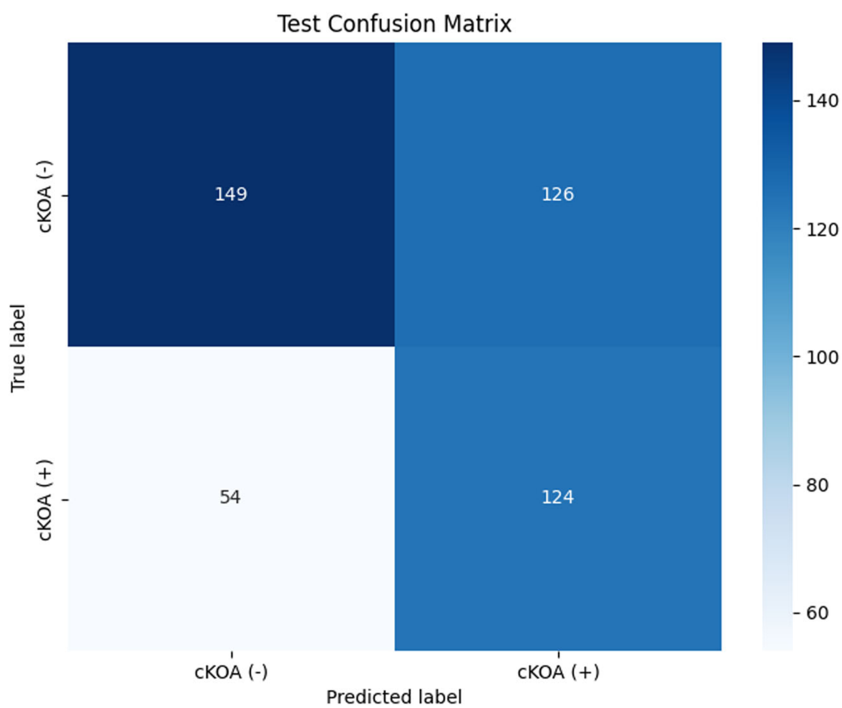


FIGURE 2 Receiver operating characteristic (ROC) curve for the validation data set from the OAI (A) and the test data set from MOST (B). AUC represents the area under the curve of the ROC curve. MOST, Multicenter Osteoarthritis Study; OAI, Osteoarthritis Initiative.

FIGURE 3 Confusion matrix for the test data set. cKOA (−) signifies participants who did not develop contralateral knee osteoarthritis after follow-up; cKOA (+) identifies participants who developed contralateral knee osteoarthritis after follow-up.



on the contralateral knee in KOA patients have indeed been reported.[14] Gait asymmetry, irrespective of whether the condition is unilateral or bilateral, has been documented in KOA patients experiencing unilateral knee pain.[32] Consequently, our findings indicate a possibility that as KOA deteriorates, the burden on the contralateral knee may augment due to altered gait mechanics induced by pain, thereby leading to the development of cKOA. The significance of the total WOMAC score of the OA knee in univariate logistic regression across both datasets emphasizes the importance of subjective discomfort in the OA knee during this process.

Well-known risk factors for KOA progression, such as BMI,[33] total WOMAC score,[34] and baseline KLG.[34] in the contralateral knee, also act as risk factors for cKOA development. However, despite previous findings indicating knee malalignment as a risk factor for KOA progression,[35] its predictive significance for cKOA development was not validated. This is consistent with earlier reports that knee alignment functions as a marker for progression but not as a predictive risk factor for incident KOA development.[36] This evidence further implies that gait alterations due to KOA may place a greater mechanical burden on the contralateral knee than any changes in the alignment of the knee itself,[37] thereby emerging as a more critical risk factor.

Our insights concerning distinct risk factors for the contralateral knee underscore the necessity of closely monitoring lateral joint space narrowing grade while addressing other well-known modifiable risk

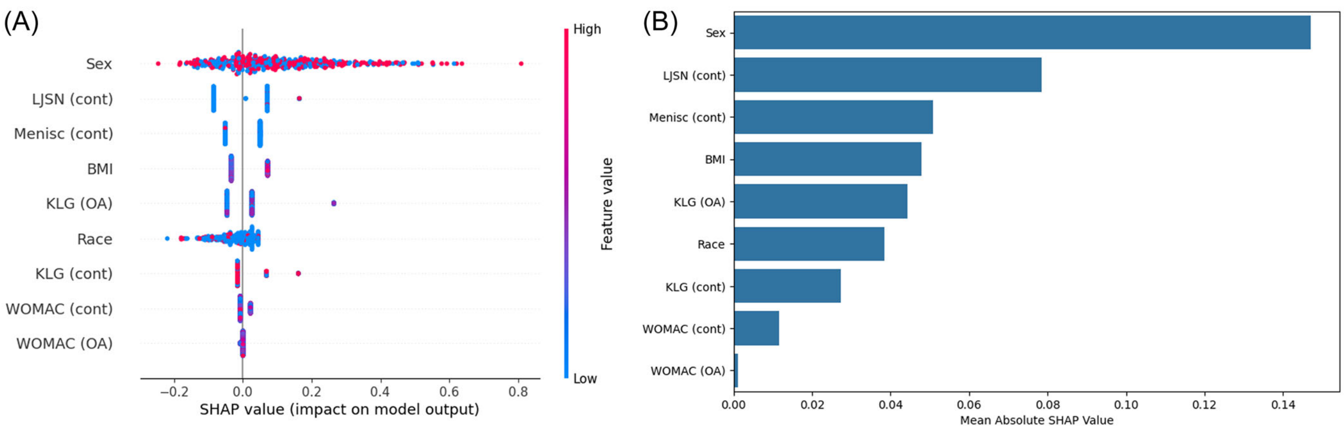


FIGURE 4 SHapley Additive exPlanations (SHAP) plot (A) and mean absolute SHAP values (B) highlight the importance of predictor variables in descending order. Included are LJSN, lateral joint space narrowing grade; Menisc, history of prior meniscectomy; BMI, body mass index; KLG, Kellgren-Lawrence grade; WOMAC, total Western Ontario and McMaster Universities Osteoarthritis Index score; cont, data from the contralateral knee at baseline; OA, data from the arthritic knee at baseline.

factors of KOA, such as BMI. Although the high incidence of cKOA development or contralateral degenerative changes in KOA patients has been previously reported,[3,5] specific risk factors for the affected OA knee influencing the contralateral knee have not been clearly identified in prior studies. Our results indicate that symptoms represented by the total WOMAC score are significant risk factors that influence the progression of cKOA. Thus, aggressive pain management for the OA knee is crucial in preventing cKOA among patients with unilateral KOA.

This study is not without certain limitations. First, variables not collected in the OAI and MOST datasets, such as the presence of osteophytes, trauma history, and gait parameters, could not be analyzed. Second, the test AUC value (0.6921) of our machine learning model was relatively lower than that of pre-existing models that predict KOA progression or development.[12,13] However, our study is valuable as it was tested on an independent data set, unlike previous models.[12] Additionally, to the best of our knowledge, this is the first trial aimed at predicting the development of cKOA. Third, OAI and MOST datasets had discrepancy in cKOA incidence rates of 19.1% and 39.3%, respectively. This is expected to result from their subtly different inclusion criteria including age,[15,17] and previous studies also reported discrepancy of KOA development between the two datasets.[38,39] This independency among the two datasets adds broader applicability to our model. Lastly, our research focuses exclusively on the development of cKOA in patients with unilateral KOA. Future studies examining patients who develop unilateral or bilateral KOA from bilateral healthy knees could enhance our understanding of the differing pathophysiologies of unilateral and bilateral KOA.

5 | CONCLUSIONS

In this study, the development of cKOA in a KOA patient could be accurately predicted using a few variables through machine learning. This study identified specific risk factors for developing cKOA in patients with unilateral KOA, including the severity of KOA in the arthritic knee, alongside the well-established risk factors for KOA development. When consulting KOA patients, surgeons should consider the potential for cKOA development and assess risk factors in both knees, such as high lateral joint space narrowing in the contralateral knee and a high total WOMAC score in the OA knee.

AUTHOR CONTRIBUTIONS

Ji-Sahn Kim, Byung Sun Choi, Sung Eun Kim, Do Weon Lee, and Du Hyun Ro were responsible for the conceptualization of the study. Data curation was carried out by Ji-Sahn Kim, Do Weon Lee, and Du Hyun Ro. Formal analysis was conducted by Ji-Sahn Kim and Do Weon Lee. Methodology was developed by Ji-Sahn Kim, Byung Sun Choi, Yong Seuk Lee, Sung Eun Kim, Do Weon Lee, and Du Hyun Ro. Software was handled by Ji-Sahn Kim and Do Weon Lee. Validation was performed by Ji-Sahn Kim and Do Weon Lee. Investigation was performed by Ji-Sahn Kim and Do Weon Lee. The original draft was written by Ji-Sahn Kim. The writing—review and editing were done by Ji-Sahn Kim, Byung Sun Choi, Sung Eun Kim, Yong Seuk Lee, Do Weon Lee, and Du Hyun Ro. All authors approved the manuscript.

ACKNOWLEDGMENTS

The study data was obtained from the OAI and MOST. OAI is a public-private partnership composed of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, executed by the OAI Study Investigators. Private funding partners comprise Merck Research Laboratories, Novartis Pharmaceuticals Corporation, GlaxoSmithKline, and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. This manuscript was prepared employing an OAI public use data set and does not represent the views or opinions of the OAI investigators, the NIH, or its private funding partners. MOST consists of four cooperative grants: U01 AG18820 David T. Felson (Boston University); U01 AG18832 James Torner (University of Iowa); U01 AG18947 Cora E. Lewis (University of Alabama at Birmingham); U01 AG19069 Michael C. Nevitt (University of California, San Francisco), sponsored by the National Institutes of Health, a division of the Department of Health and Human Services, and directed by MOST study investigators. This manuscript was prepared utilizing MOST Data and does not reflect the views or opinions of MOST, the MOST investigators, their affiliated institutions, or the University of California.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ORCID

Ji-Sahn Kim http://orcid.org/0009-0003-0047-8442
Byung Sun Choi http://orcid.org/0000-0002-4492-4358
Sung Eun Kim http://orcid.org/0000-0002-1954-9875
Yong Seuk Lee http://orcid.org/0000-0002-4641-305X
Do Weon Lee https://orcid.org/0000-0001-5139-9478
Du Hyun Ro http://orcid.org/0000-0001-6199-908X

REFERENCES

1. Cui A, Li H, Wang D, Zhong J, Chen Y, Lu H. Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. *EClinicalMedicine*. 2020;29-30:100587.
2. Murphy L, Helmick CG. The impact of osteoarthritis in the United States: a population-health perspective. *AJN, Am J Nurs*. 2012;112:S13-S19.
3. Dai Z, Yang T, Liu J. Contralateral knee osteoarthritis is a risk factor for ipsilateral knee osteoarthritis progressing: a case control study. *BMC Musculoskelet Disord*. 2024;25:190.
4. Roemer FW, Eckstein F, Duda G, et al. Frequencies of MRI-detected structural pathology in knees without radiographic OA and worsening over three years: how relevant is contralateral radiographic osteoarthritis? *Osteoarthritis Cartilage Open*. 2020;1:100014.
5. Metcalfe AJ, Andersson ML, Goodfellow R, Thorstensson CA. Is knee osteoarthritis a symmetrical disease? Analysis of a 12 year prospective cohort study. *BMC Musculoskelet Disord*. 2012;13:153.
6. Mazzuca SA, Brandt KD, Lane KA, Chakr R. Malalignment and subchondral bone turnover in contralateral knees of overweight/obese women with unilateral osteoarthritis: implications for bilateral disease. *Arthritis Care Res*. 2011;63:1528-1534.
7. Spector TD, Hart DJ, Doyle DV. Incidence and progression of osteoarthritis in women with unilateral knee disease in the general population: the effect of obesity. *Ann Rheum Dis*. 1994;53:565-568.
8. Campbell TM, Trudel G. Knee flexion contracture associated with a contracture and worse function of the contralateral knee: data from the Osteoarthritis Initiative. *Arch Phys Med Rehabil*. 2020;101:624-632.
9. Huang X, Li H, Chen B, et al. The incidence and risk factors of contralateral knee arthroplasty after primary unilateral unicompartmental knee arthroplasty. *Medicine*. 2021;100:e26825.
10. Okamoto T, Yamanashi Y, Ikemoto T, et al. Risk factors for contralateral total knee arthroplasty after unilateral total knee arthroplasty. *J Orthop Sci*. 2023;28:1311-1316.
11. Kokkotis C, Moustakidis S, Giakas G, Tsaopoulos D. Identification of risk factors and machine learning-based prediction models for knee osteoarthritis patients. *Appl Sci*. 2020;10:6797.
12. Joseph GB, McCulloch CE, Nevitt MC, Link TM, Sohn JH. Machine learning to predict incident radiographic knee osteoarthritis over 8 years using combined MR imaging features, demographics, and clinical factors: data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage*. 2022;30:270-279.
13. Tiulpin A, Klein S, Bierma-Zeinstra SMA, et al. Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. *Sci Rep*. 2019;9:20038.
14. Gustafson JA, Anderton W, Sowa GA, Piva SR, Farrokhi S. Dynamic knee joint stiffness and contralateral knee joint loading during prolonged walking in patients with unilateral knee osteoarthritis. *Gait Posture*. 2019;68:44-49.
15. Lester G. Clinical research in OA—the NIH Osteoarthritis Initiative. *J Musculoskelet Neuronal Interact*. 2008;8:313-314.
16. Schiphof D, Boers M, Bierma-Zeinstra SMA. Differences in descriptions of Kellgren and Lawrence grades of knee osteoarthritis. *Ann Rheum Dis*. 2008;67:1034-1036.
17. Segal NA, Nevitt MC, Gross KD, et al. The Multicenter Osteoarthritis Study: opportunities for rehabilitation research. *PM&R*. 2013;5:647-654.
18. Schett G, Kleyer A, Perricone C, et al. Diabetes is an independent predictor for severe osteoarthritis. *Diabetes Care*. 2013;36:403-409.
19. Deveza LA, Downie A, Tamez-Peña JG, Eckstein F, Van Spil WE, Hunter DJ. Trajectories of femorotibial cartilage thickness among persons with or at risk of knee osteoarthritis: development of a prediction model to identify progressors. *Osteoarthritis Cartilage*. 2019;27:257-265.
20. Poulsen E, Goncalves GH, Bricca A, Roos EM, Thorlund JB, Juhl CB. Knee osteoarthritis risk is increased 4-6 fold after knee injury—a systematic review and meta-analysis. *Br J Sports Med*. 2019;53:1454-1463.
21. Ratzlaff C, Ashbeck EL, Guermazi A, Roemer FW, Duryea J, Kwoh CK. A quantitative metric for knee osteoarthritis: reference values of joint space loss. *Osteoarthritis Cartilage*. 2018;26:1215-1224.
22. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage*. 2007;15(suppl A):A1-A56.
23. Radzi SFM, Karim MKA, Saripan MI, Rahman MAA, Isa INC, Ibahim MJ. Hyperparameter tuning and pipeline optimization via grid search method and tree-based AutoML in breast cancer prediction. *J Pers Med*. 2021;11:978.
24. Maniruzzaman M, Rahman MJ, Al-MehediHasan M, et al. Accurate diabetes risk stratification using machine learning: role of missing value and outliers. *J Med Syst*. 2018;42:92.
25. Ogami C, Tsuji Y, Seki H, et al. An artificial neural network-pharmacokinetic model and its interpretation using shapley additive explanations. *CPT: Pharmacometrics Syst Pharmacol*. 2021;10:760-768.
26. Zhang D, Chen Y, Chen Y, et al. Heart disease prediction based on the embedded feature selection method and deep neural network. *J Healthc Eng*. 2021;2021:6260022.
27. Iqbal MN, Haidri FR, Motiani B, Mannan A. Frequency of factors associated with knee osteoarthritis. *J Pak Med Assoc*. 2011;61:786-789.
28. Hussain SM, Cicuttini FM, Alyousef B, Wang Y. Female hormonal factors and osteoarthritis of the knee, hip and hand: a narrative review. *Climacteric*. 2018;21:132-139.
29. Englund M, Lohmander LS. Risk factors for symptomatic knee osteoarthritis fifteen to twenty-two years after meniscectomy. *Arthritis Rheumatism*. 2004;50:2811-2819.
30. Scott CEH, Nutton RW, Biant LC. Lateral compartment osteoarthritis of the knee: biomechanics and surgical management of end-stage disease. *Bone Joint J*. 2013;95-B:436-444.
31. Hunter DJ, Zhang YQ, Tu X, et al. Change in joint space width: hyaline articular cartilage loss or alteration in meniscus? *Arthritis Rheumatism*. 2006;54:2488-2495.

32. Creaby MW, Bennell KL, Hunt MA. Gait differs between unilateral and bilateral knee osteoarthritis. *Arch Phys Med Rehabil*. 2012;93: 822-827.
33. Lee R, Kean WF. Obesity and knee osteoarthritis. *Inflammopharmacology*. 2012;20:53-58.
34. Collins JE, Neogi T, Losina E. Trajectories of structural disease progression in knee osteoarthritis. *Arthritis Care Res*. 2021;73:1354-1362.
35. Brouwer GM, Tol AWV, Bergink AP, et al. Association between valgus and varus alignment and the development and progression of radiographic osteoarthritis of the knee. *Arthritis Rheumatism*. 2007;56: 1204-1211.
36. Hunter DJ, Niu J, Felson DT, et al. Knee alignment does not predict incident osteoarthritis: the Framingham osteoarthritis study. *Arthritis Rheumatism*. 2007;56:1212-1218.
37. Jansen MP, Hodgins D, Mastbergen SC, et al. Can gait patterns be explained by joint structure in people with and without radiographic knee osteoarthritis? Data from the IMI-APPROACH cohort. *Skeletal Radiol*. 2024;53:2409-2416.
38. Øiestad BE, White DK, Booton R, et al. Longitudinal course of physical function in people with symptomatic knee osteoarthritis: data from the Multicenter Osteoarthritis Study and the Osteoarthritis Initiative. *Arthritis Care Res*. 2016;68:325-331.
39. Felson DT, Niu J, Gross KD, et al. Valgus malalignment is a risk factor for lateral knee osteoarthritis incidence and progression: findings from the Multicenter Osteoarthritis Study and the Osteoarthritis Initiative. *Arthritis Rheumatism*. 2013;65:355-362.

How to cite this article: Kim J-S, Choi BS, Kim SE, Lee YS, Lee DW, Ro DH. Machine learning-based prediction of contralateral knee osteoarthritis development using the Osteoarthritis Initiative and the Multicenter Osteoarthritis Study dataset. *J Orthop Res*. 2024;1-10. doi:10.1002/jor.26018