Am J Transl Res 2026;18(8):7118-7140
www.ajtr.org /ISSN:1943-8141/AJTR0174487

Original Article

A retrospective predictive model for worsening knee osteoarthritis pain by integrating baseline gait biomechanics and wearable gait parameters

Kun Li*, Yun Yang*, Jie Zhou, Liangyu Lu

*Department of Orthopedics, Pudong New Area People's Hospital, Shanghai 201299, China. *Equal contributors and co-first authors.*

Received June 3, 2026; Accepted July 8, 2026; Epub August 15, 2026; Published August 30, 2026

Abstract: Objective: To investigate the performance of a predictive model for worsening knee osteoarthritis (KOA) pain that integrates baseline clinical characteristics, gait biomechanical parameters and wearable device-derived daily activity monitoring data and to evaluate the incremental predictive value of each module. Methods: A total of 984 KOA patients who visited the Department of Orthopedics, Pudong New Area People's Hospital between January 2018 and December 2023 were retrospectively included and randomly split into a training set (n = 689) and a test set (n = 295) at a 7:3 ratio. The primary outcome was a composite endpoint defined as an increase of ≥ 2 points in the WOMAC pain subscale score within 12 months after enrollment, accompanied by subjective patient-reported pain worsening (incidence, 28.0%). In the training set, candidate variables were initially screened by univariate analysis (P < 0.05) and final predictors were selected using LASSO regression with 10-fold cross-validation (lambda.1se = 0.012253). Four logistic regression models (clinical model, clinical + gait model, clinical + wearable model and fusion model) were constructed using a stepwise modular incremental strategy. Model discrimination was assessed using the area under the receiver operating characteristic curve (AUC), calibration using the Brier Score and calibration curves, incremental predictive value using integrated discrimination improvement (IDI) and net reclassification improvement (NRI) and clinical utility using decision curve analysis (DCA). Variable importance was assessed using multivariable logistic regression and SHAP method. Results: A total of 17 variables were selected by LASSO (7 clinical variables, 4 gait-related variables and 6 wearable-derived variables). In the test set, the AUCs of the clinical model, clinical + gait model, clinical + wearable model and fusion model were 0.673 (95% CI: 0.605-0.740), 0.753 (95% CI: 0.695-0.811), 0.756 (95% CI: 0.695-0.816) and 0.815 (95% CI: 0.761-0.869), respectively. The fusion model achieved a Brier score of 0.144, a calibration intercept of -0.261 and a calibration slope of 0.856. Compared with the clinical model, the IDIs of the clinical + gait, clinical + wearable and fusion models were 0.056, 0.123 and 0.217, respectively and the continuous NRIs were 0.606, 0.614 and 0.697, respectively (all P < 0.05). Multivariable logistic regression identified baseline WOMAC stiffness score (OR = 1.912, 95% CI: 1.583-2.331, P < 0.001), analgesic use (OR = 1.753, P = 0.009) and daytime activity fragmentation index (per 0.1-unit increase, OR = 1.948, 95% CI: 1.597-2.402, P < 0.001) as independent variables risk factors for worsening pain. Regular exercise habits (OR = 0.630, P = 0.033), 20-m walk speed (per 0.1 m/s increase, OR = 0.748, 95% CI: 0.647-0.859, P < 0.001) and sedentary break frequency (OR = 0.780 per break/hour, P = 0.023) were independently associated with decreased odds of pain worsening. SHAP analysis showed that baseline WOMAC stiffness score (mean |SHAP| = 0.101) and daytime activity fragmentation index (mean |SHAP| = 0.098) were the two most important predictors in terms of global contribution. DCA indicated that the fusion model yielded the highest net benefit within a risk threshold probability range of 0.05-0.35. Subgroup sensitivity analyses confirmed the robustness of the fusion model (AUC > 0.79 across all subgroups). Conclusion: An integrated predictive model incorporating baseline clinical characteristics, gait biomechanics and wearable device-derived activity data can markedly enhance discrimination and risk classification of worsening KOA pain. These findings suggest that assessment of the biomechanical gait and monitoring of activity through wearable devices may provide additional predictive value beyond traditional clinical variables in identifying KOA patients at higher risk of pain worsening.

Keywords: Knee osteoarthritis, pain worsening, gait biomechanics, wearable devices, prediction model, LASSO regression, SHAP analysis

https://doi.org/10.62347/ORTB8274

Introduction

Knee osteoarthritis (KOA) is the most prevalent form of degenerative joint disease worldwide, affecting an estimated 650 million individuals aged ≥ 40 years. KOA is also a leading cause of chronic pain and functional limitation among middle-aged and older adults [1]. Epidemiological surveys in China have reported that the prevalence of KOA is approximately 21.7% in individual aged ≥ 55 years. Furthermore, among KOA patients, pain worsening occurs in approximately 30%-40% of cases, manifesting as acute exacerbations characterized by a sharp increase in pain intensity and a marked decline in daily functional capacity over a short period of time [2]. Pain exacerbation is the primary reason for seeking healthcare help and represents a key contributor to loss of joint function and deterioration in quality of life. However, the pathophysiology of KOA pain deteriorations is complex, involving multidimensional interaction among structural joint degeneration, biomechanical abnormalities, central sensitization and psychosocial factors. At present, clinical practice lacks reliable tools for prospective identification of patients at high risk of pain deterioration.

Current risk assessment of KOA pain aggravating mainly relies on baseline pain severity, radiographic Kellgren-Lawrence (KL) grade and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [3], which are static clinical indicators. Although longitudinal cohort studies have established moderately strong associations between these traditional indicators and pain worsening, their predictive performance remains limited; models based solely on clinical variables typically achieve an AUC of ≤ 0.70 [4].

Gait biomechanical parameters (e.g., walking speed, chair stand time and single-leg stance time) as well as wearable device-derived daily activity indicators (e.g., step count and activity fragmentation index) have received increasing attention [5]. Alteration in load distribution resulting from gait abnormalities has been assumed to exacerbate pain, creating a vicious cycle of 'pain-gait abnormality-pain worsening' [6]. Meanwhile, continuous monitoring via wearable devices can provide ecologically valid, real-world activity data that cannot be obtained in lab settings [7]. However, few studies have jointly incorporated gait biomechanics and wearable device-derived activity parameters to predict KOA pain worsening. Thus, it remains unclear whether these two objectively measured indicators can provide incremental predictive value for worsening pain beyond traditional clinical variables.

The limitations of existing studies are as follows. First, most studies have exclusively examined either gait or wearable parameters and their cross-sectional association with KOA pain and have rarely validated their predictive utility within a unified longitudinal framework. Second, gait biomechanical parameters and wearable activity measures reflect different dimensions of functional status: while the former captures maximal exercise capacity under controlled settings, the latter reflects habitual activity patterns in free-living settings. The potential complementarity between these two has not been sufficiently evaluated. Third, modern variable selection methods such as least absolute shrinkage and selection operator (LASSO) regression remain underutilized in KOA pain worsening prediction, limiting the ability to efficiently identify optimal predictor combination from high-dimensional candidate variables [8].

Based on this background, the present single-center retrospective cohort study enrolled KOA patients and systematically collected baseline clinical characteristics, laboratory-based gait biomechanical parameters and 7-day wearable device-derived activity monitoring data. Predictive variables were selected using LASSO regression and predictive models integrating clinical, gait and wearable modules were constructed. The incremental predictive value of each module for 12-month worsening KOA pain was further evaluated, aiming to determine whether the integration of multidimensional objective assessment tools can improve early risk stratification in KOA.

Methods

Study design and sample size calculation

This study was a retrospective analysis based on a previously established KOA longitudinal follow-up database. Between January 2018 and December 2023, KOA patients who visited the Department of Orthopedics and the Department of Rehabilitation Medicine at Pudong

New Area People's Hospital were prospectively and consecutively enrolled into the database. All participants underwent standardized clinical evaluation, gait biomechanical assessment and 7-day wearable device-based activity monitoring according to a unified protocol. Follow-up assessments were completed 12 months later.

For the present analysis, a retrospective extraction of data was performed from patients meeting the predefined inclusion and exclusion criteria. The primary objective was the development of a predictive model for KOA pain worsening, which guided the sample size considerations. Among 984 patients, 276 (28.0%) experienced a pain worsening event during the 12-month follow-up. A total of 33 candidate variables were initially considered; after expansion of categorical variables (including Kellgren-Lawrence grade and affected knee laterality), the final number of features was 35. Accordingly, the events-per-variable (EPV) ratio was approximately 7.9. According to the traditional EPV ≥ 10 rule, this is considered insufficient. To mitigate overfitting during variable selection and model construction, LASSO regression with 10-fold cross-validation was adopted. Previous simulation studies have demonstrated that penalized regression methods can provide relatively stable variable selection and coefficient estimation, even at EPV levels between 5 and 10, provided that the regularization parameter is tuned via cross-validation. Furthermore, model performance was evaluated using an independent held-out test set and further internally validated using 1,000 bootstrap resamples to obtain optimism-corrected estimates of discrimination and calibration. This study was approved by the Medical Ethics Committee of Pudong New Area People's Hospital and adhered to the principles of the Declaration of Helsinki [9].

Inclusion and exclusion criteria

Inclusion criteria: (1) age ≥ 45 years; (2) diagnosis of KOA according to the 1986 American College of Rheumatology (ACR) revised clinical criteria [10]; (3) KL grade 1-3; (4) completion of the WOMAC scale, gait biomechanical testing and wearable device monitoring (valid wear days ≥ 4 days) at both baseline and 12-month follow-up; and (5) availability of complete clinical and follow-up data. Exclusion criteria: (1) KL grade 4 (end-stage KOA); (2) receipt of knee arthroplasty or other lower-limb orthopedic surgery during the 12-month follow-up; (3) comorbid rheumatoid arthritis, gouty arthritis, or other inflammatory joint diseases; (4) severe neurological disease (e.g., post-stroke sequelae, Parkinson's disease) or inability to ambulate independently; (5) severe cardiopulmonary insufficiency, malignancy, or other major systemic diseases with an expected survival of < 12 months; and (6) lower-limb fracture or acute knee trauma during the 12-month follow-up.

Clinical data collection

Baseline data were collected using an electronic medical record system and standardized questionnaires. (1) Demographic and general clinical characteristics: age, sex, body mass index (BMI). (2) KOA disease characteristics: disease duration (years), affected knee laterality (left knee/right knee/bilateral), KL grade (independently assessed by two senior radiologists, with discrepancies resolved by consensus) [11], prior knee injury history and prior knee surgery history. (3) Comorbidities and medication history: hypertension, diabetes, coronary artery disease, smoking history, alcohol consumption, regular exercise habit (defined as engagement in moderate-intensity exercise ≥ 3 times per week, ≥ 30 min per session) and baseline analgesic use. (4) Self-reported outcome measures: WOMAC subscale scores (pain, stiffness and function), Visual Analogue Scale (VAS) pain score and Center for Epidemiologic Studies Depression Scale (CES-D) score [12]. (5) Gait biomechanical parameters: 20-m walking speed (m/s), 5-times chair stand time (s), single-leg stance time (s), 400-m walking time (s) and gait speed coefficient of variation (CV, %). (6) Wearable device activity parameters: daily step count (steps/day), daily step count variability (CV, %), proportion of moderate-to-vigorous physical activity (MVPA, %), total sedentary time (h/day), sedentary break frequency (breaks/hour), daytime activity fragmentation index, sleep duration (h/night) and number of valid wearable device wear days.

Measurement methods

Gait function tests (hereafter referred to as generalized gait biomechanical performance indicators, encompassing spatiotemporal gait parameters and functional mobility tests, rather than three-dimensional kinematic/kinetic

measurements) were performed using the Zeno electronic walkway system (ProtoKinetics LLC, Havertown, PA, USA). The walkway had a length of 6.1 m, a width of 0.9 m and a sampling frequency of 120 Hz. Participants were instructed to walk 20 m at a self-selected comfortable speed across the walkway. Each assessment was repeated three times and the mean value was used for analysis.

The 5-times chair stand test was conducted using a standard armless chair with a seat height of 43 cm. Participant were instructed to rise from a seated position and sit down five consecutive times as quickly as possible and the total completion time was recorded.

The single-leg stance test required participants to stand on their dominant leg with eyes open and hands on hips. The maximum duration of independent standing was recorded (upper limit: 60 s). The test was repeated three times and the best value was taken.

The 400-m walk test was conducted in a straight, level 20-m corridor. Participants were instructed to complete 20 round trips (equivalent to 400 m) at their fastest tolerable walking speed and the total completion time was recorded.

The ActiGraph wGT3X-BT triaxial accelerometer (ActiGraph LLC, Pensacola, FL, USA) was used to measure physical activity. The device was worn on the non-dominant hip for 7 consecutive days (daytime and nighttime) at a sampling frequency of 30 Hz. Raw accelerometer data were processed using ActiLife software (version 6.13.4) and activity intensity was classified using the Freedson algorithm [13]. Specifically, sedentary behavior was defined as < 100 counts/min. Participants were included in the analysis if they had at least four valid wear days, defined as ≥ 10 hours of wear time per day. The daytime activity fragmentation index was defined as the probability of transition between sedentary bouts (≥ 30 min of consecutive < 100 counts/min) and active bouts (≥ 1 min of consecutive ≥ 100 counts/min) during waking hours. The transition probability was calculated as: transition probability = (number of sedentary-to-active transitions + number of active-to-sedentary transitions)/(total number of sedentary bouts + total number of active bouts). The index ranges from 0 to 1, with greater values implying a more fragmented activity pattern. Sleep periods were identified using the built-in Cole-Kripke algorithm in ActiLife and were excluded from the analysis. Data from valid wear days (≥ 10 h/day) were included, whereas invalid wear days were entirely excluded.

KL grade was assessed using standard anteroposterior and lateral knee radiographs. The Chinese version WOMAC scale was adopted (Likert 0-4 scoring system) [14], with a Cronbach's α coefficient of 0.85-0.95 [15]. All functional tests and questionnaire assessments were conducted by uniformly trained research personnel according to standardized protocols.

Outcome measures

The primary outcome was worsening KOA pain within 12 months after enrollment, defined as a composite binary endpoint (yes/no) requiring the presence of both of the following criteria: (1) an increase of ≥ 2 points in the WOMAC pain subscale score (range 0-20, with higher scores indicating more severe pain) from baseline to 12-month follow-up. This threshold was selected based on the minimal clinically important difference (MCID) values for the WOMAC pain subscale in KOA patients, which have been reported to range from 0.5 to 2.5 points based on study population, type and method of estimation (anchor-based vs. distribution-based approaches). A threshold of ≥ 2 points was selected to reduce the likelihood of measurement error and to increase the probability of capturing clinically meaningful change. However, MCID values have not been specifically validated in this cohort and the selected threshold may affect event classification; (2) a subjective patient-reported pain worsening at standardized follow-up, collected through a structured interview conducted by trained research staff, in response to the question: "In the past month, has your joint pain gotten worse?" (yes/no). This composite endpoint was employed to incorporate both objective changes in validated symptom scales and patient-perceived deterioration, thereby increasing clinical relevance. Notably, this approach prioritizes specificity over sensitivity, aiming to capture clinically meaningful and perceptible pain worsening. Patients who fulfilled WOMAC-based criterion but did not report subjective worsening were

classified as non-worsening cases. Sensitivity analyses were performed using alternative endpoints, including WOMAC pain increase ≥ 2 points alone and WOMAC pain increase ≥ 30% from baseline.

Secondary outcome measures were prespecified in the study protocol, including absolute changes in WOMAC pain, WOMAC function and VAS pain scores; however, these outcomes were not reported in the present analysis, which focused on development and validation of a binary prediction model for pain worsening.

Statistical analysis

All statistical analyses were performed using R software (version 4.5.1) and SPSS software (version 27.0, IBM Corp., Armonk, NY, USA). All tests were two-sided and a $P < 0.05$ was considered statistically significant.

Continuous variables were first assessed for normality using the Shapiro-Wilk test. Variables conforming to a normal distribution were presented as mean ± standard deviation and were compared using Welch's t-test; non-normally distributed variables were presented as median and interquartile range and were compared using the Wilcoxon rank-sum test, with the standardized Z statistic reported where appropriate. Categorical variables were presented as frequencies and percentages and between-group comparisons were performed using the χ^2 test. Baseline balance between the training and test sets was assessed using standardized mean difference (SMD), with SMD < 0.2 indicating acceptable balance (< 0.1 excellent, 0.1-0.2 acceptable).

The total cohort was randomly split into training and test sets in a 7:3 ratio. In the training set, 33 candidate variables were initially screened using univariate analysis ($P < 0.05$), followed by feature selection using LASSO regression with 10-fold cross-validation. The optimal penalty parameter was determined using the lambda.1se criterion. Based on the included variables, four binary logistic regression prediction models were constructed using a stepwise modular incremental strategy: a clinical model, a clinical + gait model, a clinical + wearable model and a fusion model incorporating all variables from the three modules. Model discrimination was evaluated using the area under the receiver operating characteristic curve (AUC). Differences in AUC between models were compared using the DeLong test. Calibration was evaluated using the Brier Score, calibration intercept, calibration slope and calibration plots.

Predictive improvement was assessed using integrated discrimination improvement (IDI) and net reclassification improvement (NRI). Clinical utility was evaluated using decision curve analysis (DCA), with net benefit calculated across a predefined threshold probability range of 0.05-0.35. Multivariable logistic regression analysis was used to identify independent risk factors associations with worsening KOA pain, with results reported as odds ratios (OR) with 95% confidence intervals (CIs). Confidence intervals were calculated using the profile-likelihood method. Internal model validation was conducted using 1,000 Bootstrap resamples to obtain optimism-corrected estimates of discrimination and calibration.

Subgroup analyses stratified by median baseline WOMAC pain score, KL grade and median BMI were conducted as sensitivity analyses to assess the robustness of the fusion model across diverse clinically relevant subgroups. To assess variable contributions to the fusion model, Shapley additive explanations (SHAP) were applied [16]. SHAP values were computed using the fastshap R package (version 0.1.1), which estimates Shapley values via Monte Carlo sampling. The explain() function was used with 200 simulations per observation (nsim = 200) and adjustment enabled (adjust = TRUE) to ensure additivity with model predictions. SHAP analysis was performed using all 689 observations in the training set.

In the SHAP summary plot, continuous variables were rescaled to the [0, 1] range using min-max normalization for visualization purposes only; however, all SHAP calculations were performed on the original untransformed feature scale. Since the fusion model involved continuous, binary and heterogeneous-scale variables, SHAP was used to provide a unified framework for interpreting relative feature contributions to individual predicted probabilities. SHAP analysis was conducted as a supplementary interpretive tool to regression coefficients rather than as a replacement for inferential sta-

tistical analysis. Throughout the manuscript, the terms "predictor" and "prediction" refer solely to variables included in prognostic models and the estimation of outcome probabilities, respectively and do not imply causal relationships. All associations derived from logistic regression models represent adjusted statistical associations conditional on included covariates and should not be interpreted as causal effects.

Results

Baseline characteristics of the training and test sets

A total of 984 patients with KOA were included in this study, of whom 276 (28.0%) experienced pain worsening during the 12-month follow-up. After 7:3 stratified random splitting, the training set comprised 689 patients (200 events, 29.0%) and the test set comprised 295 patients (76 events, 25.8%).

The two sets were balanced in terms of age, BMI, disease duration, WOMAC scores, VAS pain score, CES-D depression score, gait parameters, wearable device parameters and comorbidities, with most SMD < 0.21. However, KL grade showed a marginal imbalance (SMD = 0.205), slightly exceeding the prespecified threshold. All other variables demonstrated SMDs < 0.21.

Although statistically significant differences were observed in 20-m walk speed (P = 0.047), total sedentary time (P = 0.036), sex distribution (P = 0.018), KL grade (P = 0.013) and alcohol consumption (P = 0.047), the magnitude of these differences was small (maximum SMD = 0.205 for KL grade). Given the large sample size, minor between-group differences are expected to reach statistical significance; therefore, overall balance between the two datasets was considered acceptable indicating that the two sets were generally comparable (**Table 1**).

Comparison of baseline characteristics between pain worsening and non-worsening groups in the training set

In the training set, patients in the pain worsening group (n = 200) showed significantly worse baseline symptom severity compared with the non-worsening group (n = 489). In addition, higher baseline WOMAC pain, function and stiffness scores were observed in the worsening group, along with a higher VAS pain score (all P < 0.001). The CES-D depression score was also significantly elevated in the worsening group (P = 0.006).

In terms of gait biomechanical parameters, the worsening group had reduced 20-m walk speed, prolonged 5-times chair stand time, shortened single-leg stance duration and prolonged 400-m walk time (all P < 0.001).

Regarding wearable device-derived parameters, the worsening group exhibited lower daily step count, reduced proportion of MVPA, fewer sedentary breaks, higher activity fragmentation index, increased step count variability and longer total sedentary time (all P < 0.01).

In addition, the proportions of KL grade 3 (46.0% vs. 31.3%, P = 0.001) and patient-reported baseline analgesic use (44.5% vs. 32.7%, P = 0.005) were significantly higher in the worsening group compared with the non-worsening group. No significant differences were observed between the two groups in age (P = 0.057), BMI (P = 0.237), or sex (P = 0.075) (**Table 2**).

Correlation and collinearity analysis of candidate variables

Correlation and collinearity analyses were performed on 19 continuous and binary candidate variables (excluding KL grade and affected knee laterality) that showed baseline between-group differences of P < 0.05. Spearman correlation analysis showed that all pairwise correlation coefficients were below 0.7, indicating no evidence of strong collinearity. Among WOMAC subscales, the Spearman correlation coefficients for pain-stiffness, pain-function and stiffness-function pairs were 0.359, 0.560 and 0.403, respectively, consistent with the expected conceptual overlap among these domains. Variance inflation factor (VIF) analysis further confirmed the absence of substantial multicollinearity, with all variables showing VIF values < 5 (range: 1.05-2.74) (**Figure 1**).

Variable selection and final model construction

LASSO regression was subsequently performed for variable selection using the optimal tuning parameter determined by 10-fold cross-validation. Variables with P < 0.05 in at least one comparison in the univariate analysis were ini-

Table 1. Baseline characteristics of the training and test sets

Variable	Training (n = 689)	Test (n = 295)	Statistic	P	SMD
Age, years	63.00 [59.00, 68.00]	63.00 [58.00, 69.00]	Z = -0.166	0.869	0.018
BMI, kg/m^2	27.98 [25.22, 30.39]	28.11 [25.75, 30.90]	t = 1.389	0.250	0.095
Disease duration, years	7.00 [4.00, 9.00]	6.00 [4.00, 9.00]	Z = -1.206	0.228	0.074
Baseline WOMAC Pain score	5.00 [3.00, 7.00]	6.00 [4.00, 7.00]	Z = 1.391	0.164	0.070
Baseline WOMAC Function score	20.00 [13.00, 27.00]	20.00 [14.00, 27.00]	Z = 0.302	0.763	0.010
Baseline WOMAC Stiffness score	2.00 [2.00, 3.00]	3.00 [2.00, 4.00]	Z = 0.905	0.365	0.064
VAS Pain score	4.00 [3.00, 5.00]	4.00 [3.00, 5.00]	Z = 0.620	0.535	0.026
CES-D score	10.00 [7.00, 14.00]	11.00 [7.00, 15.00]	Z = 1.093	0.274	0.056
20-m walk speed, m/s	1.13 [1.00, 1.25]	1.15 [1.02, 1.29]	t = 2.103	0.047	0.145
5-times chair stand, s	13.00 [11.00, 15.00]	13.00 [11.00, 15.00]	Z = -0.360	0.719	0.036
Single-leg stance, s	20.00 [13.00, 27.00]	21.00 [15.00, 27.00]	Z = 0.837	0.403	0.051
400-m walk time, s	323.00 [283.00, 370.00]	322.00 [275.00, 368.00]	Z = -1.078	0.281	0.097
Gait speed CV, %	9.11 [6.75, 11.70]	9.42 [7.12, 11.88]	Z = 0.869	0.385	0.041
Daily steps, steps/day	6055.00 [4467.00, 7780.00]	6067.00 [4622.00, 7623.00]	Z = 0.064	0.949	0.005
Step count CV, %	32.35 [25.19, 39.79]	32.85 [24.95, 40.20]	Z = -0.091	0.928	0.014
MVPA proportion, %	9.77 [6.93, 13.44]	10.02 [6.59, 12.65]	Z = -0.975	0.329	0.069
Sedentary time, h/day	8.00 [7.00, 10.00]	9.00 [7.00, 10.00]	Z = 2.096	0.036	0.122
Sedentary breaks, /h	3.00 [3.00, 4.00]	3.00 [3.00, 4.00]	Z = -0.287	0.774	0.021
Activity fragmentation index	0.46 [0.37, 0.54]	0.46 [0.38, 0.54]	Z = 0.736	0.462	0.045
Sleep duration, h/night	7.00 [6.00, 8.00]	7.00 [6.00, 7.50]	Z = 0.162	0.871	0.017
Female sex, n (%)	435 (63.1)	210 (71.2)	χ^2 = 5.58	0.018	0.172
KL grade, n (%)			χ^2 = 8.75	0.013	0.205
Grade 1	162 (23.5)	55 (18.6)			
Grade 2	282 (40.9)	106 (35.9)			
Grade 3	245 (35.6)	134 (45.4)			
Hypertension, n (%)	270 (39.2)	112 (38.0)	χ^2 = 0.08	0.773	0.025
Diabetes, n (%)	108 (15.7)	38 (12.9)	χ^2 = 1.06	0.302	0.080
Coronary artery disease, n (%)	47 (6.8)	30 (10.2)	χ^2 = 2.76	0.096	0.120
Smoking history, n (%)	104 (15.1)	33 (11.2)	χ^2 = 2.32	0.128	0.116
Alcohol use, n (%)	124 (18.0)	70 (23.7)	χ^2 = 3.93	0.047	0.141
Analgesic use, n (%)	249 (36.1)	119 (40.3)	χ^2 = 1.38	0.240	0.086
Knee injury history, n (%)	204 (29.6)	74 (25.1)	χ^2 = 1.87	0.172	0.102
Prior knee surgery, n (%)	52 (7.5)	22 (7.5)	χ^2 = 0.00	1.000	0.003
Regular exercise, n (%)	282 (40.9)	122 (41.4)	χ^2 = 0.00	0.957	0.009
Valid wear days ≥ 6, n (%)	510 (74.0)	203 (68.8)	χ^2 = 2.55	0.110	0.115

Notes: SMD, standardized mean difference; BMI, body mass index; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; VAS, Visual Analogue Scale; CES-D, Center for Epidemiologic Studies Depression Scale; KL, Kellgren-Lawrence; MVPA, moderate-to-vigorous physical activity.

tially considered. A total of 33 candidate variables (35 features after dummy-variable encoding of KL grade and affected knee laterality) were included in the LASSO procedure. Most of the candidate features were measured as categorical and were transformed using one-hot (dummy) encoding.

The analysis produced collusion data in contrast to standardized data with no collinearity problem. The optimal penalty parameter (lambda.1se = 0.012253) resulted in the selection of 17 variables with non-zero coefficients (**Figure 2**).

The variables selected for the clinical module included disease duration, baseline WOMAC pain score, baseline WOMAC function score, baseline WOMAC stiffness score, analgesic use, regular exercise habit and KL grade. The gait biomechanics module included 20-m walk speed, 5-times chair stand time, single-leg stance time and 400-m walk time. The wearable device module included daily step count

Table 2. Comparison of baseline characteristics between pain worsening and non-worsening groups in the training set

Variable	Non-worsening group (n = 489)	Worsening group (n = 200)	Statistic	P	SMD
Age, years	63.00 [58.00, 68.00]	64.50 [60.00, 69.00]	Z = -1.902	0.057	0.163
BMI, kg/m^2	27.89 [25.01, 30.38]	28.24 [25.57, 30.43]	t = -1.628	0.237	0.138
Disease duration, years	6.00 [4.00, 9.00]	7.00 [5.00, 9.00]	Z = -2.054	0.040	0.165
Baseline WOMAC Pain score	5.00 [3.00, 7.00]	6.00 [4.00, 8.00]	Z = -3.301	0.001	0.303
Baseline WOMAC Function score	19.00 [12.00, 25.00]	23.00 [17.00, 30.00]	Z = -5.413	< 0.001	0.471
Baseline WOMAC Stiffness score	2.00 [1.00, 3.00]	3.00 [2.00, 4.00]	Z = -4.538	< 0.001	0.412
VAS Pain score	4.00 [2.00, 5.00]	4.00 [3.00, 6.00]	Z = -4.100	< 0.001	0.347
CES-D score	10.00 [6.00, 14.00]	11.50 [8.00, 15.00]	Z = -2.759	0.006	0.231
20-m walk speed, m/s	1.16 [1.04, 1.27]	1.02 [0.88, 1.19]	t = 6.824	< 0.001	0.590
5-times chair stand, s	13.00 [11.00, 15.00]	14.00 [12.00, 16.00]	Z = -4.214	< 0.001	0.356
Single-leg stance, s	21.00 [14.00, 27.00]	17.50 [11.75, 24.00]	Z = 4.103	< 0.001	0.368
400-m walk time, s	314.00 [276.00, 358.00]	348.00 [304.50, 404.25]	Z = -5.479	< 0.001	0.484
Gait speed CV, %	8.84 [6.67, 11.34]	9.83 [7.44, 12.52]	Z = -2.887	0.004	0.250
Daily steps, steps/day	6584.00 [4904.00, 8356.00]	5121.00 [3698.25, 6547.75]	t = 8.322	< 0.001	0.670
Step count CV, %	31.61 [24.99, 38.88]	34.64 [26.26, 42.66]	Z = -2.678	0.007	0.233
MVPA proportion, %	10.68 [7.55, 14.30]	8.08 [6.05, 10.85]	Z = 6.602	< 0.001	0.595
Sedentary time, h/day	8.00 [7.00, 9.00]	9.00 [7.00, 10.00]	Z = -3.148	0.002	0.269
Sedentary breaks, /h	4.00 [3.00, 4.00]	3.00 [2.00, 4.00]	Z = 6.280	< 0.001	0.560
Activity fragmentation index	0.43 [0.36, 0.51]	0.52 [0.44, 0.62]	t = -8.589	< 0.001	0.747
Sleep duration, h/night	7.00 [6.00, 8.00]	7.00 [6.00, 7.00]	Z = 1.798	0.072	0.165
Female sex, n (%)	298 (60.9)	137 (68.5)	χ^2 = 3.17	0.075	0.159
KL grade, n (%)			χ^2 = 14.79	0.001	0.325
Grade 1	128 (26.2)	34 (17.0)			
Grade 2	208 (42.5)	74 (37.0)			
Grade 3	153 (31.3)	92 (46.0)			
Hypertension, n (%)	191 (39.1)	79 (39.5)	χ^2 = 0.00	0.983	0.009
Diabetes, n (%)	78 (16.0)	30 (15.0)	χ^2 = 0.04	0.844	0.026
Coronary artery disease, n (%)	31 (6.3)	16 (8.0)	χ^2 = 0.38	0.536	0.064
Smoking history, n (%)	76 (15.5)	28 (14.0)	χ^2 = 0.16	0.692	0.043
Alcohol use, n (%)	92 (18.8)	32 (16.0)	χ^2 = 0.58	0.445	0.074
Analgesic use, n (%)	160 (32.7)	89 (44.5)	χ^2 = 8.03	0.005	0.244
Knee injury history, n (%)	136 (27.8)	68 (34.0)	χ^2 = 2.32	0.128	0.134
Prior knee surgery, n (%)	32 (6.5)	20 (10.0)	χ^2 = 1.96	0.162	0.126
Regular exercise, n (%)	212 (43.4)	70 (35.0)	χ^2 = 3.76	0.053	0.172
Valid wear days ≥ 6, n (%)	366 (74.8)	144 (72.0)	χ^2 = 0.46	0.498	0.064

Notes: SMD, standardized mean difference; BMI, body mass index; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; VAS, Visual Analogue Scale; CES-D, Center for Epidemiologic Studies Depression Scale; KL, Kellgren-Lawrence; MVPA, moderate-to-vigorous physical activity.

coefficient of variation, MVPA proportion, total sedentary time, sedentary break frequency and daytime activity fragmentation index.

Based on these variables, four logistic regression models were sequentially constructed using a modular incremental strategy: a clinical model, a clinical + gait model, a clinical + wearable model and a fusion model incorporating all 17 variables.

Univariate and multivariable logistic regression analysis

Univariate logistic regression analysis of the 17 LASSO-selected variables (**Figure 3**) showed that most variables were significantly associated with 12-month KOA pain worsening. Among clinical variables, KL grade 3, baseline WOMAC stiffness, pain and function scores, analgesic use and disease duration were associated with

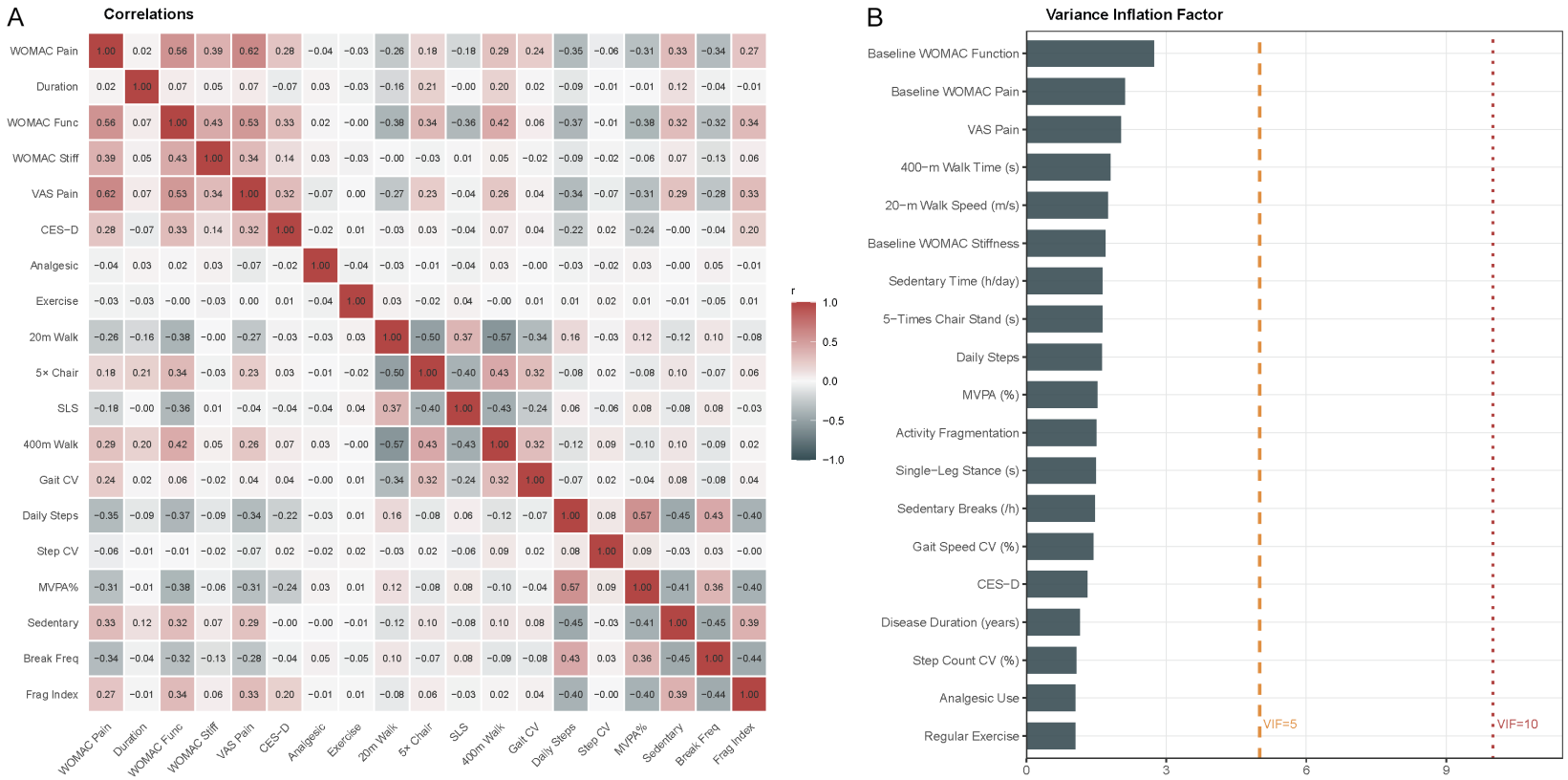

Figure 1. Correlation matrix and variance inflation factor analysis of candidate predictors. A. Spearman correlation heatmap showing the correlations among 19 continuous and binary candidate predictors with significant baseline differences between groups ($P < 0.05$). Red and blue indicate positive and negative correlations and the values within each cell represent Spearman correlation coefficients. B. Variance inflation factor values of the same 19 candidate predictors. The vertical reference lines indicate VIF = 5 and VIF = 10, representing commonly used thresholds for potential and severe multicollinearity, respectively. All predictors had VIF values below 5, suggesting no substantial evidence of multicollinearity among the candidate variables. Notes: VIF, variance inflation factor; LASSO, least absolute shrinkage and selection operator; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; VAS, visual analog scale; CES-D, Center for Epidemiologic Studies Depression Scale; BMI, body mass index; CAD, coronary artery disease; CV, coefficient of variation; MVPA, moderate-to-vigorous physical activity; SLS, single-leg stance; 20-m, 20-meter; 400-m, 400-meter; min, minutes; h/day, hours per day; Frag Index, activity fragmentation index. Correlation coefficients were calculated using Spearman rank correlation. VIF analysis was conducted to assess multicollinearity among baseline-significant continuous and binary predictors prior to model development.

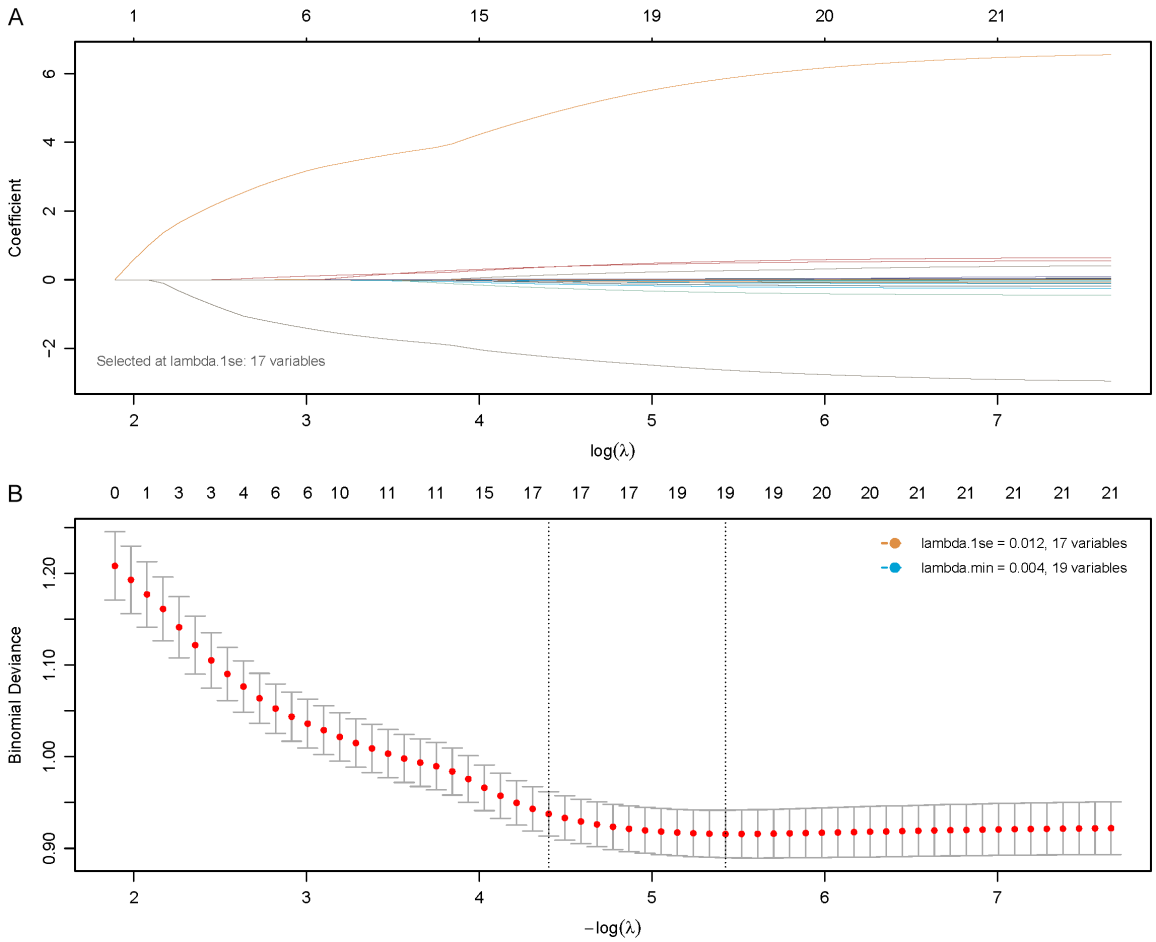

Figure 2. LASSO regression-based variable selection. A. Coefficient profiles of baseline-significant candidate variables plotted against log(λ). The vertical dashed line indicates the optimal penalty parameter selected using the 1-standard-error criterion (λ_{1se} = 0.012253), resulting in the retention of 17 variables with non-zero coefficients. B. Ten-fold cross-validation curve showing the relationship between log(λ) and binomial deviance for variable selection. The optimal λ value was determined using the 1-standard-error rule. Notes: LASSO, least absolute shrinkage and selection operator; λ, regularization parameter; λ.min, value of λ giving the minimum cross-validated binomial deviance; λ.1se, largest value of λ within 1 standard error of the minimum cross-validated deviance. Numbers displayed along the upper axis indicate the number of nonzero coefficients retained at each λ value. Binomial deviance was estimated using 10-fold cross-validation in the training set. A larger λ imposes stronger coefficient shrinkage and yields a more parsimonious model.

increased odds of pain worsening, whereas regular exercise habit was associated with reduced odds. Among gait-related variables, slower 20-m walking speed, longer 5-times chair stand time and prolonged 400-m walk time were associated with higher odds of pain worsening, while longer single-leg stance time was associated with lower odds. Among wearable device-derived variables, lower sedentary break frequency and MVPA proportion were associated with reduced odds, while higher total sedentary time, greater step count variability and higher daytime activity fragmentation index were associated with increased odds. Notably, the activity fragmentation index demonstrated a strong association with the outcome, with a univariate OR of 1.852 per 0.1-unit increase (95% CI: 1.602-2.156, P < 0.001).

In the multivariable logistic regression model including all 17 variables, 12 variables remained independently associated with KOA pain worsening. Baseline WOMAC stiffness score, analgesic use, 400-m walk time, step count CV and daytime activity fragmentation index were independently associated with increased odds of pain worsening. In contrast, regular exercise habit, 20-m walking speed,

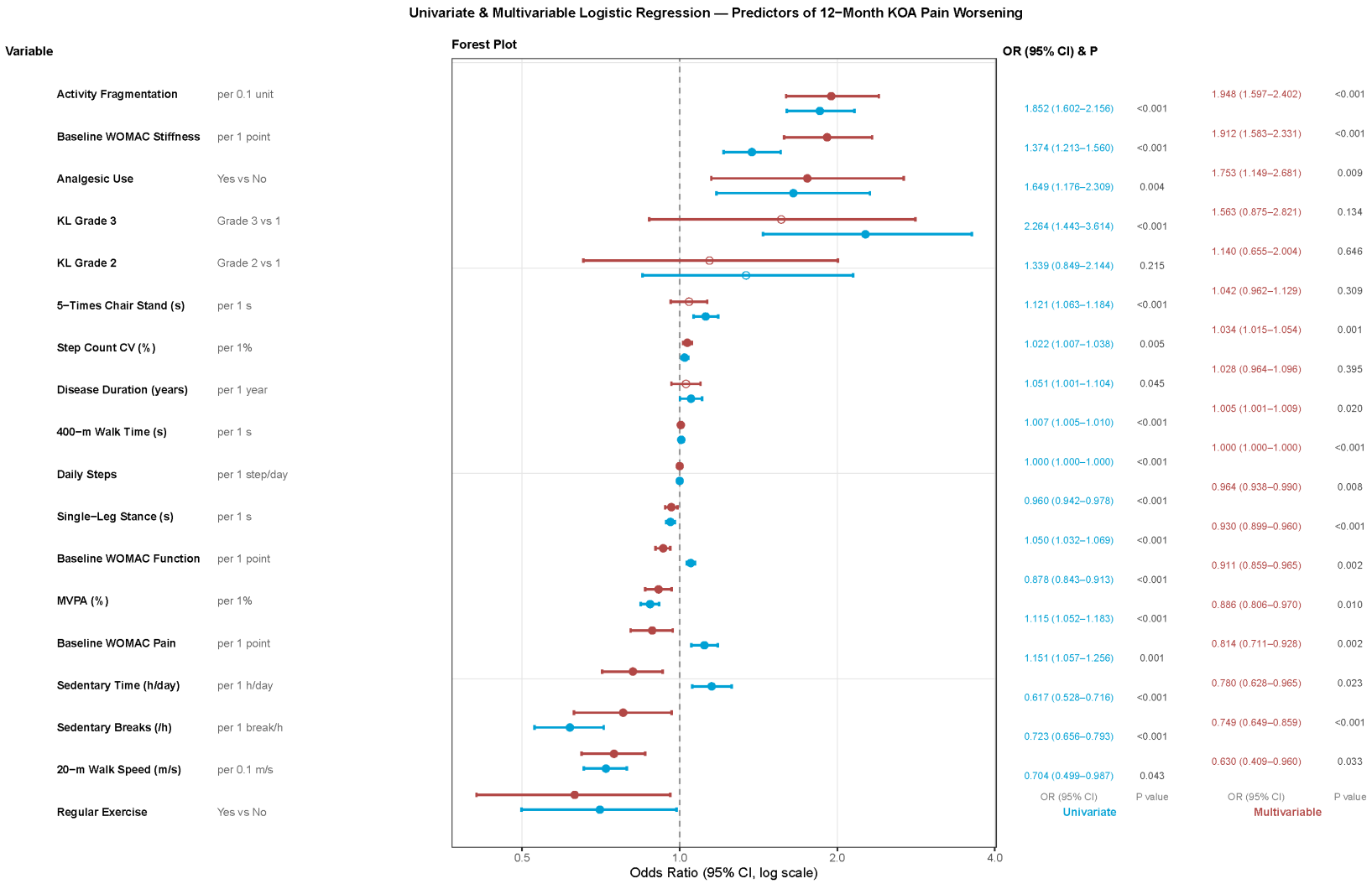

Figure 3. Univariate and multivariate logistic regression plots for predictors of 12-month KOA pain worsening. Notes: OR, odds ratio; CI, confidence interval; KOA, knee osteoarthritis; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; MVPA, moderate-to-vigorous physical activity; KL, Kellgren-Lawrence; CV, coefficient of variation; m, meter; s, seconds; h/day, hours per day.

single-leg stance time, MVPA proportion and sedentary break frequency were independently associated with reduced odds. The daytime activity fragmentation index remained significant in the multivariable model, with an OR of 1.948 per 0.1-unit increase (95% CI: 1.597-2.402, P < 0.001). Several variables, including KL grade 3, disease duration and 5-times chair stand time, were no longer statistically significant after adjustment, suggesting potential confounding or shared variance with other functional measures.

Changes in effect estimates were also observed for WOMAC subscale scores after multivariable adjustment, reflecting intercorrelations among WOMAC domains (e.g., pain-stiffness ρ = 0.359; function-stiffness ρ = 0.403). Similarly, the association between total sedentary time and the outcome was substantially attenuated after adjustment, suggesting that its crude effect was largely explained by other activity-related parameters. Therefore, the individual ORs of these variables should not be interpreted causally; their value lies in their joint contribution to the discriminative performance of the predictive model.

Comparison of discriminative performance among different prediction models

In the test set, the four models exhibited progressively improved discriminative performance (**Figure 4**; **Table 3**). The AUC of the clinical model was 0.673 (95% CI: 0.605-0.740), the AUC of clinical + gait model was 0.753 (95% CI: 0.695-0.811), the AUC of clinical + wearable model was 0.756 (95% CI: 0.695-0.816) and the AUC of fusion model was 0.815 (95% CI: 0.761-0.869). DeLong tests showed that, compared with the clinical model, all extended models demonstrated statistically significant improvements in discrimination (P = 0.002; P = 0.019; P < 0.001). The fusion model showed a small decrease in performance from the training set (AUC = 0.842) to the test set (AUC = 0.815), suggesting good generalization without obvious overfitting.

Calibration performance of different prediction models

Calibration performance (**Figure 5**) showed that the fusion model achieved the lowest Brier score (0.144), outperforming the clinical model (0.180), the clinical + gait model (0.164) and the clinical + wearable model (0.165). The fusion model had a calibration intercept of -0.261 and a calibration slope of 0.856, indicating acceptable agreement between predicted probabilities and observed outcomes.

Calibration plots showed that all models demonstrated reasonable agreement between predicted and observed probabilities in both the training and test sets, with curves generally following the 45-degree reference line.

Incremental predictive value of gait function and wearable parameters

The incremental value of each model was assessed using the clinical model as the reference (**Table 4**). Compared with the clinical model, the clinical + gait model yielded an IDI of 0.057 (95% CI: 0.029-0.085, P < 0.001) and a continuous NRI of 0.606 (95% CI: 0.351-0.846, P < 0.001); the clinical + wearable model demonstrated greater improvement, with an IDI of 0.123 (95% CI: 0.071-0.174, P < 0.001) and a continuous NRI of 0.614 (95% CI: 0.366-0.835, P < 0.001); the fusion model yielded the most pronounced incremental predictive value, with an IDI of 0.217 (95% CI: 0.151-0.279, P < 0.001) and a continuous NRI of 0.697 (95% CI: 0.446-0.932, P < 0.001).

Decision curve analysis

DCA (**Figure 6**) demonstrated that, in both the training and test sets, the fusion model consistently yielded the highest net benefit across a risk threshold range of 0.05-0.35, suggesting that clinical decisions based on the fusion model can provide greater net benefit to patients across a wide range of risk thresholds.

Variable contribution and interpretability analysis of the final model

SHAP analysis was further employed to interpret the relative contribution of each predictor in the fusion model to individual predictions of 12-month KOA pain worsening (**Figure 7**). In the global feature importance ranking (**Figure 7A**), baseline WOMAC stiffness score was identified as the most influential predictor, with a mean absolute SHAP value of 0.1006, accounting for 12.8% of the total explained contribution, followed by daytime activity fragmentation index

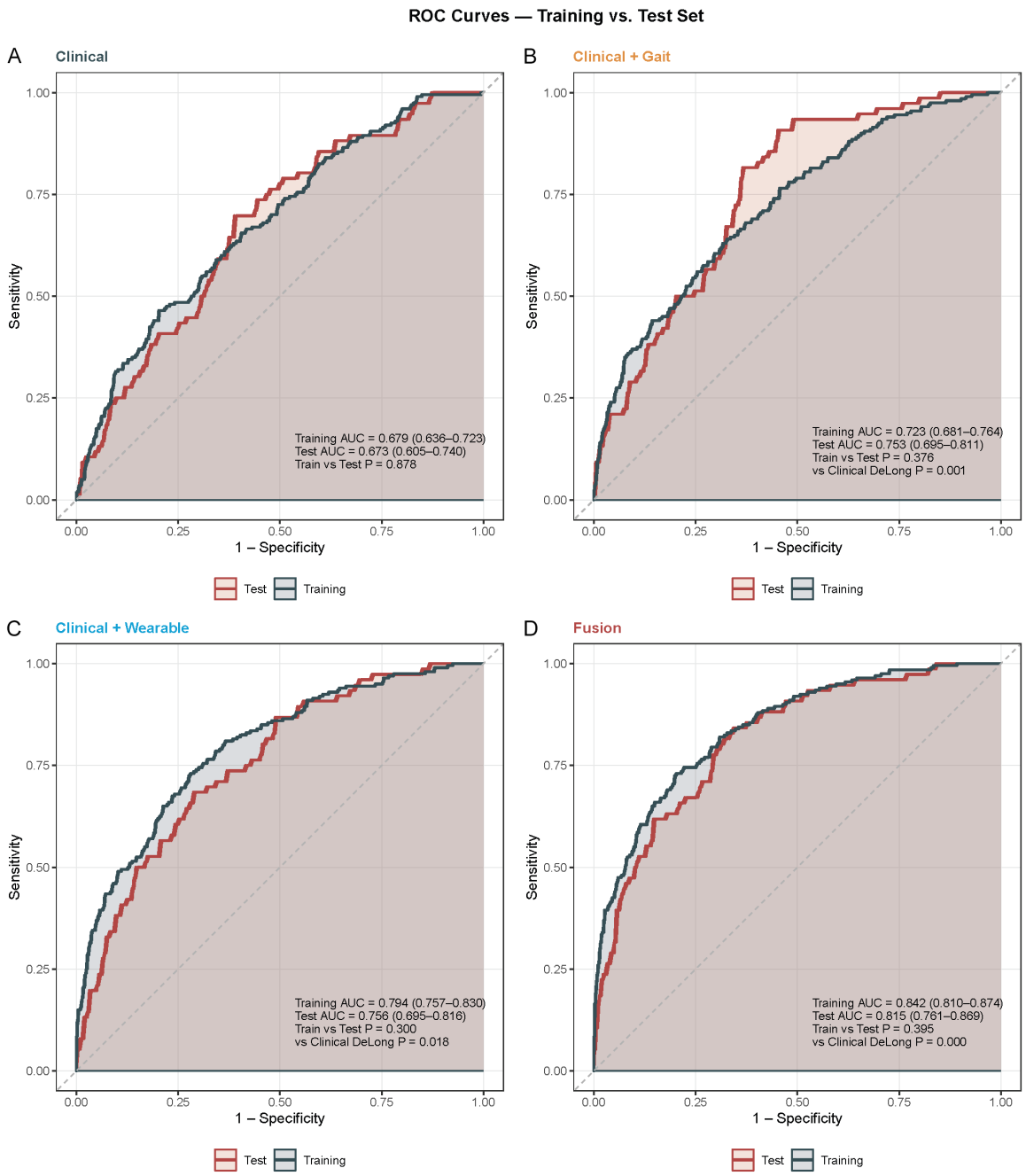

Figure 4. ROC curves for the four prediction models evaluated on training and test sets. A. Clinical model. B. Clinical + Gait model. C. Clinical + Wearable model. D. Fusion model. Notes: ROC, receiver operating characteristic; AUC, area under the ROC curve; CI, confidence interval; KOA, knee osteoarthritis.

(mean |SHAP| = 0.0977, 12.4%) and baseline WOMAC function score (mean |SHAP| = 0.0753, 9.6%). Other important contributors included 20-m walk speed, daily step count, MVPA proportion, total sedentary time, step count CV, baseline WOMAC pain score and single-leg stance time, suggesting that symptom burden, gait function and real-world activity collectively contribute to model predictions.

The SHAP summary plot (**Figure 7B**) illustrates the directionality feature effects on predicted risk. Higher baseline WOMAC stiffness score, higher daytime activity fragmentation index, higher daily step count variability, longer 400-m walk time, higher KL grade, longer 5-times chair stand time and longer disease duration were generally associated with increased risk of pain worsening. Conversely, higher 20-m walking

Table 3. Discrimination and calibration performance of the four prediction models

Model	Train AUC (95% CI)	Test AUC (95% CI)	Calib Int (Test)	Calib Slp (Test)	Calib Int (Train)	Calib Slp (Train)	Sensitivity	Specificity	Accuracy	PPV	NPV
Clinical model (n = 7)	0.679 (0.636-0.723)	0.673 (0.605-0.740)	-0.2627	0.9922	0.0000	1.0000	70.3%	46.1%	64.1%	35.0%	79.0%
Clinical + Gait model (n = 11)	0.723 (0.681-0.764)	0.753 (0.695-0.811)	0.0238	1.2563	0.0000	1.0000	64.4%	72.4%	66.4%	41.4%	87.0%
Clinical + Wearable model (n = 13)	0.794 (0.757-0.830)	0.756 (0.695-0.816)	-0.3863	0.8159	0.0000	1.0000	66.7%	69.7%	67.5%	42.1%	86.4%
Fusion model (n = 17)	0.842 (0.810-0.874)	0.815 (0.761-0.869)	-0.2614	0.8562	0.0000	1.0000	77.6%	65.8%	74.6%	50.5%	86.7%

Note: AUC, area under the receiver operating characteristic curve; CI, confidence interval; DeLong P, P value for DeLong test comparing each model against the Clinical model on the test set; Brier, Brier Score; Calib Int, calibration intercept; Calib Slp, calibration slope; Sens, sensitivity; Spec, specificity; Acc, accuracy; PPV, positive predictive value; NPV, negative predictive value. All models were built on the training set (n = 689) and evaluated on the independent test set (n = 295). Classification thresholds were determined by maximizing Youden index on the training set and applied without modification to the test set.

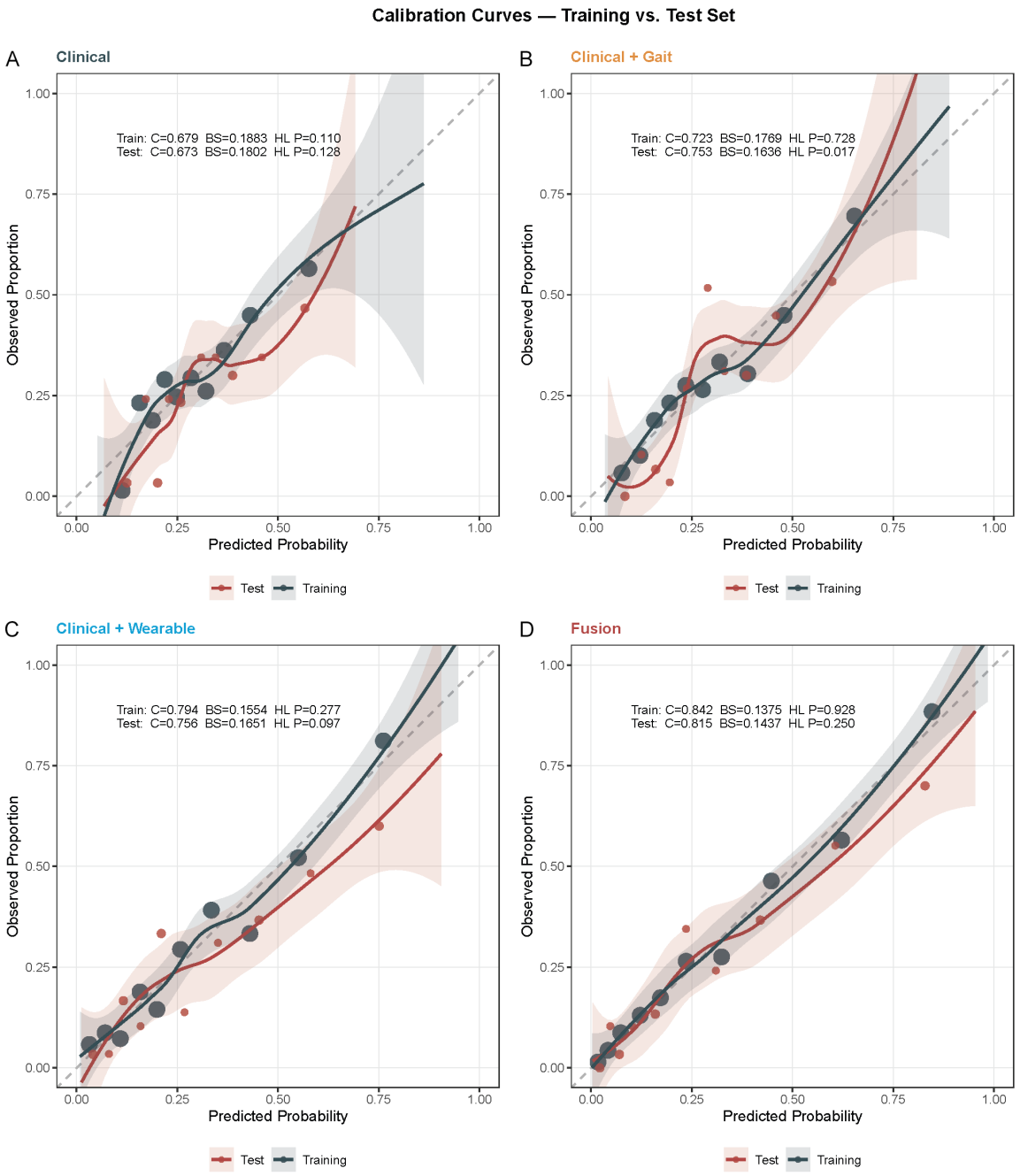

Figure 5. Calibration curves for the four prediction models. A. Clinical model. B. Clinical + Gait model. C. Clinical + Wearable model. D. Fusion model. Points represent decile-grouped observed proportions according to predicted risk, with larger points indicating larger numbers of patients within the corresponding risk group. Solid curves represent LOESS-smoothed calibration curves and shaded areas indicate the corresponding confidence bands. The diagonal dashed line represents perfect calibration, where predicted probabilities equal observed event proportions. Model-specific C-statistics, Brier Scores and Hosmer-Lemeshow test P values are annotated within each panel. Notes: KOA, knee osteoarthritis; C, C-statistic, equivalent to the area under the receiver operating characteristic curve; BS, Brier Score; HL, Hosmer-Lemeshow goodness-of-fit test; LOESS, locally estimated scatterplot smoothing.

speed, higher daily step count, higher MVPA proportion, longer single-leg stance time and higher sedentary break frequency were associated with lower predicted risk. Among binary variables, analgesic use was associated with higher mean SHAP values, whereas regular exercise habit was associated with lower mean SHAP values, suggesting that these two vari-

Table 4. Incremental predictive value of gait biomechanics and wearable-derived parameters in the test set

Comparison	IDI (95% CI)	IDI P	Continuous NRI (95% CI)	NRI P	Categorical NRI
+Gait vs. Clinical	0.057 (0.029-0.085)	< 0.001	0.606 (0.351-0.846)	< 0.001	0.142 (0.024-0.268)
+Wearable vs. Clinical	0.123 (0.071-0.174)	< 0.001	0.614 (0.366-0.835)	< 0.001	0.280 (0.123-0.438)
Fusion vs. Clinical	0.217 (0.151-0.279)	< 0.001	0.697 (0.446-0.932)	< 0.001	0.427 (0.261-0.599)

Note: IDI, integrated discrimination improvement; NRI, net reclassification improvement; CI, confidence interval. All comparisons are relative to the Clinical model.

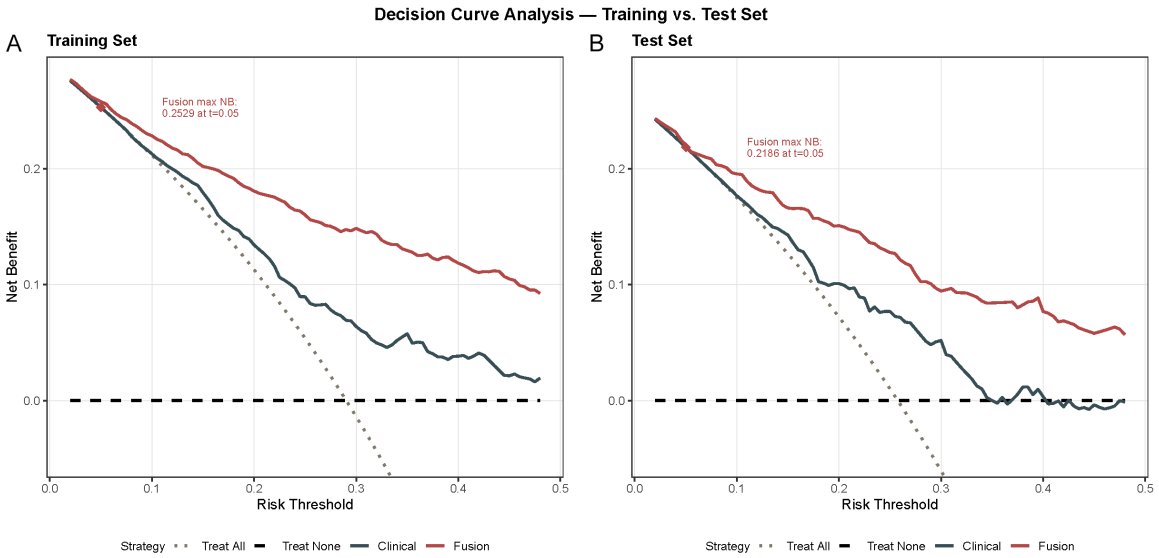

Figure 6. Decision curve analysis comparing clinical and fusion models on training and test sets. A. Decision curve analysis in the training set. B. Decision curve analysis in the independent test set. Notes: DCA, decision curve analysis; KOA, knee osteoarthritis; NB, net benefit. Net benefit was calculated by weighing true-positive classifications against false-positive classifications across different threshold probabilities. A higher net benefit indicates greater clinical usefulness at a given risk threshold. The models were developed using the training set and evaluated in both the training set and the independent test set.

ables contribute to risk prediction in opposite directions.

Dependence plots further illustrated the relationships between the top three predictors and their SHAP values (**Figure 7C-E**). Baseline WOMAC stiffness score showed a clear positive relationship with SHAP values (Spearman ρ = 0.971), with values ranging from 0 to 7 and corresponding SHAP values ranging -0.2806 to 0.4964. The mean SHAP value in the high stiffness group (highest tertile, ≥ 3 points) was 0.2013, higher than that in the low stiffness group (lowest tertile, ≤ 2 points). Similarly, the daytime activity fragmentation index showed a strong positive correlation with SHAP values (Spearman ρ = 0.986), with observed values ranging from 0.155 to 0.907 and SHAP values ranging from -0.2609 to 0.4722. The mean SHAP value in the high fragmentation group (≥ 0.510) was 0.2598 higher than that in the low fragmentation group (≤ 0.405). In contrast, baseline WOMAC function score exhibited a negative relationship with SHAP values (Spearman ρ = -0.986), with scores ranging from 0 to 55 points and SHAP values ranging from -0.2767 to 0.2392. The mean SHAP value in the highest function score group (≥ 24 points) was 0.1923 lower than that in the lowest function score group (≤ 16 points).

Nomogram or risk scoring tool construction

Based on the multivariable logistic regression analysis of the fusion model, independent predictors with P < 0.05 were incorporated into a nomogram for individualized prediction of 12-month KOA pain worsening risk (**Figure 8**). This nomogram integrates clinical symptoms, gait functional performance and wearable

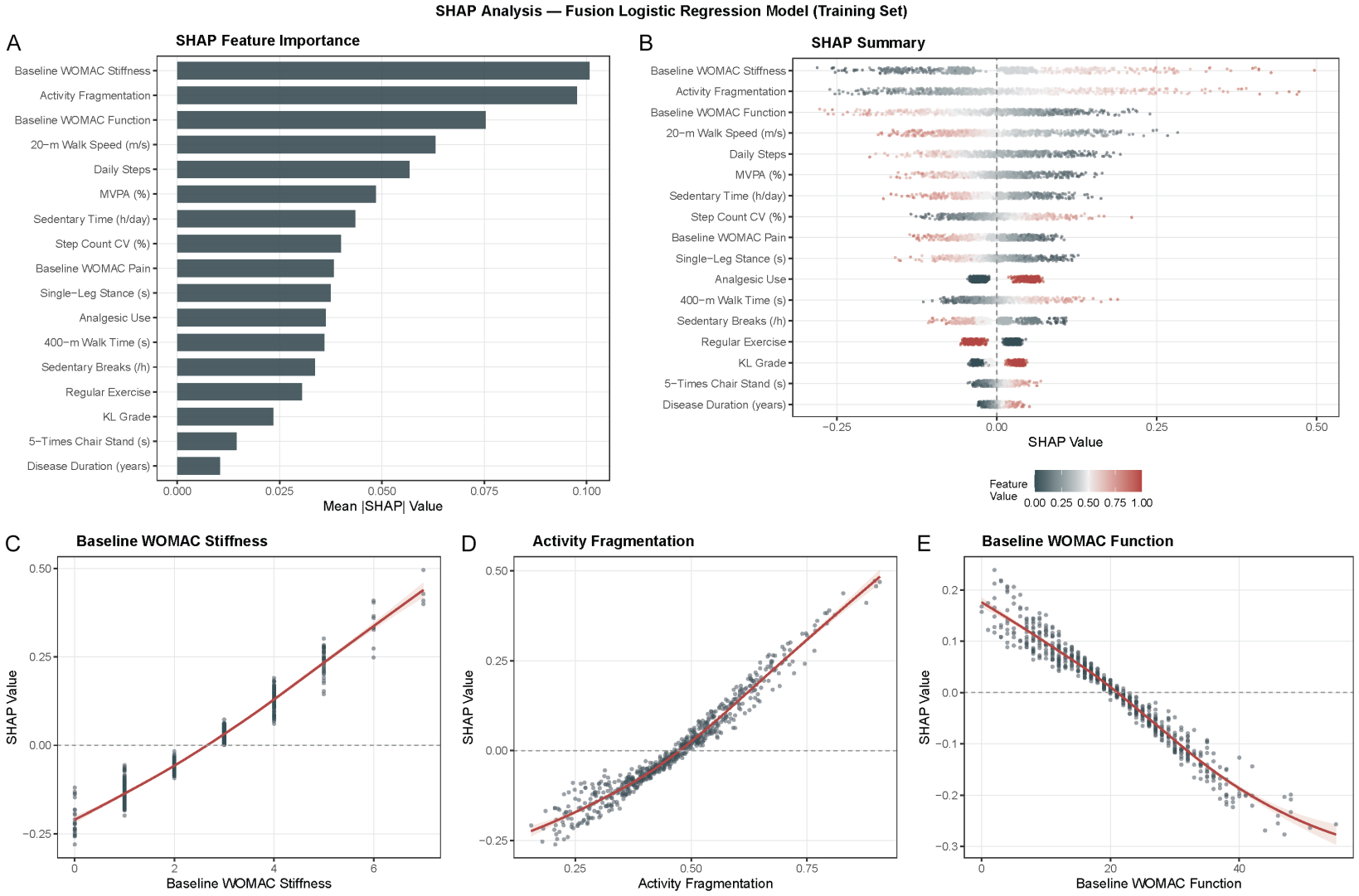

Figure 7. SHAP analysis of the fusion logistic regression model for predicting 12-month KOA pain worsening. A. SHAP feature importance plot showing the mean absolute SHAP value for each predictor, with higher values indicating greater overall contribution to model prediction. B. SHAP summary plot showing the distribution of SHAP values for each predictor across individuals; each point represents one patient and the color gradient indicates the relative feature value from low to high. Positive SHAP values indicate an increased predicted probability of 12-month worsening KOA pain, whereas negative SHAP values indicate a decreased predicted probability. C-E. SHAP dependence plots for the top-ranked predictors, showing the relationship between individual predictor values and their corresponding SHAP values. Note: SHAP, Shapley additive explanations; KOA, knee osteoarthritis; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; MVPA,

moderate-to-vigorous physical activity. Mean |SHAP| represents the average absolute SHAP value and reflects the overall importance of each predictor in the model. Feature values in the SHAP summary plot were scaled from 0 to 1 for visualization, where higher values indicate larger observed predictor values. SHAP values were derived from the fusion logistic regression model using the training set.

device-derived daily activity metrics, including baseline WOMAC pain, function and stiffness scores, analgesic use, regular exercise habit, 20-m walking speed, single-leg stance time, 400-m walk time, daily step count, daily step count CV, MVPA proportion, total sedentary time, sedentary break frequency and daytime activity fragmentation index.

Each predictor was assigned a weighted point value based on its regression coefficient in the multivariable model. In clinical practice, individual patient values are located on each variable axis to obtain corresponding scores, which are then summed to obtain a total score. Next, the total score is subsequently projected onto the risk axis to estimate the individual probability of 12-month KOA pain worsening. A higher total score indicates a higher predicted risk.

The nomogram reflects the combined contribution of symptom burden, gait function limitation and real-world physical activity patterns. Specifically, WOMAC subscale scores represent baseline symptom severity and functional limitations; gait performance measures (20-m walking speed, single-leg stance time and 400-m walk time) reflect lower-limb functional capacity; and wearable-derived metrics capture habitual activity behavior, including physical activity intensity, sedentary behavior and activity fragmentation. This nomogram integrates the above multidimensional information, which can help clinicians rapidly identify KOA patients at higher risk of potential pain worsening and provide a reference for early risk stratification, follow-up management and individualized intervention decision (**Figure 8**).

Sensitivity analysis

Sensitivity analyses were conducted to evaluate the robustness of the fusion model across prespecified clinical subgroups in the test set (**Figure 9**). The fusion model maintained robust discriminative performance across all subgroups. In analyses stratified by median baseline WOMAC pain score, the AUC values were comparable between the high-pain and low-pain subgroups. When stratified by KL grade (grade 1-2 vs. grade 3), the model demonstrated consistent performance across radiographic severity levels. Similarly, stratification by median BMI showed stable AUC values across BMI subgroups. These results indicate that the predictive performance of the fusion model is not substantially influenced by baseline pain severity, radiographic severity, or obesity status, supporting its generalizability across diverse clinical presentations of KOA.

Discussion

Drawing on a single-center retrospective cohort study, baseline clinical characteristics [2, 4] were integrated with gait biomechanical parameters and wearable device-derived daily activity data. Using LASSO regression for feature selection, 4 prediction models for 12-month KOA pain worsening were developed. The findings showed that the inclusion of objective gait and wearable device-derived parameters enhanced model discrimination compared with clinical variables alone, with both providing significant incremental predictive value. Multivariable regression and SHAP analyses jointly identified baseline WOMAC stiffness score, 20-m walk speed, daytime activity fragmentation index and sedentary break frequency as the most important contributors to model performance.

Among clinical variables, baseline WOMAC stiffness score showed the strongest and most consistent association with pain worsening in the multivariable model and ranked first in SHAP importance, consistent with prior studies reporting its prognostic relevance in KOA functional outcomes [3, 17]. The stiffness subscale may reflect overall symptom burden and functional limitation more sensitively than pain or function scores alone. It is noteworthy that the WOMAC pain score (univariate OR = 1.115 → multivariable OR = 0.756) and function score (1.050 → 0.834) exhibited changes in effect estimates after multivariable adjustment. Moderate correlations were observed among WOMAC subscales (Spearman ρ = 0.359-0.560), suggesting shared variance among these measures. Therefore, the interpretation of individual odds ratios should be restricted to statistical

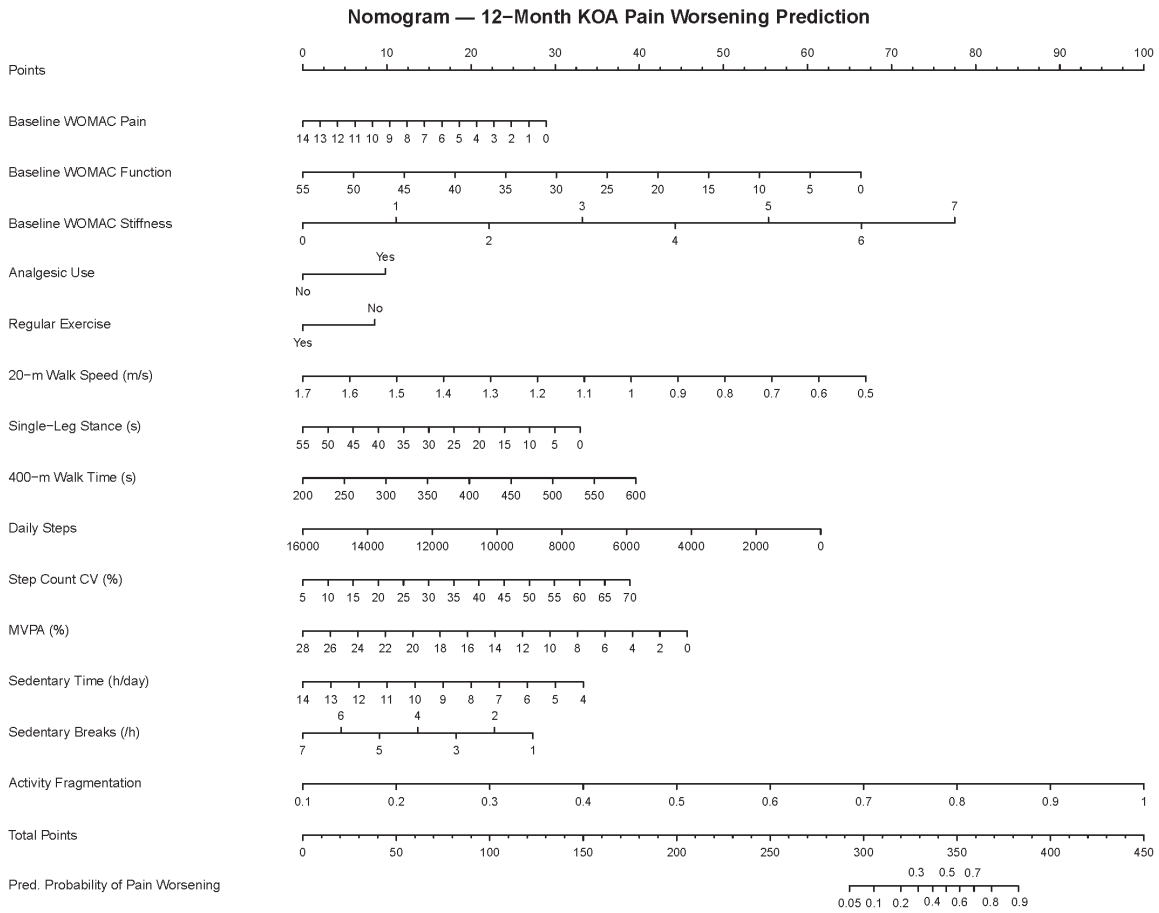

Figure 8. A nomogram developed based on independent predictors. Note: KOA, knee osteoarthritis; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; MVPA, moderate-to-vigorous physical activity; 20-m, 20-meter; 400-m, 400-meter; min, minutes; s, seconds; h/day, hours per day; Pred., predicted.

association within the context of the full model rather than causal inference. Likewise, KL grade 3 was significantly associated with pain worsening in univariate analysis (OR = 2.264) but was no longer significant after multivariable adjustment (P = 0.134), suggesting that radiographic structural damage severity may indirectly increase risk of pain worsening by causing decline in gait function and daily activities [18, 19]. Analgesic use (multivariable OR = 1.753) and regular exercise habit (multivariable OR = 0.630) remained independently associated with the outcome in the multivariable model.

The strongest independent association with reduced odds of worsening pain among gait biomechanical parameters was observed for 20 m walking speed. In the multivariable model, each 0.1 m/s increased in baseline walking speed was associated with approximately 25% lower odds of pain worsening. Clinically, walking speed is a global measure of lower-limb neuromuscular function, joint stability and functional capacity [20-22]. Reduced walking speed may reflect both pain-related functional limitation and adaptive gait strategies aimed at minimizing joint loading. Over time, such adaptations may, in theory, be associated with reduced physical activity and altered joint loading patterns, which have been hypothesized to contribute to functional decline in KOA. The associations of single-leg stance time and 400-m walking time with pain worsening further support the relevance of dynamic balance and walking endurance-related functional capacity in KOA. In contrast, the 5-times chair stand time was significantly associated with pain worsening in univariate analyses but did not remain significant in the multivariable model (P = 0.309). This may suggest that its predictive information overlaps with other functional measures, such as walking speed and WOMAC

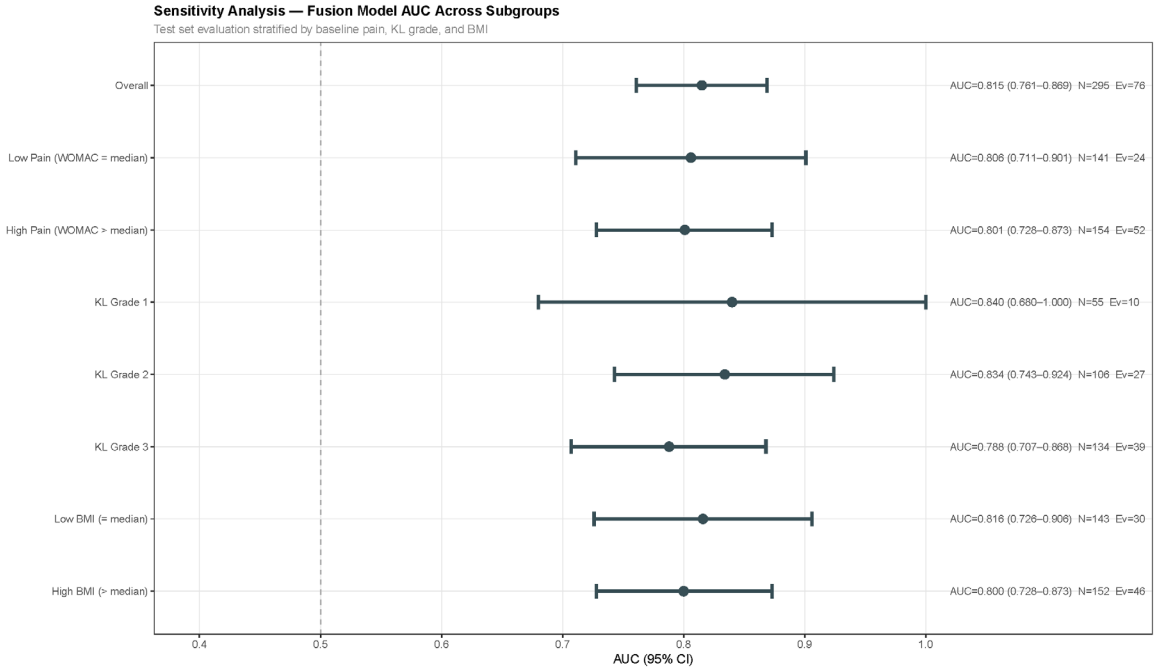

Figure 9. Sensitivity analysis - fusion model AUC across subgroups on the test set. Notes: AUC, area under the receiver operating characteristic curve; CI, confidence interval; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; KL, Kellgren-Lawrence; BMI, body mass index; N, number of patients; Ev, number of pain worsening events. Median cutoffs for WOMAC pain score and BMI were determined from the training set and applied to the independent test set. The fusion model was developed in the training set and evaluated in the independent test set. Wider confidence intervals in some subgroups reflect smaller sample sizes or fewer events.

function score, which more comprehensively capture lower-limb functional status.

Wearable device activity parameters constituted the most innovative predictive module in this study [7]. Among wearable-derived variables, the daytime activity fragmentation index showed the largest effect size and ranked second in SHAP importance, indicating that disrupted patterns of daily activity are strongly associated with pain worsening. Activity fragmentation reflects the extent to which daily physical activity is distributed across multiple short bouts rather than sustained continuous periods. Higher fragmentation may reflect reduced tolerance for prolonged activity and a tendency toward frequent rest periods. Although this pattern may reduce short-term joint loading, it may also be associated with declined periarticular muscle function and insufficient joint stability in the long term. However, these interpretations require validation in longitudinal mechanistic studies. Consistent patterns were also observed for sedentary break frequency and MVPA proportion, where more frequent interruptions of sedentary behavior and higher MVPA levels were independently associated with lower odds of pain worsening. Whether interventions targeting these behavioral patterns can reduce pain worsening risk requires confirmation in prospective interventional studies. Although daily step count was statistically significant in both univariate and multivariable analyses, its effect size was minimal (OR ≈ 1.000), suggesting limited incremental predictive value when considered alone. In contrast, measures capturing variability and temporal distribution of activity, including step count variability (CV) and activity fragmentation index, provided more informative behavioral signals for prediction.

From the perspective of model performance, the most critical finding of this study was that objectively measured gait biomechanical and wearable device-derived activity behavior parameters provided substantial incremental predictive value beyond conventional clinical variables. Compared with the clinical model alone, the fusion model improved discrimination by approximately 14%, representing a clinically meaningful improvement in the context of prognostic prediction models [23, 24]. The SHAP interpretation added further evidence to

this from a model-agnostic perspective. Gait and wearable-derived parameters accounted for six of the top 10 most important variables. The incremental predictive analysis showed that the wearable module provided greater improvement in discrimination than the gait module alone. This indicates that continuously collected daily activity patterns in free-living environments may capture additional aspects of functional status that are not fully reflected by clinic-based assessments performed under standardized conditions. These finding are consistent with the concept from ecological assessment approaches that real-world behavioral data may provide complementary information beyond measurements obtained in controlled laboratory settings [25-27].

Several limitations of this work should be acknowledged. First, this was a single-center retrospective cohort study and selection bias cannot be completely excluded. Although the model demonstrated robust performance in the independent test set and subgroup sensitivity analyses, external validation in multicenter prospective cohorts is required to evaluate its generalizability across different geographic regions and healthcare settings. Second, the current model did not include molecular-level biomarkers, such as synovial fluid or serum inflammatory and cartilage metabolism markers, which may provide additional information related to KOA pain progression and potentially improve predictive performance. Third, wearable device assessments were limited to 7 consecutive days. Although this duration met the predefined criterion of ≥ 4 valid wear days, extended periods (e.g., 14-30 days) may better capture habitual activity patterns and reduce the influence of short-term behavioral fluctuations. Fourth, the definition of the primary outcome has two limitations. The composite endpoint required both an increase in WOMAC pain score of ≥ 2 points and subjective patient-reported worsening, which may have excluded some patients with clinically meaningful deterioration but without perceived symptom worsening. However, sensitivity analyses using alternative outcome definitions yielded similar model performance (AUC: 0.802-0.809), suggesting that the findings were relatively robust to outcome classification strategies. Nevertheless, the optimal definition of worsening KOA pain remains under debate, as reported MCID values for the WOMAC pain subscale vary substantially (approximately 0.5-2.5 points) depending on population characteristics and estimation methods. Furthermore, the 12-month follow-up period may not fully capture more complex temporal patterns of pain progression, such as transient exacerbation versus persistent worsening. Fifth, the present study focused on binary prediction of pain worsening and did not evaluate the prespecified secondary outcomes, including absolute changes in WOMAC pain, WOMAC function and VAS pain scores. Future studies incorporating continuous measures of symptom trajectories may provide additional insights into the longitudinal progression of KOA symptoms. Sixth, the EPV of about 7.9 was lower than the conventional threshold of 10 recommended for standard logistic regression models. Although LASSO regression with cross-validation was applied to reduce overfitting risk during variable selection, the possibility of model optimism and limited transportability of the selected predictors cannot be fully excluded. Finally, SHAP analysis was performed using the training set and therefore represents an internal interpretation of model behavior rather than causal inference.

Conclusion

This study constructed a predictive model for 12-month KOA pain worsening by integrating baseline clinical characteristics, laboratory-based gait biomechanical parameters and wearable device-derived daily activity behavior indicators. The fusion model demonstrated superior performance compared with the clinical model alone in terms of discrimination, calibration and clinical net benefit. Objective functional measures, including 20-m walking speed, daytime activity fragmentation index and sedentary break frequency, were identified as important contributors to pain worsening prediction in multivariable regression and SHAP analyses. These findings suggest that gait biomechanical assessment and wearable device-based activity monitoring may provide additional predictive value beyond traditional clinical variables in improving risk stratification of KOA patients at higher risk of future pain worsening. Future studies should focus on external validation in multicenter prospective cohorts, longer-duration wearable monitoring, incorporation of biomarker panels, refinement of standardized definitions of KOA pain worsening and evaluation of continuous symptom trajectories in addition to binary outcome prediction.

Acknowledgements

This study was supported by the Medical Discipline Construction Program of Shanghai Pudong New Area Health Commission (the Key Disciplines Program) (PWZxk2022-16); and New Quality Clinical Specialty Program of High-end Medical Disciplinary Construction in Shanghai Pudong New Area (2025-PWXZ-10).

Disclosure of conflict of interest

None.

Address correspondence to: Liangyu Lu, Department of Orthopedics, Pudong New Area People's Hospital, 490#, South Chuanhuan Road, Pudong New Area, Shanghai 201299, China. E-mail: lly198008@163.com

References

[1] GBD 2021 Osteoarthritis Collaborators. Global, regional and national burden of osteoarthritis, 1990-2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Rheumatol 2023; 5: e508-e522.

[2] Collins JE, Yang YH, Opare-Addo MB and Losina E. Quantifying sustained pain worsening in knee osteoarthritis. Osteoarthritis Cartilage 2023; 31: 802-808.

[3] Chang AH, Almagor O, Lee JJ, Song J, Muhammad LN, Chmiel JS, Moisio KC and Sharma L. The natural history of knee osteoarthritis pain experience and risk profiles. J Pain 2023; 24: 2175-2185.

[4] Neelapala YVR, Neogi T, Kumar D, Jarraya M, Macedo L, Kobsar D, Hanna S, Frey-Law LA, Lewis CE, Nevitt M, Appleton T, Birmingham T and Carlesso LC. Exploring different models of pain phenotypes and their association with pain worsening in people with early knee osteoarthritis: the MOST cohort study. Osteoarthritis Cartilage 2024; 32: 210-219.

[5] James KA, Corrigan P, Yen SC, Hasson CJ, Davis IS and Stefanik JJ. Effects of increasing walking cadence on gait biomechanics in adults with knee osteoarthritis. J Biomech 2024; 177: 112394.

[6] Ulrich B, Pereira LC, Jolles BM and Favre J. Walking with shorter stride length could improve knee kinetics of patients with medial knee osteoarthritis. J Biomech 2023; 147: 111449.

[7] Sougué C, Diallo M, Bayala YLT, Ayouba Tinni I, Kaboré F, Zabsonré Tiendrebeogo WJS, Dakouré PWH and Ouedraogo DD. Clinical phenotypes and associated factors in knee osteoarthritis in an African black population. Osteoarthr Cartil Open 2025; 7: 100570.

[8] Castagno S, Birch M, van der Schaar M and McCaskie A. Predicting rapid progression in knee osteoarthritis: a novel and interpretable automated machine learning approach, with specific focus on young patients and early disease. Ann Rheum Dis 2025; 84: 124-135.

[9] World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013; 310: 2191-2194.

[10] Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, Christy W, Cooke TD, Greenwald R, Hochberg M and et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986; 29: 1039-1049.

[11] Kellgren JH and Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957; 16: 494-502.

[12] Radloff L. The CES-D scale: a self-report depression scale for research in the general population. Applied Psychological Measurement 1977; 1: 385-401.

[13] Freedson PS, Melanson E and Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998; 30: 777-781.

[14] Bellamy N, Buchanan WW, Goldsmith CH, Campbell J and Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988; 15: 1833-1840.

[15] Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, Bierma-Zeinstra S, Brandt KD, Croft P, Doherty M, Dougados M, Hochberg M, Hunter DJ, Kwoh K, Lohmander LS and Tugwell P. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis Cartilage 2007; 15: 981-1000.

[16] Lundberg SM and Lee SI. A unified approach to interpreting model predictions. Advances in neural information processing systems 2017; 30: 1-7.

[17] Cijs B, Stekelenburg R, Veenhof C, Knoop J, Boymans T, de Rooij M and Kloek C. Prognostic factors and changes in pain, physical functioning and participation in patients with hip and/or knee osteoarthritis: a systematic review. Arthritis Care Res (Hoboken) 2025; 77: 228-239.

[18] Wang J, Hu Q, Wu C, Li S, Deng Q, Tang R, Li K, Nie Y and Shen B. Gait asymmetry variation in kinematics, kinetics and muscle force along with the severity levels of knee osteoarthritis. Orthop Surg 2023; 15: 1384-1391.
[19] Gonzalez FF, Leporace G, Franciozi C, Cockrane M, Metsavaht L, Carpes FP, Chahla J and Luzo M. Clinical and radiographic characterization of three-dimensional gait profiles of patients with knee osteoarthritis. Knee 2023; 44: 211-219.
[20] Mehdipour A, Malouka S, Beauchamp M, Richardson J and Kuspinar A. Measurement properties of the usual and fast gait speed tests in community-dwelling older adults: a COSMIN-based systematic review. Age Ageing 2024; 53: afae055.
[21] Gilbert AL, Song J, Cella D, Chang RW and Dunlop DD. What is an important difference in gait speed in adults with knee osteoarthritis? Arthritis Care Res (Hoboken) 2021; 73: 559-565.
[22] Dewig DR, Evans-Pickett A, Pietrosimone BG and Blackburn JT. Comparison of discrete and continuous analysis approaches for evaluating gait biomechanics in individuals with anterior cruciate ligament reconstruction. Gait Posture 2023; 100: 261-267.
[23] Lohmann A, Groenwold RHH and van Smeden M. Comparison of likelihood penalization and variance decomposition approaches for clinical prediction models: a simulation study. Biom J 2024; 66: e2200108.
[24] Qi H, Zhang B, Lu D and Lian F. A nomogram model for predicting prolonged postoperative LOS after total knee arthroplasty in osteoarthritis patients. Clin Interv Aging 2025; 20: 2481-2492.
[25] Hoffman RM, Davis-Wilson HC, Hanlon S, Swink LA, Kline PW, Juarez-Colunga E, Melanson EL and Christiansen CL. Maximal daily stepping cadence partially explains functional capacity of individuals with end-stage knee osteoarthritis. PM R 2024; 16: 532-542.
[26] Small SR, Khalid S, Price AJ and Doherty A. Device-measured physical activity in 3506 individuals with knee or hip arthroplasty. Med Sci Sports Exerc 2024; 56: 805-812.
[27] Bandara A, Shimizu H, Watanabe D, Misa T, Suzuki S, Tanigawa K, Nagai-Tanima M and Aoyama T. Diurnal variations in gait parameters among older adults with early-stage knee osteoarthritis: insights from wearable sensor technology. Sci Rep 2025; 15: 8026.