Fan *et al. Arthritis Research & Therapy* (2024) 26:213
https://doi.org/10.1186/s13075-024-03450-2

Arthritis Research & Therapy

RESEARCH Open Access

XGBoost-SHAP-based interpretable diagnostic framework for knee osteoarthritis: a population-based retrospective cohort study

Zijuan Fan[1,2†], Wenzhu Song[3†], Yan Ke[4†], Ligan Jia[5], Songyan Li[1], Jiao Jiao Li[6], Yuqing Zhang[7], Jianhao Lin[4*] and Bin Wang[1*]

Abstract

Objective To use routine demographic and clinical data to develop an interpretable individual-level machine learning (ML) model to diagnose knee osteoarthritis (KOA) and to identify highly ranked features.

Methods In this retrospective, population-based cohort study, anonymized questionnaire data was retrieved from the Wu Chuan KOA Study, Inner Mongolia, China. After feature selections, participants were divided in a 7:3 ratio into training and test sets. Class balancing was applied to the training set for data augmentation. Four ML classifiers were compared by cross-validation within the training set and their performance was further analyzed with an unseen test set. Classifications were evaluated using sensitivity, specificity, positive predictive value, negative predictive value, accuracy, area under the curve(AUC), G-means, and F1 scores. The best model was explained using Shapley values to extract highly ranked features.

Results A total of 1188 participants were investigated in this study, among whom 26.3% were diagnosed with KOA. Comparatively, XGBoost with Boruta exhibited the highest classification performance among the four models, with an AUC of 0.758, G-means of 0.800, and F1 scores of 0.703. The SHAP method reveals the top 17 features of KOA according to the importance ranking, and the average of the experience of joint pain was recognized as the most important features.

Conclusions Our study highlights the usefulness of machine learning in unveiling important factors that influence the diagnosis of KOA to guide new prevention strategies. Further work is needed to validate this approach.

Keywords Knee osteoarthritis, Classification, XGBoost, Boruta, SHAP, Machine learning

†Zijuan Fan, Wenzhu Song and Yan Ke equal first authors.

*Correspondence:
Jianhao Lin
linjianhao@pkuph.edu.cn
Bin Wang
wangbin_pku@zju.edu.cn

Full list of author information is available at the end of the article

Introduction

Osteoarthritis (OA) is a widespread and debilitating musculoskeletal condition, posing a substantial and growing health burden with significant implications for both individuals and healthcare systems [1–3]. Among all types of OA, knee OA (KOA) is the greatest contributor to the chronic disease burden, leading to high morbidity and disability globally [4, 5]. KOA affects the 3 compartments of the knee joint (medial, lateral, and patellofemoral) and usually progresses slowly over 10 to 15 years, significantly impeding daily life activities [6]. In China, KOA affects approximately 14.6% of the population, with higher prevalence among females compared to males, as well as significantly higher prevalence in rural compared to urban areas [7]. Globally, an estimated 25% of the adult population is affected by KOA, making it a formidable public health challenge. Since curative clinical treatments are not available [8], current therapies focus on relieving pain and preserving joint function. The growing burden on individuals, families, and healthcare systems emphasizes the need for further research on KOA and the implementation of preventive measures.

X-ray imaging represents a traditional and widely used approach for KOA diagnosis, offering a non-invasive assessment of joint space width (JSW) and osteophytes while effectively evaluating the joint conditions to identify fractures, dislocations, and joint space narrowing (JSN) [6, 9]. According to the most widely used Kellgren and Lawrence (KL) radiographic grading system, KOA is categorized as: Grade 1, characterized by doubtful JSN and possible osteophytic lipping; Grade 2, includes definite osteophytes and possible JSN on anteroposterior weight-bearing radiographs; Grade 3, features multiple osteophytes, definite JSN, sclerosis, and potential bony deformity; Grade 4, marked by large osteophytes, severe JSN, pronounced sclerosis, and definite bony deformity [10]. However, X-ray imaging can be expensive and the correlation of image features with symptoms is unclear, especially in the early stages of KOA when the articular cartilage is not significantly deteriorated [11]. Additionally, the interpretation of radiographs requires the expertise of experienced physicians or radiologists to ensure accurate assessment and diagnosis of pathological changes [12]. To address these limitations, a new approach is to develop a KOA diagnostic model that does not rely on X-ray imaging as a routine screening tool in public healthcare settings. Using easily accessible variables including patient self-report and demographic information, this study seeks to identify individuals at high risk of developing KOA. The early identification of high-risk patients may facilitate prompt KOA diagnosis and management, hence helping to slow the disease progression. A variety of models may be considered for this purpose, however, traditional statistical models may struggle to adequately capture the intricate relationships between various features and KOA diagnosis, given their suboptimal performance in handling non-linear relationships [13, 14].

Artificial intelligence (AI) including machine learning (ML) has recently enabled a surge of technological breakthroughs, leading to significant advances and discoveries in healthcare [15]. Due to its powerful non-linear modelling capabilities and ability to process large amounts of data, ML has been increasingly applied to aid the study of rheumatology and KOA [16]. However, most existing ML models in these areas have been developed using population data from Europe and the US [16]. The evidence for their application in a broad range of clinical settings, along with interpretable risk classification models for disease prognosis, remains limited particularly in Asia-Pacific and other populations not sufficiently represented in the training data [16, 17]. For example, studies have shown that East Asian populations, particularly women, may be more vulnerable to KOA compared to Caucasian populations [18–20]. This could be attributed to anatomical differences such as the greater prevalence of valgus distal femurs in East Asians, which is less common in Caucasians [18, 19, 21, 22]. Furthermore, lifestyle and genetic factors, as well as access to healthcare can also contribute to regional differences in KOA prevalence and progression. These variations suggest that ML models developed using data from European and US populations might not fully account for the unique features of KOA in the Asia-Pacific region, limiting the applicability and accuracy of these models when used for diagnosis and prediction in other populations.

In this study, we aimed to develop diagnostic models for identifying risk of KOA among Asian individuals using data from clinical examinations, along with demographic factors such as sex, education, physical function and activity, disease and symptom history, and anatomical and functional measurements, derived from the Wu Chuan KOA Study. Additionally, SHapley Additive exPlanations (SHAP) was employed to interpret the best-performing ML model and to investigate prognostic factors associated with KOA.

Methods

Study design and participants

The Wu Chuan KOA cohort study is a retrospective study aimed at studying KOA prevalence and its determinants among rural residents aged 50 and above in Wuchuan County, Inner Mongolia, China. A total of 1228 participants completed a questionnaire at home or work on December 31, 2005. The same participants also underwent a clinical examination at Wuchuan Hospital on the same day [22–24]. At 96 months, surviving participants from the original Wu Chuan KOA cohort were

invited for a follow-up visit [25]. During the visit, 1188 participants were asked to complete the same questionnaire and received the same clinical examinations as the baseline visit [25–27].

The questionnaire was designed by the author (YQZ) based on the Framingham OA questionnaire, and was used for subsequent epidemiological investigation of OA in Beijing, Wuchuan and Xiang Ya [20, 22]. Topics covered in the questionnaire included the Chinese version of the Medical Outcome Study Short Form (SF-12), physical function and activity, disease history, past medical conditions, and symptomatic history. After collecting baseline demographic data, each participant was asked to complete the validated Chinese version of the SF-12, a widely used tool to assess health-related quality of life. The SF-12 has been extensively validated for use with OA patients [28].

Participants were also asked questions on their current daily physical activities, including cleaning, cooking, and walking, as well as about their disease history or any medical conditions such as high blood pressure, diabetes, and heart disease. To assess the condition of their knee joints and the impact on their daily activities, participants were asked if they had ever experienced a knee injury that prevented them from walking for at least a week and about the type of work they had done for the longest period of time.

Moreover, information on knee health and function was collected, including whether participants had experienced symptoms such as joint pain, stiffness, or soreness, and whether they had experienced pain in, around, or behind the knee that lasted for at least one month. The interviews were led by trained researchers who followed a standardized protocol to ensure consistency. Relevant studies that discuss the development and application of this questionnaire have been published, and further details can be found in the Supplementary text [22–27, 29, 30].

In addition to the questionnaire, clinical assessments were also performed by the participants including height and weight measurements, femorotibial alignment angulation, knee range of motion, 50-foot walk time, and weight-bearing posterior-anterior semi-flexed radiographs of both knees [24, 29]. All interviewers, clinical examiners, and radiograph technicians received training under the supervision of the study's chief investigators [26].The chief investigator, trained at Boston University, used the KL-grades criteria to assess KOA. Right knee with KL-grades≥2 was considered to have KOA. The weighted kappa for KL-grades for inter-rater reliability was 0.86 (95% CI: 0.72–0.88) and the intra-rater reliability was 0.92 (95% CI: 0.86–0.99). The follow-up data from 2013 and the questionnaire responses from 1188 individuals were used in this study.

Written informed consent was obtained for all study participants, and ethical approval was granted by the Ethics Committee of Peking University People's Hospital, Beijing, China (Approval Number: No. 2012–040). All methods were performed following the relevant guidelines and regulations.

To ensure the model's ability to generalize and avoid overfitting, the dataset was split into training and testing sets with a 7:3 ratio (831:357). The model was trained on the training set, and its performance on clean data was evaluated using the testing set. Subsequently, data preprocessing was conducted, including three feature selection methods for highly KOA-related feature elimination, as well as balancing the outcome categories (KOA and non-KOA). Next, four ML algorithms were used to develop diagnosis models in the training dataset, followed by validation in the internal testing dataset, as shown in Fig. 1.

Feature selection

Feature selection is a common preprocessing method in machine learning (ML) that involves selecting features with high predictive potential from the original data, thereby enhancing the effectiveness and performance of the ML model [31, 32]. For data processing and feature selection, we combined 78 variables from the Wu Chuan questionnaire records, including the SF-12, physical function and activity, diseases history, past medical conditions, symptomatic history, examination items, and demographic features. Variables with a high proportion of missing data (>70% missing rate) were first eliminated. Features that contributed minimally or not at all to the outcome variable were then eliminated using feature selection techniques including the Least Absolute Shrinkage and Selection Operator (LASSO), adaptive LASSO (AdaLASSO) and Boruta, respectively [33–35]. Variables with a correlation greater than 0.7 were eliminated using a correlation analysis, in conjunction with clinical knowledge, to prevent collinearity [36].

A total of 40 potential classification features were included, comprising 27 qualitative features (such as sex, education, physical function and activity, disease history, past medical and symptom history) and 13 continuous features (including age, BMI, Short Form-12 (SF-12), work years, bilateral femoral and tibial alignment angle measurement, 50-foot walk time, and bilateral knee range of motion measurement). More details and assignments of qualitative features are shown in Supplementary Table S1.

LASSO is a widely used feature selection method that identifies important features by applying an L1 regularization penalty term to the model [34]. The penalty encourages feature coefficients to shrink towards zero, resulting in a sparse model that retains only the most

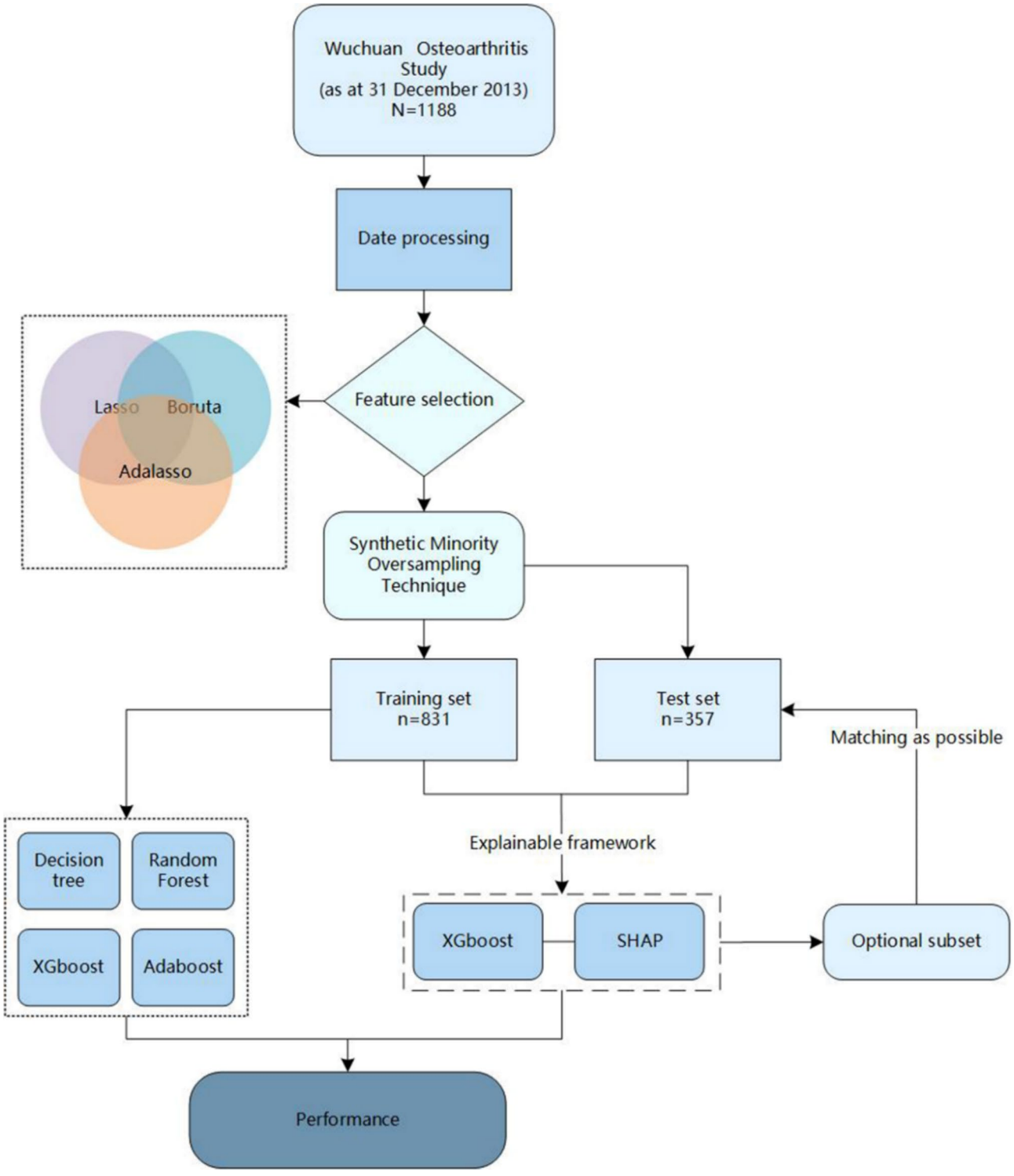

Fig. 1 Study design and workflow of cohort derivation

significant predictors for the outcome variable. The effectiveness of this method is evaluated by assessing the model's predictive performance, typically using model performance metrics (e.g., AUC, accuracy) to determine the features making the highest contribution to the prediction of the outcome variable [37].

In contrast to traditional LASSO, AdaLASSO introduces adaptive weights to address challenges posed by high-dimensional data and collinearity [35]. By adjusting

penalty weights based on feature importance, Ada-LASSO applies weaker penalties to important features and stronger penalties to less important ones [38]. Firstly, the initial feature estimates were obtained by using the traditional LASSO method. Then, based on the initial estimates, the weight coefficients of each feature were calculated, and these weights indicated the relative importance of the feature to the target variable. Next, the weight coefficients were applied to the loss function to obtain the final feature estimates by minimizing the loss function with weights. Finally, cross-validation and other techniques were used to select appropriate regularization parameters, controlling the model's degree of regularization and the stringency of feature selection. This adaptive approach enhances the accuracy and stability of feature selection, ultimately leading to improved results.

Boruta assesses feature importance by comparing the accuracy of models trained with permuted features against the accuracy of the original model [33]. This approach is not constrained by linear assumptions, making it suitable for diverse datasets. Additionally, Boruta can effectively handle high-dimensional data and highly correlated features. In contrast to other feature selection techniques, the Boruta algorithm can automatically select relevant features without the need of feature engineering or prior knowledge. It minimizes human intervention and increases the accuracy and efficiency of the feature selection process [39].

Synthetic minority oversampling technique

In our dataset, the number of non-KOA cases was almost three times higher than KOA (613:218), leading to a significant class imbalance that could translate into poor model performance. To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was employed, a method which increases the number of samples in a minority category by synthesizing new samples [40]. SMOTE creates synthetic samples for the minority class by interpolating between the nearest neighbors of each existing minority sample. This technique enhances the quantity of minority class samples, thereby aiding the model in effectively capturing the features and decision boundaries of the minority class. Specifically, by generating synthetic samples that resemble real minority instances, SMOTE allows the classifier to better learn the characteristics of the minority class. This can lead to improved model performance, especially in the context of highly imbalanced datasets, where traditional classifiers may bias towards the majority class. The effectiveness of SMOTE in addressing class imbalance has been well-documented in previous studies [39, 41].

Development of classification models

In this study, four ML algorithms (decision tree [42], random forest (RF) [43], eXtreme gradient boosting (XGBoost) [44], and adaptive boosting (Adaboost)] [45] were adopted for model training. Hyperparameters for each algorithm were optimized using 10 repeats of 5-fold cross-validation. The goal was to maximize the area under the receiver operating characteristic (ROC) curve (AUC) to enhance the predictive performance of the training model.

The decision tree algorithm offers the advantage of being easily interpretable and understandable [42]. It applies to various types of data and has a relatively short training time. RF excels in reducing overfitting and performs well with high-dimensional data [43]. XGBoost demonstrates remarkable performance and scalability, adapting regularization techniques and approximate splitting algorithms to enhance prediction accuracy [44]. For example, XGBoost has consistently been shown to consistently outperform other ML models, such as decision trees and SVMs, in accuracy and AUC when predicting disease outcomes across multiple large-scale clinical datasets [46, 47]. XGBoost also incorporates regularization techniques to avoid overfitting, making it suitable for complex, high-dimensional health data [48]. Additionally, its ability to handle missing values and efficiency in model training are significant advantages in real-world applications [49]. Meanwhile, AdaBoost is particularly suitable for handling noisy data [45]. It enhances the predictive power of decision trees through weighted majority voting. More detailed descriptions of these models are available in the Supplementary text.

Evaluation of model performance

Our metrics for model evaluation included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, G-means, AUC and F1 scores. Specificity measures the model's ability to correctly classify negative samples as negative, while sensitivity assesses model accuracy in correctly classifying positive samples as positive. By considering all classes, accuracy provides a general overview of the model's overall performance. PPV measures how accurately the model predicts positive samples, while NPV measures how accurately the model predicts negative samples. AUC, as a general measure of model discrimination, reflects the model's ability to distinguish between positive and negative classes. The G-means metric provide a balanced assessment of the model's performance between the positive and negative classes. Finally, the F1 score represents a harmonious blend of accuracy and recall, and is particularly valuable in unbalanced datasets. AUC, F1 score, and G-means are the three comprehensive metrics we prioritize in model evaluation. Specifically, we considered

a value above 0.7 for these metrics as indicative of good model performance [41]. In summary, the combination of these metrics provides a comprehensive understanding of model performance, helping to assess its accuracy and applicability.

$$Sensitivity = \frac{TP}{(TP+FN)} * 100\% \quad (1)$$

$$Specificity = \frac{TN}{(TN+FP)} * 100\% \quad (2)$$

$$PPV = \frac{TP}{TP+FP} * 100\% \quad (3)$$

$$NPV = \frac{TN}{TN+FN} * 100 \quad (4)$$

$$Accurary = \frac{TN+TP}{(TP+TN+FP+FN)} * 100\% \quad (5)$$

$$G\,mean = \sqrt{\frac{TP}{TP+FN} * \frac{TN}{TN+FP}} * 100\% \quad (6)$$

$$F1\ score = \frac{2Precision}{(Precision+Recall)} * 100\% \quad (7)$$

PPV: Positive predictive value; NPV: Negative predictive value; TP: True positive; FP: False positive; TN: True negative; FN: False negative.

Shapley additive explanations

For interpreting the classification model, the Shapley Additive Explanations (SHAP) method was employed [50]. This approach accurately calculates the contribution and influence of each feature towards the final predictions. One advantage of the SHAP method lies in its ability to visualize the relationships and interactions between features without the need for complex numerical derivations [51]. It provides graphical representations that demonstrate how the association between exposures and outcomes varies with the distribution of another feature. This allows for a more intuitive understanding of the interactions among different factors. SHAP values are computed at the individual level to account for the importance of predictors [52]. In epidemiological studies, population-level SHAP values can provide a concise numerical summary of interaction effects, expressing their direction and magnitude, thereby facilitating the interpretation of ML results [53]. Notably, larger SHAP values indicate a greater impact or contribution of a feature towards sample identification [54]. This enables a better understanding of the importance of each feature in sample identification.

Statistical analysis

All statistical analyses were performed using SPSS (version 25.0), R software (version 4.2.0), and Python (version 3.8.0). Categorical variables were presented as numbers (percentages) and tested by Chi-square (or Fisher's exact) tests. Continuous variables were presented as mean ± standard deviation or median (25–75 percentiles), and were tested by student's t-test or Wilcoxon rank sum tests. $P<0.05$ was considered statistically significant.

Results

Cohort participants

A total of 1188 participants were included in this study. The dataset was randomly divided into 2 parts to make the training set ($n=831$) and testing set ($n=357$). The baseline participant characteristics are presented in Table 1. The proportion of KOA cases in the overall derivation cohort was consistently maintained in both the training and testing cohorts. Within the training cohort, 26.2% of individuals had been diagnosed with KOA, and 58.6% had received only primary school education or below. The mean age was 59 years, with an average BMI of 23.0 ± 3.5 kg/m^2. The participants had an average work history of 35 years, and 95.8% of them engaged in heavy physical activity. Additionally, 9.3% of participants reported a history of previous fractures. Notably, there were no statistically significant differences in participant characteristics between the training and testing datasets ($P>0.05$).

Feature selection

Three feature selection techniques, LASSO, AdaLASSO, and Boruta, were utilized to select relevant features. After collinearity analysis, these methods respectively yielded 21, 24, and 17 retained features. For features with a correlation coefficient greater than 0.7, we followed the advice of clinical physicians to retain one of them. Supplementary Figure S1-3 displayed the correlation heatmap generated using different feature selection methods. Sex, Age, B01, B03, C16, BMI, right Align, WTest, RROM, RROM1, and RROM2 were the features consistently selected by all three feature selection methods. Further details regarding feature selection results can be found in Supplementary Table S2.

Model establishment and evaluation

Before modeling, the training dataset underwent oversampling using SMOTE to achieve a balanced sample distribution of KOA and non-KOA patients at a ratio of 1:1 (613:613). The distribution of Y training set sample content after SMOTE oversampling is shown in Supplementary Figure S4. Figure 2 shows the discrimination performance metrics of the classification models in the testing set. Among the different model combinations, Adaboost-AdaLASSO achieved the highest sensitivity but recorded the lowest specificity. All models, except for the Adaboost combination, exhibited an accuracy

Table 1 Baseline demographics and clinical characteristics

Characteristics	Derivation cohort	
	Training set ***n*** **= 831**	**Testing set** ***n*** **= 357[a]**
Age, years	59 ± 8.5	60 ± 9.1
BMI, kg/m^2	23.0 ± 3.5	23.4 ± 3.9
SF-12, scores	36.9 ± 5.38	37.2 ± 5.39
Work years, years	35.6 ± 8.1	35.7 ± 8.3
Left Align, degrees	0.84 ± 3.14	0.77 ± 2.97
Right Align, degrees	-0.02 ± 3.17	0.04 ± 2.98
WTest, degrees	14.1 ± 3.8	14.2 ± 4.6
Left ROM, degrees	137.5 ± 10.9	137.6 ± 9.1
Left ROM1, degrees	1.1 ± 2.6	1.3 ± 2.8
Left ROM2, degrees	1.6 ± 2.0	1.6 ± 2.0
Right ROM, degrees	136.4 ± 10.5	136.6 ± 9.7
Right ROM1, degrees	1.5 ± 3.0	1.5 ± 2.7
Right ROM2, degrees	1.6 ± 2.0	1.5 ± 2.0
KOA (%)		
Yes	218(26.2)	95(26.6)
No	613(73.8)	262(73.4)
Sex		
Female	464(55.8)	189(52.9)
Education		
Primary school or below	487(58.6)	211(59.1)
Junior high school	269(32.5)	110(30.8)
High school	70(8.5)	32(9.3)
College or above	3(0.4)	3(0.8)
Physical function and activity		
Walk a mile		
No difficulties	640(77.0)	282(79.0)
Experience challenges	170(20.5)	68(19.0)
Highly challenging	7(0.8)	5(1.4)
Unable to complete	13(1.6)	2(0.6)
Uncertain	1(0.1)	0(0.0)
Walk two miles		
No difficulties	405(48.7)	183(51.3)
Experience challenges	275(33.1)	115(32.2)
Highly challenging	81(9.7)	37(10.4)
Unable to complete	46(5.5)	17(4.8)
Uncertain	12(1.4)	3(0.8)
Need rest when walking to the 1st floor		
No difficulties	671(80.7)	293(82.1)
Experience challenges	146(17.6)	58(16.2)
Highly challenging	6(0.7)	4(1.1)
Unable to complete	7(0.8)	2(0.6)
Uncertain	1(0.1)	0(0.0)
Bend over, squat or kneel		
No difficulties	324(39.0)	148(41.5)
Experience challenges	444(53.4)	185(51.8)
Highly challenging	50(6.0)	22(6.2)
Unable to complete	13(1.6)	2(0.6)
Do housework		
No difficulties	721(86.8)	312(87.4)
Experience challenges	101(12.2)	41(11.5)
Highly challenging	5(0.6)	3(0.8)

Table 1 (continued)

Characteristics	Derivation cohort	
	Training set *n* = 831	Testing set *n* = 357[a]
Unable to complete	4(0.5)	1(0.3)
Cooking		
No difficulties	738(88.8)	323(90.5)
Experience challenges	84(10.1)	27(7.6)
Highly challenging	2(0.2)	5(1.4)
Unable to complete	6(0.7)	2(0.6)
Uncertain	1(0.1)	0(0.0)
Walk between rooms?		
No difficulties	802(96.5)	346(96.9)
Experience challenges	23(2.8)	9(2.5)
Highly challenging	2(0.2)	2(0.6)
Unable to complete	4(0.5)	0(0.0)
Stand from straight seat		
No difficulties	627(75.5)	253(70.9)
Experience challenges	185(22.3)	92(25.8)
Highly challenging	17(2.0)	12(3.4)
Unable to complete	2(0.2)	0(0.0)
Get in/out of bed		
No difficulties	671(80.7)	278(77.9)
Experience challenges	156(18.8)	74(20.7)
Highly challenging	3(0.4)	5(1.4)
Unable to complete	1(0.1)	0(0.0)
Set table/use chopsticks/drink		
No difficulties	808(97.2)	346(96.9)
Experience challenges	21(2.5)	11(3.1)
Unable to complete	2(0.2)	0(0.0)
Dress (shoes, zippers, buttons)		
No difficulties	781(94.0)	330(92.4)
Experience challenges	47(5.7)	27(7.6)
Highly challenging	1(0.1)	0(0.0)
Unable to complete	2(0.2)	0(0.0)
Disease history		
Heart disease		
No	581(69.9)	249(69.7)
Yes	181(21.8)	80(22.4)
Uncertain	69(8.3)	28(7.8)
Hypertension disease		
No	509(61.3)	236(66.1)
Yes	304(36.6)	110(30.8)
Uncertain	18(2.2)	11(3.1)
Lung disease		
No	694(83.5)	297(83.2)
Yes	123(14.8)	52(14.6)
Uncertain	14(1.7)	8(2.2)
Diabetes		
No	746(89.8)	320(89.6)
Yes	30(3.6)	13(3.6)
Uncertain	55(6.6)	24(6.7)
Kidney disease		
No	708(85.2)	295(82.6)
Yes	45(5.4)	21(5.9)

Table 1 (continued)

Characteristics	Derivation cohort	
	Training set *n* = 831	Testing set *n* = 357[a]
Uncertain	78(9.4)	41(11.5)
Malignant tumor		
No	811(97.6)	349(97.8)
Yes	7(0.8)	2(0.6)
Uncertain	13(1.6)	6(1.7)
Past situation		
Knee injury affected walking (at least a week)		
No	733(88.2)	312(87.4)
Left knee	42(5.0)	24(6.7)
Right knee	47(5.7)	19(5.3)
Bilateral knee	9(1.1)	2(0.6)
Low back injury (at least a week)		
No	765(92.0)	337(94.4)
Yes	65(8.0)	20(5.6)
Fractured		
No	754(90.7)	329(92.2)
Yes	77(9.3)	28(7.8)
Work mode		
Sedentary	1(0.1)	0(0.0)
Mild	9(1.2)	3(0.8)
Moderate	23(2.9)	13(3.8)
Heavy	796(95.8)	340(95.4)
Symptomatic history		
Knee or surrounding pain (at least a month)		
No	300(36.1)	124(34.7)
Left knee	118(14.2)	54(15.1)
Right knee	135(16.2)	65(18.2)
Bilateral knee	278(33.5)	114(31.9)
Joint pain/stiffness/soreness (at least a month in past year)		
No	310(37.3)	127(35.6)
Left knee	114(13.7)	52(14.6)
Right knee	127(15.3)	60(16.8)
Bilateral knee	280(33.7)	118(33.1)
Limited activities due to knee pain/sore/stiff (past month)		
No	500(60.2)	223(62.5)
Yes	331(39.8)	134(37.5)

[a] No significant differences observed between training and testing cohorts (all $p > 0.05$). BMI: Body Mass Index; SF-12: Short Form 12 Health Survey; Align: femur-tibia angle; WTest: timing a 50-foot walk; ROM: knee flexion angle; ROM1: knee extension angle; ROM2: knee excessive flexion or extension angle; KOA: knee osteoarthritis

above 0.7. The combinations of different feature selection methods with RF, XGBoost, and Adaboost algorithms all yielded AUC values greater than 0.7. However, the AUC values for decision tree with Lasso and decision tree with Boruta were only 0.67 and 0.66, respectively. When considering the F1 score and G-means, XGBoost with the Boruta algorithm outperformed all other model combinations. The final hyperparameters employed in the 12 ML models are detailed in Supplementary Table S3.

Explanation of XGBoost model with the SHAP

Major indicators defined by SHAP

To provide a more visually interpretable model output, we introduced SHAP to identify variables most strongly correlated with KOA as determined by the XGBoost-Boruta model. The bar charts demonstrate the most significant variables in descending order. As shown in Fig. 3A, the top five indicators were: whether participants have experienced joint pain, stiffness, or soreness for at least 1 month continuously in the past 12 months; right knee flexion angle; right knee extension angle; right knee excessive flexion or extension; and age.

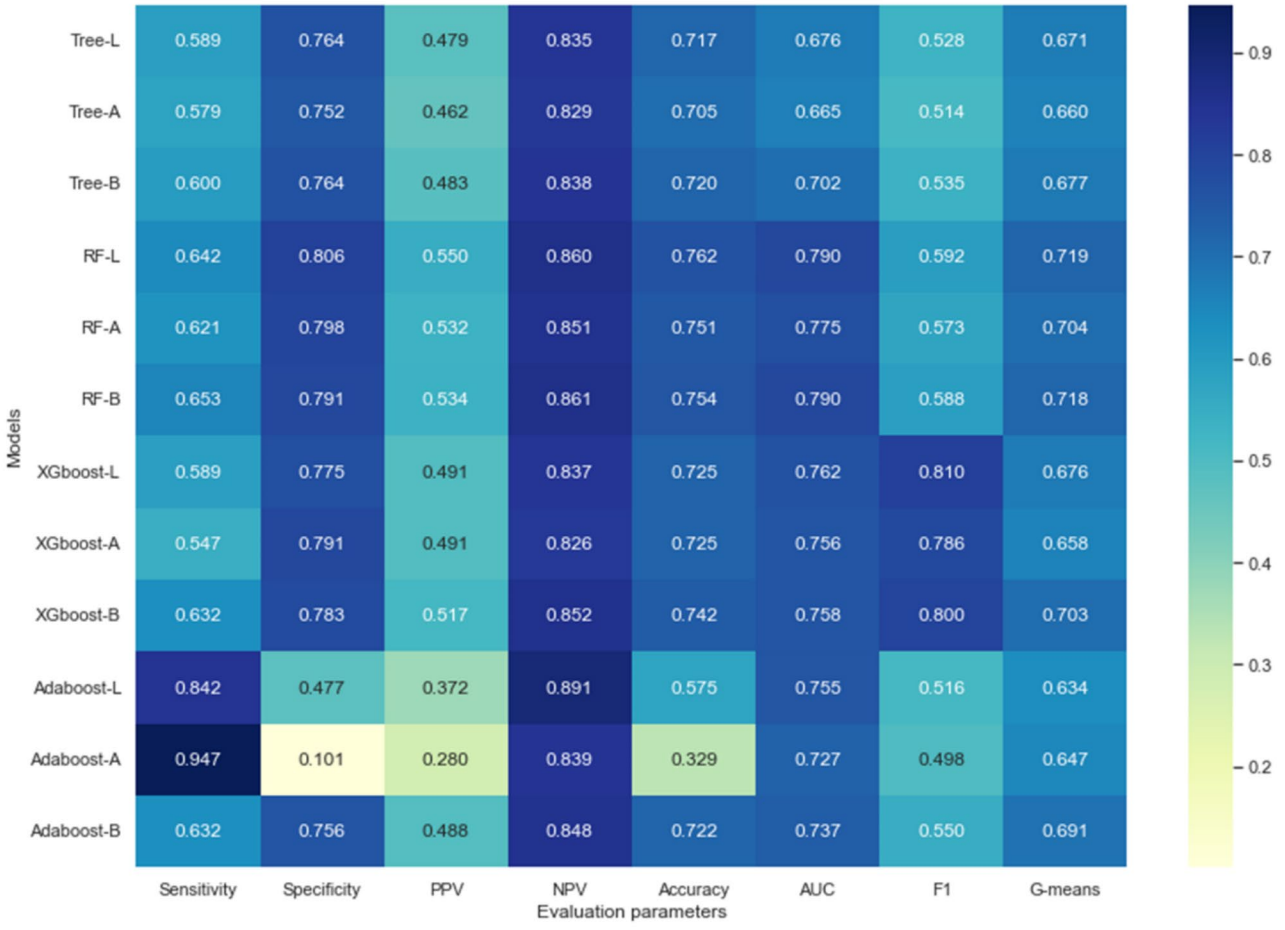

Fig. 2 Discrimination performance of KOA classification models in testing set. Tree: decision tree; L: lasso; **A** AdaLASSO; **B** boruta; XGBoost: eXtreme Gradient Boosting; Adaboost: adaptive boosting; PPV: positive predictive value; NPV: negative predictive value; AUC: area under the curve

Furthermore, to detect the positive and negative relationships between these variables and target outcomes, SHAP values were applied to uncover the risk factors associated with KOA. As shown in Fig. 3B, the horizontal position indicates whether the effect of that value is associated with higher or lower factors, while the color indicates whether that variable is high (in red) or low (in blue) for the specific observation.

Clearly, participants who experienced joint pain, stiffness, or soreness for at least 1 month continuously in the past 12 months had a positive impact, pushing the outcome toward KOA. Conversely, an increase in the right knee flexion angle had a negative impact, pushing the outcome toward non-KOA. Age and increase in knee extension angle were both factors that increased the likelihood of KOA.

SHAP individual force plot

SHAP value represents a useful approach for revealing the prediction-related characteristics of individual patients and the contribution of each feature to the KOA prediction. The bold-faced numbers are the probabilistic predicted values f(x), while the base values are the values predicted without giving input to the model. The f(x) is the log odds ratio of each observation. The red and blue features respectively indicate features that increase and decrease the KOA risk. The length of the arrows helps visualize the magnitude of the feature effect on the prediction, where a longer arrow length signifies a greater effect.

To highlight the clinical utility and translational impact of such features in KOA diagnosis, we present cases of non-KOA patients and KOA patients separately (Fig. 4). Participant No. 10,031, a 65-year-old female with KOA, experiences difficulties in bending, squatting, or kneeling, takes 15 s to walk 50 feet, and exhibits a 1-degree external rotation of the right femur-tibia angle, with a range of motion in the right knee joint of 119° flexion and 6° extension. In this example, the ML output explanations highlighted the relevant features associated with KOA, although these features are not necessarily causal or modifiable. Figure 4B shows the SHAP force plot for

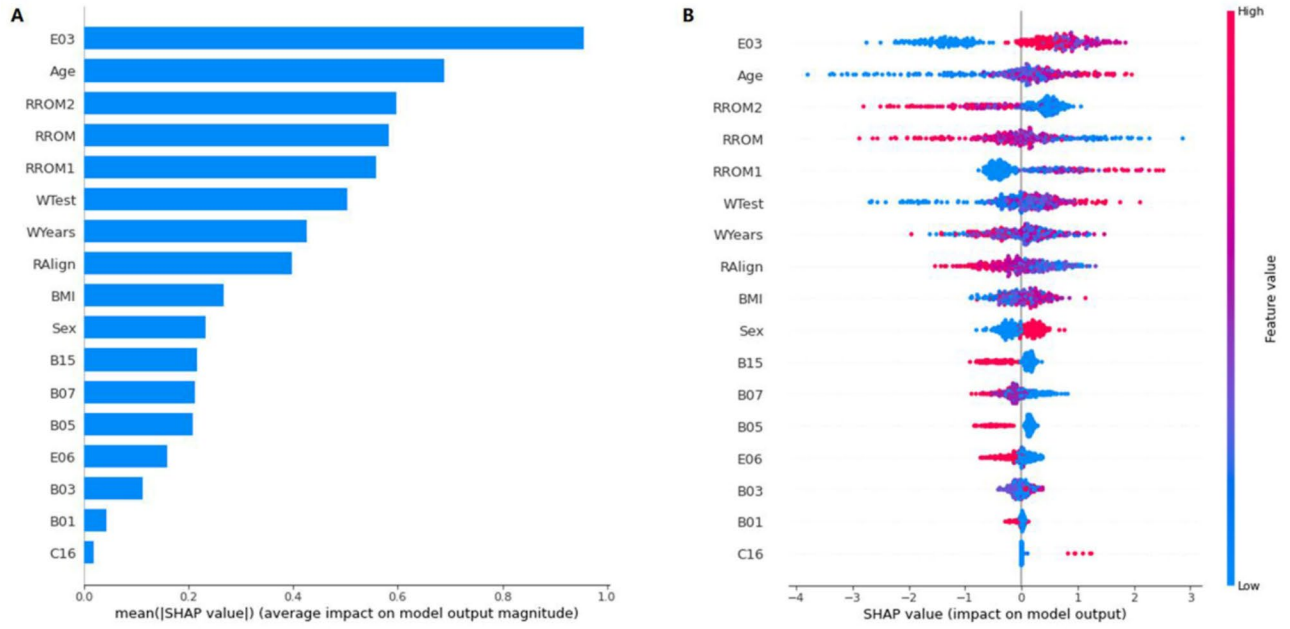

Fig. 3 SHAP summary plot. (**A**) Bar charts that rank the importance of 17 indicators identified by SHAP values. (**B**) Distribution of the impact each feature had on the full model output using SHAP values. RROM: knee flexion angle; RROM1: knee extension angle; RROM2: knee excessive flexion or extension; WTest: timing a 50-foot walk; WYears: work yesrs; RAlign: right femur-tibia angle; BMI: body mass index. E03: In the past 12 months, have you ever had joint pain, stiffness, or soreness that lasted at least 1 month? B15: Stand up from a straight-back seat without armrests? B07: Bend over, squat or kneel? B05: You don't need to rest when you walk to the first floor? E06: In the past month, have you restricted your daily activities because of knee pain, sore and stiff knees? B03: Walk two miles? B01: Walk a mile? C16: Walk a mile?

this participant, and Fig. 4D provides the corresponding X-ray image of the right knee joint.

Discussion

In the current study, we utilized data from the Wu Chuan KOA Study, a population-based cohort in China, to develop diagnosis models for KOA using 4 ML algorithms. The diagnosis models were validated in the internal testing cohort. Remarkably, the XGBoost-Boruta algorithm combination, incorporating 17 features, exhibited the best performance, achieving specificity of 78%, AUC of 0.76, and F1 score of 0.8 in the diagnosis of KOA. These findings point to the robust performance of the XGBoost-Boruta algorithm combination in KOA diagnosis, giving a high level of accuracy and reliability. Furthermore, the application of SHAP revealed that the most important features for identifying patients with KOA were related to the history of knee joint pain or stiffness, knee flexion and extension angles, and age.

Model performance

Data sources used in current ML approaches for diagnosing KOA can be broadly categorized into three types: proteomic analysis, imaging data, and clinical/demographic data [16]. However, research on diagnostic models for KOA based on clinical or demographic sociological data remains limited. In 2019, Lim et al. reported on 5749 participants from the 2015–2016 Korea National Health and Nutrition Examination Survey (KNHANES) [55]. Using participant information on medical utilization and health behavior, such as gender, age, household income, marital status, smoking status, drinking status, BMI, physical activity, and chronic diseases, a principal component analysis (PCA) with quantile transformation scaling was performed to generate characteristics from the patient's medical records for OA identification. The results showed an AUC of up to 76.8% using deep neural network and scaled PCA. However, the KNHANES cohort was not designed for KOA, and the outcome measure was limited to OA in general, which restricts the applicability of this study to KOA specifically.

In the same year, Abedin et al. used data on signs, symptoms, and medication evaluation of both knees from patients, combined with Elastic Net (EN) and RF to build a KOA prediction model, and incorporated convolutional neural network (CNN) trained by X-ray images [56]. The consistency of the outcome predictions across the three models was tested, with results showing root mean square errors (RMSE) of 0.77, 0.97, and 0.94 for the CNN, EN, and RF models, respectively. While this study highlighted the potential of EN and RF to develop an efficient KOA diagnostic model based on clinical and sociodemographic characteristics, the model's performance was poor at higher KL-grades due to severe imbalance in patient KL-grades within the dataset.

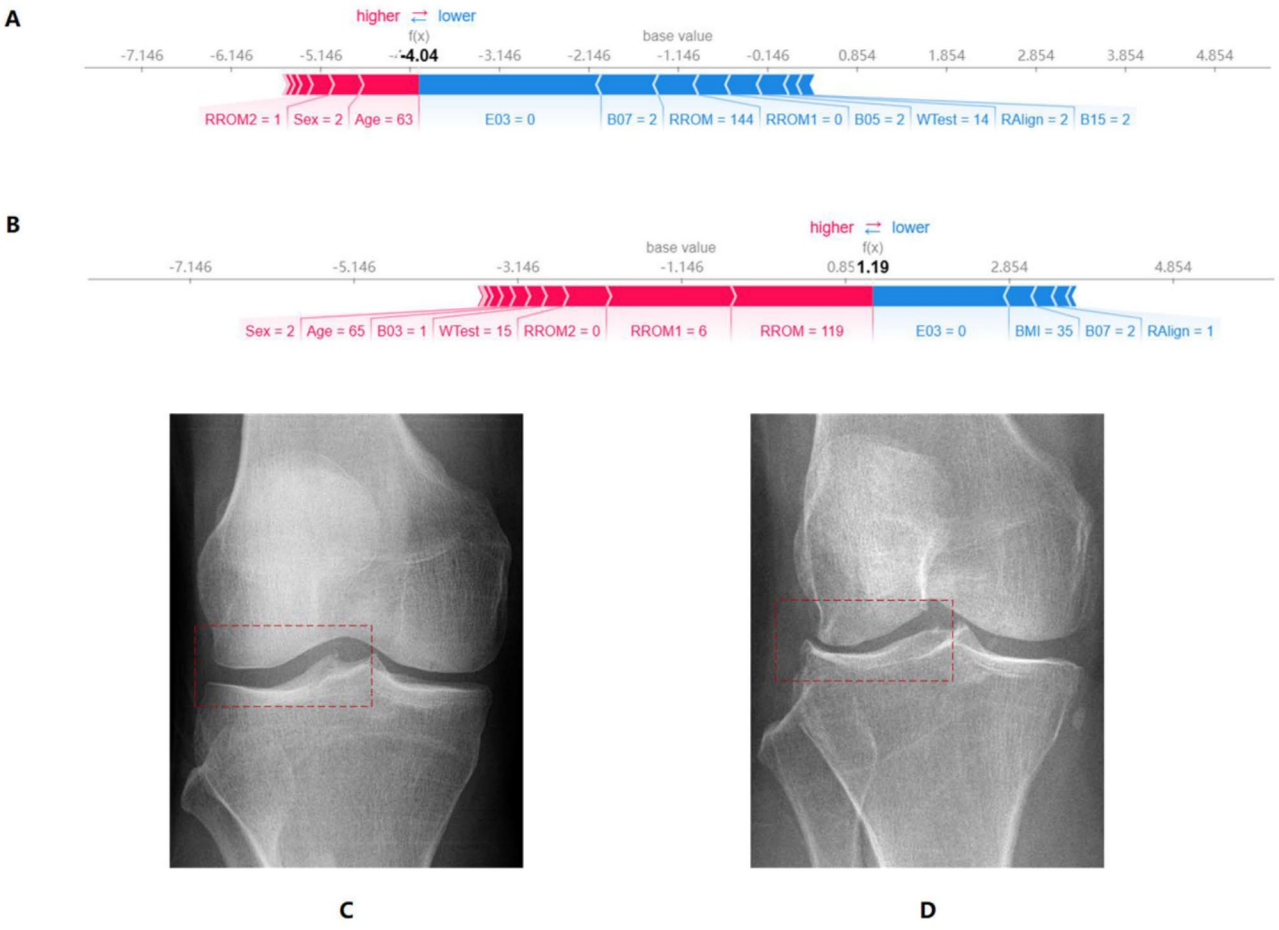

Fig. 4 SHAP force plot. (**A**) SHAP force plot for un-KOA participant (No. 10264). (**B**) SHAP force plot for KOA participant (No. 10031). (**C**) X-ray imaging of the right knee joint of non-KOA participant (No. 10264). (**D**) X-ray imaging of the right knee joint of KOA participant (No. 10031). RROM: knee flexion angle; RROM1: knee extension angle; RROM2: knee excessive flexion or extension; WTest: timing a 50-foot walk; RAlign: right femur-tibia angle; BMI: body mass index. E03: In the past 12 months, have you ever had joint pain, stiffness, or soreness that lasted at least 1 month? B15: Stand up from a straight-back seat without armrests. B07: Bend over, squat or kneel? B05: You don't need to rest when you walk to the first floor? B03: Walk two miles? B15: Stand up from a straight-back seat without armrests

The data in our study came from the Wuchuan County KOA cohort study in Inner Mongolia, China. Compared with other studies, the characteristics included in our study were easier to obtain, as they primarily involved simple data such as the participants' daily physical activity levels and disease history. Our approach of combining easily accessible variables, such as clinical data or sociodemographic characteristics, with ML algorithms to create disease diagnosis models could serve as a valuable tool for both patients and medical practitioners in prescreening for KOA, and help to reduce medical costs and time for patients.

Our study findings aligned with previous studies on the topic area. The combination of the XGBoost classifier and Boruta has been widely used across various fields, including in medicine, computer science, molecular biology, and economics, achieving better overall model performance compared to other models [57–62]. In the medical field, this approach has been used not only for disease diagnosis and prediction, but also for the identification of disease biomarkers and risk prediction [58, 60]. For example, Lslam et al. proposed a feature selection method combining Boruta and LASSO to identify common predictors of diabetic retinopathy [63]. The identified predictors were used to construct artificial neural network (ANN), SVM, RF and XGBoost models to predict diabetic retinopathy. The results showed that the combination of XGBoost classifier with Boruta and LASSO outperformed other models, with an accuracy of 90.01%, precision of 91.80%, and AUC of 0.850.

Similarly, Yue et al. applied the Boruta algorithm for feature selection and compared the performance of LR, k-nearest neighbors (KNN), SVM, decision tree, random forest, XGBoost and ANN, for constructing a prediction model for acute kidney injury [61]. Among all models, the XGBoost model showed the best prediction performance

in discrimination, calibration and clinical application. This model demonstrated strong potential to assist clinicians in identifying high-risk patients and implementing early intervention to reduce mortality. There are several reasons underlying the optimal performance observed using the combination of XGBoost with the Boruta algorithm. Firstly, XGBoost is a powerful learning algorithm based on gradient boosting [44, 64], which iteratively trains multiple decision tree models and progressively enhances the model's performance over each iteration. This iterative training process enables XGBoost to better fit the dataset, especially in high-dimensional data settings with complex relationships. Secondly, the Boruta algorithm is a feature selection method that robustly identifies important features related to the target variable [33]. When combined with XGBoost, it conducts an additional filtration step to extract the most informative features. This filtering process helps reduce the risk of overfitting and enhances the model's generalization ability. Additionally, XGBoost and the Boruta algorithm exhibit complementary effects. XGBoost considers interactions between features, while the Boruta algorithm eliminates noise features unrelated to the target variable [61]. This combination allows for more accurate capturing of patterns and rules in the data, collectively leading to improved model performance.

Factors influencing KOA diagnosis

The SHAP algorithm has identified three categories of significant factors that influence the diagnosis of KOA: the history of knee joint pain or stiffness, knee flexion and extension angles, and age.

Joint pain is one of, if not the most common complaints among KOA patients [65]. The pain associated with KOA is typically intermittent and primarily mechanical, occurring more frequently with weight-bearing activities [6]. It may also be accompanied by a sense of joint stiffness. Symptomatic KOA is characterized by knee joint pain, stiffness, and associated physical impairments. Prolonged stress and chronic compression on the knee joint can lead to gradual cartilage damage, exposing the underlying bone and causing pain [12]. As such, pain is a key factor in diagnosing KOA.

Knee flexion and extension angles are also crucial for the diagnosis and assessment of KOA [66], as they reflect the stability and functional status of the knee joint. Under normal circumstances, the knee joint can flex up to approximately 145 degrees and extend beyond 5 to 10 degrees [67]. In KOA patients, the flexion and extension angles may be limited due to cartilage deterioration and structural changes in the joint [68]. Flexion and extension angles are closely related to knee joint stability, cartilage degeneration, and joint function, thus playing an important role in evaluating KOA patients [69]. The alignment of the knee joint and the biomechanical stress on the knee joint may be affected by distal femoral valgus. Research indicates that valgus distal femurs are more common among East Asian populations, particularly in certain Asian countries, compared to Caucasian populations including those in Europe and America [21]. Due to these anatomical differences, the early diagnosis, course of progression, and presentation of KOA in Asian populations may differ from those in European and American populations. KOA diagnostic models developed for European and American populations, which are often based on the morphological and pathological characteristics of Caucasians, may not adequately account for key anatomical differences such as distal femoral valgus that are more prevalent in East Asian populations.

Age is a crucial predictive factor in the diagnosis of KOA, as the prevalence of knee joint disorders is known to increase with age [4]. With ageing, cartilage loses its elasticity and water content, accompanied by loss of synovial fluid volume, as well as ligament and muscle strength [70]. These changes make the joint more susceptible to damage and wear, increasing the risk of KOA.

Study strengths

Our study boasts several strengths. Firstly, we utilized a large sample from the Chinese KOA cohort, ensuring the representativeness of our findings. Secondly, the utilization of questionnaire data, which was readily accessible, allowed for efficient data collection without the need for intricate medical examinations. Additionally, we employed multiple feature selection methods and ML algorithms, enhancing the accuracy of our diagnostic model. The incorporation of the SHAP model further facilitated the interpretability of results, shedding light on the potential mechanisms governing KOA progression.

Limitations

Some limitations are worth noting in the interpretation of the study results. Firstly, the questionnaire used to assess KOA had not undergone formal validation for reliability and validity. Although it was designed by the author YQZ based on the Birmingham OA design and had been widely used in large-scale epidemiological studies in China, the lack of formal validation might limit the ability to fully assess its robustness across different populations. Future studies should consider conducting formal validation to further strengthen the utility and accuracy of the questionnaire. Additionally, due to limitations in data sources, the proposed model had not been externally validated. To address this issue, we employed a five-fold cross-validation technique during the analysis. This approach, along with ensuring that the test set remained undisturbed, helped to assess the stability and reliability of the model's performance.

Conclusion

Overall, we have developed an interpretable XGBoost-Boruta diagnostic model that provides optimal performance in helping to predict KOA. Its interpretability enables physicians to accurately pinpoint risk factors among patients, leading to increased confidence in making KOA diagnosis and improved precision in evaluating patient risks. Consequently, this facilitates the implementation of appropriate interventions to delay disease progression and enhance patient quality of life.

Abbreviations

OA	Osteoarthritis
KOA	Knee Osteoarthritis
JSW	Joint space width
ML	Machine learning
SHAP	SHapley Additive exPlanations
KL-grades	Kellgren and Lawrence grades
SF-12	Short Form-12
LASSO	Least Absolute Shrinkage and Selection Operator
SMOTE	Synthetic Minority Oversampling Technique
RF	Random forest
XGBoost	eXtreme gradient boosting
Adaboost	Adaptive boosting
AUC	Area under the receiver operating characteristic curve
PPV	Positive predictive value
NPV	Negative predictive value
PPV	Positive predictive value
NPV	Negative predictive value

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13075-024-03450-2.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

BW, LJH and ZJF conceived the idea for the review. BW, ZJF, and YK designed, undertook the literature search, and coordinated the study. WZS, YK, YQZ, and JJL gave crucial intellectual input and provided critical revision for the initial protocol. ZJF and SYL contributed to the implementation of the study. ZJF, WZS and YK acquired data, screened records, extracted data, and assessed risk of bias. ZJF coded the statistical analysis, figures, and appendix in collaboration with WZS and YK. LGJ, YQZ, JHL and BW analysed and interpreted the data. ZJF, WXS, and BW wrote the first draft of the manuscript. All authors gave crucial feedback on the revised report and approved the final version of the manuscript. BW obtained funding. ZJF, WZS, YK and BW are the guarantors of this manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding

This study was supported by the National Natural Science Foundation of China (81802204) and by Zhejiang University School of Medicine, The First Affiliated Hospital's Foundation (G2022010-18), Alibaba Cloud, Natural Science Foundation of Zhejiang Province (LTGY23H060007), Zhejiang University The First Affiliated Hospital of Zhejiang University School of Medicine Medical Education Research Key Project (zyjg202404)and Zhejiang Medical and Health Science and Technology Project (2023RC010). The founders had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Data availability

The guarantor (BW) is willing to examine all requests for the full dataset after a period of two years from the date of this publication. The corresponding author should be contacted at BW. wangbin_pku@zju.edu.cn .

Declarations

Ethics approval and consent to participate

Written informed consent was obtained for all study participants, and ethical approval was granted by the Ethics Committee of Peking University People's Hospital, Beijing, China (Approval Number: No. 2012–040).

Consent for publication

All authors agreed to the publication of this manuscript.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Orthopaedic Surgery, The First Affiliated Hospital, Zhejiang University School of Medicine, Qingchun Road No. 79, Hangzhou, China
[2]Department of Health Statistics, School of Public Health, Sun Yat-sen University, Guangzhou, China
[3]Department of Big Data in Health Science School of Public Health, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
[4]Arthritis Clinic & Research Center, Peking University People's Hospital, Beijing, China
[5]School of Computer Science and Technology, Xinjiang University, Urumchi, China
[6]School of Biomedical Engineering, Faculty of Engineering and IT, University of Technology Sydney, Sydney, Australia
[7]Harvard Medical School, Boston Massachusetts, USA

Received: 28 July 2024 / Accepted: 1 December 2024
Published online: 19 December 2024

References

1. Long H, Liu Q, Yin H, Wang K, Diao N, Zhang Y, et al. Prevalence trends of Site-Specific Osteoarthritis from 1990 to 2019: findings from the global burden of Disease Study 2019. Arthritis Rheumatol (Hoboken NJ). 2022;74(7):1172–83.
2. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis. 2014;73(7):1323–30.
3. Turkiewicz A, Petersson IF, Björk J, Hawker G, Dahlberg LE, Lohmander LS, et al. Current and future impact of osteoarthritis on health care: a population-based study with projections to year 2032. Osteoarthr Cartil. 2014;22(11):1826–32.
4. Sharma L. Osteoarthritis of the knee. N Engl J Med. 2021;384(1):51–9.
5. Hunter DJ, Bierma-Zeinstra S, Osteoarthritis. Lancet (London England). 2019;393(10182):1745–59.
6. Lespasio MJ, Piuzzi NS, Husni ME, Muschler GF, Guarino A, Mont MA. Knee osteoarthritis: a primer. Permanente J. 2017;21:16–183.
7. Li D, Li S, Chen Q, Xie X. The prevalence of symptomatic knee osteoarthritis in relation to Age, Sex, Area, Region, and body Mass Index in China: a systematic review and Meta-analysis. Front Med. 2020;7:304.
8. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, hip, and Knee. Volume 72. Hoboken, NJ: Arthritis & rheumatology; 2020. pp. 220–33. 2.
9. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, hip, and Knee. Arthritis Care Res. 2020;72(2):149–62.
10. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957;16(4):494–502.
11. Yázigi F, Carnide F, Espanha M, Sousa M. Development of the knee OA pre-screening questionnaire. Int J Rheum Dis. 2016;19(6):567–76.

12. Braun HJ, Gold GE. Diagnosis of osteoarthritis: imaging. Bone. 2012;51(2):278–88.
13. Harrell FE Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361–87.
14. Ahmed U, Anwar A, Savage RS, Thornalley PJ, Rabbani N. Protein oxidation, nitration and glycation biomarkers for early-stage diagnosis of osteoarthritis of the knee and typing and progression of arthritic disease. Arthritis Res Therapy. 2016;18(1):250.
15. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31–8.
16. Binvignat M, Pedoia V, Butte AJ, Louati K, Klatzmann D, Berenbaum F et al. Use of machine learning in osteoarthritis research: a systematic literature review. RMD open. 2022;8(1).
17. Nich C, Behr J, Crenn V, Normand N, Mouchère H, d'Assignies G. Applications of artificial intelligence and machine learning for the hip and knee surgeon: current state and implications for the future. Int Orthop. 2022;46(5):937–44.
18. Callahan LF, Cleveland RJ, Allen KD, Golightly Y, Racial/Ethnic. Socioeconomic, and Geographic Disparities in the epidemiology of knee and hip osteoarthritis. Rheum Dis Clin North Am. 2021;47(1):1–20.
19. Yoshida S, Aoyagi K, Felson DT, Aliabadi P, Shindo H, Takemoto T. Comparison of the prevalence of radiographic osteoarthritis of the knee and hand between Japan and the United States. J Rhuematol. 2002;29(7):1454–8.
20. Zhang Y, Xu L, Nevitt MC, Aliabadi P, Yu W, Qin M, et al. Comparison of the prevalence of knee osteoarthritis between the elderly Chinese population in Beijing and whites in the United States: the Beijing Osteoarthritis Study. Arthritis Rheum. 2001;44(9):2065–71.
21. Harvey WF, Niu J, Zhang Y, McCree PI, Felson DT, Nevitt M, et al. Knee alignment differences between Chinese and caucasian subjects without osteoarthritis. Ann Rheum Dis. 2008;67(11):1524–8.
22. Kang X, Fransen M, Zhang Y, Li H, Ke Y, Lu M, et al. The high prevalence of knee osteoarthritis in a rural Chinese population: the Wuchuan osteoarthritis study. Arthritis Rheum. 2009;61(5):641–7.
23. Lin J, Fransen M, Kang X, Li H, Ke Y, Wang Z, et al. Marked disability and high use of nonsteroidal antiinflammatory drugs associated with knee osteoarthritis in rural China: a cross-sectional population-based survey. Arthritis Res Therapy. 2010;12(6):R225.
24. Lin J, Li R, Kang X, Li H. Risk factors for radiographic tibiofemoral knee osteoarthritis: the wuchuan osteoarthritis study. Int J Rheumatol. 2010;2010:385826.
25. Wang B, Liu Q, Wise BL, Ke Y, Xing D, Xu Y, et al. Valgus malalignment and prevalence of lateral compartmental radiographic knee osteoarthritis (OA): the Wuchuan OA study. Int J Rheum Dis. 2018;21(7):1385–90.
26. Liu Q, Niu J, Huang J, Ke Y, Tang X, Wu X, et al. Knee osteoarthritis and all-cause mortality: the Wuchuan Osteoarthritis Study. Osteoarthr Cartil. 2015;23(7):1154–7.
27. Liu Q, Niu J, Li H, Ke Y, Li R, Zhang Y, et al. Knee symptomatic osteoarthritis, walking disability, NSAIDs use and all-cause Mortality: Population-based Wuchuan Osteoarthritis Study. Sci Rep. 2017;7(1):3309.
28. Gandhi SK, Salmon JW, Zhao SZ, Lambert BL, Gore PR, Conrad K. Psychometric evaluation of the 12-item short-form health survey (SF-12) in osteoarthritis and rheumatoid arthritis clinical trials. Clin Ther. 2001;23(7):1080–98.
29. Lin JH, Kang XZ, Fransen M, Li H, Ke Y, Wang ZQ, et al. Disability and common treatment strategies associated with knee pain in a rural Chinese population. Zhonghua Yi Xue Za Zhi. 2010;90(21):1477–81.
30. Jiang L, Rong J, Zhang Q, Hu F, Zhang S, Li X, et al. Prevalence and associated factors of knee osteoarthritis in a community-based population in Heilongjiang, Northeast China. Rheumatol Int. 2012;32(5):1189–95.
31. García-Domínguez A, Galván-Tejada CE, Magallanes-Quintanar R, Gamboa-Rosales H, Curiel IG, Peralta-Romero J, et al. Diabetes detection models in Mexican patients by combining machine learning algorithms and feature selection techniques for clinical and paraclinical attributes: a comparative evaluation. J Diabetes Res. 2023;2023:9713905.
32. Tang C, Bian M, Liu X, Li M, Zhou H, Wang P, et al. Unsupervised feature selection via latent representation learning and manifold regularization. Neural Networks: Official J Int Neural Netw Soc. 2019;117:163–78.
33. Kursa MB, Rudnicki WRJJSS. Feature Selection Boruta Package. 2010;36(11):1–13.
34. Tibshirani RJJRSS, Series B. Regres Shrinkage Selection via Lasso. 1996;58(1).
35. Zou AHJPAS. Adapt Lasso Its Oracle Prop. 2006;101(476):1418–29.
36. Karabis A, Nikolakopoulos S, Pandhi S, Papadimitropoulou K, Nixon R, Chaves RL, et al. High correlation of VAS pain scores after 2 and 6 weeks of treatment with VAS pain scores at 12 weeks in randomised controlled trials in rheumatoid arthritis and osteoarthritis: meta-analysis and implications. Arthritis Res Therapy. 2016;18:73.
37. Lin W, Xie X, Luo Z, Chen X, Cao H, Fang X, et al. Early identification of macrophage activation syndrome secondary to systemic lupus erythematosus with machine learning. Arthritis Res Therapy. 2024;26(1):92.
38. Alhamzawi R, Ali HTM. The bayesian adaptive lasso regression. Math Biosci. 2018;303:75–82.
39. Li J, Cui J, Wu L, Liu Y-b, Wang Q. Machine learning and molecular subtype analyses provide insights into PANoptosis-associated genes in rheumatoid arthritis. Arthritis Res Therapy. 2023;25(1):233.
40. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WPJAAF. SMOTE: synthetic minority over-sampling technique. 2002(1).
41. Ukalovic D, Leeb BF, Rintelen B, Eichbauer-Sturm G, Spellitz P, Puchner R, et al. Prediction of ineffectiveness of biological drugs using machine learning and explainable AI methods: data from the Austrian Biological Registry BioReg. Arthritis Res Therapy. 2024;26(1):44.
42. Quinlan JRJDG. System. GDotiLe, in. Induction of decision trees Machine Learning. 1986.
43. LEARN BJM. Random forests. 2001. 2001;45(1):5–32.
44. Chen T, Guestrin CJA, XGBoost:. A Scalable Tree Boosting System. 2016.
45. Freund YJMK. Experiment With a New Boosting Algorithm. 1996.
46. Clift AK, Dodwell D, Lord S, Petrou S, Brady M, Collins GS, et al. Development and internal-external validation of statistical and machine learning models for breast cancer prognostication: cohort study. BMJ (Clinical Res ed). 2023;381:e073800.
47. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. Crit Care (London England). 2019;23(1):112.
48. Huang X, Yu Z, Wei X, Shi J, Wang Y, Wang Z, et al. Prediction of Vancomycin dose on high-dimensional data using machine learning techniques. Expert Rev Clin Pharmacol. 2021;14(6):761–71.
49. Zhang X, Yan C, Gao C, Malin BA, Chen Y. Predicting Missing values in Medical Data via XGBoost Regression. J Healthc Inf Res. 2020;4(4):383–94.
50. Lundberg S, Lee S-I. Consistent feature attribution for tree ensembles. arXiv [Preprint]. 2017 Jun 16 [cited 2024 Dec 4]. Available from: https://arxiv.org/abs/1706.06060
51. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Sci Rep. 2021;11(1):6968.
52. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell. 2020;2(1):56-67.
53. Orsini N, Moore A, Wolk A. Interaction Analysis based on Shapley Values and Extreme Gradient Boosting: a realistic Simulation and Application to a large epidemiological prospective study. Front Nutr. 2022;9:871768.
54. Lundberg S, Lee SI, editors. A Unified Approach to interpreting model predictions. Nips; 2017.
55. Lim J, Kim J, Cheon S. A deep neural network-based method for early detection of Osteoarthritis using Statistical Data. Int J Environ Res Public Health. 2019;16(7).
56. Abedin J, Antony J, McGuinness K, Moran K, O'Connor NE, Rebholz-Schuhmann D, et al. Predicting knee osteoarthritis severity: comparative modeling based on patient's data and plain X-ray images. Sci Rep. 2019;9(1):5761.
57. Song C, Jiang ZQ, Hu LF, Li WH, Liu XL, Wang YY, et al. A machine learning-based diagnostic model for children with autism spectrum disorders complicated with intellectual disability. Front Psychiatry. 2022;13:993077.
58. Song W, Wu F, Yan Y, Li Y, Wang Q, Hu X, et al. Gut microbiota landscape and potential biomarker identification in female patients with systemic lupus erythematosus using machine learning. Front Cell Infect Microbiol. 2023;13:1289124.
59. Vamvakas A, Tsivaka D, Logothetis A, Vassiou K, Tsougos I. Breast Cancer classification on multiparametric MRI - increased performance of boosting ensemble methods. Technol Cancer Res Treat. 2022;21:15330338221087828.
60. Yang J, Peng H, Luo Y, Zhu T, Xie L. Explainable ensemble machine learning model for prediction of 28-day mortality risk in patients with sepsis-associated acute kidney injury. Front Med. 2023;10:1165129.
61. Yue S, Li S, Huang X, Liu J, Hou X, Zhao Y, et al. Machine learning for the prediction of acute kidney injury in patients with sepsis. J Translational Med. 2022;20(1):215.
62. Zhang S, Khattak A, Matara CM, Hussain A, Farooq A. Hybrid feature selection-based machine learning classification system for the prediction

of injury severity in single and multiple-vehicle accidents. PLoS ONE. 2022;17(2):e0262941.
63. Islam MM, Rahman MJ, Rabby MS, Alam MJ, Pollob S, Ahmed N, et al. Predicting the risk of diabetic retinopathy using explainable machine learning algorithms. Diabetes Metabolic Syndrome. 2023;17(12):102919.
64. Nwanosike EM, Conway BR, Merchant HA, Hasan SS. Potential applications and performance of machine learning techniques and algorithms in clinical practice: a systematic review. Int J Med Informatics. 2022;159:104679.
65. Katz JN, Arant KR, Loeser RF. Diagnosis and treatment of hip and knee osteoarthritis: a review. JAMA. 2021;325(6):568–78.
66. Marriott KA, Birmingham TB. Fundamentals of osteoarthritis. Rehabilitation: Exercise, diet, biomechanics, and physical therapist-delivered interventions. Osteoarthr Cartil. 2023;31(10):1312–26.
67. Mmacp SSMJP. Measurement of Joint Motion: A guide to goniometry. 1996;82(4):278-.
68. Richards R, van den Noort JC, Dekker J, Harlaar J. Gait Retraining with Real-Time Biofeedback to reduce knee adduction moment: systematic review of effects and methods used. Arch Phys Med Rehabil. 2017;98(1):137–50.
69. Booij MJ, Richards R, Harlaar J, van den Noort JC. Effect of walking with a modified gait on activation patterns of the knee spanning muscles in people with medial knee osteoarthritis. Knee. 2020;27(1):198–206.
70. Dong Y, Yan Y, Zhou J, Zhou Q, Wei H. Evidence on risk factors for knee osteoarthritis in middle-older aged: a systematic review and meta analysis. J Orthop Surg Res. 2023;18(1):634.

Publisher's note