Tian *et al.*
European Journal of Medical Research (2025) 30:317
https://doi.org/10.1186/s40001-025-02545-z

European Journal of Medical Research

REVIEW

Open Access

Accuracy of machine learning in identifying candidates for total knee arthroplasty (TKA) surgery: a systematic review and meta-analysis

Cong Tian[1†], Haifeng Chen[1†], Wenhui Shao[2], Ruikun Zhang[1], Xinmiao Yao[3*] and Jianlong Shu[4]

Abstract

Background The application of machine learning (ML) in predicting the requirement for total knee arthroplasty (TKA) at knee osteoarthritis (KOA) patients has been acknowledged. Nonetheless, the variables employed in the development of ML models are diverse and these different approaches yield inconsistent predictive performance of models. Therefore, we conducted this systematic review and meta-analysis to explore the feasibility of ML in identifying candidates for TKA.

Method This study was conducted based on the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines. This study was registered on the international prospective register of systematic reviews registration database website, PROSPERO, with a unique ID: CRD 42023443948. The study subjects were patients diagnosed with KOA. Relevant studies were searched through PubMed, Web of Science, Cochrane, and Embase until September 15, 2024. The c-index was used as the outcome measure. The risk of bias in the primary study was assessed by Prediction model Risk of Bias Assessment Tool (PROBAST). Random or fixed effects were used for the meta-analysis.

Results A total of 13 articles were included in this study, but only 11 articles with 25 models were eligible for the meta-analysis. ML models in the included studies were classified based on the source of variables, including clinical features, radiomics, and the combination of clinical features and radiomics. In the training set, the c-index was 0.713 (0.628 – 0.799) for clinical features, 0.841 (0.777 – 0.904) for radiomics, and 0.844 (0.815 – 0.873) for the combination of clinical features and radiomics. In the validation set, the c-index for ML models based on clinical features, radiomics, and the combination of clinical features and radiomics was 0.656 (0.526 – 0.786), 0.861 (0.806 – 0.916), and 0.831 (0.799 – 0.863), respectively.

Conclusion The results of this meta-analysis highlighted that the ML model is feasible in identifying candidates for TKA. X-ray-based ML models exhibit the best predictive performance among the models. However, there is currently a lack of high-level research available for clinical application. Furthermore, the accuracy of ML models in identifying candidates for TKA is significantly limited by the quality of modeling parameters and database architecture. Therefore, constructing a more targeted and professional database is imperative to promote the development and clinical application of ML models.

Keywords Machine learning, Meta-analysis, Systematic review, Total knee arthroplasty

[†]Cong Tian and Haifeng Chen have contributed equally to this work.

*Correspondence:
Xinmiao Yao
yxmzcmu@163.com
Full list of author information is available at the end of the article

Introduction

Knee osteoarthritis (KOA) is a chronic joint disease characterized by aseptic synovitis, cartilage degeneration, and osteosclerosis, which cause knee pain and restrict individual mobility [1]. KOA is one of the leading causes of disability in middle-aged and older patients [2, 3]. Epidemiological statistics show that KOA accounts for approximately 85% of all osteoarthritis cases worldwide [4]. The incidence of KOA is approximately 40% in men and 47% in women, respectively [5]. As per the guidelines of the International Association of Osteoarthritis, standard treatment is preferred to treat KOA. However, at present, no treatment method can suppress the progression of KOA.

For KOA patients who do not respond to standard treatment, total knee arthroplasty (TKA) is the ultimate treatment to alleviate pain and improve knee joint function, thus enhancing the quality of the patient's life [2, 6]. It is projected that there will be 3 million TKA cases worldwide in the next decade [7]. Despite substantial advancements in surgical techniques, prosthesis design, and materials for TKA, about 12.7% of patients have dissatisfactory outcomes and demand for revision TKA within 5 years after TKA surgery [8]. The common reasons for revision are infection (36.1%), aseptic loosening of the prosthesis (21.9%), and periprosthetic fractures (13.7%) [9]. Therefore, the prevention of TKA events should also be addressed. Moreover, it is essential to develop an effective tool to identify high-risk KOA patients who require TKA surgery. TKA cases can be delayed or even prevented by actively implementing effective interventions such as health education [10], physical therapy [11, 12], prescription medications [13], and adjusting lower limb alignment [14]. However, there is currently a lack of such tools in clinical practice.

In recent years, the application of artificial intelligence in the field of health care has garnered increasing attention from scholars. Machine learning (ML), as a unique application in the field of artificial intelligence, presents a promising avenue for data analysis. Furthermore, ML enables computer systems to learn, predict, or make informed decisions through data and pattern recognition [15, 16]. ML has been applied in various clinical contexts, including the prediction of lymph node metastasis of rectal cancer, colorectal cancer [17], and gestational diabetes mellitus [18]. Various investigations have also devised multiple ML models to stratify the risk of TKA complications and identify patients in need of TKA [19, 20].

The variables employed in the development of ML models are diverse and can be obtained from clinical features, imaging examinations [X-rays, magnetic resonance imaging (MRI) or ultrasound (US)], or a combination of both sources. These different approaches yield inconsistent predictive performance of models. Therefore, we conducted this systematic review and meta-analysis to explore the feasibility of ML in identifying candidates for TKA, providing a reference basis for formulating effective preventive measures for TKA complications and evidence-based support for developing risk prediction tools in the future.

Methods

Study registration

This study was conducted based on the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines [21]. This study was registered on the international prospective register of systematic reviews registration database website, PROSPERO, with a unique ID: CRD 42023443948.

Eligibility criteria

The studies were selected according to the following inclusion criteria:

(1) The study subjects were patients diagnosed with KOA;
(2) A risk prediction model for TKA was completely constructed;
(3) Studies without externally validated risk prediction models;
(4) Different ML studies published on the same dataset;
(5) Studies reported in English.

The following studies were excluded:

(1) Meta-analyses, reviews, guidelines, expert opinions, etc.;
(2) Only a differential factor analysis was conducted, without constructing a complete machine learning model;
(3) The following outcome indicators for predicting the accuracy of machine learning models were lacking: receiver operating characteristic (ROC), c-statistic, c-index, sensitivity, specificity, accuracy, recall, precision, confusion matrix, diagnostic four-grid table, F1 score, and calibration curve;
(4) Studies with a small sample size (< 30 cases).

Data sources and search strategy

PubMed, Web of Science, Cochrane, and Embase databases were searched for original studies published before September 15th, 2024. We used a combination of subject terms and free words to collect relevant studies, with search items including 'total knee arthroplasty', 'machine

learning', 'risk model', and 'prediction model'. We also searched for potential relevant references in the included studies. The retrieval strategy is provided in Additional File 1.

Study selection and data extraction

All retrieved studies were imported into the EndNote 20 software (Thomson ResearchSoft, USA). We used both automatic markings provided by EndNote and manual screening to exclude duplicate studies. Then, irrelevant studies were eliminated based on title and abstract screening. Next, all qualified studies entered full-text screening, where two researchers (Shao and Liu) read all the texts and evaluated all studies based on the criteria. Finally, all the selected studies were cross-checked. In case of disagreements, a third researcher assisted in the decision-making process. The detailed study selection process is presented in Fig. 1.

We also extracted key information from each eligible study, as follows:

(1) Characteristics of the study: the title, first author, publication year, author's country, and study type.

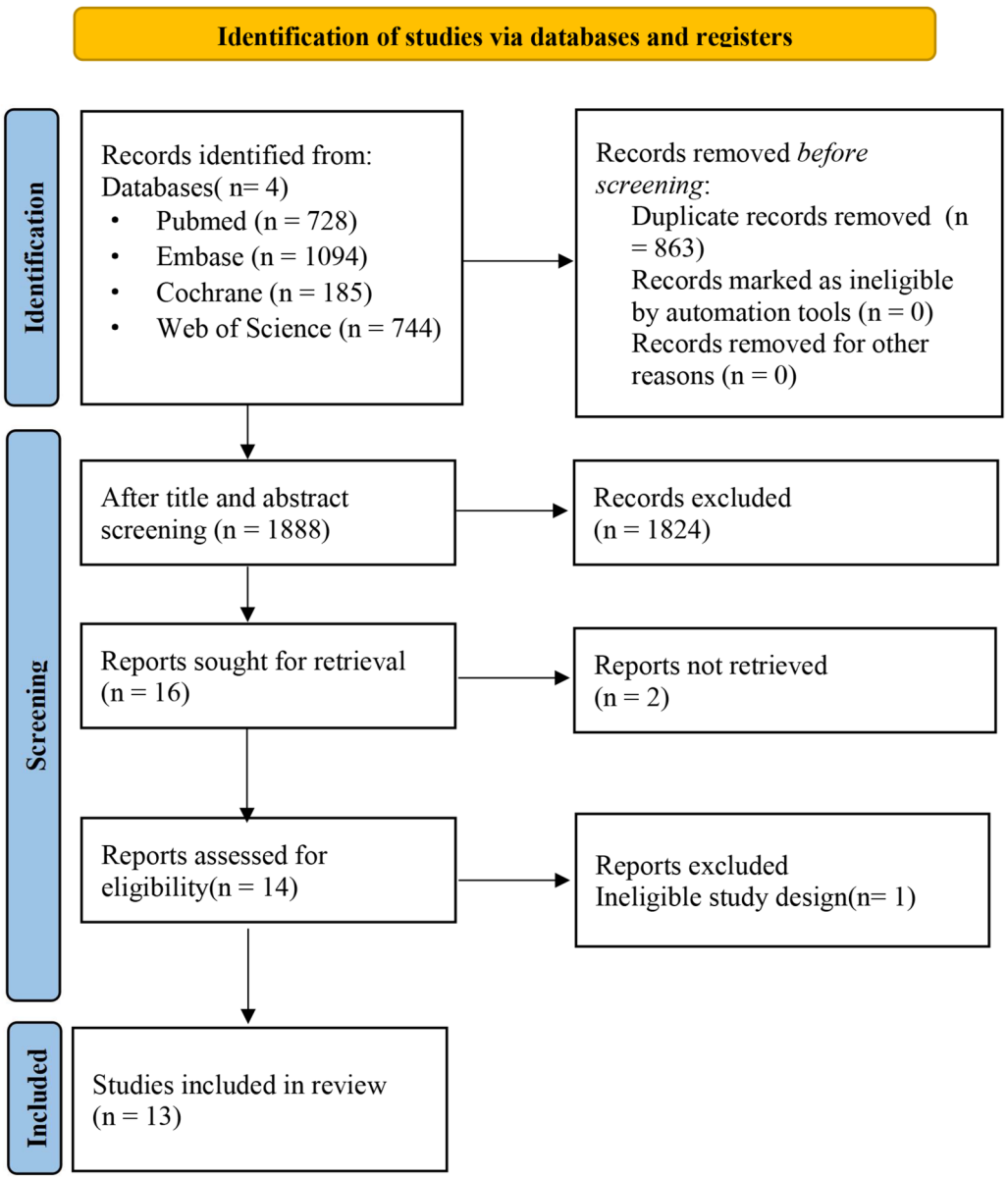

Fig. 1 PRISMA flowchart detailing the systematic search process

(2) Characteristics of cohorts: patient source, disease background, number of knee replacement cases, total case count, number of knee replacement cases in the training set, total case count in the training set, number of knee replacement cases in the validation set, and number of cases in the validation set.
(3) Method of generating the validation set, overfitting methods, missing data handling method, variable selection method, type of model used, and modeling variables.
(4) Prediction parameters: ROC, c-statistic, c-index, sensitivity, specificity, accuracy, recall, precision, confusion matrix, diagnostic four-grid table, F1 score, and calibration curve.

Risk of bias in studies

The risk of bias in the primary study was assessed by Prediction model Risk of Bias Assessment Tool (PROBAST) [22]. PROBAST provides insights into both overall bias risk and overall applicability. This assessment involved a set of questions across four domains: participants, predictive variables, outcome, and statistical analysis. Each domain included 2, 3, 6, and 9 specific questions, respectively, with three possible answers for each question (yes/probably yes, no/probably no, and no information). If a domain contained at least one question answered as yes/probably yes, it was categorized as "high risk". To be classified as low risk, a domain required all questions to have responses of yes/probably yes. When all domains were deemed to be at low risk, the overall bias risk was rated as low. Conversely, if at least one domain was classified as high risk, the overall bias risk was considered high. Two researchers independently conducted the bias risk assessment using PROBAST, followed by cross-checking upon completion. In case of disagreements, a third researcher assisted in the decision-making process.

Outcomes

The outcome is the c-index which measures the overall accuracy of models. The c-index ranges from 0.5 to 1. A value of 0.5 suggests complete randomness with no predictive value, whereas a value of 1 indicates perfect agreement between the predicted results and actual outcomes [23]. C-index greater than 0.7 suggests that the model has relatively ideal accuracy.

Synthesis methods

We meta-analyzed the c-index of machine learning models. If the 95% confidence interval and standard error of the c-index are missing, we would estimate its standard error by referring to the study by Debray et al. [24]. Considering the variability in included variables and inconsistent parameters among different machine learning models, a random-effects model (DerSimonian and Laird method) was preferred to conduct the meta-analysis of the c-index. The meta-analysis was implemented using R version 4.2.0 (R development Core Team, Vienna, http://www.R-project.org).

Results

Study selection

2,751 articles were initially retrieved, of which 863 duplicates (769 articles by automatic tools and 94 articles by manual search were excluded). Then, 1824 articles were excluded after reviewing their abstracts and titles. The remaining 14 articles underwent full-text review. Finally, 13 articles [19, 20, 25–35] were included in the study. 11 articles with 25 models were selected for meta-analysis [19, 25, 27–35].

Study characteristics

The 13 articles included were published between 2019 and 2024, with a follow-up period of 5–10 years. Regarding the study location, four studies were from the USA [19, 25–27], one study was from Austria [20], two studies were from the UK [28, 29], one study was from Australia [30], one study was from France [31], one study was from Canada [32], one study was from Finland [33], and two studies were from China[34, 35]. In terms of data source, six studies were based on the Osteoarthritis Initiative Cohort [20, 25, 26, 31, 32, 34], two studies were based on the Osteoarthritis Initiative Cohort and the Multicenter Osteoarthritis Study databases [27, 29], one study was based on the Clinical Practice Research Datalink[28], one study was based on the MedicineInsight data set [30], one study was based on the Musculoskeletal Pain in Ullensaker study MUST [33], and two studies were based on radiographic and surgical data obtained from a high-volume joint replacement practice[19, 35]. Furthermore, two articles [20, 26] did not provide outcome measures, and thus 11 articles with 25 models were included in the quantitative meta-analysis. Among these modeling variables, clinical features were used in four models [27, 28, 30, 34], radiomics was used in five models [19, 27, 31, 34], and 16 models used both clinical features and radiomics [25, 27, 29, 32–34]. In radiomics, 10 models were based on X-ray (47.6%), four models used MRI (19.0%), six models used the combination of MRI and X-ray (28.6%), and one model used the combination of US and X-ray (4.8%). The study characteristics are presented in Table 1. The features of each model and the data extracted from the models are provided in Additional File 2.

Table 1 Characteristics of included studies

No.	Study (years of publication)	Country	Study design	Follow time	Source of data	The quantity of TKA	Total quantity	Total quantity of training sets	Method of generating validation set	Total quantity of validation sets	Methods for variable screening/ feature selection	Types of models
1	Sharmala Thuraisingam [30]	Australia	Cohort	5y	MedicineInsight	15979(P)	201462(P)	201462(P)	Bootstrap	NA	Literature review, Delphi process	Fine–Gray
2	Ahmad Almhdie-Imjabbar [31]	France	Cohort	9y	OAI	375(P) /291(P)	4382(P)/4296(P)	4382(P)/4296(P)	Cross-validation	NA	Literature review	LR
3	Kevin Leung [25]	USA	Cohort	9y	OAI	364(P)	728(P)	728(P)	Cross-validation	NA	Univariable and multivariable analysis	DL
4	David J. Houserman [19]	USA	Case–control	NA	A high-volume joint replacement practice	NA	NA	NA	Random sampling	812(P)	NA	DL
5	Afshin Jamshidi [32]	Canada	Cohort	9y	OAI	413(K)	7589(K)	6071(K)	Random sampling	1518(K)	Lasso's Cox	Cox, DL, RF, SVM, LR
6	Aniket A. Tolpadi [26]	USA	Cohort	5y	OAI	NA	35482(I)	23126(I)	Random sampling	5241(I)	Literature review, multivariable analysis	DL
7	Aleksei Tiulpin [32]	Finland	Cohort	7y	MUST	30(P)	557(P)	557(P)	Cross-validation	NA	Literature review, univariable, and multivariable analysis	LR
8	Dahai Yu [28]	UK	Cohort	10y	CPRD and Multicenter	NA	NA	416030(P)	Cross-validation	NA	Literature review, Record-Wide Association Study, and panel consensus	Cox
9	Khadija Mahmoud [29]	UK	Cohort	5y	OAI and MOST	297(P)	6291(P)	3234(P)	Internal validation: random sampling External validation: MOST	3057(P)	Literature review, expert knowledge	LR, LASSO, Ridge, DT, RF, GBM
10	Stephan Heisinger [20]	Austria	Cohort	4y	OAI	165(P)	165(P)	83(P)	Random sampling	82(P)	Variables of interest were publicly available	ANN

Table 1 (continued)

No.	Study (years of publication)	Country	Study design	Follow time	Source of data	The quantity of TKA	Total quantity	Total quantity of training sets	Method of generating validation set	Total quantity of validation sets	Methods for variable screening/ feature selection	Types of models
11	Haresh Rengaraj Rajamohan [27]	USA	Cohort	9y	OAI and MOST	650(P)	5307(P)	706(P)	Cross-validation	4601(P)	Univariable and multivariable analysis	ANN, DL, Ensemble model
12	Yang Li [34]	China	Cohort	11y	OAI	273(P)	4796(P)	188(P)	Random sampling	81(P)	NA	LASSO
13	Hongzhi Liu [35]	China	Cohort	5y	A high-volume joint replacement practice	537(K)	1779(K)	1156(K)	Random sampling	623(K)	Univariate cox, multivariate cox, and Lasso's Cox	DL

USA United States of America, *UK* United Kingdom, *NA* not available, *OAI*, Osteoarthritis Initiative: A Knee Health Study, *MUST* Musculoskeletal pain in Ullensaker Study, *CPRD* Clinical Practice Research Datalink, *MOST* Multicenter Osteoarthritis Study, *P* patient, *K* knee, *I* image, *Fine–Gray* competing risk regression model, *LR* logistic regression model, *DL* deep learning model, *Cox* Cox proportional-hazards model, *SVM* support vector machine model, *RF* random forest model, *DT* decision tree model, *GBM* gradient boosting machine model *ANN*, artificial neural network model

Risk of bias in studies

In terms of predictive factors, all studies were considered to have a low risk of bias. Regarding the risk of bias, two models [19, 28] were rated unclear bias risk due to the lack of reported information on inclusion and exclusion criteria, whereas other models were rated as low risk of bias. In the results section, two models were considered unclear bias risk [19, 28, 30] due to the lack of assessment of predictive factors and the time interval for result determination, whereas other models were considered to have a low bias risk. Additionally, one model [30] did not include an independent validation set, four models [25, 31, 33] did not provide sufficient information to determine the appropriate method for handling missing data, and five models [32] did not conduct internal validation. Therefore, these studies were classified as having a high risk of bias. Five models [27] did not have enough information to determine the appropriate method for handling missing data, and six models [29] did not report information on internal validation. These studies were thus classified as having an unclear risk of bias. One model [28, 34] was judged as low risk of bias. Figure 2 gives an overall summary of PROBAST risk of bias across all included studies. The detailed assessment of each question in four domains of bias risk is presented in Additional File 3.

Synthesized results

ML models in the included studies were classified based on the source of variables, including clinical features, radiomics, and the combination of clinical features and radiomics. In the training set, the c-index for each source of variables was 0.713 (0.628 – 0.799) for clinical features, 0.841 (0.777 – 0.904) for radiomics, and 0.844 (0.815 – 0.873) for the combination of clinical features and radiomics. In particular, the radiomic variables included two categories: X-ray and MRI. The c-index value was 0.895 (0.865 – 0.924) for models based on X-ray alone and 0.755 (0.508 – 1.000) for models based on MRI alone. ML models based on the combination of clinical features and radiomics were categorized into four subgroups: clinical features + X-ray, clinical features + MRI, clinical feature + MRI + X-ray, and clinical features + US + X-ray. The c-index value for each subgroup was 0.867 (0.824–0.909), 0.772 (0.539 ~ 1.000), 0.842 (0.816–0.867), and 0.820 (0.725–0.915), respectively. In the validation set, the c-index for ML models based on clinical features, radiomics, and the combination of clinical features and radiomics was 0.656 (0.526 – 0.786), 0.861 (0.806 – 0.916), and 0.831 (0.799 – 0.863), respectively. The radiomic variables included two categories: X-ray and MRI. The c-index value was 0.882 (0.825 ~ 0.939) for models based on X-ray alone and 0.725 (0.499 – 0.950)

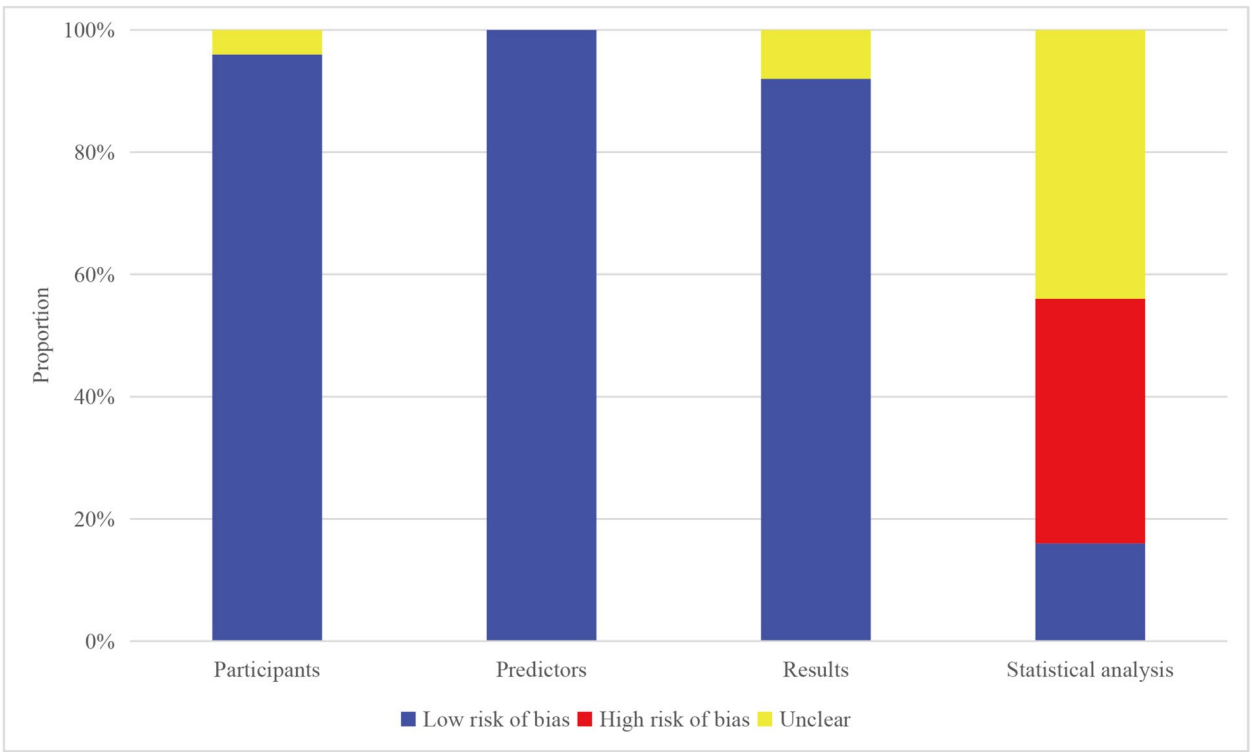

Fig. 2 Risk of bias by PROBAST criteria

for models based on MRI alone. Furthermore, the ML models based on the combination of clinical features and radiomics were also categorized into three subgroups: clinical features + X-ray, clinical features + MRI, and clinical features + MRI + X-ray. The c-index values for these subgroups were 0.837 (0.810–0.865), 0.758 (0.540 – 0.977), and 0.876 (0.842–0.910), respectively. The specific synthesized results are shown in Table 2 and Fig. 3.

Discussion

The prediction of postoperative gait conditions using ML methods in TKA patients was reported as early as 2009 [36]. In the following decades, ML research on TKA has grown increasingly and become diverse, including postoperative patient quality of life, satisfaction, duration of opioid use, and various adverse risk assessments [37–41]. The number of studies on ML prediction of TKA candidates has been increasing gradually since 2019 [28]. A study led by Lee [42] discovered that ML models could be employed to automatically assess the severity of KOA based on X-ray and predict the need for TKA. Zhong [43] et al. also investigated the potential of ML in evaluating individual conditions requiring TKA, but their study was discontinued due to objective conditions and technical problems. This is the first systematic review and meta-analysis to assess the feasibility of ML for predicting the requirement of TKA for KOA patients. The results of this meta-analysis highlighted that the ML model had an ideal performance in identifying TKA candidates. X-ray data were the most widely used among the modeling variables. In terms of model types, LR and DL models were more popular. Most ML models were constructed based on clinical features and radiomics. X-ray and MRI data in radiomics were frequently used as modeling variables. Our analysis showed no overfitting in the effect size of the training set and validation set. Furthermore, the models based on clinical features or radiomics without X-ray data had similar performance to that of models based on the combination of clinical features and radiomics. Any ML model with modeling parameters of X-ray data had better performance both in the training set and validation set, regardless of the complexity of modeling variables.

The Kellgren–Lawrence (KL) grade 4 on radiography (end-stage KOA) and persistent knee joint pain are critical indicators for the necessity of TKA for patients [44, 45]. In some cases, the radiographic evidence of end-stage KOA may not adequately reflect the symptoms and functional status of patients. Numerous endeavors have been made to improve postoperative outcomes for TKA patients by considering several parameters, such as preoperative knee joint pain, function, and survival status [46]. Although X-ray is commonly utilized as the primary diagnostic tool in orthopedics, it is not effective in diagnosing diseases affecting soft tissues. Therefore, early characteristic signs on X-rays are frequently overlooked when predicting TKA events in KOA patients. However, X-ray imaging holds clinical significance in evaluating the condition of the patient's knee joint and determining appropriate treatment strategies, including medication, physical therapy, and surgical interventions for lower limb alignment. These measures could hinder the development of end-stage KOA observed on X-rays. The results of this meta-analysis revealed that the models constructed using only X-ray as the modeling variable had the best performance, and the C-index

Table 2 Meta-analysis results of the C-index of the prediction model for KOA patients

Modeling variables	Training set			Validation set		
	No	c-index	I^2(%)	No	c-index	I^2(%)
Clinical feature	4	0.713 (0.628 – 0.799)	99.7	2	0.656 (0.526 – 0.786)	96.4
Radiomics						
X-ray	3	0.895 (0.865 – 0.924)	84.9	12	0.882 (0.825 – 0.939)	98.2
MRI	2	0.755 (0.508 – 1.000)	99.0	2	0.725 (0.499 – 0.950)	94.8
Overall	5	0.841 (0.777 – 0.904)	97.4	14	0.861 (0.806 – 0.916)	98.1
Clinical feature + radiomics						
Clinical feature + X-ray	7	0.867 (0.824–0.909)	95.4	9	0.837 (0.810–0.865)	59.4
Clinical feature + MRI	2	0.772 (0.539 – 1.000)	98.9	2	0.758 (0.540 – 0.977)	95.2
Clinical feature + MRI + X-ray	6	0.842 (0.816–0.867)	86.7	2	0.876 (0.842–0.910)	0
Clinical feature + US + X-ray	1	0.820 (0.725–0.915)	NA			
Overall	16	0.844 (0.815 – 0.873)	95.4	13	0.831 (0.799 – 0.863)	79.1

KOA knee osteoarthritis, *No.* number of models, I^2 heterogeneity, *MRI* magnetic resonance imaging, *US* ultrasound, *NA* not available

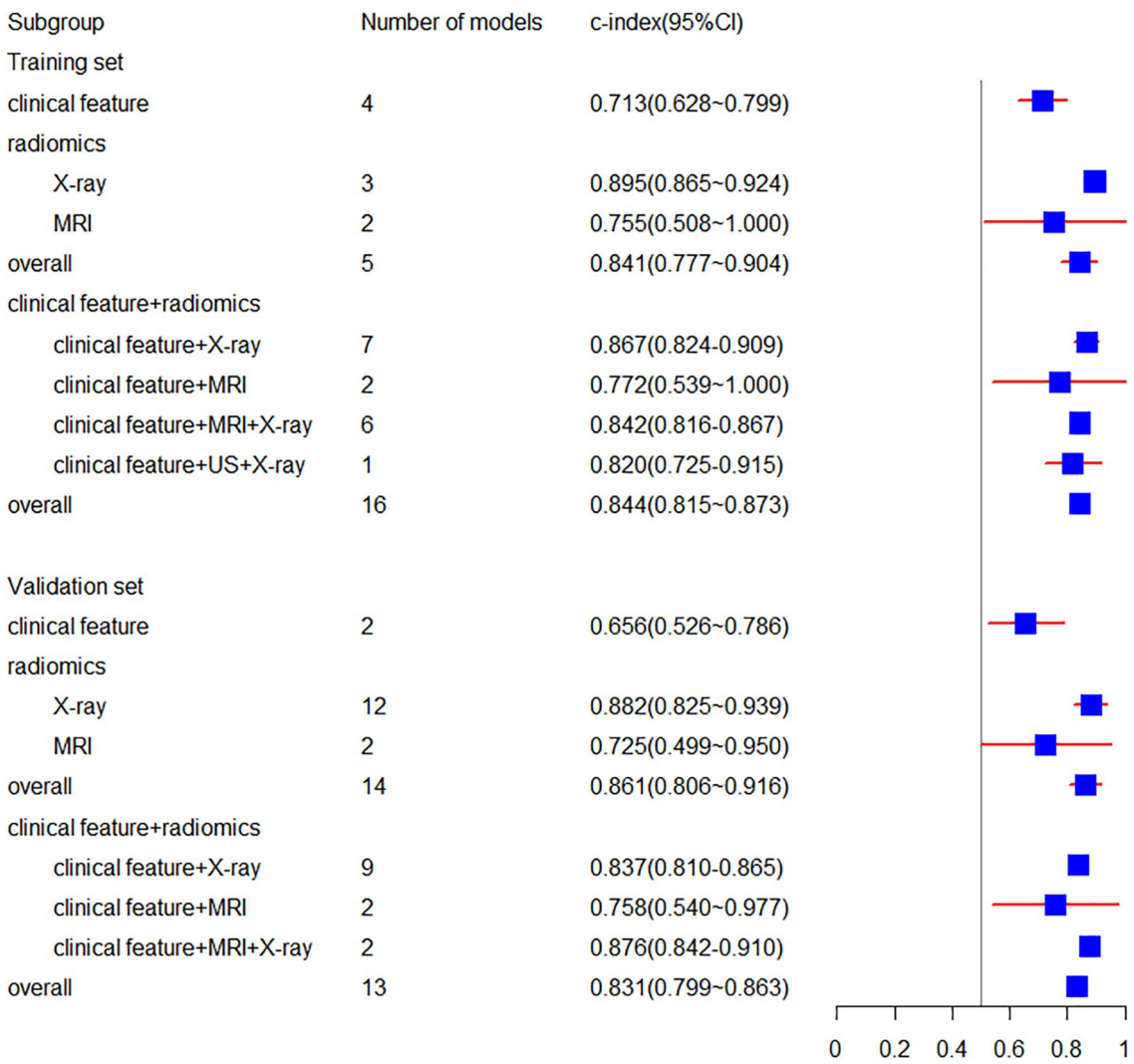

Fig. 3 Forest plot of c-index for machine learning to identify candidates for TKA surgery

was 0.895 (0.865 – 0.924) in the training set and 0.882 (0.825 – 0.939) in the validation set in the subgroup.

Limitations

However, it is imperative to recognize the presence of multiple limitations that could impact the interpretation of the results. At present, there are still few original studies eligible for the meta-analysis. For example, there was only one model in the clinical features + US + X-ray group in the subgroup of ML models based on the combination of clinical features and radiomics, which may have influenced the results. The limited use of modeling variables other than X-rays in the included studies may undermine the predictive performance. At the same time, in the context of radiomics as variables, only a few studies employed intelligent extraction tools for segmenting regions of interest, which is typically influenced by human expertise and potentially introduces bias to the model. It is worth noting that a high proportion of studies were based on modeling in the Osteoarthritis Initiative Cohort, and only two studies had modeling variables based on local databases. These databases are not designed for constructing ML models. They may lack prospective patient data and the available patient data may be incomplete. Consequently, variables for these ML models often have to be selected after the fact. In addition to the known risk factors and clinical experience, the selection of modeling variables is heavily influenced by database access rights. Given no standard database specifically for building ML models, the selection of modeling variables in many studies is complicated. The utilization of many irrelevant modeling variables not only increases the research time but also may weaken the performance of the model. Using only one indicator for

evaluating performance may result in misinterpretation. It is challenging to directly obtain or indirectly calculate enough indicators, such as sensitivity and specificity to enhance the accuracy of false positive and false negative predictions. In addition, this study did not analyze publication bias among the included literature primarily due to the diverse modeling variables and the limited number of models in the subgroups. At present, ML models designed to identify candidates for TKA are not extensively utilized within clinical settings. However, in order to maximize the potential advantages of these models in clinical practice, it is essential to build more targeted and professional prospective databases that support the establishment of ML models.

Conclusions

The ML model is feasible in identifying candidates for TKA. X-ray-based ML models exhibit the best predictive performance among the models. However, there is currently a lack of high-level research available for clinical application. Furthermore, the accuracy of ML models in identifying candidates for TKA is greatly constrained by the quality of modeling parameters and database architecture. It is crucial to construct a more targeted and professional database to promote the development and clinical application of ML models.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40001-025-02545-z.

Additional file 1. Literature search strategy.

Additional file 2. The features of each model and the data results extracted from the models.

Additional file 3. The detailed assessment of each question in four domains of bias risk.

Author contributions

All authors had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Conceptualization, C.T., X.Y. and H.C.; Methodology, C.T. and R.Z.; Investigation, W.S. and J.S.; Formal Analysis, C.T. and H.C.; Resources, W.S.; Writing—Original Draft, C.T., H.C. and W.S.; Writing—Review & Editing, X.Y. and R.Z.; Visualization, C.T.; Supervision, X.Y.

Funding

This work was supported by Zhejiang Chinese Medicine University Postgraduate Scientific Research Fund Project (No.Y202351301).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]The Third School of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou 310053, Zhejiang, China. [2]Department of Chinese Internal Medicine, Funan Hospital of Chinese Medicine, Fuyang 236300, Anhui, China. [3]Department of Orthopedics, The Third Affiliated Hospital of Zhejiang Chinese Medical University (Zhongshan Hospital of Zhejiang Province), Hangzhou 310053, Zhejiang, China. [4]The First School of Clinical Medicine, Zhejiang Chinese Medical University, Hangzhou 310053, Zhejiang, China.

Received: 23 September 2024 Accepted: 31 March 2025
Published online: 22 April 2025

References

1. Sharma L. Osteoarthritis of the knee. N Engl J Med. 2021;384(1):51–9.
2. Surakanti A, Demory Beckler M, Kesselman MM. Correction: surgical versus non-surgical treatments for the knee: which is more effective? Cureus. 2023;15(7): c129.
3. Osteoarthritis (OA) 2020 [Available from: https://www.cdc.gov/arthritis/basics/osteoarthritis.htm.
4. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1545-1602.
5. Giwnewer U, Rubin G, Orbach H, Rozen N. Treatment for osteoarthritis of the knee. Harefuah. 2016;155(7):403–6.
6. Charlesworth J, Fitzpatrick J, Perera NKP, Orchard J. Osteoarthritis- a systematic review of long-term safety implications for osteoarthritis of the knee. BMC Musculoskelet Disord. 2019;20(1):151.
7. Dossett HG. Machine learning: the future of total knee replacement. Fed Pract. 2022;39(2):62–3.
8. Ayers DC, Yousef M, Zheng H, Yang W, Franklin PD. The prevalence and predictors of patient dissatisfaction 5-years following primary total knee arthroplasty. J Arthroplasty. 2022;37(6s):S121–8.
9. Postler A, Lützner C, Beyer F, Tille E, Lützner J. Analysis of total knee arthroplasty revision causes. BMC Musculoskelet Disord. 2018;19(1):55.
10. Montilla-Herrador J, Lozano-Meca J, Lozano-Guadalajara J, Gacto-Sánchez M. The efficacy of the addition of tDCS and TENS to an education and exercise program in subjects with knee osteoarthritis: a randomized controlled trial. Biomedicines. 2024;12(6):1186.
11. Cao HT, Zhang W, Luo C, Zhao HB, Liu JM. Effect of wrist-ankle acupuncture on postoperative analgesia after total knee arthroplasty. Chin J Integr Med. 2023;29(3):253–7.
12. van Doormaal MCM, Meerhoff GA, Vliet Vlieland TPM, Peter WF. A clinical practice guideline for physical therapy in patients with hip or knee osteoarthritis. Musculoskeletal Care. 2020;18(4):575–95.
13. Elmallah RK, Chughtai M, Khlopas A, Newman JM, Stearns KL, Roche M, et al. Pain control in total knee arthroplasty. J Knee Surg. 2018;31(6):504–13.
14. Zhong H, Jin Y, Liu X, Yang J, Wu S, Liu Y. Short-term effectiveness of high tibial osteotomy combined with arthroscopic surgery for knee varus arthritis and the results of secondary arthroscopic exploration. Zhongguo xiu fu chong jian wai ke za zhi = Zhongguo xiufu chongjian waike zazhi = Chin J Repar Reconstr Sur. 2022;36(8):969–75.
15. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. J Dent Res. 2020;99(7):769–74.
16. Chen Z, Zhao M, You L, Zheng R, Jiang Y, Zhang X, et al. Developing an artificial intelligence method for screening hepatotoxic compounds in traditional Chinese medicine and Western medicine combination. Chin Med. 2022;17(1):58.
17. Bedrikovetski S, Dudi-Venkata NN, Kroon HM, Seow W, Vather R, Carneiro G, et al. Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis. BMC Cancer. 2021;21(1):1058.

18. Zhang Z, Yang L, Han W, Wu Y, Zhang L, Gao C, et al. Machine learning prediction models for gestational diabetes mellitus: meta-analysis. J Med Internet Res. 2022;24(3): e26634.
19. Houserman DJ, Berend KR, Lombardi AV Jr, Duhaime EP, Jain A, Crawford DA. The viability of an artificial intelligence/machine learning prediction model to determine candidates for knee arthroplasty. J Arthroplast. 2022;38:2075.
20. Heisinger S, Hitzl W, Hobusch GM, Windhager R, Cotofana S. Predicting total knee replacement from symptomology and radiographic structural change using artificial neural networks-data from the osteoarthritis initiative (OAI). J Clin Med. 2020;9(5):1298.
21. Mj P, McKenzie J, Bossuyt P, Boutron I, Hoffmann T, Mulrow C, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372: n71.
22. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51–8.
23. Segar MW, Vaduganathan M, Patel KV, McGuire DK, Butler J, Fonarow GC, et al. Machine learning to predict the risk of incident heart failure hospitalization among patients with diabetes: the WATCH-DM risk score. Diabetes Care. 2019;42(12):2298–306.
24. Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res. 2019;28(9):2768–86.
25. Leung K, Zhang B, Tan J, Shen Y, Geras KJ, Babb JS, et al. Prediction of total knee replacement and diagnosis of osteoarthritis by using deep learning on knee radiographs: data from the osteoarthritis initiative. Radiology. 2020;296(3):584–93.
26. Tolpadi AA, Lee JJ, Pedoia V, Majumdar S. Deep learning predicts total knee replacement from magnetic resonance images. Sci Rep. 2020;10(1):6371.
27. Rajamohan HR, Wang T, Leung K, Chang G, Cho K, Kijowski R, et al. Prediction of total knee replacement using deep learning analysis of knee MRI. Sci Rep. 2023;13(1):6922.
28. Yu D, Jordan KP, Snell KIE, Riley RD, Bedson J, Edwards JJ, et al. Development and validation of prediction models to estimate risk of primary total hip and knee replacements using data from the UK: two prospective open cohorts using the UK clinical practice research Datalink. Ann Rheum Dis. 2019;78(1):91–9.
29. Mahmoud K, Alagha MA, Nowinka Z, Jones G. Predicting total knee replacement at 2 and 5 years in osteoarthritis patients using machine learning. BMJ Surg, Interv, Health Technol. 2023;5(1): e000141.
30. Thuraisingam S, Chondros P, Manski-Nankervis JA, Spelman T, Choong PF, Gunn J, et al. Developing and internally validating a prediction model for total knee replacement surgery in patients with osteoarthritis. Osteoarthritis and cartilage open. 2022;4(3):100281.
31. Almhdie-Imjabbar A, Toumi H, Harrar K, Pinti A, Lespessailles E. Subchondral tibial bone texture of conventional X-rays predicts total knee arthroplasty. Sci Rep. 2022;12(1):8327.
32. Jamshidi A, Pelletier JP, Labbe A, Abram F, Martel-Pelletier J, Droit A. Machine learning-based individualized survival prediction model for total knee replacement in osteoarthritis: data from the osteoarthritis initiative. Arthritis Care Res. 2021;73(10):1518–27.
33. Tiulpin A, Saarakkala S, Mathiessen A, Hammer HB, Furnes O, Nordsletten L, et al. Predicting total knee arthroplasty from ultrasonography using machine learning. Osteoarthr Cartil Open. 2022;4(4):100319.
34. Li Y, Feng X, Chong C. Knee replacement risk prediction modeling for knee osteoarthritis using clinical and magnetic resonance image feathres: data from the osteoarthritis initiative. J Mech Med Biol. 2023;23(08):2340068.
35. Liu H, Wang X, Song X, Han B, Li C, Du F, et al. A multiview deep learning-based prediction pipeline augmented with confident learning can improve performance in determining knee arthroplasty candidates. Knee Surg Sports Traumatol Arthrosc. 2024;32(8):2107–19.
36. Levinger P, Lai DT, Begg RK, Webster KE, Feller JA. The application of support vector machines for detecting recovery from knee replacement surgery using spatio-temporal gait parameters. Gait Posture. 2009;29(1):91–6.
37. Tanaka S, Amano T, Uchida S, Ito H, Morikawa S, Inoue Y, et al. A clinical prediction rule for predicting a delay in quality of life recovery at 1 month after total knee arthroplasty: a decision tree model. J Orthop Sci. 2021;26(3):415–20.
38. Muertizha M, Cai X, Ji B, Aimaiti A, Cao L. Factors contributing to 1-year dissatisfaction after total knee arthroplasty: a nomogram prediction model. J Orthop Surg Res. 2022;17(1):367.
39. Zhang Y, Li Z, Su Q, Ge H, Cheng B, Tian M. The duration of postoperative analgesic use after total knee arthroplasty and nomogram for predicting prolonged analgesic use. Front Surg. 2022;9:911864.
40. Chen Y, Jiang Y. Construction of prediction model of deep vein thrombosis risk after total knee arthroplasty based on XGBoost algorithm. Comput Math Methods Med. 2022;2022:3452348.
41. Hinterwimmer F, Lazic I, Langer S, Suren C, Charitou F, Hirschmann MT, et al. Prediction of complications and surgery duration in primary TKA with high accuracy using machine learning with arthroplasty-specific data. Knee Surg Sports Traumatol Arthrosc. 2023;31(4):1323–33.
42. Lee LS, Chan PK, Wen C, Fung WC, Cheung A, Chan VWK, et al. Artificial intelligence in diagnosis of knee osteoarthritis and prediction of arthroplasty outcomes: a review. Arthroplasty. 2022;4(1):16.
43. Zhong J, Si L, Zhang G, Huo J, Xing Y, Hu Y, et al. Prognostic models for knee osteoarthritis: a protocol for systematic review, critical appraisal, and meta-analysis. Syst Rev. 2021;10(1):149.
44. Yi X, Lee J, Yu X, Yi G, Lee H. Assessing the efficacy of the early rehabilitation pathway in combination with morita therapy after hip and knee arthroplasty. J Healthc Eng. 2022;2022:4285197.
45. Goh G, Schwartz A, Friend J, Grace T, Wickes C, Bolognesi M, et al. Patients who have kellgren-lawrence grade 3 and 4 osteoarthritis benefit equally from total knee arthroplasty. J Arthroplasty. 2023;38:1714–7.
46. Gademan MG, Hofstede SN, Vliet Vlieland TP, Nelissen RG, Marang-van de Mheen PJ. Indication criteria for total hip or knee arthroplasty in osteoarthritis: a state-of-the-science overview. BMC Musculoskelet Disord. 2016;17(1):463.

Publisher's Note