SYSTEMATIC REVIEW

Prediction Models for Knee Osteoarthritis: Review of Current Models and Future Directions

Taghi Ramazanian, MD[1, 2]; Sunyang Fu, MHI[1]; Sunghwan Sohn, PhD[1]; Michael J. Taunton, MD[2]; Hilal Maradit Kremers, MD MSc[1, 2]

Research performed at the Mayo Clinic, Rochester MN, USA

Received: 13 September 2021 *Accepted: 23 February 2022*

Abstract

Background: Knee osteoarthritis (OA) is a prevalent joint disease. Clinical prediction models consider a wide range of risk factors for knee OA. This review aimed to evaluate published prediction models for knee OA and identify opportunities for future model development.

Methods: We searched Scopus, PubMed, and Google Scholar using the terms knee osteoarthritis, prediction model, deep learning, and machine learning. All the identified articles were reviewed by one of the researchers and we recorded information on methodological characteristics and findings. We only included articles that were published after 2000 and reported a knee OA incidence or progression prediction model.

Results: We identified 26 models of which 16 employed traditional regression-based models and 10 machine learning (ML) models. Four traditional and five ML models relied on data from the Osteoarthritis Initiative. There was significant variation in the number and type of risk factors. The median sample size for traditional and ML models was 780 and 295, respectively. The reported Area Under the Curve (AUC) ranged between 0.6 and 1.0. Regarding external validation, 6 of the 16 traditional models and only 1 of the 10 ML models validated their results in an external data set.

Conclusion: Diverse use of knee OA risk factors, small, non-representative cohorts, and use of magnetic resonance imaging which is not a routine evaluation tool of knee OA in daily clinical practice are some of the main limitations of current knee OA prediction models.

Level of evidence: III

Keywords: Artificial intelligence, Knee osteoarthritis, Machine learning, Prediction models

Introduction

Osteoarthritis (OA) is the most prevalent joint disease and a leading cause of chronic pain and disability in the United States and around the world.[1,2] Knee OA affects one-fifth of Americans aged 45 years and older and accounts for more than 80% of the global burden of osteoarthritis.[2,3] Since the mid-20th century, knee OA has doubled in prevalence, even after accounting for the effects of age and body mass index (BMI).[4]

Although there are several treatments that provide symptomatic relief in knee OA, their benefits are sometimes overweighed by adverse effects.[5-7] Knee OA has a progressive course, and despite extensive research, there are no effective medical treatments to slow down the disease progression.[8] Even after total knee arthroplasty (TKA), 6% to 30% still experience persistent knee pain.[9-11] Precise prediction of disease incidence and progression is important to delay or prevent the onset of cartilage degeneration by correcting modifiable risk factors such as obesity and varus malalignment. Prediction models

Corresponding Author: Hilal Maradit-Kremers, Department of Health Sciences Research, Mayo Clinic, 200 First St SW Rochester, Rochester, Minnesota, USA
Email: maradit@mayo.edu

THE ONLINE VERSION OF THIS ARTICLE ABJS.MUMS.AC.IR

could play a critical role to guide clinical decision-making and patient management as well as a screening tool to diagnose knee OA in the early stages.[12]

Risk factors for knee OA[13-15] are broadly categorized into five classes: demographics (e.g., age, gender, and education), anthropometric characteristics (BMI and waist circumference), medical history (e.g., knee pain, knee stiffness, and underlying diseases), blood biomarkers, and imaging markers.[14-16] Recognition of risk factors for knee OA is accompanied by prognostic research to better predict the knee OA progression at the individual or population level to target preventions and/or treatments to modify the disease course and improve outcomes. Therefore, clinical risk prediction models offer the opportunity to consider a wide range of confirmed and potential risk factors for knee OA and the interactions between these factors.

Historically knee OA prediction models were developed using traditional logistic regression methods.[17-32] More recently, prediction models using machine learning (ML) approaches have been introduced.[33-42] Although there is no clear distinction between ML methods and traditional statistical methods, it is generally believed that machine learning-based predictive models can handle big data and uncertainty in clinical and biological models.[16,43] We considered all regression models such as logistic regression and LASSO (Least Absolute Shrinkage and Selection Operator) regression as traditional statistical models and all models that used learning algorithms such as artificial neural network (ANN), support vector machine (SVM), K nearest neighbors (KNN) as machine learning-based predictive models.

Additionally, apart from statistical methodology, knee OA prediction models vary considerably concerning the type and definitions of risk factors included in model development, study population, data sources, and statistical modeling approaches. This variation in part hampers the generalizability and the implementation of many of these models in clinical practice. A systematic review of knee OA prediction models would improve understanding of the utility of the current models in clinical practice, clinical research, and the future research agenda.

This review aims to summarize the current prediction models for knee OA and identify strategies for future model development. Identifying opportunities for improvement in the design, conduct, analysis, validation, and reporting of prognostic research in knee OA is crucial to improve their utility in routine clinical practice and to ultimately help improve patient outcomes.

Materials and Methods

We searched Scopus, PubMed, and Google Scholar using the terms knee osteoarthritis, prediction model, deep learning, and ML after the year 2000. All the retrieved articles were reviewed by one of the researchers. We only included articles that reported a prediction model for radiographic knee OA incidence or progression and presented them as journal articles. We excluded articles that predicted other clinical outcomes (e.g., prediction of knee pain) or TKA. Included studies were divided into two groups according to statistical methodology: traditional regression-based models and ML models. We extracted the following information from each article: Author, year of publication, data source, sample size, risk factors and imaging data (predictor variables), statistical methodology, outcome definitions, calibration, results from external validation, and limitations addressed in the paper or according to the checklist of items on TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) reporting guidelines related to study development and validation.[44]

Results

From 1,645 papers published after the year 2000, we identified a total of 26 prediction models of which 16 employed traditional regression-based methods and 10 ML methods. There was substantial variation in the number and types of predictors included in knee OA prediction models. Table 1 lists the predictors included in traditional models and Table 2 lists the predictors included in ML models. While age, BMI, sex, and radiographic findings were the most common predictor variables included in traditional models, income, waist circumference, blood biomarkers, genetic markers, and MRI findings were rarely included [Table 1]. The most common predictor variables in ML models were age, BMI, and radiographic and MRI markers [Table 2]. Although none of the ML models included genetic data, medication and family history; income, nutrition characteristics, knee stiffness, and concomitant affected joints were included in some ML models. On average about 25% of traditional and 20% of ML models used demographic data other than just age and sex such as education, income, and family history. Medical history (e.g., underlying diseases and concomitant joint diseases) and blood biomarkers (e.g., serum and urine biomarkers or genetic data) were included in three and five traditional models, respectively. Only three ML models included medical history and one ML model included blood biomarkers, respectively. A total of 13 traditional models used radiographic markers and only 2 used MRI markers. The use of radiographic (5 studies) and MRI markers (4 studies) was more common in ML models. The most Most used radiographic markers were Kellgren-Lawrence (KL) grade and MRI markers of cartilage and meniscus morphology.

Tables 3 and 4 provide an overview of methodological characteristics and main findings in traditional [Table 3] and ML [Table 4] prediction models. While half of the ML models used data from the Osteoarthritis Initiative (OAI) cohort, only 4 of the 16 traditional models used OAI data, and the data sources for the traditional models were more heterogeneous than the ML models. The median sample sizes for traditional and ML models were 780 (range = 105 – 40,118) and 295 (range = 68 – 5,749), respectively. Almost all the traditional models used logistic regression for model development except one model which used Lasso regression. ML models used a variety of learning algorithms and feature engineering, including ANN (2 studies), Bayesian network (2 studies), random forest (2 studies), and KNN (2 studies) [Table 4].

Traditional and ML models achieved an area under the

Table 1. Knee Osteoarthritis Predictors included in Traditional Regression (TR) Models

		Fernandes[20]	Garriga[21]	Joseph[22]	Kerkhof[18]	Kinds[32]	Kraus[23]	Landsmeer[59]	LaVelly[24]	Magnusson[28]	Oudenaarde[26]	Riddle[19]	Schett[30]	Takahashi[31]	Halilaj[35]	Woloszynski[27]	Zhang[14]
Demographics	Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Sex	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
	Education				✓								✓				
	Occupation																✓
	Income												✓				
	Knee symptoms	✓	✓	✓				✓				✓			✓		✓
	Physical activity							✓				✓			✓		✓
	History of knee injury and/or surgery	✓		✓			✓		✓	✓		✓			✓		✓
	Family history	✓		✓				✓							✓		✓
Anthropometric measures	Body mass index	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
	Waist circumference		✓														
Medical history	Underlying diseases											✓	✓				
	Concomitant joint- affected diseases				✓							✓					
	Knee pain			✓	✓	✓	✓		✓			✓			✓		✓
	Knee stiffness		✓					✓							✓		
	Pharmacological treatment		✓				✓										
Blood biomarkers	Biomarkers[1]		✓		✓		✓						✓				
	Genetic markers[2]				✓									✓			
Imaging markers	Radiography		✓	✓	✓	✓	✓	✓	✓		✓	✓		✓	✓	✓	✓
	Magnetic Resonance Imaging			✓							✓						

Examples of some biomarkers: Urinary C-terminal cross-linked telopeptide of type II collagen (uCTX-II), cartilage oligomeric matrix protein (COMP), hyaluronan (HA), C-terminal crosslinked telopeptide of type I collagen (CTXI) and matrix metalloproteinase 3
Examples of genetic markers: Single nucleotide variants (SNPs), asporin (ASPN), growth differentiation factor 5 (GDF5), and double von Willebrand factor A domains (DVWA)

Table 2. Knee Osteoarthritis Predictors included in Machine Learning (ML) Models

		Yoo[33]	Ashinsky[41]	Du[34]	Watt[29]	Lazzarini[40]	Lim[36]	Long[42]	Sheng[37]	Tiulpin[38]	Zhong[39]
Demographics	Age	✓			✓	✓	✓	✓	✓	✓	✓
	Sex	✓			✓		✓	✓	✓	✓	✓
	Education	✓					✓		✓		
	Occupation										
	Income	✓					✓				
	Knee symptoms	✓	✓		✓						✓
	Physical activity	✓	✓			✓	✓				✓
	History of knee injury and/or surgery							✓		✓	
	Family history										

Table 2. Continued

Anthropometric measures	BMI	✓			✓	✓	✓	✓	✓	✓	
	Waist circumference	✓				✓			✓		
Medical history	Underlying diseases	✓			✓		✓				
	Concomitant joint- affected diseases				✓						
	Knee pain	✓	✓		✓	✓					✓
	Knee stiffness	✓	✓								✓
	Pharmacological treatment				✓						
Blood biomarkers	Biomarkers[1]					✓					
	Genetic markers										
Imaging markers	Radiography	✓			✓	✓		✓		✓	
	Magnetic Resonance Imaging		✓	✓		✓					✓

Examples of biomarkers: fibulin 3-1, fibulin 3-2, fibulin 3-3, nitrated type II collagen denaturation (COLL2-1NO2), type I and type II collagen metabolites (C1M, C2M)

Table 3. Traditional Regression-based Prediction Models for Knee Osteoarthritis

Author	Year	Data source	Sample size	Statistical Method	Outcome definition	Calibration	Results	External Validation & Results	Limitations
Fernandes(20)	2017	North Nottinghamshire UK	1822	Bayesian logistic regression	Incident knee pain: Self-reported knee pain on most days for at least 1 month.	HLT p-value 0.52	ROC 0.70, sensitivity 94%, specificity 32%	OAI ROC 0.54, sensitivity 73%, specificity 31%	Poor outcome definition prone to recall bias, poor performance in external validation
Garriga(21)	2019	Chingford 1000 Women study	649	logistic regression	Incident radiographic knee OA: KL grade progression to >2	Good agreement, no statistics	Radiographic model: AUC 0.797, Clinical model: AUC 0.692	No	Women only, changes in lifestyles since study performed in 1980-90's, no external validation
Joseph(22)	2018	OAI	641	logistic regression	Moderate to severe radiographic or symptomatic knee OA: worsening to KL grade 3–4. WOMAC pain score >5 or TKA	Not assessed	AUC Model-1 0.67 Model-2 0.71 Model-3 0.72	No	Inclusion of at risk-rly knee OA subjects, only use of T2 composite, potential interactions between knees were not considered, no external validation
Kerkhof(18)	2013	Rotterdam Study (RS-I)	929	logistic regression	Incident knee OA: Progression from KL grade <2 to grade≥2 at follow-up	HLT p-value Model-1 0.19 Model-2 0.46 Model-3 0.79 Model-4 0.73 Model-5 0.90	AUC Model-1 0.66 Model-2 0.66 Model-3 0.67 Model-4 0.79 Model-5 0.79	RS-II, Chingford Study AUC RS-II 0.86 Chingford 0.76	History of knee injury and physical activity were not included
Kinds(32)	2012	Cohort Hip & Cohort Knee (CHECK) Study, NL	1002	logistic regression	Radiographic knee OA: KL grade 2 Clinical knee OA: painful knee (highest 3 WOMAC quintile)	Not assessed	AUC Clinical model 0.64 Clinical&+radiographic features 0.74 KL-grade (0 and 1) 0.70	No	Inclusion of at risk- knee OA subjects, WOMAC outcomes were measured on participant level, no external validation
Kraus(23)	2017	OAI	600	logistic regression	Clinically relevant knee OA (both radiographic and pain progression)	Not assessed	AUC 0.586	No	biomarkers level is influenced by all joints, no external validation

Table 3. Continued

Landsmeer(59)	2018	PROOF	407	logistic regression	FKP (pain in/ around 1 or both knees on most days in the past month) Symptomatic knee OA(FKP and a definite tibiofemoral osteophyte in the same knee)	HLT p-value Basic model 0.92 Backward model 0.93	AUC Basic model 0.63 Backward model 0.71	Rotterdam Study AUC 0.71	High number lost to follow-up, knee pain developed between set points 2.5 and 6.5 years were not detectable
LaVelly(24)	2017	OAI	553	logistic regression	Structural knee OA progression: Loss of medial joint space within a single knee from each participant on the radiograph between the 36- and 48-month examinations	Intercept / Slope Base model -0.99 / 0.47 BMD model -0.63 / 0.63	AUC Base model 0.65 BMD model 0.73	No	Inclusion of at risk- knee OA subjects, no external validation
Magnusson(28)	2018	Swedish Conscription Registry	40118	logistic regression	Incident knee OA: first record of an OA diagnosis in inpatient or specialist care	Calibration plot but no statistics	AUC 0.60	No	OA diagnosed by specialist included might be more severe, organization of the Swedish health care system may influenced the findings, only male studied, no external validation
Oudenaarde(26)	2017	Cohort Hip & Cohort Knee (CHECK) Study, NL	148	logistic regression	Incident knee OA: KL grade ⩾ 2 or TKA	HLT p-value 0.645	AUC 0.722 Optimism corrected AUC 0.685 Sensitivity 66% Specifity 67%	No	Small sample, inclusion of subjects with OA, fair to moderate intra-observer reliability for MRI features cartilage defect and bone marrow lesions, no external validation
Riddle(19)	2016	MOST	1690	logistic regression	Knee OA with rapid progression: radiographic worsening from KL grade 0-1 at baseline to grade 3-4	HLT p-value Model-1 0.401 Model-2 0.881	AUC Model-1 0.78 Model-2 0.78	OAI AUC Model-1 0.76 Model-2 0.77	different follow up time between MOST and OAI, relative small number of OA progression, KL 3 and 4 is structural in nature and does not account for knee symptom
Schett(30)	2009	Bruneck Cohort Study	912	logistic regression	Severe knee OA as defined by TKA surgery	HLT p-value Age/sex/BMI 0.055 Soluble VCAM-1 added 0.365	ROC Age/sex/BMI 0.694 Soluble VCAM-1 added 0.734	No	No radiographic markers, no external validation
Takahashi(31)	2010	Japanese population	2158	logistic regression	Knee OA: Clinical symptoms and radiological findings (JSN, osteophytes)	Not assessed	AUC Model-1 0.554 Model-2 0.685 Model-3 0.678	No	Nonrepresentative sample, only three genes studied, no external validation
Halilaj(35)	2018	OAI	1243	LASSO regression	Knee OA progression: Joint space width and WOMAC score	Not assessed	AUC Radiographic progression 0.86 Pain progression 0.95	No	Lack of high quality data from baseline visit, LASSO model estimates are not interpretable individually, no external validation
Woloszynski(27)	2012	Lund University Hospital	105	logistic regression	Knee OA progression: Radiographic medial compartment JSN grade	Not assessed	AUC Medial Trabecular bone texture: Model-1 0.74 Model-2 0.77 Lateral Trabecular bone texture: Model-1 0.68 Model-2 0.7	No	use of 2 different radiographic protocol, all cases had prior meniscectomy, small size of cases for lateral compartment, medial JSN used for OA prediction, radiographs of cases lost to follow up did not examine, the texture parameters do not provide information about bone texture changes at individual scales, no external validation
Zhang(14)	2011	North Nottinghamshire UK	424	logistic regression	Knee OA: KL grade ≥2 in any compartment of any knee	HLT p-value Model-1 2.29 Model-2 11.76 Model-3 12.01	ROC Model-1 0.69 Model-2 0.70 Model-3 0.71	ROC OAI Model-1 0.60 Model-2 0.60 Model-3 0.52 ROC GOAL Model-1 0.74 Model-2 0.79 Model-3 NA	Small sample size, only conventional risk factors included

Abbreviations: OAI (Osteoarthritis Initiative), ROC (Receiver Operating Characteristic), AUC (Area under the ROC curve), PROOF (Prevention of Knee Osteoarthritis in Overweight Females), OAPol (Osteoarthritis Policy), MOST (Multicenter Osteoarthritis Study), ACC (Accuracy), GOAL (Genetics of Osteoarthritis and Lifestyle), LASSO (Least Absolute Shrinkage and Selection Operator), HLT(Hosmer and Lemeshow test), CHECK (Cohort Hip & Cohort Knee), BMD (Bone Mineral Density), VCAM-1(Vascular Cell Adhesion Molecule-1), RS-II(Rotterdam Study- II), JSN(Joint Space Narrowing, FKP (Frequent Knee Pain),TKA(Total Knee Arthroplasty), KL (Kellgren and Lawrence)

Table 4. Machine Learning (ML) Prediction Models for Knee Osteoarthritis

Author	Year	Data source	Sample size	Feature engineering/ data representation	Learning Algorithm	Outcome definition	Calibration	Results	External Validation	Limitations
Yoo (33)	2016	KNHANES V-1	2665	Logistic regression	ANN (architecture unspecified)	Knee OA: Radiographic OA with KL grade > 2 and symptomatic OA (knee pain)	-	AUC Scoring system: Radiographic knee OA 0.73 Symptomatic knee OA 0.88 ANN: Radiographic knee OA 0.0.81 Symptomatic knee OA 0.94	OAI AUC 0.6-0.7	Cross-sectional study, recall bias, did not consider patellofemoral OA, included knee pain
Ashinsky (41)	2017	OAI	68	_	WN(D-CHRM)	Symptomatic knee OA: defined by WOMAC score	-	ACC 75% Sensitivity 74%, Specificity 76%	No	Small sample size, long processing time, only central slice included, registration method was dependent on the target image
Du (34)	2018	OAI	100	PCA	4 methods: ANN (architecture unspecified), SVM, Random forest, Naïve Bayes	Knee OA progression: KL grade, medial and lateral joint space narrowing	-	ROC ANN 0.761 AVM 0.651 Random forest 0.677 Naïve Bayes 0.724	No	Small sample size, no external validation
Lazzarini (40)	2017	PROOF	407	Ranked Guided Iterative Feature Elimination, PCA	Random Forest	Early knee OA: incidentt knee pain, lateral JSN ≥ 1.0 mm, medial JSN ≥ 1.0 mm, incidence of KL ≥ 2	-	AUC 0.823	No	Study population limited to only obese women, no external validation
Lim (36)	2019	KNHANES	5749	PCA	feed-forward neural networks	Early knee OA: If answer to "Have you ever been diagnosed with OA by a doctor", was yes	-	AUC 0.768	No	Self-report for OA diagnosis, absence of progressive data, most input data were binary, excluded OA patients receiving treatment, no external validation
Long (42)	2017	_	176	_	KNN	KOOS	-	AUC Self-reported outcome 0.82 Biomechanical parameters 0.92 All together 1.00	No	Cross-sectional study, small sample size, no external validation
Sheng (37)	2019	Kongjiang community, China	157	_	BN	Knee OA: self-reported	-	AUC 0.78, ACC 76% Sensitivity 73%, Specificity 78%	No	Small, non-representative sample, cross-sectional, knee OA self-reported, no external validation
Tiulpin (38)	2017	OAI, MOST	2711, 2129	CNN	LR and GBM	Knee OA progression: Any increase of in KL grade	-	AUC LR 0.75 GBM 0.76	No	Only standardized radiographs acquired with frame positioning used, imputation in test set of LR model evaluation, total WOMAC used for scoring, no external validation
Zhong (39)	2016	OAI	182	_	KNN	Symptomatic knee OA progression: change in total WOMAC score > 10 by 3 year follow up	-	ACC 84% Sensitivity 77% Specificity 90%	No	Small sample size, no external validation
Watt (29)	2008	OAI	4796	_	BN	Knee pain	-	ACC 89%	No	No definition for scoring for knee pain, no external validation

Abbreviations: KNHANES (Korea National Health and Nutrition Examination Survey), OAI(Osteoarthritis Initiative), ANN (Artificial Neural Networks), WN(D-CHRM) (Weighted Neighbor Distance using Compound Hierarchy of Algorithms Representing Morphology), AUC (Area under the ROC curve), PCA (Principal Component Analysis), SVM (Support Vector Machine), PROOF (Prevention of Knee Osteoarthritis in Overweight Females), DNN (Deep Neural Network), KNN (K Nearest Neighbors), BN (Bayesian Network), MOST (Multicenter Osteoarthritis Study), LR (Logistic Regression), GBM (Gradient Boosting Machine), ACC (Accuracy), WOMAC (Western Ontario and McMaster Universities Arthritis Index)l

curve (AUC) between 0.6 – 0.9 and 0.7 – 1, respectively. Although 6 of the 16 traditional models validated their results in an external data set, only one ML model was validated in an external dataset. Traditional models that were externally validated achieved AUC between 0.6 and 0.8. The AUC for the ML model that was validated in an external dataset33 was 0.6-0.7. None of the ML models reported on calibration and 11 traditional models reported on the calibration of the logistic regression models.

Discussion

In this review paper, we provide a qualitative overview of current knee OA prediction models using both traditional regression and ML methods. While these models incorporate several risk factors for outcome prediction, there are noticeable differences in the inclusion of several well-established risk factors. Although BMI and radiographic data were widely used in these models, some of the well-known risk factors for developing knee OA were only used in one-third of traditional and ML models. The use of questionnaires to collect medical history data and relying on patients' reports on some critical risk factors such as knee alignment (varus or valgus) or history of the previous injury were some of the important inherent biases of most models.

Recently, deep learning – as a subfield of ML that structures algorithms in layers to create ANN that can learn and make intelligent decisions on its own - has dramatically improved state-of-art in several fields and attracted enormous attention in solving complex problems in healthcare due to its representation power along and automated feature learning.[45,46] One notable aspect of knee OA prediction models was the inclusion of MRI-based measures. Half of the ML models used MRI findings to predict the incidence and progression of knee OA. This is surprising since knee MRI is generally not used for knee OA diagnosis in routine clinical practice. Clinical examination combined with radiography is the current standard of practice for knee OA diagnosis, as shown by a systematic review of relevant studies.[47] Although knee MRI provides valuable information about the extent of bone and soft tissue disease in the early stages of knee OA, its use is limited to the research setting and it often reveals abnormal findings of unclear significance in asymptomatic patients.[48] Additionally, because MRI is not used in routine clinical practice for the evaluation of knee OA, it is almost impossible to perform external validation and implementation of the MRI-based models in the clinic.

As outlined in TRIPOD statement,[44] some form of external validation is essential to quantify the predictive performance of prediction models. In other words, a prediction model needs to have acceptable performance in an external dataset. Notably, prediction models are prone to overfitting. Namely, the model closely fits in a particular dataset, but it fails to predict future observations reliably in an external dataset. Furthermore, our review indicates that half of the ML and 25% of traditional models used data from the same OAI cohort. However, the OAI cohort includes individuals with either established knee OA or significant risk factors for the development of knee OA, to facilitate the identification of risk factors for progression from early knee OA to TKA.[49] Hence, the OAI cohort is not necessarily representative of the general population, and the prevalence of knee OA risk factors in the OAI population is higher than in community-based cohorts as shown by Fernandes et al.[20] Conversely, a model that performs well in a community-based population may not perform well in individuals at high risk of knee OA. This type of class imbalance can lead to erroneous predictions that are heavily biased toward the majority class.

An important methodological consideration to obtain robust predictive performance of prediction models is the sample size in relation to the number of predictors included in the model.[44] The size and quality of the dataset and the quality of image data have a significant impact, particularly the robustness of machine learning-based approaches. Both traditional and ML models had a median sample size of less than one thousand and this increases the risk of over-fitting. Small sample sizes and heterogeneous features due to the nature of available training datasets are some of the reasons that limited the use of deep learning models in knee OA prediction models. Furthermore, none of the ML models assessed calibration performance, especially the accuracy of risk estimates based on the agreement between the estimated and observed number of events. This is a major limitation since poorly calibrated models have limited clinical utility due to under or overestimation of the risk or progression of OA.[50] Unfortunately, few investigators have access to large and rich datasets to create and validate OA prediction models.

Development of knee OA prediction models historically applied traditional statistical methods. Almost all the traditional models used logistic regression except only one model which used Lasso regression which is a penalized method and allows consideration of many predictors with a small dataset. Logistic regression is a conventional statistical technique that is used to examine the relationship between a binary outcome (dependent) variable and predictor (explanatory or independent) variables.[51] Although the logistic regression model can recognize important predictors and relative rank more easily, it may fail to detect complex, nonlinear relationships, and interactions between predictor variables and knee OA outcomes. Logistic regression models generally include statistically significant variables ($P < 0.05$); however, in the setting of a very large number of variables, predictors with small effects on the outcome can also become significant.[52] Bayesian network (BN) is a graphical model that predicts a probabilistic relationship between variables. Yet, developing a BN is highly demanding and it has potential limitations in learning high-dimensional data. KNN algorithm is a simple, easy-to-implement, and nonparametric algorithm, but it is not suitable for imbalanced datasets and a high number of predictors. ANN is a complex, high-performance black box approach that can incorporate nonlinearity. Yet, ANN is prone to over-fitting (i.e., the model corresponds too closely to a particular set of data, and fails to fit new data) and the black-box nature of the algorithm limits face validity and acceptance by clinicians. Therefore, the choice of statistical methodology in predictive modeling should be carefully considered. Furthermore, emerging evidence suggests that ML algorithms may not outperform traditional regression approaches, especially in low-dimensional settings.[53]

A growing amount of heterogeneous risk factor data in knee OA research including biomedical, biomechanical, and clinical data along with the complex nature of risk factors creates challenges in the development and validation of ML approaches. Development of a robust

knee OA prediction model with good performance requires large, rich datasets and a wide range of well-documented risk factors along with radiographic data as the gold standard of knee OA evaluation in clinical settings and rigorous internal and external validation. Heterogeneity of study populations, risk factor definitions, follow-up periods, outcome measures, and reporting of results are some of the challenges for quantitative comparisons across prediction models.[54-57] For example, the out-of-sample performance of prediction models is hampered if risk factors are defined differently across studies.[58,59] Therefore, more methodological research is warranted for standardized definitions to improve the transportability of OA prediction models.

In conclusion, the current traditional and ML knee OA prediction models include a variety of clinical, image-derived, and patient-reported predictors. Most of the models are developed using data from the OAI cohort which does not necessarily represent the general population. The small sample size is one of the notable weaknesses. Furthermore, ML prediction models that include imaging data use MRI findings, but MRI is not a routine evaluation tool for knee OA in clinical practice. This limits the external validation and utility of MRI-based ML models in clinical practice. Further research is warranted to develop and validate knee OA prediction models in diverse populations incorporating a wide range of knee OA risk factors and radiographic markers from knee radiographs. Such models will offer great promise for implementing in routine clinical care and subsequently improve the clinical decision-making process.

Taghi Ramazanian MD[1, 2]
Sunyang Fu MHI[1]
Sunghwan Sohn PhD[1]
Michael J. Taunton MD[2]
Hilal Maradit Kremers MD MSc[1, 2]
1 Department of Health Sciences Research, Mayo Clinic, 200 First St SW Rochester, Rochester, Minnesota, USA
2 Department of Orthopedics, Mayo Clinic, 200 First St SW Rochester, Rochester, Minnesota, USA

References

1. Murray CJ, Atkinson C, Bhalla K, et al. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. Jama. 2013;310(6):591-608. doi: 10.1001/jama.2013.13805.
2. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2163-96. doi: 10.1016/S0140-6736(12)61729-2.
3. Lawrence RC, Felson DT, Helmick CG, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum. 2008;58(1):26-35. doi: 10.1002/art.23176.
4. Wallace IJ, Worthington S, Felson DT, et al. Knee osteoarthritis has doubled in prevalence since the mid-20th century. Proc Natl Acad Sci U S A. 2017;114(35):9332-6. doi: 10.1073/pnas.1703856114.
5. Zhang W, Moskowitz R, Nuki G, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis cartilage. 2007;15(9):981-1000. doi: 10.1016/j.joca.2007.06.014.
6. Zhang W, Nuki G, Moskowitz R, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage. 2010;18(4):476-99. doi: 10.1016/j.joca.2010.01.013.
7. Zhang W, Robertson J, Jones A, Dieppe P, Doherty M. The placebo effect and its determinants in osteoarthritis: meta-analysis of randomised controlled trials. Ann Rheum Dis. 2008;67(12):1716-23. doi: 10.1136/ard.2008.092015.
8. Losina E, Daigle ME, Suter L, et al. Disease-modifying drugs for knee osteoarthritis: can they be cost-effective? Osteoarthritis Cartilage. 2013;21(5):655-67. doi: 10.1016/j.joca.2013.01.016.
9. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KD. Patient satisfaction after total knee arthroplasty: who is satisfied and who is not? Clin Orthop Relat Res. 2010;468(1):57-63. doi: 10.1007/s11999-009-1119-9.
10. Hamel MB, Toth M, Legedza A, Rosen MP. Joint replacement surgery in elderly patients with severe osteoarthritis of the hip or knee: decision making, postoperative recovery, and clinical outcomes. Arch Intern Med. 2008;168(13):1430-40. doi: 10.1001/archinte.168.13.1430.
11. Singh JA, Gabriel S, Lewallen D. The impact of gender, age, and preoperative pain severity on pain after TKA. Clin Orthop Relat Res. 2008;466(11):2717-23. doi: 10.1007/s11999-008-0399-9.
12. McAlindon TE, Bannuru RR, Sullivan M, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage. 2014;22(3):363-88. doi: 10.1016/j.joca.2014.01.003.
13. McWilliams D, Leeb B, Muthuri S, Doherty M, Zhang W. Occupational risk factors for osteoarthritis of

the knee: a meta-analysis. Osteoarthritis Cartilage. 2011;19(7):829-39. doi: 10.1016/j.joca.2011.02.016.
14. Zhang W. Risk factors of knee osteoarthritis–excellent evidence but little has been done. Osteoarthritis Cartilage. 2010;18(1):1-2. doi: 10.1016/j.joca.2009.07.013.
15. Blagojevic M, Jinks C, Jeffery A, Jordan K. Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. Osteoarthritis Cartilage. 2010;18(1):24-33. doi: 10.1016/j.joca.2009.08.010.
16. Jamshidi A, Pelletier JP, Martel-Pelletier J. Machine-learning-based patient-specific prediction models for knee osteoarthritis. Nat Rev Rheumatol. 2019;15(1):49-60. doi: 10.1038/s41584-018-0130-5.
17. Zhang W, McWilliams DF, Ingham SL, et al. Nottingham knee osteoarthritis risk prediction models. Ann Rheum Dis. 2011;70(9):1599-604. doi: 10.1136/ard.2011.149807.
18. Kerkhof HJ, Bierma-Zeinstra SM, Arden NK, et al. Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors. Ann Rheum Dis. 2014;73(12):2116-21. doi: 10.1136/annrheumdis-2013-203620.
19. Riddle DL, Stratford PW, Perera RA. The incident tibiofemoral osteoarthritis with rapid progression phenotype: development and validation of a prognostic prediction rule. Osteoarthritis Cartilage. 2016;24(12):2100-7. doi: 10.1016/j.joca.2016.06.021.
20. Fernandes GS, Bhattacharya A, McWilliams DF, Ingham SL, Doherty M, Zhang W. Risk prediction model for knee pain in the Nottingham community: a Bayesian modelling approach. Arthritis Res Ther. 2017;19(1):59. doi: 10.1186/s13075-017-1272-6.
21. Garriga-Fuentes C, Sanchez-Santos MT, Arden N, et al. Predicting incident radiographic knee osteoarthritis in middle-aged women within four years: the importance of knee-level prognostic factors. Arthritis Care Res (Hoboken) . 2019;72(1). doi: 10.1002/acr.23932.
22. Joseph GB, McCulloch CE, Nevitt MC, et al. Tool for osteoarthritis risk prediction (TOARP) over 8 years using baseline clinical data, X-ray, and MRI: Data from the osteoarthritis initiative. J Magn Reson Imaging. 2018;47(6):1517-26. doi: 10.1002/jmri.25892.
23. Kraus VB, Collins JE, Hargrove D, et al. Predictive validity of biochemical biomarkers in knee osteoarthritis: data from the FNIH OA Biomarkers Consortium. Ann Rheum Dis. 2017;76(1):186-95. doi: 10.1136/annrheumdis-2016-209252.
24. LaValley MP, Lo GH, Price LL, Driban JB, Eaton CB, McAlindon TE. Development of a clinical prediction algorithm for knee osteoarthritis structural progression in a cohort study: value of adding measurement of subchondral bone density. Arthritis Res Ther. 2017;19(1):1-9. doi: 10.1186/s13075-017-1291-3.
25. Losina E, Klara K, Michl GL, Collins JE, Katz JN. Development and feasibility of a personalized, interactive risk calculator for knee osteoarthritis. BMC Musculoskelet Disord. 2015;16(1):1-12. doi: 10.1186/s12891-015-0771-3.
26. van Oudenaarde K, Jobke B, Oostveen AC, et al. Predictive value of MRI features for development of radiographic osteoarthritis in a cohort of participants with pre-radiographic knee osteoarthritis—the CHECK study. Rheumatology (Oxford). 2017;56(1):113-120. doi: 10.1093/rheumatology/kew368.
27. Woloszynski T, Podsiadlo P, Stachowiak G, Kurzynski M, Lohmander L, Englund M. Prediction of progression of radiographic knee osteoarthritis using tibial trabecular bone texture. Arthritis Rheum. 2012;64(3):688-95. doi: 10.1002/art.33410.
28. Magnusson K, Turkiewicz A, Timpka S, Englund M. A Prediction Model for the 40-Year Risk of Knee Osteoarthritis in Adolescent Men. Arthritis Care Res (Hoboken). 2019;71(4):558-62. doi: 10.1002/acr.23685.
29. Watt EW, Bui AA. Evaluation of a dynamic bayesian belief network to predict osteoarthritic knee pain using data from the osteoarthritis initiative. AMIA Annu Symp Proc. 2008:788-92.
30. Schett G, Kiechl S, Bonora E, et al. Vascular cell adhesion molecule 1 as a predictor of severe osteoarthritis of the hip and knee joints. Arthritis Rheum. 2009;60(8):2381-9. doi: 10.1002/art.24757.
31. Takahashi H, Nakajima M, Ozaki K, Tanaka T, Kamatani N, Ikegawa S. Prediction model for knee osteoarthritis based on genetic and clinical information. Arthritis Res Ther. 2010;12(5):R187. doi: 10.1186/ar3157.
32. Kinds MB, Marijnissen AC, Vincken KL, et al. Evaluation of separate quantitative radiographic features adds to the prediction of incident radiographic osteoarthritis in individuals with recent onset of knee pain: 5-year follow-up in the CHECK cohort. Osteoarthritis Cartilage. 2012;20(6):548-56. doi: 10.1016/j.joca.2012.02.009.
33. Yoo TK, Kim DW, Choi SB, Oh E, Park JS. Simple Scoring System and Artificial Neural Network for Knee Osteoarthritis Risk Prediction: A Cross-Sectional Study. PLoS One. 2016;11(2):e0148724. doi: 10.1371/journal.pone.0148724.
34. Du Y, Almajalid R, Shan J, Zhang M. A novel method to predict knee osteoarthritis progression on MRI using machine learning methods. IEEE Trans Nanobioscience. 2018;17(3):228-36. doi: 10.1109/TNB.2018.2840082.
35. Halilaj E, Le Y, Hicks JL, Hastie TJ, Delp SL. Modeling and predicting osteoarthritis progression: data from the osteoarthritis initiative. Osteoarthritis Cartilage. 2018;26(12):1643-50. doi: 10.1016/j.joca.2018.08.003.
36. 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):1281. doi: 10.3390/ijerph16071281.
37. Sheng B, Huang L, Wang X, et al. Identification of Knee Osteoarthritis Based on Bayesian Network: Pilot Study. JMIR Med Inform. 2019;7(3):e13562. doi: 10.2196/13562.
38. Tiulpin A, Klein S, Bierma-Zeinstra SM, et al.

Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. Scie Rep. 2019;9(1):1-11. doi: 10.1038/s41598-019-56527-3.
39. Zhong H, Miller DJ, Urish KL. T2 map signal variation predicts symptomatic osteoarthritis progression: data from the Osteoarthritis Initiative. Skeletal Radiol. 2016;45(7):909-13. doi: 10.1007/s00256-016-2360-4.
40. Lazzarini N, Runhaar J, Bay-Jensen AC, et al. A machine learning approach for the identification of new biomarkers for knee osteoarthritis development in overweight and obese women. Osteoarthritis Cartilage. 2017;25(12):2014-21. doi: 10.1016/j.joca.2017.09.001.
41. Ashinsky BG, Bouhrara M, Coletta CE, et al. Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the osteoarthritis initiative. J Orthop Res. 2017;35(10):2243-50. doi: 10.1002/jor.23519.
42. Long MJ, Papi E, Duffell LD, McGregor AH. Predicting knee osteoarthritis risk in injured populations. Clin Biomech (Bristol, Avon). 2017;47:87-95. doi: 10.1016/j.clinbiomech.2017.06.001.
43. Chen L. Overview of clinical prediction models. Ann Transl Med. 2020;8(4):71. doi: 10.21037/atm.2019.11.121.
44. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) The TRIPOD Statement. Circulation. 2015;131(2):211-9. doi: 10.1161/CIRCULATIONAHA.114.014508.
45. Zhang L, Lin J, Liu B, Zhang Z, Yan X, Wei M. A review on deep learning applications in prognostics and health management. IEEE Access. 2019;7:162415-38.
46. Schmidhuber J. Deep learning in neural networks: An overview. Neural Netw. 2015;61:85-117. doi: 10.1016/j.neunet.2014.09.003.
47. Menashe L, Hirko K, Losina E, et al. The diagnostic performance of MRI in osteoarthritis: a systematic review and meta-analysis. Osteoarthritis Cartilage. 2012;20(1):13-21. doi: 10.1016/j.joca.2011.10.003.
48. Guermazi A, Niu J, Hayashi D, et al. Prevalence of abnormalities in knees detected by MRI in adults without knee osteoarthritis: population based observational study (Framingham Osteoarthritis Study). Bmj. 2012;345:e5339. doi: 10.1136/bmj.e5339.
49. Nevitt M, Felson D, Lester G. The Osteoarthritis Initiative: A knee health study. Protocol for the cohort study. 2006 Jun:10-3.
50. Shah ND, Steyerberg EW, Kent DM. Big data and predictive analytics: recalibrating expectations. Jama. 2018;320(1):27-8. doi: 10.1001/jama.2018.5602.
51. Hosner DW, Lemeshow S. Applied logistic regression. New York: Jhon Wiley & Son. 1989;581.
52. Ayer T, Chhatwal J, Alagoz O, Kahn Jr CE, Woods RW, Burnside ES. Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. Radiographics. 2010;30(1):13-22. doi: 10.1148/rg.301095057.
53. Gravesteijn BY, Nieboer D, Ercole A, et al. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. J Clin Epidemiol. 2020;122:95-107. doi: 10.1016/j.jclinepi.2020.03.005.
54. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280-6. doi: 10.7326/0003-4819-158-4-201302190-00009.
55. Wolff RF, Moons KG, Riley RD, 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. doi: 10.7326/M18-1376.
56. Bastick AN, Belo JN, Runhaar J, Bierma-Zeinstra SM. What are the prognostic factors for radiographic progression of knee osteoarthritis? A meta-analysis. Clin Orthop Relat Res.2015;473(9):2969-89. doi: 10.1007/s11999-015-4349-z.
57. Chapple CM, Nicholson H, Baxter GD, Abbott JH. Patient characteristics that predict progression of knee osteoarthritis: a systematic review of prognostic studies. Arthritis Care Res (Hoboken). 2011;63(8):1115-25. doi: 10.1002/acr.20492.
58. Luijken K, Groenwold RH, Van Calster B, Steyerberg EW, van Smeden M. Impact of predictor measurement heterogeneity across settings on the performance of prediction models: A measurement error perspective. Stat Med. 2019;38(18):3444-59. doi: 10.1002/sim.8183.
59. Landsmeer ML, Runhaar J, van Middelkoop M, et al. Predicting knee pain and knee osteoarthritis among overweight women. J Am Board Fam Med. 2019;32(4):575-84. doi: 10.3122/jabfm. 2019.04.180302.