## ■ SYSTEMATIC REVIEW

# Understanding the role of machine learning in predicting progression of osteoarthritis

## A SYSTEMATIC REVIEW

S. Castagno,
B. Gompels,
E. Strangmark,
E. Robertson-Waters,
M. Birch,
M. van der Schaar,
A. W. McCaskie

*From University of Cambridge, Cambridge, UK*

### Aims

Machine learning (ML), a branch of artificial intelligence that uses algorithms to learn from data and make predictions, offers a pathway towards more personalized and tailored surgical treatments. This approach is particularly relevant to prevalent joint diseases such as osteoarthritis (OA). In contrast to end-stage disease, where joint arthroplasty provides excellent results, early stages of OA currently lack effective therapies to halt or reverse progression. Accurate prediction of OA progression is crucial if timely interventions are to be developed, to enhance patient care and optimize the design of clinical trials.

### Methods

A systematic review was conducted in accordance with PRISMA guidelines. We searched MEDLINE and Embase on 5 May 2024 for studies utilizing ML to predict OA progression. Titles and abstracts were independently screened, followed by full-text reviews for studies that met the eligibility criteria. Key information was extracted and synthesized for analysis, including types of data (such as clinical, radiological, or biochemical), definitions of OA progression, ML algorithms, validation methods, and outcome measures.

### Results

Out of 1,160 studies initially identified, 39 were included. Most studies (85%) were published between 2020 and 2024, with 82% using publicly available datasets, primarily the Osteoarthritis Initiative. ML methods were predominantly supervised, with significant variability in the definitions of OA progression: most studies focused on structural changes (59%), while fewer addressed pain progression or both. Deep learning was used in 44% of studies, while automated ML was used in 5%. There was a lack of standardization in evaluation metrics and limited external validation. Interpretability was explored in 54% of studies, primarily using SHapley Additive exPlanations.

### Conclusion

Our systematic review demonstrates the feasibility of ML models in predicting OA progression, but also uncovers critical limitations that currently restrict their clinical applicability. Future priorities should include diversifying data sources, standardizing outcome measures, enforcing rigorous validation, and integrating more sophisticated algorithms. This paradigm shift from predictive modelling to actionable clinical tools has the potential to transform patient care and disease management in orthopaedic practice.

Cite this article: *Bone Joint J* 2024;106-B(11):1216–1222.

## Introduction

Osteoarthritis (OA) is a complex, degenerative joint condition, and a prevailing cause of disability worldwide. Characterized by an intricate interplay of mechanical, inflammatory, and genetic pathways, OA presents a polymorphic profile, making it a challenging disease to define.[1-7] Current therapies cannot halt or reverse its progression, and traditional models are unable to predict disease trajectories or stratify patients by risk of progression.[8]

Machine learning (ML) is a subset of artificial intelligence (AI) designed to develop systems that learn and make data-driven decisions without explicit programming.[9] By analyzing large datasets and identifying patterns of disease that may elude conventional analyses, ML has the

potential to enhance the precision of predicting OA progression. Improved predictive capability is crucial for effectively stratifying patients, optimizing early intervention strategies, and selecting appropriate candidates for clinical trials, thereby advancing personalized treatment in OA.[10]

The field of ML is extensive and covers many different learning problems, from the analysis of tabular data to computer vision, which enables computers to derive meaningful information from visual inputs such as digital images,[11] and natural language processing, which is concerned with giving computers the ability to understand text and spoken words.[12] ML algorithms can be broadly categorized into supervised and unsupervised learning. In supervised learning problems, the algorithms are fed input training data with their corresponding known output values (or labels), from which the algorithm will learn to predict. Although expert labelling is required for training and testing datasets, a well-trained model can be used to perform inference on unlabelled data.[13] There are two main categories of supervised learning: classification, where the output values are categorical; and regression, where the output values are numeric.[14] In contrast to supervised learning, unsupervised learning aims to detect patterns in a dataset that are not informed by a target or label and therefore require no user supervision.[15] One of the most common unsupervised learning tasks is clustering, which groups instances in a dataset into separate clusters based upon specific combinations of their features.[15-17] Finally, deep learning (DL) represents a specialized subset of ML that employs multiple processing layers to discern data representations across a spectrum of abstraction levels.[18] A prime example is convolutional neural networks, which are commonly used in image recognition.

To understand the role of ML in predicting progression of OA, we conducted a systematic review analyzing studies that employ a range of ML techniques (from supervised learning to advanced deep learning models) to forecast disease trajectories and stratify patients by progression risk. Our aim is to assess the accuracy and utility of these models in OA management, identify research gaps, and propose directions for future study. By synthesizing diverse ML methodologies and their clinical outcomes, this review aims to guide the development of predictive models, enhance early intervention strategies, and support the shift towards personalized medicine in OA treatment, potentially transforming patient outcomes.

## Methods

We conducted a systematic review in accordance with the PRISMA guidelines,[19] and registered it in PROSPERO (ID: CRD42023446500). Ethical approval for this study was not required as it involved the analysis of published data and did not include any direct human or animal subjects.

We searched for articles in MEDLINE and Embase on 5 May 2024, by using the following search strategy on the Ovid platform: [("progression" OR "prediction" OR "incidence" OR "prognostic model" OR "predictive model").mp. OR Disease Progression/ OR Incidence/] AND ["Osteoarthritis".mp. OR Osteoarthritis/] AND ["machine learning".mp. OR Artificial Intelligence/ OR Machine Learning/ OR Deep Learning/ OR Neural Networks/ OR Algorithms/].

Terms within quotation marks specify explicit search phrases, ".mp." indicates a multi-purpose field search, and "/" denotes the use of Medical Subject Headings (MeSH) for focused topic indexing.

We included original studies that employ ML algorithms to predict the progression of OA in humans. These algorithms operate on a supervised, unsupervised, or reinforcement learning framework, and can analyze a wide array of data types including clinical, imaging, or biochemical data.

Exclusions were made for reviews, pre-prints, opinion pieces, conference abstracts, and case reports, alongside studies not employing ML, such as those using traditional statistical analysis (including logistic regression, which is often used as a benchmark for comparison with other more advanced algorithms)[20] or bioinformatics analysis. Additionally, articles including non-human subjects, investigating other types of arthritis (other than OA) or solely concentrating on OA diagnosis, phenotyping, or the identification of OA risk factors or biomarkers were omitted. The same exclusion criteria applied to studies predicting progression to joint arthroplasty or examining the impacts of surgical or medical interventions. We excluded articles focusing on the progression to arthroplasty for several reasons. First, these studies typically emphasize patient eligibility for surgery rather than the underlying biological progression of OA, which diverges from our aim to understand and predict the disease's natural progression through biological and physiological markers. Second, the criteria for recommending arthroplasty differ widely across regions and healthcare systems, introducing subjectivity into surgical decision-making that can mask the true correlation between disease severity and the decision to undergo surgery. Additionally, the choice to proceed with such surgery is influenced by a variety of factors beyond disease progression, including education level, income, fitness for surgery, and health insurance coverage, further complicating the analysis.[21]

The selection of studies for inclusion in the review followed a two-stage process. First, we screened titles and abstracts of studies identified from the search based on the inclusion and exclusion criteria. Studies that appeared to meet the criteria then underwent a full-text review before final inclusion. This process was conducted independently by SC, BG, and ES, and any disagreements were resolved through consultation with ERW.

For each article, we systematically extracted key data comprising the year of publication, the specific joint afflicted by OA under study, the dataset harnessed to train the ML models along with its public availability, and the severity of OA at baseline. We included the type of data analyzed (e.g. clinical data, imaging) and the authors' definition of OA progression. Additionally, we extracted information on the ML algorithms deployed, their accuracy in predicting disease progression alongside the validation methods employed, and any interpretability analyses of the outcomes. We employed a narrative synthesis approach to summarize and explain our findings.

Finally, we assessed the risk of bias of each article with the PROBAST tool.[22] This tool examines four key areas: participants, predictors, outcomes, and analysis. The final assessment of risk of bias and applicability concerns were categorized as "low", "high", or "unclear". As described in Abdulazeem et
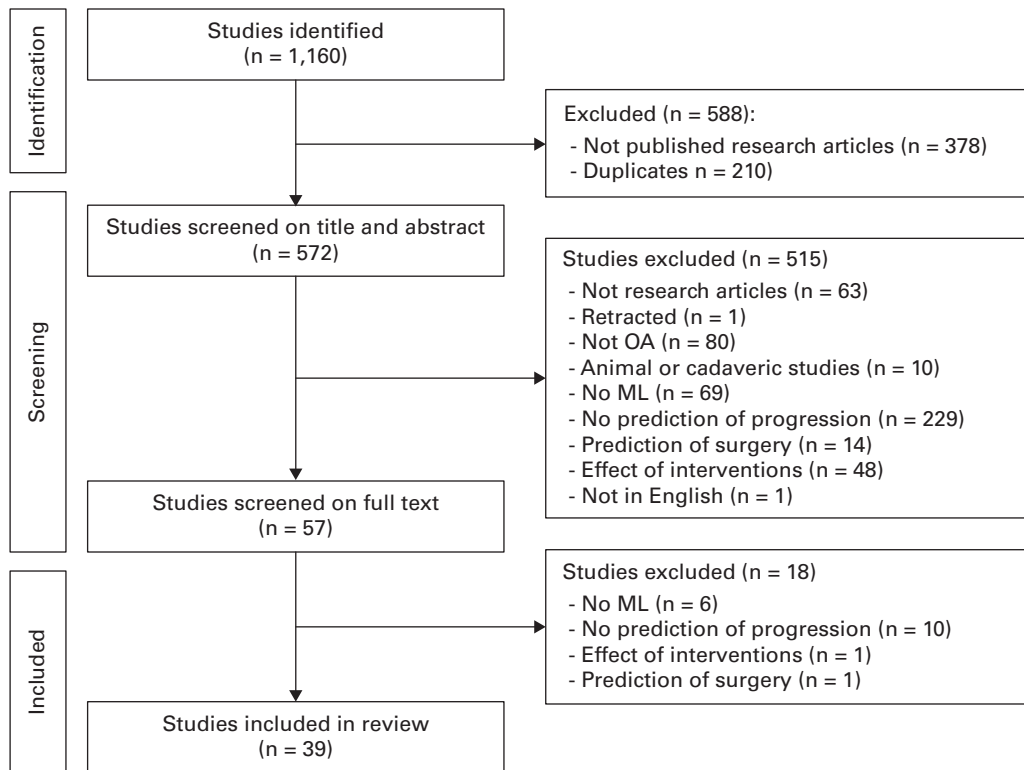
Fig. 1

Flow diagram illustrating the process of identification and selection of articles for our systematic literature review. ML, machine learning; OA, osteoarthritis.

al,[23] a model developed without external validation was judged as "high risk" independently of all other domains, unless the model was derived from an exceptionally large dataset and incorporated some type of internal validation.

## Results

Our initial search yielded 1,160 studies. Of these, 378 did not match the inclusive requirements and a further 210 were duplications. None of the articles included systematic reviews on OA progression prediction using ML. In the initial screening step, which involved evaluating titles and abstracts, an additional 515 studies did not meet the inclusion criteria. Upon a thorough review of the full text of the remaining 57 articles, 18 additional studies were excluded. Finally, 39 studies were selected for analysis.[21,24–61] This selection process is summarized in Figure 1.

All the articles spanned a publication period from 2012 to 2024. Notably, around 85% (n = 33) were published between 2020 and the present. All but three articles concentrated solely on knee OA, with the remaining focusing on shoulder OA,[24] temporomandibular joint (TMJ) OA,[25] and one study focusing on any type of OA (in the arm, foot, spine, hip, and knee).[26]

A large variability was noted among the studies in terms of severity of OA at baseline, ranging from asymptomatic or early-stage OA to end-stage. Interestingly, a minority of studies (21%; n = 8) focused not only on predicting disease progression but also on enabling early diagnosis.[26,39,44,46,47,50,53,58]

A substantial 82% of the studies (n = 32) had used publicly available datasets for the development and training of their models, with 74% (n = 29) using data from the Osteoarthritis Initiative (OAI).[62] Additional cohorts that were examined are revealed in Supplementary Table i.[63–70] A total of 21 studies (54%) employed solely tabular data for model development and training. Conversely, nine studies (23%) used only features extracted from clinical images and a further nine studies (23%) melded both image and tabular data analysis.

The diversity in data types incorporated for model training was notable. Most studies included clinical data (77%; n = 30), patient-reported outcomes (59%, n = 23); or radiological data (74%; n = 29). However, a smaller proportion used MRI data (33%; n = 13) or CT (3%; n = 1), and very few incorporated biochemical markers (10%; n = 4) or omics data (8%; n = 3). None of the studies employed movement data obtained from wearable accelerometers or gait analysis. Only one study included clinical, patient-reported outcome measure (PROM), radiograph, MRI, and biochemical data.[27]

The details of the data extracted are presented comprehensively in Supplementary Table i.

There was also considerable heterogeneity in the definitions of OA progression used in these studies (Supplementary Table ii). Among the studies, four (10%) centred on pain-only progression, 23 (59%) on structural progression, and seven (18%) on both pain and structural progression. Three studies did not distinguish between pain and structural progression, and

one did not explicitly define the type of progression, but instead concentrated on identifying the presence of the ICD-9-CM code for knee OA in patient records, thus focusing more on diagnosis coding than on progression specifics.[28]

Pain progression was predominantly determined through changes in pain scores (36%; n = 14), while structural progression was defined using radiograph findings in a substantial majority (79%; n = 31). A combination of radiograph and MRI findings were employed to characterize OA structural progression in four studies (10%), while solely MRI was the basis in two articles (5%). The full details can be found in Supplementary Table ii.

All studies included in our systematic review utilized classification algorithms to predict OA progression, and, with the exception of three, all embraced supervised learning approaches. The predominant supervised classification algorithms included random forest, support vector machine, and gradient boosting model. The three studies exploring unsupervised classification harnessed DL techniques such as convolutional neural networks (CNNs) and adversarial evolving neural networks, which can learn to identify patterns and make decisions through layers of interconnected nodes. Studies extracting features from clinical images such as MRIs and radiographs employed a diverse array of techniques, with CNN being the most prevalent (Supplementary Table ii). Overall, 44% of the studies (n = 17) used DL algorithms, while only two employed automated ML (autoML), to automatically select the most appropriate classifier in a data-driven manner.[29,30] Lastly, just over half of the studies (54%, n = 21) conducted an interpretability analysis of their findings. SHapley Additive exPlanations (SHAP) emerged as the predominant interpretability tool, whereas Gradient Class Activation Map (GradCAM) was the favoured tool for image analysis interpretability.

A variety of validation methods were employed. Internal or cross-validation were performed in 92% of studies (n = 36), hold-out validation in 36% (n = 14), and external validation in 23% (n = 9). The performance metrics used also varied greatly, with 59% of studies (n = 23) using multiple metrics. By far the most common metric used was area under the receiver operating curve (AUC-ROC), which was used in 69% of studies (n = 27). Other commonly used metrics were accuracy, specificity, sensitivity/recall, precision, and F1-score (Supplementary Table ii).

Focusing on studies assessing performance via AUC-ROC, the best models achieved validation scores ranging from 0.76 to 0.97 for predicting pain-only progression, 0.56 to 0.99 for structural progression, and 0.72 to 0.87 for both pain and structural progression. AUC-ROC of models predicting either pain or structural progression (with no distinction between the two) varied between 0.66 and 0.97. Additionally, DL techniques such as CNNs yielded AUC-ROC values ranging between 0.551 and 0.99, whereas models employing traditional ML algorithms recorded scores from 0.61 to 0.97. Notably, the two autoML models reviewed achieved an AUC-ROC between 0.80 and 0.95. The best performance scores obtained in each study are illustrated in Supplementary Table iii.

Finally, only five studies (13%) were deemed to have a low risk of bias via the PROBAST tool, as described in Supplementary Table iv.

## Discussion

The application of ML in predicting OA progression is an emerging field, with the majority of studies published in the last five years. Among the 39 studies selected for this review, 36 focused on knee OA. This predilection for the knee joint may reflect the high prevalence and burden of knee OA in the general population,[71,72] necessitating technological interventions to improve understanding and management of the condition. On the other hand, the surprising paucity of studies on hip or hand OA, despite their clinical significance, highlights a clear research gap.

A substantial portion of studies relied on data from the OAI, showcasing a common tendency to use publicly available datasets for model development and validation. This reliance on a single database, while advantageous for fostering reproducible and transparent research, may limit the diversity and generalizability of the findings. Additionally, the lack of external validation (the process of testing the model's applicability to independent datasets) in most studies raises questions regarding the possible challenges in accessing other databases, or perhaps the discrepancies in data types and format between databases that may deter researchers from pursuing external validation.

Data types used for model training predominantly included clinical data, PROMs, and radiograph data. However, the underutilization of other potentially insightful data sources such as MRI, biochemical markers, genetic data, and data from wearables, reflects a significant gap in the current research landscape.

Our analysis underscores a complex environment where traditional ML methods are still prevalent despite the technological advances represented by DL and autoML. Although DL techniques showed comparable AUC-ROC values to traditional ML algorithms, they did not demonstrate a clear superiority, which could explain the cautious adoption of these more computationally demanding methods. Similarly, the modest uptake of autoML suggests that barriers such as computational demands, expertise requirements, and inertia in adopting new methods may be hindering the broader application of these advanced technologies, despite their potential to streamline and enhance certain aspects of predictive modelling in clinical research.

A crucial point of discussion emerges from the observed heterogeneity in OA progression definitions across studies, with variations ranging from joint space narrowing (JSN) measurements to increases in pain scores. For example, one study defined progression as a reduction in joint space of more than 0.5 mm over 12 months, while another focused on an increase of nine or more points in the Western Ontario and McMaster Universities Osteoarthritis Index[73] pain score over 48 months.[31,74] This variability presents substantial challenges in comparing and synthesizing results across different studies. The absence of a standardized approach could hinder the formation of a cohesive body of evidence, essential for advancing the field and informing clinical practice.[75]

Further complicating matters, the reported model performances varied widely. While some models reliably predicted specific aspects of OA progression, such as pain or structural changes, no single model excelled across all disease dimensions. This highlights the need for multi-modal models that integrate various data types using advanced techniques (such

as DL) to capture the complex nature of OA more comprehensively. Such models could potentially lead to more robust and broadly applicable predictions across different patient demographics and disease stages.

The performance metrics used to evaluate models also varied, with AUC-ROC being the most common, followed by F1 score, accuracy, sensitivity and specificity (among others). This heterogeneity in performance metrics, each emphasizing different aspects of model accuracy and reliability, highlights a lack of consensus on the most effective evaluation methods, leading to challenges in uniformly assessing and comparing the efficacy of predictive models for OA progression.

A similar situation was observed historically in oncology clinical trials, where varying metrics for assessing cancer treatment effectiveness led to challenges in comparing study outcomes. The introduction of the Response Evaluation Criteria in Solid Tumours addressed this issue by providing standardized guidelines for tumour response assessment.[74,76] This standardization significantly improved the comparability and consistency of research outcomes, and ultimately facilitated the development of new cancer therapies.

Despite some models achieving high performance, the absence of external and clinical validation in most studies limits their translational potential for clinical practice. This is reflected by the majority of studies having a high risk of bias, with only one study demonstrating a clinically validated ML application.[32] Consequently, the reliability and generalizability of these models in diverse clinical settings remain uncertain, posing a barrier to their practical application in OA management.

Another important aspect in the development of ML models for OA progression is interpretability, which is the ability to understand and explain how a model makes its predictions or decisions, and is essential for ensuring clinical acceptance and providing actionable insights, as it elucidates the model's decision-making process. Methods like kernel SHAP exemplify this approach,[77] effectively demonstrating the contributions of individual features to predictions. Such techniques could prove particularly useful in enhancing model transparency, thereby increasing their comprehensibility and clinical applicability.

In conclusion, this systematic review highlights the significant potential of ML in predicting OA progression, especially in knee OA. However, to fully harness this potential in clinical settings, several limitations need to be addressed. These include rigorous validation of ML models across diverse populations, standardization of disease progression definitions and evaluation metrics, and the incorporation of underutilized data sources such as MRI scans, biochemical markers, and wearable technology. Additionally, adoption of advanced ML techniques, such as autoML and DL, may help in developing multi-joint and multi-modal models that accurately capture the complexity and heterogeneity of OA. This important approach could realize an important translational shift of predictive models into practical, personalized, and precise clinical tools, ultimately enhancing patient care and disease management in OA.

**Take home message**
- Machine learning (ML) holds promise for predicting osteoarthritis (OA) progression, potentially guiding the design of more efficient clinical trials and informing the development of targeted therapies, particularly in the early stages of the disease.
- Existing ML models face limitations that challenge their clinical applicability, necessitating further refinement for practical use.
- Future advancements should focus on rigorous validation across diverse populations, uniformity in definitions relating to disease progression and evaluation metrics, and the incorporation of underutilized data sources such as MRI, biochemical markers, and wearable technology.
- Enhancing these methodologies could substantially accelerate the transition from predictive modelling to practical, personalized, and precise clinical tools, thereby positively impacting patient care and disease management in OA.

## Supplementary material
Comprehensive tables of all data collected during the review and the results of the risk of bias analysis.

## References
1. **Deveza LA**, **Melo L**, **Yamato TP**, **Mills K**, **Ravi V**, **Hunter DJ**. Knee osteoarthritis phenotypes and their relevance for outcomes: a systematic review. *Osteoarthr Cartilage*. 2017;25(12):1926–1941.
2. **Glyn-Jones S**, **Palmer AJR**, **Agricola R**. Osteoarthritis. *Lancet*. 2015;386(9991):376–387.
3. **Huang Z**, **Ding C**, **Li T**, **Yu SPC**. Current status and future prospects for disease modification in osteoarthritis. *Rheumatology*. 2018;57(suppl_4):iv108–iv123.
4. **Lane NE**, **Brandt K**, **Hawker G**, **et al**. OARSI-FDA initiative: defining the disease state of osteoarthritis. *Osteoarthritis Cartilage*. 2011;19(5):478–482.
5. **Loeser RF**. Aging processes and the development of osteoarthritis. *Curr Opin Rheumatol*. 2013;25(1):108–113.
6. **Stewart HL**, **Kawcak CE**. The importance of subchondral bone in the pathophysiology of osteoarthritis. *Front Vet Sci*. 2018;5:178.
7. **Dobson GP**, **Letson HL**, **Grant A**, **et al**. Defining the osteoarthritis patient: back to the future. *Osteoarthr Cartilage*. 2018;26(8):1003–1007.
8. **Bijlsma JW**, **Berenbaum F**, **Lafeber FP**. Osteoarthritis: an update with relevance for clinical practice. *Lancet*. 2011;377(9783):2115–2126.
9. **Samuel AL**. Some studies in machine learning using the game of checkers. *IBM J Res & Dev*. 1959;3(3):210–229.
10. **Castagno S**, **Birch M**, **van der Schaar M**, **McCaskie A**. A precision health approach for osteoarthritis: prediction of rapid knee osteoarthritis progression using automated machine learning. *Bone Joint J*. 2024;106-B(SUPP_2):19.
11. **No authors listed**. What is computer vision? IBM. https://www.ibm.com/topics/computer-vision (date last accessed 21 August 2024).
12. **No authors listed**. What is NLP (natural language processing)? IBM. 2021. https://www.ibm.com/cloud/learn/natural-language-processing (date last accessed 21 August 2024).
13. **Ashraf M**, **Khalilitousi M**, **Laksman Z**. Applying machine learning to stem cell culture and differentiation. *Curr Protoc*. 2021;1(9):e261.
14. **Badillo S**, **Banfai B**, **Birzele F**, **et al**. An introduction to machine learning. *Clin Pharmacol Ther*. 2020;107(4):871–885.
15. **Choi RY**, **Coyner AS**, **Kalpathy-Cramer J**, **Chiang MF**, **Campbell JP**. Introduction to machine learning, neural networks, and deep learning. *Transl Vis Sci Technol*. 2020;9(2):14.
16. **Dridi S**. Unsupervised learning - a systematic literature review. *Open Science Framework*. 2022. https://osf.io/kpqr6
17. **Wittek P**. Unsupervised Learning. In: *Quantum Machine Learning*. 1st edition. Boston, Massachussetts, USA: Elsevier Insights, 2014: 57–62.
18. **LeCun Y**, **Bengio Y**, **Hinton G**. Deep learning. *Nature*. 2015;521(7553):436–444.
19. **Page MJ**, **McKenzie JE**, **Bossuyt PM**, **et al**. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *PLoS Med*. 2021;18(3):e1003583.
20. **Langenberger B**, **Schrednitzki D**, **Halder AM**, **Busse R**, **Pross CM**. Predicting whether patients will achieve minimal clinically important differences following hip or knee arthroplasty. *Bone Joint Res*. 2023;12(9):512–521.
21. **Salis Z**, **Driban JB**, **McAlindon TE**. Predicting the onset of end-stage knee osteoarthritis over two- and five-years using machine learning. *Semin Arthritis Rheum*. 2024;66:152433.

22. **Moons KGM**, **Wolff RF**, **Riley RD**, **et al**. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med.* 2019;170(1):W1.

23. **Abdulazeem H**, **Whitelaw S**, **Schauberger G**, **Klug SJ**. A systematic review of clinical health conditions predicted by machine learning diagnostic and prognostic models trained or validated using real-world primary health care data. *PLoS One.* 2023;18(9):e0274276.

24. **Lu Y**, **Pareek A**, **Wilbur RR**, **Leland DP**, **Krych AJ**, **Camp CL**. Understanding anterior shoulder instability through machine learning: new models that predict recurrence, progression to surgery, and development of arthritis. *Orthop J Sports Med.* 2021;9(11):232596712110533.

25. **Al Turkestani N**, **Li T**, **Bianchi J**, **et al**. A comprehensive patient-specific prediction model for temporomandibular joint osteoarthritis progression. *Proc Natl Acad Sci USA.* 2024;121(8):e2306132121.

26. **Nielsen RL**, **Monfeuga T**, **Kitchen RR**, **et al**. Data-driven identification of predictive risk biomarkers for subgroups of osteoarthritis using interpretable machine learning. *Nat Commun.* 2024;15(1):2817.

27. **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. *Osteoarthr Cartilage.* 2017;25(12):2014–2021.

28. **Ningrum DNA**, **Kung W-M**, **Tzeng I-S**, **et al**. A deep learning model to predict knee osteoarthritis based on nonimage longitudinal medical record. *J Multidiscip Healthc.* 2021;14:2477–2485.

29. **Jamshidi A**, **Leclercq M**, **Labbe A**, **et al**. Identification of the most important features of knee osteoarthritis structural progressors using machine learning methods. *Ther Adv Musculoskelet Dis.* 2020;12:1759720X20933468.

30. **Chen T**, **Or CK**. Automated machine learning-based prediction of the progression of knee pain, functional decline, and incidence of knee osteoarthritis in individuals at high risk of knee osteoarthritis: data from the osteoarthritis initiative study. *Digit Health.* 2023;9:20552076231216419.

31. **Schiratti J-B**, **Dubois R**, **Herent P**, **et al**. A deep learning method for predicting knee osteoarthritis radiographic progression from MRI. *Arthritis Res Ther.* 2021;23(1):262.

32. **Widera P**, **Welsing PMJ**, **Danso SO**, **et al**. Development and validation of a machine learning-supported strategy of patient selection for osteoarthritis clinical trials: the IMI-APPROACH study. *Osteoarthr Cartil Open.* 2023;5(4):100406.

33. **Bayramoglu N**, **Englund M**, **Haugen IK**, **Ishijima M**, **Saarakkala S**. Deep learning for predicting progression of patellofemoral osteoarthritis based on lateral knee radiographs, demographic data, and symptomatic assessments. *Methods Inf Med.* 2024.

34. **Nguyen HH**, **Blaschko MB**, **Saarakkala S**, **Tiulpin A**. Clinically-inspired multi-agent transformers for disease trajectory forecasting from multimodal data. *IEEE Trans Med Imaging.* 2024;43(1):529–541.

35. **Dunn CM**, **Sturdy C**, **Velasco C**, **et al**. Peripheral blood DNA methylation-based machine learning models for prediction of knee osteoarthritis progression: biologic specimens and data from the Osteoarthritis Initiative and Johnston County Osteoarthritis Project. *Arthritis Rheumatol.* 2023;75(1):28–40.

36. **Hu J**, **Zheng C**, **Yu Q**, **et al**. DeepKOA: a deep-learning model for predicting progression in knee osteoarthritis using multimodal magnetic resonance images from the osteoarthritis initiative. *Quant Imaging Med Surg.* 2023;13(8):4852–4866.

37. **Shen L**, **Yue S**. A clinical model to predict the progression of knee osteoarthritis: data from Dryad. *J Orthop Surg Res.* 2023;18(1):628.

38. **Yin R**, **Chen H**, **Tao T**, **et al**. Expanding from unilateral to bilateral: a robust deep learning-based approach for predicting radiographic osteoarthritis progression. *Osteoarthr Cartilage.* 2024;32(3):338–347.

39. **Yoo HJ**, **Jeong HW**, **Kim SW**, **Kim M**, **Lee JI**, **Lee YS**. Prediction of progression rate and fate of osteoarthritis: comparison of machine learning algorithms. *J Orthop Res.* 2023;41(3):583–590.

40. **Almhdie-Imjabbar A**, **Nguyen KL**, **Toumi H**, **Jennane R**, **Lespessailles E**. Prediction of knee osteoarthritis progression using radiological descriptors obtained from bone texture analysis and Siamese neural networks: data from OAI and MOST cohorts. *Arthritis Res Ther.* 2022;24(1):66.

41. **Bonakdari H**, **Pelletier J-P**, **Blanco FJ**, **et al**. Single nucleotide polymorphism genes and mitochondrial DNA haplogroups as biomarkers for early prediction of knee osteoarthritis structural progressors: use of supervised machine learning classifiers. *BMC Med.* 2022;20(1):316.

42. **Bonakdari H**, **Pelletier JP**, **Abram F**, **Martel-Pelletier J**. A machine learning model to predict knee osteoarthritis cartilage volume changes over time using baseline bone curvature. *Biomedicines.* 2022;10(6):1247.

43. **Guan B**, **Liu F**, **Mizaian AH**, **et al**. Deep learning approach to predict pain progression in knee osteoarthritis. *Skeletal Radiol.* 2022;51(2):363–373.

44. **Hu K**, **Wu W**, **Li W**, **Simic M**, **Zomaya A**, **Wang Z**. Adversarial evolving neural network for longitudinal knee osteoarthritis prediction. *IEEE Trans Med Imaging.* 2022;41(11):3207–3217.

45. **Joseph GB**, **McCulloch CE**, **Nevitt MC**, **Link TM**, **Sohn JH**. Machine learning to predict incident radiographic knee osteoarthritis over 8 years using combined MR imaging features, demographics, and clinical factors: data from the Osteoarthritis Initiative. *Osteoarthr Cartilage.* 2022;30(2):270–279.

46. **Bonakdari H**, **Jamshidi A**, **Pelletier J-P**, **Abram F**, **Tardif G**, **Martel-Pelletier J**. A warning machine learning algorithm for early knee osteoarthritis structural progressor patient screening. *Ther Adv Musculoskelet Dis.* 2021;13:1759720X21993254.

47. **Chan LC**, **Li HHT**, **Chan PK**, **Wen C**. A machine learning-based approach to decipher multi-etiology of knee osteoarthritis onset and deterioration. *Osteoarthr Cartil Open.* 2021;3(1):100135.

48. **Cheung JCW**, **Tam AYC**, **Chan LC**, **Chan PK**, **Wen C**. Superiority of multiple-joint space width over minimum-joint space width approach in the machine learning for radiographic severity and knee osteoarthritis progression. *Biology (Basel).* 2021;10(11):1107.

49. **Lee JJ**, **Liu F**, **Majumdar S**, **Pedoia V**. An ensemble clinical and MR-image deep learning model predicts 8-year knee pain trajectory: data from the osteoarthritis initiative. *Osteoarthritis Imaging.* 2021;1:100003.

50. **Ntakolia C**, **Kokkotis C**, **Moustakidis S**, **Tsaopoulos D**. Identification of most important features based on a fuzzy ensemble technique: evaluation on joint space narrowing progression in knee osteoarthritis patients. *Int J Med Inform.* 2021;156:104614.

51. **Ntakolia C**, **Kokkotis C**, **Moustakidis S**, **Tsaopoulos D**. Prediction of joint space narrowing progression in knee osteoarthritis patients. *Diagnostics (Basel).* 2021;11(2):285.

52. **Guan B**, **Liu F**, **Haj-Mirzaian A**, **et al**. Deep learning risk assessment models for predicting progression of radiographic medial joint space loss over a 48-month follow-up period. *Osteoarthr Cartilage.* 2020;28(4):428–437.

53. **Kundu S**, **Ashinsky BG**, **Bouhrara M**, **et al**. Enabling early detection of osteoarthritis from presymptomatic cartilage texture maps via transport-based learning. *Proc Natl Acad Sci USA.* 2020;117(40):24709–24719.

54. **Morales Martinez A**, **Caliva F**, **Flament I**, **et al**. Learning osteoarthritis imaging biomarkers from bone surface spherical encoding. *Magn Reson Med.* 2020;84(4):2190–2203.

55. **Wang Y**, **You L**, **Chyr J**, **et al**. Causal discovery in radiographic markers of knee osteoarthritis and prediction for knee osteoarthritis severity with attention-long short-term memory. *Front Public Health.* 2020;8:604654.

56. **Widera P**, **Welsing PMJ**, **Ladel C**, **et al**. Multi-classifier prediction of knee osteoarthritis progression from incomplete imbalanced longitudinal data. *Sci Rep.* 2020;10(1):8427.

57. **Tiulpin A**, **Klein S**, **Bierma-Zeinstra SMA**, **et al**. Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. *Sci Rep.* 2019;9(1):20038.

58. **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):2243–2250.

59. **Hafezi-Nejad N**, **Guermazi A**, **Roemer FW**, **et al**. Prediction of medial tibiofemoral compartment joint space loss progression using volumetric cartilage measurements: data from the FNIH OA biomarkers consortium. *Eur Radiol.* 2017;27(2):464–473.

60. **Marques J**, **Genant HK**, **Lillholm M**, **Dam EB**. Diagnosis of osteoarthritis and prognosis of tibial cartilage loss by quantification of tibia trabecular bone from MRI. *Magn Reson Med.* 2013;70(2):568–575.

61. **Woloszynski T**, **Podsiadlo P**, **Stachowiak G**, **Kurzynski M**. A dissimilarity-based multiple classifier system for trabecular bone texture in detection and prediction of progression of knee osteoarthritis. *Proc Inst Mech Eng H.* 2012;226(11):887–894.

62. **Wirth W**, **Hunter DJ**, **Nevitt MC**, **et al**. Predictive and concurrent validity of cartilage thickness change as a marker of knee osteoarthritis progression: data from the Osteoarthritis Initiative. *Osteoarthr Cartilage.* 2017;25(12):2063–2071.

63. **Segal NA**, **Nevitt MC**, **Gross KD**, **et al**. The Multicenter Osteoarthritis Study: opportunities for rehabilitation research. *PM R.* 2013;5(8):647–654.

64. **Wesseling J**, **Boers M**, **Viergever MA**, **et al**. Cohort profile: Cohort Hip and Cohort Knee (CHECK) study. *Int J Epidemiol.* 2016;45(1):36–44.

65. **Damman W**, **Liu R**, **Kroon FPB**, **et al**. Do comorbidities play a role in hand osteoarthritis disease burden? Data from the Hand Osteoarthritis in Secondary Care Cohort. *J Rheumatol.* 2017;44(11):1659–1666.

66. **Sellam J**, **Maheu E**, **Crema MD**, **et al**. The DIGICOD cohort: a hospital-based observational prospective cohort of patients with hand osteoarthritis-methodology and baseline characteristics of the population. *Joint Bone Spine*. 2021;88(4):105171.

67. **Oreiro-Villar N**, **Raga AC**, **Rego-Pérez I**, **et al**. Descripción de la cohorte PROCOAC (PROspective COhort of A CoruñA): Cohorte prospectiva española para el estudio de la osteoartritis. *Reum Clín*. 2022;18(2):100–104. [Article in Spanish].

68. **Østerås N**, **Risberg MA**, **Kvien TK**, **et al**. Hand, hip and knee osteoarthritis in a Norwegian population-based study--the MUST protocol. *BMC Musculoskelet Disord*. 2013;14(1):1–16.

69. **Runhaar J**, **van Middelkoop M**, **Reijman M**, **et al**. Prevention of knee osteoarthritis in overweight females: the first preventive randomized controlled trial in osteoarthritis. *Am J Med*. 2015;128(8):888–895.

70. **Kremers HM**, **Myasoedova E**, **Crowson CS**, **Savova G**, **Gabriel SE**, **Matteson EL**. The Rochester Epidemiology Project: exploiting the capabilities for population-based research in rheumatic diseases. *Rheumatology*. 2011;50(1):6–15.

71. **Hunter DJ**, **Bierma-Zeinstra S**. Osteoarthritis. *Lancet*. 2019;393(10182):1745–1759.

72. **Michael JWP**, **Schlüter-Brust KU**, **Eysel P**. The epidemiology, etiology, diagnosis, and treatment of osteoarthritis of the knee. *Dtsch Arztebl Int*. 2010;107(9):152–162.

73. **Bellamy N**, **Buchanan WW**, **Goldsmith CH**, **Campbell J**, **Stitt LW**. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol*. 1988;15(12):1833–1840.

74. **Therasse P**, **Arbuck SG**, **Eisenhauer EA**, **et al**. New guidelines to evaluate the response to treatment in solid tumors. *JNCI*. 2000;92(3):205–216.

75. **Kunze KN**, **Orr M**, **Krebs V**, **Bhandari M**, **Piuzzi NS**. Potential benefits, unintended consequences, and future roles of artificial intelligence in orthopaedic surgery research : a call to emphasize data quality and indications. *Bone Jt Open*. 2022;3(1):93–97.

76. **Eisenhauer EA**, **Therasse P**, **Bogaerts J**, **et al**. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228–247.

77. **Lundberg SM**, **Lee SI**. A unified approach to interpreting model predictions: advances in neural information processing systems. NeurIPS Proceedings. https://papers.nips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf (date last accessed 21 August 2024).

**Author information:**
S. Castagno, MBBS, BEng, Orthopaedic Surgeon, Research Fellow
B. Gompels, MBBS, PhD Student
E. Strangmark, BA, MPhil Student
E. Robertson-Waters, MBBS, Orthopaedic Surgeon
M. Birch, PhD, Deputy Research Director
A. W. McCaskie, MD, FRCS, Professor of Orthopaedic Surgery, Interim Deputy Head of School of Clinical Medicine
Department of Surgery, University of Cambridge, Cambridge, UK.

M. van der Schaar, PhD, John Humphrey Plummer Professor of Machine Learning, Artificial Intelligence and Medicine, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK.

**Author contributions:**
S. Castagno: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing.
B. Gompels: Data curation, Investigation, Validation.
E. Strangmark: Data curation, Investigation, Validation.
E. Robertson-Waters: Data curation, Investigation, Validation.
M. Birch: Conceptualization, Supervision, Writing – review & editing.
M. van der Schaar: Project administration, Supervision.
A. W. McCaskie: Conceptualization, Project administration, Supervision, Writing – review & editing.