

INTELIGÊNCIA ARTIFICIAL NA PREDIÇÃO DA PROGRESSÃO ESTRUTURAL DA OSTEOARTRITE DE JOELHO: DESEMPENHO, RISCO DE VIÉS E APLICABILIDADE CLÍNICA

ARTIFICIAL INTELLIGENCE IN PREDICTING STRUCTURAL PROGRESSION OF KNEE OSTEOARTHRITIS: PERFORMANCE, RISK OF BIAS, AND CLINICAL APPLICABILITY

Diego dos Santos Cahet¹; Roberta Monik Cavalcante Lima Cahet¹; Victória Khrystty Calheiros de Amorim¹; Bruna Gabriela dos Santos Silva¹; Silvia Letícia da Conceição Abreu¹; Jaim Simões de Oliveira²

1 Acadêmicos do curso de Medicina do Centro Universitário Afya Maceió; 2 Professor do curso de Medicina do Centro Universitário Afya Maceió.

E-mail do autor principal e correspondente: jaim.oliveira@afya.com.br.

Eixo Temático: 01 - Saúde

RESUMO

Introdução: A progressão estrutural da osteoartrite (OA) do joelho apresenta grande variabilidade individual, dificultando a identificação de pacientes com maior risco de deterioração. **Objetivo:** Avaliar o desempenho preditivo, o risco de viés e a aplicabilidade clínica de modelos de inteligência artificial (IA) destinados à predição da progressão estrutural da OA do joelho em adultos. **Método:** Revisão integrativa com busca estruturada da literatura até 15 de agosto de 2026, priorizando modelos prognósticos de machine learning ou deep learning com desfechos longitudinais, como progressão de Kellgren-Lawrence, estreitamento do espaço articular e perda de cartilagem. A análise crítica foi orientada pelo PROBAST+AI e por critérios de aplicabilidade clínica. **Resultados:** A literatura mostrou discriminação geralmente moderada a boa, com desempenho superior em alguns modelos baseados em ressonância magnética e biomarcadores. Entretanto, observou-se forte dependência da base Osteoarthritis Initiative, baixa frequência de validação externa, heterogeneidade dos desfechos e relato insuficiente de calibração. **Conclusão:** Embora a IA tenha potencial para estratificação de risco e enriquecimento de ensaios clínicos, seu uso assistencial permanece limitado. Validação externa, calibração, padronização dos desfechos e estudos prospectivos de impacto clínico são prioridades para sua translação à prática.

Palavras-chave: Osteoartrite do Joelho; Inteligência Artificial; Prognóstico.

ABSTRACT

Introduction: Structural progression of knee osteoarthritis (OA) shows substantial individual variability, making it difficult to identify patients at higher risk of deterioration. **Objective:** To evaluate the predictive performance, risk of bias, and clinical applicability of artificial intelligence (AI) models designed to predict structural progression of knee OA in adults. **Methods:** Integrative review with a structured literature search through August 15, 2026, prioritizing prognostic machine learning or deep learning models using longitudinal outcomes such as Kellgren-Lawrence progression, joint space narrowing, and cartilage loss. Critical appraisal was guided by PROBAST+AI and clinical applicability criteria. **Results:** The literature showed generally

moderate-to-good discrimination, with higher performance in some magnetic resonance imaging- and biomarker-based models. However, strong dependence on the Osteoarthritis Initiative dataset, infrequent external validation, heterogeneous outcomes, and insufficient reporting of calibration were observed.

Conclusion: Although AI has potential for risk stratification and clinical trial enrichment, routine clinical use remains limited. External validation, calibration, outcome standardization, and prospective studies of clinical impact are priorities for translation into practice.

Keywords: Osteoarthritis, Knee; Artificial Intelligence; Prognosis.

INTRODUÇÃO

A osteoartrite (OA) do joelho é uma condição estrutural e clínica heterogênea. Pacientes com achados basais semelhantes podem apresentar trajetórias distintas de estreitamento do espaço articular, progressão radiográfica e perda de cartilagem. Essa variabilidade dificulta a identificação dos indivíduos que evoluirão mais rapidamente e reduz a eficiência de ensaios de terapias modificadoras da doença. Nesse cenário, modelos de inteligência artificial (IA), incluindo machine learning e deep learning, vêm sendo utilizados para integrar dados clínicos, radiografias, ressonância magnética (RM) e biomarcadores com o objetivo de estimar o risco futuro de progressão. Revisões recentes demonstram crescimento expressivo do campo, porém também evidenciam heterogeneidade de desfechos, fontes de dados e estratégias de validação.^{1,2}

A comparação entre modelos não deve se restringir à área sob a curva ROC (AUC). Um algoritmo pode apresentar boa discriminação em uma amostra interna e perder desempenho quando aplicado a outra população, equipamento de imagem ou contexto assistencial. Para utilização clínica, também são relevantes calibração, validação externa, disponibilidade dos preditores, transparência metodológica e demonstração de benefício para decisões. Assim, esta revisão teve como pergunta norteadora: qual é o desempenho preditivo, o risco de viés e a aplicabilidade clínica dos modelos de IA desenvolvidos para prever a progressão estrutural da OA de joelho em adultos? O objetivo foi sintetizar o desempenho desses modelos, identificar suas principais fontes de viés e discutir sua prontidão para aplicação clínica.

MÉTODO

Foi realizada revisão integrativa da literatura, com busca estruturada até 15 de agosto de 2026. A estratégia combinou os conceitos knee osteoarthritis, artificial intelligence, machine learning, deep learning, prediction, prognosis, progression, joint space narrowing, joint space width, Kellgren-Lawrence e cartilage loss. A busca principal foi realizada no PubMed/MEDLINE e complementada pelo rastreamento de referências de revisões sistemáticas contemporâneas que incluíram bases como Embase, Scopus, Web of Science e

Cochrane Library. A organização do relato considerou princípios do PRISMA 2020 e do TRIPOD-SRMA.^{3,4}

Foram elegíveis estudos em adultos que desenvolveram ou avaliaram modelos de IA destinados à predição de um desfecho estrutural futuro do joelho. Consideraram-se progressão do grau de Kellgren-Lawrence, surgimento de OA radiográfica, estreitamento ou redução da largura do espaço articular e perda de volume ou espessura de cartilagem. Foram excluídos estudos restritos à classificação contemporânea da OA, segmentação de imagens, previsão exclusiva de dor ou função e previsão isolada de artroplastia, por esta depender também de sintomas, acesso ao sistema e decisão terapêutica. Extraíram-se fonte dos dados, preditores, algoritmo, desfecho, horizonte temporal, estratégia de validação e métricas de desempenho. O risco de viés e a aplicabilidade foram analisados segundo os domínios conceituais do PROBAST+AI, e a transparência do relato foi apreciada à luz do TRIPOD+AI.^{5,6} Devido à heterogeneidade dos estudos e à sobreposição entre coortes, os resultados foram integrados por meio de síntese qualitativa estruturada e comparativa, sem realização de metanálise própria.

RESULTADOS

As revisões recentes indicam que os modelos de IA discriminam progressão estrutural acima do acaso, mas sem desempenho uniforme. Liu et al. incluíram 32 estudos: oito apresentaram baixo risco de viés, 13 alto risco e 11 risco incerto. Os índices de concordância agrupados variaram conforme a fonte de dados, alcançando aproximadamente 0,72 para radiografia, 0,80 para RM e 0,82 para combinações de dados clínicos, radiografia e RM; os amplos intervalos de predição sugeriram incerteza sobre a generalização para novos cenários.¹

Em revisão de deep learning publicada em 2026, 33 estudos e 50 modelos foram analisados, dos quais 14 avaliavam deterioração estrutural. A AUC mediana dos modelos estruturais foi próxima de 0,78. Entretanto, 97% dos estudos utilizaram a OAI no treinamento e apenas 27% realizaram validação externa. Nos modelos externamente avaliados, a AUC mediana caiu de 0,81 na avaliação interna para 0,75 em dados externos. Modelos baseados em RM exibiram desempenho interno mais alto, porém sem validação externa na síntese analisada. O domínio de análise concentrou a maior proporção de risco alto ou incerto, especialmente por validação insuficiente e possibilidade de sobreajuste.²

Os estudos primários reforçam esse padrão. Modelos radiográficos podem alcançar AUC superior a 0,80 em avaliações internas.⁷ Em RM, Schiratti et al. relataram AUC de 0,65 para progressão de estreitamento do espaço articular em 12 meses, ilustrando a dificuldade de

prever eventos estruturais em horizontes curtos.⁸ Em contraste, Almhdie-Imjabbar et al. testaram um modelo de textura óssea e rede siamesa entre as coortes OAI e MOST, mantendo AUC acima de 0,70 em cenários cruzados, resultado metodologicamente relevante por demonstrar transportabilidade.⁹ Modelos moleculares recentes também apresentaram desempenho elevado: um modelo metabolômico com seis metabólitos séricos, idade e sexo obteve AUC de 0,98 no teste e 0,89 em coorte externa independente.¹⁰

Apesar dos resultados promissores, quatro limitações foram recorrentes: dependência de poucas bases de dados, validação externa infrequente, definições não uniformes de progressão e relato incompleto de calibração e utilidade clínica. A reutilização da OAI por diferentes grupos pode produzir uma aparência de replicação sem verdadeira independência populacional. Além disso, a AUC descreve discriminação, mas não demonstra se as probabilidades previstas são bem calibradas ou se o modelo melhora decisões em um limiar clínico definido.

Os achados mostram que o desempenho preditivo da IA para progressão estrutural da OA do joelho é promissor, mas ainda insuficiente para uso rotineiro. O resultado mais relevante não é a obtenção isolada de AUCs elevadas, e sim a redução de desempenho quando os modelos são avaliados fora do ambiente de desenvolvimento. A queda de AUC interna para externa observada nas sínteses recentes evidencia que a generalização permanece um dos principais obstáculos.²

Outro problema é a própria definição de progressão. Aumento de Kellgren-Lawrence, redução da largura do espaço articular, estreitamento ordinal e perda de cartilagem por RM medem componentes relacionados, porém não intercambiáveis. Misturá-los dificulta a comparação entre algoritmos. O risco de viés concentra-se especialmente na análise, pois modelos de alta dimensionalidade, pequeno número de progressores, seleção extensa de variáveis e validação no mesmo ecossistema de dados aumentam a chance de sobreajuste. O PROBAST+AI foi atualizado para avaliar especificamente essas fontes de viés em modelos desenvolvidos por regressão ou IA, enquanto o TRIPOD+AI orienta a transparência do relato.^{5,6}

A modalidade de entrada com maior AUC interna também não é necessariamente a mais aplicável. A RM fornece informação estrutural detalhada, mas exige maior custo, padronização de aquisição e validação entre equipamentos. Radiografias são amplamente disponíveis e podem facilitar implementação em escala, ainda que apresentem desempenho discretamente inferior. Biomarcadores moleculares oferecem potencial para identificar risco antes de alterações radiográficas avançadas, porém dependem de infraestrutura laboratorial e validação multicêntrica.¹⁰ No curto prazo, a aplicação mais plausível desses modelos é a

estratificação de risco e o enriquecimento de ensaios clínicos, selecionando participantes com maior probabilidade de progressão. Para uso individual, ainda são necessários estudos prospectivos multicêntricos, avaliação de calibração, curvas de decisão, comparação com modelos clínicos simples e demonstração de impacto sobre condutas e desfechos.

Esta revisão apresenta limitações inerentes ao desenho integrativo: não substitui uma revisão sistemática com busca independente em todas as bases, dupla triagem e metanálise própria. A síntese quantitativa foi apoiada em revisões sistemáticas recentes e complementada por estudos primários representativos. Além disso, a forte sobreposição da OAI impede considerar cada publicação como evidência estatisticamente independente. Ainda assim, a análise integrada de desempenho, risco de viés e aplicabilidade permite destacar que a translação clínica depende mais de transportabilidade e validação rigorosa do que de novos recordes de AUC.

CONCLUSÃO

Os modelos de IA destinados à predição da progressão estrutural da OA do joelho apresentam discriminação geralmente moderada a boa, com resultados elevados em alguns modelos baseados em RM e biomarcadores. Contudo, a dependência de poucas coortes, a escassez de validação externa, a heterogeneidade dos desfechos e o relato insuficiente de calibração limitam sua aplicabilidade clínica. A evidência atual sustenta a IA como ferramenta promissora para estratificação de risco e pesquisa, mas ainda não como instrumento pronto para decisão assistencial rotineira. A transição para a prática dependerá de validação externa multicêntrica, padronização dos desfechos, transparência metodológica e demonstração prospectiva de benefício clínico.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Liu Y, Xiao G, Zhang Y, et al. Predictive value of machine learning in knee osteoarthritis progression: systematic review and meta-analysis. *J Med Internet Res*. 2025;27:e80430. doi:10.2196/80430.
2. Gillani A, von Eisenhart-Rothe R, Woertler K, Hirschmann MT, Rueckert D, Hinterwimmer F. Promises and limitations of deep learning for predicting knee osteoarthritis progression from medical imaging: a systematic review. *Knee Surg Sports Traumatol Arthrosc*. 2026. doi:10.1002/ksa.70555.
3. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71.

4. Snell KIE, Levis B, Damen JAA, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). *BMJ*. 2023;381:e073538. doi:10.1136/bmj-2022-073538.
5. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*. 2025;388:e082505. doi:10.1136/bmj-2024-082505.
6. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378.
7. 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. *Osteoarthritis Cartilage*. 2020;28(4):428-437.
8. Schiratti JB, 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. doi:10.1186/s13075-021-02634-4.
9. 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. doi:10.1186/s13075-022-02743-8.
10. Jamshidi A, Zhai G, Liu M, et al. An interpretable metabolomic-driven machine learning model for early prediction of knee structural OA progression. *Rheumatology (Oxford)*. 2026;65(1):keaf686. doi:10.1093/rheumatology/keaf686.