

Desenvolvimento de rede neural convolucional para o diagnóstico radiográfico de osteoartrite dos joelhos no ELSA-Brasil Musculoesquelético

Development of a convolutional neural network for diagnosing osteoarthritis, trained with knee radiographs from the ELSA-Brasil Musculoskeletal

Júlio Guerra Domingues^{1,a}, Daniella Castro Araujo^{2,3,b}, Luciana Costa-Silva^{4,c}, Alexei Manso Corrêa Machado^{1,d}, Luciana Andrade Carneiro Machado^{5,e}, Adriano Alonso Veloso^{2,f}, Sandhi Maria Barreto^{1,5,g}, Rosa Weiss Telles^{1,5,h}

1. Faculdade de Medicina da Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil. 2. Instituto de Ciências Exatas da Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil. 3. Huna-AI, São Paulo, SP, Brasil. 4. Instituto Hermes Pardini, Belo Horizonte, MG, Brasil. 5. Hospital das Clínicas da Universidade Federal de Minas Gerais (UFMG)/Empresa Brasileira de Serviços Hospitalares (EBSERH), Belo Horizonte, MG, Brasil.

Correspondência: Dr. Júlio Guerra Domingues. Departamento de Anatomia e Imagem – Universidade Federal de Minas Gerais. Avenida Professor Alfredo Balena, 190, sala 179, Santa Efigênia. Belo Horizonte, MG, Brasil, 30130-100. E-mail: juliogdomingues@gmail.com.

a. <https://orcid.org/0000-0002-5542-3380>; b. <https://orcid.org/0000-0002-2748-4776>; c. <https://orcid.org/0000-0003-0034-7285>; d. <https://orcid.org/0000-0001-8077-3377>; e. <https://orcid.org/0000-0001-6303-2753>; f. <https://orcid.org/0000-0002-9177-4954>; g. <https://orcid.org/0000-0001-7383-7811>; h. <https://orcid.org/0000-0003-4027-2943>.

Submetido em 1/3/2023. Revisado em 5/5/2023. Aceito em 19/7/2023.

Como citar este artigo:

Domingues JG, Araujo DC, Costa-Silva L, Machado AMC, Machado LAC, Veloso AA, Barreto SM, Telles RW. Desenvolvimento de rede neural convolucional para o diagnóstico radiográfico de osteoartrite dos joelhos no ELSA-Brasil Musculoesquelético. Radiol Bras. 2023 Set/Out;56(5):248–254.

Resumo **Objetivo:** Desenvolver um modelo computacional – rede neural convolucional (RNC) – treinado com radiografias da linha de base do Estudo Longitudinal de Saúde do Adulto Musculoesquelético (ELSA-Brasil Musculoesquelético), para a classificação automática de osteoartrite dos joelhos.

Materiais e Métodos: Trata-se de um estudo transversal abrangendo todos os exames da linha de base do ELSA-Brasil Musculoesquelético (5.660 radiografias dos joelhos em incidência posteroanterior). Os exames foram interpretados por médico radiologista com treinamento específico e calibração previamente publicada.

Resultados: A RNC desenvolvida apresentou área sob a curva característica de operação do receptor de 0,866 (IC 95%: 0,842–0,882). O modelo pode ser calibrado para alcançar, não simultaneamente, valores máximos de 0,907 para acurácia, 0,938 para sensibilidade e 0,994 para especificidade.

Conclusão: A RNC desenvolvida pode ser utilizada como ferramenta de triagem, reduzindo o número total de exames avaliados pelos radiologistas do estudo, e/ou como ferramenta de segunda leitura, contribuindo com a redução de possíveis erros de interpretação.

Unitermos: Osteoartrite do joelho; Radiografia; Redes neurais de computação; Aprendizado de máquina; Diagnóstico por computador; Estudos epidemiológicos.

Abstract **Objective:** To develop a convolutional neural network (CNN) model, trained with the Brazilian “Estudo Longitudinal de Saúde do Adulto Musculoesquelético” (ELSA-Brasil MSK, Longitudinal Study of Adult Health, Musculoskeletal) baseline radiographic examinations, for the automated classification of knee osteoarthritis.

Materials and Methods: This was a cross-sectional study carried out with 5,660 baseline posteroanterior knee radiographs from the ELSA-Brasil MSK database (5,660 baseline posteroanterior knee radiographs). The examinations were interpreted by a radiologist with specific training, and the calibration was as established previously.

Results: The CNN presented an area under the receiver operating characteristic curve of 0.866 (95% CI: 0.842–0.882). The model can be optimized to achieve, not simultaneously, maximum values of 0.907 for accuracy, 0.938 for sensitivity, and 0.994 for specificity.

Conclusion: The proposed CNN can be used as a screening tool, reducing the total number of examinations evaluated by the radiologists of the study, and as a double-reading tool, contributing to the reduction of possible interpretation errors.

Keywords: Osteoarthritis, knee; Radiography; Neural networks, computer; Machine learning; Diagnosis, computer-assisted; Epidemiologic studies.

INTRODUÇÃO

A osteoartrite é um dos problemas de saúde mais prevalentes na população mundial, especialmente em idosos⁽¹⁾. A osteoartrite dos joelhos destaca-se não apenas pela

alta prevalência, mas também pela morbidade associada, sendo uma das principais causas de anos vividos com incapacidade no mundo⁽²⁾. Identificou-se osteoartrite dos joelhos radiográfica em 18,1% dos participantes do maior

estudo longitudinal para pesquisa de doenças musculoesqueléticas do Brasil⁽³⁾.

A osteoartrite dos joelhos pode causar dor, rigidez articular, redução na amplitude de movimento e fraqueza muscular⁽⁴⁾. As consequências no longo prazo incluem redução no grau de atividade física, alterações do sono, depressão e incapacidade⁽⁴⁾. Ressaltam-se as repercussões econômicas e sociais relacionadas à osteoartrite, que incluem custos diretos (como tratamentos e cirurgias), indiretos (absenteísmo, menor empregabilidade e aposentadoria precoce) e intangíveis (dor, redução da qualidade de vida e menor participação social)⁽⁵⁾. Estima-se que os custos totais relacionados à osteoartrite podem chegar a 1% a 2,5% do produto interno bruto de países desenvolvidos⁽⁶⁾, com tendência de elevação de tais custos, decorrente do aumento na prevalência de sobrepeso e obesidade e do envelhecimento da população⁽⁵⁾.

O diagnóstico da osteoartrite dos joelhos pode ser realizado a partir de critérios clínicos e radiográficos, em associação ou não, sendo o critério radiográfico considerado um dos mais sensíveis⁽⁷⁾. Em estudos epidemiológicos longitudinais, o diagnóstico é usualmente realizado a partir de achados das radiografias de joelhos⁽⁸⁾, sendo o sistema de gradação mais utilizado o de Kellgren e Lawrence (KL)⁽⁹⁾. Joelhos com KL 0 e 1 indicam ausência de osteoartrite dos joelhos definitiva, enquanto escores KL 2, 3 e 4 indicam a presença de osteoartrite.

A classificação das radiografias em estudos longitudinais é usualmente realizada por médicos especialistas e requer treinamento, padronização e calibração rigorosos⁽³⁾. A análise das imagens consiste na gradação semiquantitativa dos osteófitos e espaços articulares, de acordo com atlas radiográfico. Em pesquisas de larga escala, tal processo se torna demasiadamente demorado e oneroso, estando sujeito ao grau de experiência dos observadores. Por esta razão, estudos têm sido desenvolvidos com o objetivo de verificar a viabilidade na utilização de modelos computacionais para classificação automática e semiautomática da osteoartrite dos joelhos⁽¹⁰⁾, reduzindo o número total de exames a serem avaliados por humanos⁽¹¹⁾.

Diversos algoritmos de inteligência artificial (IA) vêm sendo utilizados para avaliação de imagens médicas. O aprendizado de máquina é uma subárea da IA que inclui modelos que conseguem aprender padrões e se aperfeiçoar a partir de banco de dados fornecido^(10,12).

Classicamente, o desenvolvimento de algoritmos de análise de imagens baseava-se em atributos relevantes previamente selecionados. Existe, todavia, abordagem mais recente de aprendizado de máquina, conhecida como aprendizado profundo, que utiliza algoritmos que identificam, por si só, as características que melhor classificariam os dados diretamente das imagens⁽¹¹⁾. Dentre as arquiteturas de aprendizado profundo utilizadas na análise de exames de imagem, destacam-se as redes neurais convolucionais (RNCs). As RNCs têm demonstrado melhores resultados

em relação a outras técnicas de IA para essa tarefa, especialmente a partir de 2012⁽¹¹⁾, permitindo maior rapidez e reprodutibilidade das leituras⁽¹¹⁾. A relação entre áreas da IA está ilustrada na Figura 1.

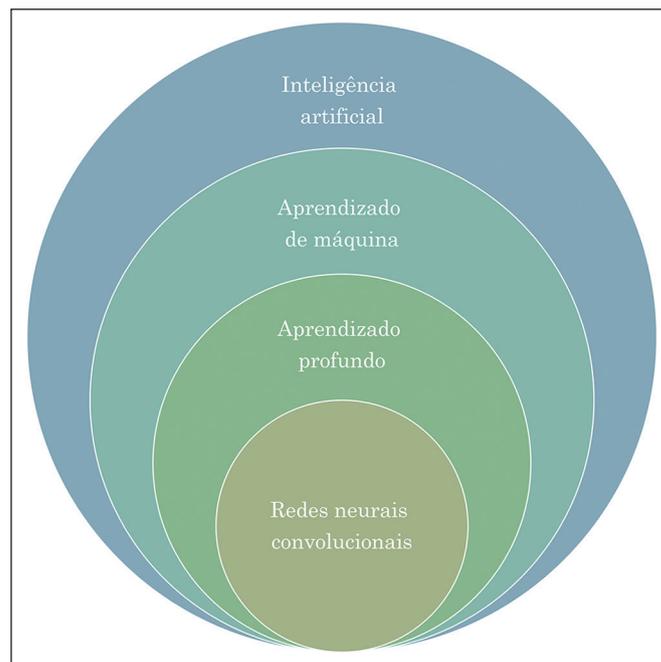


Figura 1. Diagrama de Venn empilhado demonstrando a relação entre as várias áreas da IA.

Na radiologia musculoesquelética, diversos estudos têm verificado a utilização da IA, incluindo o diagnóstico e classificação de fraturas, identificação de lesões ligamentares e meniscais, melhorias no fluxo de trabalho dos radiologistas, dentre outros benefícios⁽¹²⁾.

O treinamento e verificação da acurácia desses modelos computacionais, todavia, concentram-se em estudos clinicoepidemiológicos dos Estados Unidos^(10,13), sendo escassas as ferramentas validadas na população de outros países. Duas revisões recentes sobre o tema^(10,13) não identificaram estudos que abrangessem a população brasileira, ou mesmo latino-americana, utilizados no treinamento das RNCs atualmente existentes para o diagnóstico radiográfico de osteoartrite dos joelhos, tendo demonstrado a necessidade de maior validação externa.

O Estudo Longitudinal de Saúde do Adulto (ELSA-Brasil), maior estudo epidemiológico longitudinal da América Latina⁽¹⁴⁾, abrange, desde 2012, a avaliação de doenças musculoesqueléticas, por meio do estudo ancilar ELSA-Brasil Musculoesquelético (ELSA-Brasil MSK). Este incorpora, além das avaliações já realizadas pelo ELSA-Brasil, questionários de incapacidade e sintomas musculoesqueléticos, verificação de fatores de risco para doenças musculoesqueléticas, testes de *performance* física e radiografias das mãos e joelhos⁽³⁾.

O objetivo do presente trabalho é propor um modelo computacional para classificação de osteoartrite em radiografias de joelho, treinado com as radiografias do ELSA-

Brasil MSK. O *software* desenvolvido (código-fonte e modelo pré-treinado) está disponível no repositório GitHub (<https://github.com/jgdjuly/kneelsa>).

MATERIAIS E MÉTODOS

Amostra

O desenvolvimento do modelo computacional para análise automática de radiografias foi realizado a partir dos exames realizados na primeira visita da coorte ELSA-Brasil MSK. O ELSA-Brasil MSK incluiu, em sua linha de base, 2.901 servidores ativos ou aposentados de duas grandes instituições de ensino e pesquisa de Minas Gerais, com idades entre 38 e 79 anos (média de 56,0 anos), sendo 52,9% mulheres. Desses participantes, 2.830 realizaram radiografias de ambos os joelhos, resultando em imagens de 5.660 joelhos. Detalhes sobre o delineamento e perfil da coorte ELSA-Brasil MSK foram previamente publicados⁽³⁾. O estudo foi aprovado pelos comitês de ética e pesquisa das instituições envolvidas e os dados dos participantes são mantidos confidenciais no centro de dados do ELSA-Brasil.

Exame radiográfico

As radiografias dos joelhos com processamento digital foram realizadas em uma clínica de radiologia conveniada ao ELSA-Brasil, localizada nas adjacências do centro de investigação. A aquisição foi realizada em incidência posteroanterior bilateral em flexão fixa, com a utilização de posicionador (INPI BR102013033625-4) desenvolvido pela equipe de pesquisadores do ELSA-Brasil MSK⁽¹⁵⁾. Todos os exames foram realizados por técnico ou tecnólogo em radiologia devidamente treinado e certificado segundo o protocolo do estudo.

O protocolo de aquisição radiográfica foi avaliado em estudo anterior com delineamento teste-reteste, tendo sido demonstradas adequada qualidade da imagem e repetibilidade de parâmetros quantitativos⁽¹⁶⁾. Tal protocolo é compatível com os praticados em outros estudos longitudinais como Osteoarthritis Initiative (OAI) e Multicenter Osteoarthritis Study (MOST)^(17,18).

Interpretação humana

A interpretação das radiografias foi realizada de acordo com o seguinte protocolo, publicado e validado⁽³⁾: a) triagem de “osteoartrite possível” por dois tecnólogos de forma independente; b) todos os exames categorizados, por pelo menos um dos tecnólogos, como “osteoartrite possível”

foram revisados por um médico radiologista com treinamento específico. A concordância da leitura do médico radiologista do ELSA-Brasil MSK em relação a um leitor externo (radiologista musculoesquelético com *background* acadêmico, responsável pelas leituras do Framingham Osteoarthritis Study e MOST) e em relação a si próprio foi publicada previamente⁽³⁾, tendo sido estimado kappa inter-observador de 0,755 (IC 95%: 0,663–0,847) e intraobservador de 0,891 (IC 95%: 0,807–0,975) para diagnóstico radiográfico da osteoartrite dos joelhos.

Radiografias KL 0 e 1 foram consideradas negativas para osteoartrite e radiografias KL 2, 3 e 4 foram consideradas positivas para osteoartrite. A classificação binária (osteoartrite = 0; osteoartrite = 1) foi utilizada como valor de referência pela rede neural.

Modelo computacional

Avaliando-se as técnicas de IA mais utilizadas para a avaliação de imagens médicas atualmente^(11,19), destacam-se as RNCs, modelos de rede neural artificial compostos por camadas interconectadas (conceitualmente análogas aos neurônios biológicos), que implementam processo de classificação. As primeiras camadas detectam e extraem os atributos primitivos das imagens (tais como bordas e elementos de textura), que são, por sua vez, processados e selecionados nas camadas seguintes. Tais características são integradas, com diferentes pesos, na camada de saída, que fornece a predição de classe/desfecho de maior probabilidade⁽¹¹⁾.

O modelo de RNC aqui proposto utiliza arquitetura convolucional densamente conectada (DenseNet) pré-treinada de 161 camadas proposta por Huang et al.⁽²⁰⁾ (Figura 2). Nessa arquitetura, as camadas posteriores recebem também informações das camadas iniciais, o que evita a perda de informações importantes (detalhes das imagens) e permite modelos computacionalmente mais eficientes.

Foi realizado pré-processamento das imagens a partir dos dados brutos das radiografias em posteroanterior bilateral dos joelhos (arquivos DICOM). Inicialmente foram individualizados os joelhos direito e esquerdo, seguido de ampliação e redimensionamento das imagens, em matriz quadrada, com localização das regiões de interesse (compartimentos femorotibiais) (Figura 3).

Para aumento do número de imagens disponíveis para o treinamento da rede neural, foram realizados os seguintes mecanismos de *data augmentation* randômico do módulo *torchvision.transforms* da biblioteca PyTorch,

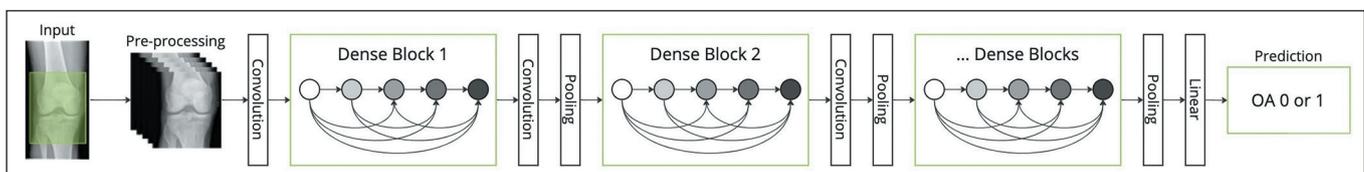


Figura 2. Ilustração esquemática da arquitetura DenseNet. Os pares de camadas são conectados, permitindo que elementos das primeiras camadas (como bordas) possam ser também utilizados nas últimas. Adaptado de Huang et al.⁽²⁰⁾.

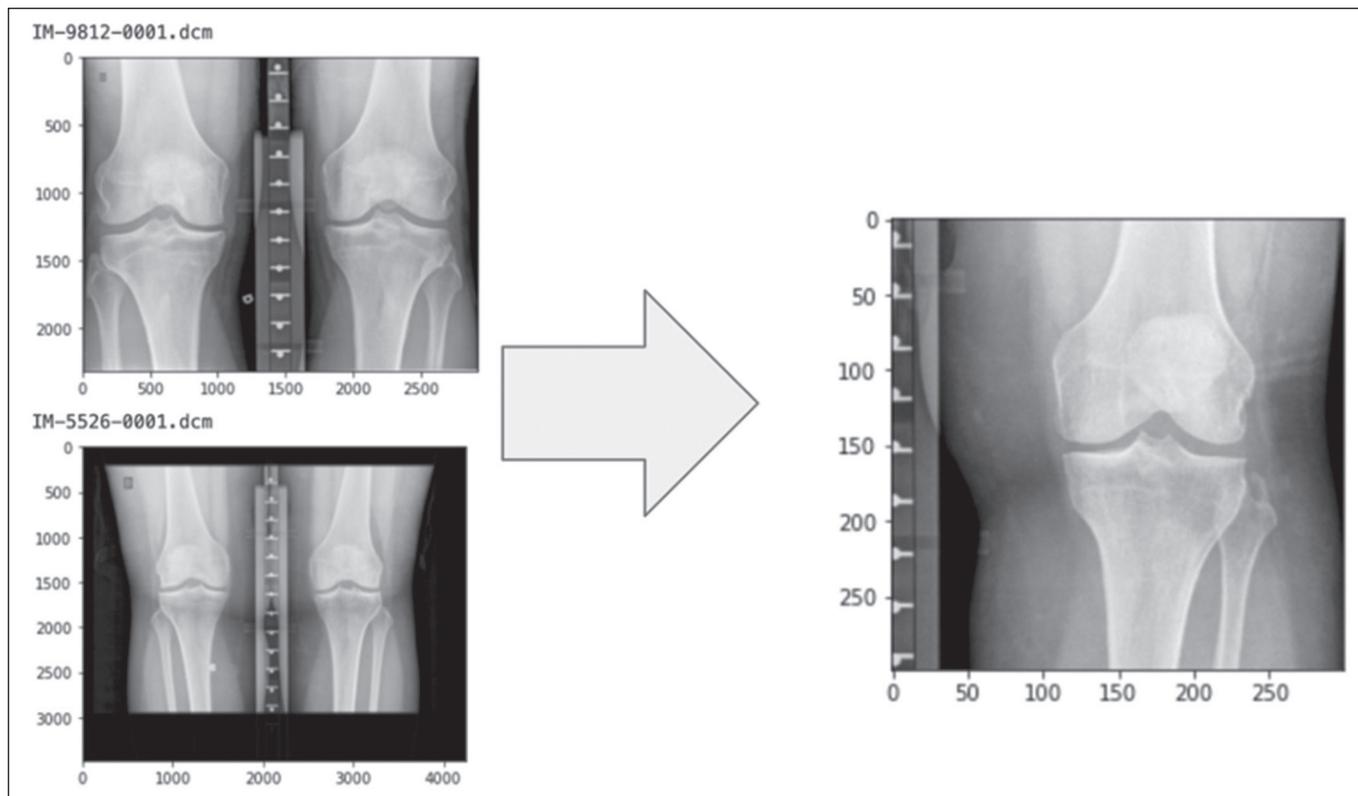


Figura 3. Demonstração do pré-processamento.

aplicados sobre a amostra de treinamento: rotação ($0,5^\circ$) e desfoque gaussiano, inversão horizontal, ajuste de nitidez (fator = 0,5) e desfoque gaussiano. Seguiram-se redimensionamento, corte central (*CenterCrop*), além de normalização. Os exames da amostra foram divididos em duas partições mutuamente exclusivas (*folds*) de treino e teste.

Como a saída do modelo é uma probabilidade para cada imagem, ele pode ser calibrado por meio da otimização de limiares (*threshold-moving*) que variam de 0 a 1. Em problemas de classificação binária, o limiar de decisão padrão é de 0,5; caso a probabilidade seja superior a esse valor, considera-se a classe 1, e caso contrário, a classe 0.

Análise dos dados

As classificações binárias (osteoartrite = 0; osteoartrite = 1) realizadas pela RNC foram comparadas com a análise do radiologista (valor de referência). Realizou-se a verificação do desempenho da RNC utilizando o módulo *metrics* da biblioteca *scikitlearn* (versão 1.0.2). Para cada limiar, as proporções de resultados verdadeiro-positivos, verdadeiro-negativos, falso-positivos e falso-negativos foram armazenadas em vetores, a partir dos quais foram calculados os valores de sensibilidade, especificidade, precisão, acurácia, acurácia balanceada, acurácia balanceada ponderada e escores F1 e F2, considerando-se a média das *folds*.

A acurácia é calculada pela razão entre o número de acertos (verdadeiro-positivos e verdadeiro-negativos) e o total de exames avaliados. Em amostras desbalanceadas, porém, como no caso do presente estudo, em que existem

muito mais exemplos de exames normais que alterados, essa métrica pode não demonstrar adequadamente o desempenho do modelo. Nesse contexto, a utilização da acurácia balanceada permite uma melhor estimativa do rendimento da RNC⁽²¹⁾, sendo calculada segundo a fórmula: $(\text{sensibilidade} + \text{especificidade}) / 2$. Alguns autores defendem, ainda, a utilização da acurácia balanceada ponderada^(21,22), que permite a atribuição de diferentes pesos para cada métrica, tendo sido calculada da seguinte forma: $(2 \times \text{sensibilidade} + \text{especificidade}) / 3$.

Para o cálculo da área sob a curva das características de operação do receptor (AUC) do modelo foram consideradas as probabilidades previstas para cada imagem, calculadas e armazenadas em lotes de 128 exames, que foram comparados com o valor-verdade por meio da função *roc_auc_score* do *scikit-learn*. Essa função traça a taxa de positivos classificados corretamente entre todas as previsões positivas (ou seja, a taxa de verdadeiro-positivos) em função de positivos incorretos entre todos os negativos (ou seja, a taxa de falso-positivos), em limites variados⁽²³⁾.

RESULTADOS

A RNC desenvolvida apresentou, considerando a média simples das duas *folds*, acurácia de 0,814 (no ponto de acurácia balanceada máxima), com sensibilidade de 0,755 e especificidade de 0,821. A AUC do modelo foi de 0,866 (IC 95%: 0,854–0,883) (Figura 4).

Seguindo a técnica acima exposta, o modelo pode ser calibrado para alcançar, não simultaneamente, valores

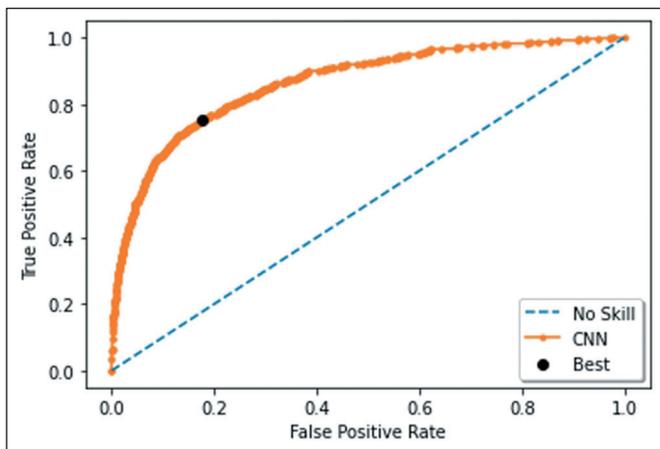


Figura 4. Curva característica de operação do receptor do modelo. O ponto preto demonstra o limiar de maior acurácia balanceada.

máximos de 0,907 para acurácia, 0,938 para sensibilidade e 0,994 para especificidade. Os valores máximos dos escores F1 e F2 alcançados foram de, respectivamente, 0,553 e 0,619. A Tabela 1 demonstra os valores máximos alcançados pela RNC, de acordo com a métrica otimizada.

DISCUSSÃO

O modelo desenvolvido apresentou bom desempenho⁽²⁴⁾ para o diagnóstico de osteoartrite dos joelhos radiográfica na incidência posteroanterior flexão-fixa. A comparação entre a eficiência de diferentes modelos de IA não está ainda padronizada na literatura⁽²⁵⁾. Apesar de haver uma predileção histórica pela acurácia, considera-se a AUC a métrica mais adequada atualmente para avaliação de desempenho⁽²⁵⁾.

Na principal revisão recente sobre a utilização de algoritmos de aprendizado de máquina para avaliação de osteoartrite, Binvignat et al.⁽¹⁰⁾ apontam apenas dois estudos que se propuseram diagnosticar osteoartrite dos joelhos somente em radiografias^(26,27). Brahim et al.⁽²⁶⁾ alcançaram 82,98% de acurácia (sensibilidade: 87,15%; especificidade: 80,65%) para a diferenciação de KL 0 e 2, tendo sido treinado em 1.024 imagens (514 KL = 0; 514 KL = 2) da OAI. Não foi calculada a AUC. O modelo conta, no processo de segmentação, com a delimitação manual de marcos anatômicos ósseos na tíbia, o que limita sua utilização em estudos de larga escala. A inclusão

de radiografias KL = 1 e a verificação da acurácia em uma amostra com maior número de pacientes sem osteoartrite (como ocorre na população em geral) seria interessante para verificar a acurácia do modelo em um contexto mais próximo do real. Tiulpin et al.⁽²⁷⁾ criaram uma rede neural siamesa para a classificação KL automática de radiografias dos joelhos. Foram utilizadas 18.376 radiografias do MOST para o treinamento da rede, com 2.957 e 5.960 imagens da OAI sendo empregadas para a validação e teste. Para estimar o desempenho do modelo para o diagnóstico de osteoartrite dos joelhos consideraram $KL \geq 2$, tendo alcançado AUC de 0,93. No treinamento, foram utilizados exames seriados dos participantes (de todas as visitas de controle) e todas as angulações de feixes de raios X disponíveis (5°, 10° e 15°), o que contribuiu para a robustez do modelo.

Algumas técnicas para tratar o desequilíbrio entre as classes foram tentadas, como a realização de *data augmentation* utilizada no presente trabalho apenas sobre os dados positivos e mudança na função de perda (penalizando 10 vezes mais os erros do tipo II), sem repercussões na AUC do modelo. De fato, estudos recentes em dados tabulares⁽²⁸⁾ demonstraram que esses e outros métodos de correção podem até reduzir a AUC, especialmente para modelos com bom desempenho⁽²²⁾.

No presente estudo, a calibração da rede neural por meio da definição de limiares foi o mecanismo que mais impactou nas métricas de desempenho. De fato, a definição de um menor limiar para a definição de osteoartrite dos joelhos permitiu incremento na sensibilidade do modelo, ao passo que o aumento do limiar ocasionou incremento na especificidade.

A calibração do modelo deve ser feita de acordo com a aplicação que se pretende dar a ela. Nesse sentido, modelo com maior acurácia balanceada seria mais apropriado caso sua aplicação seja como ferramenta de segunda leitura, enquanto modelo mais sensível seria preferível para sua utilização como método de triagem⁽¹¹⁾. Uma mesma rede neural com dois ou mais limiares ou, ainda, mais de uma rede neural podem ser também utilizadas em conjunto, especialmente quando se considera o baixo custo computacional e de tempo relacionado à utilização de modelos pré-treinados.

Tabela 1—Valores de acurácia, acurácia balanceada, acurácia balanceada ponderada, sensibilidade, especificidade, precisão e escores F1 e F2 para cada limiar definido.

Métrica maximizada	Limiar	Acurácia	Acurácia balanceada	Acurácia balanceada ponderada	Sensibilidade	Especificidade	Precisão	Escore F1	Escore F2
Sensibilidade	0,010	0,400	0,646	0,755	0,973	0,319	0,758	0,272	0,479
Acurácia balanceada ponderada	0,040	0,649	0,758	0,805	0,899	0,617	0,235	0,373	0,575
Acurácia balanceada	0,140	0,814	0,789	0,778	0,756	0,822	0,357	0,485	0,618
Escore F1	0,314	0,881	0,774	0,728	0,636	0,913	0,490	0,553	0,600
Acurácia	0,712	0,907	0,671	0,568	0,363	0,978	0,687	0,475	0,401
Especificidade	0,900	0,902	0,599	0,467	0,203	0,994	0,806	0,325	0,239

Considerando-se a especificidade alcançada pelo modelo, sua aplicação é viável em tarefas como a verificação de possíveis inconsistências (falso-negativos) no banco de dados ou definição de prioridade na fila de exames a serem analisados. A sensibilidade apresentada permite seu uso como possível ferramenta de triagem de exames normais, o que reduziria o volume de exames a serem avaliados pelos radiologistas. Ressalta-se que o diagnóstico de algumas doenças, como é o caso da osteoartrite dos joelhos, usualmente exige a associação dos achados radiográficos com dados clinicoepidemiológicos, laboratoriais e/ou de outros métodos de imagem, o que não foi testado neste estudo.

Por se tratar de modelo do tipo *blackbox*, é importante que as conclusões da RNC sejam pautadas em aspectos considerados relevantes para o diagnóstico, de forma compreensível pelo ser humano⁽²⁹⁾. Esse fator é chamado de explicabilidade ou interpretabilidade da rede, podendo ser expresso sob a forma de mapas de atenção, que destacam as regiões da imagem mais relacionadas à predição do modelo (como osteófitos, espaços articulares, esclerose). Ferramentas de explicabilidade da RNC criada ainda estão em desenvolvimento, sendo uma limitação atual do modelo.

O treinamento e a validação da RNC foram realizados com base na interpretação de dois tecnólogos e um radiologista, de acordo com fluxo de classificação de radiografias do ELSA-Brasil MSK, atendendo a rigoroso controle de qualidade⁽³⁾. A inclusão de exames provenientes de outros estudos longitudinais, com laudos realizados por um comitê de radiologistas, no entanto, pode contribuir para a robustez da rede, sendo um passo futuro no seu desenvolvimento. Ainda assim, considerando-se a acurácia alcançada, pensamos que o modelo conseguiu aprender a interpretar as radiografias dos joelhos.

As radiografias do ELSA-Brasil MSK foram realizadas por tecnólogos com treinamento, de forma padronizada, em incidência específica e com a utilização de posicionador próprio para a avaliação de osteoartrite dos joelhos. Na prática médica, entretanto, a incidência historicamente mais utilizada é a anteroposterior com extensão dos joelhos⁽⁸⁾, apesar de apresentar menor acurácia para a avaliação de osteoartrite dos joelhos, não sendo possível extrapolar os resultados para ambientes ambulatoriais ou hospitalares em geral. Ainda, foi utilizada apenas uma incidência (posteroanterior flexão-fixa) para desenvolvimento da RNC. No entanto, apenas 9,9% dos indivíduos com osteoartrite dos joelhos radiográfica no ELSA-Brasil MSK apresentavam osteoartrite isolada identificada na incidência lateral⁽³⁾. O desempenho dessa RNC ainda não foi testado em populações de outros estudos (como OAI ou MOST).

Outra limitação está na redução da resolução das imagens no pré-processamento, o que é usual no desenvolvimento de modelos de IA. Apesar de possibilitar maior tempo-efetividade, tal mecanismo pode limitar os resultados pela perda de informações sutis dos exames.

Melhorias no modelo apresentado, a partir da utilização de imagens das ondas subseqüentes do ELSA-Brasil MSK, bem como de bancos de imagens de outros estudos (como do OAI e MOST), devem contribuir para o aumento do desempenho e robustez da rede. Tais aprimoramentos estão sendo implementados e poderão ser aplicados em trabalhos futuros.

CONCLUSÕES

A RNC desenvolvida apresenta desempenho compatível com a das redes neurais treinadas com radiografias de estudos internacionais. A acurácia e AUC alcançadas permitem sua utilização como ferramenta de segunda leitura no ELSA-Brasil MSK, contribuindo para contornar a limitada disponibilidade de radiologistas treinados e os elevados custos e tempo gastos na interpretação de radiografias do joelho.

A validação do modelo em populações distintas daquela em que foi treinado, tanto de outros estudos longitudinais quanto da prática clínica, é importante para a sua adoção futura. Nesse sentido, reitera-se que o *software* desenvolvido está publicamente disponível no repositório GitHub (<https://github.com/jgdjulio/kneelsa>), possibilitando sua validação externa em estudos futuros.

Agradecimentos

O ELSA-Brasil é financiado pelo Ministério da Saúde (Departamento de Ciência e Tecnologia) e pelo Ministério da Ciência, Tecnologia e Inovação (Financiadora de Estudos e Projetos – Finep – e Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq) do Brasil [01 10 0742-00 BA, 01 12 0284-00 ES, 01 10 0746-00 MG, 01 11 0093-01 RJ, 01 10 0643-03 RS, 01 10 0773-00 SP]. O ELSA-Brasil MSK recebe apoio e agradece o recebimento de fomento dos seguintes órgãos/fundações: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) [post-doc/SUS 054/2010], da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig) [APQ-00921-16, APQ-00549-22] e do CNPq [423585/2016-9, 404728/2021-9].

REFERÊNCIAS

1. Santo L, Okeyode T. National Ambulatory Medical Care Survey: 2018 National Summary Tables. [cited 2022 Mar 3]. Available from: https://www.cdc.gov/nchs/data/ahcd/names_summary/2018-names-web-tables-508.pdf.
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;380:2163–96.
3. Telles RW, Machado LAC, Costa-Silva L, et al. Cohort profile update: the Brazilian Longitudinal Study of Adult Health Musculoskeletal (ELSA-Brasil MSK) cohort. *Int J Epidemiol*. 2022;51:e391–e400.
4. Sharma L. Osteoarthritis of the knee. *N Engl J Med*. 2021;384:51–9.
5. Hunter DJ, Schofield D, Callander E. The individual and socioeconomic impact of osteoarthritis. *Nat Rev Rheumatol*. 2014;10:437–41.
6. March LM, Bachmeier CJ. Economics of osteoarthritis: a global perspective. *Baillieres Clin Rheumatol*. 1997;11:817–34.

7. Miguel RCC, Machado LA, Costa-Silva L, et al. Performance of distinct knee osteoarthritis classification criteria in the ELSA-Brasil musculoskeletal study. *Clin Rheumatol*. 2019;38:793–802.
8. Buckland-Wright C. Which radiographic techniques should we use for research and clinical practice? *Best Pract Res Clin Rheumatol*. 2006;20:39–55.
9. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. *Ann Rheum Dis*. 1957;16:494–502.
10. Binvignat M, Padoia V, Butte AJ, et al. Use of machine learning in osteoarthritis research: a systematic literature review. *RMD Open*. 2022;8:e001998.
11. Chartrand G, Cheng PM, Vorontsov E, et al. Deep learning: a primer for radiologists. *Radiographics*. 2017;37:2113–31.
12. Román-Belmonte JM, Corte-Rodríguez H, Rodríguez-Merchán EC. Artificial intelligence in musculoskeletal conditions. *Front Biosci (Landmark Ed.)*. 2021;26:1340–8.
13. Yeoh PSQ, Lai KW, Goh SL, et al. Emergence of deep learning in knee osteoarthritis diagnosis. *Comput Intell Neurosci*. 2021;2021:4931437.
14. Schmidt MI, Duncan BB, Mill JG, et al. Cohort profile: longitudinal study of adult health (ELSA-Brasil). *Int J Epidemiol*. 2015;44:68–75.
15. Machado LAC, Barreto SM, Costa-Silva L, et al., inventores. Posicionador para aquisição e controle de qualidade de imagem radiográfica de joelhos em flexão fixa. Brasil. Instituto Nacional da Propriedade Industrial. Carta Patente Nº BR 102013033625-4; 2013.
16. Telles RW, Costa-Silva L, Machado LAC, et al. Fixed-flexion knee radiography using a new positioning device produced highly repeatable measurements of joint space width: ELSA-Brasil Musculoskeletal Study (ELSA-Brasil MSK). *Rev Bras Reumatol*. 2017;57:154–61.
17. Nevitt MC, Felson DT, Lester G. The osteoarthritis initiative. Protocol for the cohort study. [cited 2022 Mar 24]. Available from: <https://nda.nih.gov/static/docs/StudyDesignProtocolAndAppendices.pdf>.
18. Segal NA, Nevitt MC, Gross KD, et al. The Multicenter Osteoarthritis Study: opportunities for rehabilitation research. *PMR*. 2013; 5:647–54.
19. Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA; 2009.
20. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA; 2017.
21. Gupta A, Tatbul N, Marcus R, et al. Class-weighted evaluation metrics for imbalanced data classification. [cited 2022 Out 12]. Available from: <https://arxiv.org/abs/2010.05995v1>.
22. Araújo DC, Veloso AA, Borges KBG, et al. Prognosing the risk of COVID-19 death through a machine learning-based routine blood panel: a retrospective study in Brazil. *Int J Med Inform*. 2022;165:104835.
23. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;27:861–74.
24. Carter JV, Pan J, Rai SN, et al. ROC-ing along: evaluation and interpretation of receiver operating characteristic curves. *Surgery*. 2016; 159:1638–45.
25. Ling CX, Huang J, Zhang H. AUC: a better measure than accuracy in comparing learning algorithms. In: XiangY, Chaib-draa B, editors. *Advances in artificial intelligence. Canadian AI 2003. Lecture Notes in Computer Science*, vol 2671. Berlin, Heidelberg: Springer; 2003. p. 329–41.
26. Brahim A, Jennane R, Riad R, et al. A decision support tool for early detection of knee osteoarthritis using X-ray imaging and machine learning: data from the OsteoArthritis Initiative. *Comput Med Imaging Graph*. 2019;73:11–8.
27. Tiulpin A, Thevenot J, Rahtu E, et al. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. *Sci Rep*. 2018;8:1727.
28. van den Goorbergh R, van Smeden M, Timmerman D, et al. The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. *J Am Med Inform Assoc*. 2022;29:1525–34.
29. Fan FL, Xiong J, Li M, et al. On interpretability of artificial neural networks: a survey. *IEEE Trans Radiat Plasma Med Sci*. 2021;5: 741–60.

