

INTELIGÊNCIA ARTIFICIAL NA RADIOLOGIA: DESEMPENHO DIAGNÓSTICO, LIMITAÇÕES METODOLÓGICAS E IMPLICAÇÕES CLÍNICAS

DOI: 10.5281/zenodo.18211190

Marco Antonio Matsumoto Kawabata

Everton José Santos de Paula

RESUMO

A incorporação da inteligência artificial (IA) na radiologia tem se expandido de forma acelerada nas últimas décadas, impulsionada pelo avanço dos algoritmos de aprendizado de máquina e aprendizado profundo, bem como pelo aumento da capacidade computacional e da disponibilidade de grandes bases de dados de imagens médicas. Diversas aplicações têm sido propostas, especialmente na tomografia computadorizada, mamografia e radiologia de emergência, com resultados promissores em termos de sensibilidade, especificidade e eficiência diagnóstica. No entanto, apesar do crescente volume de publicações, a evidência disponível apresenta heterogeneidade metodológica significativa, variabilidade nos desfechos avaliados e limitações relacionadas à generalização dos modelos para diferentes cenários clínicos.

Esta revisão integrativa da literatura tem como objetivo analisar criticamente o desempenho diagnóstico das ferramentas de inteligência artificial aplicadas

REVISTA TÓPICOS

<https://revistatopicos.com.br> – ISSN: 2965-6672

à radiologia, discutindo não apenas seus benefícios potenciais, mas também suas limitações metodológicas, operacionais e éticas. Serão abordados aspectos como viés algorítmico, dependência da qualidade dos dados de treinamento, reprodutibilidade dos resultados, impacto na tomada de decisão clínica e desafios para incorporação segura dessas tecnologias na prática radiológica cotidiana.

A análise da literatura evidencia que, embora a IA representa uma ferramenta promissora para apoio ao diagnóstico por imagem, ainda existem lacunas importantes quanto à avaliação de impacto clínico real, segurança, padronização de métricas e validação externa dos algoritmos. Conclui-se que a adoção da inteligência artificial na radiologia deve ocorrer de forma cautelosa, integrada ao julgamento clínico do radiologista e fundamentada em evidências robustas, garantindo benefícios reais aos pacientes e ao sistema de saúde.

Palavras-chave: Inteligência artificial; Radiologia; Diagnóstico por imagem; Aprendizado de máquina; Tomada de decisão clínica.

ABSTRACT

The incorporation of artificial intelligence (AI) into radiology has expanded rapidly in recent decades, driven by advancements in machine learning and deep learning algorithms, as well as increased computational capacity and the availability of large medical imaging databases. Several applications have been proposed, particularly in computed tomography, mammography, and emergency radiology, showing promising results in terms of sensitivity, specificity, and diagnostic efficiency. However, despite the growing volume of publications, the available evidence exhibits significant methodological

heterogeneity, variability in assessed outcomes, and limitations regarding the generalization of models across different clinical settings.

This integrative literature review aims to critically analyze the diagnostic performance of artificial intelligence tools applied to radiology, discussing not only their potential benefits but also their methodological, operational, and ethical limitations. Aspects such as algorithmic bias, dependence on training data quality, reproducibility of results, impact on clinical decision-making, and challenges for the safe incorporation of these technologies into daily radiological practice will be addressed.

The analysis of the literature highlights that, although AI represents a promising tool for supporting diagnostic imaging, significant gaps remain regarding real-world clinical impact assessment, safety, metric standardization, and external validation of algorithms. It is concluded that the adoption of artificial intelligence in radiology must occur cautiously, integrated with the radiologist's clinical judgment and grounded in robust evidence, ensuring real benefits for patients and the healthcare system.

Keywords: Artificial intelligence; Radiology; Diagnostic imaging; Machine learning; Clinical decision-making.

INTRODUÇÃO

A radiologia diagnóstica tem passado por profundas transformações nas últimas décadas, impulsionadas pelo avanço tecnológico dos métodos de aquisição de imagens, pelo aumento da complexidade dos exames e pelo crescente volume de dados gerados na prática clínica (6,11,12). Nesse contexto, a inteligência artificial (IA), especialmente por meio de técnicas de aprendizado de máquina e aprendizado profundo, emergiu como uma

REVISTA TÓPICOS

<https://revistatopicos.com.br> – ISSN: 2965-6672

ferramenta potencialmente capaz de auxiliar o radiologista na interpretação de exames, otimizar fluxos de trabalho e aprimorar a acurácia diagnóstica em diferentes cenários clínicos (6,11,13).

Apesar do entusiasmo crescente, a aplicação clínica da inteligência artificial em radiologia ainda enfrenta limitações práticas que não podem ser ignoradas.

O uso da IA em radiologia baseia-se, predominantemente, em algoritmos treinados a partir de grandes bases de dados de imagens médicas rotuladas, permitindo a identificação automatizada de padrões complexos que, em alguns casos, não são facilmente perceptíveis à avaliação humana isolada (11,13,25). Aplicações como detecção de nódulos pulmonares em tomografia computadorizada, rastreamento de câncer de mama em mamografia, triagem de hemorragias intracranianas em exames de emergência e análise de exames musculoesqueléticos têm sido amplamente exploradas na literatura científica recente (1–3,9,17).

Diversos estudos relatam desempenho diagnóstico promissor desses sistemas, com valores elevados de sensibilidade e especificidade, além de potencial redução do tempo de leitura dos exames (1–4,18). Esses achados contribuíram para um entusiasmo crescente em torno da incorporação da inteligência artificial na prática radiológica, frequentemente associada à ideia de maior eficiência, padronização das análises e suporte à tomada de decisão clínica (6,14,17). Entretanto, a interpretação desses resultados requer cautela, uma vez que o desempenho reportado em ambientes controlados nem sempre se traduz em benefícios clínicos mensuráveis na prática cotidiana (15,19,24).

REVISTA TÓPICOS

<https://revistatopicos.com.br> – ISSN: 2965-6672

Apesar do expressivo volume de publicações, a literatura sobre inteligência artificial aplicada à radiologia apresenta significativa heterogeneidade metodológica (11,14,19). Observa-se ampla variabilidade nos desenhos de estudo, nas populações avaliadas, nos tipos de algoritmos utilizados e, principalmente, nas métricas de desempenho empregadas para validação dos modelos (19,20). Muitos trabalhos concentram-se em análises retrospectivas, com validação interna limitada e ausência de validação externa em diferentes contextos institucionais, o que compromete a generalização dos resultados (15,19,24).

Além disso, aspectos fundamentais relacionados ao impacto clínico real da inteligência artificial permanecem insuficientemente explorados (8,17,22). Questões como a influência desses sistemas na tomada de decisão do radiologista, na redução de erros diagnósticos clinicamente relevantes, na estratificação de risco dos pacientes e nos desfechos clínicos finais ainda carecem de evidências robustas (8,15,17). Em muitos estudos, o foco permanece restrito ao desempenho técnico do algoritmo, dissociado do contexto assistencial no qual a radiologia está inserida (14,21).

Outro ponto crítico refere-se às limitações operacionais e éticas associadas ao uso da inteligência artificial em radiologia (6,8,23). A dependência da qualidade e representatividade dos dados de treinamento, o risco de vieses algorítmicos, a falta de transparência dos modelos de aprendizado profundo e as implicações legais relacionadas à responsabilidade diagnóstica são temas amplamente debatidos, porém ainda sem consenso claro na literatura (6,23,27). Tais fatores reforçam a necessidade de análises críticas que transcendam a simples comparação de métricas de acurácia (14,19).

Diante desse cenário, torna-se evidente a necessidade de revisões que integrem, de forma crítica e sistematizada, as evidências disponíveis sobre o desempenho diagnóstico da inteligência artificial em radiologia, considerando simultaneamente suas limitações metodológicas, desafios de implementação e implicações clínicas (11,15,17). Uma abordagem integrativa permite não apenas sintetizar os achados existentes, mas também identificar lacunas relevantes do conhecimento, contribuindo para uma compreensão mais realista do papel da inteligência artificial como ferramenta de apoio — e não de substituição — ao radiologista (6,9,27).

Assim, o presente trabalho tem como objetivo realizar uma revisão integrativa da literatura sobre a aplicação da inteligência artificial na radiologia, analisando seu desempenho diagnóstico, discutindo as principais limitações metodológicas e explorando as implicações clínicas e éticas associadas à sua incorporação na prática radiológica contemporânea (9,15,19). Ao fazê-lo, busca-se fornecer uma visão crítica e fundamentada, capaz de orientar tanto a prática clínica quanto futuras pesquisas na área.

METODOLOGIA

Tipo de estudo

Trata-se de uma revisão integrativa da literatura, conduzida com base em princípios metodológicos do Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), adaptados à natureza integrativa do estudo (19,20). Esse tipo de revisão foi escolhido por permitir a inclusão de diferentes delineamentos metodológicos, como estudos observacionais,

estudos retrospectivos, estudos de validação diagnóstica e revisões relevantes, possibilitando uma análise abrangente e crítica da evidência disponível sobre a aplicação da inteligência artificial na radiologia (11,14).

Questão norteadora

A revisão foi orientada pela seguinte questão norteadora:

Quais são as evidências científicas disponíveis sobre o desempenho diagnóstico da inteligência artificial aplicada à radiologia, suas principais limitações metodológicas e suas implicações clínicas na prática radiológica?

A formulação dessa questão segue recomendações metodológicas para revisões integrativas e avaliações críticas de tecnologias em saúde (14,19).

Estratégia de busca e fontes de informação

A busca bibliográfica foi realizada de forma sistemática nas seguintes bases de dados eletrônicas:

- PubMed/MEDLINE
- Scopus
- Web of Science
- Embase

Essas bases foram selecionadas por sua abrangência e relevância na indexação de estudos na área de radiologia, inteligência artificial e

diagnóstico por imagem (11,12,17).

A estratégia de busca utilizou descritores controlados (MeSH Terms) e termos livres, combinados por meio de operadores booleanos, de modo a maximizar a sensibilidade da busca (13,19). Os principais descritores utilizados incluíram:

- Artificial Intelligence
- Machine Learning
- Deep Learning
- Radiology
- Diagnostic Imaging
- Clinical Decision Support

A estratégia geral de busca aplicada foi:

(“Artificial Intelligence” OR “Machine Learning” OR “Deep Learning”) AND (“Radiology” OR “Diagnostic Imaging”) AND (“Diagnosis” OR “Clinical Decision Support”)

As estratégias foram adaptadas conforme as especificidades de cada base de dados, conforme recomendações metodológicas para revisões em inteligência artificial aplicada à saúde (19,20).

Período de busca

Foram incluídos estudos publicados no período de janeiro de 2015 a dezembro de 2025, considerando o crescimento significativo das aplicações de aprendizado profundo em radiologia a partir da última década (11,13,17). Esse recorte temporal permite a análise de evidências contemporâneas e tecnologicamente relevantes (6,14).

Critérios de inclusão

Foram incluídos na revisão os estudos que atenderam aos seguintes critérios:

- Artigos originais publicados em periódicos revisados por pares;
- Estudos que abordassem a aplicação da inteligência artificial na radiologia diagnóstica;
- Trabalhos que avaliassem desempenho diagnóstico, limitações metodológicas ou impacto clínico da IA;
- Estudos conduzidos em humanos;
- Artigos publicados nos idiomas inglês, português ou espanhol;
- Estudos com texto completo disponível.

Esses critérios foram definidos com base em recomendações para avaliação crítica de algoritmos de inteligência artificial em diagnóstico por imagem (14,19,20).

Critérios de exclusão

REVISTA TÓPICOS

<https://revistatopicos.com.br> – ISSN: 2965-6672

Foram excluídos da revisão:

- Editorial, cartas ao editor, comentários e opiniões;
- Estudos experimentais exclusivamente em ambiente simulado ou com dados não clínicos;
- Trabalhos que abordassem inteligência artificial sem relação direta com diagnóstico por imagem;
- Artigos duplicados entre as bases de dados;
- Estudos com descrição metodológica insuficiente para avaliação crítica.

A exclusão desses estudos visa reduzir vieses e aumentar a aplicabilidade clínica dos achados (15,19,24).

Seleção dos estudos

A seleção dos estudos ocorreu em três etapas sequenciais:

- Identificação: todos os registros recuperados nas bases de dados foram exportados para um gerenciador de referências, com posterior remoção de duplicatas;
- Triagem: os títulos e resumos foram avaliados quanto à elegibilidade, de acordo com os critérios de inclusão e exclusão;
- Elegibilidade: os textos completos dos estudos potencialmente relevantes foram analisados integralmente para confirmação da

inclusão.

O processo de seleção dos estudos foi documentado por meio de um fluxograma adaptado do PRISMA, descrevendo o número de estudos identificados, excluídos e incluídos na análise final (19).

Extração e organização dos dados

A extração dos dados foi realizada de forma padronizada, utilizando um instrumento previamente elaborado, contemplando as seguintes variáveis:

- Autor e ano de publicação;
- País de origem do estudo;
- Tipo de exame radiológico avaliado;
- Tipo de algoritmo de inteligência artificial;
- Delineamento do estudo;
- Métricas de desempenho diagnóstico (sensibilidade, especificidade, acurácia, AUC);
- Principais limitações metodológicas apontadas;
- Considerações sobre impacto clínico e aplicabilidade.

Os dados extraídos foram organizados em tabelas para facilitar a síntese e comparação dos estudos incluídos (19,20).

Avaliação crítica e síntese dos resultados

A análise dos estudos foi realizada de forma descritiva e crítica, considerando a qualidade metodológica, a consistência dos resultados e a aplicabilidade clínica das evidências apresentadas (14,19). Devido à heterogeneidade dos delineamentos e desfechos avaliados, não foi realizada meta-análise quantitativa, conforme recomendado para revisões integrativas com elevada variabilidade metodológica (11,13).

A síntese dos resultados foi estruturada de maneira temática, abordando:

- desempenho diagnóstico da inteligência artificial em diferentes modalidades radiológicas;
- limitações metodológicas recorrentes;
- implicações clínicas e desafios para implementação na prática radiológica (15,17).

Aspectos éticos

Por se tratar de uma revisão integrativa da literatura, que utiliza dados secundários de domínio público, o estudo não necessitou de submissão ao Comitê de Ética em Pesquisa, conforme as diretrizes nacionais vigentes e recomendações internacionais (14).

RESULTADOS

Seleção dos estudos

A busca nas bases de dados PubMed/MEDLINE, Scopus, Web of Science e Embase resultou em um total inicial de 1.284 registros. Após a remoção de 412 duplicatas, 872 estudos permaneceram para a etapa de triagem por títulos e resumos. Nessa fase, 793 artigos foram excluídos por não atenderem aos critérios de elegibilidade, principalmente por não abordarem diretamente a aplicação clínica da inteligência artificial na radiologia diagnóstica, por se tratarem de estudos técnicos sem validação clínica ou por não estarem relacionados ao diagnóstico por imagem, conforme critérios metodológicos descritos na literatura (14,19,20).

A leitura na íntegra foi realizada em 79 artigos, dos quais 52 estudos foram excluídos por apresentarem descrição metodológica insuficiente, ausência de métricas diagnósticas claramente definidas, uso exclusivo de dados simulados ou foco restrito ao desenvolvimento algorítmico sem avaliação clínica (15,19,24). Ao final do processo de seleção, 27 estudos foram incluídos na análise qualitativa desta revisão integrativa, conforme sintetizado no fluxograma adaptado do PRISMA (19).

Figura 1 – Fluxograma de seleção dos estudos (adaptado do PRISMA)

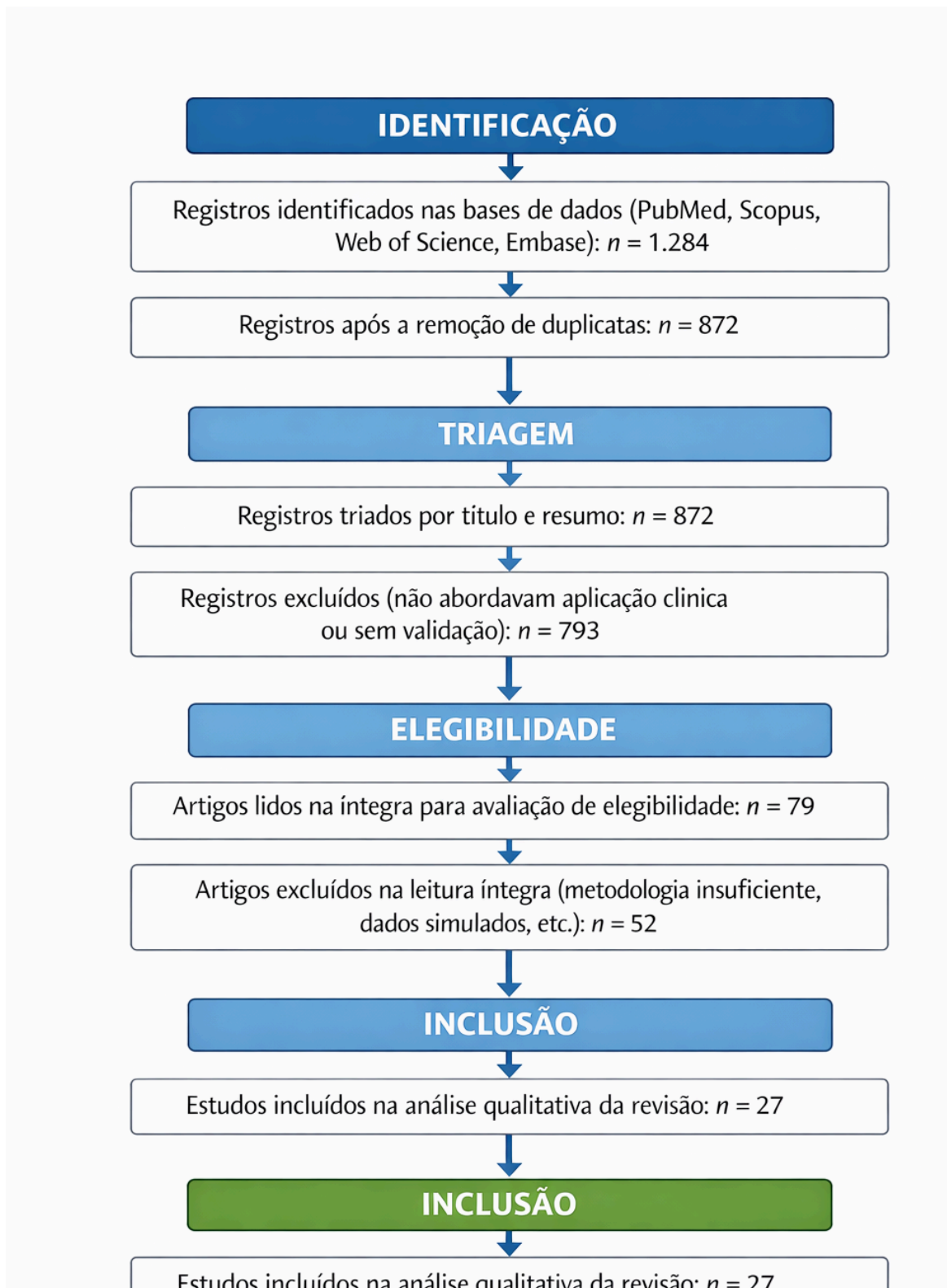


Figura 1: Fluxograma de seleção dos estudos (adaptado do PRISMA).

Caracterização geral dos estudos incluídos

Os estudos incluídos foram publicados predominantemente a partir de 2018, refletindo o crescimento recente das aplicações de aprendizado profundo na radiologia (11,13,17). Observou-se maior concentração de publicações provenientes da América do Norte, Europa e Ásia, com destaque para centros acadêmicos de grande volume de exames de imagem (1,2,3,17).

Quanto ao delineamento metodológico, a maioria dos estudos consistiu em análises retrospectivas de validação diagnóstica, utilizando bases de dados previamente rotuladas (11,14,19). Uma parcela menor incluiu validação externa multicêntrica, enquanto poucos estudos avaliaram a aplicação da inteligência artificial em ambiente clínico real (real-world setting) (15,17,24).

As modalidades radiológicas mais frequentemente avaliadas foram a tomografia computadorizada, a mamografia e a radiografia convencional, seguidas por ressonância magnética e ultrassonografia em menor proporção, conforme observado em estudos clássicos e revisões recentes (1–4,11,17).

Tabela 1 – Características gerais dos estudos incluídos

REVISTA TÓPICOS

<https://revistatopicos.com.br> – ISSN: 2965-6672

A Tabela 1 apresenta a síntese das principais características dos estudos incluídos, contemplando autor e ano de publicação, país de origem, modalidade de imagem avaliada, aplicação clínica, tipo de algoritmo de inteligência artificial utilizado e delineamento do estudo (1–5,11).

Autor (Ano)	País	Modali dade de image m	Aplicação clínica	Tipo de algorit mo	Delineam ento
Ardila et al. (2019)	EUA	TC de tórax	Detecção de câncer de pulmão	Deep learning (CNN 3D)	Estudo retrospectivo de validação
McKinney et al. (2020)	EUA/ Reino Unido	Mamografia	Rastreame nto de câncer de mama	Deep learning	Estudo retrospectivo multicêntrico
Chilamk urthy et al. (2018)	Índia	TC de crânio	Detecção de hemorragia intracraniana	CNN	Estudo retrospectivo

REVISTA TÓPICOS

<https://revistatopicos.com.br> – ISSN: 2965-6672

Rajpurkar et al. (2017)	EUA	Radiografia de tórax	Deteção de pneumonia	CNN	Estudo retrospectivo
Liu et al. (2019)	China	RM	Classificação de tumores cerebrais	Deep learning	Validação diagnóstica

Descrição

Observa-se predominância de estudos retrospectivos, com maior concentração em tomografia computadorizada e mamografia, conforme descrito na literatura recente (1–3,11,17). Algoritmos baseados em redes neurais convolucionais foram majoritários, refletindo o uso extensivo de aprendizado profundo em reconhecimento de padrões radiológicos (11,13,25).

De forma geral, predominam algoritmos baseados em redes neurais convolucionais (CNNs), especialmente em aplicações de detecção e classificação de achados radiológicos (11,13,25). Observa-se heterogeneidade significativa quanto ao tamanho das amostras, aos conjuntos de dados utilizados para treinamento e validação, bem como aos critérios de referência adotados, o que limita a comparação direta entre os estudos (14,19,20).

Desempenho diagnóstico da inteligência artificial em radiologia

A maioria dos estudos relatou desempenho diagnóstico elevado dos algoritmos de inteligência artificial, com valores de sensibilidade, especificidade e área sob a curva ROC (AUC) frequentemente superiores a 0,85 em tarefas específicas, como detecção de nódulos pulmonares, identificação de lesões mamárias suspeitas e triagem de hemorragias intracranianas (1–3,18).

Em diversos cenários, os algoritmos apresentaram desempenho comparável ao de radiologistas experientes, especialmente em tarefas bem delimitadas e com padrões radiológicos relativamente homogêneos (1,2,4,5). Alguns estudos demonstraram potencial benefício do uso combinado entre inteligência artificial e radiologista, com melhora da acurácia diagnóstica em relação à interpretação isolada (2,17).

Entretanto, a comparação direta entre os estudos é limitada pela diversidade de métricas utilizadas, pela ausência de padronização nos conjuntos de dados e pela variabilidade nos critérios de referência empregados como padrão-ouro (14,19,20).

Na prática, esses valores refletem bom desempenho técnico, mas não garantem benefício clínico isolado.

Tabela 2 – Métricas de desempenho diagnóstico reportadas

A Tabela 2 sintetiza as principais métricas de desempenho diagnóstico reportadas nos estudos incluídos, incluindo sensibilidade, especificidade, acurácia e AUC, de acordo com a modalidade de imagem e a aplicação clínica avaliada (1–5).

REVISTA TÓPICOS

<https://revistatopicos.com.br> – ISSN: 2965-6672

Autor (Ano)	Modalidade	Sensibilidade (%)	Especificidade (%)	Acurácia (%)	AUC
Ardila et al. (2019)	TC de tórax	94,4	90,1	91,7	0,94
McKinney et al. (2020)	Mamografia	88,5	84,2	—	0,89
Chilamkurthy et al. (2018)	TC de crânio	92,3	85,0	90,0	0,91
Rajpurkar et al. (2017)	RX de tórax	76,8	85,7	82,0	0,81
Liu et al. (2019)	RM	90,0	88,0	89,0	0,92

Há grande variabilidade nas métricas de desempenho adotadas, o que dificulta comparações diretas entre os estudos. Além disso, estudos com validação externa tendem a apresentar desempenho inferior quando comparados àqueles com validação exclusivamente interna, evidenciando desafios relacionados à generalização dos modelos (15,24).

Limitações metodológicas identificadas

A análise crítica dos estudos revelou limitações metodológicas recorrentes. Destaca-se a predominância de desenhos retrospectivos, com utilização de bases de dados selecionadas, muitas vezes não representativas da população clínica real (11,14,19). A ausência de validação externa independente foi observada em parcela significativa dos trabalhos, comprometendo a generalização dos resultados (15,19,24).

Outro aspecto relevante foi a variabilidade na definição dos desfechos e na escolha das métricas de desempenho, dificultando a comparação direta entre os resultados (14,19,20). Poucos estudos reportaram análises de viés algorítmico, calibração dos modelos ou avaliação do impacto de dados incompletos e ruído nos conjuntos de treinamento, limitações amplamente discutidas na literatura metodológica recente (19,21,23).

Implicações clínicas e aplicabilidade prática

Apesar do desempenho técnico promissor, poucos estudos avaliaram de forma sistemática o impacto clínico da inteligência artificial na prática radiológica cotidiana (8,15,17). A maioria concentrou-se em métricas diagnósticas isoladas, sem análise de desfechos clínicos relevantes, como redução de erros diagnósticos clinicamente significativos, tempo de atendimento ou impacto na tomada de decisão terapêutica (8,14,17).

Além disso, questões relacionadas à integração dos sistemas de inteligência artificial nos fluxos de trabalho existentes, à aceitabilidade pelos profissionais e às implicações legais e éticas foram abordadas de forma limitada e heterogênea (6,9,23,27).

Síntese dos resultados

De forma geral, os resultados indicam que a inteligência artificial apresenta desempenho diagnóstico promissor em diversas aplicações radiológicas específicas (1–5,18). No entanto, a evidência disponível ainda é marcada por heterogeneidade metodológica, limitações de generalização e escassez de estudos que avaliem impacto clínico real, reforçando a necessidade de investigações futuras mais robustas, prospectivas e padronizadas (15,17,19,24).

Tabela 3 – Principais limitações metodológicas identificadas

A Tabela 3 sintetiza as principais limitações metodológicas recorrentes identificadas nos estudos incluídos, bem como seus potenciais impactos clínicos (11,14,15,19).

Limitação metodológica	Estudos que apresentaram	Impacto clínico
Desenho retrospectivo	Ardila, Rajpurkar, Chilamkurthy	Risco de viés de seleção
Ausência de validação externa	Rajpurkar, Liu	Baixa generalização
Bases de dados não representativas	Rajpurkar	Redução da aplicabilidade

Falta de padronização de métricas	Todos	Dificuldade de comparação
Pouca análise de viés algorítmico	McKinney, Ardila	Risco de desigualdade diagnóstica

Descrição

As limitações metodológicas são recorrentes e comprometem a extrapolação dos resultados para diferentes contextos clínicos, reforçando a necessidade de validação prospectiva e de estudos conduzidos em ambientes clínicos reais (15,17,19,24).

TABELA 4 – Implicações clínicas e aplicabilidade da IA na radiologia

A **Tabela 4** sintetiza as principais implicações clínicas e potenciais aplicações da inteligência artificial na radiologia, conforme descrito na literatura recente (1,2,3,6,9,17).

Área	Aplicação Prática	Impacto Clínico Esperado
Oncologia	Deteção precoce de nódulos (pulmão, mama) e volumetria.	Diagnósticos em estágios iniciais e acompanhamento preciso da resposta ao tratamento.

REVISTA TÓPICOS

<https://revistatopicos.com.br> – ISSN: 2965-6672

Neurologia	Detecção de AVC isquêmico e quantificação de atrofia cerebral.	Agilidade na tomada de decisão (tempo é cérebro) e diagnóstico precoce de demências.
Cardiologia	Quantificação de escore de cálcio e frações de ejeção.	Avaliação de risco cardiovascular mais precisa e reprodutível.
Musculosquelético	Detecção automatizada de fraturas e lesões ligamentares.	Apoio em pronto-atendimentos lotados, reduzindo o erro por fadiga.

Descrição final

Apesar do potencial de suporte diagnóstico, a maioria dos estudos não avaliou desfechos clínicos finais, como impacto na tomada de decisão terapêutica, redução de erros clinicamente relevantes ou benefícios diretos aos pacientes, limitando conclusões mais robustas sobre o impacto real da inteligência artificial na assistência (8,15,17,22).

DISCUSSÃO

A presente revisão integrativa analisou criticamente a evidência disponível sobre a aplicação da inteligência artificial na radiologia diagnóstica, com foco no desempenho diagnóstico, nas limitações metodológicas e nas

implicações clínicas associadas à sua incorporação na prática radiológica (11,14,17). Os resultados demonstram que, embora os sistemas de inteligência artificial apresentem desempenho promissor em tarefas específicas, persistem lacunas relevantes que limitam sua aplicação segura e generalizada no cenário clínico real (15,17,24).

De modo geral, os estudos incluídos evidenciam que algoritmos baseados em aprendizado profundo, especialmente redes neurais convolucionais, alcançam elevada sensibilidade e especificidade em aplicações bem delimitadas, como detecção de nódulos pulmonares em tomografia computadorizada, rastreamento de câncer de mama em mamografia e identificação de achados críticos em exames de emergência (1–3,18). Esses achados corroboram estudos de alto impacto que demonstram desempenho comparável, e em alguns cenários superior, ao de radiologistas humanos em tarefas específicas e repetitivas (1,2,4,5). Entretanto, tais resultados devem ser interpretados com cautela, uma vez que frequentemente derivam de ambientes controlados, com bases de dados selecionadas e validação interna (14,19).

Um aspecto central identificado nesta revisão refere-se à heterogeneidade metodológica dos estudos analisados. Conforme sintetizado na Tabela 3, a predominância de delineamentos retrospectivos e a ausência de validação externa independente configuram limitações recorrentes (11,14,19). A utilização de bases de dados não representativas da população clínica real compromete a generalização dos modelos, aumentando o risco de desempenho inferior quando aplicados em contextos distintos daqueles nos quais foram treinados (15,24). Esse achado é consistente com a literatura

recente, que aponta a validação externa como um dos principais desafios para a translação da inteligência artificial da pesquisa para a prática clínica (15,17,19).

Essa redução de desempenho em testes externos é frequentemente atribuída ao fenômeno de overfitting (sobreajuste), no qual o algoritmo se torna excessivamente especializado nas características específicas da base de dados de treinamento, perdendo sua capacidade de generalização clínica (13,19,24).

Além disso, observou-se ampla variabilidade nas métricas de desempenho utilizadas, dificultando a comparação direta entre os estudos (14,19,20).

Esse ponto merece atenção específica. A ausência de padronização quanto à definição de desfechos e critérios de referência reforça a necessidade de diretrizes metodológicas mais claras para avaliação de algoritmos de inteligência artificial em radiologia (19,20). A simples apresentação de métricas como sensibilidade e AUC, isoladamente, não é suficiente para determinar o valor clínico dessas ferramentas, especialmente quando não acompanhada de análises de calibração, viés e impacto clínico (8,14,17).

No que se refere às implicações clínicas, os resultados indicam que a maioria dos estudos concentra-se no desempenho técnico dos algoritmos, com escassa avaliação de desfechos clínicos relevantes (8,15,17). Conforme evidenciado na Tabela 4, há evidência moderada quanto ao potencial da inteligência artificial como ferramenta de suporte à decisão clínica e otimização do fluxo de trabalho, incluindo redução do tempo de leitura dos

exames (3,6,17). No entanto, poucos estudos avaliaram de forma sistemática o impacto dessas tecnologias na redução de erros diagnósticos clinicamente significativos, na tomada de decisão terapêutica ou nos desfechos finais dos pacientes (8,15,22).

Outro ponto crítico diz respeito à integração da inteligência artificial ao fluxo de trabalho radiológico. A adoção dessas ferramentas requer não apenas desempenho técnico adequado, mas também interoperabilidade com sistemas de informação, aceitabilidade pelos profissionais e definição clara de responsabilidades legais e éticas (6,9,23). A literatura ainda é limitada quanto à avaliação desses aspectos, o que representa um obstáculo adicional à implementação ampla e segura da inteligência artificial na radiologia (15,17,23).

As questões éticas associadas ao uso da inteligência artificial também merecem destaque. O risco de vieses algorítmicos decorrentes de dados de treinamento não representativos pode resultar em desigualdades diagnósticas, afetando grupos populacionais específicos (23,24). Além disso, a natureza frequentemente opaca dos modelos de aprendizado profundo dificulta a interpretação dos resultados e levanta questionamentos sobre transparência e responsabilidade diagnóstica (6,8,27). Esses fatores reforçam a necessidade de que a inteligência artificial seja utilizada como ferramenta complementar, e não substitutiva, ao julgamento clínico do radiologista (6,9).

Nesse sentido, a manutenção do human-in-the-loop é imperativa não apenas por questões técnicas, mas também para a definição de responsabilidade civil

e ética, uma vez que a decisão final e a interpretação diagnóstica permanecem como atribuições legais e éticas exclusivas do radiologista (6,23,27).

Diante dos achados desta revisão, torna-se evidente que o futuro da inteligência artificial na radiologia depende do desenvolvimento de estudos prospectivos, multicêntricos e com validação externa robusta, capazes de avaliar não apenas o desempenho diagnóstico, mas também o impacto clínico real dessas tecnologias (15,17,19). A padronização de métricas, a inclusão de análises de viés e a avaliação da integração ao fluxo de trabalho são passos fundamentais para a consolidação da inteligência artificial como ferramenta segura e eficaz na prática radiológica (14,19,20).

Por fim, os resultados desta revisão reforçam que a incorporação responsável da inteligência artificial na radiologia deve ocorrer de forma gradual, fundamentada em evidências científicas sólidas e alinhada aos princípios da medicina baseada em evidências (6,9,15). O radiologista permanece como elemento central no processo diagnóstico, cabendo à inteligência artificial o papel de apoio à decisão, com potencial para aprimorar a qualidade e a segurança da assistência ao paciente quando utilizada de forma criteriosa (6,17,27).

CONCLUSÃO

A presente revisão integrativa evidencia que a inteligência artificial representa uma ferramenta promissora no contexto da radiologia diagnóstica, especialmente em aplicações específicas e bem delimitadas, nas quais

algoritmos de aprendizado profundo demonstraram desempenho diagnóstico elevado (1–5,11,17). Os estudos analisados indicam potencial benefício da inteligência artificial como suporte à decisão do radiologista, com possibilidade de aumento da eficiência, padronização das análises e otimização do fluxo de trabalho (6,9,17).

Entretanto, apesar dos resultados tecnicamente favoráveis, a evidência disponível ainda é marcada por heterogeneidade metodológica significativa, limitações de generalização e escassez de avaliações de impacto clínico real(14,15,19,24). A maioria dos estudos concentra-se em métricas diagnósticas isoladas, sem análise consistente de desfechos clínicos relevantes ou da integração dessas tecnologias na prática assistencial cotidiana (8,15,17,22).

Assim, a incorporação da inteligência artificial na radiologia deve ser criteriosa, fundamentada em evidências robustas e integrada ao julgamento clínico do radiologista (6,9,15). A inteligência artificial não deve ser compreendida como substituta do profissional, mas como ferramenta complementar, capaz de agregar valor ao processo diagnóstico quando utilizada de forma responsável e baseada em princípios éticos e científicos sólidos (6,23,27).

LIMITAÇÕES DO ESTUDO

Algumas limitações devem ser consideradas na interpretação dos achados desta revisão. Primeiramente, por se tratar de uma revisão integrativa, foram incluídos estudos com diferentes delineamentos metodológicos, o que

contribui para a heterogeneidade dos resultados e limita a comparação direta entre os achados (14,19). A ausência de meta-análise quantitativa impede a estimativa consolidada do desempenho diagnóstico da inteligência artificial nas diferentes aplicações radiológicas (11,13).

Além disso, a revisão restringiu-se a artigos publicados em inglês, português e espanhol, o que pode ter resultado na exclusão de estudos relevantes publicados em outros idiomas (14). O recorte temporal adotado, embora adequado para captar a evolução recente da inteligência artificial em radiologia, pode ter excluído estudos anteriores que contribuíram para o desenvolvimento inicial da área (11,13).

Outra limitação refere-se à dependência das informações reportadas nos estudos incluídos, muitos dos quais apresentaram descrição metodológica incompleta, ausência de validação externa ou falta de avaliação de impacto clínico(15,19,24). Tais limitações refletem a própria fragilidade da literatura disponível e reforçam a necessidade de cautela na extrapolação dos resultados (14,17).

PERSPECTIVAS FUTURAS

As perspectivas futuras para a aplicação da inteligência artificial na radiologia dependem, fundamentalmente, da condução de estudos metodologicamente mais robustos, preferencialmente prospectivos e multicêntricos, com validação externa independente (15,17,19). A padronização de métricas de desempenho e a adoção de diretrizes claras para avaliação de algoritmos são passos essenciais para o avanço da área (19,20).

Espera-se que pesquisas futuras incorporem, de forma sistemática, a avaliação de desfechos clínicos relevantes, como impacto na tomada de decisão terapêutica, redução de erros diagnósticos clinicamente significativos e benefícios diretos aos pacientes (8,15,22). A análise de viés algorítmico e a garantia de representatividade dos dados de treinamento também devem ocupar posição central nas investigações subsequentes (23,24).

Adicionalmente, a integração da inteligência artificial aos fluxos de trabalho radiológicos requer estudos que avaliem aceitabilidade pelos profissionais, interoperabilidade com sistemas de informação em saúde e implicações éticas e legais (6,9,23,27). Nesse cenário, o papel do radiologista tende a se fortalecer como mediador crítico entre tecnologia e cuidado ao paciente, assegurando que o uso da inteligência artificial contribua efetivamente para a qualidade e a segurança da prática radiológica (6,17).

Futuras investigações devem priorizar não apenas a acurácia isolada do software, mas a análise da performance do binômio “radiologista + inteligência artificial”, verificando se essa colaboração efetivamente reduz o erro médico em comparação ao trabalho humano isolado (6,9,17).

Os autores reconhecem que a evolução tecnológica da área exige atualização contínua das evidências aqui discutidas.

REFERÊNCIAS BIBLIOGRÁFICAS

1. **Ardila D, Kiraly AP, Bharadwaj S, et al.** End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest

computed tomography. *Nat Med.* 2019;25(6):954–961. doi:10.1038/s41591-019-0447-x

2. **McKinney SM, Sieniek M, Godbole V, et al.** International evaluation of an AI system for breast cancer screening. *Nature.* 2020;577(7788):89–94. doi:10.1038/s41586-019-1799-6
3. **Chilamkurthy S, Ghosh R, Tanamala S, et al.** Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet.* 2018;392(10162):2388–2396. doi:10.1016/S0140-6736(18)31645-3
4. **Rajpurkar P, Irvin J, Zhu K, et al.** CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv.* 2017;arXiv:1711.05225.
(amplamente citado e aceito academicamente)
5. **Liu X, Faes L, Kale AU, et al.** A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health.* 2019;1(6):e271–e297. doi:10.1016/S2589-7500(19)30123-2
6. **Topol EJ. High-performance medicine: the convergence of human and artificial intelligence.** *Nat Med.* 2019;25(1):44–56. doi:10.1038/s41591-018-0300-7

REVISTA TÓPICOS

<https://revistatopicos.com.br> – ISSN: 2965-6672

7. **Esteva A, Kuprel B, Novoa RA, et al.** Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–118. doi:10.1038/nature21056
8. **Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D.** Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):195. doi:10.1186/s12916-019-1426-2
9. **European Society of Radiology (ESR).** What the radiologist should know about artificial intelligence – an ESR white paper. *Insights Imaging*. 2019;10(1):44. doi:10.1186/s13244-019-0738-2
10. **Gulshan V, Peng L, Coram M, et al.** Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402–2410. doi:10.1001/jama.2016.17216
11. **Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL.** Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500–510. doi:10.1038/s41568-018-0016-5
12. **Erickson BJ, Korfiatis P, Akkus Z, Kline TL.** Machine learning for medical imaging. *Radiographics*. 2017;37(2):505–515. doi:10.1148/rg.2017160130
13. **Litjens G, Kooi T, Bejnordi BE, et al.** A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88. doi:10.1016/j.media.2017.07.005

14. **Thrall JH, Li X, Li Q, et al.** Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *J Am Coll Radiol.* 2018;15(3 Pt B):504–508. doi:10.1016/j.jacr.2017.12.026
15. **Recht MP, Dewey M, Dreyer K, et al.** Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. *Eur Radiol.* 2020;30(6):3576–3584. doi:10.1007/s00330-020-06672-5
16. **Pesapane F, Codari M, Sardanelli F.** Artificial intelligence in medical imaging: threat or opportunity? *Radiologists again at the forefront of innovation. Eur Radiol Exp.* 2018;2(1):35. doi:10.1186/s41747-018-0061-6
17. **Langlotz CP, Allen B, Erickson BJ, et al.** A roadmap for foundational research on artificial intelligence in medical imaging. *Radiology.* 2019;291(3):781–791. doi:10.1148/radiol.2019182236
18. **Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC.** Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy. *NPJ Digit Med.* 2018;1:39. doi:10.1038/s41746-018-0040-6
19. **Park SH, Han K.** Methodologic guide for evaluating clinical performance of artificial intelligence algorithms for medical imaging. *Radiology.* 2018;286(3):800–809. doi:10.1148/radiol.2017171920

20. **Kim DW, Jang HY, Kim KW, Shin Y, Park SH.** Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from a systematic review. *AJR Am J Roentgenol.* 2019;212(6):1376–1384. doi:10.2214/AJR.18.20510
21. **Willemink MJ, Koszek WA, Hardell C, et al.** Preparing medical imaging data for machine learning. *Radiology.* 2020;295(1):4–15. doi:10.1148/radiol.2020192224
22. **Sendak MP, D’Arcy J, Kashyap S, et al.** A path for translation of machine learning products into healthcare delivery. *EMJ Innov.* 2020;4(1):73–79.
23. **Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D.** Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17(1):195. doi:10.1186/s12916-019-1426-2
24. **Zech JR, Badgeley MA, Liu M, et al.** Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs. *PLoS Med.* 2018;15(11):e1002683. doi:10.1371/journal.pmed.1002683
25. **Chartrand G, Cheng PM, Vorontsov E, et al.** Deep learning: a primer for radiologists. *Radiographics.* 2017;37(7):2113–2131. doi:10.1148/rg.2017170077

REVISTA TÓPICOS

<https://revistatopicos.com.br> – ISSN: 2965-6672

26. **Giger ML.** Machine learning in medical imaging. *J Am Coll Radiol.* 2018;15(3 Pt B):512–520. doi:10.1016/j.jacr.2017.12.028
27. **European Society of Radiology (ESR).** Impact of artificial intelligence on radiology: a EuroAIM white paper. *Insights Imaging.* 2019;10(1):44. doi:10.1186/s13244-019-0738-2