REVISTA TÓPICOS
ISSN 2965-6672 | Qualis A2

O AUXÍLIO DA INTELIGÊNCIA ARTIFICIAL NO DIAGNÓSTICO MÉDICO

THE ROLE OF ARTIFICIAL INTELLIGENCE IN MEDICAL DIAGNOSIS

Ciências da Saúde • 06/11/2026

REGISTRO DOI: 10.70773/revistatopicos/793751152

Guilherme Andrade Soares Moreno[1]

Gustavo Oliveira Andrade[2]

Pedro Henrique Nascimento de Almeida[3]

Raoni Fraga Abreu[4]

Clarissa Leal Silva e Souza[5]

RESUMO

Este estudo analisa as aplicações da inteligência artificial (IA) no diagnóstico médico, com ênfase em seus benefícios e limitações. Trata-se de uma revisão integrativa da literatura, conduzida a partir da pergunta norteadora sobre quais são as aplicações da IA no diagnóstico médico. A busca ocorreu entre agosto e setembro de 2026 nas bases Cochrane, SciELO e BVS, com descritores combinados por operadores booleanos, e considerou publicações de 2021 a 2025 disponíveis na íntegra e gratuitamente. Após as etapas de filtragem, leitura de títulos e resumos e leitura integral, a amostra final reuniu 13 artigos, cujos dados foram organizados em dois quadros. Os resultados apontam desempenho técnico elevado em tarefas de análise de imagens e de predição de risco, além de viabilidade de implantação em diferentes contextos de saúde. Em contrapartida, os estudos relatam falsos negativos, baixa sensibilidade de modelos de linguagem em casos reais, variação de desempenho entre coortes, amostras pequenas, delineamentos retrospectivos e escassa validação externa. O único ensaio clínico randomizado registrou diferenças estatisticamente não significativas nos desfechos globais. Conclui-se que a IA atua como ferramenta complementar ao raciocínio clínico, exige supervisão profissional e validação rigorosa, e demanda estudos prospectivos com pacientes reais para a comprovação de benefício clínico.

Palavras-chave: Inteligência artificial; Diagnóstico médico; Diagnóstico por imagem; Aprendizado de máquina; Revisão integrativa.

ABSTRACT

This study analyzes the applications of artificial intelligence (AI) in medical diagnosis, emphasizing its benefits and limitations. It is an integrative literature review guided by the question of which

applications of AI exist in medical diagnosis. The search took place between August and September 2026 in the Cochrane, SciELO and BVS databases, using descriptors combined with Boolean operators, and included publications from 2021 to 2025 available in full and free of charge. After filtering, title and abstract screening, and full-text reading, the final sample comprised 13 articles, whose data were organized into two tables. The results show high technical performance in image analysis and risk prediction tasks, as well as feasibility of implementation in different healthcare settings. Conversely, the studies report false negatives, low sensitivity of language models in real cases, performance variation between cohorts, small samples, retrospective designs and scarce external validation. The only randomized clinical trial recorded statistically nonsignificant differences in overall outcomes. It is concluded that AI acts as a complementary tool to clinical reasoning, requires professional supervision and rigorous validation, and calls for prospective studies with real patients to confirm clinical benefit.
Keywords: Artificial intelligence; Medical diagnosis; Diagnostic imaging; Machine learning; Integrative review.

1. INTRODUÇÃO

O diagnóstico médico é uma das etapas mais desafiadoras da prática clínica, pois dele dependem a definição do quadro clínico, a escolha terapêutica e o prognóstico do paciente. Precisão e rapidez são fundamentais para reduzir complicações, evitar tratamentos inadequados e otimizar os recursos em saúde (Singh *et al.*, 2016). Contudo, o processo enfrenta desafios como a sobrecarga de informações, a variabilidade entre profissionais, a limitação de tempo nos atendimentos e a necessidade de interpretar dados clínicos e laboratoriais cada vez mais complexos (Singh *et al.*, 2016).

Nas últimas décadas, a tecnologia tem transformado a prática médica e ampliado a capacidade do profissional, dos exames de imagem aos prontuários eletrônicos (Topol, 2023). Nesse contexto, a inteligência artificial (IA) surge como um dos principais recursos contemporâneos para auxiliar o diagnóstico, trazendo inovação à tomada de decisão clínica.

A IA é um conjunto de tecnologias capazes de aprender, interpretar dados e realizar tarefas quase instantaneamente. Na saúde, sua aplicação é ampla: análise de grandes bases de dados, apoio a pesquisas, triagem automatizada e, especialmente, suporte ao diagnóstico médico (Simonassi *et al.*, 2024). Na radiologia, já é usada para detectar padrões em imagens; na clínica médica, algoritmos que integram sinais e sintomas do paciente podem otimizar o fechamento do diagnóstico (Topol, 2023).

Entre os benefícios da IA na saúde, destacam-se a maior precisão diagnóstica, a agilidade na interpretação de resultados e a redução de falhas ligadas à subjetividade humana (Soares *et al.*, 2023). Além disso, ela pode democratizar o acesso à saúde, oferecendo suporte em tempo hábil a regiões isoladas. No entanto, persistem desafios, como questões éticas, necessidade de validação científica robusta, limitações técnicas e capacitação dos profissionais para o uso adequado dessas ferramentas (Iranilda *et al.*, 2024).

Diante desse cenário, ainda existem lacunas na literatura quanto à efetividade e aos limites do uso da IA na prática cotidiana, além de importantes questões éticas e legais, como a responsabilidade profissional, a transparência dos algoritmos e a proteção de dados sensíveis (Iranilda *et al.*, 2024). Torna-se, portanto, relevante investigar de que forma essa tecnologia pode auxiliar o diagnóstico

médico, analisando suas contribuições, limites e perspectivas futuras, uma vez que compreender como ela pode complementar o trabalho do profissional de saúde é fundamental para aumentar a segurança tanto do médico quanto do paciente (Beam *et al.*, 2023; Li *et al.*, 2024). Assim, o presente estudo tem como objetivo analisar as aplicações da Inteligência Artificial (IA) no diagnóstico médico.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. História da Inteligência Artificial

A IA consolidou-se como um dos principais vetores de transformação tecnológica e social da contemporaneidade. Seu conceito remonta à década de 1950, quando pesquisadores como Alan Turing, John McCarthy e Marvin Minsky iniciaram os estudos sobre máquinas capazes de pensar e aprender. A conferência de Dartmouth, em 1956, é tida como o marco inicial da IA como campo científico (Teixeira, 2025). Com o avanço da capacidade computacional e o acúmulo de grandes volumes de dados, surgiram abordagens como o aprendizado de máquina (*machine learning*) e o aprendizado profundo (*deep learning*), que permitiram desempenhos antes restritos à cognição humana. Isso viabilizou aplicações em setores como medicina, educação, indústria e segurança pública, com ganhos de eficiência, produtividade e inovação (Barbosa; Portes, 2023).

Seu uso envolve sistemas capazes de reconhecer padrões, realizar previsões e automatizar processos complexos, contribuindo, quando bem aplicada, para a reestruturação de processos, a criação de novos modelos de negócio e a reorganização social baseada na inovação (Barbosa; Portes, 2023; Teixeira, 2025). Contudo, sua

expansão traz desafios éticos e sociais, como a concentração de poder tecnológico, o risco de reforço das desigualdades e os impactos sobre a autonomia e a privacidade dos indivíduos, o que reforça a necessidade de regulamentação, transparência e monitoramento contínuo (Teixeira, 2025). Assim, a IA deve ser compreendida de forma multidimensional, considerando seus aspectos técnicos, sociais, econômicos e éticos (Barbosa; Portes, 2023; Teixeira, 2025).

2.2. Diagnóstico Médico

O diagnóstico médico é um dos pilares da prática clínica, pois consiste em identificar a condição ou doença do paciente a partir da análise de sinais, sintomas e resultados de exames. Por ser a etapa central da atuação médica, dele derivam as decisões terapêuticas que influenciam diretamente a evolução clínica e o prognóstico, exigindo do profissional a combinação de conhecimento científico, experiência clínica e raciocínio lógico (Braun et al., 2017).

Tradicionalmente, o processo diagnóstico se estrutura em três etapas: a anamnese, que consiste na entrevista clínica para coleta do histórico de saúde, dos hábitos de vida e dos sintomas; o exame físico, que busca alterações corporais por meio de métodos semiológicos padronizados; e os exames complementares, que incluem análises laboratoriais (bioquímicas, microbiológicas e imunológicas) e exames de imagem, como tomografia computadorizada (TC), ultrassonografia (USG) e ressonância magnética (RM), que ampliam a acurácia da investigação ao permitir a visualização de estruturas internas.

Apesar dos avanços científicos e tecnológicos, essa etapa ainda enfrenta desafios relevantes, como a subjetividade da interpretação clínica, a sobrecarga de informações, a variabilidade entre profissionais e a possibilidade de erros diagnósticos, que podem causar danos irreversíveis ao paciente. Como o diagnóstico fundamenta todo o processo terapêutico, a maior parte dos erros médicos tem origem nele, seja por falhas de comunicação, insuficiência de dados ou limitações humanas (Zhang *et al.*, 2023). Esses desafios reforçam a necessidade de ferramentas que auxiliem o profissional na tomada de decisões mais assertivas e baseadas em evidências.

2.3. Uso da Inteligência Artificial no Diagnóstico Médico

O diagnóstico médico adequado é a base da prática clínica, pois orienta condutas terapêuticas, reduz riscos de iatrogenia e otimiza recursos em saúde. Diagnósticos incorretos ou tardios estão presentes em parcela significativa dos atendimentos, com impacto direto na morbidade, na mortalidade e nos custos hospitalares (Graber; Russell; Graber, 2018). Um bom diagnóstico exige, portanto, a integração de dados clínicos, laboratoriais e de imagem, aliada ao raciocínio clínico criterioso e ao uso de ferramentas de apoio.

Watari e colaboradores (2022) apontam os vieses cognitivos entre as principais causas de erro de diagnóstico. No estudo, os autores identificaram em média três vieses por caso de erro lembrado pelos médicos, com risco ainda maior em ambientes como pronto-atendimentos. Em radiologia, a análise do espectro de vieses cognitivos mostrou sua sobreposição em contextos de alta demanda de imagens e sugeriu estratégias de mitigação, como

conscientização, revisão sistemática da interpretação e uso de ferramentas auxiliares (Lee; Bezuidenhout; Kruskal, 2024).

Nesse cenário, a inteligência artificial (IA) surge como aliada, ao processar grandes volumes de dados, identificar padrões que podem passar despercebidos ao médico e contrabalançar vieses cognitivas. Sua incorporação, contudo, deve respeitar princípios éticos, a responsabilidade profissional e a validação rigorosa, mantendo o médico como decisor final, com o suporte da tecnologia para um diagnóstico mais preciso, rápido e seguro (Rajkomar *et al.*, 2019; Topol, 2019).

2.4. Benefícios e Riscos

A inteligência artificial auxilia na identificação de padrões e de pequenas alterações em exames e informações clínicas, favorecendo o reconhecimento mais precoce e preciso de doenças, como o câncer, cujo diagnóstico precoce é fundamental para aumentar as chances de cura (Silva et al., 2026). Entre seus principais benefícios estão o aumento da precisão diagnóstica, com redução de erros humanos e maior segurança ao paciente, e a agilidade na obtenção de resultados, essencial em situações de emergência (Ferreira; Lima; Sousa, 2022). A IA também apoia a tomada de decisão médica, sobretudo em locais com poucos especialistas (Carvalho; Araújo, 2022), permite personalizar o atendimento ao prever riscos e sugerir condutas mais adequadas ao perfil de cada paciente (Moura; Costa, 2023) e, associada à telemedicina, amplia o acesso ao diagnóstico em regiões remotas e com recursos limitados (Marques *et al.*, 2025).

Por outro lado, sua adoção exige análise crítica dos desafios técnicos, éticos e legais (Topol, 2019). Algoritmos treinados com bases de

dados não representativas podem reproduzir vieses e comprometer a equidade diagnóstica entre grupos populacionais (Jiang *et al.*, 2017; WHO, 2021). Muitos sistemas funcionam como "caixas-pretas", o que dificulta compreender os critérios das recomendações e limita a transparência e a validação científica (WHO, 2021). Há ainda o risco de dependência excessiva da tecnologia, que pode reduzir o exercício do raciocínio clínico e aumentar a chance de erros em casos atípicos ou complexos (Esr, 2019), além de questões sobre a responsabilidade por falhas diagnósticas e a proteção de dados sensíveis (WHO, 2021).

Assim, a IA deve ser entendida como apoio, e não substituta do médico, sendo fundamental ponderar seus benefícios e riscos para embasar protocolos que integrem a tecnologia de forma segura, ética e eficaz, preservando o papel central do profissional e a segurança do paciente (WHO, 2021).

3. METODOLOGIA

3.1. Tipo de Estudo

Trata-se de uma revisão integrativa da literatura, método que reúne, avalia e sintetiza de forma sistemática os resultados de pesquisas sobre um tema. O estudo foi conduzido a partir da seguinte pergunta norteadora: "Quais são as aplicações da Inteligência Artificial no diagnóstico médico?".

3.2. Bases de Dados e Estratégia de Busca

A busca dos artigos foi realizada entre agosto e setembro de 2026, nas bases de dados Cochrane library, Scientific Electronic Library Online (SciELO) e Biblioteca Virtual em Saúde (BVS). Foram

utilizados os Descritores em Ciências da Saúde (DeCS) e seus correspondentes no *Medical Subject Headings* (MeSH). Os cruzamentos foram estruturados em português e inglês por meio dos operadores booleanos AND e OR, conforme as seguintes estratégias: "Inteligência Artificial" AND "Diagnóstico Clínico" OR "Diagnóstico por Imagem"; e seus correspondentes "*Artificial Intelligence*" AND "*Clinical Diagnosis*" OR "*Diagnostic Imaging*".

3.3. Critérios de Inclusão e Exclusão

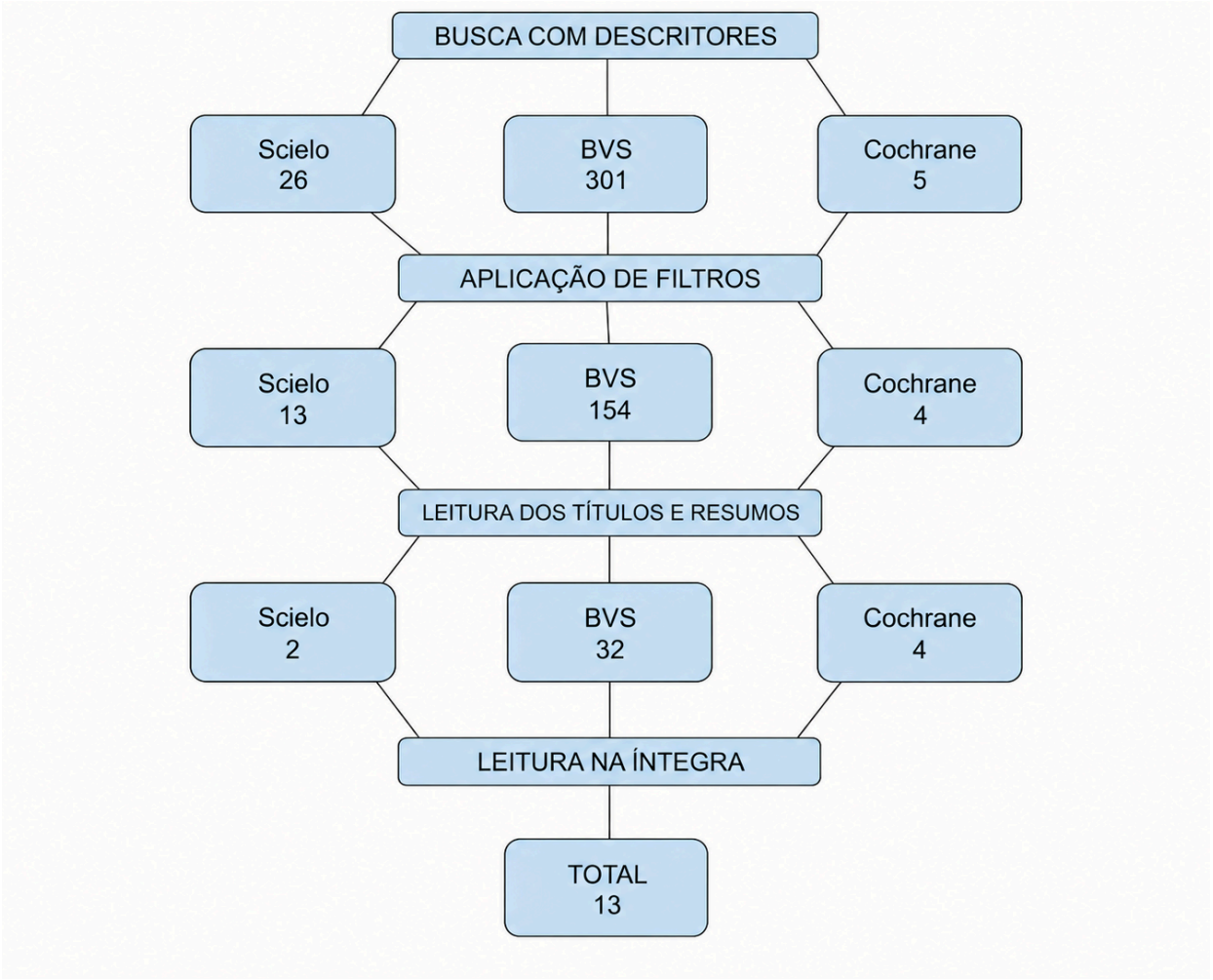
Foram incluídos estudos publicados entre 2021 e 2025, disponíveis na íntegra e de forma gratuita, que respondessem à pergunta norteadora e estivessem alinhados ao objetivo do estudo. Foram excluídos revisões e *overviews* , artigos duplicados e estudos que fugiam do objetivo da pesquisa.

3.4. Seleção dos Estudos

A busca inicial resultou em 332 artigos: 301 na BVS, 5 na Cochrane e 26 na SciELO. Após a aplicação dos filtros de ano de publicação (2021 a 2025) e de tipo de estudo, com exclusão de revisões e overviews, restaram 171 artigos (154 na BVS, 4 na Cochrane e 13 na SciELO).

Em seguida, realizou-se a leitura dos títulos e resumos, que resultou na pré-seleção de 38 artigos (32 na BVS, 4 na Cochrane e 2 na SciELO). Os artigos em língua estrangeira foram traduzidos, e as duplicatas foram removidas. Por fim, procedeu-se à leitura na íntegra dos textos, com exclusão dos que não atendiam ao objetivo do estudo e dos indisponíveis gratuitamente. A amostra final foi composta por 13 artigos, e o processo de seleção está representado na Figura 1.

Figura 1. Fluxograma das etapas de identificação, triagem e seleção dos estudos incluídos na revisão.



3.5. Extração e Organização dos Dados

Dos 13 artigos incluídos, extraíram-se as informações organizadas em dois quadros: o Quadro 1, com autor e ano, título, tipo de estudo e revista; e o Quadro 2, com artigo, metodologia e evidências. Esses dados fundamentam a análise e a discussão desta revisão integrativa.

3.6. Declaração de Uso de Inteligência Artificial

Durante a elaboração deste trabalho, foram utilizadas as ferramentas de Inteligência Artificial Claude (Anthropic) e ChatGPT (OpenAI) como apoio na formatação, na tradução e na sintetização de

informações. O uso dessas ferramentas se restringiu a tarefas de organização textual e estruturação do conteúdo, não tendo sido empregadas para geração de dados, resultados ou conclusões da pesquisa, que são de inteira responsabilidade dos autores.

4. RESULTADOS E DISCUSSÕES

Com a aplicação dos descritores e filtros nas bases de dados Cochrane, SciELO e BVS, foram identificados 332 artigos iniciais. Após as etapas de triagem, filtragem por ano de publicação, aplicação dos critérios de elegibilidade, leitura de títulos, resumos e textos na íntegra, foram incluídos no presente estudo 13 artigos (Figura 1). Os resultados da pesquisa foram organizados em dois quadros, detalhando: 1) A caracterização dos estudos selecionados conforme autores, ano de publicação, título, revista e tipo de estudo (Quadro 1); e 2) Os objetivos, metodologias empregadas e as principais evidências apontadas por cada estudo (Quadro 2).

Quadro 1. Caracterização dos artigos selecionados conforme autores e ano de publicação, título, revista e tipo de estudo dos respectivos estudos incluídos na revisão.

Autor/Ano	Título	Revista	Tipo de Estudo
Restini *et al.* (2024)	AI tool for predicting MGMT methylation in glioblastoma for clinical decision support in resource limited settings	Scientific Reports	Estudo retrospectivo de desenvolvimento e validação de modelos de aprendizado de máquina para predição do estado de metilação do gene MGMT em pacientes com glioblastoma.

Daza *et al.* (2024)	Evaluation of four chatbots in autoimmune liver disease: A comparative analysis	Annals of Hepatology	Estudo comparativo de avaliação de desempenho de quatro chatbot de IA, realizado por especialistas.
Andrade Júnior *et al.* (2023)	A new model based on artificial intelligence to screening preterm birth	The Journal of Maternal-Fetal & neonatal medicine	Estudo observacional, retrospectivo, de coorte, com desenvolvimento e avaliação de modelo de inteligência artificial para rastreamento de parto prematuro espontâneo.
Nogueira *et al.* (2025)	Artificial intelligence applied in identifying left ventricular walls in myocardial perfusion scintigraphy images: Pilot study	PLOS ONE	Estudo piloto prospectivo de desenvolvimento e validação de modelo de inteligência artificial para segmentação das paredes do ventrículo esquerdo em imagens de cintilografia de perfusão miocárdica.
Ureta *et al.* (2024)	Capacidad diagnóstica de algoritmo basado en la inteligencia artificial en la detección de patologías malignas mediante análisis de Mamografías	Journal of Health and Medical Sciences	Estudo observacional retrospectivo de avaliação diagnóstica, baseado na aplicação de algoritmo de inteligência artificial a imagens mamográficas e comparação com os

			resultados dos laudos médicos.
Prada *et al.* (2024)	Assessing Fuchs Corneal Endothelial Dystrophy Using Artificial Intelligence–Derived Morphometric Parameters From Specular Microscopy Images.	Cornea	Estudo observacional analítico de avaliação diagnóstica, utilizando inteligência artificial para análise de imagens de microscopia especular na avaliação da distrofia endotelial de Fuchs.
F. Segura; P. Segura; F. Segura. (2023)	Determinación del grado de agresividad tumoral mediante técnicas de inteligencia artificial aplicadas a imágenes de resonancia magnética.	Revista de la Asociación Argentina de Ortopedia y Traumatología	Estudo observacional retrospectivo de desenvolvimento e avaliação de modelo de inteligência artificial para classificação do grau de agressividade de sarcomas de partes moles por meio de imagens de ressonância magnética.
Castro *et al.* (2025)	Enhanced performance in automated diabetic retinopathy diagnosis achieved through Voronoi diagrams and artificial intelligence	Scientific Reports	Estudo experimental de desenvolvimento e avaliação de algoritmos de inteligência artificial para classificação da retinopatia diabética, utilizando diagramas de Voronoi e imagens da base MESSIDOR.

Bayona-Daza *et al.* (2025)	Concordancia de Inteligencia Artificial (ChatGPT, Bing) como predictor de apendicitis versus las escalas de Alvarado y AIR en pacientes adultos del Hospital Militar Central de Bogotá, Colombia	Revista Colombiana de Cirugía	Estudo observacional analítico de avaliação de desempenho diagnóstico, comparando ferramentas de inteligência artificial com as escalas clínicas de Alvarado e AIR na identificação de apendicite aguda.
Kilci *et al.* (2025)	Artificial intelligence vs. emergency physicians: who diagnoses better?	Revista da Associação Médica Brasileira (RAMB)	Estudo comparativo de desempenho diagnóstico entre modelos de inteligência artificial e médico especialista, utilizando casos clínicos simulados.
Santos *et al.* (2025)	White light imaging versus artificial intelligence-assisted white light imaging for colorectal neoplasia detection: a randomised trial	Revista de Gastroenterología del Perú	Ensaio clínico randomizado, prospectivo e unicêntrico, que comparou a colonoscopia com e sem auxílio de inteligência artificial.
Zaldivar *et al.* (2025)	Evaluating the Predictive Accuracy of an AI-Based Tool for Postoperative Vault Estimation in	Clinical Ophthalmology	Estudo retrospectivo, multicêntrico, de validação de ferramenta de inteligência artificial para predição do vault pós-operatório.

	Phakic Intraocular Lens Implantation		
Mendez-Avila *et al.* (2025)	Artificial intelligence in radiology, nuclear medicine and radiotherapy: Perceptions, experiences and expectations from the medical radiation technologists in Central and South America	Journal of Medical Imaging and Radiation Sciences	Estudo observacional transversal, realizado por meio de questionário eletrônico, para investigar percepções, experiências e expectativas de profissionais da área de radiação médica sobre a utilização da inteligência artificial.

Fonte: Elaborado pelos autores, 2026.

Quadro 2. Caracterização dos artigos selecionados conforme autores e ano de publicação, objetivo do estudo, metodologia empregada e principais evidências apontadas.

Autor/Ano	Objetivo do Estudo	Metodologia	Evidências
Restini *et al.* (2024)	Desenvolver e validar modelo de IA para predizer metilação do MGMT utilizando RM T1GD/T2-FLAIR, características radiômicas e dados clínicos.	Em uma coorte de treino selecionou-se 100 pacientes do banco público UPENN-GBM. Em outra coorte de validação, 46 pacientes de uma instituição brasileira. Foram extraídas 343 características radiômicas e clínicas, e aplicados 8	O modelo de aprendizado de máquina com melhor desempenho apresentou acurácia de 82%, Área Sob a Curva (Area Under the Curve – AUC) de 0,81, sensibilidade de 75% e precisão de 75% na predição

		métodos de seleção de variáveis. Foram testados 18 classificadores de machine learning (PyCaret) e 3 modelos de deep learning (CNN).	do estado de metilação do gene MGMT. Os resultados demonstram potencial para auxiliar a avaliação molecular de pacientes com glioblastoma, especialmente em contextos com recursos limitados, embora a ferramenta não substitua o exame molecular.
Daza *et al.* (2024)	Avaliar a utilidade de quatro chatbots como ferramentas de suporte à decisão clínica nas doenças hepáticas autoimunes.	Foi aplicado um questionário de 56 perguntas sobre hepatite autoimune, colangite biliar primária e colangite esclerosante primária (conhecimento geral, terapia e casos clínicos). As respostas, limitadas a 300 palavras, foram avaliadas por 10 hepatologistas da Europa e das Américas, em escala Likert de 1 a 10.	As respostas dos quatro chatbots foram avaliadas por dez hepatologistas, obtendo médias de 7,37 para Claude, 7,17 para ChatGPT, 6,63 para Copilot e 6,52 para Bard. Embora tenham apresentado desempenho considerado satisfatório em determinadas questões, foram identificadas respostas imprecisas e limitações na qualidade das orientações, evidenciando a necessidade de avaliação crítica

			por profissionais de saúde.
Andrade Júnior *et al.* (2023)	Desenvolver modelo de IA para rastreamento de parto prematuro espontâneo <35	Foram analisadas 524 gestações únicas, com dados clínicos e ultrassonográficos coletados entre 18 e 24 semanas de gestação. Essas informações foram utilizadas para desenvolver e avaliar um modelo de inteligência artificial capaz de prever o risco de parto prematuro, combinando diferentes variáveis maternas e o comprimento cervical.	O modelo de inteligência artificial apresentou AUC de 0,808, sensibilidade de 47,3% e especificidade de 92,8% para predição do parto prematuro espontâneo. O desempenho foi superior ao rastreamento baseado exclusivamente no comprimento cervical inferior a 25 mm, cuja sensibilidade foi de 33,3%. Apesar dos resultados promissores, a sensibilidade limitada indica a necessidade de validação adicional.
Nogueira *et al.* (2025)	Desenvolver e avaliar uma técnica de inteligência artificial, especificamente uma rede neural convolucional nnU-Net, para identificar automaticam	Foram coletados 83 exames de rotina (50 cortes cada, 4.150 imagens). Dois especialistas fizeram a segmentação manual no 3D Slicer, usada como padrão de referência. O modelo nnU-Net 2D foi treinado em 31 exames, testado em	O modelo de inteligência artificial demonstrou elevada concordância com a segmentação manual de referência na identificação das paredes do ventrículo

	ente as paredes do ventrículo esquerdo em imagens de cintilografia de perfusão miocárdica.	9 (validação interna) e validado em 43 (validação externa). O desempenho foi medido pelo coeficiente de Dice. Três médicos nucleares avaliaram, de forma cega, as máscaras geradas pela IA e as manuais.	esquerdo em imagens de cintilografia de perfusão miocárdica, avaliada pelo coeficiente de Dice. Os resultados indicam potencial para automatizar etapas do processamento de imagens cardíacas, mas não demonstram diretamente melhora da acurácia diagnóstica de doenças cardiovasculares.
Ureta *et al.* (2024)	Avaliar a capacidade diagnóstica da inteligência artificial na detecção de patologias malignas por meio da análise de mamografias.	Foram analisadas 100 mamografias digitais de pacientes do sexo feminino, realizadas entre 2022 e 2023. As imagens foram processadas por um algoritmo de IA baseado em rede neural convolucional, denominado Breast Cancer Classifier, que estimou a probabilidade de malignidade e gerou mapas de calor para localizar áreas suspeitas. Os resultados foram comparados com os	Na análise retrospectiva de 100 exames mamográficos, o algoritmo de inteligência artificial demonstrou capacidade de identificar e localizar achados suspeitos de malignidade. Entretanto, foram observadas divergências em relação aos laudos médicos, incluindo um resultado falso negativo. Os achados sugerem

		laudos radiológicos e as classificações BI-RADS.	potencial como ferramenta complementar à interpretação mamográfica, mas indicam a necessidade de aprimoramento e validação antes de sua utilização independente.
Prada *et al.* (2024)	Avaliar a eficácia de parâmetros morfométricos derivados de inteligência artificial na caracterização da distrofia endotelial corneana de Fuchs a partir de imagens de microscopia especular.	Estudo observacional transversal que recrutou 52 pacientes, correspondentes a 104 olhos, dos quais 76 foram analisados. Foram utilizadas imagens de microscopia especular para obtenção de parâmetros morfométricos por inteligência artificial, posteriormente comparados com a avaliação clínica da distrofia endotelial de Fuchs.	Os parâmetros morfométricos obtidos por inteligência artificial apresentaram associação com a gravidade clínica da distrofia endotelial de Fuchs. A razão da área ocupada por guttae apresentou correlação de 0,60 com a classificação clínica modificada de Krachmer na região central da córnea, indicando potencial para auxiliar a avaliação objetiva da doença.
F. Segura; P. Segura; F. Segura (2023)	Avaliar a capacidade de técnicas de inteligência artificial aplicadas a imagens de ressonância magnética	Foram analisadas imagens de ressonância magnética de 73 pacientes com sarcomas de partes moles, provenientes de duas bases de dados. Foram	O modelo alcançou precisão média de 84,3% e sensibilidade de 73,3% na classificação da agressividade tumoral. Os resultados

	para determinar o grau de agressividade de sarcomas de partes moles, verificando seu potencial como ferramenta auxiliar no diagnóstico e na avaliação pré-operatória.	selecionadas 134 imagens representativas dos tumores, classificadas conforme o grau histológico, para treinar, validar e testar um modelo de IA capaz de diferenciar tumores de baixo e alto grau de agressividade.	indicaram potencial para auxiliar na avaliação pré-operatória dos sarcomas, oferecendo uma segunda opinião antes da biópsia. Entretanto, o número reduzido de imagens e a utilização de dados retrospectivos limitaram os resultados, sendo necessária validação com bases de dados maiores.
Castro *et al.* (2025)	Desenvolver e avaliar um algoritmo de IA para diagnóstico automatizado da retinopatia diabética, utilizando diagramas de Voronoi para analisar a distribuição de microaneurismas em imagens de fundo de olho.	A metodologia do estudo consistiu em uma pesquisa experimental, baseada no desenvolvimento e na validação de modelos de inteligência artificial para análise de imagens oftalmológicas. Foram utilizadas 800 imagens do banco de dados MESSIDOR, submetidas ao processamento e à análise computacional. Posteriormente, diferentes	A utilização de diagramas de Voronoi associados a algoritmos de inteligência artificial demonstrou desempenho promissor na classificação da retinopatia diabética, com AUC de 0,964 para o classificador de melhor desempenho. Os resultados indicam potencial para auxiliar a análise automatizada de imagens retinianas, embora sejam

		algoritmos de aprendizado de máquina foram aplicados e comparados, considerando seu desempenho na classificação das imagens e sua capacidade de auxiliar no diagnóstico de alterações retinianas.	necessários estudos adicionais para avaliar sua aplicabilidade na prática clínica.
Bayona-Daza *et al.* (2025)	Avaliar o desempenho de ferramentas de inteligência artificial na identificação de apendicite aguda em pacientes adultos, comparando seus resultados com as escalas clínicas de Alvarado e AIR.	Estudo observacional analítico de avaliação diagnóstica, realizado com 172 pacientes adultos atendidos por suspeita de apendicite aguda. Foram comparados os resultados de ferramentas de inteligência artificial com os das escalas clínicas de Alvarado e AIR, utilizando o diagnóstico anatomopatológico como referência para avaliar o desempenho diagnóstico.	Em 172 pacientes, 129 (75%) tiveram apendicite confirmada por anatomopatológico. O ChatGPT apresentou baixa concordância com o diagnóstico (Kappa = 0,005), sensibilidade de 10,1% e especificidade de 90,7%, apresentando desempenho inferior às escalas de Alvarado e AIR. Conclui-se que o ChatGPT apresentou baixa acurácia diagnóstica para apendicite, com desempenho inferior às escalas clínicas de Alvarado e AIR.

Kilci *et al.* (2025)	Comparar a acurácia diagnóstica e a capacidade de escolher o exame diagnóstico inicial entre modelos de IA e um médico especialista em medicina de emergência.	Estudo comparativo realizado com casos clínicos simulados, no qual foram avaliados o desempenho diagnóstico e a escolha do primeiro exame complementar por um médico especialista em medicina de emergência e por três modelos de inteligência artificial: GPT-4, GPT-4o e o3-mini.	A acurácia diagnóstica foi de 92% para o médico, 97% para ChatGPT-4 e 99% para ChatGPT-4o e o3-mini. Para escolha do primeiro exame: 88% médico, 80% ChatGPT-4, 87% ChatGPT-4o e 89% o3-mini, sem diferença estatisticamente significativa. Importante: foram casos simulados, sem pacientes reais, e apenas um médico participou.
Santos *et al.* (2025)	Comparar a detecção de neoplasias colorretais (taxa de detecção de adenomas e outros indicadores) entre colonoscopia com luz branca (WLI) e colonoscopia com luz branca assistida por inteligência artificial (WLI+IA).	O ensaio foi realizado em Bagé (Brasil) de maio a dezembro de 2024, com pacientes encaminhados para colonoscopia de rastreamento, vigilância ou diagnóstico. Dos 844 elegíveis, 711 foram analisados: 357 no grupo luz branca (WLI) e 354 no grupo luz branca assistida por IA (WLI+IA, sistema CAD EYE, Fujifilm). Todos os exames foram feitos por um único endoscopista experiente. O	A taxa de detecção de adenomas foi de 45,9% no grupo de colonoscopia convencional e 50,8% no grupo assistido por inteligência artificial, sem diferença estatisticamente significativa (p = 0,20). Também não foram observadas diferenças significativas nos demais desfechos avaliados. Entre os pacientes submetidos à colonoscopia de rastreamento,

		desfecho primário foi a taxa de detecção de adenomas (ADR). Os secundários foram a taxa de detecção de lesões serradas sésseis (SDR), o número médio de adenomas por paciente (MAP), a taxa de detecção de neoplasias (NDR), a taxa de detecção de adenomas avançados (AADR) e o tempo de retirada do colonoscópio. A análise usou teste exato de Fisher ou qui-quadrado e Mann-Whitney, com significância de 5%.	houve diferença numérica favorável à IA (61,5% versus 49,2%), mas sem significância estatística (p = 0,06).
Zaldivar *et al.* (2025)	Avaliar a acurácia preditiva de uma ferramenta baseada em inteligência artificial (ICLGuru) que integra biomicroscopia ultrassônica (UBM) com dados biométricos para estimar o vault pós-operatório em pacientes	O estudo foi realizado com 347 olhos de 228 pacientes míopes (idade média de 31,3 ± 7,7 anos), submetidos a implante de lente intraocular fácica (ICL EVO Visian V4c) em quatro centros (Argentina, Irlanda, Bélgica e Filipinas). Os parâmetros biométricos pré-operatórios e as imagens de biomicroscopia ultrassônica (UBM)	No 1º dia, o vault clínico médio foi de 520,97 ± 178,73 µm e o previsto, de 508,16 ± 163,00 µm, com diferença média de −12,81 µm (r^2 = 0,621; p < 0,001). Em 53,45% dos olhos a diferença foi menor que 100 µm. Em 1 mês (291 olhos), a correlação foi menor (r^2 = 0,530). Não houve diferença significativa entre os centros nos

	míopes submetidos a implante de lente intraocular fácica (ICL).	alimentaram a ferramenta de IA ICLGuru, que previu o vault (distância entre a ICL e o cristalino). As previsões foram comparadas com as medidas clínicas de 1 dia e 1 mês de pós-operatório. A análise usou correlação de Spearman, regressão linear multivariável, gráficos de Bland-Altman e testes para comparar os centros.	vaults clínico e previsto ($p > 0,05$). ARise e poder esférico foram preditores significativos da discrepância, mas o modelo explicou só 9,3% da variância. Também houve viés proporcional ($r^2 = 0,254$). Os autores concluem que a ferramenta prevê o vault com acurácia clinicamente aceitável. Limitações: estudo retrospectivo, medidas sob iluminação fotópica e conflito de interesse.
Mendez-Avila *et al.* (2025)	Compreender as percepções, experiências e expectativas de tecnólogos em radiação médica (MRTs) da América Central e do Sul sobre a implementação da inteligência artificial em radiologia,	Aplicou-se um questionário online (Microsoft Forms) em espanhol e português, adaptado de um levantamento global, a tecnólogos em radiação médica que atuam em radiologia, medicina nuclear e radioterapia na América Central e do Sul. A coleta ocorreu entre setembro e outubro de 2024, com divulgação em redes sociais e	Entre os 398 participantes, 98,5% concordaram com a implementação da inteligência artificial na prática clínica, embora apenas 14,6% relatassem experiência prévia com programas específicos de IA e 30,2% se considerassem totalmente preparados para sua utilização. Os

	medicina nuclear	WhatsApp e incentivo para que os participantes repassassem o convite a colegas. Das 631 respostas, 398 foram consideradas válidas. O instrumento reunia perguntas fechadas e abertas sobre dados demográficos, experiência com IA, impacto na profissão, percepções e opiniões. As respostas fechadas foram analisadas por estatística descritiva (frequências absolutas e relativas) no Excel, e as abertas, por codificação qualitativa no ATLAS.ti 25.	resultados evidenciam elevada aceitação da tecnologia, acompanhada de limitações relacionadas à capacitação profissional. Entretanto, o estudo avaliou percepções e experiências dos profissionais, não o desempenho diagnóstico da inteligência artificial.

Fonte: Elaborado pelos autores, 2026.

4.1. Caracterização dos Estudos e Qualidade Metodológica

Os 13 estudos incluídos concentram-se nos anos mais recentes: dois são de 2023, três de 2024 e oito de 2025, sem nenhum de 2021 ou 2022. É natural associar esse crescimento ao interesse por chatbots, presentes em Daza *et al.* (2024), Bayona-Daza *et al.* (2025) e Kilcin *et al.* (2025).

Os desenhos metodológicos dos estudos analisados foram diversos, sendo a maioria estudos retrospectivos e voltada ao desenvolvimento e à validação de modelos, como em Restini *et al.*, Andrade Júnior *et al.*, F. Segura; P. Segura; F. Segura (2023) e Zaldivar *et al.* Por sua vez, Santos e colaboradores (2025) conduziram um ensaio clínico randomizado, enquanto Nogueira *et al.* fizeram um piloto prospectivo, e Kilcin *et al.* e Mendez-Avila *et al.* realizaram estudos transversais. Essa variedade metodológica é útil, pois cada formato responde a uma pergunta diferente: se o algoritmo funciona, se supera médicos ou escalas clínicas e como os profissionais o recebem. Entretanto, o peso clínico dos dados é desigual. Santos *et al.* acompanharam 711 pacientes e Bayona-Daza *et al.* confirmaram o diagnóstico de 172 por anatomopatológico, enquanto Castro *et al.* trabalharam com um banco público de imagens e Kilcin *et al.* com casos simulados, que mostram o potencial técnico da IA, mas pouco dizem sobre o que acontece com pacientes reais.

A validação externa é rara. Restini *et al.* e Nogueira *et al.* a fizeram, e Zaldivar *et al.* incluíram quatro centros, mas mesmo assim há fragilidades: Restini *et al.* observaram desempenho diferente entre as coortes, com possível overfitting, e F. Segura; P. Segura; F. Segura (2023) partiu de apenas 134 imagens de 73 pacientes.

Também é preciso cuidado para não confundir bons números com eficácia clínica. A AUC de 0,964 de Castro *et al.* e a acurácia de 99% do ChatGPT-4o em Kilcin *et al.* vêm de um banco de imagens e de casos simulados. Por outro lado, Bayona-Daza *et al.* encontraram sensibilidade de apenas 10,1% do ChatGPT na apendicite, e Santos et al. não observaram diferença significativa na detecção de adenomas com IA (45,9% vs. 50,8%; p = 0,20).

O conjunto é promissor, mas heterogêneo e ainda insuficiente para uso clínico. Os achados sustentam a IA como apoio ao julgamento médico, não como substituta, e justificam estudos prospectivos em cenários reais.

4.2. Principais Aplicações e Especialidades Médicas

A inteligência artificial tem se consolidado como uma ferramenta de suporte à decisão clínica e de predição em cenários de recursos limitados, auxiliando na identificação de marcadores prognósticos em neuro-oncologia (Restini *et al.*, 2024). Modelos de aprendizado de máquina aplicados a dados clínicos e de imagem demonstram potencial na triagem e no auxílio ao manuseio de patologias neurológicas complexas (Restini *et al.*, 2024). Desse modo, o uso dessas tecnologias funciona como um teste preliminar que otimiza a triagem de pacientes de alto risco, embora ainda não substitua os exames moleculares definitivos (Restini *et al.*, 2024).

No campo da hepatologia e da obstetrícia, algoritmos e modelos baseados em inteligência artificial auxiliam na avaliação de doenças hepáticas crônicas e no rastreamento precoce de intercorrências gestacionais (Daza *et al.*, 2024; Andrade Júnior *et al.*, 2023). Ferramentas computacionais e chatbots integrados demonstram utilidade no suporte a diagnósticos hepáticos, exigindo, contudo, validação crítica do corpo clínico (Daza *et al.*, 2024). Paralelamente, o cruzamento de variáveis clínicas e ultrassonográficas permite prever desfechos desfavoráveis na gravidez com alta especificidade, otimizando o acompanhamento pré-natal (Andrade Júnior *et al.*, 2023).

A cardiologia e a oftalmologia destacam-se pelo uso de redes neurais convolucionais e ferramentas de segmentação avançadas voltadas à análise minuciosa de exames diagnósticos (Nogueira *et al.*, 2025; Prada *et al.*, 2024; Castro *et al.*, 2025). A identificação automatizada de estruturas cardíacas em exames de imagem e a caracterização morfométrica de afecções corneanas ou retinianas evidenciam a precisão dos algoritmos na detecção de alterações anatômicas sutis (Castro *et al.*, 2025; Nogueira *et al.*, 2025; Prada *et al.*, 2024). Essas aplicações podem reduzir a subjetividade na interpretação de exames oftalmológicos e cardiológicos, favorecendo o diagnóstico precoce de lesões orgânicas (Nogueira *et al.*, 2025; Prada *et al.*, 2024; Castro *et al., 20*25).

Na radiologia e na medicina de emergência, a implementação de softwares de inteligência artificial abrange desde a detecção de malignidades mamárias e sarcomas até o suporte na interpretação de imagens torácicas e na condução de casos de urgência (BaBayona-Daza *et al.*, 2025; F. Segura; P. Segura; F. Segura, 2023; Kilcin *et al.*, 2025; eta *et al.*, 2024). Algoritmos de classificação e mapas de calor auxiliam na triagem de lesões suspeitas e na otimização de fluxos de atendimento em diferentes contextos globais de saúde (F. Segura; P. Segura; F. Segura, 2023; Ureta *et al.*, 2024). No entanto, o desempenho variável de grandes modelos de linguagem e ferramentas de suporte em emergências e cirurgias aponta para a necessidade de validações rigorosas antes de sua ampla adoção na rotina de pronto-socorro (Bayona-Daza *et al.*, 2025; Kilcin *et al.*, 2025).

Por fim, especialidades como a gastroenterologia e a oftalmologia cirúrgica beneficiam-se de inovações voltadas ao rastreamento endoscópico e ao planejamento de procedimentos oculares

complexos (Mendez-Avila *et al.*, 2025; Santos *et al.*, 2025; Zaldivar *et al.*, 2025). Por sua vez, sistemas de auxílio em colonoscopias e ferramentas preditivas baseadas em biomicroscopia ultrassônica aumentam a acurácia na detecção de neoplasias e na estimativa de parâmetros pós-operatórios (Santos *et al.*, 2025; Zaldivar et al., 2025). Contudo, profissionais da área ressaltam que a efetividade dessas inovações depende diretamente de programas de capacitação adequados e de uma sinergia contínua entre a experiência humana e a automação digital (Mendez-Avila *et al.*, 2025).

4.3. Principais Benefícios e Limitações Demonstrados Pela IA nos Estudos Analisados

Os estudos incluídos mostram que a IA tem bom desempenho técnico em tarefas bem delimitadas, principalmente na análise de imagens (Castro *et al,* 2025; F. Segura; P. Segura; F. Segura, 2023; Nogueira *et al.,* 2025; Restini *et al.,* 2024) Entretanto, quase todas as evidências dizem respeito a métricas de desempenho, e poucos trabalhos comparam a ferramenta com a conduta clínica estabelecida. No único ensaio clínico randomizado da amostra, o benefício não foi demonstrado (Santos *et al, 2025*).

Entre os indicadores diagnósticos mais favoráveis está o de Castro *et al.* (2025), que combinaram o classificador *Decision Tree* a diagramas de Voronoi para o diagnóstico automatizado de retinopatia diabética e obtiveram Área Sob a Curva (Area Under the Curve – AUC) de 0,964, sensibilidade de 93,6%, especificidade de 87,6%, acurácia de 90,9% e F1-score de 92,6%. Os próprios autores apontam que os dados vêm de uma base de imagens (MESSIDOR), e não de pacientes avaliados prospectivamente. Esse tipo de resultado indica que o algoritmo separa bem os casos dentro de uma amostra

conhecida, mas nada diz sobre o efeito na conduta ou no desfecho do paciente, e os valores dependem da prevalência da doença e da qualidade das imagens.

No estudo realizado por Restini e colaboradores. (2024), o melhor modelo para predizer a metilação do MGMT alcançou, em dados não vistos, acurácia de 0,82, AUC de 0,81, sensibilidade de 0,75 e precisão de 0,75. Assim, os autores o propõem como teste preliminar, e não como substituto do teste molecular.

No contexto da detecção e caracterização de alterações patológicas. Ureta *et al*. (2024) relataram que o algoritmo de IA identificou achados malignos em 12 das 100 mamografias, com concordância de 99% nos casos com BI-RADS ≥ 4 considerados na comparação apresentada pelos autores. No estudo realizado por F. Segura; P. Segura; F. Segura (2023), a IA classificou a agressividade de sarcomas de partes moles com precisão média de 84,3% e sensibilidade de 73,3%, o que sugere uma possível segunda opinião antes da biópsia. No âmbito da avaliação de alterações corneana, Prada e colaboradores (2024) observaram que a proporção da área ocupada por *guttae* acompanhava a gravidade clínica da distrofia endotelial de Fuchs e que a IA estimou a densidade celular endotelial de modo distinto do software convencional. Em todos esses estudos, a capacidade de detectar ou classificar foi demonstrada; o efeito dessa capacidade sobre a evolução dos pacientes, não.

A padronização de tarefas repetitivas aparece com mais clareza em Nogueira *et al.* (2025). A rede nnu-Net segmentou as paredes do ventrículo esquerdo na cintilografia de perfusão miocárdica com coeficiente de Dice de 87,01% na validação interna e de 88,46% na externa, e os especialistas concordaram de 80% a 100% com as

máscaras de segmentação geradas. Para os autores, a precisão é comparável à da segmentação manual, o que poderia facilitar a futura detecção de artefatos. Esse ganho, porém, permanece hipotético, já que a detecção de artefatos não foi avaliada.

Nas aplicações de predição, Andrade Júnior *et al.* (2023) obtiveram AUC de 0,808 e especificidade de 92,8% para o parto prematuro espontâneo em 524 gestações únicas. Como o quadro não informa a sensibilidade, não se sabe quantas gestantes de fato em risco foram identificadas. No estudo de Zaldivar *et al.* (2025), foi comparado o vault previsto pela ferramenta ICLGuru com o medido no primeiro dia: 508,16 ± 163,00 µm contra 520,97 ± 178,73 µm, com diferença média de −12,81 µm (r^2 = 0,621; $p < 0,001$). A diferença foi menor que 100 µm em 53,45% dos olhos, de modo que quase metade das estimativas ficou fora desse limiar, e o modelo explicou apenas 9,3% da variância da discrepância.

No apoio à decisão clínica, Daza *et al.* (2024) pediram a 10 hepatologistas que avaliassem, em escala Likert de 1 a 10, as respostas de quatro chatbots sobre doenças hepáticas autoimunes. As médias foram 7,37 (Claude), 7,17 (ChatGPT), 6,63 (Copilot) e 6,52 (Bard), valores intermediários e relativamente próximos. No estudo conduzido por Kilcin *et al.* (2025) foi verificado acurácia diagnóstica de 92% para o médico, 97% para o ChatGPT-4 e 99% para o ChatGPT-4o e o o3-mini. Esses números não sustentam a superioridade da IA, pois os casos eram simulados e apenas um médico participou. Na escolha do primeiro exame (88% para o médico; 80%, 87% e 89% para os modelos), a diferença não foi estatisticamente significativa. Quanto ao potencial de implementação, 98,5% (n = 392) dos 398 respondentes válidos no estudo de Mendez-Avila *et al.* (2025)

concordaram com a adoção da IA na prática clínica. A aceitação dos profissionais, entretanto, não é prova de eficácia.

As limitações começam pelos erros de classificação. Em Ureta *et al.* (2024), houve um falso negativo, o que mostra que o algoritmo pode deixar passar lesões suspeitas. Em Bayona-Daza *et al.* (2025), entre 172 pacientes, dos quais 129 (75%) tinham apendicite confirmada por anatomopatológico, o ChatGPT apresentou Kappa de 0,005, sensibilidade de 10,1% e especificidade de 90,7%; ou seja, deixou de identificar a maior parte dos casos confirmados e teve desempenho inferior ao das escalas de Alvarado e AIR. Também ficou aquém do esperado o desempenho dos modelos de *deep learning* de Restini *et al.* (2024), com acurácia de apenas 51% a 54%.

A generalização dos resultados é outro ponto frágil. Restini *et al.* (2024) observaram variação de desempenho entre as coortes, com possível *overfitting,* enquanto Nogueira *et al*. (2025) não relataram queda do Dice da validação interna para a externa, embora a amostra fosse pequena. Andrade Júnior *et al.* (2023) ressaltam que o desempenho deve ser interpretado no contexto da população analisada e exige validação adicional. Os limites de desenho e amostra são frequentes: Nogueira *et al.* (2025) reconhecem a amostra pequena e o uso de limiarização; F. Segura; P. Segura; F. Segura (2023) trabalhou com 134 imagens de 73 pacientes, em dados retrospectivos; Zaldivar *et al.* (2025) apontam o caráter retrospectivo, as medidas sob iluminação fotópica e um conflito de interesse; e Castro *et al.* (2025) usaram uma única base de imagens.

Sobre a qualidade dos dados, Prada *et al*. (2024) excluíram imagens de baixa qualidade, e Mendez-Avila *et al*. (2025) advertem que a amostragem não probabilística e o provável maior conhecimento

em IA dos respondentes limitam a generalização. Nos chatbots, Daza *et al.* (2024) descreveram como falhas mais comuns a listagem de informações no lugar de orientações específicas, imprecisões (como na gestação), dados desatualizados e pouca discussão sobre uso *off-label*; e o acesso à internet em tempo real (Copilot e Bard) não melhorou a qualidade.

Colocados lado a lado, os resultados favoráveis e desfavoráveis indicam que o desempenho varia com a tarefa e com o modelo. Na própria pesquisa de Restini *et al.* (2024), um modelo com acurácia de 0,82 convive com redes de *deep learning* em torno de 51% a 54%, o que sugere que algoritmos mais sofisticados não garantem melhor resultado quando a amostra é reduzida. Em Ureta *et al.* (2024), o potencial para identificar e localizar alterações mamárias foi acompanhado de um falso negativo, argumento a favor do uso como apoio, e não como instância independente, sobretudo em rastreamento. O contraste entre Bayona-Daza *et al.* (2025) e Kilcin *et al.* (2025) é o mais instrutivo. No primeiro, com pacientes reais e confirmação anatomopatológica, o ChatGPT teve baixa sensibilidade e ficou atrás das escalas clínicas; no segundo, em casos simulados, os modelos superaram numericamente o médico. É provável que a diferença reflita o cenário, pois casos simulados tendem a trazer informações organizadas, enquanto a prática real lida com dados incompletos e ambíguos.

O ensaio de dos Santos *et al.* (2025) é o único que testou a IA como intervenção, e seu resultado destoa das métricas favoráveis dos estudos de imagem. A taxa de detecção de adenomas (ADR) foi de 45,9% com luz branca (WLI) e de 50,8% com luz branca assistida por IA (WLI+IA), $p = 0,20$, e os demais desfechos secundários, incluindo o tempo de retirada (12,4 vs. 12,2 min; $p = 0,32$), também não diferiram

significativamente. Entre os pacientes de rastreamento, a ADR foi de 61,5% com IA e 49,2% sem IA (p = 0,06), significância marginal. Esse resultado não permite concluir que as estratégias sejam equivalentes, nem tratar a diferença numérica do subgrupo como efeito demonstrado. Os autores atribuem a ausência de diferença global ao fato de o endoscopista ser um alto detector de adenomas e citam como limitações a participação de um único profissional e sua grande experiência com IA.

A percepção dos profissionais também merece leitura cuidadosa. Em Mendez-Avila *et al.* (2025), apenas 14,6% já haviam usado programas específicos de IA, 30,2% se sentiam totalmente preparados e somente 5,6% concordaram fortemente que há treinamento suficiente. Ainda assim, 65,8% apoiaram fortemente a inclusão da IA no currículo e 82,4% defenderam a sinergia entre IA e profissional na interpretação das imagens. Entre os respondentes, 19,6% consideraram a IA uma ameaça, e as respostas abertas, nas quais predominou a visão de oportunidade, trouxeram preocupação com a substituição de funções e a desumanização do cuidado. A adesão, portanto, parece anteceder a experiência e a capacitação. Em Daza *et al.* (2024), os autores concluem que os chatbots exigem julgamento crítico do profissional e aprimoramento antes do uso clínico, já que as limitações de qualidade atingiram todas as ferramentas, apesar das diferenças entre elas.

Do conjunto, depreende-se que a IA, no estágio atual das evidências, funciona como ferramenta complementar ao raciocínio clínico, à interpretação de exames e à tomada de decisão, e não como substituta do profissional. É a posição de vários autores: Restini *et al.* (2024) ressalvam que a ferramenta não substitui o teste molecular, Ureta *et al.* (2024) apontam a necessidade de validação antes do uso

clínico independente e Daza *et al.* (2024) enfatizam o julgamento crítico. Como foram documentados falsos negativos, desempenho inferior ao de escalas clínicas e respostas imprecisas de chatbots, confiar de modo acrítico em resultados automatizados é arriscado, e a supervisão profissional e a validação local e externa tornam-se condições para o uso responsável. Convém, ainda, não confundir três planos: o potencial tecnológico, ilustrado pela segmentação automatizada de Nogueira *et al.* (2025); o desempenho diagnóstico, representado por métricas como as de Castro *et al.* (2025) e Restini *et al.* (2024); e o benefício clínico comprovado, que, entre os estudos analisados, não foi demonstrado.

5. CONCLUSÃO E CONSIDERAÇÕES FINAIS

Os 13 estudos incluídos nesta revisão indicam que a IA pode auxiliar o diagnóstico principalmente como ferramenta de apoio à interpretação de exames de imagem, à detecção e classificação de alterações patológicas, à predição de riscos e desfechos clínicos e ao suporte à decisão, inclusive por meio de chatbots. Essas aplicações distribuíram-se por especialidades como radiologia, oftalmologia, cardiologia, gastroenterologia, obstetrícia, hepatologia e neuro-oncologia. Entretanto, a interpretação dos achados apontados exige cautela, pois o desempenho técnico de um algoritmo não equivale a benefício clínico comprovado. Nesse cenário, a IA deve ser compreendida como ferramenta complementar ao raciocínio clínico, e não como substituta do médico. O uso responsável pressupõe supervisão profissional, validação local e externa dos sistemas e capacitação dos profissionais. Acrescentam-se a proteção dos dados dos pacientes e a atenção aos vieses algorítmicos, de modo que a incorporação da tecnologia preserve a segurança, a equidade e a responsabilidade profissional.

Esta revisão apresenta limitações. A busca restringiu-se a três bases de dados (Cochrane, SciELO e BVS) e a estudos disponíveis na íntegra e de forma gratuita, o que pode ter excluído publicações relevantes. A amostra final, de 13 artigos, é pequena, e a heterogeneidade dos delineamentos, das especialidades e das aplicações dificultou a comparação dos resultados entre diferentes sistemas e contextos clínicos, impedindo qualquer síntese quantitativa.

Diante disso, são necessários estudos prospectivos e multicêntricos, realizados com pacientes reais, com validação externa em diferentes populações e comparação com os métodos diagnósticos já estabelecidos. Tais pesquisas devem avaliar desfechos clinicamente relevantes, a segurança, a aplicabilidade e o impacto da tecnologia na rotina dos serviços de saúde.

Conclui-se, portanto, que a inteligência artificial apresenta potencial para auxiliar o diagnóstico médico, sobretudo na análise de imagens, na predição de riscos e no apoio à decisão clínica, mas as evidências reunidas ainda não demonstraram benefício clínico consistente para os pacientes. Seu emprego deve ser prudente, supervisionado e subordinado ao julgamento do médico, até que estudos mais robustos confirmem sua efetividade e segurança na prática assistencial.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE JÚNIOR, Valter Lacerda de et al. A new model based on artificial intelligence to screening preterm birth. **The journal of maternal-fetal & neonatal medicine: the official journal of the European Association of Perinatal Medicine, the Federation of**

Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians, v. 36, n. 2, p. 2241100, dez. 2023.

BARBOSA, L. M.; PORTES, L. A. F. A inteligência artificial. **Revista Tecnologia Educacional,** Rio de Janeiro, n. 236, p. 16–27, 2023.

BAYONA-DAZA, Álvaro Javier; MENDIZÁBAL-VARONA, Erika Tatiana; MARÍN, Jaime Andrés. Concordancia de Inteligencia Artificial (ChatGPT, Bing) como predictor de apendicitis versus las escalas de Alvarado y AIR en pacientes adultos del Hospital Militar Central de Bogotá, Colombia. **Revista Colombiana de Cirugía,** v. 40, n. 6, p. 1080–1087, 14 out. 2025.

BEAM, A. L. et al. Artificial intelligence in medicine. **New England Journal of Medicine,** v. 388, n. 13, p. 1220–1221, 30 mar. 2023.

BRAUN, Leah T. et al. Diagnostic errors by medical students: results of a prospective qualitative study. **BMC Medical Education,** v. 17, n. 1, 9 nov. 2017.

CARVALHO, T. R.; ARAÚJO, F. M. Aplicações da inteligência artificial no diagnóstico médico remoto. **Revista Multidisciplinar UNIPACTO,** v. 15, n. 2, p. 233–242, 2022.

CASTRO, Mac Gayver da Silva et al. Enhanced performance in automated diabetic retinopathy diagnosis achieved through Voronoi diagrams and artificial intelligence. **Scientific Reports,** v. 15, n. 1, 14 out. 2025.

DAZA, Jimmy et al. Evaluation of four chatbots in autoimmune liver disease: A comparative analysis. **Annals of Hepatology,** p. 101537–101537, 1 ago. 2024.

ESR. European Society of Radiology. The impact of artificial intelligence on clinical practice and diagnostic safety. **Insights into Imaging,** v. 10, n. 1, p. 1–8, 2019.

FERREIRA, J. L.; LIMA, R. S.; SOUSA, D. P. O uso da inteligência artificial no diagnóstico clínico: revisão integrativa. **Brazilian Journal of Health and Science,** v. 8, n. 4, p. 45–53, 2022.

GRABER, M. L.; RUSSELL, A. T.; GRABER, M. A. The epidemiology of diagnostic error in medicine. **BMJ Quality & Safety,** v. 27, n. 6, p. 465–472, 2018.

IRANILDA, M. et al. Impacto da inteligência artificial na diagnóstico médico: desafios e oportunidades. **Revista Ibero-Americana de Humanidades, Ciências e Educação,** v. 10, n. 1, p. 1477–1485, 26 fev. 2024.

JIANG, Fei et al. Artificial Intelligence in Healthcare: Past, Present and Future. **Stroke and Vascular Neurology,** v. 2, n. 4, p. 230–243, 21 jun. 2017.

KILCI, Ali İhsan et al. Artificial intelligence vs. emergency physicians: who diagnoses better? **Revista da Associação Médica Brasileira,** v. 71, n. 11, 2025.

LEE, K. S.; BEZUIDENHOUT, A. F.; KRUSKAL, J. B. Spectrum of cognitive biases in diagnostic radiology. **RadioGraphics,** publicado online em jun. 2024.

LI, Y.-H. et al. Innovation and challenges of artificial intelligence technology in personalized healthcare. **Scientific Reports,** v. 14, n. 1, 16 ago. 2024.

MARQUES, D. N. et al. Saúde Digital: O potencial da telemedicina e da inteligência artificial. **Interference: A journal of audio culture,** v. 11, n. 2, p. 6954–6973, 7 nov. 2025.

MENDEZ-AVILA, Catalina et al. Artificial intelligence in radiology, nuclear medicine and radiotherapy: Perceptions, experiences and expectations from the medical radiation technologists in Central and South America. **Journal of Medical Imaging and Radiation Sciences,** v. 56, n. 6, p. 102081, dez. 2025.

MOURA, A. C.; COSTA, V. P. Aprendizado de máquina e predição de desfechos clínicos: uma revisão sistemática. **Revista de Neurociências,** v. 31, n. 1, p. 12–21, 2023.

NOGUEIRA, Solange Amorim et al. Artificial intelligence applied in identifying left ventricular walls in myocardial perfusion scintigraphy images: Pilot study. **PLOS ONE,** v. 20, n. 1, p. e0312257, 17 jan. 2025.

PRADA, Angelica M. et al. Assessing Fuchs Corneal Endothelial Dystrophy Using Artificial Intelligence–Derived Morphometric Parameters From Specular Microscopy Images. **Cornea,** v. 43, n. 9, p. 1080–1087, 9 fev. 2024.

RAJKOMAR, A.; DEAN, J.; KOHANE, I. Machine learning in medicine. **New England Journal of Medicine,** v. 380, p. 1347–1358, 2019.

RESTINI, Felipe Cicci et al. AI tool for predicting MGMT methylation in glioblastoma for clinical decision support in resource limited settings. **Scientific Reports,** v. 14, n. 1, 14 nov. 2024.

SANTOS, Carlos Eduardo Oliveira dos et al. White light imaging versus artificial intelligence-assisted white light imaging for

colorectal neoplasia detection: a randomised trial. **Revista de Gastroenterología del Perú,** v. 45, n. 4, p. 359–366, 30 dez. 2025.

SEGURA, Facundo; SEGURA, Pablo; SEGURA, Florencio. Determinación del grado de agresividad tumoral mediante técnicas de inteligencia artificial aplicadas a imágenes de resonancia magnética. **Revista de la Asociación Argentina de Ortopedia y Traumatología,** v. 88, n. 6, p. 653–661, 20 dez. 2023.

SILVA, E. et al. O Uso de Algoritmos de Inteligência Artificial no Diagnóstico Precoce do Câncer: Avanços Tecnológicos, Desafios Éticos e Perspectivas para a Prática Clínica. **Periódicos Brasil. Pesquisa Científica,** v. 5, n. 6, p. 257–273, 7 ago. 2026.

SIMONASSI, G. S. et al. O impacto da inteligência artificial no diagnóstico médico: avanços, desafios e oportunidades. **Revista Ibero-Americana de Humanidades, Ciências e Educação,** v. 10, n. 10, p. 2233–2242, 14 out. 2024.

SINGH, Hardeep et al. The global burden of diagnostic errors in primary care. **BMJ Quality & Safety,** v. 26, n. 6, p. 484–494, 16 ago. 2016. Disponível em: https://qualitysafety.bmj.com/content/26/6/484. Acesso em: 03 out. 2026.

SOARES, R. A. et al. O uso da inteligência artificial na medicina: aplicações e benefícios. **Research, Society and Development,** v. 12, n. 4, p. e5012440856–e5012440856, 26 mar. 2023.

TEIXEIRA. A evolução tecnológica da inteligência artificial e seus impactos sociais. **Cadernos de Tecnologia,** v. 22, n. 1, p. 15–30, 2025.

TOPOL, E. **Deep medicine:** how artificial intelligence can make healthcare human again. New York: Basic Books, 2019.

TOPOL, E. J. As artificial intelligence goes multimodal, medical applications multiply. **Science,** v. 381, n. 6663, 15 set. 2023.

URETA, Nicolás et al. Capacidad diagnóstica de algoritmo basado en la inteligencia artificial en la detección de patologías malignas mediante análisis de Mamografías. **Journal of Health and Medical Sciences,** v. 10, n. 4, p. 17–23, 2024.

WATARI et al. Cognitive biases in diagnostic errors: a study among physicians in clinical practice. **Internal Medicine,** v. 61, n. 5, p. 655–661, 2022.

WHO. World Health Organization. **Ethics and governance of artificial intelligence for health:** WHO guidance. Geneva: World Health Organization, 2021.

ZALDIVAR, Roger et al. Evaluating the Predictive Accuracy of an AI-Based Tool for Postoperative Vault Estimation in Phakic Intraocular Lens Implantation. **Clinical Ophthalmology,** v. 19, p. 1945–1956, 1 jun. 2025.

ZHANG et al. Desafios e erros diagnósticos na prática médica contemporânea. **Revista de Clínica e Pesquisa Médica,** v. 29, n. 4, p. 301–310, 2023.

---

[1] Discente do Curso Superior de Medicina da Afya Faculdades de Ciências Médicas de Vitória da Conquista, BA. E-mail: acesse o artigo original para visualizar o e-mail

[2] Discente do Curso Superior de Medicina da Afya Faculdades de Ciências Médicas de Vitória da Conquista, BA. E-mail: acesse o artigo original para visualizar o e-mail

[3] Discente do Curso Superior de Medicina da Afya Faculdades de Ciências Médicas de Vitória da Conquista, BA. E-mail: acesse o artigo original para visualizar o e-mail

[4] Discente do Curso Superior de Medicina da Afya Faculdades de Ciências Médicas de Vitória da Conquista, BA. E-mail: acesse o artigo original para visualizar o e-mail

[5] Docente do Curso Superior de Medicina da Afya Faculdades de Ciências Médicas de Vitória da Conquista, BA. E-mail: acesse o artigo original para visualizar o e-mail