

DESEMPENHO DA INTELIGÊNCIA ARTIFICIAL NO DIAGNÓSTICO DE RADIOGRAFIAS TORÁCICAS: uma revisão de literatura

PERFORMANCE OF ARTIFICIAL INTELLIGENCE IN THE DIAGNOSIS OF CHEST RADIOGRAPHS: a literature review

ANDRÉ GALANTE ALENCAR ARANHA (Docente na Faculdade de Ciências Médicas de Santos do curso de medicina, ORCID 0009-0007-5994-4726, e-mail: drandrearanha@icloud.com)

LUCAS GOMES NASCIMENTO (Discente na Faculdade de Ciências Médicas de Santos do curso de medicina, ORCID 0009-0008-5178-6417, e-mail: lucasgomesn2002@gmail.com)

LUIS FERNANDO ELIAS MIZOBUCHI (Discente na Faculdade de Ciências Médicas de Santos do curso de medicina, ORCID 0009-0006-6497-2206, e-mail: luisfmizobuchi@hotmail.com)

PEDRO SIMÕES VIGNA (Discente na Faculdade de Ciências Médicas de Santos do curso de medicina, ORCID 0009-0009-7670-0103, e-mail: Pedrosimoesvigna@gmail.com)

RESUMO

Introdução: A inteligência artificial, desde o pioneiro sistema MYCIN na década de 1970, tem avançado significativamente na medicina e, especialmente na radiologia, seu uso em radiografias torácicas evidencia o potencial de apoiar médicos, otimizar fluxos de trabalho e aprimorar o cuidado ao paciente. **Objetivo:** Executar uma revisão qualitativa da literatura a respeito do papel da inteligência artificial no diagnóstico de diversas patologias e padrões radiológicos, utilizando exclusivamente radiografias torácicas. **Métodos:** Trata-se de uma Revisão de Literatura por meio das bases de dados *PubMed* e *PubMed Central* por meio dos Descritores em Ciências da Saúde (DeCS/MeSH): “Inteligência artificial” (*Artificial intelligence*), Radiografia torácica (*Radiography Thoracic*), Diagnóstico (*Diagnosis*), Revisão de literatura (*Review*); Esses termos foram associados aos operadores booleanos *AND* e *OR*, por meio de todas as combinações possíveis das palavras-chave mencionadas. Foram incluídos 10 artigos, selecionados de 651 registros iniciais, com base em critérios de idioma em inglês, filtro temporal de 2018 a 2025 e acesso gratuito. **Resultados:** Os 10 estudos indicam que a inteligência artificial apresenta desempenho comparável ou superior ao de médicos na interpretação de radiografias torácicas, destacando-se em triagem rápida e suporte a profissionais menos experientes, mas enfrenta limitações como detecção de achados sutis, falsos positivos e dependência de bases de dados robustas, devendo ser usada como complemento ao julgamento clínico. **Conclusão:** A revisão indica que a inteligência artificial melhora a precisão e eficiência do diagnóstico por imagem, mas enfrenta desafios como integração clínica, necessidade de bases robustas e custos, exigindo pesquisas prospectivas para avaliar seu impacto real.

Palavras-chave: Inteligência artificial; Radiografia torácica; Diagnóstico; Revisão de literatura.

ABSTRACT

Introduction: Artificial intelligence (AI), since the pioneering MYCIN system in the 1970s, has advanced significantly in medicine, and its application in radiology, particularly in chest radiographs, demonstrates its potential to support physicians, optimize workflow, and enhance patient care. **Objective:** To conduct a qualitative literature review on the role of AI in the diagnosis of various pathologies and radiological patterns, focusing exclusively on chest radiographs. **Methods:** This literature review was conducted using the PubMed and PubMed Central databases with the following Health Science Descriptors (DeCS/MeSH): “Artificial Intelligence,” “Chest Radiography,” “Diagnosis,” and “Review,” combined with the Boolean operators AND and OR in all possible configurations. Ten articles were included, selected from an initial 651 records based on inclusion criteria of English language, publication between 2018 and 2025, and open access. **Results:** The ten studies indicate that AI demonstrates comparable or superior performance to physicians in interpreting chest radiographs, showing particular strength in rapid triage and support for less experienced professionals, but facing limitations such as detection of subtle findings, false positives, and dependence on robust datasets, highlighting its role as a complement to clinical judgment. **Conclusion:** The review suggests that AI enhances diagnostic accuracy and efficiency in medical imaging, yet faces challenges including clinical integration, the need for high-quality datasets, and costs, emphasizing the necessity of prospective studies to evaluate its real-world impact.

Keywords: Artificial intelligence; Radiography, Thoracic; Diagnosis; Review.

1 INTRODUÇÃO

A Inteligência Artificial (IA), caracterizada pela capacidade de máquinas ou programas realizarem tarefas que normalmente exigem raciocínio humano, tornou-se proeminente na medicina a partir de 2021, com ferramentas como *ChatGPT*, *Mid Journey* e sistemas baseados em redes convolucionais [1]. Uma das primeiras aplicações significativas foi o sistema *MYCIN*, desenvolvido por Edward H. Shortliffe em 1972 para auxiliar no diagnóstico de infecções graves e recomendação de antibioticoterapia, publicado em 1977 [2][3]. Entre 2018 e 2021, o número de estudos em IA cresceu 270,20% na PubMed, impulsionado por avanços em processamento de linguagem e hardware [4].

Desde 2018, algoritmos de aprendizado profundo (*deep learning*) e aprendizado de máquina (*machine learning*) têm transformado a prática médica, especialmente na radiologia, com uso de Convolutional Neural Networks (CNNs) e Graphics Processing Units (GPUs), melhorando a precisão no diagnóstico por imagem e a extração automatizada de padrões complexos [5], e aplicações incluem doenças musculoesqueléticas, cânceres e doenças neurodegenerativas, com eficácia na detecção precoce e personalização do tratamento, embora existam desafios relacionados a conjuntos de dados e ética [6], de modo que, na radiologia, a IA reduz a carga de trabalho de radiologistas, aumentando a eficiência [7][8].

Radiografias torácicas (RXT) são exames amplamente utilizados para identificar pneumonia, insuficiência cardíaca e neoplasias pulmonares; estudos recentes propõem *frameworks* multietapas para classificação de doenças respiratórias e avaliação de gravidade, com foco em *Coronavirus Disease 2019 (COVID-19)* [9][10]. A IA se mostrou eficaz na triagem e interpretação dessas imagens, priorizando casos críticos para análise humana [11]. Ferramentas como *ChatGPT*, *Midjourney* e redes neurais têm otimizado fluxos clínicos, auxiliando especialistas, especialmente em alta demanda ou entre profissionais menos experientes [12].

Entretanto, desafios incluem necessidade de bases de dados robustas, integração aos fluxos clínicos e predominância de estudos retrospectivos, evidenciando a necessidade de abordagens mais rigorosas [13]. Diante disso, justifica-se uma revisão sistemática para sintetizar evidências, identificar lacunas e subsidiar a adoção segura e eficiente da IA na interpretação de Radiografias Torácicas (RXT). A hipótese é que a IA apresenta desempenho comparável ou superior ao de profissionais sem auxílio, podendo otimizar precisão, eficiência e consistência. O objetivo geral é revisar qualitativamente o papel da IA em RXT, analisando efetividade autônoma e colaborativa, comparações com especialistas e interação com profissionais, respondendo à pergunta norteadora: “Em pacientes submetidos a RXT, a interpretação por IA baseada em deep learning e machine learning, autônoma ou colaborativa, apresenta desempenho superior ou comparável ao de profissionais sem auxílio da tecnologia?”.

2 MÉTODO

Este estudo trata-se de uma Revisão de Literatura, caracterizada pela busca, seleção, análise e interpretação de produções científicas previamente publicadas, com o objetivo de oferecer compreensão ampla e crítica do conhecimento existente, identificar lacunas e posicionar a proposta de estudo de maneira fundamentada [14]. A revisão focou na avaliação do diagnóstico de

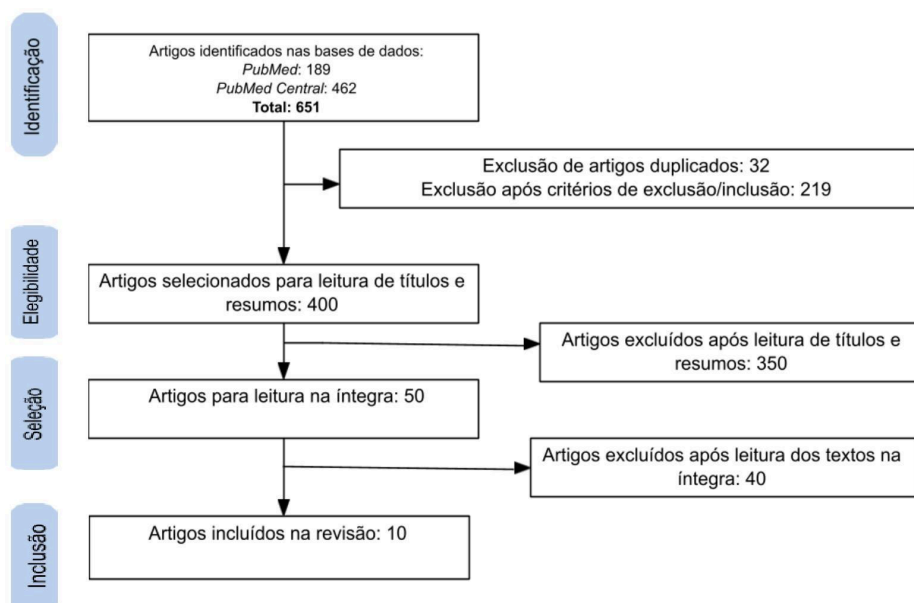
radiografias torácicas com o uso de inteligência artificial, selecionando artigos entre maio e junho de 2025 nas bases *PubMed* e *PubMed Central*.

A busca seguiu a estrutura PICO: P – pacientes submetidos a radiografias torácicas; I – interpretação por sistemas de IA baseados em *deep learning* e *machine learning*, de forma autônoma ou colaborativa; C – profissionais de saúde sem auxílio tecnológico; O – desempenho diagnóstico (sensibilidade, especificidade, valores preditivos, *Area Under the Curve (AUC)*, índice *Kappa* e eficiência clínica). A pergunta norteadora foi: “Em pacientes submetidos a radiografias torácicas, a interpretação por sistemas de IA apresenta desempenho superior ou comparável ao de profissionais sem tecnologia?”.

Foram utilizados descritores DeCS/MeSH em português e inglês: “Inteligência artificial” (*Artificial Intelligence*), Radiografia torácica (*Radiography Thoracic*), Diagnóstico (*Diagnosis*), Revisão de literatura (*Review*), combinados com operadores booleanos *AND* e *OR*. A estratégia incluiu: (“*Artificial Intelligence*” *OR* “*Deep Learning*” *OR* “*Machine Learning*”) *AND* (“*Chest X-Ray*” *OR* “*Chest Radiograph*” *OR* “*CXR*”) *AND* (“*Diagnostic Accuracy*” *OR* “*Diagnosis*” *OR* “*Diagnostic Performance*”) *AND* (“*health professionals*” *OR* “*Radiologists*” *OR* “*Physicians*” *OR* “*Health Professionals*” *OR* “*Thoracic Surgeons*”) *AND* (“*Comparison*” *OR* “*Comparative Study*” *OR* “*Assisted Interpretation*” *OR* “*Augmented Interpretation*” *OR* “*With and Without AI*”).

Os critérios de inclusão foram artigos em inglês, publicados entre 2018 e 2025, com acesso gratuito; os de exclusão foram estudos sem padrão-ouro comparativo entre intervenção e controle ou que não apresentassem ao menos uma das métricas: sensibilidade, especificidade ou *AUC*. Após a busca seguindo a metodologia *Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA)*, foram selecionados 10 estudos originais sobre a eficácia da IA no diagnóstico de patologias e padrões radiológicos utilizando exclusivamente RXT [14].

Tabela 1 - Identificação de novos estudos via base de dados e registros documentados no *PRISMA*.



Fonte: Autoria própria.

A busca inicial identificou 651 estudos (189 na *PubMed* e 462 no *PubMed Central – PMC*). Após exclusões por critérios e duplicatas via *AI-Powered Systematic Review Management Platform – RSRMP (Rayyan)*, restaram 400 artigos para triagem, dos quais 350 foram eliminados. Assim, 50 foram lidos na íntegra e 10 incluídos na análise final, destacando comparações e colaborações entre IA e profissionais humanos (radiologistas, cirurgiões torácicos e residentes).

3 RESULTADOS

Dos dez estudos incluídos na revisão, oito apresentaram delineamento retrospectivo, analisando dados previamente coletados, enquanto dois foram prospectivos, monitorando participantes a partir de um ponto específico com coleta planejada, permitindo avaliação mais próxima da prática clínica. A seção de resultados será apresentada seguindo essa divisão.

3.1 Estudos retrospectivos

Ahn e colaboradores [4] investigaram o impacto do sistema de inteligência artificial *Lunit INSIGHT CXR v3.1.2.0* na interpretação de 497 radiografias de tórax por seis leitores humanos (dois residentes, dois fellows e dois radiologistas assistentes), utilizando delineamento multicêntrico, cruzado e randomizado, com avaliação cega em duas sessões (com e sem IA) separadas por quatro semanas para reduzir viés de memória, e padrão-ouro definido por consenso de dois radiologistas torácicos experientes.

Foram identificados 538 achados-alvo: pneumonia (36,2%), derrame pleural (27,7%), nódulo pulmonar (21,2%) e pneumotórax (14,9%), além de 193 achados adicionais não contemplados pelo estudo. O uso da IA melhorou o desempenho dos leitores, com aumento da sensibilidade média de 56,7% para 62,9%, especificidade de 88,5% para 88,1% e *AUC* média de 0,724 para 0,752 na detecção de nódulos pulmonares, sendo que a IA isolada alcançou sensibilidade de 81,6%, especificidade de 73,1% e *AUC* de 0,858.

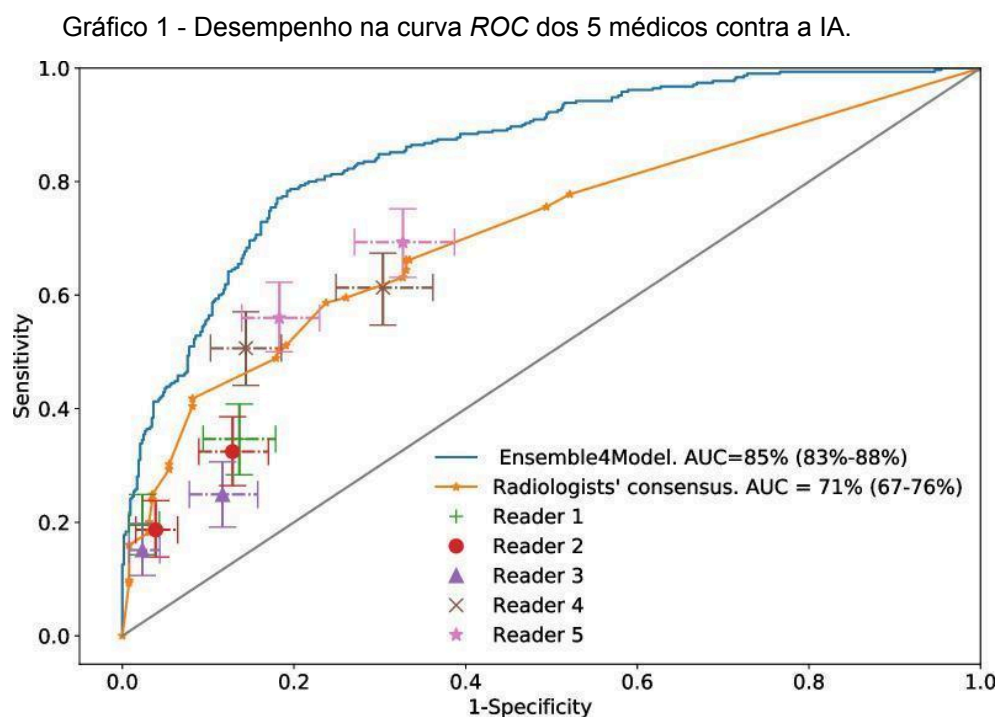
Na pneumonia, a sensibilidade aumentou de 67,3% para 71,9%, especificidade de 86,2% para 83,2% e *AUC* de 0,768 para 0,776, com a IA isolada alcançando sensibilidade de 88,7%, especificidade de 72,8% e *AUC* de 0,880. No derrame pleural, a sensibilidade subiu de 88,9% para 89,5%, especificidade de 92,8% para 93,7% e *AUC* de 0,909 para 0,916, enquanto a IA isolada obteve sensibilidade de 87,2%, especificidade de 96,0% e *AUC* de 0,983.

No pneumotórax, a sensibilidade aumentou de 79,2% para 96,5%, a especificidade manteve-se elevada (98,8% - 98,6%) e a *AUC* passou de 0,885 para 0,969, com a IA isolada atingindo 98,8% de sensibilidade, 98,6% de especificidade e *AUC* de 0,999. O tempo médio de leitura por imagem reduziu de 40,8 para 36,9 segundos [4].

De forma complementar, Albiol e colaboradores [15] avaliaram retrospectivamente o algoritmo *Ensemble4Covid* em 6.962 RXT (4.566 positivas e 2.396 negativas para *COVID-19*) de 23 hospitais na Espanha, utilizando um subconjunto de 480 imagens (225 positivas e 255 negativas) para análise comparativa entre o algoritmo e cinco radiologistas com experiência de 1 a 14 anos.

O *Ensemble4Covid* foi desenvolvido por aprendizado por transferência com quatro redes convolucionais pré-treinadas (*ResNet50*, *DenseNet121*, *InceptionV3* e *InceptionResNetV2*), treinado com 9.188 imagens positivas e 2.551 negativas do banco *BIMCV-COVID19* e validado em conjunto

independente de 310 positivas e 527 negativas, garantindo que as 480 imagens da fase prática fossem inéditas. As métricas avaliadas foram *AUC* e índice *Kappa*, permitindo comparar a concordância entre radiologistas e entre a IA e o padrão-ouro, com a curva *Receiver Operating Characteristic (ROC)* apresentada para ilustrar o desempenho comparativo [15].



Fonte: Albiol e colaboradores [15].

O gráfico mostra as curvas ROC do *Ensemble4Covid* em azul, com intervalo de confiança de 95%, comparadas ao consenso de cinco radiologistas em laranja, numerados por experiência, apresentando sensibilidade (eixo Y) e especificidade (eixo X) para diferentes limiares de decisão. O *Ensemble4Covid* obteve desempenho superior, com sensibilidade de 78,2%, especificidade de 81,3%, *AUC* de 0,856 e *Kappa* de 0,60, enquanto os radiologistas tiveram média de sensibilidade de 59,0%, especificidade de 76,5%, *AUC* 0,713 e *Kappa* 0,36 [15].

Chen e colaboradores [16] desenvolveram um modelo de *deep learning* para detecção de tuberculose pulmonar (TB) usando 4.154 radiografias no treinamento e validação externa com 250 imagens, comparando o desempenho com cinco pneumologistas. A IA alcançou *AUC* de 0,951 na validação interna e 0,975 na externa, mantendo desempenho estável mesmo com casos de outras patologias (*AUC* 0,828), enquanto os especialistas apresentaram *AUCs* variando de 0,843 a 0,995. A combinação IA com especialistas, mostrou leve melhoria nos resultados (*AUC* 0,862–0,885).

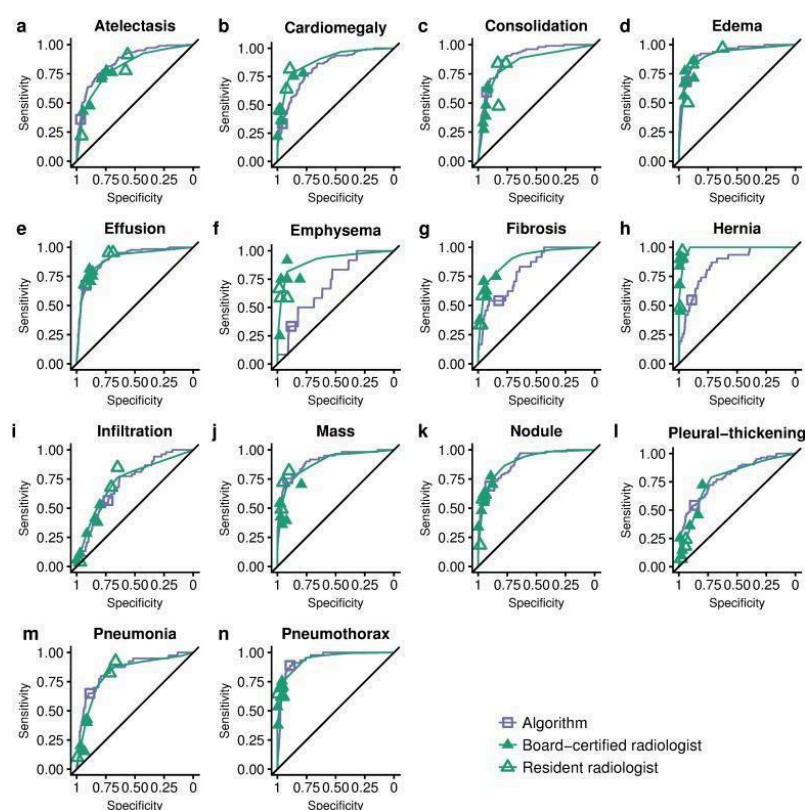
Homayounieh e outros [17] avaliaram o *AI Rad Companion Chest X-ray* em 100 radiografias, comparando leituras com e sem IA. O suporte da IA aumentou a sensibilidade média de 45% para 55%, especificidade de 93% para 95% e acurácia de 69% para 75%, com ganhos mais evidentes entre radiologistas juniores.

López e colaboradores [18] analisaram o *Arterys MICA* em 784 radiografias, comparando com um residente de radiologia. A IA apresentou desempenho variável: excelente em fraturas ósseas

e pneumotórax (sensibilidade e Valor Preditivo Negativo (VPN) 100%), aceitável em opacidades (sensibilidade 76%, AUC 0,86) e derrame pleural (sensibilidade 60%, AUC 0,79), e limitado em nódulos pulmonares (sensibilidade 33%, AUC 0,73), enquanto o residente teve melhor desempenho em nódulos (sensibilidade 75%, AUC 0,90).

Rajpurkar e outros autores [19] avaliaram a *CheXNeXt* na detecção de 14 patologias torácicas em 420 radiografias. A IA apresentou desempenho equivalente aos radiologistas em 11 patologias, inferior em cardiomegalia, enfisema e hérnia hiatal, e superior em atelectasias (AUC 0,862 vs. 0,808). No total, a IA se destacou em 8 das 14 patologias e completou a análise em 1 min 30 s, enquanto os humanos gastaram em média 240 min [19].

Gráfico 2 - Curva ROC Das 14 Patologias Abordadas.



Fonte: Rajpurkar e colaboradores [19].

O gráfico do estudo de Rajpurkar e outros [19] apresenta as curvas ROC das 14 patologias avaliadas, com a IA *CheXNeXt* indicada pela linha roxa, os seis radiologistas especialistas por triângulos verdes sólidos e os três residentes por triângulos com centro branco. A IA forneceu probabilidades percentuais, enquanto os radiologistas atribuíram escores de 1 a 3.

Teng e colaboradores [20] avaliaram quatro algoritmos de IA (heat-map, probabilidade anômala, probabilidade de nódulo e probabilidade de massa) para detecção de nódulos e massas pulmonares em 100 radiografias de adultos, comparando-os com seis radiologistas de diferentes níveis de experiência. O estudo retrospectivo e cego utilizou tomografia computadorizada como padrão-ouro e análises estatísticas via ROC, índice de Youden, sensibilidade, especificidade, Valor Preditivo Positivo (VPP) e acurácia.

Os resultados mostraram que o algoritmo de probabilidade de massa teve o melhor desempenho geral (AUC 0,916; sensibilidade 76,6%; especificidade 90,57%; acurácia 83%), superando os radiologistas (AUC 0,778; sensibilidade 77,3%; especificidade 78,3%; acurácia 78%). O algoritmo de probabilidade de nódulo apresentou alta sensibilidade (85,11%) mas baixa especificidade (67,92%), enquanto o algoritmo de probabilidade anômala teve desempenho semelhante aos radiologistas (AUC 0,810; acurácia 78%). O *heat-map* obteve alta especificidade (98,11%) e VPP (95%) porém baixa sensibilidade (38,3%) e AUC inferior (0,682), limitando sua utilidade isolada [20].

Tabela 2 – Tabela De Desempenho Entre IA e Radiologistas.

Algoritmos	Algoritmo (IC 95%)	Radiologistas (IC 95%)	Diferença entre áreas (Algoritmo - Radiologistas 95% IC)	Vantagem
Algoritmo de mapa de calor	0,682 (0,643–0,719)	0,778 (0,743–0,811)	-0,096 (0,0561–0,136)	Radiologistas
Algoritmo de probabilidade anormal	0,810 (0,776–0,841)	0,778 (0,743–0,811)	0,0321 (-0,0136–0,0777)	Nenhuma diferença
Algoritmo de probabilidade de massa	0,916 (0,891–0,937)	0,778 (0,743–0,811)	0,138 (0,100–0,176)	Algoritmo de massa
Algoritmo de probabilidade de nódulo	0,813 (0,780–0,844)	0,778 (0,743–0,811)	0,0353 (-0,00911–0,0797)	Nenhuma diferença

Fonte: Teng e colaboradores [20].

Wu e outros [21] compararam o desempenho do algoritmo de IA qXR com cinco residentes de radiologia do terceiro ano na interpretação de 1.998 radiografias torácicas frontais AP em pronto-socorro. O estudo retrospectivo e cego contou com padrão-ouro definido por três radiologistas especializados, considerando dados clínicos e exames prévios. O algoritmo, treinado com imagens *Medical Information Mart for Intensive Care IV (MIMIC-IV)* e *National Institutes of Health (NIH)*, foi avaliado quanto à sensibilidade, especificidade e VPP. Os resultados mostraram que a IA apresentou sensibilidade de 71,6% (IC 95%: 70,4–72,9), especificidade de 98,0% (IC 95%: 97,9–98,1) e VPP de 73,0% (IC 95%: 71,8–74,2), enquanto os residentes atingiram sensibilidade de 72,0% (IC 95%: 70,9–73,2), especificidade de 97,3% (IC 95%: 97,1–97,4) e VPP de 68,2% (IC 95%: 67,0–69,4). Não houve diferença significativa na sensibilidade ($p = 0,66$), mas a IA apresentou especificidade ($p < 0,001$) e VPP ($p = 0,001$) superiores [21].

3.2 Estudos prospectivos

O estudo de Hwang e colaboradores [22] avaliou se a IA (*Lunit INSIGHT CXR v2.0.2.0*) poderia aumentar a sensibilidade diagnóstica e a eficácia clínica em radiografias torácicas de pacientes com sintomas respiratórios agudos em pronto-socorro. O ensaio clínico prospectivo incluiu 3.576 participantes distribuídos aleatoriamente entre grupo com suporte de IA ($n = 1.761$) e grupo controle ($n = 1.815$), com 20 residentes interpretando as imagens. O padrão ouro foi definido por radiologista torácico experiente. Não houve diferença significativa entre os grupos em sensibilidade (67,2% vs. 66,0%), taxa de falsos positivos (19,3% vs. 18,5%) ou medidas de eficiência clínica, incluindo tempo de emissão de laudos, frequência de Tomografia Computadorizada (TC), permanência no pronto-socorro e taxa de retorno em 30 dias [22].

Novak e outros [23] investigaram o efeito da IA (*GEHC Critical Care Suite*) na detecção de pneumotórax por 18 médicos de seis especialidades em quatro hospitais do Reino Unido, avaliando

395 radiografias. O algoritmo isolado apresentou sensibilidade de 73%, especificidade de 96% e *AUROC* de 0,94. Com suporte da IA, a sensibilidade global aumentou de 66,8% para 78,1% ($p = 0,002$), reduzindo falsos negativos, enquanto a especificidade teve aumento discreto não significativo (93,9% para 95,8%; $p = 0,247$). A maior melhora ocorreu entre leitores juniores (56,0% para 77,7%), alcançando desempenho semelhante a intermediários e seniores, sem alteração significativa no tempo médio de leitura, mas com aumento subjetivo da confiança diagnóstica [23].

4 DISCUSSÃO

A análise dos dez estudos evidencia que a inteligência artificial (IA) evoluiu significativamente desde sua primeira aplicação na medicina em 1972, apresentando desempenho comparável ou superior ao de médicos em radiografias torácicas, consolidando-se como ferramenta promissora, embora limitada em alguns cenários. Rajpurkar e outros [19] mostraram desempenho satisfatório da IA em 11 das 14 patologias avaliadas, com superioridade em atelectasias e análise extremamente rápida, enquanto López e colaboradores [18] confirmaram resultados semelhantes, ressaltando que a IA complementa, mas não substitui, o julgamento clínico. Albiol e outros [15] demonstraram desempenho superior da IA na triagem de *COVID-19*, com modelo *Ensemble4Covid* apresentando sensibilidade de 78,2%, especificidade de 81,3% e *AUC* de 0,856, usando quase 7 mil radiografias e validação externa robusta. Chen e colaboradores [16] destacaram modelo acessível para triagem de tuberculose com *AUC* de 0,975, útil em regiões com escassez de especialistas. Wu e outros [21], Homayounieh e outros [17] e Teng e outros [20] evidenciaram que a IA iguala resultados de profissionais experientes e auxilia no treinamento de novos radiologistas. Novak e colaboradores [23] mostraram que médicos assistidos por IA, inclusive de outras especialidades, aumentaram a precisão diagnóstica. Ahn e colaboradores [4] e Hwang e colaboradores [22] indicaram que a efetividade da IA depende do contexto clínico e da experiência do profissional, sendo que Hwang e outros [22] não observaram benefícios significativos em residentes de pronto-socorro, devido a alta taxa de falsos positivos (61,7%), uso paralelo da IA sem integração ao fluxo clínico e relevância clínica limitada dos achados.

Em síntese, a IA apresenta benefícios claros em triagem rápida, priorização de casos urgentes e suporte a profissionais menos experientes, mas seu impacto depende da integração eficiente ao ambiente clínico, calibração adequada e experiência dos usuários, reforçando que a combinação de IA e conhecimento humano constitui a abordagem mais eficaz para otimizar diagnósticos radiológicos e o cuidado ao paciente.

5 CONSIDERAÇÕES FINAIS

O estudo objetivou avaliar o desempenho da inteligência artificial (IA) na interpretação de radiografias torácicas, de forma autônoma ou complementar ao trabalho médico, com base em dez estudos publicados entre 2018 e 2025 que utilizaram algoritmos de deep learning e machine learning, comparando-os a radiologistas, residentes e outros profissionais de saúde. Os resultados indicam que a IA apresenta desempenho equivalente ou superior ao de humanos, especialmente em tarefas repetitivas, triagem e profissionais menos experientes, aumentando a precisão diagnóstica e reduzindo a variabilidade. Limitações incluem dificuldade de aplicação clínica, altas taxas de falsos

positivos, dependência de grandes bases de dados e ausência de estudos prospectivos e de sistemas de acesso aberto. Conclui-se que a IA é promissora para aprimorar diagnósticos, padronizar práticas e otimizar a eficiência dos serviços de saúde, sendo necessários esforços técnicos, financeiros e metodológicos para sua ampla e segura implementação.

REFERÊNCIAS

- [1] Chan KY, Yuen TH, Co M. Using ChatGPT for medical education: the technical perspective. *BMC Med. Educ.* 2025; 25(201), 1-29. DOI: 10.1186/s12909-025-06785-9.m
- [2] Shortliffe EH, Buchanan BG. Mycin: a knowledge-based computer program applied to infectious diseases. *Math Biosci.* 1975; 23(3-4): 351–79.
- [3] Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J.* 2019; 6(2): 94-98. DOI: 10.7861/futurehosp.6-2-94.
- [4] Ahn JS, Park SH, Kim H, Park J, Bae WJ, Lee SS, *et al.* Association of artificial intelligence-aided chest radiograph interpretation with reader performance and efficiency. *JAMA Netw Open.* 2022; 5(8): 1-15. DOI: 10.1001/jamanetworkopen.2022.29289.
- [5] Sadr H, Nazari M, Khodaverdian Z, Farzan R, Yousefzadeh-Chabok S, Ashoobi MT, Hemmati H, Hendi A, Ashraf A, Pedram MM, Hasannejad-Bibalan M, Yamaghani MR. Unveiling the potential of artificial intelligence in revolutionizing disease diagnosis and prediction: a comprehensive review of machine learning and deep learning approaches. *Eur. J. Med. Res.* 2025; 30:(418), 1-36. DOI: 10.1186/s40001-025-02680-7.
- [6] Obuchowicz R, Lasek J, Wodziński M, Piórkowski A, Strzelecki M, Nurzynska K. Artificial Intelligence-Empowered Radiology - Current Status and Critical Review. *Diagnostics (Basel).* 2025;15(3): 1-28. DOI: 10.3390/diagnostics15030282.
- [7] Achour N, Zapata T, Saleh Y, Pierscionek B, Azzopardi-Muscat N, Novillo-Ortiz D, Morgan C, Chaouali M. The role of AI in mitigating the impact of radiologist shortages: a systematised review. *Health Technol.* 2025; 15(3): 489–501. DOI: 10.1007/s12553-025-00970-y.
- [8] Byberg E, Crimi M. Preparing hospitals and health organizations for AI: practical guidelines for the required infrastructure. *Front Digit Health.* 2025; 7:(1), 1-6. DOI: 10.3389/fdgh.2025.1605006.
- [9] Anderson PG, Tarder-Stoll H, Alpaslan M, Keathley N, Levin DL, Venkatesh S, Bartel E, Sicular S, Howell S, Lindsey RV, Jones RM. Deep learning improves physician accuracy in the comprehensive detection of abnormalities on chest X-rays. *Sci. Rep.* 2024;14:(25151), 1-13. DOI: 10.1038/s41598-024-76608-2.
- [10] Sahoo P, Sharma SK, Saha S, Jain D, Mondal S. A multistage framework for respiratory disease detection and assessing severity in chest X-ray images. *Sci Rep.* 2024; 14 (12380), 1-10. DOI: 10.1038/s41598-024-60861-6.
- [11] Annarumma M, Withey SJ, Bakewell RJ, Pesce E, Goh V, Montana G. Automated triaging of adult chest radiographs with deep artificial neural networks. *Radiology.* 2019; 291(1): 196-202. DOI: 10.1148/radiol.2018180921.
- [12] Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare (Basel).* 2023; 11(6): 887, 1-20. DOI: 10.3390/healthcare11060887.

- [13] Antonissen N, Tryfonos O, Houben IB, Jacobs C, de Rooij M, van Leeuwen KG. Artificial intelligence in radiology: 173 commercially available products and their scientific evidence. *Imaging Inform Artificial Int.* 2025; 25:1, 1-11. DOI: 10.1007/s00330-025-11830-8.
- [14] Breviário AG. Fusões e aquisições: uma revisão da literatura. *Rev. Digital Gest. Neg.* 2024;1(6):1-16.
- [15] Albiol A, Montell JA, Monserrat C, Martí-Bonmatí L, Laguarda J, Puertas A *et al.* A comparison of Covid-19 early detection between convolutional neural networks and radiologists. *Comput Biol Med.* 2022; 150:(1), 1-12. DOI: 10.1186/s13244-022-01250-3.
- [16] Chen CF, Hsu CH, Jiang YC, Lin WR, Hong WC, Chen IY, Lin MH, Chu KA, Lee CH, Lee DL, Chen PF. A deep learning-based algorithm for pulmonary tuberculosis detection in chest radiography. *Sci Rep.* 2024, 28;14(1): 1-10. DOI: 10.1038/s41598-024-65703-z.
- [17] Homayounieh F, Digumarthy S, Ebrahimian S *et al.* An Artificial Intelligence-Based Chest X-ray Model on Human Nodule Detection Accuracy From a Multicenter Study. *JAMA Netw Open.* 2021; 4(12): 1-10. DOI: 10.1001/jamanetworkopen.2021.41096.
- [18] López Alcolea J, Fernández Alfonso A, Cano Alonso R, et al. Diagnostic Performance of Artificial Intelligence in Chest Radiographs Referred from the Emergency Department. *Diagnostics (Basel).* 2024; 14(22): 1-14. DOI:10.3390/diagnostics14222592.
- [19] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T *et al.* Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med.* 2018 Nov; 15(11): 1-17. DOI: 10.1371/journal.pmed.1002686.
- [20] Teng PH, Chou YH, Liu YJ, Chen HC, Hsieh HH, Hsieh KH. Performance and educational training of radiographers in lung nodule or mass detection. *PLoS One.* 2021; 16(6): 1-5. DOI:10.1097/MD.00000000000026270.
- [21] Wu JT, Wong KC, Gur Y, Ansari N, Karargyris A, Sharma A *et al.* Comparison of Chest Radiograph Interpretations by Artificial Intelligence Algorithm vs Radiology Residents. *JAMA Netw Open.* 2020; 3(10): 1-10. DOI: 10.1001/jamanetworkopen.2020.22779.
- [22] Hwang EJ, Park S, Jin KN, Im DJ, Goo JM, Choi BW *et al.* Conventional Versus Artificial Intelligence-Assisted Interpretation of Chest Radiographs in Patients With Acute Respiratory Symptoms in Emergency Department: A Pragmatic Randomized Clinical Trial. *J Thorac Imaging.* 2023; 38(2): 79–87. DOI: 10.3348/kjr.2022.0651.
- [23] Novak A, Johnstone H, Smith B, Hamilton C, McCoubrie P, Mirsadraee S *et al.* Evaluation of the impact of artificial intelligence-assisted image interpretation on the diagnostic performance of clinicians in identifying pneumothoraces on plain chest X-ray: a multi-case multi-reader study. *Br J Radiol.* 2024;97(1159): 1-8. DOI: 10.1136/emered-2023-213620.