

O impacto da inteligência artificial na clínica médica

The impact of Artificial Intelligence on medical practice

KALIL RIBEIRO NUNES
Discente de Medicina (UNIPAM)
kalilnunes@unipam.edu.br

HUMBERTO BRAGA SILVA FILHO
Bacharel em Enfermagem
humbertobraga@unipam.edubr

MARILENE RIVANY NUNES
Professora orientadora (UNIPAM)
maryrivany@unipam.edu.br

ELCIMAR DOS REIS CAIXETA
Professora coorientadora (UNIPAM)
elcimarcaixeta@unipam.edu.br

Resumo: As Inteligências Artificiais (IA) têm desempenhado papel significativo na transformação da medicina, sendo capazes de influenciar as etapas do processo saúde-doença. O estudo objetivou identificar, analisar e explorar as diversas maneiras pelas quais a IA influencia a prática da clínica médica. Foi feita uma revisão integrativa, usando os descritores em Ciências da Saúde: inteligência artificial, aprendizado de máquina, aprendizado profundo, prevenção, diagnóstico e tratamento. Foram incluídos artigos publicados entre janeiro/2019 e dezembro/2023. A busca encontrou 657 artigos; apenas 13 artigos foram incluídos no estudo. No geral, as IAs contribuíram para a interpretação de diversos exames de imagem e para predição, diagnóstico e prognóstico de diversas comorbidades. Uma porção significativa dos estudos apresentou falta de uma padronização metodológica, visto que precederam a publicação da ferramenta TRIPOD-AI. Espera-se que estudos futuros sejam padronizados visando facilitar a comparação dos estudos através de meta-análises.

Palavras-chave: inteligência artificial; medicina; clínica médica.

Abstract: Artificial Intelligence (AI) has played a significant role in transforming medicine, influencing multiple stages of the health-disease process. This study aimed to identify, analyze, and explore the various ways in which AI impacts clinical medical practice. An integrative review was conducted using the Health Sciences descriptors: artificial intelligence, machine learning, deep learning, prevention, diagnosis, and treatment. Articles published between January 2019 and December 2023 were considered. The search yielded 657 articles, of which only 13 met the inclusion criteria. Overall, AI contributed to the interpretation of various imaging tests and to the prediction, diagnosis, and prognosis of several comorbidities. A substantial portion of the studies lacked methodological standardization, as they preceded the publication of the TRIPOD-AI

guideline. Future studies are expected to follow standardized methodologies to enable comparison through meta-analyses.

Keywords: artificial intelligence; medicine; clinical practice.

1 INTRODUÇÃO

A inteligência artificial (IA) é um termo geral que implica o uso de um computador para modelar um comportamento inteligente e automatizado, com o mínimo de intervenção humana possível (Hamet; Tremblay, 2017). As IAs têm desempenhado um papel cada vez mais significativo na transformação de diversas áreas da sociedade e uma das ciências em que mais possui potencial é a medicina. Conforme a tecnologia avança e uma maior quantidade de dados médicos são disponibilizados, as IAs também tornam-se capazes de influenciar todas as etapas do processo saúde-doença (Ahuja, 2019; Li, 2020).

A IA é utilizada para melhor precisão diagnóstica e prognóstica ao utilizar-se de algoritmos e técnicas de aprendizado de máquina para relacionar grandes conjuntos de dados, como exames de imagem, prontuários ou exames laboratoriais, podendo encontrar padrões sutis, não perceptíveis através do viés da percepção humana. Isso resulta em diagnósticos mais precoces e precisos, permitindo intervenções mais eficazes e aumentando as chances de sucesso no tratamento (Briganti; Moine, 2020; Li, 2020).

Além disso, a IA pode ser útil no desenvolvimento de terapias personalizadas, com a aplicação de informações genéticas, históricos médicos e dados epidemiológicos, sendo possível identificar a existência de fatores de risco específicos e sugerir tratamentos únicos baseados nas necessidades individuais de cada paciente. Com o uso dessa abordagem personalizada, as chances de continuidade e sucesso no tratamento podem se tornar mais eficientes, minimizando efeitos colaterais e otimizando os resultados do tratamento (He *et al.*, 2019; Li; 2020; Farhud; Zokaei, 2021).

Um outro aspecto importante da IA na medicina é sua influência na aceleração da pesquisa médica. Ao analisar grandes bancos de dados científicos, a IA permite mais rapidamente identificar a existência de correlações e padrões que podem gerar descobertas importantes na medicina. A eficiência da IA pode ajudar a identificar terapias promissoras, acelerar o desenvolvimento de medicamentos e aprimorar a eficiência dos ensaios clínicos, reduzindo o tempo necessário para o desenvolvimento e aprovação de novos tratamentos (Li, 2020).

Em suma, a IA está se consolidando como uma ferramenta poderosa na medicina, permitindo avanços significativos na precisão diagnóstica, personalização do tratamento e aceleração da pesquisa médica. Deve-se garantir um uso ético e regulamentado, por profissionais de saúde qualificados para tal, de modo a aproveitar ao máximo suas vantagens e minimizar seus desafios. Por fim, o surgimento da IA na medicina é um caminho sem volta, e seu potencial para melhorar a saúde e salvar vidas é imenso.

A IA tem se tornado uma área de pesquisa e aplicação cada vez mais relevante na medicina. Compreender a sua influência nesse campo é de suma importância, devido aos seus potenciais benefícios para a sociedade através de maior precisão diagnóstica,

terapêuticas customizadas e utilização na pesquisa médica. Além disso, explorar os avanços, desafios e impactos da IA na prática médica auxilia na tomada de decisões e na busca de melhores cuidados de saúde ao trazer insights valiosos para profissionais e pesquisadores da área da saúde. Considerando-se o exposto, o presente estudo buscou identificar o impacto da IA na clínica médica, desde o diagnóstico até o tratamento, visando aperfeiçoar e otimizar a prática médica e conseqüentemente melhorar a qualidade de vida da sociedade. Ou seja, o estudo pretendeu identificar, analisar e explorar as diversas maneiras pelas quais a IA influencia a prática da clínica médica.

2 METODOLOGIA

Foi feita uma revisão integrativa, que consistiu na construção de uma análise ampla da literatura, contribuindo para discussões sobre métodos e resultados de pesquisas, com a finalidade de sintetizar resultados obtidos em pesquisas sobre um tema ou questão de forma sistemática, ordenada e abrangente, proporcionando a síntese do conhecimento e a incorporação da aplicabilidade de resultados de estudos significativos na prática (Souza; Silva; Carvalho, 2010; Camargo *et al.*, 2018).

Para a realização da revisão integrativa foram seguidas as seguintes etapas: identificação do tema e seleção da hipótese ou questão norteadora de pesquisa; amostragem ou busca na literatura com estabelecimento de critérios de inclusão e exclusão de artigos; extração de dados estabelecendo as informações a serem extraídas, categorização dos dados extraídos, análise e interpretação crítica dos dados ou resultados e, por fim, a apresentação da revisão integrativa com uma síntese do conhecimento elencado (Souza; Silva; Carvalho, 2010; Camargo *et al.*, 2018).

O estudo foi guiado pela pergunta norteadora “Quais são as formas de atuação da inteligência artificial e seu impacto na clínica médica?”. Para direcionar a busca dos artigos, foi adotado o uso dos Descritores em Ciências da Saúde: inteligência artificial, aprendizado de máquina, aprendizado profundo, prevenção, diagnóstico e tratamento em português e artificial intelligence, machine learning, deep learning, prevention, diagnosis, treatment em inglês, utilizando-se dos operadores “OR” e “AND” para guiar a busca pelos bancos de dados selecionados.

As bases de dados eletrônicas consultadas foram o Medical Literature Analysis and Retrieval System Online (MEDLINE), por meio da ferramenta de busca PubMed, a Biblioteca Virtual em Saúde (BVS), e o EBSCOHost, que inclui uma série de outras bases de dados em sua pesquisa. Essas bases de dados foram escolhidas pelo alcance científico na área da saúde. A busca foi realizada com vistas a incluir os artigos publicados no período de janeiro de 2019 a dezembro de 2023, em língua portuguesa e inglesa.

Foi realizado o levantamento dos artigos conforme critérios de inclusão estabelecidos, e os artigos que não apresentaram coerência com o objetivo do estudo e a questão norteadora ou duplicados foram excluídos.

Na sequência, foi realizada a análise e seleção de cada artigo, por dois pesquisadores, observando-se o título e resumos dos artigos. Nos casos em que estes não forem suficientes para determinar a elegibilidade da publicação, foi verificada a publicação na íntegra. Posteriormente, os pesquisadores se encontraram para comparar a amostragem selecionada, discutir as discrepâncias e chegar a um consenso acerca da

amostragem dos artigos. Foi adotada a descrição do processo de seleção dos artigos seguindo as etapas de identificação, triagem, elegibilidade, exclusão e inclusão dos artigos.

Então, os artigos selecionados foram analisados e os dados relevantes para direcionar a análise e discussão da temática proposta foram coletados. Para a coleta de dados, foi utilizado um instrumento desenvolvido pelos autores, o qual contempla as seguintes variáveis: ano de publicação, título, autores, delineamento do tipo de estudo e nível de evidência científica do artigo.

Para determinar o nível de evidência científica, foram adotadas as diretrizes metodológicas do sistema Grading of Recommendations Assessment, Development and Evaluation (GRADE), disposto no manual de graduação da qualidade da evidência e força de recomendação para tomada de decisão em saúde (Brasil, 2014). Assim, foi possível elencar o nível de evidência científica dos artigos e o grau de recomendação das IAs.

O GRADE é um sistema universal, transparente e sensível para classificar a qualidade das evidências e a força das recomendações de estudos. O nível de evidência representa a confiança na informação utilizada em apoio a uma determinada recomendação. No sistema GRADE, a avaliação da qualidade da evidência é classificada em quatro níveis: alto, moderado, baixo, muito baixo, conforme mostrado no Quadro 1 (Brasil, 2014).

Quadro 1: Níveis de evidências de acordo com o sistema GRADE

Nível	Definição	Implicações	Fonte de informação
Alto	Há forte confiança de que o verdadeiro efeito esteja próximo daquele estimado.	É improvável que trabalhos adicionais irão modificar a confiança na estimativa do efeito.	- Ensaios clínicos bem delineados, com amostra representativa. - Em alguns casos, estudos observacionais bem delineados, com achados consistentes*.
Moderado	Há confiança moderada no efeito estimado.	Trabalhos futuros poderão modificar a confiança na estimativa de efeito, podendo, inclusive, modificar a estimativa.	- Ensaios clínicos com limitações leves**. - Estudos observacionais bem delineados, com achados consistentes*.
Baixo	A confiança no efeito é limitada.	Trabalhos futuros provavelmente terão um impacto importante em nossa confiança na estimativa de efeito.	- Ensaios clínicos com limitações moderadas**. - Estudos observacionais comparativos: coorte e caso controle.
Muito baixo	A confiança na estimativa de efeito é muito limitada. Há importante grau de incerteza nos achados.	Qualquer estimativa de efeito é incerta.	- Ensaios clínicos com limitações graves**. - Estudos observacionais comparativos com limitações**. - Estudos observacionais não comparados***. - Opinião de especialistas.

Fonte: Brasil, 2014.

* Estudos de coorte sem limitações metodológicas, com achados consistentes apresentando tamanho de efeito grande e/ou gradiente dose resposta.

** Limitações: vieses no delineamento do estudo, inconsistência nos resultados, desfechos substitutos ou validade externa comprometida.

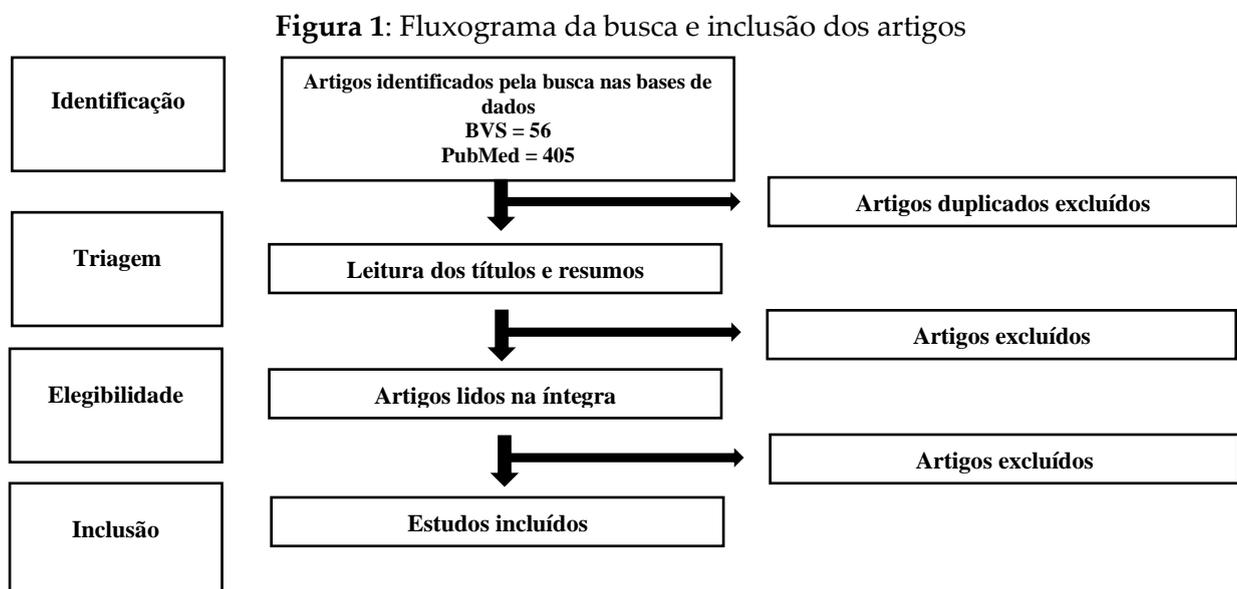
***Séries e relatos de caso.

Além disso, para definir o nível de confiança de cada determinado modelo de IA, foi priorizada a análise da área sob a curva (AUC) e a característica operacional do receptor (ROC), valores obtidos a partir da comparação dos resultados do programa com avaliações realizadas por médicos, em um gráfico com as taxas de verdadeiro-positivo versus falso-positivo, para determinar a diferença de acurácia entre a IA e o médico humano, assim permitindo expressar o nível de precisão das IAs (Topol, 2019).

A análise dos artigos será realizada pelo método de análise descritiva, possibilitando observar, analisar e correlacionar os dados com a maior precisão possível. Por fim, os artigos selecionados foram organizados e apresentados em uma tabela descritiva, objetivando sintetizar as informações, bem como propiciar uma melhor visualização dos dados essenciais que merecem destaque.

3 RESULTADOS

A profunda busca na literatura por artigos que abordam o tema em questão propiciou encontrar de forma abrangente 657 artigos únicos, publicados entre 2019 e 2023, no entanto, desse montante, apenas 13 artigos foram incluídos ao fim da avaliação pelos autores, sendo representado na Figura 1.



Fonte: Instrumento de coleta de dados elaborado pelos autores, 2024.

Desse modo, as publicações selecionadas para análise foram estruturadas e descritas no Quadro 2, citando ano de publicação, título e autores dos artigos analisados.

Quadro 2: Organização dos artigos por título, autores e ano de publicação

Nº	Ano	Título	Autores
1	2019	Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound: a systematic review and meta-analysis.	Zhao <i>et al.</i>
2	2020	Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy.	Fleuren <i>et al.</i>
3	2021	Application of machine learning in predicting hospital readmissions: a scoping review of the literature.	Huang <i>et al.</i>
4	2021	Machine learning versus usual care for diagnostic and prognostic prediction in the emergency department: a systematic review	Kareemi <i>et al.</i>
5	2022	Accuracy of machine learning classification models for the prediction of type 2 diabetes mellitus: a systematic survey and meta-analysis approach	Olusanya <i>et al.</i>
6	2022	Artificial intelligence in fracture detection: a systematic review and meta-analysis	Kuo <i>et al.</i>
7	2022	Convolutional neural network performance compared to radiologists in detecting intracranial hemorrhage from brain computed tomography: A systematic review and meta-analysis	Jørgensen <i>et al.</i>
8	2023	Are current machine learning applications comparable to radiologist classification of degenerate and herniated discs and Modic change? A systematic review and meta-analysis	Compte <i>et al.</i>
9	2023	Early prediction of ventilator-associated pneumonia with machine learning models: a systematic review and meta-analysis of prediction model performance	Frondelius <i>et al.</i>
10	2023	Machine and deep learning for tuberculosis detection on chest x-rays: systematic literature review.	Hansun <i>et al.</i>
11	2023	Machine learning for prediction of asthma exacerbations among asthmatic patients: a systematic review and meta-analysis.	Xiong <i>et al.</i>
12	2023	Machine-learning versus traditional approaches for atherosclerotic cardiovascular risk prognostication in primary prevention cohorts: a systematic review and meta-analysis.	Liu <i>et al.</i>
13	2023	Unassisted clinicians versus deep learning-assisted clinicians in image-based cancer diagnostics: systematic review with meta-analysis	Xue <i>et al.</i>

Fonte: elaborado pelos autores, 2024.

Em seguida, apresentaram-se os artigos e as informações contidas nos textos completos, incluindo tipo de estudo, tamanho da amostra e sua duração e o nível de evidência científica da publicação, representado no Quadro 3.

Quadro 3: Distribuição dos artigos de acordo com delineamento do estudo, níveis de evidência científica do artigo e os fármacos identificados

Nº do artigo	Delineamento do estudo	Nível de evidência científica
1	Revisão sistemática e meta-análise incluindo 5 estudos únicos, 536 pacientes, com 723 nódulos de tireoide, analisando sistemas com assistência diagnóstica por computadores para o diagnóstico de nódulos de tireoide por ultrassom.	Alta
2	Revisão sistemática com meta-análise, com 28 publicações diferentes e 130 modelos de IA únicas, avaliando sua capacidade em prever o surgimento de sepse dentro de UTIs, hospitais e departamentos de emergência.	Alta
3	Revisão da literatura incluindo 43 artigos em busca de algoritmos de aprendizado de máquina capazes de prever a remissão hospital de pacientes do Estados Unidos da América.	Moderada
4	Revisão sistemática com 23 estudos, com o total de 16.274.647 pacientes avaliados, analisando a capacidade de algoritmos de aprendizado de máquina em prever o prognóstico de pacientes no caráter de emergência.	Alta
5	Meta-análise classificando diversos modelos de inteligência artificial no rastreamento, diagnóstico e prognóstico relacionados a Diabetes Mellitus tipo 2, incluindo 34 estudos únicos.	Alta
6	Revisão sistemática e meta-análise com 42 estudos e 55.061 imagens, comparando a diferença de performance entre clínicos e algoritmos na detecção de fraturas em caráter de emergência.	Alta
7	Revisão de literatura com 47 estudos e meta-análise com 6 estudos comparando a performance entre redes neurais convolucionais e radiologistas para detecção de hemorragia intracraniana utilizando tomografia computadorizada.	Alta
8	Revisão sistemática e meta-análise com 27 estudos, comparando a performance de radiologistas e algoritmos de aprendizado de máquina na classificação de degeneração discal, hérnias discais e alteração na classificação Modic.	Alta
9	Revisão sistemática e meta-análise de 10 artigos, buscando identificar modelos de aprendizado de máquina capazes de identificar precocemente pneumonia relacionada à ventilação.	Alta
10	Revisão sistemática com 47 estudos, incluindo 41 algoritmos únicos de aprendizado de máquina e aprendizado profundo, buscando avaliar a sua capacidade de detectar tuberculose em Raio-X de tórax.	Alta
11	Revisão sistemática e meta-análise com 11 estudos, incluindo 23 modelos de IA avaliando a capacidade de esses modelos prever novos episódios de exacerbações em pacientes asmáticos.	Alta
12	Revisão sistemática com meta-análise, incluindo 16 artigos únicos, com o total de 3.302.515 pessoas avaliadas, comparando a performance de algoritmos de aprendizado de máquina para avaliação prognóstica de pacientes com doença aterosclerótica cardiovascular	Alta

13	Revisão sistemática com meta-análise de 48 artigos únicos, com 25 deles apresentando comparação entre a performance de médicos e de algoritmos de aprendizado profundo.	Alta
----	---	------

Fonte: elaborado pelos autores, 2024.

4 DISCUSSÃO

Essa revisão da literatura buscou identificar, analisar e explorar as diversas maneiras pelas quais a IA influencia a prática da clínica médica, bem como os benefícios, malefícios e presentes limitações associados ao seu uso. De maneira geral, os estudos identificaram modelos de IA com eficácia similar aos clínicos/especialistas, tendo grande impacto na análise de exames de imagem e predição de ocorrência de eventos através de dados clínicos.

As IAs alcançaram um grande sucesso nos últimos anos na interpretação dos diversos exames de imagem, devido à sua capacidade de aprender padrões diretamente de dados brutos previamente nomeados, e, por isso, tem sido utilizada nas especialidades que dependem fortemente da interpretação de imagens, como na radiologia. Sabe-se que a ultrassonografia é uma técnica de imagem essencial para detecção de diversas patologias, porém apresenta como maior desvantagem a acurácia ser operador dependente. Zhao *et al.* (2019) analisou sistemas com IA para o diagnóstico de nódulos de tireoide por ultrassom e encontrou uma sensibilidade similar à de radiologistas experientes, porém com menor especificidade do que o radiologista experiente.

Xue *et al.* (2023) compararam acurácia diagnóstica de clínicos sem assistência de IA e os que usavam tais modelos para o diagnóstico de diversas modalidades de câncer, e encontrou uma sensibilidade e especificidade superior para clínicos assistidos por DL, com sensibilidade de 87%, especificidade de 86% e AUC de 0,92.

Xue *et al.* (2023) também analisaram outros exames de imagens no diagnóstico de câncer, incluindo imagens de radiografia, endoscopia, ressonância magnética e exames de ultrassom, encontrando no geral uma sensibilidade e especificidade combinados para clínicos assistidos por DL maiores do que para clínicos não assistidos, em proporções de 1,07 e 1,03, respectivamente. Em relação à radiografia, Hansun *et al.* (2023) analisando modelos de ML e DL para detecção de tuberculose através do exame, encontrou que, em termos de resultados de desempenho, o ML apresentou maior acurácia e sensibilidade, enquanto, em média, os modelos DL alcançaram melhor AUC e especificidade. Apesar disso, como limitação, o estudo apresentava conjuntos de dados com volume relativamente pequenos.

Compte *et al.* (2023), em uma revisão sistemática com meta-análise, comparou a eficácia entre IAs e radiologistas na classificação de hérnias de disco e percepção de alterações na classificação Modic entre exames, e encontrou que os estudos que examinaram hérnias de disco tiveram métricas de desempenho mais altas do que aqueles que examinaram protuberâncias discais. No geral, os modelos de DL apresentaram uma precisão média de 87,0%, especificidade de 90,4% e sensibilidade de 88,2%.

Por fim, Kuo *et al.* (2022) produziu uma revisão sistemática com meta-análise sobre a detecção de fraturas em exames de imagem com IA, e encontrou uma acurácia

significativa, com especificidade de 94% e sensibilidade de 91% para a detecção das diversas fraturas, além de os clínicos encontrarem o diagnóstico mais rapidamente com o uso da IA.

Além do diagnóstico através da imagiologia, a IA também tem sido utilizada para prever a ocorrência de diversas doenças e suas complicações. Fleuren *et al.* (2020), buscou IAs capazes de prever a ocorrência de sepse, sabendo que o diagnóstico precoce de sepse é essencial para a sobrevivência do paciente. Os critérios utilizados previamente, como o SIRS e (q)SOFA possuíam suas limitações, e as IAs se mostraram como uma alternativa à utilização desses critérios. O estudo encontrou que modelos de IA podem prever a sepse com acurácia significativa, porém a minoria (4; 16,6%) dos estudos eram prospectivos, e nenhum deles realizou intervenções médicas baseadas na predição de sepse, mas deixaram a critério do médico atuante.

A revisão de Kareemi *et al.* (2019) encontrou que modelos de aprendizado de máquina (ML) apresentavam capacidade diagnóstica superior ao tratamento usual no departamento de emergência, prevendo com mais precisão um diagnóstico de infarto agudo do miocárdio do que o médico de emergência usando as mesmas variáveis. Além disso, os ML foram capazes de avaliar a mortalidade hospitalar, sendo que os modelos ML superaram as regras de decisão clínica na discriminação na previsão diagnóstica e prognóstica para pacientes do departamento de emergência. Apesar disso, questões metodológicas nos estudos incluídos introduziram risco de viés, e devido à heterogeneidade clínica, não foi possível realizar meta-análise dos estudos.

A IA também tem grande potencial na atenção primária, prevendo o prognóstico de diversas comorbidades. Liu *et al.* (2023) analisou a eficácia dos modelos de IA na predição de risco de pacientes portadores de doença cardiovascular, e encontrou que modelos de ML eram consistentemente superiores em relação aos escores de risco tradicionais na discriminação de prognóstico de risco da doença. A revisão sistemática, porém, não apresentava nenhum estudo prospectivo capaz de validar os modelos testados.

Olusanya *et al.* (2022), por sua vez, em uma meta-análise analisou a eficácia de modelos para determinar a ocorrência de diabetes mellitus tipo 2 (DM2), sendo que na rede neural e *decision tree* foram os modelos mais eficazes para prever o DM2, com acurácia de 0.88 e 0.85, respectivamente. Além disso, Xiong *et al.* (2023) em uma meta-análise sobre a predição de exacerbações asmáticas com IA, encontrou uma sensibilidade e especificidade média de 61,0% e 82,0%, respectivamente, porém, os estudos incluídos eram muito heterogêneos, com grande diferença de tamanho de amostra e janelas de predição (alguns dias a 4 anos).

Jørgensen *et al.* (2022), em sua meta-análise sobre algoritmos para detecção de hemorragia intracraniana (HIC) através de tomografia computadorizada sem contraste, encontraram uma acurácia média ótima, com sensibilidade de 96% e especificidade de 97% para diagnóstico da HIC, porém os estudos eram, em sua maioria retrospectivos, e poucos utilizavam ferramentas de validação externa.

Modelos de IA capazes de prever readmissões hospitalares também foram testados. Huang *et al.* (2021), em sua revisão sistemática, encontrou que os modelos de rede neurais e *gradient boosting* foram os mais eficazes para discriminar os prognósticos dos pacientes, e a maioria dos estudos (28, 65%) tinha algoritmos com AUC acima de

0,70, indicativo de capacidade de discriminação modesta a alta de readmissão hospitalar após até 30 dias. Apesar da IA conseguir identificar pacientes em risco de readmissão, o estudo não conseguiu definir quais variáveis estão mais relacionadas com sua ocorrência.

Embora contribui para salvar vidas, a ventilação mecânica predispõe os pacientes à pneumonia associada à ventilação mecânica (PAV). Frondelius *et al.* (2023) desenvolveram uma revisão sistemática com meta-análise para identificar os fatores de risco mais fortemente associados à ocorrência da PAV e a capacidade de modelos de IA de prever a ocorrência da pneumonia nos pacientes. Ao fim do estudo, foi encontrado uma sensibilidade e especificidade combinadas para PAV de 72,0% e 90,0%, respectivamente, e, comparados às ferramentas de pontuação clínica, os modelos superaram as ferramentas de pontuação PIRO e CPIS em todos os tempos de predição.

Esta revisão teve algumas limitações. Primeiro, alguns estudos tinham tamanhos de amostra muito pequenos, e os estudos incluídos nas revisões tinham também suas limitações, nomeadamente a carência de estudos prospectivos para validação e quantificação do impacto dos modelos IA produzidos previamente. Além disso, os estudos não apresentavam uma metodologia padronizada para analisar estatisticamente a eficácia das IAs, o que dificultou a comparação em diversas revisões aqui incluídas. Os estudos reportavam que os ensaios realizados apresentavam alto risco de viés e baixos índices de qualidades ao aplicar escores como o PROBAST. Com a chegada da ferramenta TRIPOD-AI, que visa avaliar a qualidade de estudos realizados com IAs, espera-se que futuros estudos apresentem metodologias mais firmes e facilitem o processo de comparação em futuras meta-análises.

5 CONCLUSÃO

O campo da IA médica está avançando consideravelmente em direção à adoção em larga escala, especialmente por meio de estudos prospectivos, como os ensaios clínicos randomizados (RCTs), e na área de análise de imagens médicas. No entanto, poucos estudos possuem metodologias com validação externa, avaliação prospectiva e métricas padronizadas para esclarecer o verdadeiro impacto da IA em cenários clínicos reais. A IA demonstrou sua capacidade de extrair insights de fontes variadas, e por vezes, estabeleceu conexões que os humanos normalmente não antecipariam. A união entre IA e clínico é mais benéfica do que uma competição entre os dois; espera-se que surjam abordagens ainda mais criativas e inovadoras na IA médica no futuro. Apesar do grande potencial, questões técnicas e éticas significativas ainda existem na IA médica. À medida que essas questões forem abordadas de forma sistemática, o potencial da IA para melhorar o futuro da medicina se tornará cada vez mais evidente.

REFERÊNCIAS

AHUJA, A. S. The impact of artificial intelligence in medicine on the future role of the physician. **PeerJ**, v. 7, n. 1, p. e7702, 2019.

ASAN, O.; BAYRAK, A. E.; CHOUDHURY, A. Artificial intelligence and human trust in healthcare: focus on clinicians. **Journal of medical Internet research**, v. 22, n. 6, p. e15154, 2020.

BEEDE, E. *et al.* A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. *In: Proceedings of the 2020 CHI conference on human factors in computing systems*. 2020. v.1, n. 1, p. 1-12.

BEJNORDI, B. E. *et al.* Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. **Jama**, v. 318, n. 22, p. 2199-2210, 2017.

BERA, K. *et al.* Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. **Nature reviews Clinical oncology**, v. 16, n. 11, p. 703-715, 2019.

BRASIL. Ministério da Saúde. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia. Diretrizes metodológicas: Sistema GRADE. **Manual de graduação da qualidade da evidência e força de recomendação para tomada de decisão em saúde**. 2014.

BRIGANTI, G.; MOINE, O. L. Artificial intelligence in medicine: today and tomorrow. **Frontiers in medicine**, v. 7, n. 1, p. 27, 2020.

CAMARGO, F. C. *et al.* Competências e barreiras para prática baseada em evidências na enfermagem: revisão integrativa. **Revista Brasileira de Enfermagem**, v. 71, n. 1, p. 2030- 2038, 2018.

COMPTE, Roger *et al.* Are current machine learning applications comparable to radiologist classification of degenerate and herniated discs and Modic change? A systematic review and meta-analysis. **European Spine Journal**, v. 32, n. 11, p. 3764-3787, 2023.

ESTEVA, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, v. 542, n. 7639, p. 115-118, 2017.

FARHUD, D. D.; ZOKAEI, S. Ethical issues of artificial intelligence in medicine and healthcare. **Iranian Journal of Public Health**, v. 50, n. 11, p. i, 2021.

FLEUREN, Lucas M. *et al.* Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. **Intensive care medicine**, v. 46, p. 383-400, 2020.

FRONDELIUS, Tuomas *et al.* Early prediction of ventilator-associated pneumonia with machine learning models: A systematic review and meta-analysis of prediction model performance. **European Journal of Internal Medicine**, 2023.

GHORBANI, A. *et al.* Deep learning interpretation of echocardiograms. **NPJ digital medicine**, v. 3, n. 1, p. 10, 2020.

GONG, E. *et al.* Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. **Journal of magnetic resonance imaging**, v. 48, n. 2, p. 330-340, 2018.

HAENSSLE, H. A. *et al.* Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. **Annals of oncology**, v. 29, n. 8, p. 1836-1842, 2018.

HAMET, P.; TREMBLAY, J. Artificial intelligence in medicine. **Metabolism**, v. 69, n. 1, p. S36-S40, 2017.

HANSUN, Seng *et al.* Machine and deep learning for tuberculosis detection on chest x-rays: systematic literature review. **Journal of medical Internet research**, v. 25, p. e43154, 2023.

HE, J. *et al.* The practical implementation of artificial intelligence technologies in medicine. **Nature medicine**, v. 25, n. 1, p. 30-36, 2019.

HUANG, Yinan *et al.* Application of machine learning in predicting hospital readmissions: a scoping review of the literature. **BMC medical research methodology**, v. 21, p. 1-14, 2021.

JØRGENSEN, Mia Daugaard *et al.* Convolutional neural network performance compared to radiologists in detecting intracranial hemorrhage from brain computed tomography: a systematic review and meta-analysis. **European journal of radiology**, v. 146, p. 110073, 2022.

KAREEMI, Hashim *et al.* Machine learning versus usual care for diagnostic and prognostic prediction in the emergency department: a systematic review. **Academic Emergency Medicine**, v. 28, n. 2, p. 184-196, 2021.

KUO, R. Y *et al.* Artificial intelligence in fracture detection: a systematic review and meta-analysis. **Radiology**, 304(1):50-62, 2022. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/35348381/>.

LARSON, D. B. *et al.* Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. **Radiology**, v. 287, n. 1, p. 313-322, 2018.

LI, H. Impact of artificial intelligence based on big data on medical care. *In: Journal of Physics: Conference Series*. IOP Publishing, 2020. v.1, n. 1, p. 032077.

LIU, Weber *et al.* Machine-learning versus traditional approaches for atherosclerotic cardiovascular risk prognostication in primary prevention cohorts: a systematic review and meta-analysis. **European Heart Journal - Quality of Care and Clinical Outcomes**, v. 9, n. 4, p. 310-322, 2023.

OLUSANYA, Micheal O. *et al.* Accuracy of machine learning classification models for the prediction of type 2 diabetes mellitus: A systematic survey and meta-analysis approach. **International Journal of Environmental Research and Public Health**, v. 19, n. 21, p. 14280, 2022.

RAJPURKAR, P. *et al.* AI in health and medicine. **Nature Medicine**, v. 28, n. 1, p. 31-38, 2022.

SOUZA, M. T. de; SILVA, M. D. da; CARVALHO, Rachel de. Revisão integrativa: o que é e como fazer. **Einstein (São Paulo)**, v. 8, n. 1, p. 102-106, 2010.

SUN, Xingxing *et al.* Diagnostic accuracy of different ECG-based algorithms in wide QRS complex tachycardia: a systematic review and meta-analysis. **BMJ open**, v. 13, n. 7, p. e069273, 2023.

THIAN, Y. L. *et al.* Convolutional neural networks for automated fracture detection and localization on wrist radiographs. **Radiology: Artificial Intelligence**, v. 1, n. 1, p. e180001, 2019.

TOPOL, E. J. High-performance medicine: the convergence of human and artificial intelligence. **Nature Medicine**, v. 25, n. 1, p. 44-56, 2019.

WU, N. *et al.* Deep neural networks improve radiologists' performance in breast cancer screening. **IEEE transactions on medical imaging**, v. 39, n. 4, p. 1184-1194, 2019.

XIAO, Xuan *et al.* Health care cost and benefits of artificial intelligence-assisted population-based glaucoma screening for the elderly in remote areas of China: a cost-offset analysis. **BMC Public Health**, v. 21, n. 1, p. 1065, 2021.

XIONG, Shiqiu *et al.* Machine learning for prediction of asthma exacerbations among asthmatic patients: a systematic review and meta-analysis. **BMC Pulmonary Medicine**, v. 23, n. 1, p. 278, 2023.

XUE, Peng *et al.* Unassisted Clinicians Versus Deep Learning–Assisted Clinicians in Image-Based Cancer Diagnostics: Systematic Review With Meta-analysis. **Journal of Medical Internet Research**, v. 25, p. e43832, 2023.

ZHAO, Wan-Jun *et al.* Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound: A systematic review and meta-analysis. **Medicine**, v. 98, n. 32, p. e16379, 2019.

ZHOU, D. *et al.* Diagnostic evaluation of a deep learning model for optical diagnosis of colorectal cancer. **Nature communications**, v. 11, n. 1, p. 2961, 2020.