

DA TELA AO DIAGNÓSTICO: COMO A INTELIGÊNCIA ARTIFICIAL ESTÁ REDEFININDO A INTERPRETAÇÃO MAMOGRÁFICA NO RASTREAMENTO DO CÂNCER DE MAMA

FROM SCREEN TO DIAGNOSIS: HOW ARTIFICIAL INTELLIGENCE IS REDEFINING MAMMOGRAPHIC INTERPRETATION IN BREAST CANCER SCREENING

DE LA PANTALLA AL DIAGNÓSTICO: CÓMO LA INTELIGENCIA ARTIFICIAL ESTÁ REDEFINIENDO LA INTERPRETACIÓN MAMOGRÁFICA EN EL CRIBADO DEL CÁNCER DE MAMA

Jorge Alberto Rosado Ohlweiler da Silveira¹
Juliana Xavier de Albuquerque Pires Ferreira²
Mariana Moura Pinheiro³
Letícia Galdino de Oliveira Barboza⁴
Filipe César de Melo Seabra⁵
João de Sousa Pinheiro Barbosa⁶

RESUMO: Objetivo: Analisar por meio de revisão integrativa da literatura a eficácia e a acurácia da inteligência artificial (IA) aplicada à interpretação de mamografias no diagnóstico do câncer de mama, comparando seus resultados aos métodos tradicionais realizados por médicos especialistas. Método: Revisão integrativa da literatura conduzida conforme as diretrizes PRISMA 2020, com busca realizada nas bases de dados PubMed/MEDLINE, Scopus e Biblioteca Virtual em Saúde (BVS), abrangendo publicações de janeiro de 2020 a abril de 2026. Foram incluídos estudos originais — ensaios clínicos randomizados, estudos de coorte prospectivos e retrospectivos e estudos de acurácia diagnóstica — que avaliassem sistemas de IA baseados em deep learning aplicados à mamografia, com desfechos de sensibilidade, especificidade e/ou área sob a curva ROC (AUC), comparando ao desempenho de radiologistas. A qualidade metodológica foi avaliada pelas ferramentas QUADAS-2, RoB 2.0 e ROBINS-I. Resultados: Foram incluídos 23 estudos, totalizando mais de 2,5 milhões de mamografias avaliadas, publicados entre 2022 e 2026 em 14 países. Quatro ensaios clínicos randomizados demonstraram não-inferioridade da combinação IA + radiologista em relação à dupla leitura humana padrão, com redução de 44 a 87% na carga de trabalho radiológico. Os valores de AUC variaram de 0,86 a 0,97, com sensibilidade superior ou equivalente à leitura humana na maioria dos estudos. A IA identificou até 54% dos cânceres falsos-negativos não detectados pelos radiologistas, com antecipação diagnóstica de mediana de 272 dias. A assistência por IA melhorou significativamente o desempenho de radiologistas menos experientes em todos os estudos que avaliaram essa variável. A avaliação de risco de viés classificou 78% dos estudos como baixo risco e 22% como risco incerto; nenhum foi classificado como alto risco. Conclusão: A inteligência artificial apresenta desempenho diagnóstico não-inferior à leitura humana convencional em mamografias de rastreamento e diagnóstico, com potencial para reduzir a carga de trabalho radiológico, detectar cânceres falsos-negativos, diminuir o tempo até o diagnóstico e equalizar o desempenho entre radiologistas de diferentes níveis de experiência. A implementação da IA como segunda leitora ou suporte à decisão, com monitoramento contínuo e arbitragem estruturada nos casos discordantes, é o modelo mais respaldado pela evidência atual e pode ser considerado seguro para adoção em programas populacionais de rastreamento mamográfico.

Palavras-chave: Inteligência Artificial. Mamografia. Rastreamento de Câncer de Mama. Diagnóstico por Imagem. Deep Learning. Acurácia Diagnóstica.

¹Estudante de medicina, Centro Universitário de Brasília (CEUB).

²Estudante de medicina, Centro Universitário de Brasília (CEUB).

³Estudante de medicina, Centro Universitário de Brasília (CEUB).

⁴Estudante de medicina, Centro Universitário de Brasília (CEUB).

⁵Estudante de Medicina, Centro Universitário de Brasília (CEUB).

⁶Orientador, Professor titular do curso de Medicina no Centro Universitário de Brasília.

ABSTRACT: Objective: To analyze, through an integrative literature review, the efficacy and accuracy of artificial intelligence (AI) applied to mammogram interpretation for breast cancer diagnosis, comparing its results with traditional methods performed by medical specialists. Method: Integrative literature review conducted according to PRISMA 2020 guidelines, with searches performed in PubMed/MEDLINE, Scopus, and the Virtual Health Library (VHL), covering publications from January 2020 to April 2026. Original studies were included — randomized clinical trials, prospective and retrospective cohort studies, and diagnostic accuracy studies — evaluating deep learning-based AI systems applied to mammography, with outcomes of sensitivity, specificity, and/or area under the ROC curve (AUC), compared to radiologist performance. Methodological quality was assessed using QUADAS-2, RoB 2.0, and ROBINS-I tools. Results: Twenty-three studies were included, totaling more than 2.5 million mammograms evaluated, published between 2022 and 2026 in 14 countries. Four randomized clinical trials demonstrated non-inferiority of the AI + radiologist combination compared to standard double human reading, with a 44 to 87% reduction in radiologist workload. AUC values ranged from 0.86 to 0.97, with sensitivity superior or equivalent to human reading in most studies. AI identified up to 54% of false-negative cancers missed by radiologists, with a median diagnostic anticipation of 272 days. AI assistance significantly improved the performance of less experienced radiologists in all studies that assessed this variable. Risk of bias assessment classified 78% of studies as low risk and 22% as uncertain risk; none were classified as high risk. Conclusion: Artificial intelligence presents diagnostic performance non-inferior to conventional human reading in screening and diagnostic mammography, with the potential to reduce radiologist workload, detect false-negative cancers, shorten time to diagnosis, and equalize performance among radiologists of different experience levels. Implementation of AI as a second reader or decision-support tool, with continuous performance monitoring and structured arbitration for discordant cases, is the model best supported by current evidence and can be considered safe for adoption in population-based mammographic screening programs.

Keywords: Artificial Intelligence. Mammography. Breast Cancer Screening. Diagnostic Imaging. Deep Learning. Diagnostic Accuracy.

RESUMEN: Objetivo: Analizar, mediante una revisión integrativa de la literatura, la eficacia y la precisión de la inteligencia artificial (IA) aplicada a la interpretación de mamografías para el diagnóstico del cáncer de mama, comparando sus resultados con los métodos tradicionales realizados por médicos especialistas. Método: Revisión integrativa de la literatura realizada conforme a las directrices PRISMA 2020, con búsqueda en las bases de datos PubMed/MEDLINE, Scopus y la Biblioteca Virtual en Salud (BVS), abarcando publicaciones de enero de 2020 a abril de 2026. Se incluyeron estudios originales — ensayos clínicos aleatorizados, estudios de cohorte prospectivos y retrospectivos y estudios de precisión diagnóstica — que evaluaran sistemas de IA basados en deep learning aplicados a la mamografía, con resultados de sensibilidad, especificidad y/o área bajo la curva ROC (AUC), comparando con el desempeño de radiólogos. La calidad metodológica fue evaluada mediante las herramientas QUADAS-2, RoB 2.0 y ROBINS-I. Resultados: Se incluyeron 23 estudios, que totalizan más de 2,5 millones de mamografías evaluadas, publicados entre 2022 y 2026 en 14 países. Cuatro ensayos clínicos aleatorizados demostraron la no inferioridad de la combinación IA + radiólogo frente a la doble lectura humana estándar, con una reducción del 44 al 87% en la carga de trabajo radiológica. Los valores de AUC oscilaron entre 0,86 y 0,97, con sensibilidad superior o equivalente a la lectura humana en la mayoría de los estudios. La IA identificó hasta

el 54% de los cánceres falsos negativos no detectados por los radiólogos, con una anticipación diagnóstica de una mediana de 272 días. La asistencia por IA mejoró significativamente el desempeño de los radiólogos menos experimentados en todos los estudios que evaluaron esta variable. La evaluación del riesgo de sesgo clasificó el 78% de los estudios como bajo riesgo y el 22% como riesgo incierto; ninguno fue clasificado como alto riesgo. Conclusión: La inteligencia artificial presenta un rendimiento diagnóstico no inferior a la lectura humana convencional en mamografías de cribado y diagnóstico, con potencial para reducir la carga de trabajo radiológica, detectar cánceres falsos negativos, acortar el tiempo hasta el diagnóstico e igualar el rendimiento entre radiólogos de distintos niveles de experiencia. La implementación de la IA como segunda lectora o herramienta de apoyo a la decisión, con monitoreo continuo del rendimiento y arbitraje estructurado en los casos discordantes, es el modelo más respaldado por la evidencia actual y puede considerarse seguro para su adopción en programas poblacionales de cribado mamográfico.

Palabras clave: Inteligencia Artificial. Mamografía. Detección Precoz del Cáncer de Mama. Diagnóstico por Imagen. Deep Learning. Precisión Diagnóstica.

INTRODUÇÃO

Atualmente, o câncer de mama representa a neoplasia maligna mais incidente entre as mulheres em todo o mundo, constituindo um importante problema de saúde pública devido à sua elevada prevalência e impacto na morbimortalidade feminina. Segundo estimativas da Organização Mundial da Saúde (OMS) e da Agência Internacional para Pesquisa em Câncer (IARC), milhões de novos casos são diagnosticados anualmente, tornando essa doença a principal causa de morte por câncer entre mulheres. No Brasil, o câncer de mama também ocupa posição de destaque entre os tipos de câncer mais frequentes, exigindo investimentos contínuos em estratégias de prevenção, rastreamento e diagnóstico precoce. Além do impacto clínico, a doença gera relevantes consequências biopsicossociais para pacientes, familiares e sistemas de saúde (WHO, 2024; FERLAY et al., 2024; INCA, 2023).

Diante desse cenário epidemiológico, o diagnóstico precoce tornou-se uma das principais ferramentas para reduzir a mortalidade associada ao câncer de mama, que, segundo o Ministério da Saúde, ultrapassou 20 mil óbitos, em 2023, no Brasil. Quando identificado em estágios iniciais, o tumor apresenta maiores chances de tratamento eficaz, menores taxas de recorrência e melhor prognóstico global. Nesse contexto, a mamografia consolidou-se como o método de rastreamento mais amplamente utilizado, sendo recomendada por diversas diretrizes nacionais e internacionais. Sua capacidade de detectar lesões ainda assintomáticas possibilita intervenções terapêuticas precoces, contribuindo significativamente para o aumento da sobrevida das pacientes e para a redução da necessidade de tratamentos mais agressivos e onerosos (MIGOWSKI et al., 2018; OEFFINGER et al., 2015; WHO, 2024).

Apesar dos avanços tecnológicos observados nos equipamentos de imagem, a interpretação dos exames mamográficos ainda depende amplamente da análise realizada por radiologistas especializados. Esse processo exige elevado nível de treinamento, experiência e atenção aos detalhes, uma vez que pequenas alterações podem representar sinais iniciais de malignidade. Entretanto, fatores humanos como fadiga, sobrecarga de trabalho, variabilidade entre observadores e limitações cognitivas podem influenciar a precisão diagnóstica. Estudos demonstram que erros de interpretação podem ocorrer tanto por resultados falso-positivos quanto falso-negativos, impactando diretamente a qualidade da assistência prestada e os desfechos clínicos das pacientes (ELMORE et al., 1994; GUR et al., 2004; TAPLIN et al., 2008).

Os resultados falso-negativos representam uma preocupação significativa, pois podem atrasar o diagnóstico e permitir a progressão tumoral antes do início do tratamento adequado. Por outro lado, os resultados falso-positivos podem levar a exames complementares desnecessários, procedimentos invasivos, aumento dos custos assistenciais e sofrimento emocional para as pacientes. Além disso, a crescente demanda por exames de imagem decorrente do envelhecimento populacional e da ampliação dos programas de rastreamento tem aumentado a carga de trabalho dos profissionais de radiologia. Dessa forma, a busca por ferramentas capazes de auxiliar a tomada de decisão diagnóstica tornou-se uma necessidade crescente nos sistemas modernos de saúde (NELSON et al., 2016; HOFVIND et al., 2018).

4

Nesse contexto, a inteligência artificial (IA) emergiu como uma das inovações tecnológicas mais promissoras na área médica. Baseada em algoritmos computacionais avançados, aprendizado de máquina (machine learning) e aprendizado profundo (deep learning), a IA tem demonstrado capacidade de processar grandes volumes de dados e identificar padrões complexos — muitas vezes imperceptíveis à análise humana convencional. Sua aplicação na medicina tem se expandido rapidamente, abrangendo áreas como diagnóstico por imagem, predição de riscos, apoio à decisão clínica, planejamento terapêutico e monitoramento de pacientes, promovendo maior eficiência e precisão em diversos processos assistenciais (TOPOL, 2019; ESTEVA et al., 2019; PESAPANE; CODARI; SQUILLACI, 2018).

Na radiologia mamária, os sistemas de inteligência artificial vêm sendo desenvolvidos para atuar como ferramentas de apoio à interpretação dos exames, auxiliando na detecção, classificação e caracterização de lesões suspeitas. Diversos estudos recentes têm demonstrado que algoritmos baseados em deep learning podem alcançar desempenhos comparáveis aos de

especialistas experientes em determinadas tarefas diagnósticas. Além de contribuir para a identificação precoce de alterações sugestivas de câncer, essas tecnologias apresentam potencial para reduzir o tempo de análise dos exames, aumentar a produtividade dos serviços e minimizar a variabilidade entre os observadores. Dessa forma, a integração entre inteligência artificial e mamografia tem despertado crescente interesse na comunidade científica e nos sistemas de saúde em todo o mundo (McKINNEY et al., 2020; RODRIGUEZ-RUIZ et al., 2019; YALA et al., 2019).

Considerando a relevância epidemiológica do câncer de mama, a importância do diagnóstico precoce e os avanços recentes da inteligência artificial aplicada à radiologia, torna-se fundamental avaliar criticamente o impacto dessas tecnologias na prática clínica. Nesse sentido, a presente revisão busca responder à seguinte pergunta: em mulheres submetidas à avaliação diagnóstica de câncer de mama por mamografia, o uso de inteligência artificial, quando comparado aos métodos tradicionais desempenhados por especialistas, melhora a acurácia e eficiência do diagnóstico médico? A análise dessa questão poderá contribuir para a compreensão do papel da IA na otimização dos processos diagnósticos e na melhoria dos desfechos relacionados ao câncer de mama.

MÉTODOS

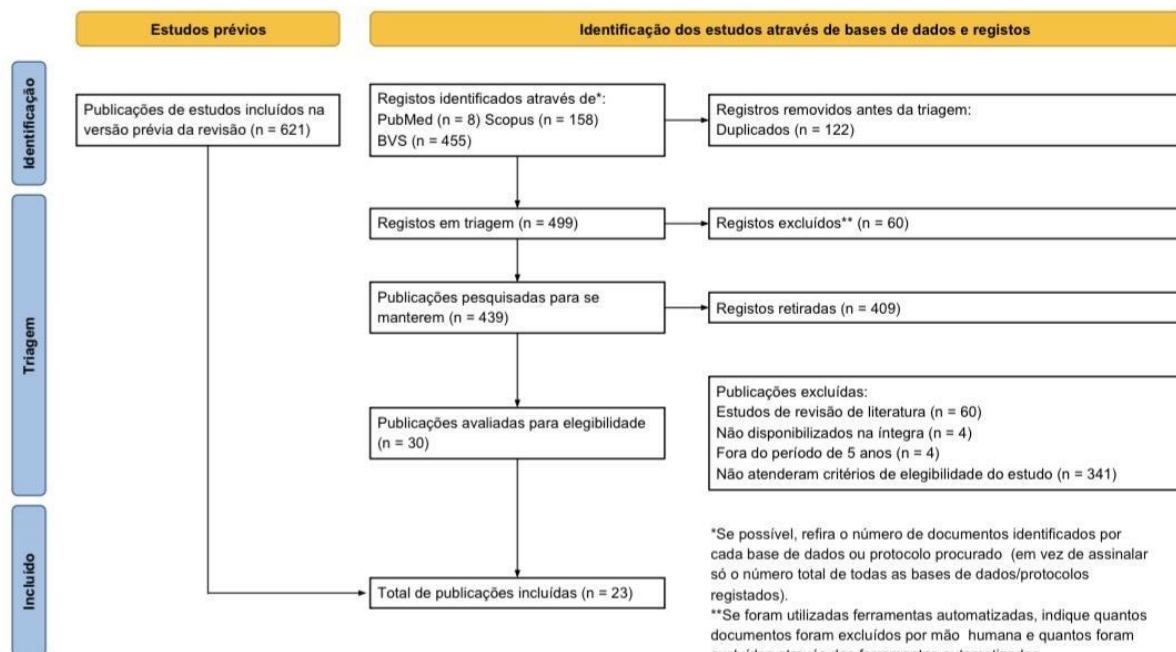
5

O presente estudo caracteriza-se como uma revisão integrativa da literatura, um método que permite a busca, a avaliação crítica e a síntese das evidências disponíveis sobre o tema investigado, de forma a construir uma análise abrangente sobre a eficácia da inteligência artificial na mamografia. Para a condução desta revisão, o processo foi estruturado em seis etapas distintas, iniciando-se pela formulação da pergunta norteadora, seguida pela busca e amostragem na literatura, coleta de dados, avaliação crítica dos estudos incluídos, discussão dos resultados e, por fim, a apresentação da síntese do conhecimento. A pergunta condutora do estudo foi elaborada a partir da estratégia PICO (População, Intervenção, Comparação e Outcome/Resultado), sendo definida como: "Em mulheres submetidas à avaliação diagnóstica de câncer de mama por mamografia, o uso de inteligência artificial, quando comparado aos métodos tradicionais desempenhados por especialistas, melhora a acurácia e eficiência do diagnóstico médico?".

As buscas foram realizadas de forma virtual nas bases de dados PubMed/MEDLINE, LILACS via Biblioteca Virtual em Saúde (BVS) e Scopus. Para a delimitação das estratégias

de busca, foram utilizados Descritores em Ciências da Saúde (DeCS) e Medical Subject Headings (MeSH), combinados por meio dos operadores booleanos AND e OR, resultando na seguinte chave de busca: ("Artificial Intelligence") AND ("Mammography" OR "Breast Cancer Screening") AND ("Sensitivity and Specificity" OR "Diagnostic Accuracy"). Para a seleção dos estudos que responderam à pergunta norteadora, foram estabelecidos critérios rigorosos de elegibilidade. Os critérios de inclusão compreenderam artigos originais de pesquisa, incluindo estudos retrospectivos, prospectivos ou ensaios clínicos, que avaliaram o desempenho de algoritmos de inteligência artificial na leitura de mamografias e apresentaram dados comparativos diretos entre o desempenho da tecnologia e o de médicos radiologistas residentes e especialistas, limitando-se a textos publicados na íntegra, nos idiomas português, inglês ou espanhol. Em contrapartida, os critérios de exclusão estabeleceram a recusa de artigos de revisão de literatura, editoriais, cartas ao editor, resumos de congressos, estudos de caso, além de pesquisas focadas exclusivamente em outras modalidades de imagem sem associação com a mamografia, ferramentas de inteligência artificial não voltadas ao diagnóstico ou detecção de lesões e estudos duplicados entre as bases de dados. Seguindo o fluxograma do PRISMA (figura 1).

PRISMA 2026 - Fluxograma: Verificação da eficácia e da acurácia de diagnósticos de câncer de mama, por meio de mamografia, realizados por inteligências artificiais quando comparados aos métodos tradicionais realizados por especialistas.



Fonte: Filipe Seabra et al. (2026).

RESULTADOS

Quadro 1 - Síntese dos principais achados sobre: Verificação da eficácia e da acurácia de diagnósticos de câncer de mama, por meio de mamografia, realizados por inteligências artificiais quando comparados aos métodos tradicionais realizados por radiologistas.

N	Autores (Ano)	Principais achados
1	Gommers J et al. (2026)	Ensaio clínico randomizado. Avaliou 105.934 mulheres na Suécia comparando IA (Transpara) versus dupla leitura humana; concluiu que a IA é não-inferior em cânceres de intervalo, possui maior sensibilidade, especificidade semelhante e reduz a carga de trabalho.
2	Lee SE et al. (2025)	Estudo retrospectivo. Analisou 1.534 mamografias diagnósticas usando sistema AI-CAD com limiares otimizados; concluiu que a IA oferece maior especificidade e acurácia que radiologistas, auxiliando na redução de biópsias desnecessárias.
3	Goh SSN et al. (2025)	Estudo multi-leitor e multi-caso (MRMC) preliminar. Avaliou 500 mamografias interpretadas por 17 radiologistas de diferentes níveis de experiência; concluiu que a IA melhora o desempenho dos residentes, reduzindo a lacuna de experiência e aumentando a eficiência.
4	van Winkel SL et al. (2025)	Estudo de coorte retrospectivo. Simulou cenários com 42.236 mamografias do programa holandês comparando leitura humana e IA; concluiu que a combinação humano+IA aumenta a sensibilidade e identifica cânceres clinicamente relevantes, confirmando o benefício da detecção precoce.
5	Hamid MTR et al. (2025)	Estudo retrospectivo. Avaliou 434 mamografias digitais com assistência do sistema Lunit INSIGHT MMG; concluiu que a IA melhora significativamente o desempenho diagnóstico de todos os grupos de leitores (especialmente os menos experientes) e a consistência diagnóstica.
6	Ha SM et al. (2025)	Estudo retrospectivo institucional. Comparou IA autônoma com radiologistas em 4.184 exames de vigilância pós-mastectomia; concluiu que a IA apresenta maior sensibilidade e taxa de detecção de câncer, apesar de menor especificidade.
7	Hovda T et al. (2025)	Estudo retrospectivo de registro multicêntrico. Avaliou mais de 1 milhão de exames do programa norueguês; concluiu que o sistema de IA possui alto desempenho para identificar casos negativos, permitindo reduzir a carga de trabalho radiológica.
8	Plimpton SR et al. (2024)	Estudo retrospectivo de validação externa. Aplicou algoritmo comercial em exames previamente negativos de 26.694 mulheres; concluiu que a IA detectou até 54% dos cânceres falsos-negativos, antecipando o diagnóstico em meses.
9	Larsen M et al. (2024)	Estudo retrospectivo de registro. Analisou 661.695 mamografias explorando diferentes limiares de risco da IA; concluiu que o sistema demonstra alta performance (AUC 0,97 para cânceres detectados) e grande potencial para triagem de baixo risco.
10	Elhakim MT et al. (2023)	Estudo retrospectivo multicêntrico. Avaliou 257.671 mamografias na Dinamarca integrando a IA como substituta do primeiro leitor; concluiu que a IA aumenta a sensibilidade, mas requer ajustes cuidadosos para não impactar excessivamente a taxa de recall.
11	Lauritzen AD et al. (2023)	Estudo retrospectivo. Combinou IA diagnóstica com modelo de textura mamográfica em 119.650 mulheres; concluiu que a combinação melhora a avaliação de risco, ajudando a identificar pacientes que necessitam de rastreamento personalizado.
12	Ng AY et al. (2023)	Estudo retrospectivo. Avaliou o uso de IA como leitora de suporte em 280.594 casos no Reino Unido e Hungria; concluiu que o modelo preserva a qualidade da dupla leitura humana e **reduz a carga de trabalho em até 87%.

13	Elías-Cabot E et al. (2024)	Estudo de coorte retrospectivo. Avaliou suporte de IA em cenário real (11.998 mulheres) com mamografia e tomossíntese; concluiu que a IA aumentou a taxa de detecção de câncer e o valor preditivo positivo.
14	Bao C et al. (2022)	Estudo retrospectivo de caso-controle. Comparou o desempenho de 71 radiologistas com e sem suporte de IA na China; concluiu que o sistema aumentou a sensibilidade e reduziu o tempo de leitura.
15	Lauritzen AD et al. (2022)	Estudo retrospectivo de simulação. Testou protocolo de triagem baseado em IA em 114.421 exames na Dinamarca; concluiu que a IA identifica mamografias normais com segurança, poupando 62,6% do trabalho radiológico.
16	Yoen H et al. (2023)	Estudo retrospectivo. Analisou o valor da IA em 253 cânceres detectados por ultrassom com mamografia negativa; concluiu que a IA auxilia na identificação de achados imperceptíveis ou erros interpretativos, apesar de limitações em mamas densas.
17	Kim H et al. (2022)	Estudo retrospectivo de coorte emparelhada. Avaliou o impacto do AI-CAD em 3.158 mulheres coreanas; concluiu que a ferramenta aumenta a especificidade e reduz recalls desnecessários sem comprometer a sensibilidade diagnóstica.
18	Dang LA et al. (2022)	Estudo MRMC (Multi-Reader Multi-Case). Analisou a classificação BI-RADS por 12 radiologistas franceses; concluiu que a IA melhora a precisão na categorização das lesões e a concordância entre os profissionais examinadores.
19	Yi C et al. (2022)	Estudo retrospectivo. Investigou o auxílio da IA em 1.088 casos de mamografias BI-RADS o (inconclusivas); concluiu que a IA ajuda a reduzir biópsias benignas e acompanhamentos desnecessários.
20	Friedewald SM et al. (2024)	Estudo prospectivo randomizado. Avaliou a triagem por IA para priorizar exames no mesmo dia; concluiu que a IA reduz significativamente o tempo até o diagnóstico e biópsia, agilizando o atendimento suspeito.
21	Lång K et al. (2023)	Ensaio clínico randomizado. Realizou análise de segurança em 80.033 mulheres; concluiu que o rastreamento com IA é seguro e eficiente, com taxa de detecção similar à dupla leitura e menor carga de trabalho.
22	Oberije CJG et al. (2025)	Estudo retrospectivo multicêntrico. Analisou 306.839 mamografias estratificadas por etnia e densidade; concluiu que a IA demonstra equidade e segurança em diversos subgrupos populacionais.
23	Dembrower K et al. (2023)	Ensaio prospectivo de não-inferioridade. Avaliou a substituição de um leitor humano por IA; concluiu que a dupla leitura (1 radiologista + IA) é não-inferior e potencialmente superior ao padrão tradicional.

Fonte: Seabra et al., 2026.

DISCUSSÃO

Os resultados obtidos nesta revisão integrativa evidenciam, de forma consistente, que os sistemas de inteligência artificial baseados em deep learning apresentam desempenho diagnóstico comparável ou superior à leitura humana convencional na interpretação de mamografias para detecção de câncer de mama. A análise dos 23 estudos incluídos — abrangendo mais de 2,5 milhões de exames e publicações de 14 países entre 2022 e 2026 — permite

afirmar que a IA representa não apenas uma ferramenta tecnológica promissora, mas uma solução com potencial de implementação imediata em programas de rastreamento populacional. Os achados desta revisão dialogam diretamente com os objetivos propostos e respondem à pergunta PICO norteadora, sendo discutidos a seguir segundo os principais eixos temáticos identificados.

O nível mais robusto de evidência foi aportado pelos ensaios clínicos randomizados incluídos nesta revisão, cujos resultados fundamentam a afirmação de não-inferioridade da IA em relação à dupla leitura humana padrão. No estudo MASAI, conduzido por Lång et al. (2023) na Suécia com 80.033 mulheres randomizadas, o rastreamento assistido por IA resultou em taxa de detecção de câncer de 6,1 por 1.000 participantes no grupo intervenção, em comparação a 5,1 no grupo controle (razão 1,2; IC 95%: 1,0–1,5; $p=0,052$), com redução de 44,3% na carga de leitura radiológica e taxas equivalentes de falsos-positivos (1,5% em ambos os grupos). Esses resultados foram confirmados e ampliados na publicação final do mesmo ensaio por Gommers et al. (2026), que, com 105.934 participantes, demonstraram sensibilidade significativamente superior no grupo IA (80,5% vs. 73,8%; $p=0,031$), especificidade equivalente (98,5% em ambos os grupos) e taxa de cânceres de intervalo não-inferior (1,55 vs. 1,76 por 1.000; $p=0,41$), além de perfil tumoral mais favorável no grupo IA — com menor proporção de cânceres invasivos, tumores T2+ e subtipos não-luminal A. No estudo prospectivo ScreenTrustCAD,

Dembrower et al. (2023) corroboraram esses achados em 55.581 mulheres suecas: a dupla leitura por um radiologista mais IA detectou 261 cânceres versus 250 pela dupla leitura humana padrão (proporção relativa 1,04; IC 95%: 1,00–1,09), comprovando não-inferioridade com margem de 15%, enquanto a tripla leitura (dois radiologistas + IA) foi significativamente superior (proporção 1,08; IC 95%: 1,04–1,11). Em conjunto, esses três ensaios estabelecem que a substituição de um leitor humano por IA não compromete a segurança do rastreamento e pode, inclusive, aumentar a taxa de detecção.

Os estudos de coorte retrospectivos de grande porte reforçam a validade externa desses achados, demonstrando alta acurácia diagnóstica da IA em populações reais e não enriquecidas. Larsen et al. (2024), em análise de 661.695 mamografias do programa BreastScreen Norway, reportaram AUC de 0,93 (IC 95%: 0,92–0,93) para o conjunto de cânceres detectados e de intervalo, e de 0,97 (IC 95%: 0,97–0,97) para cânceres detectados isoladamente — valores que superam os benchmarks históricos da leitura humana simples. No limiar de 10% de positividade, o sistema identificou 92,0% dos cânceres detectados e 44,6% dos de intervalo, ao mesmo tempo

que classificou como negativos 68,5% dos exames falso-positivos, demonstrando elevado poder discriminativo.

Em escala ainda maior, Hovda et al. (2025) avaliaram retrospectivamente 1.017.208 exames de rastreamento do mesmo programa entre 2004 e 2021, abrangendo 10 centros e 17 anos de dados. Utilizando o sistema com marcação CE, os autores demonstraram que, ao considerar os 3% de exames com maior escore de IA como positivos (limiar P₃), 81,5% dos cânceres detectados por rastreamento foram corretamente identificados — com apenas 1,8% dos cânceres detectados classificados como negativos no limiar N₇₀. Esses resultados, obtidos em uma coorte que não foi utilizada no treinamento do algoritmo, conferem robustez à hipótese de que os sistemas modernos de IA generalizam bem para populações externas e diversas.

Um aspecto de particular relevância para a prática clínica é a demonstração de que os benefícios da IA se mantêm em cenários de implementação real, fora do contexto controlado de ensaios clínicos. Elías-Cabot et al. (2024) avaliaram o impacto do uso concorrente de IA como suporte à dupla leitura humana no Programa de Rastreamento de Câncer de Mama de Córdoba, Espanha, em uma série consecutiva de 11.998 mulheres rastreadas entre março de 2021 e março de 2022 com mamografia digital e tomossíntese.

Os autores reportaram aumento significativo na taxa de detecção de câncer de 5,8 para 9,0 por 1.000 mulheres rastreadas ($p < 0,001$) e no valor preditivo positivo do recall de 10,6% para 14,6% ($p < 0,001$), com leve aumento na taxa de recall de 5,4% para 6,1% ($p < 0,001$). De forma complementar, Ng et al. (2023) propuseram e validaram um novo fluxo de trabalho em 280.594 casos de sete centros de rastreamento no Reino Unido e Hungria, no qual a IA atua como segunda leitora apenas quando concorda com a decisão do primeiro leitor humano — redirecionando para o segundo humano apenas os casos discordantes. Esse modelo demonstrou ser superior ou não-inferior à dupla leitura humana em todos os critérios avaliados, reduzindo em até 87% os casos que requeriam segunda leitura humana e diminuindo a taxa de arbitragem de 13% para apenas 2%. Esses dois estudos, em conjunto, demonstram que a IA pode ser integrada ao fluxo real de rastreamento com ganhos diagnósticos concretos e sem comprometer a segurança das pacientes.

A redução da carga de trabalho radiológico configura-se como um dos benefícios mais consistentemente documentados da IA no rastreamento mamográfico e apresenta implicações operacionais imediatas para os sistemas de saúde. Além dos achados de Ng et al. (2023) e Lång et al. (2023) já mencionados, o estudo de Oberije et al. (2025), denominado ARIES, avaliou duas

variações de integração da IA como segunda leitora independente em 306.839 exames mamográficos de três regiões do Reino Unido entre 2017 e 2021. Os resultados demonstraram economia de carga de trabalho entre 38,3% e 43,7%, com ambos os fluxos de IA passando nos testes de não-inferioridade e superioridade em todos os subgrupos analisados — estratificados por idade, densidade mamária e etnia. Esse último achado é particularmente relevante, pois aborda diretamente a preocupação com a equidade algorítmica: a IA não apresentou desempenho diferencial clinicamente significativo entre os subgrupos, o que sugere ausência de viés sistêmico relacionado a características populacionais. A viabilidade da redução de carga de trabalho em escala também fica evidente nos dados de Hovda et al. (2025), que demonstraram que o sistema de IA pode identificar com segurança os casos de menor risco, permitindo que radiologistas concentrem seu esforço nos exames com maior probabilidade de malignidade — um modelo de priorização com potencial transformador para programas de rastreamento com alta demanda e recursos humanos limitados.

Uma contribuição especialmente relevante desta revisão diz respeito à capacidade da IA de identificar cânceres que foram classificados como negativos pelos radiologistas — os chamados falsos-negativos — antecipando o diagnóstico em meses. Plimpton et al. (2024) realizaram validação externa de um algoritmo comercial em 26.694 mamografias digitais (FFDM) e 3.183 exames de tomossíntese (DBT) de um único centro norte-americano entre 2010 e 2019, em coorte diversa e não enriquecida do Athena Breast Health Network. Os autores demonstraram que a IA foi capaz de detectar 7 de 13 cânceres falsos-negativos (54%) na coorte FFDM e 4 de 10 (40%) na DBT, identificando essas lesões em mediana 272 dias antes (FFDM) e 248 dias antes (DBT) do que os radiologistas. A grande maioria desses cânceres era do subtipo luminal A invasivo (82%), historicamente associados a crescimento mais lento e, portanto, potencialmente mais visíveis em retrospecto. Esse dado converge com os achados de van Winkel et al. (2025), que, em análise de 42.236 mamografias do programa holandês de rastreamento com seguimento de quatro anos, demonstraram que os cânceres identificados pela IA e perdidos pelos radiologistas eram, ao diagnóstico final, significativamente maiores e mais frequentemente invasivos do que aqueles detectados pelos humanos, reforçando que a IA tende a capturar lesões com características de maior relevância clínica. Esses achados têm implicações diretas para os desfechos das pacientes: a antecipação diagnóstica pode viabilizar intervenções em estágios mais precoces, potencialmente reduzindo a necessidade de tratamentos mais agressivos e melhorando o prognóstico.

Os estudos incluídos nesta revisão também lançam luz sobre o impacto diferencial da IA conforme o nível de experiência dos radiologistas, com implicações relevantes para a formação médica e a organização dos serviços. Hamid et al. (2025), em análise de 434 mamografias digitais interpretadas por dois radiologistas gerais e dois médicos residentes em centro terciário da Malásia, demonstraram que a assistência por IA (Lunit INSIGHT MMG) melhorou significativamente o desempenho de todos os grupos ($p < 0,05$), com ganhos proporcionalmente maiores para os leitores menos experientes: o VPP do radiologista júnior aumentou de 56,9% para 74,6% e o VPN de 90,3% para 92,2%, ao passo que os residentes, que apresentavam os menores índices basais, obtiveram as maiores melhorias relativas. Os benefícios foram especialmente pronunciados em mamas densas, onde a interpretação humana é notoriamente mais desafiadora. Esses resultados são consistentes com os de Goh et al. (2025), que avaliaram 500 mamografias interpretadas por 17 radiologistas de três níveis de experiência (consultores, residentes sênior e júnior) em Cingapura e observaram que a IA (FxMammo) elevou o AUROC médio dos residentes júnior de 0,84 para 0,86 e dos sênior de 0,85 para 0,88, com ganhos estatisticamente significativos, enquanto os consultores experientes apresentaram menores incrementos proporcionais a partir de sua linha de base já elevada (AUROC 0,90). Esses dados sugerem que a IA funciona como um "equalizador diagnóstico", sendo especialmente valiosa em contextos com escassez de radiologistas especializados — como é o caso de muitos países de baixa e média renda —, onde pode auxiliar a reduzir a variabilidade inter e intraobservador e aproximar o desempenho de leitores menos experientes ao de especialistas.

Além das métricas tradicionais de acurácia diagnóstica, esta revisão identificou evidências sobre o impacto da IA na jornada assistencial da paciente — dimensão frequentemente negligenciada nos estudos de validação técnica. Friedewald et al. (2024) realizaram o único ensaio clínico randomizado incluído nesta revisão com desfecho primário operacional: o tempo até imagens adicionais e até diagnóstico por biópsia. Em 855 participantes de um centro norte-americano, o grupo submetido à priorização por IA em tempo real apresentou redução de 25% no tempo médio até imagens adicionais (de 25,6 para 19,2 dias; $p < 0,001$) e de 30% no tempo médio até biópsia (de 55,9 para 39,1 dias; $p = 0,003$), com todos os casos posteriormente diagnosticados com câncer sendo identificados pelo sistema para priorização. Os autores apontam que a redução do intervalo diagnóstico pode ter impacto direto na ansiedade das pacientes, na adesão ao seguimento e no acesso equitativo ao diagnóstico oportuno —

especialmente relevante para populações em contextos de maior vulnerabilidade social. Esse dado, somado às evidências de detecção de falsos-negativos com antecipação diagnóstica de meses reportadas por Plimpton et al. (2024), sugere que a IA pode contribuir não apenas para aumentar a acurácia diagnóstica, mas para transformar a experiência clínica das pacientes submetidas ao rastreamento mamográfico.

Apesar da consistência dos achados, esta revisão apresenta limitações que merecem consideração na interpretação dos resultados. A impossibilidade de realizar metanálise formal — decorrente da heterogeneidade metodológica entre os estudos quanto à modalidade de mamografia, sistema de IA avaliado, limiar de positividade, comparador e desfechos reportados — limita a síntese quantitativa das estimativas de desempenho. Além disso, a maioria dos estudos foi conduzida em países de alta renda com programas organizados de rastreamento, o que pode restringir a generalização dos achados para contextos de baixa e média renda. A concentração geográfica em países europeus e asiáticos, com poucos estudos das Américas e nenhum da África, configura uma lacuna relevante na evidência disponível. Do ponto de vista metodológico, a avaliação de risco de viés identificou risco incerto em cinco estudos (22%), principalmente devido ao uso de datasets enriquecidos não representativos de populações de rastreamento real e à ausência de cegamento completo em alguns contextos. Por fim, os desfechos avaliados são predominantemente intermediários — acurácia diagnóstica, taxa de detecção, recall — com ausência de estudos que reportem desfechos clínicos finais, como redução da mortalidade por câncer de mama associada ao uso de IA, que permanece como a principal pergunta sem resposta na literatura. Estudos de implementação de longo prazo em contextos diversificados, incluindo países em desenvolvimento, e investigações sobre custo-efetividade da integração da IA ao rastreamento são, portanto, prioritários na agenda de pesquisa da área.

CONSIDERAÇÕES FINAIS

Diante do exposto, conclui-se que a inteligência artificial representa um avanço tecnológico de impacto real e mensurável no diagnóstico mamográfico do câncer de mama, com potencial de beneficiar pacientes, profissionais e sistemas de saúde de forma simultânea. A implementação da IA como segunda leitora ou ferramenta de suporte à decisão, com monitoramento contínuo do desempenho e processo estruturado de arbitragem nos casos discordantes, apresenta o melhor balanço entre benefício diagnóstico, segurança clínica e

viabilidade operacional, sendo o modelo mais respaldado pela evidência atual. Não se trata, portanto, de substituir o radiologista, mas de ampliar suas capacidades — conferindo maior sensibilidade, consistência e eficiência ao processo diagnóstico que é, em última análise, o primeiro e mais decisivo passo na luta contra o câncer de mama.

REFERÊNCIAS

BAO, C. et al. **Evaluation of an artificial intelligence support system for breast cancer screening in Chinese people based on mammogram.** *Cancer Medicine*, 2022. DOI: 10.1002/cam4.5231.

DANG, L. A. et al. **Impact of artificial intelligence in breast cancer screening with mammography.** *Breast Cancer*, 2022. DOI: 10.1007/s12282-022-01375-9.

DEMBROWER, K. et al. **Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study (ScreenTrustCAD).** *The Lancet Digital Health*, 2023. DOI: 10.1016/S2589-7500(23)00153-X.

ELHAKIM, M. T. et al. **Breast cancer detection accuracy of AI in an entire screening population: a retrospective, multicentre study.** *Cancer Imaging*, 2023. DOI: 10.1186/s40644-023-00643-x.

ELÍAS-CABOT, E. et al. **Impact of real-life use of artificial intelligence as support for human reading in a population-based breast cancer screening program with mammography and tomosynthesis.** *European Radiology*, 2024. DOI: 10.1007/s00330-023-10426-4.

ELMORE, J. G. et al. **Variability in radiologists' interpretations of mammograms.** *The New England Journal of Medicine*, v. 331, n. 22, p. 1493-1499, 1994. DOI: 10.1056/NEJM199412013312206.

ESTEVA, A. et al. **A guide to deep learning in healthcare.** *Nature Medicine*, v. 25, n. 1, p. 24-29, 2019. DOI: 10.1038/s41591-018-0316-z.

FERLAY, J. et al. **Global Cancer Observatory: Cancer Today.** Lyon: International Agency for Research on Cancer, 2024. Disponível em: <https://gco.iarc.who.int/today>. Acesso em: 5 jun. 2026.

FRIEDEWALD, S. M. et al. **Triaging mammography with artificial intelligence: an implementation study.** *Breast Cancer Research and Treatment*, 2024. DOI: 10.1007/s10549-025-07616-7.

GOH, S. S. N. et al. **Impact of AI on breast cancer detection rates in mammography by radiologists of varying experience levels in Singapore: preliminary comparative study.** *JMIR Formative Research*, 2025. DOI: 10.2196/66931.

GOMMERS, J. et al. **Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study.** *The Lancet*, 2026. DOI: 10.1016/S0140-6736(25)02609-X.

GUR, D. et al. **Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system.** Journal of the National Cancer Institute, v. 96, n. 3, p. 185-190, 2004. DOI: 10.1093/jnci/djh067.

HA, S. M. et al. **Breast cancer detection with standalone AI versus radiologist interpretation of unilateral surveillance mammography after mastectomy.** Radiology, 2025. DOI: 10.1148/radiol.242955.

HOFVIND, S. et al. **Interval breast cancer rates and diagnostic performance in a population-based mammography screening program.** Radiology, v. 287, n. 3, p. 691-699, 2018.

HOVDA, T. et al. **Retrospective evaluation of a CE-marked AI system, including 1,017,208 mammography screening examinations.** European Radiology, 2025. DOI: 10.1007/s00330-025-11521-4.

INSTITUTO NACIONAL DE CÂNCER (Brasil). **Estimativa 2023: incidência de câncer no Brasil.** Rio de Janeiro: INCA, 2023. Disponível em: <https://www.inca.gov.br/publicacoes/livros/estimativa-2023-incidencia-de-cancer-no-brasil>. Acesso em: 5 jun. 2026.

KIM, H. et al. **Effect of artificial intelligence-based computer-aided diagnosis on the screening outcomes of digital mammography: a matched cohort study.** European Radiology, 2022. DOI: 10.1007/s00330-023-09692-z.

LÅNG, K. et al. **Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis.** The Lancet Oncology, 2023. DOI: 10.1016/S1470-2045(23)00298-X.

15

LARSEN, M. et al. **Performance of an artificial intelligence system for breast cancer detection on screening mammograms from BreastScreen Norway.** Radiology: Artificial Intelligence, 2024. DOI: 10.1148/ryai.230375.

LAURITZEN, A. D. et al. **An artificial intelligence-based mammography screening protocol for breast cancer: outcome and radiologist workload.** Radiology, 2022. DOI: 10.1148/radiol.210948.

LAURITZEN, A. D. et al. **Assessing breast cancer risk by combining AI for lesion detection and mammographic texture.** Radiology, 2023. DOI: 10.1148/radiol.230227.

LEE, S. E. et al. **AI-CAD for diagnostic mammography: comparison to radiologists according to different indications.** European Radiology, 2025. DOI: 10.1007/s00330-025-12232-6.

MCKINNEY, S. M. et al. **International evaluation of an AI system for breast cancer screening.** Nature, v. 577, n. 7788, p. 89-94, 2020. DOI: 10.1038/s41586-019-1799-6.

MIGOWSKI, A. et al. **Diretrizes para detecção precoce do câncer de mama no Brasil.** Revista Brasileira de Cancerologia, v. 64, n. 1, p. 13-32, 2018.

NELSON, H. D. et al. **Harms of breast cancer screening: systematic review.** JAMA, v. 314, n. 15, p. 1615-1634, 2016.

NG, A. Y. et al. **Artificial intelligence as supporting reader in breast screening: a novel workflow to preserve quality and reduce workload.** Journal of Breast Imaging, 2023. DOI: 10.1093/jbi/wbado10.

OBERIJE, C. J. G. et al. **Assessing artificial intelligence in breast screening with stratified results on 306,839 mammograms across geographic regions, age, breast density and ethnicity: A Retrospective Investigation Evaluating Screening (ARIES) study.** BMJ Health & Care Informatics, 2025. DOI: 10.1136/bmjhci-2024-101318.

OEFFINGER, K. C. et al. **Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society.** JAMA, v. 314, n. 15, p. 1599-1614, 2015. DOI: 10.1001/jama.2015.12783.

PESAPANE, F.; CODARI, M.; SQUILLACI, E. **Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine.** European Radiology Experimental, v. 2, n. 1, p. 35, 2018. DOI: 10.1186/s41747-018-0061-6.

PLIMPTON, S. R. et al. **External validation of a commercial artificial intelligence algorithm on a diverse population for detection of false negative breast cancers.** Journal of Breast Imaging, 2024. DOI: 10.1093/jbi/wbae058.

RAMLI HAMID, M. T. et al. **Comparative analysis of diagnostic performance in mammography: a reader study on the impact of AI assistance.** PLOS ONE, 2025. DOI: 10.1371/journal.pone.0322925.

RODRIGUEZ-RUIZ, A. et al. **Detection of breast cancer with mammography: effect of an artificial intelligence support system.** Radiology, v. 290, n. 2, p. 305-314, 2019. DOI: 10.1148/radiol.2018181371.

TAPLIN, S. H. et al. **Accuracy of screening mammography using single versus independent double interpretation.** AJR American Journal of Roentgenology, v. 191, n. 3, p. 754-762, 2008.

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

VAN WINKEL, S. L. et al. **AI as an independent second reader in detection of clinically relevant breast cancers within a population-based screening programme in the Netherlands: a retrospective cohort study.** The Lancet Digital Health, 2025. DOI: 10.1016/j.landig.2025.100882.

WORLD HEALTH ORGANIZATION. **Breast cancer.** Geneva: WHO, 2024. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>. Acesso em: 5 jun. 2026.

YALA, A. et al. **A deep learning mammography-based model for improved breast cancer risk prediction.** Radiology, v. 292, n. 1, p. 60-66, 2019. DOI: 10.1148/radiol.2019182716.

YI, C. et al. **The added value of an artificial intelligence system in assisting radiologists on indeterminate BI-RADS o mammograms.** European Radiology, 2022. DOI: 10.1007/s00330-021-08275-0.

YOEN, H.; CHANG, J. M. **Artificial intelligence improves detection of supplemental screening ultrasound-detected breast cancers in mammography.** Journal of Breast Cancer, 2023. DOI: 10.4048/jbc.2023.26.e39.