REVISTA TÓPICOS
ISSN 2965-6672 | Qualis A2

INTELIGÊNCIA ARTIFICIAL NA MEDICINA: BENEFÍCIOS, LIMITAÇÕES E DESAFIOS ÉTICOS PARA A PRÁTICA CLÍNICA

ARTIFICIAL INTELLIGENCE IN MEDICINE: BENEFITS, LIMITATIONS AND ETHICAL CHALLENGES FOR CLINICAL PRACTICE

Ciências da Saúde • 05/11/2026

REGISTRO DOI: 10.70773/revistatopicos/793768043

Clara Luquetti Paiva[1]

Alexandre Fleury Curado de Faria[2]

Ana Clara Machado Bariani[3]

Brunna Layla Santos de Andrade[4]

Daniel Gonçalves da Silva[5]

Felipe Schmaltz Zalaf[6]

Gabriel Alcino Souza Ferreira[7]

Gabriela de Paula Vaz Candido Gomes[8]

Gustavo Carlos de Alvarenga[9]

Gustavo Teixeira de Souza[10]

Iasmin Silva Bomtempo[11]

João Pedro Gonçalves Souza[12]

João Vitor Lelis Faria[13]

José Nicolas Andraos Filho[14]

Lucas Henrique Melo Guimarães[15]

Luis Fellipe Diniz Fernandes[16]

Mariana Bandeira Junqueira[17]

Matheus Lúcio Fernandes Gomes[18]

Pedro Henrique Miranda de Mendonça[19]

Ruan Carlos Vicente Barbosa dos Santos[20]

Sarah Elias Landim[21]

Vitor Davi Ferreira[22]

RESUMO

A inteligência artificial vem sendo incorporada à medicina em velocidade crescente, com aplicações que vão da interpretação de exames de imagem ao apoio à decisão clínica e à comunicação com pacientes por meio de modelos de linguagem de grande escala. Este artigo analisa os benefícios, as limitações e os desafios éticos do uso da inteligência artificial na prática clínica. Trata-se de uma revisão integrativa da literatura, realizada nas bases PubMed/MEDLINE, SciELO, LILACS e Scopus, com seleção de estudos publicados entre 2016 e 2025. A amostra final reuniu 12 estudos, entre ensaios clínicos randomizados, estudos de validação, metanálises e estudos transversais. Os resultados mostram desempenho diagnóstico comparável ao de especialistas em tarefas bem delimitadas, aumento da detecção de câncer no rastreamento mamográfico com redução da carga de trabalho e maior detecção de adenomas na colonoscopia. Por outro lado, evidenciam queda de desempenho em validação externa, escassez de ensaios clínicos robustos, viés algorítmico com impacto sobre grupos vulneráveis, viés de automação entre profissionais e limitações dos modelos de linguagem, como respostas incorretas apresentadas com aparente segurança. Os desafios éticos concentram-se na responsabilidade profissional, na transparência, na privacidade de dados, na equidade e na preservação da autonomia do paciente. Conclui-se que a inteligência artificial deve ser compreendida como ferramenta de apoio, cuja incorporação segura depende de validação clínica rigorosa, supervisão humana, governança institucional e formação médica adequada.

Palavras-chave: Inteligência artificial; Aprendizado de máquina; Tomada de decisão clínica; Ética médica; Viés algorítmico.

ABSTRACT

Artificial intelligence is being incorporated into medicine at an increasing pace, with applications ranging from the interpretation of imaging studies to clinical decision support and patient communication through large language models. This article analyzes the benefits, limitations and ethical challenges of using artificial intelligence in clinical practice. It is an integrative literature review carried out in the PubMed/MEDLINE, SciELO, LILACS and Scopus databases, selecting studies published between 2016 and 2025. The final sample comprised 12 studies, including randomized clinical trials, validation studies, meta-analyses and cross-sectional studies. The results show diagnostic performance comparable to that of specialists in well-defined tasks, increased cancer detection in mammography screening with reduced workload, and higher adenoma detection in colonoscopy. On the other hand, they reveal performance loss in external validation, a shortage of robust clinical trials, algorithmic bias affecting vulnerable groups, automation bias among professionals and limitations of language models, such as incorrect answers delivered with apparent confidence. The ethical challenges center on professional accountability, transparency, data privacy, equity and the preservation of patient autonomy. It is concluded that artificial intelligence should be understood as a support tool whose safe adoption depends on rigorous clinical validation, human oversight, institutional governance and adequate medical training.

Keywords: Artificial intelligence; Machine learning; Clinical decision-making; Medical ethics; Algorithmic bias.

1. INTRODUÇÃO

Poucas transformações tecnológicas chegaram à medicina com a velocidade e a amplitude da inteligência artificial (IA). Em pouco mais de uma década, algoritmos de aprendizado de máquina passaram de experimentos acadêmicos a ferramentas aprovadas por agências reguladoras para ler retinografias, triar exames de imagem, prever deterioração clínica e auxiliar na redação de documentos médicos. A popularização dos modelos de linguagem de grande escala, a partir de 2022, ampliou ainda mais esse alcance, colocando sistemas capazes de dialogar em linguagem natural ao alcance de médicos e pacientes (Rajpurkar et al., 2022).

O entusiasmo com essas tecnologias tem razões concretas. Os sistemas de saúde enfrentam envelhecimento populacional, aumento das doenças crônicas, escassez de especialistas em determinadas regiões e sobrecarga de trabalho dos profissionais. Nesse contexto, a promessa de diagnósticos mais rápidos e precisos, de redução de tarefas repetitivas e de ampliação do acesso a cuidados especializados é bastante atraente. Topol (2019) descreveu esse movimento como a convergência entre a inteligência humana e a artificial, na qual as máquinas assumiriam tarefas de reconhecimento de padrões e liberariam o médico para aquilo que é essencialmente humano na relação de cuidado.

A experiência acumulada nos últimos anos, contudo, mostrou que o caminho entre o bom desempenho em laboratório e o benefício real ao paciente é mais longo do que se imaginava. Algoritmos com excelente acurácia em bases de dados retrospectivas tiveram resultados decepcionantes quando aplicados em hospitais diferentes daqueles em que foram desenvolvidos (Wong et al., 2021), e ferramentas amplamente utilizadas revelaram vieses capazes de aprofundar desigualdades em saúde (Obermeyer et al., 2019). Ao

mesmo tempo, os modelos de linguagem trouxeram novos riscos, como a produção de informações falsas com aparência de verdade e a reprodução de concepções médicas ultrapassadas (Omiye et al., 2023).

Essas questões ganharam relevância normativa. A Organização Mundial da Saúde publicou orientações sobre ética e governança da IA em saúde (World Health Organization, 2021; 2024), a União Europeia aprovou um regulamento específico para sistemas de IA (União Europeia, 2024) e, no Brasil, o Conselho Federal de Medicina editou, em 2026, resolução que disciplina o uso dessas ferramentas na prática médica (Conselho Federal de Medicina, 2026), enquanto tramita no Congresso Nacional o projeto de lei que pretende estabelecer o marco legal da IA no país (Brasil, 2023).

Diante desse cenário, formula-se a seguinte questão norteadora: quais são os benefícios, as limitações e os desafios éticos do uso da inteligência artificial na prática clínica, segundo a literatura científica recente? A justificativa deste estudo reside na necessidade de que médicos e estudantes de medicina compreendam criticamente essas ferramentas, já presentes no cotidiano assistencial, para utilizá-las de forma segura, reconhecendo tanto o seu potencial quanto os seus limites.

O objetivo geral deste trabalho é analisar as evidências científicas sobre os benefícios, as limitações e os desafios éticos da IA na prática clínica. Como objetivos específicos, busca-se descrever as principais aplicações clínicas com evidência de desempenho; identificar as limitações metodológicas e práticas relatadas na literatura; e discutir os desafios éticos e as respostas regulatórias, com ênfase no contexto brasileiro.

1.1. Delimitação do Problema e Pertinência Clínica

A incorporação de inteligência artificial aos serviços de saúde exige distinguir a disponibilidade tecnológica de sua utilidade assistencial. Um sistema pode produzir classificações convincentes em um conjunto de imagens e, ainda assim, não oferecer benefício quando incorporado à rotina. O problema científico desta revisão não se resume, portanto, a perguntar se um algoritmo acerta determinada resposta: é necessário identificar em quais tarefas, populações e condições de uso a ferramenta produz informação confiável e como essa informação interfere na decisão clínica. Essa distinção organiza a interpretação dos estudos selecionados (Liu et al., 2019; Plana et al., 2022).

A relevância prática desse recorte torna-se evidente quando se contrastam tecnologias destinadas a tarefas restritas e ferramentas generalistas. Um algoritmo voltado à detecção de retinopatia diabética opera com entrada, objetivo e encaminhamento relativamente definidos, enquanto um modelo de linguagem pode receber solicitações clínicas heterogêneas, formuladas sem padronização. Consequentemente, as exigências de validação, monitoramento e supervisão não são idênticas. O caráter aberto das interações em linguagem natural amplia o número de situações em que uma resposta aparentemente apropriada pode estar incorreta ou incompleta (Abràmoff et al., 2018; World Health Organization, 2024).

A avaliação da IA também precisa reconhecer que a assistência médica não consiste em uma sequência isolada de classificações. A interpretação de um exame depende da história clínica, das condições de coleta, da prevalência da doença, das possibilidades de

confirmação e das consequências de cada decisão. Mesmo uma medida elevada de sensibilidade ou especificidade não esclarece, isoladamente, se haverá menos atrasos diagnósticos, menor utilização desnecessária de recursos ou melhores desfechos para pacientes. É por isso que os estudos de desempenho técnico e os estudos de implementação devem ser examinados como etapas complementares, e não como evidências equivalentes (Rajpurkar et al., 2022; Topol, 2019).

1.2. Objetivos Analíticos e Organização da Revisão

A presente discussão organiza a literatura em três planos interdependentes. O primeiro corresponde à demonstração de capacidade técnica em aplicações específicas, sobretudo em exames de imagem. O segundo examina o comportamento dos sistemas e dos profissionais quando a tecnologia é transferida para ambientes reais, incluindo perda de desempenho, alertas inadequados e confiança excessiva. O terceiro reúne as consequências éticas e institucionais do uso de ferramentas que participam da produção de informações clínicas. Essa organização permite interpretar resultados favoráveis sem desconsiderar as condições necessárias para sua utilização responsável.

Uma consequência metodológica dessa escolha é evitar a comparação direta de indicadores que medem fenômenos diferentes. A taxa de detecção de adenomas, por exemplo, expressa um resultado operacional do exame; a área sob a curva de um sistema de alerta descreve discriminação estatística; a preferência por respostas escritas reflete avaliação de comunicação; e o desempenho em questões médicas não equivale à segurança de uma conduta. Embora todos esses resultados sejam relevantes, não

podem ser agregados como se representassem uma medida única de eficácia da inteligência artificial na medicina (Hassan et al., 2021; Wong et al., 2021; Ayers et al., 2023; Singhal et al., 2023).

2. REFERENCIAL TEÓRICO

2.1. Conceitos Fundamentais

Inteligência artificial é o campo da computação dedicado ao desenvolvimento de sistemas capazes de executar tarefas que, tradicionalmente, exigiriam inteligência humana, como reconhecer padrões, fazer previsões e interpretar linguagem. Na medicina, a maior parte das aplicações atuais baseia-se no aprendizado de máquina, em que o algoritmo aprende a partir de grandes volumes de dados em vez de seguir regras programadas explicitamente. Uma de suas vertentes, o aprendizado profundo, utiliza redes neurais artificiais com múltiplas camadas e mostrou-se particularmente eficaz na análise de imagens, como radiografias, fotografias de retina, lâminas de patologia e lesões de pele (Topol, 2019; Rajpurkar et al., 2022).

Mais recentemente, ganharam destaque os modelos de linguagem de grande escala, treinados com enormes quantidades de texto e capazes de gerar respostas coerentes a perguntas formuladas em linguagem natural. Esses modelos alcançaram desempenho expressivo em questões de exames de licenciamento médico (Singhal et al., 2023) e vêm sendo testados em tarefas como apoio ao raciocínio diagnóstico, elaboração de resumos clínicos e resposta a mensagens de pacientes (Ayers et al., 2023; Goh et al., 2024). Diferentemente dos algoritmos tradicionais, que executam uma tarefa específica, esses sistemas têm uso geral, o que torna sua

avaliação e regulação mais complexas (World Health Organization, 2024).

2.2. Do Desempenho Técnico Ao Benefício Clínico

Um conceito central para compreender a literatura sobre IA em saúde é a distinção entre desempenho técnico e impacto clínico. A maioria dos estudos mede a acurácia do algoritmo, geralmente expressa por sensibilidade, especificidade ou área sob a curva ROC, em conjuntos de dados retrospectivos. Esse tipo de avaliação é necessário, mas insuficiente, pois não responde se o uso da ferramenta melhora desfechos relevantes para o paciente, como diagnóstico mais precoce, redução de mortalidade ou melhor qualidade de vida (Liu et al., 2019; Plana et al., 2022).

Entre a acurácia e o benefício existem etapas críticas: a validação externa, em populações e serviços diferentes daqueles usados no desenvolvimento; a avaliação prospectiva, em condições reais de uso; e os ensaios clínicos randomizados, que permitem estimar o efeito da intervenção sobre desfechos. Também é preciso considerar a interação entre o profissional e a máquina, já que o resultado final depende de como o médico interpreta e utiliza a recomendação do sistema (Dratsch et al., 2023).

2.3. Bioética e Regulação da Inteligência Artificial em Saúde

A Organização Mundial da Saúde estabeleceu seis princípios para orientar o uso ético da IA em saúde: proteger a autonomia humana; promover o bem-estar, a segurança e o interesse público; garantir transparência, explicabilidade e inteligibilidade; promover responsabilidade e prestação de contas; assegurar inclusão e equidade; e estimular uma IA responsiva e sustentável (World Health

Organization, 2021). Esses princípios dialogam diretamente com os pilares clássicos da bioética, como a beneficência, a não maleficência, a autonomia e a justiça.

No plano regulatório, o Regulamento Europeu de Inteligência Artificial adotou abordagem baseada em risco, classificando grande parte das aplicações em saúde como de alto risco e exigindo gestão de riscos, qualidade dos dados, documentação técnica, transparência e supervisão humana (União Europeia, 2024). No Brasil, além da Lei Geral de Proteção de Dados Pessoais, que considera os dados de saúde como dados sensíveis (Brasil, 2018), o Projeto de Lei nº 2.338/2023, aprovado pelo Senado Federal e em tramitação na Câmara dos Deputados, segue lógica semelhante (Brasil, 2023). No âmbito profissional, a Resolução CFM nº 2.454/2026 estabeleceu que a IA deve ser utilizada como ferramenta de apoio, que a decisão clínica permanece humana e que o médico continua integralmente responsável pelos atos praticados com auxílio dessas tecnologias (Conselho Federal de Medicina, 2026).

2.4. Tipologias de Aplicação e Diferentes Níveis de Autonomia

A expressão inteligência artificial reúne tecnologias com graus distintos de autonomia e exposição ao risco. Em um extremo encontram-se sistemas que identificam padrões e apresentam uma sugestão ao profissional; em outro, ferramentas concebidas para emitir resultados em tarefas clínicas estritamente delimitadas. A diferença não é apenas operacional: altera a forma de distribuir responsabilidades, definir critérios de encaminhamento e explicar ao paciente o papel desempenhado pelo software. O ensaio de Abràmoff et al. (2018) ilustra a importância de especificar a finalidade

de um sistema autônomo, enquanto Rajpurkar et al. (2022) discutem a diversidade de funções da IA na assistência.

Outra distinção essencial é a existente entre algoritmos treinados para reconhecer imagens e modelos que produzem linguagem. Os primeiros geralmente são avaliados em relação a uma condição-alvo e a um padrão de referência; os segundos podem gerar sínteses, hipóteses e orientações com múltiplas afirmações verificáveis. Para esses modelos, uma resposta linguisticamente adequada não garante precisão factual. A análise de Singhal et al. (2023) e as orientações da Organização Mundial da Saúde (2024) reforçam a necessidade de avaliar tanto o conteúdo clínico quanto os riscos de interpretações equivocadas.

A autonomia de um sistema deve ser entendida em relação à tarefa concreta, e não como atributo geral da tecnologia. Um programa pode ser autorizado a classificar imagens em determinado fluxo de rastreamento e, ao mesmo tempo, não possuir evidência para orientar tratamento ou atender situações fora do protocolo. A delimitação do uso pretendido impede que resultados positivos em um cenário sejam convertidos em autorização informal para outros. A governança clínica depende de documentar essas fronteiras e estabelecer mecanismos de revisão quando a aplicação real se afasta da finalidade originalmente avaliada.

2.5. Medidas de Desempenho e Interpretação Clínica

Sensibilidade e especificidade descrevem propriedades importantes de testes classificatórios, mas sua interpretação requer atenção ao contexto. A sensibilidade se relaciona à capacidade de identificar casos com a condição investigada, enquanto a especificidade

informa a capacidade de classificar corretamente os casos sem a condição. Entretanto, o significado de um resultado positivo para a pessoa atendida também depende da frequência da doença no grupo avaliado. Assim, a transposição de métricas de um estudo para outro serviço pode produzir expectativas inadequadas mesmo quando o algoritmo permanece tecnicamente inalterado.

A área sob a curva ROC resume a capacidade discriminatória ao longo de diferentes limiares, mas não define automaticamente o melhor ponto de operação para um hospital. A escolha do limiar modifica a proporção de falsos positivos e falsos negativos e deve considerar a gravidade dos erros, a capacidade de resposta da equipe e a disponibilidade de exames confirmatórios. O estudo de Wong et al. (2021) é particularmente instrutivo porque mostra que a avaliação de um alerta de sepse precisa incorporar não apenas sua discriminação, mas também a frequência de notificações e a proporção de eventos não identificados.

Indicadores de processo devem ser separados dos desfechos finais. A redução da carga de leitura mamográfica observada no estudo MASAI pode representar ganho operacional relevante, mas não substitui o acompanhamento de resultados clínicos em horizontes mais longos. Da mesma forma, o aumento da taxa de detecção de adenomas é um achado favorável para a qualidade da colonoscopia, sem que a metanálise permita afirmar automaticamente a magnitude de eventual redução futura da mortalidade. A interpretação responsável preserva a relação entre o indicador efetivamente medido e a conclusão formulada (Lång et al., 2023; Hassan et al., 2021).

2.6. Validação Externa, Mudança de Contexto e Monitoramento

Um sistema treinado em determinado conjunto de dados pode aprender relações associadas à população, aos equipamentos e às rotinas de registro daquele ambiente. Quando essas condições se modificam, o desempenho pode se alterar sem que o código tenha sido atualizado. A validação externa busca justamente investigar se o modelo mantém utilidade fora do contexto de desenvolvimento. Liu et al. (2019) e Plana et al. (2022) identificam limitações importantes na qualidade e na extensão dessas avaliações, tornando inadequada a presunção de generalização automática.

A transferência de tecnologia envolve diferenças que nem sempre aparecem nas características demográficas dos pacientes. A forma de registrar sintomas, a disponibilidade de exames, o fluxo de atendimento e os critérios de confirmação diagnóstica também afetam a qualidade das informações utilizadas. Um sistema de apoio pode produzir alertas tecnicamente consistentes com seus dados de entrada, mas pouco úteis para a equipe que deve agir sobre eles. A avaliação local precisa considerar o percurso completo entre coleta de dados, processamento, apresentação do resultado e decisão subsequente.

O monitoramento após a implantação constitui continuação lógica da validação. Mesmo quando o desempenho inicial é satisfatório, mudanças na população atendida ou no fluxo assistencial podem comprometer a utilidade da ferramenta. Instituições precisam definir responsáveis pela observação de falhas, critérios de reavaliação e procedimentos para suspender usos inadequados. A necessidade de acompanhamento permanente decorre das evidências de desempenho variável e dos princípios de segurança e prestação de contas apresentados pela Organização Mundial da

Saúde (2021) e pelo Regulamento Europeu de Inteligência Artificial (União Europeia, 2024).

2.7. Supervisão Humana Como Processo Verificável

A presença formal de um médico no fluxo não assegura, por si, julgamento independente. A sugestão de um sistema pode funcionar como âncora cognitiva e modificar a leitura de dados clínicos, inclusive quando o profissional possui experiência. O estudo experimental de Dratsch et al. (2023) demonstra que recomendações incorretas podem reduzir a acurácia de leitores experientes. Portanto, supervisão humana efetiva deve envolver oportunidade real de discordância, acesso aos dados relevantes e condições de trabalho que permitam avaliar criticamente a saída do algoritmo.

A organização da interface e do processo de decisão pode favorecer ou dificultar essa revisão crítica. Quando o sistema apresenta apenas uma conclusão categórica, sem contexto ou limites de aplicação, o profissional pode ser levado a aceitar o resultado por conveniência. Em contrapartida, a exigência de cotejar a recomendação com o quadro clínico e registrar divergências relevantes torna a intervenção mais auditável. Essas medidas constituem implicações práticas dos resultados de Dratsch et al. (2023) e Goh et al. (2024), e não garantias isoladas de eliminação do viés.

A capacitação deve incluir o reconhecimento de situações em que o sistema pode falhar. Isso compreende compreender a população em que foi validado, identificar informações ausentes, reconhecer contradições com achados clínicos e saber quando recorrer a avaliação especializada. A educação em IA não deve ser reduzida ao

domínio de comandos de interface; trata-se de formação para interpretar evidências e preservar a responsabilidade pelo cuidado. Esse enfoque é coerente com a perspectiva de integração entre inteligência humana e artificial discutida por Topol (2019).

2.8. Equidade, Dados e Responsabilidade Institucional

Os algoritmos reproduzem escolhas feitas ao definir objetivos, selecionar dados e estabelecer indicadores de sucesso. O estudo de Obermeyer et al. (2019) mostra que utilizar gasto em saúde como aproximação da necessidade assistencial pode perpetuar desigualdades, pois custos observados não representam necessariamente gravidade clínica equivalente entre grupos. Esse caso evidencia que a avaliação ética precisa começar antes do treinamento do modelo, examinando o significado das variáveis e as consequências distributivas das decisões automatizadas.

A qualidade de um banco de dados não pode ser reduzida à quantidade de registros. É necessário considerar quem foi incluído, quais grupos aparecem em proporções insuficientes, quais condições são medidas de modo desigual e como erros de registro podem repercutir na previsão. O estudo de Omiye et al. (2023) acrescenta outra dimensão: modelos de linguagem podem reproduzir pressupostos médicos discriminatórios presentes nos textos utilizados durante o treinamento. A avaliação de equidade deve, portanto, alcançar tanto os dados estruturados quanto as respostas textuais.

A responsabilidade institucional abrange aquisição, implantação, acompanhamento e resposta a incidentes. Mesmo quando a decisão final é atribuída ao médico, a instituição participa da escolha

da ferramenta, do desenho do fluxo e das condições de treinamento da equipe. Os princípios de governança da OMS (2021) e as obrigações de documentação e gestão de risco previstas no regulamento europeu oferecem referências para examinar essa distribuição de deveres. A proteção do paciente exige que responsabilidade profissional e responsabilidade organizacional sejam tratadas como dimensões complementares.

3. METODOLOGIA

Trata-se de uma revisão integrativa da literatura, método que permite reunir e sintetizar resultados de estudos com diferentes delineamentos, contribuindo para a compreensão abrangente de um fenômeno e para a prática baseada em evidências (Mendes; Silveira; Galvão, 2008; Souza; Silva; Carvalho, 2010). O estudo percorreu seis etapas: formulação da questão norteadora; definição dos critérios de inclusão e exclusão e busca na literatura; definição das informações a serem extraídas; avaliação dos estudos incluídos; interpretação dos resultados; e apresentação da síntese do conhecimento.

A questão norteadora foi construída com base na estratégia PICo, considerando como população pacientes e profissionais de saúde, como fenômeno de interesse o uso de ferramentas de IA e como contexto a prática clínica. A busca foi realizada nas bases PubMed/MEDLINE, Scientific Electronic Library Online (SciELO), Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS) e Scopus, com descritores dos vocabulários DeCS e MeSH combinados por operadores booleanos. A estratégia principal utilizada no PubMed foi: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "large language models") AND

("clinical decision-making" OR "diagnosis" OR "clinical practice") AND ("validation" OR "randomized controlled trial" OR "bias" OR "ethics"), adaptada às demais bases.

Foram incluídos estudos originais, ensaios clínicos, estudos de validação e revisões sistemáticas com metanálise publicados entre janeiro de 2016 e dezembro de 2025, em português, inglês ou espanhol, que avaliassem o desempenho, o impacto clínico, as limitações ou as implicações éticas de ferramentas de IA aplicadas à assistência ao paciente. Excluíram-se editoriais, cartas, resumos de eventos, estudos exclusivamente técnicos sem avaliação clínica e trabalhos voltados apenas à gestão administrativa ou financeira. Documentos normativos e diretrizes de organismos internacionais e nacionais foram utilizados como referencial teórico, sem compor a amostra analisada.

A seleção ocorreu em duas fases, com leitura de títulos e resumos e posterior leitura integral dos textos potencialmente elegíveis, e orientou-se pelas recomendações do PRISMA 2020 (Page et al., 2021). Ao final, 12 estudos compuseram a amostra. Os dados foram extraídos por meio de instrumento contendo autoria, ano, periódico, delineamento, área de aplicação e principais resultados, e os achados foram analisados de forma descritiva e organizados em eixos temáticos. Por se tratar de pesquisa baseada em dados secundários de domínio público, dispensou-se a apreciação por Comitê de Ética em Pesquisa.

3.1. Coerência Entre Questão, Seleção e Síntese

A opção pela revisão integrativa é pertinente porque a questão norteadora reúne desfechos e delineamentos heterogêneos. Ensaios

clínicos, estudos de validação, revisões sistemáticas e investigações sobre comportamento humano contribuem com evidências de natureza distinta. A síntese temática permite preservar essas diferenças e identificar convergências sem calcular uma estimativa combinada artificial. As etapas descritas no manuscrito seguem o propósito de reunir e interpretar achados diversos, conforme a abordagem metodológica apresentada por Mendes, Silveira e Galvão (2008) e Souza, Silva e Carvalho (2010).

A amostra de 12 estudos não deve ser confundida com a totalidade das publicações citadas. Documentos normativos, diretrizes e textos de fundamentação oferecem o contexto ético e conceitual, mas não integram a mesma categoria analítica dos estudos empíricos apresentados no Quadro 1. Manter essa distinção impede que recomendações institucionais sejam interpretadas como demonstrações experimentais de eficácia. Também torna mais clara a origem das afirmações sobre desempenho, risco e responsabilidades profissionais.

3.2. Heterogeneidade e Limites de Comparabilidade

Os estudos selecionados diferem em população, finalidade clínica, tipo de algoritmo e métrica de avaliação. O ensaio sobre retinopatia diabética avalia uma tarefa diagnóstica específica, enquanto a investigação sobre sepse analisa um sistema de alerta e o trabalho sobre comunicação compara respostas escritas. Essa heterogeneidade torna inadequado hierarquizar todas as aplicações por um único percentual de acerto. Na síntese, cada resultado deve ser interpretado segundo a pergunta do estudo, a referência utilizada e o ambiente de coleta dos dados.

As conclusões desta revisão também dependem da qualidade dos relatos disponíveis. Uma síntese descritiva pode identificar padrões consistentes, mas não elimina problemas de seleção, comparabilidade ou generalização dos estudos originais. Por isso, a discussão distingue achados diretamente observados, implicações plausíveis e recomendações de prudência. O uso do PRISMA 2020 como orientação de apresentação não transforma, por si só, a revisão em metanálise nem substitui a descrição transparente dos procedimentos de seleção (Page et al., 2021).

4. RESULTADOS E DISCUSSÃO

Os 12 estudos incluídos foram publicados entre 2018 e 2024 e compreendem dois ensaios clínicos randomizados, um ensaio clínico prospectivo de validação, duas revisões sistemáticas com metanálise, uma revisão sistemática sem metanálise, dois estudos de validação externa ou análise retrospectiva de algoritmos em uso, dois estudos experimentais com leitores ou avaliadores e dois estudos de avaliação de modelos de linguagem. As áreas de aplicação incluíram oftalmologia, radiologia, gastroenterologia, medicina intensiva, gestão de saúde populacional e atenção clínica geral. O Quadro 1 apresenta a síntese dos estudos analisados.

Quadro 1. Síntese dos estudos incluídos na revisão integrativa (n = 12)

Autor/ano	Periódico	Delineamento e aplicação	Principais resultados

Abràmoff et al. (2018)	npj Digital Medicine	Ensaio clínico prospectivo; 900 pacientes diabéticos na atenção primária (EUA)	Sistema autônomo para retinopatia diabética: sensibilidade de 87,2% e especificidade de 90,7%.
Liu et al. (2019)	The Lancet Digital Health	Revisão sistemática e metanálise; aprendizado profundo vs. profissionais em imagens médicas	Desempenho equivalente (sensibilidade 87,0% vs. 86,4%; especificidade 92,5% vs. 90,5%); poucos estudos com validação externa.
Obermeyer et al. (2019)	Science	Análise retrospectiva de algoritmo comercial de gestão populacional (EUA)	Uso do custo como indicador de necessidade gerou viés contra pacientes negros; correção do critério reduziu o viés.
Wong et al. (2021)	JAMA Internal Medicine	Validação externa de modelo proprietário de sepse; 38.455 hospitalizações (EUA)	AUC de 0,63; 67% dos casos de sepse não identificados; alertas em 18% dos internados.
Hassan et al. (2021)	Gastrointestinal Endoscopy	Revisão sistemática e metanálise de ensaios com detecção assistida em colonoscopia	Taxa de detecção de adenomas de 36,6% com IA vs. 25,2% sem IA (RR 1,44).
Plana et al. (2022)	JAMA Network Open	Revisão sistemática de ensaios randomizados com aprendizado de máquina	Poucos ensaios publicados, com risco de viés relevante, baixa diversidade e adesão incompleta ao CONSORT-AI.

Lång et al. (2023)	The Lancet Oncology	Ensaio clínico randomizado; 80.020 mulheres em rastreamento mamográfico (Suécia – MASAI)	Detecção de 6,1 vs. 5,1 cânceres/1.000; reconvocação semelhante; redução de 44% na carga de leitura.
Dratsch et al. (2023)	Radiology	Estudo experimental com leitores; sugestões BI-RADS de IA corretas e incorretas	Viés de automação: acurácia de experientes caiu de 82% para 45,5% com sugestões incorretas.
Ayers et al. (2023)	JAMA Internal Medicine	Estudo transversal; 195 perguntas de pacientes em fórum público	Respostas do chatbot preferidas em 79% dos casos; melhor qualidade e maior empatia que as dos médicos.
Singhal et al. (2023)	Nature	Avaliação de modelos de linguagem em bases de perguntas médicas	Flan-PaLM com 67,6% no MedQA; lacunas em factualidade, raciocínio e potencial de dano.
Omiye et al. (2023)	npj Digital Medicine	Avaliação de modelos de linguagem comerciais	Reprodução de conceitos de medicina baseada em raça já desacreditados.
Goh et al. (2024)	JAMA Network Open	Ensaio clínico randomizado; 50 médicos (EUA)	Acesso ao modelo de linguagem não melhorou o raciocínio diagnóstico (76% vs. 74%); modelo isolado superou o grupo controle.

Fonte: elaborado pela autora (2026). Legenda: IA – inteligência artificial; AUC – área sob a curva ROC; RR – risco relativo; BI-RADS – Breast Imaging Reporting and Data System.

A análise permitiu organizar os achados em quatro eixos: benefícios demonstrados; limitações técnicas e metodológicas; limitações na interação entre humanos e máquinas e nos modelos de linguagem; e desafios éticos e regulatórios.

4.1. Benefícios Demonstrados na Prática Clínica

As evidências mais sólidas de benefício concentram-se em tarefas de reconhecimento de padrões em imagens. Um marco nesse campo foi o ensaio clínico de Abràmoff et al. (2018), que avaliou prospectivamente, em consultórios de atenção primária, um sistema autônomo para detecção de retinopatia diabética. Com sensibilidade de 87,2% e especificidade de 90,7%, o sistema tornou-se o primeiro dispositivo de IA autônomo autorizado pela agência reguladora norte-americana para emitir um resultado diagnóstico sem interpretação de um especialista, o que tem implicações relevantes para a ampliação do rastreamento em locais com poucos oftalmologistas.

No rastreamento do câncer de mama, o ensaio randomizado sueco MASAI avaliou mais de 80 mil mulheres e mostrou que a leitura apoiada por IA detectou 6,1 cânceres por mil mulheres rastreadas, contra 5,1 na dupla leitura convencional, sem aumento relevante da taxa de reconvocação e com redução de 44% na carga de leitura dos radiologistas (Lång et al., 2023). Trata-se de um exemplo em que a tecnologia não substitui o especialista, mas reorganiza o fluxo de

trabalho, direcionando a atenção humana aos exames de maior risco.

Na colonoscopia, a metanálise de Hassan et al. (2021) demonstrou que os sistemas de detecção assistida por computador elevaram a taxa de detecção de adenomas de 25,2% para 36,6%, um aumento relativo de 44%. Como a taxa de detecção de adenomas é um indicador de qualidade associado à redução do câncer colorretal de intervalo, esse achado sugere potencial impacto em desfechos, embora ainda sejam necessários estudos de longo prazo.

De forma mais ampla, a metanálise de Liu et al. (2019), que analisou estudos comparando algoritmos de aprendizado profundo com profissionais de saúde na interpretação de imagens médicas, encontrou desempenho equivalente entre ambos, com sensibilidade de 87,0% e especificidade de 92,5% para os algoritmos, contra 86,4% e 90,5% para os profissionais. Os modelos de linguagem também mostraram potencial: o Flan-PaLM atingiu 67,6% de acurácia em questões no formato do exame de licenciamento médico norte-americano (Singhal et al., 2023), e respostas de um assistente conversacional a perguntas de pacientes foram preferidas por avaliadores em 79% dos casos, sendo consideradas de melhor qualidade e mais empáticas que as respostas de médicos (Ayers et al., 2023).

4.2. Limitações Técnicas e Metodológicas

Apesar desses resultados, a literatura revela limitações importantes. A própria metanálise de Liu et al. (2019) ressaltou que, entre milhares de estudos identificados, apenas uma pequena parcela realizou validação externa e comparou algoritmo e profissionais na mesma

amostra, e que a qualidade metodológica e o padrão de relato eram frequentemente insuficientes. Na mesma linha, a revisão de Plana et al. (2022) identificou um número reduzido de ensaios clínicos randomizados com intervenções de aprendizado de máquina, a maioria com risco de viés relevante, pouca diversidade geográfica e adesão incompleta às diretrizes de relato CONSORT-AI.

O problema da generalização ficou evidente no estudo de Wong et al. (2021), que avaliou externamente um modelo proprietário de predição de sepse amplamente implantado em hospitais norte-americanos. Em mais de 38 mil hospitalizações, o modelo apresentou área sob a curva de apenas 0,63, deixou de identificar 67% dos casos de sepse e gerou alertas para 18% de todos os pacientes internados, desempenho muito inferior ao divulgado pelo fabricante. Além de não beneficiar os pacientes, esse tipo de ferramenta pode contribuir para a fadiga de alarmes, levando os profissionais a ignorar alertas, inclusive os corretos.

Esses achados mostram que o desempenho de um algoritmo depende do contexto em que foi treinado. Diferenças na população atendida, nos equipamentos, nos protocolos de registro e na prevalência das doenças podem comprometer o funcionamento de uma ferramenta transferida para outro serviço. Para países como o Brasil, que frequentemente importam soluções desenvolvidas com dados de populações norte-americanas, europeias ou asiáticas, a validação local é uma exigência, e não uma formalidade.

4.3. Interação Humano-Máquina e Limitações dos Modelos de Linguagem

Um aspecto frequentemente subestimado é o comportamento do profissional diante da recomendação do sistema. No estudo de Dratsch et al. (2023), radiologistas avaliaram mamografias acompanhadas de sugestões de classificação BI-RADS supostamente geradas por IA. Quando a sugestão estava correta, a acurácia dos leitores inexperientes foi de quase 80%; quando estava incorreta, caiu para menos de 20%. Mesmo os radiologistas muito experientes passaram de 82% para 45,5% de acerto diante de sugestões erradas. Esse fenômeno, conhecido como viés de automação, demonstra que a presença de um profissional não garante, por si só, uma supervisão humana efetiva.

Os modelos de linguagem trazem desafios adicionais. No ensaio randomizado de Goh et al. (2024), médicos que tiveram acesso a um modelo de linguagem não apresentaram melhora significativa no raciocínio diagnóstico em relação aos que usaram apenas recursos convencionais, embora o modelo, isoladamente, tenha obtido desempenho superior ao do grupo controle. O resultado sugere que o simples acesso à ferramenta não basta e que é preciso aprender a integrá-la ao raciocínio clínico. Além disso, Omiye et al. (2023) mostraram que modelos de linguagem amplamente disponíveis reproduziram conceitos de medicina baseada em raça já desacreditados, como fórmulas de função renal e de capacidade pulmonar ajustadas por raça, evidenciando o risco de perpetuar erros e vieses presentes nos textos usados no treinamento.

Somam-se a isso as chamadas alucinações, isto é, a geração de informações falsas, como referências bibliográficas inexistentes ou condutas inadequadas, apresentadas com linguagem fluente e segura. Embora as respostas de assistentes conversacionais tenham sido bem avaliadas quanto à empatia (Ayers et al., 2023), esse

mesmo tom convincente pode induzir pacientes e profissionais a confiar em informações incorretas, razão pela qual a Organização Mundial da Saúde recomenda cautela e avaliação rigorosa antes do uso clínico desses sistemas (World Health Organization, 2024).

4.4. Desafios Éticos e Respostas Regulatórias

O estudo de Obermeyer et al. (2019) tornou-se referência no debate sobre equidade algorítmica. Os autores analisaram um algoritmo comercial utilizado para identificar pacientes que se beneficiariam de programas de cuidado intensivo e verificaram que, para um mesmo nível de risco estimado, pacientes negros apresentavam mais doenças crônicas que pacientes brancos. O viés decorria da escolha do custo com saúde como indicador de necessidade: como historicamente se gasta menos com pacientes negros, o algoritmo os classificava como mais saudáveis. A correção desse critério reduziu substancialmente a desigualdade, o que mostra que o viés algorítmico frequentemente nasce de escolhas humanas no desenho do sistema e pode ser corrigido.

O Quadro 2 sintetiza os principais desafios éticos identificados na literatura e as respostas normativas correspondentes.

Quadro 2. Principais desafios éticos da inteligência artificial na prática clínica e respostas normativas

Desafio	Descrição	Evidência na literatura	Resposta normativa

Equidade e viés algorítmico	Algoritmos treinados com dados enviesados podem prejudicar grupos vulneráveis	Obermeyer et al. (2019); Omiye et al. (2023)	Princípio de inclusão e equidade (OMS); qualidade e representatividade dos dados (UE)
Segurança e validação	Queda de desempenho fora do contexto de desenvolvimento	Wong et al. (2021); Plana et al. (2022)	Classificação de risco e avaliação prévia (UE; PL 2.338/2023; Resolução CFM 2.454/2026)
Supervisão humana	Viés de automação e confiança excessiva nas recomendações	Dratsch et al. (2023); Goh et al. (2024)	Decisão clínica sempre humana; vedação à delegação completa (Resolução CFM 2.454/2026)
Responsabilidade	Definição de quem responde pelos danos associados ao uso da IA	Rajpurkar et al. (2022)	Responsabilidade integral do médico; deveres institucionais de governança (Resolução CFM 2.454/2026)
Transparência e autonomia	Funcionamento opaco dos modelos e desconhecimento do paciente sobre seu uso	Topol (2019); OMS (2021)	Dever de informar o paciente e direito de recusa informada (Resolução CFM 2.454/2026)
Privacidade de dados	Tratamento de dados sensíveis e envio a servidores externos	OMS (2021; 2024)	Dados de saúde como dados sensíveis (LGPD); privacidade desde a concepção

Fonte: elaborado pela autora (2026). Legenda: OMS – Organização Mundial da Saúde; UE – Regulamento (UE) 2024/1689; LGPD – Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018).

A questão da responsabilidade é especialmente sensível. Quando um algoritmo contribui para um erro, é preciso definir as atribuições do desenvolvedor, da instituição que o adotou e do médico que o utilizou. A Resolução CFM nº 2.454/2026 optou por reafirmar a responsabilidade integral do médico, vedando a delegação completa da decisão clínica a sistemas autônomos, garantindo ao profissional o direito de recusar ferramentas que considere inadequadas e exigindo que o paciente seja informado quando a IA exercer papel relevante em seu diagnóstico ou tratamento, com direito de recusa informada (Conselho Federal de Medicina, 2026). Essa abordagem protege a autonomia do paciente, mas também impõe ao médico o dever de conhecer as limitações das ferramentas que utiliza, sob pena de responder pelo uso acrítico.

A privacidade de dados constitui outro eixo central. O desenvolvimento de algoritmos depende de grandes volumes de dados clínicos, e o uso de modelos de linguagem comerciais pode envolver o envio de informações de pacientes a servidores externos. No Brasil, a Lei Geral de Proteção de Dados classifica os dados de saúde como sensíveis e exige bases legais específicas para seu tratamento (Brasil, 2018). Na prática, isso significa que médicos não devem inserir dados identificáveis de pacientes em ferramentas que não ofereçam garantias adequadas de segurança e confidencialidade.

Por fim, a formação profissional emerge como condição para o uso seguro da tecnologia. Os resultados de Dratsch et al. (2023) e de Goh

et al. (2024) indicam que o benefício da IA depende da capacidade do médico de interpretar criticamente suas recomendações. Incorporar conteúdos de letramento em IA, estatística aplicada à avaliação de algoritmos e ética digital aos currículos de graduação e residência médica parece, portanto, tão importante quanto o desenvolvimento das próprias ferramentas (Rajpurkar et al., 2022; World Health Organization, 2021).

4.5. Comparação Transversal das Aplicações de Imagem

As aplicações em retinografia, mamografia e colonoscopia compartilham a vantagem de operar sobre dados visuais e objetivos clínicos relativamente delimitados. Contudo, os estudos não investigam exatamente o mesmo tipo de intervenção. Abràmoff et al. (2018) avaliaram um sistema autônomo em atenção primária; Lång et al. (2023) examinaram a leitura de mamografias apoiada por IA em um fluxo de rastreamento; Hassan et al. (2021) sintetizaram estudos de detecção assistida durante colonoscopia. A diferença entre autonomia diagnóstica, apoio à leitura e detecção em tempo real influencia a interpretação dos resultados.

O resultado favorável de uma dessas modalidades não valida automaticamente as demais. Na retinopatia diabética, a questão central é a classificação de um exame e o encaminhamento adequado. No rastreamento mamográfico, a tecnologia participa da distribuição do trabalho e da leitura dos casos. Na colonoscopia, o apoio atua durante um procedimento em que a resposta do profissional ocorre de maneira imediata. Cada cenário requer critérios próprios de segurança, treinamento e avaliação de efeitos adversos, além da medição do desfecho específico apresentado pelo estudo.

A metanálise de Liu et al. (2019) oferece uma perspectiva agregada sobre a capacidade de algoritmos de aprendizado profundo em imagens, mas a semelhança das métricas globais com as de profissionais não equivale a substituição clínica generalizada. O desempenho depende da seleção dos estudos e dos contextos de comparação. A presença de poucos estudos com validação externa adequada reforça que o valor principal da síntese é identificar potencial técnico e lacunas metodológicas, não estabelecer superioridade universal dos sistemas automatizados.

4.6. Desempenho de Algoritmos de Alerta e Riscos Operacionais

O estudo de Wong et al. (2021) representa um contraponto importante aos resultados favoráveis de sistemas de reconhecimento de imagem. A ferramenta de predição de sepse apresentou desempenho limitado quando avaliada externamente e produziu alertas para parcela expressiva dos pacientes internados. Esse achado permite discutir um risco operacional específico: o alerta não é um diagnóstico definitivo, mas uma demanda adicional de atenção e decisão para a equipe. Quando muitos avisos são pouco informativos, o sistema pode aumentar a carga cognitiva sem melhorar a identificação dos casos.

As consequências de falsos positivos e falsos negativos não são simétricas em todas as situações. Um falso negativo pode retardar a avaliação de um paciente que necessita de intervenção, enquanto um falso positivo pode mobilizar recursos e expor a pessoa a investigações desnecessárias. A decisão sobre limiares de alerta exige equilibrar esses riscos em função da finalidade clínica. O desempenho estatístico, portanto, precisa ser traduzido em consequências práticas para pacientes e profissionais, evitando que

a implantação seja justificada apenas por indicadores apresentados pelo fornecedor.

A experiência com o sistema de sepse evidencia a necessidade de avaliar tecnologias já incorporadas ao cotidiano. A ampla utilização de um produto não constitui evidência de efetividade. Da mesma forma, a existência de um algoritmo comercial não dispensa a comparação com o padrão assistencial vigente. O estudo mostra a relevância de avaliações independentes, transparentes e contextualizadas, especialmente quando o software influencia decisões urgentes e compete pela atenção de equipes submetidas a elevada demanda assistencial.

4.7. Modelos de Linguagem: Conhecimento, Comunicação e Decisão

Os estudos sobre modelos de linguagem analisam competências que devem permanecer conceitualmente separadas. Singhal et al. (2023) investigaram desempenho em perguntas médicas; Ayers et al. (2023) avaliaram a percepção de respostas a pacientes; Goh et al. (2024) examinaram o efeito do acesso à ferramenta sobre o raciocínio diagnóstico de médicos. Esses resultados não formam uma escala única de competência clínica. Um modelo pode responder corretamente a questões padronizadas e ainda produzir informações inadequadas em casos ambíguos ou incompletos.

A preferência dos avaliadores por respostas de chatbot em um fórum público não demonstra, por si, segurança para aconselhamento clínico individualizado. Uma resposta pode ser considerada clara e empática sem incorporar todos os dados necessários para orientar uma decisão. A comunicação clínica

envolve também reconhecer incertezas, obter informações adicionais, avaliar riscos e encaminhar o paciente quando apropriado. Assim, o resultado de Ayers et al. (2023) sustenta potencial comunicacional, mas não autoriza equiparar qualidade percebida de texto à qualidade global da assistência.

O ensaio de Goh et al. (2024) acrescenta um problema de implementação: a capacidade isolada do modelo não se converte necessariamente em melhor desempenho do profissional que o utiliza. A interação pode ser influenciada por confiança, tempo disponível, forma de formular perguntas e maneira de integrar respostas ao raciocínio clínico. Não se deve concluir que a ferramenta seja inútil, mas que o benefício depende de condições de uso que precisam ser investigadas. Esse resultado dialoga com o viés de automação demonstrado por Dratsch et al. (2023).

As respostas fabricadas e os pressupostos discriminatórios representam riscos distintos. A alucinação envolve a geração de conteúdo incorreto com aparência plausível; o viés pode envolver afirmações sistematicamente desfavoráveis a determinados grupos. Omiye et al. (2023) demonstraram a reprodução de conceitos inadequados de medicina baseada em raça, enquanto a OMS (2024) enfatiza a necessidade de governança dos grandes modelos multimodais. A verificação clínica deve abranger tanto a precisão factual quanto os pressupostos implícitos nas recomendações.

4.8. Viés de Automação e Desenho do Trabalho Clínico

O estudo de Dratsch et al. (2023) permite compreender que a interação humano-máquina pode alterar o desempenho mesmo quando o profissional dispõe de conhecimentos técnicos. A

sugestão incorreta funciona como referência inicial e pode deslocar a interpretação do leitor. O problema não se resolve apenas com a declaração de que a decisão final é humana. É necessário examinar como a recomendação é apresentada, quanto tempo existe para conferência e se a organização do trabalho favorece a revisão independente.

A literatura analisada sugere que a supervisão deve ser operacionalizada por procedimentos concretos. O profissional precisa conhecer a finalidade do sistema, identificar os elementos clínicos que podem contradizer sua saída e ter condições de registrar divergências. A instituição, por sua vez, deve permitir a análise de incidentes e evitar incentivos que transformem o uso da ferramenta em obrigação acrítica. Essas propostas decorrem da interpretação dos riscos observados nos estudos de Dratsch et al. (2023) e Wong et al. (2021), sem constituir evidência experimental de que uma medida isolada elimine falhas.

O treinamento também deve incluir casos em que a IA apresenta sugestões erradas. A formação restrita a demonstrações bem-sucedidas pode produzir expectativas exageradas e dificultar a identificação de erros. Exercícios de comparação entre avaliação independente e recomendação automatizada podem tornar visíveis os limites de confiança e os critérios de discordância. O objetivo não é estimular rejeição sistemática da tecnologia, mas preservar a capacidade de reconhecer quando a resposta do algoritmo é incompatível com os achados clínicos.

4.9. Justiça Distributiva e Vieses nos Sistemas de Saúde

A investigação de Obermeyer et al. (2019) demonstra que o viés não precisa decorrer de uma variável explicitamente discriminatória. Quando um sistema utiliza custos anteriores como substitutos da necessidade de cuidado, desigualdades históricas no acesso podem ser reproduzidas como se fossem diferenças clínicas legítimas. A escolha do desfecho a ser previsto constitui, portanto, decisão ética e metodológica central. Corrigir um algoritmo exige compreender o que ele otimiza e se essa meta corresponde ao interesse do paciente.

O estudo de Omiye et al. (2023) mostra que a questão da equidade alcança também modelos capazes de produzir explicações e recomendações em linguagem natural. A reprodução de conceitos ultrapassados pode ser particularmente prejudicial quando a resposta parece técnica e confiável. A identificação desses problemas requer análise de conteúdo e não apenas testes globais de acurácia. A avaliação de grupos potencialmente afetados deve considerar como diferentes erros podem produzir consequências desiguais.

Para a prática brasileira, a discussão sobre equidade tem implicação metodológica direta: resultados obtidos em uma população não dispensam exame de adequação ao contexto de uso. A diversidade de condições assistenciais e de perfis de pacientes torna relevante avaliar a representatividade dos dados, a acessibilidade do sistema e os efeitos de suas decisões. Essa é uma inferência de implementação fundamentada nos exemplos de viés e generalização discutidos na revisão, e não um resultado específico de estudo brasileiro incluído na amostra.

4.10. Privacidade, Transparência e Autonomia do Paciente

A utilização de IA na assistência frequentemente depende de informações sensíveis, como exames, histórico de saúde e registros de atendimento. A proteção desses dados exige que profissionais e instituições compreendam quais informações são processadas, com que finalidade e sob quais condições de segurança. A LGPD estabelece tratamento especial para dados pessoais sensíveis, enquanto as orientações da OMS (2021; 2024) situam a privacidade como componente da governança. O envio indiscriminado de dados identificáveis a ferramentas externas é incompatível com uma postura prudente de proteção do paciente.

A transparência deve ser entendida de maneira proporcional à participação da ferramenta na decisão. O paciente precisa receber informação compreensível quando a IA exerce papel relevante em seu cuidado, sem que isso signifique transferir a ele a responsabilidade técnica de avaliar o algoritmo. Explicar finalidade, limitações e supervisão humana contribui para o exercício da autonomia. O debate regulatório apresentado no manuscrito, incluindo a Resolução CFM nº 2.454/2026 e o Regulamento Europeu, reforça a importância de atribuições claras e mecanismos de prestação de contas.

Não basta disponibilizar uma política geral de privacidade se a rotina assistencial não define como os dados serão introduzidos e armazenados. A instituição precisa estabelecer orientações para seleção de ferramentas, permissões de acesso e resposta a incidentes. O profissional deve saber quando uma informação não deve ser compartilhada e como obter suporte institucional diante de dúvidas. Essas medidas representam desdobramentos operacionais dos princípios de proteção de dados e governança já presentes nas referências normativas do artigo.

4.11. Regulação e Governança Clínica em Diferentes Níveis

Os documentos normativos analisados abordam responsabilidades em planos diferentes. A Organização Mundial da Saúde formula princípios para orientar decisões de desenvolvimento e utilização; o Regulamento Europeu de Inteligência Artificial adota mecanismos de gestão de risco; a LGPD disciplina o tratamento de dados pessoais; e a regulamentação profissional brasileira se concentra nas condições de exercício da medicina. Essas camadas não são intercambiáveis. Sua leitura conjunta permite compreender que segurança clínica, privacidade e responsabilidade profissional exigem instrumentos complementares.

Uma instituição que adota tecnologia de IA precisa avaliar não apenas o contrato ou a alegação de desempenho, mas também a finalidade assistencial e a capacidade de acompanhar resultados. A documentação do uso pretendido, dos grupos de pacientes atendidos, das situações de contraindicação e dos responsáveis pelo monitoramento torna mais claro o percurso da decisão. Esse tipo de governança traduz em procedimentos os princípios de transparência e responsabilização defendidos pela OMS (2021) e a lógica de risco presente no Regulamento Europeu (União Europeia, 2024).

O debate normativo também precisa reconhecer os limites da supervisão individual. Mesmo profissionais diligentes podem ser afetados por interfaces inadequadas, alertas excessivos ou informações incompletas. A atribuição de responsabilidade deve considerar a cadeia de desenvolvimento, aquisição, implantação e utilização, sem apagar os deveres próprios do médico. Essa leitura é compatível com a discussão do manuscrito sobre responsabilidade

profissional e com os riscos empíricos de automação e validação externa identificados na amostra.

4.12. Formação Médica, Avaliação Crítica e Educação Continuada

O ensino de inteligência artificial na medicina deve priorizar competências de avaliação e uso seguro, e não apenas familiaridade com ferramentas. A leitura crítica de estudos de validação, o entendimento de métricas diagnósticas e o reconhecimento de vieses são habilidades diretamente relacionadas aos problemas observados nesta revisão. Liu et al. (2019) e Plana et al. (2022) apontam limitações metodológicas recorrentes, enquanto Dratsch et al. (2023) demonstram que a experiência clínica isolada não impede influência indevida de recomendações automatizadas.

Na graduação e na residência, situações simuladas podem ser utilizadas para discutir diferenças entre respostas plausíveis e decisões sustentadas por evidência. O aluno deve aprender a identificar dados ausentes, questionar pressupostos e documentar razões para concordar ou discordar de uma recomendação. A experiência com modelos de linguagem, analisada por Singhal et al. (2023) e Goh et al. (2024), mostra que bom desempenho em perguntas e disponibilidade de acesso não são equivalentes à integração segura ao raciocínio diagnóstico.

A educação continuada também precisa acompanhar mudanças nas ferramentas e nos fluxos assistenciais. Uma instituição pode substituir versões de software, modificar a população atendida ou reorganizar processos, alterando as condições em que a tecnologia foi inicialmente avaliada. A formação não deve ocorrer apenas na implantação: precisa incluir discussão de incidentes, revisão de

resultados e atualização dos critérios de uso. A continuidade desse processo está alinhada à governança responsiva recomendada pela Organização Mundial da Saúde (2021).

4.13. Implicações para o Sistema Único de Saúde

A adoção de IA em sistemas públicos deve ser analisada em relação às necessidades assistenciais e às desigualdades de acesso. Ferramentas de apoio ao rastreamento podem ter interesse particular em contextos com escassez de especialistas, mas o benefício depende da existência de encaminhamento, confirmação diagnóstica e continuidade do cuidado. O estudo de Abràmoff et al. (2018) oferece exemplo de aplicação delimitada em atenção primária; sua transferência para outro sistema requer avaliação própria de recursos, população e fluxos.

A promessa de eficiência não pode ser separada dos custos organizacionais. Sistemas de alerta exigem equipes capazes de responder, enquanto ferramentas de imagem dependem de infraestrutura e de integração ao processo assistencial. Quando a implantação amplia notificações sem utilidade proporcional, pode produzir sobrecarga. O caso estudado por Wong et al. (2021) evidencia a necessidade de avaliar a relação entre alertas, recursos disponíveis e consequências clínicas antes de presumir que a automatização representa economia ou melhoria.

No contexto brasileiro, é especialmente importante evitar que a adoção de tecnologias reforce diferenças entre serviços com alta capacidade de investimento e unidades com infraestrutura limitada. A avaliação institucional pode considerar segurança, adequação populacional, treinamento, privacidade e sustentabilidade do

acompanhamento. Essas dimensões derivam dos princípios de equidade e governança da OMS (2021) e dos problemas de generalização e viés identificados nos estudos incluídos, sem implicar que a revisão tenha mensurado diretamente impactos no SUS.

4.14. Critérios de Implementação e Avaliação Contínua

A análise da literatura permite propor uma sequência de perguntas para orientar a avaliação institucional, sem confundi-la com protocolo formalmente validado. Inicialmente, deve-se definir qual problema assistencial a ferramenta pretende resolver e qual desfecho será considerado relevante. Em seguida, é necessário verificar se os pacientes e dados do serviço se aproximam daqueles utilizados nos estudos de desenvolvimento e validação. Somente depois dessa delimitação é possível interpretar métricas técnicas em relação às decisões que o sistema pretende apoiar. Os contrastes entre Abràmoff et al. (2018) e Wong et al. (2021) mostram a importância de relacionar finalidade e contexto.

Uma segunda dimensão envolve as consequências de erros e a capacidade de resposta. A implantação de um sistema de rastreamento deve considerar os encaminhamentos produzidos por resultados positivos e negativos, enquanto um sistema de alerta requer análise de notificações, tempo de resposta e efeitos sobre a carga de trabalho. A avaliação deve identificar quem receberá a informação e quais ações serão esperadas. Sem esse encadeamento, o desempenho de classificação pode permanecer desconectado de qualquer melhora assistencial mensurável, problema coerente com as limitações metodológicas descritas por Plana et al. (2022).

A terceira dimensão é a interação humana. O serviço precisa examinar se a apresentação da recomendação favorece revisão independente ou confiança excessiva. O estudo de Dratsch et al. (2023) demonstra que sugestões incorretas podem modificar decisões mesmo entre leitores experientes, ao passo que Goh et al. (2024) não identificaram melhora significativa do raciocínio diagnóstico apenas com acesso ao modelo de linguagem. Esses achados indicam que a avaliação deve incluir o desempenho do conjunto profissional-ferramenta, e não somente a capacidade isolada do software.

A quarta dimensão refere-se à justiça e à proteção de dados. A definição dos objetivos do algoritmo deve ser examinada quanto à possibilidade de reproduzir desigualdades, como ocorreu com o uso de custos de assistência como indicador de necessidade no estudo de Obermeyer et al. (2019). Ao mesmo tempo, o tratamento de informações clínicas exige procedimentos compatíveis com a LGPD e com os princípios de governança da OMS (2021). A existência de dados disponíveis não significa que qualquer compartilhamento seja apropriado; finalidade e necessidade devem orientar sua utilização.

Finalmente, a avaliação precisa continuar após a implantação. Alterações de fluxo, população ou versão do sistema podem modificar o desempenho e gerar riscos não identificados no momento da aquisição. A instituição deve dispor de mecanismos para receber relatos de falhas, analisar eventos e rever critérios de utilização. Trata-se de consequência prática da preocupação com generalização presente em Liu et al. (2019) e Wong et al. (2021), associada à exigência de governança responsável descrita pela OMS (2021) e pelo regulamento europeu (União Europeia, 2024).

4.15. Delimitação das Conclusões e Prudência Interpretativa

Uma questão transversal é o risco de converter resultados específicos em afirmações genéricas sobre toda a inteligência artificial. Os estudos reunidos abordam tecnologias diferentes e foram conduzidos com desenhos e objetivos próprios. Por isso, um resultado de rastreamento mamográfico não determina a utilidade de um chatbot, assim como a falha de um alerta de sepse não invalida automaticamente algoritmos de imagem. A análise deve permanecer vinculada à tarefa, ao ambiente e ao indicador observado, preservando a diversidade das evidências.

A revisão também diferencia evidência de eficácia técnica, evidência de melhora de processo e evidência de benefício clínico. A precisão em uma classificação é importante, mas não substitui a demonstração de que a ferramenta modifica decisões de maneira favorável. O aumento de detecção ou a redução de trabalho podem representar avanços, mas seus efeitos sobre resultados posteriores exigem investigação específica. Essa distinção impede que desfechos intermediários sejam apresentados como comprovação de redução de mortalidade ou de outras consequências não mensuradas.

Ao considerar regulação e ética, convém distinguir princípios normativos de resultados empíricos. Uma norma pode estabelecer obrigações de supervisão e transparência, mas não comprova que a simples existência dessas obrigações elimine viés de automação, erros de generalização ou desigualdades. Da mesma forma, os estudos que demonstram falhas não definem sozinhos o modelo regulatório mais adequado. A contribuição desta síntese é aproximar

essas dimensões, explicitando seus limites e evitando inferências causais não sustentadas pelos trabalhos examinados.

4.16. Síntese Crítica dos Resultados e Lacunas Persistentes

O conjunto dos estudos sugere uma distribuição desigual da maturidade das aplicações. Há resultados clínicos promissores em tarefas delimitadas de rastreamento e detecção, mas permanecem dificuldades para demonstrar generalização, impacto sustentado e integração ao julgamento profissional. O contraste entre os estudos de imagem, a validação externa do alerta de sepse e os ensaios de interação com modelos de linguagem impede uma conclusão indiferenciada sobre eficácia da IA. A tecnologia precisa ser avaliada segundo sua função específica e seu contexto de uso.

Os resultados também mostram que riscos técnicos e éticos se relacionam. Um modelo que falha em nova população pode gerar dano por insuficiência de validação; um algoritmo que utiliza indicador inadequado pode produzir desigualdade; e uma recomendação incorreta pode ser aceita por viés de automação. Em todos esses casos, a qualidade do resultado depende de escolhas de desenvolvimento e de condições institucionais de utilização. A análise ética não é etapa posterior à demonstração técnica, mas parte da própria avaliação de segurança.

Entre as lacunas relevantes estão a necessidade de estudos prospectivos em cenários diversos, a avaliação de desfechos centrados no paciente, a compreensão das interações humano-máquina e a investigação de consequências distributivas. A revisão não fornece estimativas próprias para essas questões e não deve apresentá-las como resolvidas. Sua contribuição está em organizar

evidências e indicar critérios de leitura crítica que permitam distinguir potencial tecnológico, benefício demonstrado e risco ainda não suficientemente caracterizado.

5. CONSIDERAÇÕES FINAIS

Esta revisão integrativa permite responder à questão norteadora proposta: a inteligência artificial já oferece benefícios concretos à prática clínica, sobretudo na interpretação de imagens e na organização de fluxos de rastreamento, mas apresenta limitações relevantes de generalização, de validação clínica e de interação com os profissionais, além de impor desafios éticos relacionados à responsabilidade, à transparência, à privacidade, à equidade e à autonomia do paciente.

Os objetivos do estudo foram atingidos. Verificou-se que as aplicações com evidência mais robusta são aquelas voltadas a tarefas bem delimitadas e avaliadas prospectivamente, como a detecção de retinopatia diabética, o rastreamento mamográfico e a detecção de adenomas na colonoscopia. Em contrapartida, ferramentas implantadas sem validação externa adequada podem falhar em condições reais, e algoritmos aparentemente neutros podem reproduzir desigualdades presentes nos dados. Os modelos de linguagem, embora promissores, ainda carecem de evidências de benefício clínico e exigem uso cauteloso.

A inteligência artificial deve ser compreendida como ferramenta de apoio, e não como substituta do julgamento clínico. Sua incorporação segura depende de validação local, de ensaios clínicos com desfechos relevantes, de supervisão humana efetiva, de governança institucional e de formação médica adequada. No Brasil,

a Resolução CFM nº 2.454/2026 representa avanço importante, mas sua efetividade dependerá da capacitação dos profissionais e da aprovação de um marco legal que estabeleça regras claras para desenvolvedores e instituições.

Como limitações, destacam-se a rápida evolução do campo, que torna parte das evidências rapidamente desatualizada, e a escassez de estudos brasileiros. Recomenda-se a realização de pesquisas nacionais que avaliem o desempenho de algoritmos na população brasileira e o impacto de seu uso no Sistema Único de Saúde.

5.1. Desdobramentos para Pesquisa e Prática

A leitura integrada das evidências aponta que a decisão de incorporar IA deve ser orientada pela tarefa clínica e pelo benefício esperado, e não apenas pela novidade tecnológica. Aplicações com resultados favoráveis em rastreamento e interpretação de imagens demonstram potencial, mas sua utilidade depende de fluxos assistenciais adequados e avaliação no contexto de implantação. Para ferramentas de alerta e modelos de linguagem, a revisão destaca a importância de examinar falhas, incertezas e a maneira pela qual profissionais utilizam as recomendações.

A principal consequência prática é a necessidade de estabelecer critérios explícitos de validação, supervisão e governança. O uso responsável exige compreender a população avaliada, acompanhar desempenho, proteger dados sensíveis e preservar a autonomia do paciente. A formação médica deve incorporar essas competências como parte da avaliação crítica das tecnologias. Tais conclusões permanecem circunscritas aos estudos e documentos analisados,

sem pretensão de demonstrar benefícios não mensurados ou substituir a avaliação clínica individualizada.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRÀMOFF, M. D. et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. **npj Digital Medicine,** London, v. 1, 39, 2018. DOI: https://doi.org/10.1038/s41746-018-0040-6.

AYERS, J. W. et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. **JAMA Internal Medicine,** Chicago, v. 183, n. 6, p. 589-596, 2023. DOI: https://doi.org/10.1001/jamainternmed.2023.1838.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018.** Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 29 out. 2026.

BRASIL. Senado Federal. **Projeto de Lei nº 2.338, de 2023.** Dispõe sobre o uso da inteligência artificial. Brasília, DF: Senado Federal, 2023. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 29 out. 2026.

CONSELHO FEDERAL DE MEDICINA (Brasil). **Resolução CFM nº 2.454, de 11 de fevereiro de 2026.** Dispõe sobre o uso da inteligência artificial na medicina. Brasília, DF: CFM, 2026.

DRATSCH, T. et al. Automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance. **Radiology,** Oak Brook, v. 307, n. 4, e222176, 2023. DOI: https://doi.org/10.1148/radiol.222176.

GOH, E. et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. **JAMA Network Open,** Chicago, v. 7, n. 10, e2440969, 2024. DOI: https://doi.org/10.1001/jamanetworkopen.2024.40969.

HASSAN, C. et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. **Gastrointestinal Endoscopy,** New York, v. 93, n. 1, p. 77-85, 2021. DOI: https://doi.org/10.1016/j.gie.2020.06.059.

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 of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. **The Lancet Oncology,** London, v. 24, n. 8, p. 936-944, 2023. DOI: https://doi.org/10.1016/S1470-2045(23)00298-X.

LIU, X. et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. **The Lancet Digital Health,** London, v. 1, n. 6, p. e271-e297, 2019. DOI: https://doi.org/10.1016/S2589-7500(19)30123-2.

MENDES, K. D. S.; SILVEIRA, R. C. C. P.; GALVÃO, C. M. Revisão integrativa: método de pesquisa para a incorporação de evidências na saúde e na enfermagem. **Texto & Contexto Enfermagem,** Florianópolis, v. 17, n. 4, p. 758-764, 2008.

OBERMEYER, Z. et al. Dissecting racial bias in an algorithm used to manage the health of populations. **Science,** Washington, v. 366, n. 6464, p. 447-453, 2019. DOI: https://doi.org/10.1126/science.aax2342.

OMIYE, J. A. et al. Large language models propagate race-based medicine. **npj Digital Medicine,** London, v. 6, 195, 2023. DOI: https://doi.org/10.1038/s41746-023-00939-z.

PAGE, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **BMJ,** London, v. 372, n71, 2021. DOI: https://doi.org/10.1136/bmj.n71.

PLANA, D. et al. Randomized clinical trials of machine learning interventions in health care: a systematic review. **JAMA Network Open,** Chicago, v. 5, n. 9, e2233946, 2022. DOI: https://doi.org/10.1001/jamanetworkopen.2022.33946.

RAJPURKAR, P. et al. AI in health and medicine. **Nature Medicine,** New York, v. 28, n. 1, p. 31-38, 2022. DOI: https://doi.org/10.1038/s41591-021-01614-0.

SINGHAL, K. et al. Large language models encode clinical knowledge. **Nature,** London, v. 620, n. 7972, p. 172-180, 2023. DOI: https://doi.org/10.1038/s41586-023-06291-2.

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

TOPOL, E. J. High-performance medicine: the convergence of human and artificial intelligence. **Nature Medicine,** New York, v. 25, n. 1, p. 44-56, 2019. DOI: https://doi.org/10.1038/s41591-018-0300-7.

UNIÃO EUROPEIA. Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial. **Jornal Oficial da União Europeia,** Bruxelas, série L, 12 jul. 2024.

WONG, A. et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. **JAMA Internal Medicine,** Chicago, v. 181, n. 8, p. 1065-1070, 2021. DOI: https://doi.org/10.1001/jamainternmed.2021.2626.

WORLD HEALTH ORGANIZATION. **Ethics and governance of artificial intelligence for health:** guidance on large multi-modal models. Geneva: WHO, 2024.

WORLD HEALTH ORGANIZATION. **Ethics and governance of artificial intelligence for health:** WHO guidance. Geneva: WHO, 2021.

---

[1] Discente do Curso Superior de Medicina da Universidade de Rio Verde Campus Rio Verde. E-mail: acesse o artigo original para visualizar o e-mail

[2] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[3] Discente do Curso Superior de Medicina da Faculdade Morgana Potrich (FAMP) Campus Mineiros. E-mail: acesse o artigo original para visualizar o e-mail

[4] Discente do Curso Superior de Medicina da Faculdade Morgana Potrich (FAMP) Campus Mineiros. E-mail: acesse o artigo original para visualizar o e-mail

[5] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[6] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[7] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[8] Médica pela Faculdade de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[9] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[10] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[11] Discente do Curso Superior de Medicina da Faculdade Morgana Potrich (FAMP) Campus Mineiros. E-mail: acesse o artigo original para visualizar o e-mail

[12] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[13] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[14] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[15] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[16] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[17] Discente do Curso Superior de Medicina do Centro Universitário Atenas Campus Paracatu. E-mail: acesse o artigo original para visualizar o e-mail

[18] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[19] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[20] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail

[21] Discente do Curso Superior de Medicina da Zarns Campus Itumbiara. E-mail: acesse o artigo original para visualizar o e-mail

[22] Discente do Curso Superior de Medicina da Universidade Federal de Goiás Campus Colemar Natal e Silva. E-mail: acesse o artigo original para visualizar o e-mail