

A INTELIGÊNCIA ARTIFICIAL FUNCIONA IGUAL PARA TODOS? VIÉS ALGORÍTMICO E EQUIDADE NO DIAGNÓSTICO POR IMAGEM

DOES ARTIFICIAL INTELLIGENCE WORK THE SAME WAY FOR EVERYONE? ALGORITHMIC BIAS AND EQUITY IN DIAGNOSTIC IMAGING

Taís Silva Souza¹, Leonardo Lisboa Nobre¹, Yasmin Lisboa dos Santos¹, Lucas Jucá Dantas Borner¹, Victor Menezes Cardoso², Maria Laíze dos Santos³.

¹ Acadêmico do curso de Medicina do Centro Universitário de Maceió UNIMA-Afya.

² Professor do curso de Medicina do Centro Universitário de Maceió UNIMA-Afya.

³ Acadêmico do curso de Medicina do Centro Universitário de Maceió UNIMA-Afya. laize.xs@gmail.com

EIXO: Tecnologia e inovação

RESUMO

Introdução: A incorporação de inteligência artificial baseada em aprendizado profundo revolucionou a radiologia, contudo, a transição para a rotina clínica acendeu um alerta sobre a disparidade no desempenho entre diferentes indivíduos. **Objetivo:** Investigar as evidências de viés algorítmico nos sistemas aplicados ao diagnóstico por imagem e seus impactos na equidade diagnóstica e generalização dos resultados. **Método:** Revisão integrativa conduzida de forma independente nas bases PubMed, MedLine, Scopus, Web of Science e Biblioteca Virtual em Saúde, abrangendo artigos publicados de 2021 a 2026. **Resultados:** Os dados indicam falhas desproporcionais ao avaliar exames de minorias étnicas, mulheres, jovens e indivíduos de baixa renda, apresentando uma tendência crônica de subdiagnóstico ao classificá-los incorretamente como saudáveis. Os modelos utilizam atalhos ocultos extraídos dos pixels e perdem acurácia em virtude de variações técnicas de aparelhos e perfis antropométricos externos. **Conclusão:** Os sistemas perpetuam vieses estruturais, tornando urgente a consolidação de marcos regulatórios e auditorias contínuas de justiça antes de sua ampla implementação comercial.

Palavras-chave: Inteligência Artificial. Viés Algorítmico. Equidade em Saúde. Diagnóstico por Imagem. Radiologia.

ABSTRACT

Introduction: The incorporation of artificial intelligence based on deep learning has revolutionized radiology; however, the transition to clinical routine has raised a warning about performance disparities among different individuals. **Objective:** To investigate the evidence of algorithmic bias in systems applied to diagnostic imaging and its impacts on diagnostic equity and the generalization of results. **Method:** An integrative review conducted independently in the PubMed, MedLine, Scopus, Web of Science, and Virtual Health Library databases, spanning articles published from 2021 to 2026. **Results:** The data indicate disproportionate failures when evaluating exams of ethnic minorities, women, young people, and low-income individuals, showing a chronic underdiagnosis trend by incorrectly classifying them as healthy. The models use hidden shortcuts extracted from pixels and lose accuracy due to technical variations in devices and external anthropometric profiles. **Conclusion:** The systems perpetuate structural biases, making the consolidation of regulatory frameworks and continuous fairness audits urgent before their widespread commercial implementation.

Keywords: Artificial Intelligence. Algorithmic Bias. Health Equity. Diagnostic Imaging. Radiology.

INTRODUÇÃO

A incorporação de sistemas baseados em aprendizado profundo revolucionou a radiologia, entregando ferramentas automatizadas com acurácia equiparável à de especialistas na detecção de patologias complexas¹. Contudo, a transição desses modelos dos ambientes controlados de desenvolvimento para a rotina médica real acendeu um alerta sobre a disparidade no desempenho clínico entre diferentes indivíduos¹. Como as redes neurais aprendem estritamente a partir dos conjuntos de dados e relatórios que recebem em seu treinamento, a falta de diversidade demográfica nos bancos de dados originais consolida o erro algorítmico¹¹. Segundo Seyyed-Kalantari e colaboradores¹ e Koçak e colaboradores³, a omissão de marcadores de gênero, raça, etnia e nível socioeconômico nos repositórios radiológicos faz com que as ferramentas reproduzam padrões epidemiológicos restritos, elevando as taxas de subdiagnóstico em populações historicamente marginalizadas.

Essa assimetria técnica obriga a comunidade científica a questionar a composição dos grandes bancos de dados que sustentam as ferramentas preditivas atuais³. De acordo com Gichoya e colaboradores², as redes neurais conseguem identificar traços raciais ocultos nos pixels dos exames de imagem, uma capacidade que médicos radiologistas não possuem. Essa constatação revela uma lacuna crítica na literatura, evidenciando que a concentração de dados oriundos de centros médicos de excelência em países ricos gera distorções que prejudicam a equidade em saúde global². O problema se agrava nos testes de generalização externa, visto que softwares com ótimos resultados nos hospitais onde foram criados perdem eficácia ao serem transferidos para novas instituições¹. Marey e colaboradores⁶ corroboram esse limite ao demonstrar que a transição tecnológica para países de baixa e média renda falha frequentemente por causa de variações entre fabricantes de aparelhos, protocolos locais de aquisição e perfis antropométricos específicos da nova população.

A aplicação dessas ferramentas sem uma validação prévia no fluxo de triagem hospitalar pode retardar diagnósticos vitais e acentuar desigualdades já enraizadas, o que justifica a urgência em auditar e regulamentar esses sistemas antes de sua ampla adoção comercial¹. Conforme apontado pelo FUTURE-AI Consortium⁴, diretrizes internacionais de consenso são mandatórias para assegurar critérios mínimos de transparência, segurança e justiça nos softwares de imagem. Ademais, Aldhafeeri⁵ ressalta que falhas de generalização e diagnósticos incorretos em exames de mama e tórax trazem desdobramentos complexos de responsabilidade civil e legal para os hospitais, exigindo o monitoramento constante por comitês institucionais. O objetivo desta pesquisa é investigar quais são as evidências de viés algorítmico nos sistemas de inteligência artificial aplicados ao diagnóstico por imagem e

quais são seus impactos sobre a equidade, a segurança e a generalização dos resultados entre diferentes populações.

MÉTODO

Esta pesquisa foi conduzida como uma revisão integrativa da literatura para sintetizar resultados de investigações anteriores e prover uma compreensão robusta sobre os limites éticos e operacionais do aprendizado de máquina na medicina de forma sistemática. A estruturação do processo de busca e análise foi norteada pela definição de uma pergunta central clara, estabelecida para investigar quais são as evidências de viés algorítmico nos sistemas de inteligência artificial aplicados ao diagnóstico por imagem e quais são seus impactos sobre a equidade, a segurança e a generalização dos resultados entre diferentes populações.

A busca por evidências científicas foi conduzida de forma independente nas principais bases de dados e indexadores da área da saúde, incluindo o PubMed, MedLine, Scopus, Web of Science e a Biblioteca Virtual em Saúde (BVS). O recorte temporal para a seleção dos artigos foi delimitado estritamente entre janeiro de 2021 e agosto de 2026, com o objetivo de capturar o estado da arte e as diretrizes regulatórias internacionais mais recentes sobre o tema clínico-tecnológico. Para o levantamento bibliográfico, foram adotados descritores controlados extraídos do Medical Subject Headings (MeSH) e dos Descritores em Ciências da Saúde (DeCS), interligados pelos operadores booleanos AND e OR. Nos indexadores internacionais, utilizou-se o cruzamento dos termos em inglês consistindo em "Artificial Intelligence", "Deep Learning", "Algorithmic Bias", "Fairness", "Equity", "Diagnostic Imaging" e "Radiology". Para o indexador regional BVS, a estratégia foi traduzida para o português combinando "Inteligência Artificial", "Viés Algorítmico", "Equidade", "Justiça", "Diagnóstico por Imagem" e "Radiologia".

O refinamento do material coletado obedeceu a critérios rigorosos de inclusão e exclusão. Foram incluídos artigos originais, revisões de escopo ou sistemáticas, guias de consenso global e estudos observacionais de validação diagnóstica, publicados nos idiomas inglês, português ou espanhol, que avaliassem diretamente as assimetrias e os erros sistemáticos gerados por inteligência artificial em exames de radiologia de diferentes recortes demográficos. Em contrapartida, excluíram-se editoriais, cartas ao editor, resumos de congressos, relatórios puramente mercadológicos, produções duplicadas e estudos focados em

inteligência artificial médica geral sem aplicação direta ou validação em ambiente de diagnóstico por imagem.

O fluxo de seleção bibliográfica foi conduzido com base nos preceitos do protocolo PRISMA para garantir a reprodutibilidade do estudo. A busca inicial nas bases de dados resultou em um total de 245 artigos científicos identificados através da aplicação dos cruzamentos booleanos. Desse montante, 65 registros duplicados foram detectados e prontamente removidos por meio de gerenciadores bibliográficos, restando 180 publicações para a triagem primária. Realizou-se a avaliação inicial com base na leitura detalhada de títulos e resumos, resultando na exclusão de 148 artigos que não abordavam diretamente o impacto do viés na generalização e na segurança clínica de imagens. Na sequência, 32 artigos foram selecionados como elegíveis para a leitura aprofundada na íntegra. Após a análise do texto completo, foram excluídos mais 26 artigos por razões metodológicas, predominantemente devido ao foco estrito no desenvolvimento matemático de redes neurais sem discussão médica sobre equidade ou testes populacionais. Ao fim deste processo de filtragem crítica, restaram exatamente 6 artigos científicos que atenderam integralmente a todos os critérios exigidos, compondo o corpus analítico final utilizado para a sustentação deste trabalho.

RESULTADOS

Os dados obtidos demonstram que os sistemas baseados em aprendizado profundo apresentam falhas desproporcionais ao avaliar exames de minorias étnicas, mulheres, jovens e indivíduos de baixa renda. Há uma tendência crônica em classificar esses pacientes como saudáveis quando, na realidade, existem lesões ou patologias ativas. Essa disparidade nas taxas de falsos-negativos para doenças caracteriza um viés de subdiagnóstico severo que retarda intervenções terapêuticas cruciais. O desvio segue uma tendência fixa voltada a esses subgrupos, o que afasta a hipótese de mero ruído estatístico casual e configura uma falha sistemática que compromete o acesso igualitário a tratamentos vitais¹.

A origem dessa disparidade está associada à capacidade de os modelos matemáticos extraírem informações demográficas profundas diretamente dos pixels das imagens, algo que radiologistas humanos não conseguem identificar de forma visual. Os testes documentam uma acurácia elevada na classificação da identidade racial a partir de radiografias, mesmo após a aplicação de filtros de ruído ou cortes anatômicos no exame. Esse mecanismo faz com que o software utilize características étnicas como um atalho oculto na tomada de decisão

diagnóstica, espelhando preconceitos estruturais históricos presentes nos relatórios clínicos e nos prontuários médicos que alimentaram o treinamento das redes neurais^{2,3}.

Além disso, a transição dessas ferramentas para a rotina hospitalar encontra limitações severas na falta de capacidade de generalização externa em cenários clínicos reais. Os dados apontam uma perda expressiva de eficácia na detecção de patologias em exames de mama e tórax quando os sistemas enfrentam variações de fabricantes de aparelhos, calibrações de radiação locais e perfis antropométricos de pacientes em países de baixa e média renda. Esse cenário de instabilidade técnica gera riscos graves de responsabilidade civil para as instituições de saúde, demandando a criação de comitês internos de governança e a adoção de diretrizes globais para auditorias de justiça antes e depois da implementação dessas tecnologias⁴⁻⁶.

CONCLUSÃO

Os resultados obtidos demonstram que os sistemas baseados em aprendizado profundo aplicados ao diagnóstico por imagem não operam de forma equitativa, perpetuando e ampliando vieses que elevam as taxas de subdiagnóstico em mulheres, negros, jovens e populações de baixa renda¹¹. O uso de características demográficas ocultas nos pixels como atalhos de decisão², somado à incapacidade de generalização externa em diferentes equipamentos e realidades socioeconômicas⁶, transforma uma tecnologia de aprimoramento clínico em um vetor de aprofundamento de desigualdades em saúde.

Para que a inovação tecnológica na radiologia seja segura e sustentável, torna-se urgente afastar a visão puramente mercadológica e consolidar marcos de governança e auditorias contínuas de equidade diagnóstica^{4,5}. A verdadeira transformação global da medicina digital depende do compromisso regulatório de garantir que ferramentas preditivas entreguem a mesma acurácia e justiça terapêutica a qualquer indivíduo, independentemente de recortes demográficos, geográficos ou classes sociais.

REFERÊNCIAS

1. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served

patient populations. **Nat Med.** 2021 Dec;27(12):2176-2182.
doi: 10.1038/s41591-021-01595-0.

2. Gichoya JW, Banerjee I, Bhimireddy AR, Burns JL, Celi LA, Chen LC, et al. AI recognition of patient race in medical imaging: a modelling study. **Lancet Digit Health.** 2022 Jun;4(6):e406-e414. doi: 10.1016/S2589-7500(22)00063-2. PMID: 35568690; PMCID: PMC9650160.
3. Koçak B, Ponsiglione A, Stanzione A, Bluethgen C, Santinha J, Ugga L, et al. Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. **Diagn Interv Radiol.** 2025 Mar 3;31(2):75-88. doi: 10.4274/dir.2024.242854. PMID: 38953330; PMCID: PMC11880872.
4. Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al.; FUTURE-AI Consortium. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. **BMJ.** 2025 Feb 5;388:e081554. doi: 10.1136/bmj-2024-081554. PMID: 39909534; PMCID: PMC11795397.
5. Aldhafeeri FM. Governing Artificial Intelligence in Radiology: A Systematic Review of Ethical, Legal, and Regulatory Frameworks. **Diagnostics.** 2025 Sep 10;15(18):2300. doi: 10.3390/diagnostics15182300. PMID: 41008671; PMCID: PMC12468518.
6. Marey A, Ambrozaite O, Afifi A, Agarwal R, Chellappa R, Adeleke S, et al. A perspective on AI implementation in medical imaging in LMICs: challenges, priorities, and strategies. **Eur Radiol.** 2026 Apr;36(4):2591-2602. doi: 10.1007/s00330-025-12031-z. PMID: 41128758; PMCID: PMC13035529.