DISCRIMINAÇÃO ALGORÍTMICA EM SAÚDE: desafio contemporâneo aos direitos humanos, à cidadania sanitária e à justiça no cuidado

Codinome: LEITO DE PROCUSTO

Resumo: *Em sistemas de saúde digitalizados, a desigualdade pode anteceder a consulta: populações historicamente menos diagnosticadas produzem menos registros, aparecem menos nos dados e podem ser classificadas como menor risco por algoritmos de aparente neutralidade técnica. Este artigo analisa a discriminação algorítmica em saúde como desafio contemporâneo aos direitos humanos, tomando a inteligência artificial cardiovascular como campo sentinela. Realizou-se revisão integrativa de 99 estudos publicados entre 2020 e 2025. Os achados revelam baixa operacionalização da equidade: 90,9% omitiram determinantes sociais e 6,1% realizaram recalibração. Propõe-se a cidadania sanitária algorítmica para defender que representatividade, validação externa, análise por subgrupos, recalibração, transparência e supervisão humana são garantias de igualdade material no direito à saúde. Conclui-se que a IA só será compatível com os direitos humanos se ampliar o cuidado aos grupos invisibilizados pelos dados.*

Palavras-chave: *Direitos Humanos 1. Equidade Algorítmica 2. Inteligência artificial em saúde 3.*

1. Introdução

A violação do direito à saúde nem sempre começa com a recusa explícita de atendimento; em sistemas digitalizados, pode surgir quando um algoritmo atribui menor risco, menor urgência ou menor necessidade assistencial a quem a própria rede historicamente não aprendeu a enxergar. Uma pessoa negra, indígena, ribeirinha, pobre ou moradora de região remota pode chegar ao serviço com risco de saúde real, mas com poucos registros prévios, exames incompletos, trajetória assistencial fragmentada e menor acesso a especialistas. Se o modelo foi treinado majoritariamente com dados de populações mais diagnosticadas e mais presentes nos prontuários, essa pessoa pode ser classificada como menos grave não porque esteja protegida, mas porque grupos semelhantes ao seu foram sub-representados nos dados que sustentam o cálculo. Essa presunção algorítmica de baixo risco é especialmente grave porque transfere ao paciente invisibilizado o custo da própria invisibilidade: quem foi menos diagnosticado no passado pode ser novamente menos priorizado no presente. Nesse instante, a desigualdade deixa de ser apenas social ou assistencial: torna-se computacional.

Esse risco não é apenas teórico. Em estudo paradigmático, Obermeyer *et al.* (2019) demonstraram que um algoritmo utilizado para gestão populacional em saúde subpriorizava pacientes negros ao empregar gastos prévios em saúde como substituto de necessidade clínica. Como o menor gasto refletia, em parte, desigualdade histórica de acesso ao cuidado, o sistema interpretava menor utilização de serviços como menor necessidade assistencial. Revisões e estruturas recentes sobre viés algorítmico em saúde reforçam que esse mecanismo não é episódico: modelos treinados com dados desiguais podem reproduzir sub-representação, erros diferenciais e decisões menos seguras para grupos racial, social ou territorialmente vulnerabilizados (CHIN *et al*., 2023; NORORI *et al*., 2021).

A inteligência artificial, entendida como o conjunto de métodos computacionais capazes de reconhecer padrões, produzir inferências e apoiar decisões a partir de grandes volumes de dados, tem avançado rapidamente na área da saúde e, de modo particularmente relevante, na cardiologia, com aplicações na interpretação de eletrocardiogramas, na análise de imagens cardiovasculares, na predição de eventos adversos e na estratificação de risco (TOPOL, 2019; RAJPURKAR et al., 2022). Em regiões com carência estrutural de especialistas, essas ferramentas podem expressar potencial democratizador do acesso a serviços especializados. A crítica aqui desenvolvida não rejeita esse potencial; afirma, porém, que tecnologias capazes de ampliar acesso também podem aprofundar exclusões quando incorporadas sem representatividade, validação contextual, recalibração e supervisão humana. O problema, portanto, não está na inovação em si, mas na adoção acrítica de sistemas algorítmicos em redes assistenciais historicamente marcadas por desigualdades (CHAR; SHAH; MAGNUS, 2018; CHEN et al., 2021).

Essa forma de discriminação possui uma particularidade relevante. Sem hierarquizar gravidades em relação a outros usos discriminatórios da tecnologia, é necessário reconhecer que os modelos aplicados à saúde produzem um modo específico de lesão. Enquanto aplicações algorítmicas em segurança pública tendem a gerar violações mais imediatamente visíveis, como abordagem, identificação, restrição de liberdade ou persecução penal, ferramentas usadas no cuidado podem operar de modo mais silencioso, difuso e cumulativo. A classificação de baixo risco, a menor prioridade assistencial, o atraso no encaminhamento, a não indicação de exame ou a falsa segurança clínica dificilmente aparecem ao paciente como discriminação. O dano, portanto, não está apenas no erro técnico, mas na dificuldade de percebê-lo, demonstrá-lo e contestá-lo. Justamente por isso, as tecnologias algorítmicas em saúde exigem atenção própria no campo dos direitos humanos: seu potencial lesivo pode estar menos na violência explícita e mais na distribuição opaca e desigual das oportunidades de cuidado.

As doenças cardiovasculares constituem a principal causa de morte no mundo. Em 2022, estima-se que 19,8 milhões de pessoas morreram por causas cardiovasculares, aproximadamente 32% de todas as mortes globais, com concentração desproporcional em países de baixa e média renda, onde o acesso à detecção precoce e ao tratamento adequado permanece estruturalmente limitado (WORLD HEALTH ORGANIZATION, 2025a). Esse panorama não decorre apenas de fatores clínicos individuais. A literatura reconhece que determinantes sociais, como condição socioeconômica, escolaridade, território, raça, experiências de discriminação e acesso a serviços, influenciam incidência, tratamento e desfechos das doenças cardiovasculares (HAVRANEK *et al*., 2015). No Brasil, esse problema

assume densidade própria em razão das desigualdades regionais de acesso a especialistas, diagnóstico e registro, especialmente em territórios como a Amazônia, nos quais barreiras geográficas e assistenciais podem reduzir a produção de dados clínicos qualificados.

Este trabalho parte da premissa de que modelos de IA cardiovascular avaliados apenas por métricas globais de desempenho podem operar como um novo leito de Procusto da saúde digital: ajustam populações desiguais a uma medida média aparentemente neutra e, com isso, ocultam erros diferenciais contra grupos historicamente sub-representados nos dados. Por isso, analisa-se a discriminação algorítmica em saúde como desafio contemporâneo aos direitos humanos, investigando se a literatura recente sobre tecnologias cardiovasculares baseadas em dados incorpora salvaguardas de equidade, validade contextual, ajuste populacional e avaliação estratificada do desempenho. Embora o recorte empírico esteja concentrado na cardiologia, a área funciona como campo sentinela para examinar um risco mais amplo da saúde digital: a possibilidade de que modelos treinados com dados desiguais reproduzam, em diferentes especialidades, desigualdades prévias de acesso, diagnóstico e cuidado.

2. Direito à saúde e legitimidade jurídico-sanitária da inteligência artificial

A inteligência artificial em saúde deve ser analisada não apenas por critérios de desempenho técnico, mas também por critérios jurídico-sanitários. Quando sistemas algorítmicos participam da estratificação de risco, do encaminhamento diagnóstico, da priorização assistencial ou da alocação de recursos, passam a integrar a cadeia institucional de concretização do direito à saúde. Por isso, sua legitimidade depende de compatibilidade com igualdade, não discriminação, transparência, qualidade assistencial e responsabilização.

A Constituição Federal de 1988 estabelece, em seu artigo 196, que a saúde é direito de todos e dever do Estado, garantido por políticas voltadas à redução do risco de doença e ao acesso universal e igualitário; a Lei nº 8.080/1990 reafirma esse compromisso ao estruturar o Sistema Único de Saúde (SUS) com base na universalidade, integralidade e equidade. Em ambiente assistencial digitalizado, esses comandos alcançam também tecnologias que participam da decisão clínica, da estratificação de risco e da priorização do cuidado (BRASIL, 1988; BRASIL, 1990). Além disso, há fundamento no plano internacional: o Pacto Internacional sobre Direitos Econômicos, Sociais e Culturais reconhece o direito de toda pessoa ao mais alto padrão possível de saúde física e mental (ORGANIZAÇÃO DAS NAÇÕES UNIDAS, 1966), e o Comentário Geral nº 14 [1]densifica esse direito a partir das dimensões de

[1] Os Comentários Gerais são interpretações autorizadas dos tratados internacionais de direitos humanos emitidas pelos respectivos Comitês das Nações Unidas. Embora não tenham força vinculante formal equivalente ao tratado, possuem relevante valor interpretativo para a compreensão das obrigações estatais em matéria de direitos humanos.

disponibilidade, acessibilidade, aceitabilidade e qualidade (COMITÊ DE DIREITOS ECONÔMICOS, SOCIAIS E CULTURAIS DAS NAÇÕES UNIDAS, 2000). Na saúde digital, esse enquadramento torna insuficiente avaliar a inteligência artificial apenas por eficiência ou acurácia média, exigindo que seus efeitos sejam analisados à luz do acesso igualitário, da qualidade assistencial e da não discriminação.

Outrossim, a Organização Mundial da Saúde (OMS) tem advertido que tecnologias digitais e modelos de inteligência artificial em saúde só são compatíveis com a proteção de direitos quando incorporam transparência, supervisão humana, inclusão, equidade e mecanismos de responsabilização. Suas orientações sobre ética e governança da inteligência artificial em saúde alertam que modelos treinados com dados enviesados podem reproduzir desigualdades segundo raça, etnia, sexo, território e outras condições sociais, recomendando avaliação desagregada por subgrupos e auditoria após a implantação (WORLD HEALTH ORGANIZATION, 2021; WORLD HEALTH ORGANIZATION, 2025b).

3. Cidadania sanitária algorítmica: categoria jurídico-sanitária e garantias mínimas

A *cidadania sanitária algorítmica*[2] é proposta, neste artigo, como categoria jurídico-sanitária destinada a avaliar se tecnologias de inteligência artificial em saúde ampliam direitos ou apenas digitalizam desigualdades prévias de acesso, diagnóstico e cuidado. Por essa expressão, entende-se o conjunto de garantias necessárias para que sistemas algorítmicos incorporados à assistência não convertam desigualdades históricas de registro, representação e acesso em decisões clínicas automatizadas de menor qualidade para grupos vulnerabilizados.

Essa categoria difere da imparcialidade algorítmica em sentido estrito. Enquanto a imparcialidade costuma descrever uma propriedade técnica relacionada à distribuição de erros, acertos e oportunidades entre grupos, a cidadania sanitária algorítmica constitui categoria jurídica, sanitária e política: afirma que o exercício pleno do direito à saúde, em sistemas digitalizados, depende da garantia de que modelos computacionais não distribuam cuidado de forma desigual sob aparência de neutralidade técnica.

O conceito ancora-se na obrigação constitucional de acesso universal e igualitário ao SUS, nas dimensões de acessibilidade e qualidade do direito à saúde e na literatura sobre responsabilização algorítmica, que exige sistemas automatizados auditáveis, explicáveis e passíveis de contestação (DIAKOPOULOS, 2016). O conceito também dialoga com a crítica de Eubanks (2018) à automação de desigualdades por sistemas classificatórios. Na saúde, essa

[2] A expressão "cidadania sanitária algorítmica" é empregada neste artigo como categoria analítica própria para designar o conjunto de garantias necessárias à proteção do direito à saúde quando decisões clínicas, diagnósticas ou assistenciais passam a ser mediadas por sistemas de inteligência artificial.

automação tende a operar menos pela exclusão explícita e mais pela modulação silenciosa da prioridade, da oportunidade e da qualidade do cuidado. As garantias operacionais decorrentes dessa categoria são desenvolvidas na seção 6.

4. Método

Trata-se de revisão integrativa da literatura, delineamento adequado quando o objetivo envolve compreender a configuração de um campo de conhecimento, identificar lacunas e interpretar implicações clínicas, éticas e sociais a partir de estudos com diferentes desenhos metodológicos (WHITTEMORE; KNAFL, 2005; SOUZA; SILVA; CARVALHO, 2010).

A pergunta norteadora foi: em que medida os estudos sobre inteligência artificial cardiovascular publicados entre 2020 e 2025 incorporam equidade, validade contextual e avaliação de desempenho entre populações, e quais são suas implicações para o direito à saúde?

A busca foi realizada em janeiro de 2026 nas bases PubMed, Scopus, SciELO e LILACS, com estratégias adaptadas à sintaxe de cada base. Foram incluídos estudos originais publicados entre 2020 e 2025, nos idiomas português, inglês ou espanhol, que avaliassem aplicações de inteligência artificial em contextos cardiológicos diagnósticos, prognósticos ou de suporte à decisão, desde que abordassem ou permitissem inferir aspectos relacionados à equidade, desempenho diferencial entre subgrupos, validação externa, calibração ou transportabilidade. Foram identificados 1.405 registros nas bases consultadas: PubMed (597), Scopus (802), SciELO (2) e LILACS (4). Os registros foram organizados no Zotero e deduplicados no Rayyan, conforme Ouzzani *et al*. (2016). Após a remoção de 467 duplicatas, 938 registros seguiram para triagem por título e resumo, etapa em que 746 foram excluídos. Assim, 192 estudos foram avaliados por leitura integral; destes, 93 foram excluídos, resultando na inclusão final de 99 estudos originais. A seleção foi conduzida em planilha auditável, com critérios previamente definidos e resolução consensual das divergências entre avaliadores, conforme fluxo PRISMA 2020 (PAGE *et al*., 2021).

A extração foi orientada por domínios do CHARMS para revisões de modelos preditivos, adaptados para cinco categorias analíticas: contexto de aplicação clínica; subgrupos populacionais contemplados; forma de aparecimento da equidade; fragilidades metodológicas; e garantias de igualdade no cuidado (MOONS *et al*., 2014). A identificação dos fatores de equidade foi qualificada pela lente PROGRESS-Plus, que orienta a análise de desigualdades em saúde a partir de dimensões como local de residência, raça, ocupação, escolaridade, condição socioeconômica, sexo e outras vulnerabilidades sociais relevantes (O'NEILL *et al*., 2014). O PROBAST+AI foi utilizado como lente conceitual para identificação de fragilidades

metodológicas, sem aplicação formal de julgamento de risco de viés por domínio, opção compatível com o objetivo analítico desta revisão (MOONS *et al*., 2025).

A presença da equidade foi classificada em três níveis analíticos. Considerou-se "equidade explícita direta" quando o estudo tinha como objetivo central avaliar viés, justiça algorítmica, desempenho diferencial ou desigualdades entre grupos populacionais. A categoria "equidade explícita indireta" foi atribuída aos estudos em que a equidade não era o foco principal, mas aparecia por meio de validação externa, comparação entre populações, análise por subgrupos ou discussão sobre generalização do modelo. Por fim, classificou-se como "equidade inferível" o estudo que não empregava explicitamente os termos equidade, viés ou justiça algorítmica, mas permitia identificar lacunas metodológicas relevantes, como a ausência de salvaguardas de equidade na avaliação do desempenho do modelo.

A síntese dos achados foi predominantemente narrativa, sem metanálise, em razão da heterogeneidade dos estudos e do objetivo de interpretar lacunas metodológicas como potenciais riscos à igualdade material no direito à saúde.

5. Resultados

Na amostra final, houve concentração temporal recente: 53 estudos (53,5%) foram publicados em 2024 ou 2025, indicando aceleração do debate sobre equidade algorítmica em cardiologia. O corpus foi geograficamente diverso; o Brasil respondeu por oito estudos (8,1%), maior representação individual do Sul Global na amostra.

A equidade apareceu de forma heterogênea nos estudos analisados: foi explícita e direta em 18,2%, como objeto central da investigação; explícita indireta em 53,5%; e apenas inferível em 28,3%, a partir de lacunas metodológicas relevantes. Assim, 71,7% dos estudos reconheceram a equidade para além do nível meramente inferível, mas esse reconhecimento não se traduziu em plena operacionalização metodológica, pois a maioria omitiu determinantes sociais e raramente realizou recalibração para novas populações. O achado sugere um campo em transição: a equidade começa a aparecer como preocupação científica, mas ainda é tratada mais como critério tardio de avaliação do que como princípio estruturante da concepção, validação e implementação dos modelos.

Entre os subgrupos contemplados, o eixo geográfico ou institucional foi o mais frequente (65,7%), seguido por sexo ou gênero (40,4%), raça ou etnia (28,3%), vulnerabilidade social ou clínica (21,2%) e faixa etária (18,2%). Embora 90,9% dos estudos tenham contemplado ao menos um subgrupo demográfico, apenas 9,1% incorporaram determinantes sociais à modelagem. Essa diferença é central para o argumento deste artigo: reconhecer que

populações são distintas não equivale a incorporar, no desenho do modelo, as desigualdades que produzem risco, acesso tardio, diagnóstico incompleto e piores desfechos.

No campo das fragilidades, 94 estudos (95,0%) apresentaram ao menos uma. A omissão de determinantes sociais foi a mais frequente (90/99; 90,9%), seguida pela ausência de validação externa (21/99; 21,2%) e pela dependência de base local ou única (20/99; 20,2%). As garantias de igualdade no cuidado mais frequentes foram validação externa ou multicêntrica (76/99; 76,8%), estratificação formal por subgrupos (32/99; 32,3%), interpretabilidade algorítmica (23/99; 23,2%) e recalibração para nova população (6/99; 6,1%). A baixa frequência de recalibração é particularmente relevante, pois revela que muitos estudos ainda tratam o risco previsto como se pudesse ser transferido entre populações, instituições e territórios sem ajuste contextual. A síntese visual dos principais achados segundo os blocos analíticos da revisão está apresentada na Figura 1.

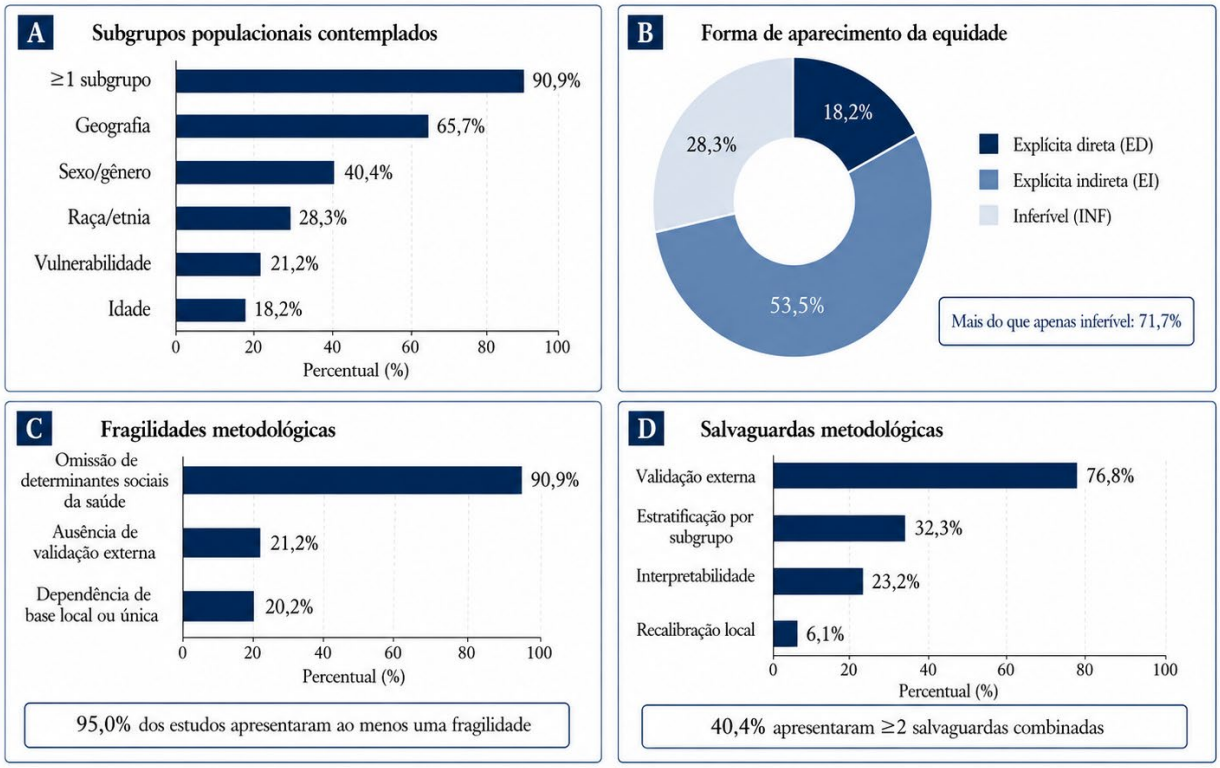

FIGURA 1 - Reconhecimento da equidade e lacunas de operacionalização metodológica na IA cardiovascular.
FONTE – Elaboração própria, com base na planilha-resumo codificada dos 99 estudos incluídos, 2026.
NOTA - ED = explícita direta; EI = explícita indireta; INF = inferível.

Quanto ao contexto de aplicação clínica, predominaram estudos voltados à predição de risco cardiovascular (50 estudos; 50,5%), seguidos por imagem cardiovascular e eletrocardiografia (28 estudos; 28,3%) e por cenários de cuidado intensivo, síndromes coronarianas agudas e insuficiência cardíaca aguda (21 estudos; 21,2%). Na predição de risco, a fragilidade central foi a baixa transportabilidade entre contextos epidemiológicos distintos,

isto é, a dificuldade de o modelo manter validade quando muda a população, o território ou o serviço de saúde. Em imagem cardiovascular e eletrocardiografia, destacou-se o viés de infraestrutura: diferenças no equipamento disponível, na qualidade do sinal e na capacidade técnica dos serviços podem produzir variações de desempenho que permanecem ocultas nos estudos de desenvolvimento. Já em terapia intensiva e urgência, emergiram riscos associados à superconfiança algorítmica e à comunicação insuficiente da incerteza do modelo. Em conjunto, esses achados mostram que a discriminação algorítmica em saúde pode assumir formas distintas conforme o contexto de uso.

Em conjunto, os achados do estudo revelam uma tensão central: a literatura cardiovascular começa a reconhecer diferenças populacionais, mas ainda não as converte de modo suficiente em arquitetura metodológica de proteção contra erro diferencial. A equidade, portanto, permanece mais visível como vocabulário de validação do que como arquitetura de proteção. É nessa distância entre reconhecimento e operacionalização que se localiza o risco jurídico-sanitário da discriminação algorítmica.

6. Discussão

6.1 Da fragilidade técnica à implicação em direitos humanos

Os achados desta revisão indicam que fragilidades metodológicas em inteligência artificial cardiovascular não são neutras. Quando modelos são desenvolvidos, testados ou implementados sem atenção ao contexto populacional, ao desempenho diferencial e às desigualdades que estruturam o acesso ao cuidado, podem influenciar quem será considerado de maior ou menor risco, quem será priorizado ou postergado e quem receberá cuidado oportuno ou tardio. A Tabela 1 sintetiza essa passagem da fragilidade técnica à implicação em direitos humanos.

TABELA 1
Da fragilidade técnica à implicação em direitos humanos

Fragilidade técnica	Consequência clínica possível	Implicação em direitos humanos	Garantia necessária
Ausência de validação externa	Perda de desempenho em nova população	Acesso desigual a cuidado seguro	Validação local e multicêntrica
Omissão de determinantes sociais	Estimativa incompleta do risco cardiovascular	Invisibilização de desigualdades estruturais	Incorporação ética de variáveis sociais
Falta de recalibração	Superestimação ou subestimação de risco	Distribuição injusta de prioridade assistencial	Recalibração populacional
Ausência de análise por subgrupos	Erros ocultos pela média global	Discriminação indireta	Avaliação estratificada
Baixa interpretabilidade	Dificuldade de auditoria clínica	Fragilização da transparência e responsabilização	Explicabilidade e supervisão humana
Falta de monitoramento pós-implantação	Degradação silenciosa do desempenho	Risco continuado para grupos vulnerabilizados	Auditoria contínua e revisão periódica

FONTE - Elaboração própria, 2026.

6.2 A desigualdade que os dados não veem: Amazônia, Norte, subdiagnóstico e invisibilidade algorítmica

No Brasil, as doenças cardiovasculares registraram 398.605 óbitos em 2024, segundo o Sistema de Informações sobre Mortalidade, constituindo o principal grupo de causas de morte no país (BRASIL, 2025). Esses óbitos, contudo, distribuem-se de forma desigual no território nacional. A Região Sudeste concentrou 182.582 óbitos, correspondentes a 45,8% do total nacional, enquanto a Região Norte registrou 23.405 óbitos, equivalentes a 5,9% (BRASIL, 2025). Em taxa por 100 mil habitantes, o Sudeste apresentou 215 mortes por 100 mil; o Sul, 206; o Nordeste, 190; e o Norte, 135 (BRASIL, 2025). Essa menor taxa registrada no Norte não pode ser lida automaticamente como menor risco cardiovascular: pode representar, justamente, a face estatística do subdiagnóstico, da menor disponibilidade de especialistas, das barreiras territoriais amazônicas e da menor qualidade do registro das causas de morte (BRASIL, 2025; BRASIL, 2026).

Essa assimetria ocorre em uma população racialmente diversa e territorialmente dispersa. O Censo 2022 revela que, dos 203,1 milhões de habitantes do país, 55,5% se declaram negros ou pardos, 43,5% brancos, 0,6% indígenas e 0,4% amarelos; além disso, o país possui 1,7 milhão de indígenas e 1,3 milhão de quilombolas (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2023). Parte expressiva desses grupos vive em territórios marcados por barreiras geográficas, sociais e institucionais de acesso ao cuidado especializado. Esses dados não indicam, isoladamente, maior risco cardiovascular, mas demonstram que modelos de inteligência artificial cardiovascular serão aplicados em uma população heterogênea, cujas condições de acesso ao diagnóstico, ao cuidado especializado e ao registro adequado de informações clínicas variam de forma expressiva entre grupos e regiões.

Os dados do Cadastro Nacional de Estabelecimentos de Saúde reforçam essa preocupação. Em março de 2026, o Brasil contava com 13.212 cardiologistas vinculados a estabelecimentos com atendimento ao SUS; desses, 48,3% estavam concentrados no Sudeste e apenas 4,3% no Norte, correspondendo a 7,53 cardiologistas por 100 mil habitantes no Sudeste contra 3,27 no Norte (BRASIL, 2026). Em territórios amazônicos, essa menor densidade de especialistas pode reduzir a produção de diagnósticos registrados e, consequentemente, a disponibilidade de dados clínicos qualificados para treinamento, validação e recalibração de modelos de inteligência artificial cardiovascular. Forma-se, assim, um ciclo potencialmente autoalimentado: menor acesso gera menor produção de dados; menor produção de dados favorece modelos menos sensíveis às populações que o sistema já subatende. Esses modelos,

nesse cenário, correm o risco de aprender não a realidade da saúde populacional brasileira, mas a desigualdade de registro produzida pelo próprio sistema.

Os resultados desta revisão confirmam que esse risco não é hipotético. Ignorar determinantes sociais em 90,9% dos estudos não é uma limitação metodológica menor: é uma falha que pode transformar desigualdade histórica em discriminação automatizada. Como demonstrado por Obermeyer et al. (2019), algoritmos em saúde reproduzem desigualdades quando utilizam variáveis marcadas por acesso desigual ao cuidado como substitutos de necessidade clínica real. A automação da desigualdade não exige intenção discriminatória; ocorre quando sistemas técnicos são aplicados a contextos socialmente desiguais sem que essa desigualdade seja reconhecida, medida e corrigida.

6.3 O argumento do benefício médio e sua refutação

Uma objeção recorrente sustenta que erros diferenciais seriam inevitáveis em qualquer modelo estatístico e que o benefício clínico agregado justificaria sua aceitação. Esse argumento é metodologicamente frágil porque estudos recentes indicam que modelos adaptados ao contexto de uso, submetidos a validação externa, recalibração e análise por subgrupos, podem melhorar a transportabilidade e reduzir perdas de desempenho em populações sub-representadas sem abandonar a acurácia global (LIU *et al*., 2025; ZINZUWADIA *et al*., 2024). O erro diferencial, portanto, não deve ser tratado como custo inevitável da inovação, mas como risco prevenível por melhores escolhas metodológicas.

O argumento também é eticamente insustentável. A Constituição Federal não garante saúde apenas à maioria, mas acesso universal e igualitário; do mesmo modo, o Comentário Geral nº 14 não legitima desigualdades de qualidade em nome de ganhos médios de cobertura. Uma tecnologia que melhora o desempenho médio enquanto aprofunda iniquidades não cumpre o mandato constitucional do SUS (BRASIL, 1988; DALLARI, 2003). Reconhecer o potencial emancipatório da inteligência artificial em saúde exige avaliá-la não apenas pelo que entrega em média, mas pelo que distribui e para quem.

6.4 Validade contextual e risco concreto de erro diferencial

A transportabilidade de um modelo não é garantida. Estudos de predição cardiovascular demonstram que modelos desenvolvidos em países de alta renda exigem recalibração quando transportados para outros sistemas de saúde com perfis epidemiológicos distintos (LIU *et al*., 2025). Investigações em eletrocardiografia e imagem indicam que diferenças de sexo, raça, equipamento e contexto institucional alteram o desempenho algorítmico de forma significativa (SAU *et al*., 2025; SUFIAN *et al*., 2024). No Brasil, modelos desenvolvidos com bases nacionais demonstram potencial de adaptação ao perfil dos pacientes, todavia evidenciam a

necessidade de validação externa antes de qualquer generalização assistencial (FERREIRA *et al*., 2025).

Uma falha de calibração pode significar subtratamento de condições cardiovasculares graves ou falsa segurança clínica em populações que já enfrentam barreiras históricas de acesso. A recalibração, realizada em apenas 6,1% dos estudos desta revisão, constitui uma das principais salvaguardas contra esse problema. Um algoritmo desenvolvido em hospital universitário de alta complexidade pode subestimar risco quando aplicado em unidade básica de saúde, município remoto ou contexto amazônico de acesso descontínuo, sem que esse erro apareça nas métricas globais. Transportar modelos sem validação e ajuste local significa presumir que populações diferentes podem ser medidas com o mesmo instrumento — uma presunção metodologicamente frágil e eticamente perigosa (SUBBASWAMY; SARIA, 2020; STEYERBERG; HARRELL, 2016).

6.5 Implicações para o sistema de saúde brasileiro

O Sistema Único de Saúde constitui a principal infraestrutura de cuidado no Brasil, e sua centralidade é reforçada pela Pesquisa Nacional de Saúde 2019, segundo a qual apenas 28,5% da população possuía plano de saúde (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2020). A incorporação de modelos algorítmicos cardiovasculares no SUS, portanto, não pode ser tratada apenas como inovação tecnológica: trata-se de decisão com impacto distributivo sobre acesso, qualidade, segurança e cidadania.

Como alertam Bahia e Chamas (2025), reverter a exclusão da América Latina das cadeias de valor da inteligência artificial em saúde exige que as demandas da saúde coletiva estejam refletidas na produção tecnológica e na distribuição dos resultados. Essa exigência dialoga com a Estratégia de Saúde Digital para o Brasil 2020-2028, que orienta a incorporação de tecnologias digitais no SUS; neste artigo, sustenta-se que tal incorporação só é legítima quando acompanhada de validação contextual, equidade e responsabilização (BRASIL, 2020). Para o SUS, isso significa que a adoção de modelos deve ser processo ativo de validação, adaptação e monitoramento, e não importação passiva de ferramentas desenvolvidas em outros contextos.

7. Garantias mínimas para uma inteligência artificial cardiovascular compatível com o direito à saúde

Dos achados desta revisão derivam garantias mínimas para uma inteligência artificial cardiovascular compatível com direitos humanos, cidadania sanitária algorítmica e direito à saúde.

a) *Representatividade dos dados*. Os modelos devem incluir diversidade de sexo, idade, raça ou cor, território, nível socioeconômico, perfil epidemiológico e contexto assistencial. A ausência de representatividade compromete a validade do modelo e permite que ele aprenda a ignorar quem já é ignorado pelo sistema de saúde (MARMOT *et al*., 2008; HAVRANEK *et al*., 2015).

b) *Validação externa*. Todo modelo deve ser testado na população e no cenário em que será utilizado. Sem validação externa, a população real passa a funcionar como campo de teste de uma tecnologia ainda não comprovada em seu próprio contexto.

c) *Recalibração populacional*. O modelo deve ser ajustado quando houver discrepância entre risco previsto e risco observado. Diante da baixa frequência de recalibração identificada nesta revisão, essa garantia é central para evitar superestimação ou subestimação sistemática de risco.

d) *Análise por subgrupos.* O desempenho deve ser reportado separadamente para grupos clinicamente e socialmente relevantes. Sem avaliação estratificada, erros contra mulheres, pessoas negras, indígenas, idosos, populações periféricas ou territórios remotos podem permanecer ocultos pela média global.

e) *Incorporação ética dos determinantes sociais da saúde*. Variáveis sociais devem ser consideradas de modo responsável, sem reforçar estigmas ou naturalizar desigualdades. Omissões também discriminam quando tratam como biologia individual aquilo que é produto de desigualdade histórica.

f) *Interpretabilidade e supervisão humana*. Profissionais de saúde devem compreender finalidade, dados utilizados, limites e contexto de validação do modelo, mantendo capacidade de questionar resultados e divergir do sistema quando a história clínica e social do paciente assim exigir (UNESCO, 2021; WORLD HEALTH ORGANIZATION, 2021).

g) *Monitoramento pós-implantação*. O desempenho deve ser acompanhado após a incorporação, com auditoria contínua e resultados desagregados por subgrupos. Sem monitoramento, a degradação de desempenho pode ocorrer silenciosamente e acumular riscos sobre grupos menos visíveis.

Essas garantias não são recomendações técnicas opcionais. No contexto do SUS, representam condições de legitimidade para que a digitalização da saúde cumpra o mandato constitucional de acesso universal e igualitário. A incorporação de inteligência artificial cardiovascular deve ser precedida por documentação das bases de treinamento, demonstração de validade no contexto brasileiro, avaliação estratificada de desempenho, plano de auditoria contínua e possibilidade de revisão humana. Sem esses requisitos, a inovação deixa de ser

instrumento de ampliação de acesso e passa a funcionar como tecnologia de distribuição opaca do risco.

8. Conclusão

Esta revisão analisou 99 estudos sobre inteligência artificial cardiovascular publicados entre 2020 e 2025 e identificou uma assimetria relevante entre reconhecimento e operacionalização da equidade. Embora parte da literatura já reconheça a importância de testar modelos em diferentes populações e examinar seu desempenho entre grupos, esse reconhecimento ainda não se converte em proteção metodológica suficiente contra erro diferencial: 90,9% dos estudos omitiram determinantes sociais e apenas 6,1% realizaram ajuste para novas populações.

No contexto brasileiro, esses achados assumem especial gravidade diante das desigualdades regionais de acesso a especialistas, diagnóstico e registro. Em territórios marcados por barreiras geográficas, sociais e institucionais, a inteligência artificial em saúde pode ampliar o acesso ao cuidado, mas também automatizar a invisibilidade de populações historicamente subdiagnosticadas. A pergunta decisiva, portanto, não é apenas se os algoritmos funcionam, mas para quem funcionam, em quais contextos, com quais dados e com quais consequências distributivas.

Ainda que o recorte empírico tenha se concentrado na cardiologia, a tese defendida ultrapassa esse campo. Sempre que sistemas algorítmicos participarem da triagem, da classificação de risco, da priorização diagnóstica ou da orientação terapêutica, haverá risco de discriminação indireta se populações vulnerabilizadas permanecerem ausentes dos dados, dos testes e do monitoramento. A principal contribuição deste trabalho é propor a cidadania sanitária algorítmica como categoria jurídico-sanitária para deslocar o debate da acurácia média para uma questão constitucional: quem é visto pelos dados, quem é protegido pelo modelo e quem permanece exposto a decisões automatizadas de menor qualidade.

Se a inteligência artificial for treinada apenas com os rastros dos grupos que o sistema já consegue diagnosticar, ela não corrigirá a desigualdade: irá automatizá-la. A inovação tecnológica em saúde só é compatível com os direitos humanos quando amplia a capacidade do Estado de cuidar daqueles que historicamente foram menos vistos, menos diagnosticados e menos protegidos. A acurácia média não pode se converter em um novo leito de Procusto, no qual populações desiguais só são reconhecidas quando cabem nos padrões dos grupos mais assistidos. Nenhuma pessoa deve ser invisível para o cuidado porque foi invisível para os dados.

Referências

BAHIA, B.; CHAMAS, C. **Inteligência artificial, saúde global e desigualdades.** Cadernos CRIS — Informe 09/2025. Rio de Janeiro: Centro de Estudos Estratégicos da Fundação Oswaldo Cruz, 2025.

BRASIL. **Constituição da República Federativa do Brasil de 1988.** Brasília, DF: Presidência da República, 1988.

BRASIL. **Lei nº 8.080, de 19 de setembro de 1990.** Dispõe sobre as condições para a promoção, proteção e recuperação da saúde, a organização e o funcionamento dos serviços correspondentes. Brasília, DF: Presidência da República, 1990.

BRASIL. Ministério da Saúde. **Estratégia de Saúde Digital para o Brasil 2020-2028**. Brasília, DF: Ministério da Saúde, 2020.

BRASIL. Ministério da Saúde. **Sistema de Informações sobre Mortalidade — SIM.** Óbitos por doenças do aparelho circulatório: Brasil, 2024. Brasília, DF: MS/SVSA/CGIAE, 2025.

BRASIL. Ministério da Saúde. **Cadastro Nacional dos Estabelecimentos de Saúde — CNES.** Recursos humanos: médicos cardiologistas que atendem no SUS, por região. Brasília, DF: Ministério da Saúde, 2026.

CHAR, D. S.; SHAH, N. H.; MAGNUS, D. Implementing machine learning in health care: addressing ethical challenges. **The New England Journal of Medicine,** Boston, v. 378, n. 11, p. 981-983, 2018.

CHEN, I. Y. *et al*. Ethical machine learning in healthcare. **Annual Review of Biomedical Data Science,** Palo Alto, v. 4, p. 123-144, 2021.

CHIN, M. H. et al. Guiding principles to address the impact of algorithm bias on racial and ethnic disparities in health and health care. **JAMA Network Open**, Chicago, v. 6, n. 12, e2345057, 2023.

COMITÊ DE DIREITOS ECONÔMICOS, SOCIAIS E CULTURAIS DAS NAÇÕES UNIDAS. **Comentário Geral nº 14:** o direito ao mais alto padrão possível de saúde. Genebra: Organização das Nações Unidas, 2000.

DALLARI, S. G. **Direito sanitário e saúde pública.** São Paulo: Ministério da Saúde, 2003. v. 1.

DIAKOPOULOS, N. Algorithmic accountability: journalistic investigation of computational power structures. **Digital Journalism,** Londres, v. 3, n. 3, p. 398-415, 2016.

EUBANKS, V. **Automating inequality: how high-tech tools profile, police, and punish the poor.** New York: St. Martin's Press, 2018.

FERREIRA, M. B. D. *et al.* Um novo escore de risco baseado em aprendizado de máquina em pacientes com insuficiência cardíaca aguda: o escore ML-HF. **Arquivos Brasileiros de Cardiologia,** São Paulo, v. 122, n. 11, e20250136, 2025.

HAVRANEK, E. P. *et al*. Social determinants of risk and outcomes for cardiovascular disease: a scientific statement from the American Heart Association. **Circulation,** Dallas, v. 132, n. 9, p. 873-898, 2015.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico 2022:** resultados do universo. Rio de Janeiro: IBGE, 2023.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional de Saúde 2019:** atenção à saúde, acesso e uso dos serviços de saúde. Rio de Janeiro: IBGE, 2020.

LIU, X. *et al*. A population-based recalibration method for updating survival neural networks models for cardiovascular risk prediction in United Kingdom and China. **BMC Medicine,** London, v. 23, n. 1, 142, 2025.

MARMOT, M. *et al*. Closing the gap in a generation: health equity through action on the social determinants of health. **The Lancet,** London, v. 372, n. 9650, p. 1661-1669, 2008.

MOONS, K. G. M. *et al*. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. **PLoS Medicine,** San Francisco, v. 11, n. 10, e1001744, 2014.

MOONS, K. G. M. *et al*. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. **BMJ,** London, v. 388, e082505, 2025.

NORORI, N. *et al*. Addressing bias in big data and AI for health care: a call for open science. **Patterns**, Cambridge, v. 2, n. 10, 100347, 2021.

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.

O'NEILL, J. *et al*. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. **Journal of Clinical Epidemiology,** Oxford, v. 67, n. 1, p. 56-64, 2014.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Pacto Internacional sobre Direitos Econômicos, Sociais e Culturais.** Nova York: ONU, 1966.

OUZZANI, M. *et al*. Rayyan: a web and mobile app for systematic reviews. **Systematic Reviews,** London, v. 5, n. 1, art. 210, 2016.

PAGE, M. J. *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **BMJ,** London, v. 372, n. 71, 2021.

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

SAU, A. *et al.* Artificial intelligence-enhanced electrocardiography for the identification of a sex-related cardiovascular risk continuum. **The Lancet Digital Health,** London, v. 7, n. 7, p. e495-e505, 2025.

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.

STEYERBERG, E. W.; HARRELL, F. E. Prediction models need appropriate internal, internal-external, and external validation. **Journal of Clinical Epidemiology,** Oxford, v. 69, p. 245-247, 2016.

SUBBASWAMY, A.; SARIA, S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. **Biostatistics,** Oxford, v. 21, n. 2, p. 345-352, 2020.

SUFIAN, M. A. *et al*. Mitigating algorithmic bias in AI-driven cardiovascular imaging for fairer diagnostics. **Diagnostics,** Basel, v. 14, n. 23, 2675, 2024.

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.

UNESCO. **Recommendation on the Ethics of Artificial Intelligence.** Paris: United Nations Educational, Scientific and Cultural Organization, 2021.

WHITTEMORE, R.; KNAFL, K. The integrative review: updated methodology. **Journal of Advanced Nursing,** Oxford, v. 52, n. 5, p. 546-553, 2005.

WORLD HEALTH ORGANIZATION. **Cardiovascular diseases (CVDs): fact sheet.** Geneva: World Health Organization, 31 jul. 2025a.

WORLD HEALTH ORGANIZATION. **Ethics and governance of artificial intelligence for health: guidance on large multi-modal models.** Geneva: World Health Organization, 2025b.

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

ZINZUWADIA, A. N. *et al*. Tailoring risk prediction models to local populations. **Circulation: Cardiovascular Quality and Outcomes,** Dallas, v. 17, n. 10, e011789, 2024.