Capítulo 4

Sesgos algorítmicos en aplicaciones de la IA en el sector salud: una revisión sistemática de la literatura

Yaidy Sughey Solís Valdez, José Alberto Hernández Aguilar, Laura Cruz Abarca

Resumen

La Inteligencia Artificial (IA) está transformando el sector salud, especialmente en la detección temprana de enfermedades y el diagnóstico médico; sin embargo, este avance enfrenta desafíos significativos derivados de los sesgos algorítmicos. Conjuntos de datos pequeños, desbalanceados o no representativos utilizados en el entrenamiento y validación de modelos pueden generar errores que afectan la equidad, la precisión y la confiabilidad de los resultados clínicos, además de plantear dilemas éticos y técnicos que requieren atención urgente. Esta investigación busca identificar y analizar de manera sistemática los principales desafíos asociados con los sesgos algorítmicos en aplicaciones de IA en la atención médica, con el propósito de aportar evidencia y sentar bases para futuras investigaciones integrales. Se empleó la metodología Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), ampliamente utilizada en revisiones sistemáticas en el ámbito de la salud. Los resultados derivados de su aplicación se presentan y discuten a lo largo del capítulo.

Palabras clave: Inteligencia Artificial; Sesgos; Aplicación Informática; Salud; metodología PRISMA.

Solís Valdez, Y. S., Hernández Aguilar, J. A., & Cruz Abarca, L. (2025). Sesgos algorítmicos en aplicaciones de la IA en el sector salud: una revisión sistemática de la literatura. En N. Roque Nieto, A. R. Pérez Mayo, P. Guerrero Sánchez, N. Betanzos Díaz, & C. Rodríguez Leana, (Coord). *Organizaciones, Salud y Bienestar: Perspectivas Transdisciplinares sobre lo Social, lo Tecnológico y lo Emocional.* (pp. 98-128). Religación Press. http://doi.org/10.46652/religacionpress.312.c721

Introducción

La IA (inteligencia artificial) está revolucionando el sector salud mediante la mejora en la realización de diagnósticos precisos, la detección en sus primeras etapas de enfermedades, y la optimización de procesos y tratamientos. Sin embargo, a medida que estas tecnologías avanzan, surgen desafíos críticos relacionados con los sistemas algorítmicos que soportan estas aplicaciones. Particularmente los sesgos originados en gran medida por conjuntos de datos no representativos o métodos de entrenamiento inadecuados, ya que no solo comprometen la equidad en los resultados, sino que también perpetúan desigualdades sociales, económicas y de acceso a los servicios médicos (Ali y Nikberg, 2024).

En esta investigación se busca analizar cómo los sistemas algorítmicos de aplicaciones basadas en IA pueden impactar en el acceso a los servicios de salud y en su calidad, particularmente en poblaciones vulnerables. Para ello se utilizará la metodología PRISMA, para identificar patrones en la literatura reciente que permitan categorizar los desafíos más recurrentes en las aplicaciones de IA en el sector salud, estableciendo bases robustas para la generación de estrategias que mitiguen los sesgos algorítmicos y fomenten la creación de sistemas más inclusivos y éticos.

Planteamiento del problema

A medida que la IA continúa transformando el sector salud, particularmente en áreas como la detección temprana y el diagnóstico médico. Se desea identificar los desafíos críticos relacionados con los sesgos algorítmicos en las aplicaciones de la IA en el sector salud. Estos sesgos no solo pueden comprometer la equidad y precisión de los resultados, sino que también representan barreras éticas y técnicas que requieren atención urgente. A pesar de la creciente adopción de la IA, existe una necesidad en la sistematización en la literatura que

identifique, categorice y analice los desafíos específicos relacionados con los sistemas algorítmicos en aplicaciones médicas. Por ello, esta revisión sistemática tiene como objetivo recopilar y sintetizar la evidencia existente que permita identificar los principales desafíos en este ámbito, sentando las bases para investigaciones futuras que aborden estos problemas de manera integral.

Justificación

La inteligencia artificial ha comenzado a implementarse en países desarrollados para incrementar la velocidad y exactitud en la detección de enfermedades, así como en el diagnóstico médico. Además, de acuerdo con Zeng et al. (2021), su aplicación en el sector salud permite mejorar la calidad de la atención provista, fomentar la investigación y el desarrollo de medicamentos, así como coadyuvar en la vigilancia de enfermedades y la respuesta a brotes infecciosos. Otra ventaja potencial de la IA es su capacidad para ofrecer a los pacientes un mejor control sobre su atención y facilitar el acceso a la salud en áreas rurales o de bajos recursos, donde el acceso a personal médico puede ser limitado.

No obstante, un reciente informe de la OMS (2021), alerta sobre los riesgos de confiar demasiado en las promesas de la IA en el sector salud, especialmente si esto implica descuidar inversiones y estrategias tradicionales que permitan lograr la cobertura universal. La OMS destaca que la implementación de la IA conlleva desafíos y riesgos, que incluyen la recolección y el uso ético de los datos, los sesgos algoritmos, las posibles amenazas a la ciberseguridad, así como a la integridad del paciente y el medio ambiente en el que se desenvuelve.

En el sector de salud, los algoritmos de IA que muestran sesgos suelen utilizar datos biométricos y genéticos, en parte debido a la naturaleza de los datos y a la forma en que son procesados. Por ejemplo, los algoritmos de reconocimiento biométrico, que se basan

en patrones como huellas dactilares, reconocimiento facial o análisis de voz, pueden reflejar sesgos cuando los datos de entrenamiento no incluyen una representación demográfica diversa. Esto puede llevar a errores en la identificación de personas de grupos minoritarios y a un decremento al acceso de servicios de salud para estos grupos.

En los algoritmos de diagnóstico por imágenes, que emplean aprendizaje profundo para identificar anomalías en estudios como radiografías y tomografías, los sesgos pueden derivarse de bases de datos que carecen de diversidad étnica y demográfica, lo que afecta la precisión en ciertos grupos poblacionales. Algo similar ocurre en los algoritmos de genética aplicada a la medicina personalizada, que pueden fallar en el diagnóstico de personas de etnias subrepresentadas debido a que la mayoría de las bases de datos genéticas contienen información de personas de ascendencia europea.

Asimismo, los algoritmos de predicción de riesgos de salud y asignación de recursos pueden reflejar sesgos socioeconómicos, favoreciendo a grupos con mayor acceso histórico a servicios de salud y con recursos económicos perpetuando desigualdades. En "El informe mundial sobre Inteligencia Artificial aplicada a la salud" la (OMS, 2021) advierte que el uso no regulado de la IA puede priorizar intereses comerciales o gubernamentales por encima de los derechos de los pacientes, y recalca la importancia de diseñar sistemas de IA que reflejen la diversidad socioeconómica y de salud.

Así mismo la OMS (2021), subraya la necesidad de capacitar a los profesionales de salud en competencias digitales y crear conciencia en las comunidades sobre la implicación de la IA en sus vidas, asegurando que estas herramientas no comprometen la autonomía y las decisiones tanto de los pacientes como de los proveedores de salud.

Los objetivos de esta investigación son: Identificar y analizar de manera sistemática los principales desafíos relacionados con los sistemas algorítmicos en aplicaciones médicas de Inteligencia

Artificial, con el fin de sintetizar la evidencia existente y así establecer bases para investigaciones futuras que discutan estas problemáticas de manera integral.

Para ello se establecen los siguientes objetivos específicos: 1) Realizar una revisión de la literatura existente del uso de la Inteligencia Artificial en la atención médica para identificar los sesgos algorítmicos reportados. 2) Categorizar y clasificar los desafíos éticos, técnicos y operativos asociados con los sistemas algorítmicos en aplicaciones médicas. 3) Recopilar y comparar los datos de los estudios revisados a través de gráficos y tablas, con el fin de identificar la frecuencia de los sesgos reportados y su impacto potencial en la equidad y precisión de los resultados médicos.

La pregunta de investigación que sustenta este trabajo es la siguiente: ¿De qué manera los sesgos algorítmicos en la inteligencia artificial perpetúan y amplifican desigualdades preexistentes en el área de la salud?

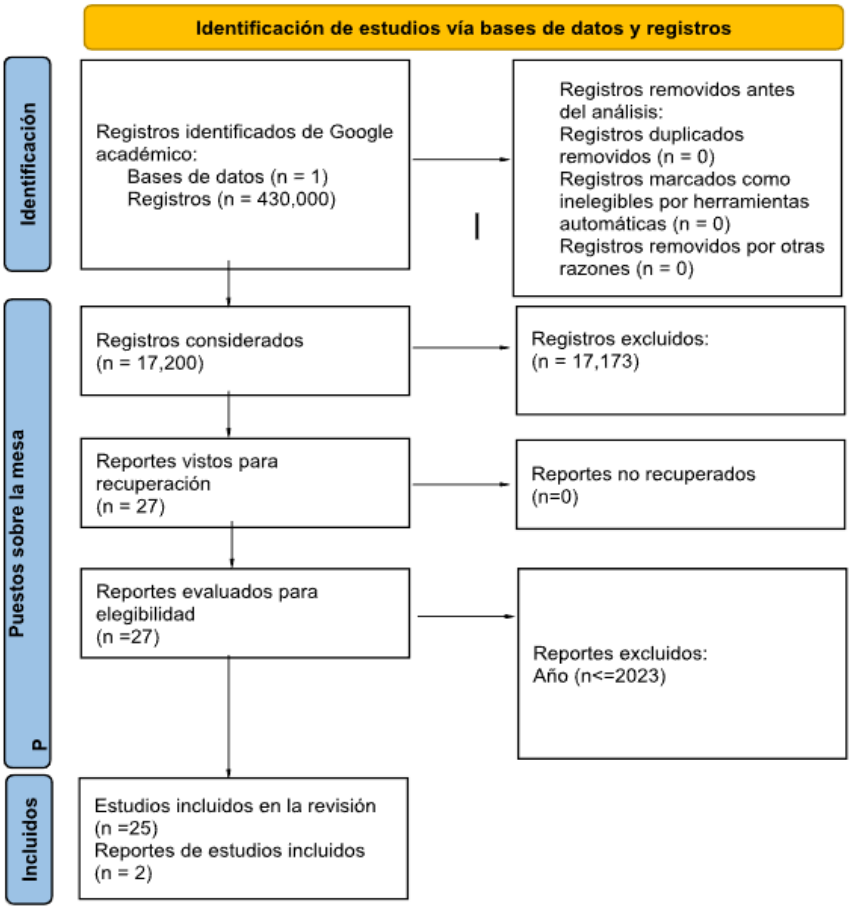
La Hipótesis que se propone es H1. Los sistemas de inteligencia artificial en el sector salud, al ser entrenados con datos no representativos, perpetúan y amplifican sesgos sociales preexistentes. La hipótesis nula se formula de la siguiente manera H0. Los sistemas de inteligencia artificial en el sector salud, al ser entrenados con datos no representativos, no perpetúan y amplifican sesgos sociales preexistentes.

En cuanto a los alcances y limitaciones de esta investigación se destaca que el repositorio académico que se ocupará para la adquisición de datos es el de Google Académico. Para asegurarnos que los contenidos son actuales se buscará analizar investigaciones recientes, del periodo 2023-2024 preferentemente.

Metodología

Para este trabajo de investigación se empleó la metodología PRISMA (Page et al., 2021), declaración 2020. Esta metodología es muy utilizada en la creación de revisiones sistemáticas de la literatura para el área de salud. Los pasos que se siguieron para la presente investigación se pueden observar en la Figura 1. El repositorio de datos que se eligió para esta investigación fue la de Google Académico, las palabras clave que se eligieron están en Inglés: "machine learning bias" + "healthcare applications" ("sesgo de aprendizaje automático" + "aplicaciones sanitarias") lo que nos dio un resultado de 430,000 registros; posteriormente se agregaron las palabras clave " + case studies" (+ estudios de caso) y los siguientes filtros de fecha para asegurarnos de considerar las investigaciones más recientes: período de tiempo 2023-2024, lo que nos arrojó un resultado de 17,200 registros, los cuales posteriormente fueron ordenados por relevancia (los más citados). Se decidió tomar los primeros veintisiete estudios más relevantes. Los estudios se clasificaron por temáticas, para cada temática se generó un resumen de los casos y temas más relevantes que se encontraron. Posteriormente, estas temáticas se tabularon y graficaron en Excel.

Figura 1. Metodología PRISMA aplicada a la consulta: “machine learning bias” + “healthcare applications” + “case studies.” en Google académico.



Fuente: adaptado de Prisma-statement.org (2024).

Resultados y discusión

De los artículos que se analizaron utilizando la metodología PRISMA, se identificaron las siguientes grandes temáticas:

Sesgos en el diagnóstico dermatológico

Abhari y Ashok (2023), destacan que las herramientas de detección de cáncer de piel basadas en aprendizaje automático (ML) tienen el potencial de revolucionar el diagnóstico temprano de cáncer de piel, democratizando el acceso. Sin embargo, estas tecnologías enfrentan un problema significativo: el sesgo racial sistémico derivado de la agrupación de datos aplicados en el entrenamiento. Las comunidades raciales y étnicas subrepresentadas en dichos conjuntos no pueden beneficiarse plenamente de estas herramientas, lo que intensifica las desigualdades en salud. Los autores señalan que los datos de entrenamiento tienden a incluir predominantemente imágenes de personas con tonos de piel más claros, a pesar de que el cáncer de piel resulta más mortal en personas de color.

En su investigación, emplean técnicas de adaptación de dominio, utilizando CycleGAN, para abordar estos sesgos al modificar imágenes de minorías para que se asemejen a las de las comunidades mayoritarias. Este enfoque permite aumentar los conjuntos de datos de minorías, beneficiando significativamente la eficiencia de estos modelos de aprendizaje automático en la clasificación sobre el cáncer de piel. Como resultado, la precisión en la clasificación de imágenes de tonos de piel de minorías se eleva del 50 % al 79 %. También presentan una aplicación móvil como demostración práctica de esta metodología.

Montoya et al. (2024), realizaron un estudio profundo acerca de la literatura para el diagnóstico precoz del melanoma utilizando inteligencia artificial entre 2013 y 2024. En su estudio, señalaron que, aunque la IA ofrece avances significativos en la identificación de esta enfermedad, aún persiste un sesgo hacia tonos de piel claros. Para mitigar este problema, sugirieron incorporar un modelo de evaluación más integral que complemente los tonos representados por el mapa de colores de L'Oréal. Los autores subrayan la importancia de emplear conjuntos de datos variados y métricas sólidas para diseñar modelos

equitativos, destacando que aplicar un marco adaptado de equidad, como PRISMA, podría reducir las desigualdades en el diagnóstico del melanoma.

En Salinas et al. (2024), se presenta una revisión sistemática y un metaanálisis para comparar el desempeño de los algoritmos de IA en la clasificación del cáncer de piel con el de médicos con diferentes niveles de experiencia. Aplicando las directrices PRISMA, se analizaron tres repositorios digitales (PubMed, Embase y Cochrane Library) hasta agosto de 2022 para identificar estudios relevantes. La calidad de los trabajos se evaluó mediante el instrumento QUADAS-2. Los autores subrayan la necesidad de considerar las limitaciones actuales de la IA en entornos clínicos y proponen que futuros estudios se centren en evaluar su utilidad en escenarios reales y en su capacidad para apoyar el trabajo médico.

Holtz et al. (2024), revisaron la literatura sobre modelos de IA aplicados a imágenes no invasivas para la detección temprana de cánceres de piel no melanoma, evaluando precisión, sensibilidad, especificidad y riesgo de sesgo. Usaron bases como MEDLINE y PubMed para recopilar estudios publicados entre 2018 y 2023, seleccionando 44 trabajos: 40 sobre dermatoscopia, 3 con microscopía confocal de reflectancia (RCM) y 1 con imágenes hiperespectrales (HEI). La precisión promedio fue del 86,8 %, similar en dermatoscopia; RCM alcanzó un 87 %, y HEI no reportó datos. Concluyen que los algoritmos de IA tienen buen desempeño, pero es necesaria más investigación para aislar su precisión específica en cánceres no melanoma.

Furriel et al. (2024), realizaron una revisión sistemática para analizar el impacto de la inteligencia artificial (IA) y las tecnologías avanzadas en la detección, clasificación y diagnóstico de imágenes de cáncer de piel en contextos clínicos. Los autores se enfocaron en evaluar el potencial de herramientas como las redes neuronales convolucionales y los sistemas de detección asistida por computadora, para detectar y clasificar lesiones cutáneas, con especial énfasis

en enfermedades como el melanoma. Además, investigaron las técnicas convencionales de diagnóstico, incluyendo la inspección visual, el análisis histopatológico y el uso de dispositivos como el dermatoscopio, reconociendo sus fortalezas y limitaciones.

El estudio destacó los principales desafíos asociados con la implementación de tecnologías de IA en dermatología, tales como la necesidad de estandarizar los métodos de adquisición y procesamiento de imágenes, la falta de conjuntos de datos amplios y representativos, y la validación insuficiente de estas herramientas en entornos clínicos reales. También señalaron la importancia de considerar la diversidad étnica y genética en el desarrollo de estas tecnologías, así como la urgencia de adherirse a estándares éticos y científicos. Finalmente, los autores subrayan que las herramientas basadas en IA deben entenderse como un complemento del diagnóstico médico, no como un reemplazo, y resaltaron su potencial para mejorar la precisión diagnóstica y brindar acceso al cuidado especializado para regiones de recursos limitados. A través de esta revisión, propusieron recomendaciones para futuras investigaciones, destacando la relevancia de abordar las brechas en el conocimiento existente y fomentar el desarrollo de tecnologías más inclusivas y accesibles.

Wei et al. (2024), presentaron que los sesgos en la IA aplicada a la dermatología incluyen limitaciones en los datos de entrenamiento, como la predominancia de imágenes de piel clara, tipos de lesiones restringidos y calidad variable de las imágenes, lo que reduce su capacidad para generalizar en tonos de piel oscuros y lesiones diversas. Además, la falta de seguimiento adecuado dificulta identificar diagnósticos falsos negativos, mientras que las diferencias regulatorias, como las menores exigencias en Europa frente a la FDA, afectan la precisión. También, la exclusión de factores relevantes como la exposición a rayos UV y antecedentes familiares limita su efectividad, y el desempeño desigual en poblaciones con baja

prevalencia de cáncer de piel restringe su aplicabilidad global, comprometiendo la equidad y la confiabilidad clínica.

Sesgos en el diagnóstico cardiológico

Garcha y Phillips (2023), realizaron un estudio centrado en la equidad para identificar sesgos relacionados con raza, género y estrato social en algoritmos de aprendizaje automático (Machine Learning - ML) diseñados para evaluar el riesgo de enfermedades cardiovasculares (ECV) en comparación con el índice de riesgo de Framingham (FRS). Mediante una búsqueda exhaustiva en MEDLINE, Embase e IEEE. Los autores investigaron si estos algoritmos abordan los sesgos inherentes al FRS. Sin restricciones de fecha, pero limitados a artículos en inglés, incluyeron estudios que evalúan algoritmos enfocados en enfermedades cardiovasculares específicas y comparan su desempeño con el FRS. Se excluyeron los algoritmos centrados en diagnóstico para la realización de un análisis narrativo estructurado de los estudios seleccionados.

Thamman et al. (2023), destacan cómo la pandemia de COVID-19 exacerbó las desigualdades sociales y de salud, con un aumento del 20 % en las muertes por enfermedades cardíacas en poblaciones negras, hispanas y asiáticas, en comparación con las blancas. Estas desigualdades se profundizan debido a factores como la menor evaluación de presión arterial y colesterol, así como la necesidad de migrar los servicios médicos a modalidades virtuales. Aunque la inteligencia artificial tiene potencial para reducir estas brechas, también puede generar sesgos de muestreo debido a la carencia de acceso justo a los servicios de salud y a los altos costos asociados con su implementación, afectando especialmente a centros con menos recursos.

Van Assen et al. (2024), analizan de manera exhaustiva los sesgos implícitos y explícitos presentes en los datos utilizados para la atención de enfermedades cardiovasculares (ECV) y cómo estos

afectan al avance en el desempeño de los modelos de IA. Las ECV, siendo la causa principal de mortalidad a nivel global, presentan desigualdades significativas en los resultados según género y raza, derivadas de diferencias en biomarcadores, representación en ensayos clínicos, diagnóstico y tratamiento. Estas disparidades, especialmente evidentes en comunidades históricamente marginadas, se ven amplificadas por la limitada representación de estos grupos en los datos clínicos utilizados para entrenar los modelos de IA. El estudio revisa además estrategias actuales, como herramientas de auditoría, para mitigar estas inequidades, subrayando la importancia de abordar estos sesgos en la selección de datos, las métricas utilizadas y la evaluación de los algoritmos, que permitan mejorar la equidad en la atención cardiovascular.

Muzammil et al. (2024), destacan que la electrocardiografía (ECG) mejorada con inteligencia artificial (IA) representa una poderosa herramienta para la predicción, la detección y el tratamiento de enfermedades cardiovasculares. Aunque el ECG convencional es accesible, económico y útil para evaluar la condición del corazón, su interpretación varía según la experiencia del médico, lo que puede dificultar el diagnóstico. La integración de IA, especialmente a través de redes neuronales convolucionales (CNN) de aprendizaje profundo, ha permitido desarrollar modelos automatizados que analizan los ECG con alta precisión, detectando patrones sutiles que podrían pasar desapercibidos para los humanos. Estos sistemas no invasivos destacan por su capacidad de identificar problemas como arritmias, enfermedades silenciosas e insuficiencia ventricular izquierda, lo que mejora la detección y tratamiento, especialmente en emergencias. Sin embargo, su implementación enfrenta desafíos como sesgos en los datos (edad, género, raza), problemas de generalización, barreras regulatorias y falta de interpretabilidad. Para garantizar su uso eficaz, se requiere mejorar la representatividad de los datos, mitigar sesgos y validar continuamente los modelos. A pesar de estas limitaciones, la IA en ECG tiene el potencial de transformar la cardiología, facilitando

diagnósticos precisos y mejorando los resultados clínicos, aunque se necesita más investigación para perfeccionar su aplicación.

Chen et al. (2024), destacan el uso de la inteligencia artificial y los registros médicos electrónicos (EHR) con el objetivo de optimizar el servicio médico, resaltando la importancia de abordar estos sesgos en los modelos de IA para evitar agravar las disparidades en salud. Mediante la revisión sistemática de la literatura publicada entre 2010 y 2023, se identificaron seis tipos principales de sesgos (algorítmico, de confusión, implícito, de medición, de selección y temporal) en 20 estudios seleccionados. Además, se propusieron estrategias para mitigar estos sesgos, enfocándose en técnicas de preprocesamiento de datos como remuestreo y ponderación, y se destacó la importancia de informes estandarizados y evaluaciones prácticas en entornos reales para garantizar una IA ética y equitativa en la atención sanitaria.

Sesgos en la interpretación de imágenes médicas

Wang et al. (2023), mencionan que, aunque el aprendizaje automático ha mostrado un gran potencial en la medicina, también genera preocupaciones por posibles sesgos relacionados con género, edad, etnicidad, hospitales y protocolos de adquisición de datos. En su estudio sobre tres enfermedades cerebrales, demuestran que los modelos entrenados adecuadamente pueden generalizar correctamente y minimizar sesgos. Utilizando resonancias magnéticas de múltiples estudios para diagnosticar Alzheimer, esquizofrenia y trastorno del espectro autista, lograron altos valores de Área bajo la curva o AUC (Area Under Curve) en diferentes subgrupos, siendo imparciales según métricas de equidad. Además, los modelos que incluyen datos demográficos, clínicos y genéticos suelen ser más precisos, aunque no siempre mejoran el desempeño en todos los casos.

Vrudhula et al. (2024), destacan que las imágenes médicas son fundamentales en el diagnóstico, pero están influenciadas por sesgos

relacionados con el acceso, la captura y la interpretación. El aprendizaje automático posee la capacidad de optimizar diagnósticos, detectar condiciones infra diagnosticadas y reducir sesgos cognitivos, aunque también puede perpetuar desigualdades si no se diseña y aplica adecuadamente. Los autores proponen un marco para equilibrar riesgos y beneficios, promoviendo un uso responsable del aprendizaje automático para garantizar una atención médica más equitativa.

Tejani et al. (2024), destacan que los algoritmos de IA pueden presentar sesgos en diversas etapas de su desarrollo, lo que podría agravar desigualdades en salud. En el ámbito de imágenes médicas, el sesgo incluye preferencias desiguales, desviaciones cognitivas y errores estadísticos que afectan la precisión y representatividad de los modelos. Esto puede dañar a los pacientes al basarse en resultados inexactos o perpetuar inequidades entre grupos. Sin embargo, un enfoque consciente del "sesgo equitativo" puede abordar la subrepresentación de minorías o enfermedades raras. Además, advierten sobre el sesgo de automatización, que lleva a aceptar decisiones automatizadas sin cuestionarlas.

Los autores revisan fuentes de sesgo en el ciclo de vida de la IA y proponen medidas de control de calidad para mitigarlas, simplificando conceptos técnicos para radiólogos generales. Comprender estos términos es clave para prevenir y abordar proactivamente el sesgo en la IA aplicada a imágenes médicas.

Gurmessa y Jimma (2024), analizaron el uso de inteligencia artificial explicable (XAI) en la identificación del cáncer de mama, revisando 646 artículos, de los cuales se incluyeron 79 tras evaluar calidad y sesgos. Solo 14 estudios utilizaron XAI, y uno evaluó la confianza en su aplicación. El 92,86 % identificó problemas en los conjuntos de datos como una brecha clave. Concluyen que XAI aún no genera suficiente confianza ni ha sido evaluado sistemáticamente, limitando su aplicación práctica debido a sesgos y carencias en la investigación.

Zeng et al. (2024), evaluaron los errores de la IA en mamografías para detección del cáncer de mama, analizando su impacto en la precisión diagnóstica. En una revisión de siete estudios retrospectivos (2019-2022, 447,676 exámenes), encontraron que los falsos positivos (FPP) disminuyen y los falsos negativos (FNP) aumentan a medida que se eleva el umbral de positividad de la IA, con variaciones influenciadas por la versión del algoritmo y la calidad del estándar de referencia. Otros tipos de errores, como problemas técnicos o de localización, se reportaron escasamente. Concluyeron que ampliar el análisis de errores podría ofrecer una perspectiva más completa sobre la utilidad de la IA en la práctica clínica.

Thomassin-Naggara et al. (2024), destacan que los errores en el diagnóstico por imágenes mamarias pueden tener consecuencias graves, como retrasos en el tratamiento, procedimientos innecesarios, costos sanitarios elevados y desconfianza en el sistema de salud. Estos errores pueden afectar negativamente los resultados de salud, generar angustia emocional y ocasionar problemas legales para los proveedores. Para prevenirlos, es crucial el uso adecuado de técnicas diagnósticas, la formación continua de los profesionales y una comunicación efectiva entre los equipos médicos. En caso de error, se deben tomar medidas como repetir estudios, realizar biopsias adicionales o derivar a especialistas.

Ahmed et al. (2024), proponen un marco que combina redes neuronales convolucionales (CNN) e inteligencia artificial explicable (XAI) con el fin de ir perfeccionando el diagnóstico del cáncer de mama mediante el conjunto de datos CBIS-DDSM. Utilizando técnicas de preprocesamiento, aumento de datos y transferencia de aprendizaje con modelos como VGG-16, Inception-V3 y ResNet, el estudio evalúa cómo XAI ayuda a interpretar las predicciones del modelo mediante la medida de Hausdorff para comparar explicaciones generadas por IA con anotaciones de expertos. Este enfoque busca promover la confianza, la ética y la integración fluida de IA en entornos clínicos, fomentando una mejor colaboración entre médicos y sistemas de

IA, y sentando las bases para investigaciones futuras sobre datos multimodales y explicaciones adaptadas a la práctica clínica.

Sesgos en datos genéticos y biomarcadores

Winchester et al. (2023), destacan que el aumento de cohortes multimodales de gran escala y el avance en tecnologías de alto rendimiento han ampliado significativamente las posibilidades de descubrir nuevos biomarcadores, eliminando la limitación impuesta por el tamaño de los conjuntos de datos. Para ello, se han implementado enfoques de IA y ML capaces de identificar biomarcadores e interacciones complejas en estos datos. En su análisis, revisan casos ejemplares y evalúan tanto las aplicaciones actuales como las limitaciones de estos métodos.

Entre los desafíos principales se mencionan la falta de diversidad en los conjuntos de datos, la complejidad inherente a la investigación de interacciones, el carácter invasivo y el alto costo de algunos biomarcadores, así como problemas en la calidad de los informes de ciertos estudios. Para superar estos retos, propone estrategias como la inclusión de poblaciones subrepresentadas, el desarrollo de IA más robusta, la validación de biomarcadores no invasivos y el cumplimiento de estándares rigurosos de presentación de resultados. Al combinar datos multimodales con enfoques de IA y fomentar la colaboración internacional, existe un gran potencial para identificar biomarcadores clínicos que sean precisos, generalizables, imparciales y adecuados para la práctica clínica.

Song et al. (2023), mencionan que los avances en la digitalización de portaobjetos de tejido y el progreso acelerado en inteligencia artificial, particularmente en el aprendizaje profundo, han revolucionado el campo de la patología computacional. Este ámbito presenta un gran potencial para automatizar diagnósticos clínicos, prever el pronóstico de los pacientes, anticipar la respuesta a tratamientos y descubrir nuevos biomarcadores morfológicos a partir de imágenes de tejidos.

Aunque algunos sistemas basados en inteligencia artificial ya están siendo aprobados para apoyar el diagnóstico clínico, persisten desafíos técnicos que dificultan su adopción generalizada en la práctica médica y su integración como herramienta de estudio. En este trabajo se discuten los avances metodológicos más recientes en patología computacional para predecir resultados clínicos a partir de imágenes completas de portaobjetos y resalta cómo estos desarrollos promueven la automatización en la aplicación clínica y el hallazgo de biomarcadores innovadores. Asimismo, se exploran perspectivas futuras a medida que el campo se amplía para abordar una mayor variedad de tareas clínicas y de investigación, integrando datos clínicos cada vez más diversos.

Kather et al. (2023), destacan que prever la respuesta a la inmunoterapia es uno de los retos principales en oncología. Los ICI o inhibidores de puntos de control inmunitarios, indican resultados revolucionarios en algunos pacientes con melanoma, lo que ha llevado a su uso generalizado en diversos tipos de cáncer. Sin embargo, mientras algunos pacientes responden de manera excepcional, otros no muestran respuesta y pueden experimentar efectos adversos significativos. Por ello, es crucial identificar biomarcadores predictivos que permitan una selección más precisa de los pacientes, optimizando los tratamientos y minimizando los riesgos. Aunque existen biomarcadores como la inestabilidad de microsatélites (MSI), la carga mutacional tumoral (TMB), la expresión de PD-L1 y los linfocitos infiltrantes de tumores (TIL), su capacidad predictiva individual o combinada sigue siendo insuficiente. Esto resalta la necesidad de biomarcadores mejorados, con alta precisión, reproducibilidad y costos reducidos.

Rajpal et al. (2023), destacan como la principal causa de mortalidad femenina el cáncer de mama, y que de acuerdo con (Sung et al., 2021) presentó su mayor incidencia en 2020. Debido a su heterogeneidad, se clasifica en subtipos clínicos y moleculares, siendo crucial identificar biomarcadores específicos para diagnóstico

y tratamiento. La metilación del ADN, un cambio epigenético clave, afecta la progresión del cáncer: la hipermetilación silencia genes supresores, mientras que la hipometilación activa oncogenes (Holm et al., 2010). Los autores proponen XAI-Methyl Marker, un marco basado en inteligencia artificial explicable, que combina reducción de dimensionalidad y redes neuronales profundas para clasificar cinco subtipos de cáncer de mama e identificar 52 biomarcadores clave. Este enfoque alcanzó una precisión del 81.45% y reveló genes asociados con subtipos, tratamientos farmacológicos y pronósticos, ofreciendo nuevas posibilidades para intervenciones terapéuticas personalizadas.

Hajjar et al. (2023), destacan que los avances en procesamiento del lenguaje natural, la identificación de voz y el aprendizaje automático, han abierto nuevas posibilidades para analizar cambios lingüísticos y acústicos que antes eran difíciles de estudiar. En su investigación, diseñaron procesos para obtener medidas léxico-semánticas y acústicas que sirvan como biomarcadores digitales de voz para la enfermedad de Alzheimer (EA).

Marwala (2024), examina las implicaciones éticas de los algoritmos de inteligencia artificial (IA), diferenciando entre discriminación evitable e inevitable. La primera puede reducirse con mejor gobernanza de datos, diseño de algoritmos y regulaciones, mientras que la segunda surge de limitaciones tecnológicas o estándares legales y éticos. A través de casos en sectores como salud, finanzas y conflictos internacionales, se ilustra el impacto de ambas. Se propone un marco multidisciplinario para identificar y mitigar la discriminación, destacando estrategias como mejorar la calidad de los datos, garantizar la transparencia algorítmica y aplicar técnicas de aprendizaje automático equitativas. El capítulo incluye recomendaciones para promover sistemas de IA innovadores, eficientes y justos.

Adeoye y Su (2023), destacan que los biomarcadores salivales pueden contribuir a la mejora de la exactitud, velocidad y

efectividad en el diagnóstico y seguimiento de enfermedades orales y maxilofaciales. Estas herramientas se han empleado en patologías como enfermedades periodontales, caries, cáncer oral, disfunción de la articulación temporomandibular y trastornos de las glándulas salivales. Sin embargo, debido a las limitaciones en la precisión durante su validación, el uso de técnicas analíticas modernas, basadas en datos multiómicos, podría optimizar su desempeño. Entre estas innovaciones, la inteligencia artificial surge como un enfoque clave para maximizar el potencial de los biomarcadores salivales en la identificación y abordaje de estas patologías. En este contexto, su revisión analiza el papel y las aplicaciones actuales de la inteligencia artificial en el descubrimiento y validación de estos biomarcadores.

Green et al. (2024), destacan que las disparidades en salud, influenciadas por factores como raza, etnia, género, idioma, discapacidades, nivel socioeconómico y ambiente, los cuales evidencian sesgos sistémicos y desigualdades estructurales que afectan el acceso equitativo a atención médica. A pesar de los avances, estas disparidades siguen siendo un reto, especialmente en comunidades marginadas. El sistema de salud de EE. UU. puede empeorar estas desigualdades al generar barreras en el acceso a servicios de calidad. Las herramientas de inteligencia artificial (IA), respaldadas por evidencia científica, ofrecen un gran potencial para abordar estos problemas, permitiendo analizar factores sociales, genéticos y ambientales, y posicionándose como un recurso fundamental para fomentar la equidad en salud.

Williams (2023), analiza los dilemas éticos relacionados con el sesgo algorítmico y su posible impacto en las poblaciones que enfrentan desigualdades en salud, investigando las raíces históricas del sesgo explícito e implícito, el papel de las variables sociales de la salud y la representación de minorías raciales y étnicas en los datos. En los últimos 25 años, los avances en diagnóstico y tratamiento de enfermedades han sido notables, con tecnologías como la telemedicina, la medicina de precisión, los macrodatos y la IA en la

evolución del campo médico. Estas innovaciones, especialmente la IA, prometen mejorar la calidad de la atención, reducir costos, optimizar los resultados de tratamiento y disminuir la mortalidad. Aunque la inteligencia artificial podría contribuir a reducir las desigualdades en salud, el sesgo algorítmico podría limitar su efectividad. Este trabajo profundiza los desafíos de aplicar la IA en el contexto de las disparidades en salud, dirigido a investigadores en servicios de salud, expertos en ética, analistas de políticas, científicos sociales, investigadores en desigualdades en salud y responsables de políticas en inteligencia artificial.

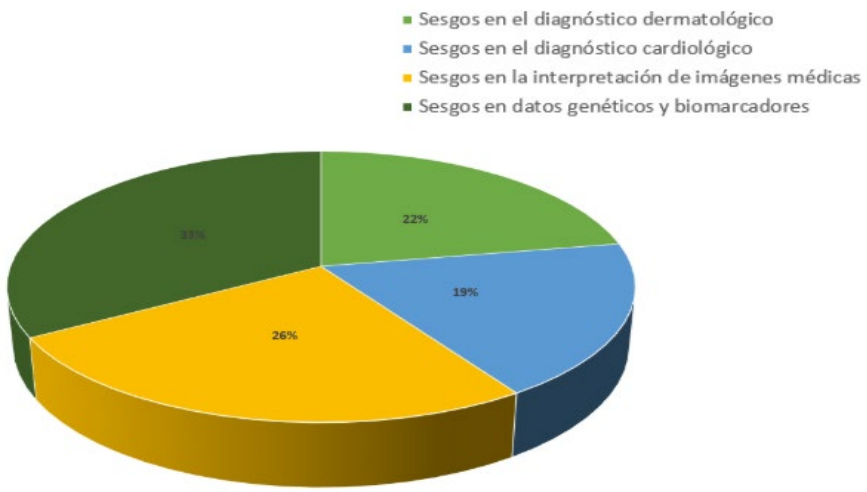
Tabla 1. Clasificación de los principales temas encontrados en la revisión de la literatura Fuente propia.

Muestra (n=27)	Frecuencia	Porcentaje
Sesgos en el diagnóstico dermatológico	6	22%
Sesgos en el diagnóstico cardiológico	5	19%
Sesgos en la interpretación de imágenes médicas	7	26%
Sesgos en datos genéticos y biomarcadores	9	33%
Total	**27**	**100%**

Fuente: elaboración propia

La Tabla 1 y la Figura 2, resumen los hallazgos obtenidos en esta investigación. En la primera columna de la Tabla 1 se describe la temática encontrada, en la segunda columna la frecuencia, y en la tercera columna el porcentaje con respecto a la base (n=27).

Figura 2. Distribución de las temáticas encontradas de los sesgos algorítmicos en inteligencia artificial reportados en el sector salud.



Fuente: elaboración propia

La inteligencia artificial se ha convertido en una herramienta esencial en el sector de la salud, aportando avances significativos en áreas como el análisis de imágenes médicas, la genética y la predicción de riesgos de salud. Sin embargo, los sistemas algorítmicos son una barrera importante para su implementación efectiva. Estos sesgos pueden surgir por diversas causas, como datos que no reflejan la diversidad demográfica o la falta de estándares éticos en el diseño y entrenamiento de algoritmos.

Por ejemplo, los sistemas de diagnóstico dermatológico basados en aprendizaje automático han mostrado una precisión significativamente menor para personas con tonos de piel oscuros debido a la falta de imágenes representativas dentro de los conjuntos de entrenamiento. Si bien se reconoce que los algoritmos de IA tienen buen desempeño en esta área, es necesaria más investigación para aislar su precisión específica, por ejemplo, en la detección de cánceres no melanoma, así como en lesiones diversas. El desempeño desigual en poblaciones con baja prevalencia de cáncer de piel restringe su

aplicabilidad global, comprometiendo la equidad y la confiabilidad clínica. En esa área existe la necesidad de estandarizar los métodos de adquisición y procesamiento de imágenes, la obtención de conjuntos de datos amplios y representativos, así como la validación de estas herramientas en entornos clínicos reales.

Del mismo modo, los modelos utilizados en cardiología frecuentemente sub representan a mujeres y minorías, amplificando disparidades históricas en el acceso equitativo a los servicios de salud. El COVID-19 exacerbó las desigualdades sociales y de salud, con un aumento importante en las muertes por enfermedades cardíacas en poblaciones negras, hispanas y asiáticas, en comparación con las blancas. La inteligencia artificial tiene potencial para reducir estas brechas de atención, sin embargo, puede generar sesgos de muestreo derivados a la falta de inclusión equitativa a la atención médica y a los altos costos asociados con su implementación, afectando especialmente a centros hospitalarios con menos recursos.

En este mismo sentido, los algoritmos de interpretación de imágenes médicas pueden estar influenciados por sesgos inherentes a los protocolos de captura de datos y los contextos hospitalarios. En esta área se identificaron sesgos relacionados con el acceso, la captura y la interpretación de las imágenes, se presentan desigualdades significativas en los resultados según edad, género y raza, representación en ensayos clínicos, diagnóstico y tratamiento. Estas disparidades son especialmente evidentes en comunidades históricamente marginadas, y se ven amplificadas por la limitada representación de estos grupos en los datos. Para garantizar su uso eficaz, se puede mejorar la representatividad de los datos, mitigar los sesgos y validar continuamente los modelos. Para ello se pueden utilizar técnicas de preprocesamiento de datos como remuestreo y ponderación.

En el ámbito de los biomarcadores genéticos, la subrepresentación de ciertas etnias en bases de datos globales limita la aplicabilidad de las tecnologías a nivel global. Este panorama evidencia la necesidad

de un enfoque interdisciplinario que combine avances tecnológicos en el área, medidas regulatorias y práctica ética que permitan garantizar una implementación más equitativa y efectiva de la IA en el sector salud. En esta área se presentan desigualdades significativas en los resultados según edad, género y raza, hospitales y protocolos de adquisición derivadas de diferencias en los biomarcadores, por lo que se requieren de biomarcadores mejorados, con alta precisión, reproducibilidad y costos reducidos. Para ello es crucial identificar biomarcadores específicos para diagnóstico y tratamiento; en este contexto, la inteligencia artificial juega un papel relevante en el descubrimiento y validación de estos biomarcadores.

Conclusiones y trabajos futuros

El uso de la inteligencia artificial en el sector salud representa un cambio paradigmático en la forma de abordar la atención médica, desde la prevención hasta el diagnóstico y tratamiento. Sin embargo, los sistemas algorítmicos identificados en esta revisión subrayan la importancia de abordar las limitaciones actuales para maximizar el impacto positivo de estas tecnologías.

En primer lugar, es fundamental garantizar la representatividad en los conjuntos de datos utilizados para el entrenamiento de los algoritmos. Esto incluye no solo diversificar las bases de datos existentes, sino también establecer colaboraciones internacionales para la recolección de datos en contextos diversos. En segundo lugar, es imprescindible desarrollar estándares regulatorios que evalúen la transparencia y equidad de los sistemas de IA, considerando su impacto ético y social.

La capacitación continua de los profesionales de la salud en competencias digitales y la concientización de las comunidades sobre el uso de la IA son pasos esenciales para asegurar que estas herramientas sean aceptadas y utilizadas de manera efectiva. Solo a través de un enfoque integral, que combine tecnología, ética y

educación, será posible aprovechar el potencial de la IA para reducir desigualdades y transformar la salud a nivel global.

De acuerdo con los hallazgos de esta revisión de la literatura utilizando la metodología PRISMA, se identificaron desafíos claves relacionados con los sistemas algorítmicos en áreas como el diagnóstico dermatológico y cardiológico, la interpretación de imágenes médicas, y el análisis de datos genéticos y biomarcadores. Estos sesgos afectan la precisión, equidad y aplicabilidad de los sistemas de IA. Por ejemplo, la falta de representatividad de los conjuntos de datos en el diagnóstico dermatológico limita la efectividad de los algoritmos para personas con tonos de piel oscuros. De manera similar, en cardiología, las mujeres y las minorías étnicas suelen estar subrepresentadas, amplificando las desigualdades históricas en el acceso a servicios médicos. Con esta investigación y los resultados arriba expuestos se confirma la hipótesis H1. Los sistemas de inteligencia artificial en el sector salud, al ser entrenados con datos no representativos, perpetúan y amplifican sesgos sociales preexistentes.

Frente a este panorama, se identifican varias líneas para investigaciones futuras. Es necesario ampliar la diversidad de los conjuntos de datos utilizados en el entrenamiento de algoritmos, incorporando información de poblaciones subrepresentadas en términos de etnia, género y condiciones socioeconómicas, para mejorar la equidad y precisión de los modelos. Asimismo, se requiere el desarrollo de estándares regulatorios y éticos internacionales que garanticen la transparencia, la equidad, y el impacto social de los sistemas de inteligencia artificial aplicados a la salud.

Por otra parte, se deben realizar investigaciones que validen el desempeño de los algoritmos en entornos clínicos reales, asegurando su aceptación y uso por parte de los profesionales de la salud. La formación del personal sanitario en competencias digitales también es una prioridad, mediante programas educativos que les permitan comprender y manejar estas herramientas de manera efectiva en su

práctica clínica. Además, se sugiere explorar enfoques innovadores, como técnicas de inteligencia artificial explicable (XAI) y modelos adaptativos, para mitigar los sesgos detectados en áreas críticas como diagnóstico médico e interpretación de imágenes.

Finalmente, se destaca la importancia de analizar cómo las comunidades perciben y adoptan estas tecnologías, y cómo la educación y estrategias de comunicación pueden fomentar la confianza y el uso ético de la IA. También es fundamental diseñar herramientas que prioricen la diversidad y la equidad, asegurando que las aplicaciones beneficien a comunidades marginadas y con recursos limitados. Estas líneas de trabajo contribuirán a desarrollar sistemas de IA más inclusivos, éticos y efectivos, capaces de transformar la atención médica global de manera equitativa.

Referencias

Abhari, J., & Ashok, A. (2023). Mitigación de sesgos raciales en la detección del cáncer de piel basada en aprendizaje automático. En *Actas del Vigésimo Cuarto Simposio Internacional sobre Teoría, Fundamentos Algorítmicos y Diseño de Protocolos para Redes Móviles y Computación Móvil* (pp. 556–561). Association for Computing Machinery. https://doi.org/10.1145/3565287.3617639

Adeoye, J., & Su, Y. (2023). Artificial intelligence in salivary biomarker discovery and validation for oral diseases. *Oral Diseases, 30*(1), 23–37. https://doi.org/10.1111/odi.14641

Ahmed, M., Bibi, T., Khan, R. A., & Nasir, S. (2024, 05 de abril). *Enhancing Breast Cancer Diagnosis in Mammography: Evaluation and Integration of Convolutional Neural Networks and Explainable AI*. arXiv. https://arxiv.org/abs/2404.03892

Ali, O., Abdelbaki, W., Shrestha, A., Elbasi, E., Alryalat, M. A. A., & Dwivedi, Y. K. (2023). A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. *Journal of Innovation & Knowledge, 8*(1). https://doi.org/10.1016/j.jik.2023.100333

Ali, H., & Nikberg, N. (2024). Bias in AI-Driven Healthcare: Navigating Ethical Challenges at an Early Stage.

Chen, F., Wang, L., Hong, J., Jiang, J., & Zhou, L. (2024). Unmasking bias in artificial intelligence: A systematic review of bias detection and mitigation strategies in electronic health record-based models. *Journal of the American Medical Informatics Association, 31*(5), 1230–1242. https://doi.org/10.1093/jamia/ocae060

Foltz, E. A., Witkowski, A., Becker, A. L., Latour, E., Lim, J. Y., Hamilton, A., & Ludzik, J. (2024). Artificial Intelligence Applied to Non-Invasive Imaging Modalities in Identification of Nonmelanoma Skin Cancer: A Systematic Review. *Cancers, 16*(3). https://doi.org/10.3390/cancers16030629

Furriel, B. C. R. S., Oliveira, B. D., Prôa, R., Paiva, J. Q., Loureiro, R. M., Calixto, W. P., Reis, M. R. C., & Giavina-Bianchi, M. (2024). Artificial intelligence for skin cancer detection and classification for clinical environment: A systematic review. *Frontiers in Medicine, 10*. https://doi.org/10.3389/fmed.2023.1305954

Garcha, I., & Phillips, S. P. (2023). Social bias in artificial intelligence algorithms designed to improve cardiovascular risk assessment relative to the Framingham Risk Score: A protocol for a systematic review. *BMJ Open, 13*(5). https://doi.org/10.1136/bmjopen-2022-067638

Green, B. L., Murphy, A., & Robinson, E. (2024). Accelerating health disparities research with artificial intelligence. *Frontiers in Digital Health, 6*. https://doi.org/10.3389/fdgth.2024.1330160

Gurmessa, D. K., & Jimma, W. (2024). Explainable machine learning for breast cancer diagnosis from mammography and ultrasound images: A systematic review. *BMJ Health & Care Informatics, 31*(1). https://doi.org/10.1136/bmjhci-2023-100954

Hajjar, I., Okafor, M., Choi, J. D., Moore, E., Abrol, A., Calhoun, V. D., & Goldstein, F. C. (2023). Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 15*(1). https://doi.org/10.1002/dad2.12393

Kather, J. N., & Perez-Lopez, R. (2023). Data and hopes for the use of artificial intelligence for predictive immunotherapy biomarkers in cancer. *Clinical Cancer Research, 29*(2), 316–323. https://doi.org/10.1158/1078-0432.CCR-22-0390

Marwala, T. (2024). Avoidable and Inevitable AI Algorithmic Bias. En *The Problem of Balance in AI Governance* (pp. 85–102). Springer. https://doi.org/10.1007/978-981-97-9251-1_8

Montoya, L. N., Roberts, J. S., & Hidalgo, B. S. (2024, 19 de noviembre). *Towards Fairness in AI for Melanoma Detection: Systemic Review and Recommendations*. arXiv. https://arxiv.org/abs/2411.12846

Muzammil, M. A., Javid, S., Afridi, A. K., Siddineni, R., Shahabi, M., Haseeb, M., Fariha, F. N. U., Kumar, S., Zaveri, S., & Nashwan, A. J. (2024). AI-enhanced electrocardiography for accurate diagnosis and treatment of cardiovascular diseases. *Journal of Electrocardiology, 83*, 30–40. https://doi.org/10.1016/j.jelectrocard.2024.01.006

Rajpal, S., Rajpal, A., Saggar, A., Vaid, A. K., Kumar, V., Agarwal, M., & Kumar, N. (2023). XAI-MethylMarker: Explainable AI approach for biomarker discovery for breast cancer subtype classification using methylation data. *Expert Systems with Applications, 225*. https://doi.org/10.1016/j.eswa.2023.120130

Salinas, M. P., Sepúlveda, J., Hidalgo, L., Peirano, D., Morel, M., Uribe, P., Rotemberg, V., Briones, J., Mery, D., & Navarrete-Dechent, C. (2024). A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis. *NPJ Digital Medicine, 7*(1). https://doi.org/10.1038/s41746-024-01103-x

Song, A. H., Jaume, G., Williamson, D. F. K., Lu, M. Y., Vaidya, A., Miller, T. R., & Mahmood, F. (2023). Artificial intelligence for digital and computational pathology. *Nature Reviews Bioengineering, 1*(12), 930–949. https://doi.org/10.1038/s44222-023-00096-8

Tejani, A. S., Ng, Y. S., Xi, Y., & Rayan, J. C. (2024). Understanding and Mitigating Bias in Imaging Artificial Intelligence. *Radiographics, 44*(5). https://doi.org/10.1148/rg.230067

Thamman, R., Yong, C. M., Tran, A. H., Tobb, K., & Brandt, E. J. (2023). Role of Artificial Intelligence in Cardiovascular Health Disparities. *JACC: Advances, 2*(7). https://doi.org/10.1016/j.jacadv.2023.100578

Thomassin-Naggara, I., Kilburn-Toppin, F., Athanasiou, A., Forrai, G., Ispas, M., Lesaru, M., Giannotti, E., Pinker-Domenig, K., Van Ongeval, C., Gilbert, F., Mann, R. M., & Pediconi, F. (2024). Misdiagnosis in breast imaging: A statement paper from European Society Breast Imaging (EUSOBI)—Part 1: The role of common errors in radiology in missed breast cancer and implications of misdiagnosis. *European Radiology, 35, 2387–2396*. https://doi.org/10.1007/s00330-024-11128-1

Van Assen, M., Beecy, A., Gershon, G., Newsome, J., Trivedi, H., & Gichoya, J. (2024). Implications of Bias in Artificial Intelligence: Considerations for Cardiovascular Imaging. *Current Atherosclerosis Reports, 26*(4), 91–102. https://doi.org/10.1007/s11883-024-01190-x

Vrudhula, A., Kwan, A. C., Ouyang, D., & Cheng, S. (2024). Machine Learning and Bias in Medical Imaging: Opportunities and Challenges. *Circulation: Cardiovascular Imaging, 17*(2). https://doi.org/10.1161/CIRCIMAGING.123.015495

Wang, R., Chaudhari, P., & Davatzikos, C. (2023). Bias in machine learning models can be significantly mitigated by careful training: Evidence from neuroimaging studies. *Proceedings of the National Academy of Sciences, 120*(6). https://doi.org/10.1073/pnas.2211613120

Wei, M. L., Tada, M., So, A., & Torres, R. (2024). Artificial intelligence and skin cancer. *Frontiers in Medicine, 11*. https://doi.org/10.3389/fmed.2024.1331895

Williams, N. H. (2023). Artificial Intelligence and Algorithmic Bias. En *The International Library of Bioethics* (pp. 1–18). Springer. https://doi.org/10.1007/978-3-031-48262-5_1

Winchester, L. M., Harshfield, E. L., Shi, L., Badhwar, A., Khleifat, A. A., Clarke, N., Dehsarvi, A., Lengyel, I., Lourida, I., Madan, C. R., Marzi, S. J., Proitsi, P., Rajkumar, A. P., Rittman, T., Silajdžić, E., Tamburin, S., Ranson, J. M., & Llewellyn, D. J. (2023). Artificial intelligence for biomarker discovery in Alzheimer's disease and dementia. *Alzheimer's & Dementia, 19*(12), 5860–5871. https://doi.org/10.1002/alz.13390

World Health Organization. (2021, 28 de junio). *WHO issues first global report on AI in health and six guiding principles for its design and use*. https://n9.cl/7bhcl

Zeng, A., Houssami, N., Noguchi, N., Nickel, B., & Marinovich, M. L. (2024). Frequency and characteristics of errors by artificial intelligence (AI) in reading screening mammography: A systematic review. *Breast Cancer Research and Treatment, 207*(1), 1–13. https://doi.org/10.1007/s10549-024-07353-3

Zeng, D., Cao, Z., & Neill, D. B. (2021). Artificial intelligence–enabled public health surveillance—From local detection to global epidemic monitoring and control. En *Artificial intelligence in medicine* (pp. 437–453). Academic Press.

Algorithmic Biases in Healthcare AI Applications: A Systematic Literature Review

Vieses Algorítmicos em Aplicações de IA no Setor Saúde: Uma Revisão Sistemática da Literatura

Yaidy Sughey Solís Valdez
Universidad Autónoma del Estado de Morelos | Morelos | México
https://orcid.org/0009-0008-1168-0239
sugheyvldz@gmail.com
sugheyvldz@gmail.com
José Alberto Hernández Aguilar
Universidad Autónoma del Estado de Morelos | Morelos | México
https://orcid.org/0000-0002-5184-0005
jose_hernandez@uaem.mx
jose_hernandez@uaem.mx
Laura Cruz Abarca
Universidad Autónoma del Estado de Morelos | Morelos | México
https://orcid.org/0000-0001-5770-5974
laura.cruz@uaem.mx
laura.cruz@uaem.mx

Abstract

Artificial Intelligence (AI) is transforming the healthcare sector, particularly in early disease detection and medical diagnosis. However, this progress faces significant challenges stemming from algorithmic biases. Small, unbalanced, or unrepresentative datasets used in model training and validation can generate errors that affect the fairness, accuracy, and reliability of clinical outcomes. These issues also raise ethical and technical dilemmas that require urgent attention. This research systematically aims to identify and analyze the main challenges associated with algorithmic biases in AI applications in healthcare, with the purpose of providing evidence and establishing a foundation for future comprehensive research. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology, widely used in systematic reviews in the health field, was employed. The results derived from its application are presented and discussed throughout the chapter.
Keywords: Artificial Intelligence; Bias; Computer Applications; Health; PRISMA methodology.

Resumo

A Inteligência Artificial (IA) está transformando o setor da saúde, especialmente na detecção precoce de doenças e no diagnóstico médico; no entanto, esse avanço enfrenta desafios significativos derivados dos vieses algorítmicos. Conjuntos de dados pequenos, desbalanceados ou não representativos utilizados no treinamento e validação de modelos podem gerar erros que afetam a equidade, a precisão e a confiabilidade dos resultados clínicos, além de apresentar dilemas éticos e técnicos que requerem atenção urgente. Esta pesquisa busca identificar e analisar de maneira sistemática os principais desafios associados aos vieses algorítmicos em aplicações de IA na atenção médica, com o propósito de aportar evidências e estabelecer bases para futuras pesquisas abrangentes. Empregou-se a metodologia Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), amplamente utilizada em revisões sistemáticas no

âmbito da saúde. Os resultados derivados de sua aplicação são apresentados e discutidos ao longo do capítulo.
Palavras-chave: Inteligência Artificial; Viés; Aplicação Informática; Saúde; metodologia PRISMA.