

Estudio documental de los estándares de validación algorítmica en radiología digital:

Retos para la equidad en el diagnóstico automatizado

Samuel Sneider Caicedo Cardona

Asesor

Edna Rocio Jamaica Guio

Universidad Nacional Abierta y a Distancia - UNAD

Escuela de Ciencias de la Salud - ECISA

Tecnología en Radiología e Imágenes Diagnosticas.

2026

Resumen

La radiografía de tórax es una herramienta fundamental en el diagnóstico clínico debido a su accesibilidad y utilidad en la evaluación de patologías pulmonares y cardiovasculares. Su calidad depende de factores técnicos y de las características anatómicas del paciente, las cuales pueden variar entre diferentes grupos poblacionales.

El presente estudio tuvo como objetivo analizar la relación entre la variabilidad anatómica según grupos étnicos y los posibles sesgos en algoritmos de inteligencia artificial aplicados a la interpretación de radiografías de tórax. Se empleó un enfoque cualitativo, de tipo documental y diseño descriptivo-analítico, basado en la revisión de literatura científica.

Los resultados evidencian que las diferencias antropométricas influyen en la formación de la imagen radiográfica y pueden afectar el desempeño de los sistemas de inteligencia artificial. Asimismo, se identificó que la falta de representatividad de ciertos grupos poblacionales en los conjuntos de datos contribuye a la aparición de sesgos algorítmicos, lo que impacta la precisión diagnóstica.

Se concluye que es necesario promover la inclusión de poblaciones diversas en las bases de datos médicas y desarrollar modelos de inteligencia artificial más equitativos, con el fin de mejorar la calidad del diagnóstico y garantizar la equidad en salud.

Palabras Clave: Radiografía de tórax, Inteligencia artificial, Sesgos algorítmicos, Variabilidad antropométrica; Equidad en salud.

Abstract

Chest radiography is a fundamental diagnostic tool widely used in clinical practice due to its accessibility and effectiveness in evaluating pulmonary and cardiovascular conditions. However, image quality and interpretation are influenced by technical factors and anatomical variability among different population groups.

This study aimed to analyze the relationship between anatomical variability across ethnic groups and potential biases in artificial intelligence algorithms applied to chest radiograph interpretation. A qualitative, documentary approach with a descriptive- analytical design was employed, based on a review of scientific literature.

The results show that anthropometric differences can affect image formation and influence the performance of artificial intelligence systems. Furthermore, the lack of representation of certain population groups in medical datasets contributes to algorithmic bias, impacting diagnostic accuracy.

It is concluded that promoting diversity in medical datasets and developing more inclusive artificial intelligence models are essential to improve diagnostic accuracy and ensure equity in healthcare.

Keywords: Chest radiography, Artificial intelligence, Algorithmic bias, Anthropometric variability, Health equity.

Contenido

Introducción	8
Planteamiento del Problema	9
Justificación	11
Objetivos	13
Objetivo General	13
Objetivos Específicos.....	13
Marco Teórico.....	14
Fundamentos de la Radiografía de Tórax y Calidad de Imagen	15
<i>Principios de Formación de la Imagen</i>	15
<i>Factores Técnicos: kVp y mAs</i>	16
Interacciones Físicas y Criterios de Calidad Diagnóstica	18
Variabilidad Antropométrica y su Impacto en la Radiografía	19
<i>Variabilidad Anatómica y Antropométrica Según la Etnia</i>	20
<i>Densidad Mineral Ósea (DMO)</i>	20
<i>Composición del Tejido Blando</i>	21
Desempeño Asimétrico en Subgrupos Poblacionales	22
Bases de Datos Utilizadas en Inteligencia Artificial Radiológica	22
Limitaciones de las Bases de Datos Utilizadas	23
<i>Sesgo de Datos (Data Bias)</i>	24
<i>Precisión Diagnóstica Diferencial</i>	24
Representatividad en la Literatura Médica y Bases de Datos	25
Integración Crítica y Aportes Investigativos.....	26
Impacto en el Contexto Colombiano	27
Marco Metodológico.....	28

	5
Descriptores de Búsqueda (términos MeSH y DeCS)	28
Criterios de Inclusión	28
Criterios de Exclusión	28
Diagrama de flujo.....	29
Resultados	30
Variabilidad en la Densidad Mineral Ósea (DMO)	30
Influencia del Hábito Corporal y Grosor de la Pared Torácica.....	30
Sesgo de Subdiagnóstico por Homogeneidad de los Datasets	31
Conclusiones	33
Referencias Bibliográficas	34

Lista de Tablas

Tabla 1 <i>Análisis Comparativo del Impacto de las Variables Anatómicas en la Predicción de la IA</i>	31
--	----

Lista de Figuras

Figura 1	<i>Principios Formación de la Imagen</i>	16
Figura 2	<i>Factores Técnicos de la Imagen</i>	17
Figura 3	<i>Radiografía Subpenetrada que Afecta el Contraste Estructural</i>	19
Figura 4	<i>Densidad Mineral Ósea (DMO)</i>	21
Figura 5	<i>Variabilidad Anatómica Torácica</i>	22
Figura 6	<i>Inteligencia Artificial en Radiología</i>	24
Figura 7	<i>Radiografía de Tórax con Penetración Adecuada</i>	25
Figura 8	<i>Diagrama de Flujo del Procedimiento de Análisis Documental</i>	29

Introducción

En la última década, la integración de la inteligencia artificial (IA) en la radiología digital ha transformado el flujo de trabajo clínico, optimizando los tiempos de lectura y mejorando el triaje en los servicios de urgencias (Esteve et al., 2019). No obstante, la eficacia de estos sistemas de aprendizaje profundo se encuentra estrictamente supeditada a la representatividad y calidad de los datos utilizados durante su fase de entrenamiento. De acuerdo con Mehrabi et al. (2021), los sesgos en los modelos automatizados surgen con frecuencia cuando los conjuntos de datos omiten la diversidad biológica de la población real.

En el ámbito de la radiografía de tórax, esta problemática adquiere una dimensión física y ética crítica; los algoritmos operan bajo patrones estadísticos que suelen ignorar variables antropométricas como la densidad mineral ósea o el grosor de la pared torácica, características que varían significativamente entre grupos étnicos (Bushberg et al., 2020). Cuando estos sistemas se despliegan de forma estandarizada, se produce un incremento en las tasas de su diagnóstico en poblaciones minoritarias, perpetuando las brechas de inequidad en el acceso a servicios de salud de calidad (Seyyed-Kalantari et al., 2021). Por consiguiente, la presente investigación documental analiza estos factores desde una perspectiva de control de la calidad tecnológica, buscando proponer estándares de validación que salvaguarden la seguridad del paciente.

Planteamiento del Problema

La inteligencia artificial (IA) se ha incorporado de manera gradual en el ámbito de la radiología como un recurso auxiliar para la interpretación de imágenes diagnósticas, sobre todo en radiografía torácica. Los progresos en aprendizaje profundo han facilitado la creación de algoritmos que puedan identificar patologías con precisiones similares a las de expertos, lo que constituye una oportunidad significativa para reforzar el diagnóstico temprano, optimizar tiempos y mejorar la atención sanitaria (Esteva et al., 2019).

No obstante, la efectividad de estos sistemas se relaciona directamente con los datos utilizados para su entrenamiento. La IA identifica patrones a partir de bases de datos actuales, por lo que su desempeño puede verse influenciado cuando esos datos tienen desbalances o carecen de representatividad. Según Mehrabi et al. (2021), los sesgos en los algoritmos frecuentemente surgen cuando los conjuntos de datos de entrenamiento no representan de manera adecuada la diversidad de la población real, lo que produce resultados desiguales entre diferentes grupos.

En el sector de la salud, esta cuestión cobra una importancia crucial ya que los fallos impactan no solo en indicadores técnicos, sino también en decisiones médicas. Obermeyer et al. (2019) mostraron que un algoritmo destinado a manejar necesidades médicas subestimaba la atención necesaria para pacientes afroamericanos, destacando cómo estos sistemas pueden perpetuar desigualdades estructurales. De forma similar, Buolamwini y Gebru (2018) encontraron tasas de error considerablemente mayores en individuos con tonos de piel oscuros, evidenciando que la escasez de diversidad afecta el rendimiento de la tecnología.

En la radiografía de tórax, la escasa representación de grupos étnicos minoritarios en los conjuntos de datos de formación puede resultar en una menor precisión diagnóstica para estas poblaciones, elevando los falsos negativos o los diagnósticos tardíos. Esto representa un desafío no solo técnico, sino también ético, dado que una herramienta diagnóstica con

rendimiento inconsistente pone en riesgo el principio de equidad en salud. En Colombia, el tema ha sido tratado por entidades académicas y científicas que alertan sobre el peligro de la discriminación racial en la creación y uso de IA (Academia Nacional de Medicina de Colombia, 2024; UNAD, 2025). Incluso medios de comunicación han informado sobre fallas y errores relacionados con el uso de IA en cuestiones médicas, lo que demuestra la necesidad de una evaluación crítica antes de su implementación clínica (Pulzo, 2025).

Por lo tanto, el problema principal se centra en entender de qué manera la ausencia de representatividad étnica en conjuntos de datos de entrenamiento influye en la precisión diagnóstica de algoritmos utilizados para radiografías de tórax, así como sus repercusiones clínicas en términos de equidad, seguridad del paciente y justicia en salud. Para organizar de manera adecuada este análisis, se utilizan enfoques metodológicos de investigación que facilitan la delimitación del problema y el examen de sus causas y efectos, como el árbol de problemas (González et al., 2023; Ortiz, 2022).

Justificación

La investigación sobre los sesgos algorítmicos en inteligencia artificial aplicada a imágenes diagnósticas es pertinente en cuatro aspectos: científico, clínico, social y metodológico.

Desde una óptica científica y tecnológica, examinar la conexión entre la representatividad étnica y la exactitud diagnóstica facilita entender de qué manera los modelos de IA se generalizan a poblaciones variadas. La literatura señala que los sesgos en el aprendizaje automático a menudo surgen de desequilibrios en los conjuntos de datos, impactando la validez y la fiabilidad de los sistemas (Mehrabi et al., 2021). Como resultado, este análisis ayuda a la creación de algoritmos más sólidos y justos.

En segundo lugar, desde la perspectiva clínica, el sesgo algorítmico puede afectar de manera directa la seguridad del paciente. Si un algoritmo muestra menor sensibilidad hacia ciertos grupos étnicos, incrementará la probabilidad de falsos negativos, lo que conlleva diagnósticos tardíos, demoras en tratamientos y posibles complicaciones. Esto tiene particular relevancia en la radiografía de tórax, debido a su uso común en enfermedades respiratorias de alta incidencia.

En tercer lugar, bajo una óptica ética y social, la aplicación de IA con resultados desiguales puede agravar las inequidades históricas en cuanto al acceso y la calidad de atención sanitaria. Si un instrumento de diagnóstico tiene un rendimiento superior en determinados grupos poblacionales e inferior en otros, se crea una desigualdad tecnológica que va en contra del principio de equidad en salud. Por lo tanto, se estima esencial incluir auditorías de equidad, validación en diferentes poblaciones y estrategias de reducción del sesgo antes de su aplicación clínica (Academia Nacional de Medicina de Colombia, 2024; UNAD, 2025).

Finalmente, esta labor se fundamenta en el enfoque metodológico, dado que refuerza. El proceso de investigación a través de la aplicación de herramientas de revisión bibliográfica y la organización del problema. Hernández et al. (2022) subrayan que una revisión bibliográfica exhaustiva es crucial en estudios clínicos, ya que proporciona un respaldo teórico al problema y asegura la calidad del análisis. De igual manera, los recursos didácticos creados en la UNAD guían la elaboración de un enfoque robusto, la generación de interrogantes de investigación y el análisis causal a través del diagrama de problemas (Corredor, 2018; Guerrero, 2024; Lancheros, 2023; Ortiz, 2022).

En resumen, este estudio es relevante ya que la incorporación de IA en radiología progresa aceleradamente, mientras que la validación ética y poblacional de estas tecnologías aún enfrenta desafíos. Examinar el sesgo causado por la escasez de diversidad étnica proporcionará pruebas clave para fomentar el uso responsable de la IA, asegurando seguridad, equidad y justicia en el ámbito de la salud.

Objetivos

Objetivo General

Analizar, mediante una revisión documental, cómo la falta de diversidad poblacional en las bases de datos utilizadas para entrenar algoritmos de inteligencia artificial influye en la precisión diagnóstica de las radiografías de tórax.

Objetivos Específicos

Identificar las principales variables anatómicas y antropométricas que pueden influir en la interpretación de radiografías de tórax mediante algoritmos de inteligencia artificial.

Describir los sesgos diagnósticos reportados en la literatura científica relacionados con la falta de representatividad de diferentes grupos poblacionales en los conjuntos de datos de entrenamiento.

Analizar las implicaciones que estos sesgos pueden tener en la precisión diagnóstica y en la equidad de la atención en salud.

Marco Teórico

La radiografía de tórax constituye uno de los estudios de imagen más utilizados en la práctica clínica para la evaluación de patologías respiratorias, cardiovasculares y torácicas. Su utilidad diagnóstica se debe a su rapidez, bajo costo y capacidad para ofrecer una visión general de las estructuras pulmonares, mediastínicas y óseas. Sin embargo, la calidad diagnóstica de una radiografía no depende únicamente del equipo utilizado, sino también de factores físicos, técnicos y biológicos que influyen en la formación y la interpretación de la imagen.

En las últimas décadas, la inteligencia artificial (IA) ha comenzado a incorporarse en la interpretación de imágenes médicas, especialmente en radiología. Los modelos de aprendizaje profundo han demostrado un desempeño prometedor en la detección automatizada de patologías en radiografías de tórax, incluyendo neumonía, nódulos pulmonares y cardiomegalia. No obstante, el rendimiento de estos algoritmos depende directamente de la calidad y representatividad de los datos utilizados durante su entrenamiento.

En este contexto, es necesario comprender tres dimensiones fundamentales: los fundamentos físicos de la formación de la imagen radiográfica, la variabilidad anatómica entre diferentes poblaciones y el impacto de los sesgos presentes en los conjuntos de datos utilizados para entrenar modelos de inteligencia artificial. La interacción entre estos factores puede influir tanto en la calidad de la imagen como en la precisión diagnóstica de los sistemas automatizados.

Por lo tanto, este marco teórico analiza la relación entre los principios físicos de la radiografía de tórax, las diferencias antropométricas entre grupos poblacionales y los posibles sesgos presentes en los algoritmos de inteligencia artificial utilizados en radiología.

Fundamentos de la Radiografía de Tórax y Calidad de Imagen

La radiografía convencional se basa en la atenuación diferencial del haz de rayos X al interactuar con tejidos de distinta composición. El efecto fotoeléctrico predomina en estructuras de mayor número atómico como el hueso, mientras que la dispersión Compton predomina en tejidos blandos y contribuye al ruido en la imagen (Bushberg et al., 2012).

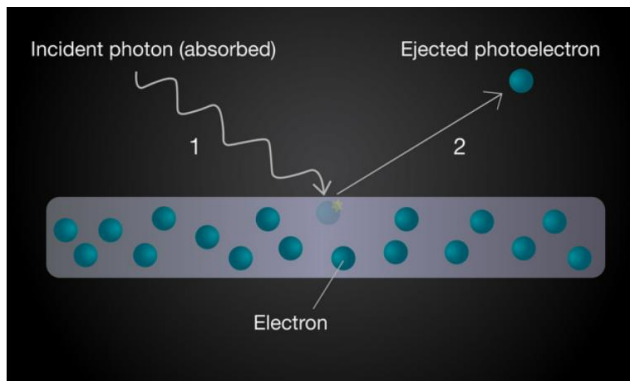
Principios de Formación de la Imagen

En radiología diagnóstica, la formación de la imagen se basa en la interacción de los rayos X con los tejidos del cuerpo humano. Cuando el haz de radiación atraviesa el organismo, parte de la energía es absorbida y otra parte alcanza el detector, generando diferencias de atenuación que permiten representar las estructuras anatómicas (Bushberg et al., 2012). En el rango de energías utilizado en radiografía diagnóstica predominan principalmente dos fenómenos físicos: el efecto fotoeléctrico y la dispersión Compton. El efecto fotoeléctrico ocurre cuando un fotón de rayos X es completamente absorbido por un electrón de una capa interna del átomo, fenómeno que se presenta con mayor frecuencia en materiales con mayor número atómico, como el calcio presente en el tejido óseo (Bushberg et al., 2012).

Por otra parte, la dispersión Compton se produce cuando un fotón interactúa con un electrón de la capa externa del átomo, transfiriendo parte de su energía y cambiando su dirección. Este fenómeno es predominante en tejidos blandos y contribuye a la radiación dispersa, lo que puede reducir el contraste de la imagen radiográfica (Shepard et al., 2014).

Figura 1

Principios Formación de la Imagen



Nota. Esquema de interacciones de rayos X con la materia. Fuente. Radiopaedia 2026.

Factores Técnicos: kVp y mAs

En radiografía diagnóstica, los parámetros técnicos del equipo influyen directamente en la calidad de la imagen obtenida. Entre los factores más importantes se encuentran el kilovoltaje pico (kVp) y el miliamperaje por segundo (mAs), los cuales determinan las características físicas del haz de rayos X y su interacción con los tejidos del paciente. El kilovoltaje (kVp) define la energía de los fotones emitidos por el tubo de rayos X y, por lo tanto, su capacidad de penetración a través de las estructuras anatómicas. Un incremento en el kVp produce fotones con mayor energía, capaces de atravesar tejidos más densos, lo que se traduce en una mayor penetración del haz y una disminución del contraste radiográfico debido al aumento de la radiación dispersa (Bushberg et al., 2012; Shepard et al., 2014). Por otro lado, el miliamperaje- segundo (mAs) controla la cantidad total de fotones producidos durante la exposición. A diferencia del kVp, este parámetro no modifica la energía de los fotones, sino la intensidad del haz de radiación. Un aumento en el mAs incrementa la cantidad de radiación que alcanza el detector, lo que contribuye a reducir el ruido cuántico y mejorar la relación señal-ruido de la imagen radiográfica (Bushberg et al., 2012).

La adecuada selección de estos parámetros resulta fundamental para garantizar imágenes diagnósticas de calidad, evitando tanto la subexposición como la sobreexposición del paciente.

En la práctica clínica, los técnicos en radiología deben ajustar los valores de kVp y mAs teniendo en cuenta las características anatómicas y la complejión del paciente, ya que factores como el grosor de la pared torácica, la densidad mineral ósea y la distribución del tejido adiposo pueden modificar el grado de atenuación del haz de rayos X (Shepard et al., 2014). En este sentido, las variaciones antropométricas observadas entre diferentes poblaciones pueden influir en la selección de los parámetros técnicos. Por ejemplo, pacientes con mayor masa corporal o con mayor densidad ósea pueden requerir ajustes en los niveles de kVp o mAs para lograr una penetración adecuada del haz y evitar imágenes con bajo contraste o alto nivel de ruido. Esta adaptación técnica resulta esencial para garantizar una adecuada calidad diagnóstica en contextos clínicos diversos.

Figura 2

Factores Técnicos de la Imagen



Nota. Criterios de calidad radiográfica: Comparación de Penetración e Inspiración. Fuente. Choy et al. 2018.

Interacciones Físicas y Criterios de Calidad Diagnóstica

La calidad diagnóstica de una radiografía de tórax depende no solo de los parámetros técnicos utilizados durante la adquisición de la imagen, sino también del cumplimiento de criterios radiológicos estandarizados que permiten una interpretación adecuada.

Entre los principales criterios de calidad se encuentran la penetración, inspiración y simetría torácica (Felson, 2014). La penetración adecuada permite visualizar las estructuras vertebrales a través de la silueta cardíaca. Cuando la radiografía se encuentra subpenetrada (kVp insuficiente), la imagen aparece excesivamente blanca, lo que puede ocultar lesiones pulmonares o generar falsos positivos por aumento artificial de la densidad. Por el contrario, una sobrepenetración puede disminuir el contraste y dificultar la identificación de estructuras vasculares finas. La inspiración adecuada se evalúa observando el número de arcos costales posteriores visibles. En condiciones óptimas deben observarse aproximadamente entre 9 y 10 costillas posteriores, lo que indica una expansión pulmonar suficiente. Una inspiración deficiente puede simular cardiomegalia o aumentar la densidad pulmonar, generando interpretaciones erróneas.

Finalmente, la simetría torácica constituye un criterio fundamental para garantizar la correcta posición del paciente. Esta se evalúa comparando la distancia entre los extremos mediales de las clavículas y las apófisis espinosas de las vértebras torácicas. Una rotación del paciente puede producir una aparente asimetría pulmonar o modificar el tamaño aparente del mediastino, lo que podría confundirse con patologías como atelectasias o desviaciones mediastínicas. Por esta razón, el control técnico de estos parámetros es fundamental para garantizar que las imágenes radiográficas reflejen con precisión la anatomía del paciente y permitan una interpretación diagnóstica confiable.

Además de los factores técnicos propios del equipo radiográfico, las características anatómicas y antropométricas del paciente influyen significativamente en la formación de la

imagen. Variables como el grosor de la pared torácica, la densidad mineral ósea y la distribución del tejido adiposo pueden modificar los niveles de atenuación del haz de rayos X. Por esta razón, el tecnólogo en radiología debe ajustar parámetros como el kVp y el mAs de acuerdo con la complejión física del paciente, con el fin de garantizar una penetración adecuada y evitar artefactos o interpretaciones diagnósticas erróneas.

Figura 3

Radiografía Subpenetrada que Afecta el Contraste Estructural



Nota. Menor penetración: la radiografía se denomina placa blanda, ya que la imagen se observa demasiado blanca. Fuente. Ashesh 2026.

Variabilidad Antropométrica y su Impacto en la Radiografía

Las características anatómicas de los pacientes pueden variar significativamente entre individuos y poblaciones, lo cual tiene implicaciones directas en la adquisición y calidad de las imágenes radiográficas. Factores como el grosor de la pared torácica, la densidad mineral ósea y la distribución del tejido adiposo influyen en el grado de atenuación del haz de rayos X, lo que puede modificar la apariencia de las estructuras anatómicas en la imagen resultante (Bushberg et al., 2012).

El grosor de la pared torácica constituye uno de los factores más relevantes en la radiografía de tórax, ya que determina la cantidad de radiación necesaria para atravesar

adecuadamente el cuerpo del paciente. Estudios antropométricos han evidenciado que este parámetro puede variar según factores como el sexo, la edad y el origen étnico, lo que puede requerir ajustes en los parámetros técnicos del equipo radiológico, particularmente en los niveles de kVp y mAs, con el fin de garantizar una penetración adecuada del haz de rayos X (Shepard et al., 2014). De igual manera, la densidad mineral ósea (DMO) representa otro factor que puede influir en la absorción de la radiación.

Variabilidad Anatómica y Antropométrica Según la Etnia

Las diferencias anatómicas entre individuos y poblaciones pueden influir en la apariencia radiográfica de las estructuras torácicas. Estas variaciones incluyen diferencias en la densidad mineral ósea, el grosor de la pared torácica y la distribución del tejido adiposo.

Desde el punto de vista radiológico, estas características pueden modificar la atenuación del haz de rayos X y alterar la apariencia de las estructuras en la imagen, lo que puede afectar tanto la interpretación clínica como el desempeño de los sistemas de inteligencia artificial entrenados para analizar estas imágenes.

Densidad Mineral Ósea (DMO)

Diversos estudios han demostrado que existen diferencias significativas en la densidad mineral ósea (DMO) entre diferentes grupos étnicos. Investigaciones epidemiológicas han reportado que las poblaciones afrodescendientes tienden a presentar una mayor densidad mineral ósea en comparación con poblaciones caucásicas, lo cual se asocia con una mayor resistencia estructural del tejido óseo (Borrell & Elhawary, 2021).

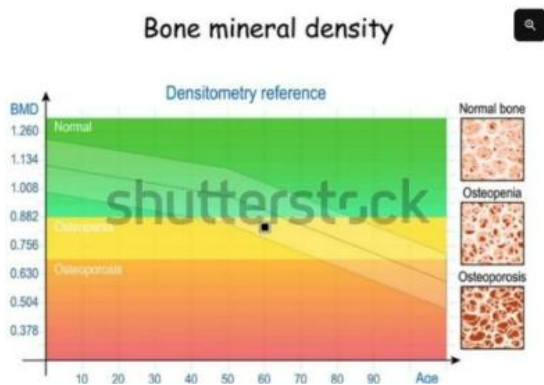
Desde el punto de vista radiológico, estas diferencias pueden modificar la atenuación del haz de rayos X, ya que los tejidos con mayor densidad mineral absorben una mayor cantidad de radiación. En radiografía de tórax, esto puede traducirse en una mayor opacidad

de las estructuras óseas, lo que potencialmente altera el contraste entre hueso y tejido pulmonar.

Adicionalmente, estudios de densitometría han mostrado que la densidad ósea promedio en población afrodescendiente puede ser entre 5 % y 15 % superior en comparación con poblaciones caucásicas, particularmente en estructuras como costillas y columna vertebral. Estas variaciones pueden influir en la apariencia radiográfica del tórax y deben ser consideradas tanto en la interpretación clínica como en el desarrollo de algoritmos de inteligencia artificial. Cuando estas diferencias no se encuentran representadas adecuadamente en los conjuntos de datos utilizados para entrenar modelos de inteligencia artificial, existe el riesgo de que el algoritmo interprete estas variaciones fisiológicas como anomalías patológicas o reduzca su sensibilidad diagnóstica en determinados grupos poblacionales.

Figura 4

Densidad Mineral Ósea (DMO)



Nota. Shutterstock, 2020.

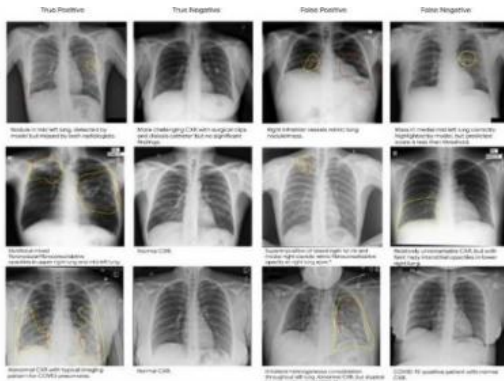
Composición del Tejido Blando

Las variaciones en el espesor de la pared torácica y en la distribución de grasa pectoral pueden producir sombras mamarias o superposiciones anatómicas que imitan enfermedades pulmonares Oakden-Rayner et al. (2020) señalan que la estratificación

encubierta en los conjuntos de datos puede generar errores clínicamente relevantes si no se tienen en cuenta subgrupos anatómicamente diferentes durante el entrenamiento de modelos de inteligencia artificial.

Figura 5

Variabilidad Anatómica Torácica



Nota. Afzali, 2018.

Desempeño Asimétrico en Subgrupos Poblacionales

Un fenómeno crítico documentado en la validación de algoritmos médicos es la pérdida de rendimiento diagnóstico al segmentar a los pacientes por sus características físicas. Mientras un modelo de inteligencia artificial puede exhibir una precisión superior al 90% en la detección global de neumonía, las tasas de falsos positivos aumentan considerablemente al evaluar radiografías de pacientes con un mayor grosor de la pared torácica (Oakden-Rayner et al., 2020). Este desempeño asimétrico evidencia que las redes neuronales a menudo memorizan atajos visuales del grupo mayoritario de entrenamiento, ignorando la atenuación diferencial que generan los tejidos blandos en pacientes con fenotipos distintos, lo que compromete la fiabilidad clínica del sistema.

Bases de Datos Utilizadas en Inteligencia Artificial Radiológica

El desarrollo de algoritmos de inteligencia artificial en radiología se ha apoyado principalmente en grandes bases de datos públicas de imágenes médicas. Entre las más

utilizadas se encuentran MIMIC-CXR y ChestX-ray14, las cuales contienen cientos de miles de radiografías de tórax utilizadas para el entrenamiento y validación de modelos de aprendizaje profundo (Johnson et al., 2019; Wang et al., 2017).

La base de datos MIMIC-CXR contiene más de 370.000 radiografías de tórax correspondientes a aproximadamente 65.000 pacientes, acompañadas de reportes radiológicos estructurados. Por su parte, ChestX-ray14 incluye más de 100.000 estudios radiográficos etiquetados con 14 patologías torácicas comunes. Aunque estas bases han sido fundamentales para el avance de la inteligencia artificial en radiología, presentan limitaciones importantes relacionadas con la disponibilidad de variables demográficas detalladas, como etnia o características antropométricas. Esta falta de información dificulta evaluar si los algoritmos presentan un rendimiento diagnóstico uniforme en todos los grupos poblacionales.

Limitaciones de las Bases de Datos Utilizadas

A pesar de que bases de datos como MIMIC-CXR y ChestX-ray14 han jugado un papel crucial en el avance de los algoritmos de inteligencia artificial en radiología, estas tienen ciertas limitaciones relevantes. En primer lugar, gran parte de estas bases se edificaron con datos originados principalmente de hospitales en Estados Unidos, lo que sugiere que la información representa primordialmente poblaciones concretas y no necesariamente muestra la diversidad global.

Asimismo, estas bases de datos tienden a incluir información demográfica restringida, lo que complica el análisis del rendimiento de los algoritmos en función de variables como etnia, rasgos antropométricos o situaciones socioeconómicas. La ausencia de representatividad puede conducir a sesgos en los modelos de inteligencia artificial, perjudicando su exactitud diagnóstica cuando se utilizan en contextos distintos, como el colombiano o el latinoamericano.

Sesgo de Datos (Data Bias)

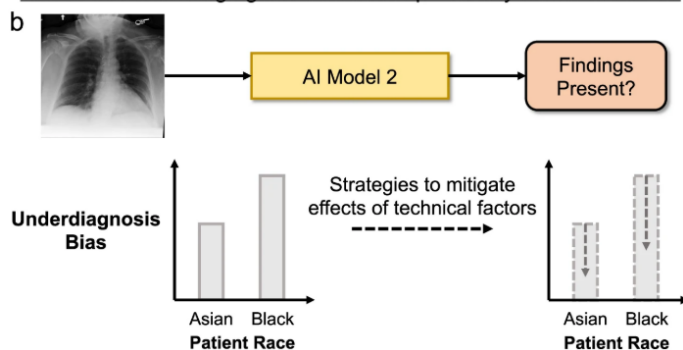
El sesgo de datos ocurre cuando los conjuntos de entrenamiento no representan adecuadamente la diversidad poblacional. Bases de datos ampliamente utilizadas como: MIMIC-CXR (Johnson et al., 2019) NIH ChestX-ray14 (Wang et al., 2017).

Han sido fundamentales en el desarrollo de modelos de inteligencia artificial en radiología. No obstante, estas bases presentan limitaciones en la documentación de variables demográficas, lo que dificulta analizar el desempeño diferencial de los algoritmos según grupo étnico. Zechetal (2018) demostraron que los modelos pueden aprender patrones institucionales en lugar de características clínicas reales, lo que reduce su capacidad de generalización entre hospitales.

Figura 6

Inteligencia Artificial en Radiología

Aim 2: Use knowledge gained to reduce previously identified AI bias



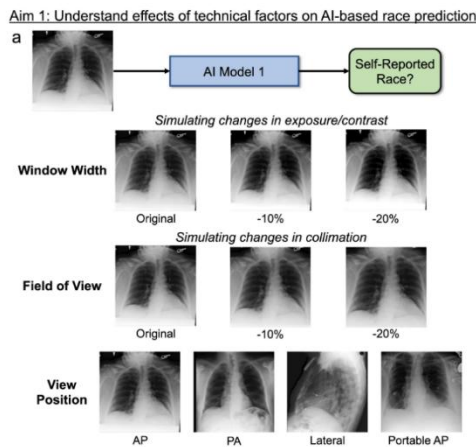
Nota. Nature Medicine, 2024.

Precisión Diagnóstica Diferencial

Obermeyer et al. (2019) demostraron que un algoritmo del sistema de salud de Estados Unidos subestimaba el riesgo en pacientes afrodescendientes, debido a variables indirectas utilizadas como indicadores de salud. Esto evidencia que los algoritmos pueden reproducir inequidades existentes en los sistemas de salud.

Figura 7

Radiografía de Tórax con Penetración Adecuada



Nota. Nature Medicine, 2024.

Representatividad en la Literatura Médica y Bases de Datos

Johnson et al. (2019) analizan la estructura de la base de datos MIMIC-CXR, destacando su gran magnitud y utilidad para el desarrollo de algoritmos. Sin embargo, la falta de información demográfica detallada limita el análisis de equidad diagnóstica. De manera similar, Wang et al. (2017) desarrollaron la base ChestX-ray14, ampliamente utilizada en investigación, aunque con limitaciones similares en representatividad poblacional. La carencia de diversidad en los conjuntos de datos dificulta determinar si el rendimiento del algoritmo es uniforme en todos los grupos de población, lo que tiene implicaciones éticas y clínicas importantes.

La representación en la literatura médica es esencial para asegurar que los progresos tecnológicos en salud favorezcan de forma justa a todas las comunidades. No obstante, varios estudios han indicado que una gran parte de la investigación médica y tecnológica se fundamenta en datos de poblaciones de naciones desarrolladas, lo que puede provocar una subrepresentación de comunidades de otras áreas del planeta. Este escenario puede influir en

la creación de algoritmos de inteligencia artificial, puesto que los modelos desarrollados con datos poco variados pueden mostrar menor precisión diagnóstica al usarse en poblaciones con diferentes características biológicas y sociales.

Integración Crítica y Aportes Investigativos

La integración crítica de la literatura analizada muestra que el progreso de la inteligencia artificial en radiología no se basa solo en avances tecnológicos, sino también en la calidad y variedad de los datos empleados para entrenar los modelos. Varios autores concuerdan en que los algoritmos pueden perpetuar las desigualdades existentes en los sistemas de salud si no se gestionan correctamente los sesgos que hay en los conjuntos de datos.

De igual manera, los estudios revisados muestran que elementos como la variabilidad anatómica, las diferencias antropométricas entre grupos y la ausencia de representación de determinados grupos étnicos pueden afectar el rendimiento diagnóstico de los sistemas automatizados. Por esta razón, la literatura científica sugiere la adopción de estrategias como auditorías de equidad, validación multicéntrica de algoritmos y creación de bases de datos más variadas.

Estos estudios indican que el uso responsable de la inteligencia artificial en el ámbito de la salud demanda no solo avances tecnológicos, sino también un enfoque ético y social que asegure la igualdad en la atención médica. La integración de evidencia obtenida en español con literatura internacional permite identificar tres aspectos clave:

La calidad técnica de una radiografía de tórax depende de factores físicos, biológicos y sociales.

La variabilidad anatómica poblacional tiene impacto diagnóstico real.

El uso de inteligencia artificial en diagnóstico requiere validación contextual y diversidad poblacional para evitar sesgos.

Impacto en el Contexto Colombiano

En el contexto colombiano, la implementación de inteligencia artificial en radiología debe considerar la diversidad étnica y antropométrica de la población. Colombia es un país con una composición poblacional heterogénea que incluye poblaciones mestizas, afrodescendientes e indígenas, cada una con características biológicas y socioeconómicas particulares. Si los algoritmos utilizados en sistemas de apoyo diagnóstico son entrenados principalmente con bases de datos provenientes de poblaciones norteamericanas o europeas, existe el riesgo de que su desempeño no sea equivalente en la población colombiana. Esto podría generar diferencias en la precisión diagnóstica y afectar la equidad en la atención en salud.

Por lo tanto, resulta necesario promover investigaciones que evalúen el rendimiento de estos sistemas en poblaciones locales y fomentar la creación de bases de datos médicas representativas de la diversidad demográfica del país.

En Colombia, la integración de inteligencia artificial en radiología todavía está en una etapa de investigación y desarrollo. No obstante, la nación muestra una amplia diversidad cultural y étnica que abarca comunidades indígenas, afrodescendientes y mestizas, lo que requiere una evaluación cuidadosa de la aplicabilidad de los algoritmos creados en otros contextos.

Si los sistemas de inteligencia artificial aplicados en el diagnóstico radiológico son entrenados mayormente con datos de poblaciones europeas o norteamericanas, hay un riesgo de que su rendimiento no sea del todo apropiado para la población colombiana. Por ello, es crucial impulsar estudios nacionales que faciliten la validación de estos sistemas en entornos locales y que propicien la creación de bases de datos médicas que representen la diversidad de la población del país.

Marco Metodológico

La presente investigación se desarrolló bajo un enfoque cualitativo con un diseño documental de alcance descriptivo-analítico. Este diseño fue seleccionado debido a que permite la sistematización, evaluación y síntesis de la evidencia científica disponible, facilitando la identificación de patrones, vacíos de conocimiento y tendencias respecto a los sesgos algorítmicos en radiología.

Descriptores de Búsqueda (términos MeSH y DeCS)

Se utilizaron las siguientes cadenas de búsqueda, combinando operadores booleanos: *"Artificial intelligence" AND "chest radiography" AND "algorithmic bias" OR "anthropometric variability"*.

Criterios de Inclusión

Artículos científicos originales, revisiones sistemáticas y guías de práctica clínica publicadas entre el año 2020 y 2026. Se priorizó literatura en inglés y español por ser los idiomas predominantes en la producción científica de vanguardia sobre IA médica.

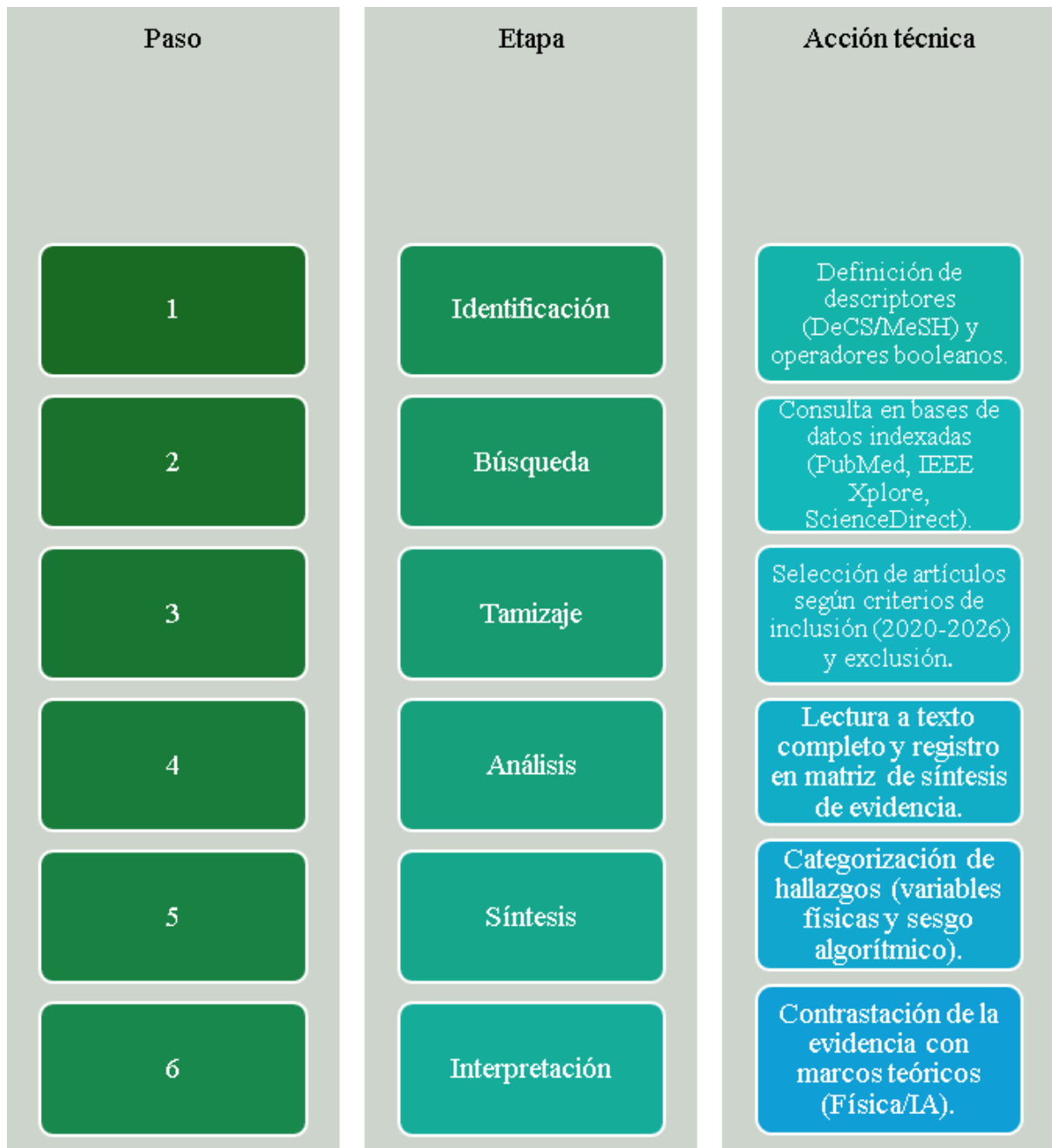
Criterios de Exclusión

Se descartaron artículos de opinión, blogs, repositorios informales, noticias o documentos sin revisión por pares, con el fin de evitar fuentes de baja credibilidad científica.

Diagrama de flujo

Figura 8

Diagrama de Flujo del Procedimiento de Análisis Documental



Nota. Elaboración propia 2026, Hernández-Sampieri & Mendoza (2018).

Resultados

Tras realizar la revisión documental, los resultados permiten concluir que el rendimiento de los algoritmos de IA en radiología no es estático ni universal, sino que se degrada significativamente cuando se enfrenta a variabilidades biofísicas que no fueron consideradas durante su fase de entrenamiento. Al contrastar los principios de la física de la exposición, como los descritos por Bushong (2020), con los hallazgos epidemiológicos sobre sesgos algorítmicos en salud (Gichoya et al., 2022), es evidente que los sistemas de IA actuales presentan una limitación técnica fundamental: la incapacidad de decodificar correctamente las variaciones en la atenuación del haz de rayos X cuando el paciente posee características anatómicas distintas a las del dataset de referencia.

La evidencia analizada destaca tres puntos de quiebre donde la IA falla sistemáticamente:

Variabilidad en la Densidad Mineral Ósea (DMO)

Se ha documentado que poblaciones de origen afrodescendiente presentan una DMO entre un 5% y un 15% superior a poblaciones caucásicas (Looker et al., 2012). Desde la física radiológica, esta diferencia anatómica genera una mayor absorción de fotones mediante el efecto fotoeléctrico. Cuando un algoritmo, entrenado con imágenes de menor densidad ósea, procesa esta mayor atenuación, el sistema suele interpretar la densidad aumentada como una falta de penetración del haz o como un artefacto, lo que deriva en falsos negativos al ocultar patologías subyacentes en el tejido pulmonar.

Influencia del Hábito Corporal y Grosor de la Pared Torácica

La literatura consultada señala que variaciones en la composición de los tejidos blandos y el grosor de la pared torácica incrementan la interacción de los fotones mediante el efecto Compton, aumentando el ruido de fondo en la imagen digital (Shepard et al., 2014; Bushong, 2020). Los algoritmos de IA, al no haber sido expuestos a esta variabilidad en el

entrenamiento, fallan al clasificar patrones intersticiales o efusiones pleurales, categorizando erróneamente hallazgos anatómicos normales como hallazgos patológicos.

Sesgo de Subdiagnóstico por Homogeneidad de los Datasets

El problema raíz identificado es la falta de representatividad en las bases de datos de entrenamiento (Seyyed-Kalantari et al., 2021). Al ser modelos construidos sobre *datasets* homogéneos, el algoritmo desarrolla una "visión" limitada que no contempla la diversidad fenotípica global, traducándose en una brecha de seguridad para los pacientes que no encajan en el perfil estándar, exacerbando las desigualdades en salud ya existentes (Gichoya et al., 2022).

Tabla 1

Análisis Comparativo del Impacto de las Variables Anatómicas en la Predicción de la IA

Variable anatómica	Efecto físico y respuesta algorítmica	Fuentes
Mayor Densidad Mineral Ósea (DMO)	Incrementa la absorción fotoeléctrica. La IA puede subdiagnosticar lesiones al interpretar la alta densidad como un artefacto.	Loocker et al. 2012; Bushong 2020.
Grosor de la Pared Torácica	Incrementa el efecto Compton y el ruido de la imagen. La IA puede confundir patrones normales con patologías intersticiales.	Bushong 2020; Shepard et al. 2014.
Homogeneidad de los Datos	La omisión de variables fenotípicas genera sesgo de subdiagnóstico sistemático en minorías étnicas.	Seyyed-Kalantari et al. 2021; Gichoya et al. 2022.

Nota. Elaboración propia 2026, basada en los marcos regulatorios y de investigación consultados.

Para la gestión de calidad institucional, estos resultados exigen que la validación del software médico incluya auditorías demográficas antes de su despliegue clínico (Banerjee et al., 2021).

Conclusiones

La revisión documental permitió identificar que la representatividad de los datos utilizados para entrenar algoritmos de inteligencia artificial es un factor determinante en su desempeño diagnóstico. Cuando los conjuntos de datos no incluyen poblaciones diversas, pueden presentarse diferencias en la precisión de los resultados obtenidos.

Asimismo, se evidenció que variables anatómicas y antropométricas, como la densidad mineral ósea y el grosor de la pared torácica, pueden influir en la formación de la imagen radiográfica y afectar el comportamiento de los sistemas automatizados de interpretación.

Los estudios analizados coinciden en que la falta de diversidad en las bases de datos médicas puede generar sesgos algorítmicos que impactan la equidad en salud, especialmente en grupos poblacionales con menor representación durante el entrenamiento de los modelos.

Finalmente, se concluye que es necesario promover la validación de los algoritmos en poblaciones diversas y fortalecer el desarrollo de bases de datos más representativas, con el fin de mejorar la precisión diagnóstica y favorecer una atención sanitaria más equitativa.

Referencias Bibliográficas

- Academia Nacional de Medicina de Colombia. (2024). *Discriminación racial en la inteligencia artificial*. <https://anmdocolombia.org.co/>
- Afzali, A., Babapour-Mofrad, F., & Pouladian, M. (2018). Inter-patient modelling of 2D lung variations from chest X-ray imaging via Fourier descriptors. *Journal of Medical Systems*, 42(11), 233. <https://doi.org/10.1007/s10916-018-1058-7>
- American College of Radiology. (2022). *ACR guidance on artificial intelligence in radiology*. <https://www.acr.org/>
- Banerjee, I., Bhimireddy, A. R., Krishnan, R., & Pan, I. (2021). Reading race: AI recognises patient's racial identity in medical images. *The Lancet Digital Health*, 3(10), e626–e627. [https://doi.org/10.1016/S2589-7500\(21\)00240-7](https://doi.org/10.1016/S2589-7500(21)00240-7)
- Borrell, L. N., & Elhawary, J. R. (2021). Race and genetic ancestry in medicine—A time for reckoning with history. *The New England Journal of Medicine*, 384(5), 474–480. <https://doi.org/10.1056/NEJMms2029562>
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Choy, G., Khalilzadeh, O., Michalski, M., Do, S., Samir, A. E., Pianykh, O. S., Geis, J. R., Pandharipande, P. V., Brink, J. A., & Dreyer, K. J. (2018). Current applications and future impact of machine learning in radiology. *Radiology*, 288(2), 318–328. <https://doi.org/10.1148/radiol.2018171820>
- Corredor, E. S. (2018). *Técnicas de investigación: Identificación del problema* [Objeto virtual de aprendizaje]. Repositorio Institucional UNAD. <https://repository.unad.edu.co/handle/10596/22997>

- Drukker, K., Giger, M. L., & Chen, H. (2023). Toward fairness in artificial intelligence for medical image analysis: Identification and mitigation of potential biases from data collection to model deployment. *Radiology: Artificial Intelligence*, 5(3).
<https://doi.org/10.1148/ryai.220296>
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
- Felson, B. (2014). *Felson's principles of chest roentgenology: A programmed text* (4th ed.). Elsevier Saunders.
- Gichoya, J. W., Banerjee, I., Bhimireddy, A. R., Burns, J. L., Celi, L. A., Chen, L. C., Correa, R., Dullerud, N., Ghassemi, M., Huang, S. C., Kuo, P. C., Lungren, M. P., Oakden-Rayner, L., Okechukwu, C., Seyyed-Kalantari, L., Trivedi, H., Wang, M., & Wilson, R. (2022). AI recognition of patient race in medical imaging: A modelling study. *The Lancet Digital Health*, 4(6), e406–e414. [https://doi.org/10.1016/S2589-7500\(22\)00063-2](https://doi.org/10.1016/S2589-7500(22)00063-2)
- González, S., Sánchez, M., & Hernández, R. (2023). Árbol de problemas como base en la investigación. *Educación y Salud*, 12(23), 125–129.
- Guerrero, K. A. (2024). *Cómo plantear una pregunta de investigación* [Objeto virtual de aprendizaje]. Repositorio Institucional UNAD.
<https://repository.unad.edu.co/handle/10596/64468>
- Hernández-Sampieri, R., & Mendoza, C. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Education.
- Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, 18(8), 500–510.
<https://doi.org/10.1038/s41568-018-0016-5>

- Hussain, S. A., Bresnahan, M., & Zhuang, J. (2025). The bias algorithm: How AI in healthcare exacerbates ethnic and racial disparities. *Ethnicity & Health, 30*(1), 45–62. <https://doi.org/10.1080/13557858.2025.2301456>
- Johnson, A. E. W., Pollard, T. J., Berkowitz, S. J., Greenbaum, N. R., Lungren, M. P., Deng, C. Y., Mark, R. G., & Horng, S. (2019). MIMIC-CXR: A large publicly available database of labeled chest radiographs. *Scientific Data, 6*, 317. <https://doi.org/10.1038/s41597-019-0322-0>
- Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine, 17*(1), 195. <https://doi.org/10.1186/s12916-019-1426-2>
- Lancheros, E. (2023). *Cómo elaborar un planteamiento de problema sólido y convincente* [Objeto virtual de información]. Repositorio Institucional UNAD. <https://repository.unad.edu.co/handle/10596/57735>
- Larrazabal, A. J., Nieto, N., Peterson, V., Milone, D. H., & Ferrante, E. (2020). Gender imbalance in medical imaging datasets produces biased classifiers. *Proceedings of the National Academy of Sciences, 117*(23), 12592–12594. <https://doi.org/10.1073/pnas.1919012117>
- Leslie, D., Mazumder, A., Peppin, A., Wolters, M. K., & Hagerty, A. (2021). Does “AI” stand for augmenting inequality in the era of COVID-19 healthcare? *BMJ, 372*. <https://doi.org/10.1136/bmj.n304>
- Looker, A. C., Melton, L. J., Harris, T. B., Borrud, L. G., & Shepherd, J. A. (2012). Prevalence and trends in low femur bone density among older U.S. adults. *Journal of Bone and Mineral Research, 27*(1), 64–71. <https://doi.org/10.1002/jbmr.523>

- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), 1–35.
<https://doi.org/10.1145/3457607>
- Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, *2*(10), 100347.
<https://doi.org/10.1016/j.patter.2021.100347>
- Oakden-Rayner, L., Dunnmon, J., Carneiro, G., & Ré, C. (2020). Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. *Proceedings of the ACM Conference on Health, Inference, and Learning*, 151–159.
<https://doi.org/10.1145/3368555.3384468>
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453.
<https://doi.org/10.1126/science.aax2342>
- Ortiz, J. (2022). *Árbol de problemas como metodología para la identificación de problemas de investigación en salud* [Objeto virtual de información]. Repositorio Institucional UNAD. <https://repository.unad.edu.co/handle/10596/50047>
- Panch, T., Mattie, H., & Celi, L. A. (2019). The “inconvenient truth” about AI in healthcare. *npj Digital Medicine*, *2*(1), 77. <https://doi.org/10.1038/s41746-019-0155-4>
- Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine*, *27*(12), 2176–2182.
<https://doi.org/10.1038/s41591-021-01595-0>
- Shepard, S. J., Linnau, K. F., & Starr, F. (2014). *Radiographic imaging and exposure* (5th ed.). Elsevier.

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly supervised classification and localization of common thorax diseases. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

<https://doi.org/10.1109/CVPR.2017.369>

World Health Organization. (2021). *Ethics and governance of artificial intelligence for health*. World Health Organization.

<https://www.who.int/publications/i/item/9789240029200>