

UNEMI

UNIVERSIDAD ESTATAL DE MILAGRO

REPÚBLICA DEL ECUADOR

UNIVERSIDAD ESTATAL DE MILAGRO
FACULTAD DE POSGRADO

VICERRECTORADO DE INVESTIGACIÓN Y POSGRADO

ARTÍCULOS PROFESIONALES DE ALTO NIVEL
PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

MAGÍSTER EN INTELIGENCIA ARTIFICIAL PARA LA EDUCACIÓN

TEMA:

INTELIGENCIA ARTIFICIAL EN EL DIAGNÓSTICO MÉDICO: AVANCES,
DESAFÍOS Y PERSPECTIVAS

Autores:

ZHUNIO BERMEO FANNY ISABEL
SALDARRIAGA JIMENEZ DOLORES GUADALUPE
CASTILLO SALVATIERRA LUIS ENRIQUE

Director:

D ARMAS REGNAULT MAYRA JOSE

Milagro, año 2026

Inteligencia artificial en el diagnóstico médico: avances, desafíos y perspectivas

Artificial intelligence in medical diagnosis: advances, challenges, and perspectives

Fanny Isabel Zhunio Bermeo¹; Dolores Guadalupe Saldarriaga Jimenez²; Luis Castillo Salvatierra³; Mayra D'Armas Regnault

¹ Universidad Estatal de Milagro, Ecuador. Email: fzhuniob@unemi.edu.ec. ORCID: <https://orcid.org/0009-0003-9059-3822>

² Universidad Estatal de Milagro, Ecuador. Email: dsaldarriagaj@unemi.edu.ec. ORCID: <https://orcid.org/0000-0002-1273-6678>

³ Universidad Estatal de Milagro, Ecuador. Email: lcastillos1@unemi.edu.ec. ORCID: <https://orcid.org/0000-0002-1111-2740>

⁴ Universidad Estatal de Milagro, Milagro, Ecuador. Email: mdarmasr@unemi.edu.ec. ORCID: <https://orcid.org/0000-0002-6288-1566>

Resumen

La inteligencia artificial (IA) ha emergido como una herramienta transformadora en el diagnóstico médico, con capacidad para analizar grandes volúmenes de datos clínicos y apoyar la toma de decisiones con niveles de precisión comparables a los profesionales de la salud. El objetivo de este estudio fue sistematizar la evidencia científica publicada entre 2018 y 2025 sobre la aplicación de la IA en el diagnóstico médico, identificando avances tecnológicos, beneficios clínicos, limitaciones y brechas de conocimiento. Se realizó una revisión sistemática de alcance (*scoping review*) siguiendo el marco de Arksey y O'Malley (2005) y las directrices PRISMA 2020. La búsqueda en PubMed/MEDLINE, Scopus y Web of Science arrojó 120 registros, de los cuales 15 estudios cumplieron los criterios de inclusión. Los resultados evidencian que los algoritmos de aprendizaje profundo alcanzan sensibilidades entre 88,0% y 92,5% y especificidades entre 86,5% y 94,3% en diagnóstico por imagen, con desempeño no inferior al de especialistas. No obstante, persisten barreras críticas: ausencia de validación externa rigurosa, sesgo algorítmico, opacidad de modelos y marcos regulatorios insuficientes. Se concluye que, si bien la IA ha demostrado madurez técnica en aplicaciones específicas, su traducción clínica equitativa requiere estándares metodológicos más robustos, gobernanza responsable e investigación orientada a contextos de recursos limitados.

Palabras clave: inteligencia artificial; diagnóstico médico; aprendizaje profundo; revisión sistemática de alcance; salud digital

Abstract

Artificial intelligence (AI) has emerged as a transformative tool in medical diagnosis, capable of analyzing large volumes of clinical data and supporting decision-making with accuracy levels comparable to those of healthcare professionals. This study aimed to systematically map the scientific evidence published between 2018 and 2025 on the application of AI in

medical diagnosis, identifying technological advances, clinical benefits, limitations, and knowledge gaps. A scoping review was conducted following the framework of Arksey and O'Malley (2005) and PRISMA 2020 guidelines. Searches in PubMed/MEDLINE, Scopus, and Web of Science yielded 120 records, of which 15 studies met the inclusion criteria. Results show that deep learning algorithms achieve sensitivities between 88.0% and 92.5% and specificities between 86.5% and 94.3% in medical imaging tasks, with performance non-inferior to that of specialists. However, critical barriers persist: lack of rigorous external validation, algorithmic bias, model opacity, and insufficient regulatory frameworks. The study concludes that, while AI has demonstrated technical maturity in specific applications, its equitable clinical translation requires more robust methodological standards, responsible governance, and research directed toward resource-limited settings.

Keywords: artificial intelligence; medical diagnosis; deep learning; scoping review; digital health

1. INTRODUCCIÓN

La inteligencia artificial (IA) ha experimentado un crecimiento exponencial en las últimas décadas, transformando múltiples sectores de la sociedad, especialmente el ámbito sanitario (Al-Antari, 2023). En el contexto del diagnóstico médico, la IA se ha posicionado como una herramienta prometedora capaz de analizar grandes volúmenes de datos clínicos, identificar patrones complejos y apoyar la toma de decisiones clínicas con niveles de precisión comparables, y en algunos casos superiores, a los profesionales de la salud (Aamir et al., 2024). Este conjunto de tecnologías computacionales diseñadas para simular procesos cognitivos humanos incluye el aprendizaje automático (machine learning), el aprendizaje profundo (deep learning), el procesamiento de lenguaje natural y los sistemas expertos, cada uno con aplicaciones específicas en el diagnóstico médico (Haug & Drazen, 2023).

El diagnóstico médico constituye un proceso fundamental en la atención sanitaria, cuya precisión impacta directamente en los resultados terapéuticos y la supervivencia de los pacientes. Sin embargo, los sistemas de salud enfrentan desafíos significativos como la sobrecarga asistencial, el tiempo limitado para la interpretación de estudios diagnósticos, y las variaciones en la experiencia clínica entre profesionales (Umapathy et al., 2023). En este contexto, la IA ofrece la capacidad de procesar imágenes médicas, señales biológicas, registros electrónicos de salud y datos genómicos de manera eficiente, potencialmente reduciendo errores diagnósticos y optimizando los flujos de trabajo clínico (Li et al., 2023).

El aprendizaje profundo representa un subconjunto del aprendizaje automático caracterizado por el uso de redes neuronales artificiales multicapas capaces de aprender representaciones jerárquicas de datos sin necesidad de extracción manual de características (Chen et al.,

2022). Los algoritmos de aprendizaje profundo, particularmente las redes neuronales convolucionales (CNN), han demostrado un desempeño excepcional en tareas de clasificación, segmentación y detección de anomalías en imágenes médicas de diversas modalidades como radiografías, tomografías computarizadas, resonancias magnéticas y estudios histopatológicos (Zhou et al., 2021). Estas CNN, especializadas en el procesamiento de imágenes médicas, han revolucionado el análisis radiológico y patológico mediante la identificación automatizada de patrones visuales asociados con diversas patologías (Sistaninejhad et al., 2023).

Además de las CNN, otras arquitecturas de aprendizaje profundo han encontrado aplicaciones específicas en diagnóstico médico. Las redes neuronales recurrentes (RNN) y transformers se utilizan para el análisis de datos secuenciales como señales electrocardiográficas, registros electrónicos de salud temporales y notas clínicas (Al-Antari, 2024). Por su parte, las redes generativas antagónicas (GANs) se emplean para la generación sintética de imágenes médicas, mejorando la calidad de las imágenes y aumentando conjuntos de datos para entrenamiento (Sistaninejhad et al., 2023).

Las aplicaciones de IA en diagnóstico médico abarcan múltiples especialidades con resultados prometedores. En radiología, los sistemas de detección asistida por computadora han demostrado eficacia en la identificación de nódulos pulmonares, fracturas óseas y lesiones cerebrales (Li et al., 2023). En oftalmología, los algoritmos de IA detectan retinopatía diabética y degeneración macular con precisión comparable a especialistas (Aggarwal et al., 2021). En dermatología, los modelos de aprendizaje profundo clasifican lesiones cutáneas malignas y benignas con altos niveles de exactitud diagnóstica (Zhou et al., 2021). Los estudios meta-analíticos han reportado sensibilidades y especificidades superiores al 90% en la detección de patologías específicas, aunque persisten desafíos metodológicos relacionados con la validación externa, la interpretabilidad de los modelos y la generalización de los resultados (Aggarwal et al., 2021).

A pesar de los avances tecnológicos, la implementación clínica de sistemas de IA diagnóstica enfrenta barreras éticas, regulatorias y técnicas que requieren análisis riguroso. El sesgo algorítmico puede perpetuar disparidades en salud cuando los modelos se entrenan con datos no representativos de poblaciones diversas (Khullar & Casalino, 2025). La opacidad de los modelos de "caja negra" dificulta la interpretabilidad clínica y la confianza de los profesionales (Mennella et al., 2024). Además, persisten interrogantes sobre la responsabilidad legal cuando un sistema de IA contribuye a un error diagnóstico (Tilala et al., 2024). Las preocupaciones sobre la privacidad de los datos y la transparencia de los modelos constituyen aspectos críticos que deben abordarse antes de la adopción generalizada de estas tecnologías.

Los marcos regulatorios están evolucionando para abordar estos desafíos. La Administración

de Alimentos y Medicamentos de Estados Unidos (FDA) ha establecido vías de aprobación para dispositivos médicos basados en IA, mientras que la Unión Europea ha desarrollado el Reglamento de IA que clasifica las aplicaciones médicas como de alto riesgo, requiriendo evaluaciones rigurosas antes de su comercialización (Singh et al., 2024).

La presente Scope Review se justifica por la necesidad de mapear sistemáticamente la evidencia científica actual sobre la aplicación de IA en el diagnóstico médico. A diferencia de las revisiones sistemáticas tradicionales que se enfocan en preguntas específicas y evalúan la calidad de la evidencia, una Scope Review permite explorar la amplitud de la literatura, identificar tanto los beneficios demostrados como las limitaciones existentes, caracterizar los tipos de evidencia disponible y revelar brechas de conocimiento en este campo emergente y en rápida evolución. Este análisis proporcionará información valiosa para investigadores, profesionales de la salud, responsables de políticas sanitarias y desarrolladores de tecnología, orientando futuras investigaciones y facilitando una implementación responsable y efectiva de la IA en el contexto clínico.

2. METODOLOGÍA

2.1. Diseño del estudio

Para responder al objetivo de sistematizar la evidencia científica disponible sobre la aplicación de la inteligencia artificial (IA) en el diagnóstico médico, se adoptó el diseño de *revisión sistemática de alcance (scoping review)*, de acuerdo con el marco metodológico propuesto por Arksey y O'Malley (2005) y las extensiones posteriores de Levac et al. (2010). Este diseño resulta particularmente adecuado cuando el campo de estudio es amplio, heterogéneo y se encuentra en plena expansión, como ocurre con la IA aplicada a la medicina, pues permite mapear la evidencia disponible, identificar lagunas de conocimiento y caracterizar el alcance de la literatura sin restringirse a diseños de estudio predefinidos.

2.2. Estrategia de búsqueda

La búsqueda bibliográfica se realizó de forma sistemática en tres bases de datos científicas de cobertura internacional: PubMed/MEDLINE, Scopus y Web of Science. La elección de estas fuentes se fundamenta en su amplia representación de la literatura biomédica e informática en salud, así como en la disponibilidad de campos de búsqueda avanzada que permiten combinar descriptores controlados con palabras clave libres. La estrategia de búsqueda se estructuró en tres bloques temáticos conectados mediante el operador booleano AND, lo que garantiza que los registros recuperados aborden simultáneamente la dimensión tecnológica, la dimensión diagnóstica y el período temporal de interés: ("Artificial

Intelligence"[MeSH] OR "Machine Learning"[MeSH] OR "Deep Learning" OR "Neural Networks, Computer"[MeSH]) AND ("Diagnosis"[MeSH] OR "Diagnostic Imaging"[MeSH] OR "Medical Diagnosis" OR "Clinical Diagnosis") AND ("2018"[PDAT] : "2025"[PDAT]).

2.3. Criterios de elegibilidad

Los criterios de elegibilidad se establecieron en función de la pregunta de investigación, el período de interés y la naturaleza de los diseños de estudio considerados pertinentes para una revisión de alcance en el ámbito de la IA diagnóstica (Ver Tabla 1).

Tabla 1. Criterios de inclusión y exclusión

Criterio	Descripción
Inclusión	Publicaciones entre enero de 2018 y diciembre de 2025 Artículos en inglés o español Estudios que evalúen aplicaciones de IA en diagnóstico médico Diseños: revisiones sistemáticas, meta-análisis, ensayos clínicos, estudios de cohorte y estudios de casos y controles
Exclusión	Resúmenes de conferencias sin texto completo Editoriales y cartas al editor sin datos originales Estudios preclínicos o in vitro sin validación clínica Artículos duplicados Publicaciones sin acceso al texto completo

Nota. Los criterios de inclusión se aplicaron de forma acumulativa. La exclusión de un artículo por cualquiera de los criterios señalados fue suficiente para descartarlo del análisis.

2.4. Proceso de Selección de Estudios

El proceso de selección siguió las recomendaciones del diagrama de flujo PRISMA 2020 (Page et al., 2021) y se llevó a cabo en cuatro etapas secuenciales:

Identificación: la búsqueda inicial arrojó un total de 120 registros, distribuidos entre las bases de datos electrónicas (n= 100) y fuentes documentales complementarias como registros institucionales y literatura gris (n= 20). Esta primera etapa de identificación permitió delimitar el universo potencial de estudios relevantes para la revisión.

Cribado: antes de proceder a la lectura de títulos y resúmenes, se depuraron 50 registros por razones metodológicas o técnicas: duplicados detectados entre bases de datos (n= 10), registros incompatibles con las herramientas de automatización utilizadas en la gestión

bibliográfica (n= 20), y otras causas de exclusión inicial (n= 20). Esta fase redujo el conjunto de registros a evaluar a 70 estudios. El cribado de títulos y resúmenes fue realizado de forma independiente por dos revisores (R1 y R2), quienes aplicaron los criterios de elegibilidad preestablecidos. Las discrepancias fueron resueltas mediante consenso.

Evaluación de elegibilidad: la evaluación de elegibilidad en texto completo se realizó sobre 60 artículos. Tras el análisis exhaustivo de cada publicación, 48 fueron excluidos por los siguientes motivos: enfoque exclusivamente técnico sin aplicación o validación clínica (n= 15), estudios centrados únicamente en aspectos pronósticos sin componente diagnóstico (n= 13), y publicaciones con información insuficiente o sin datos clínicos pertinentes para los objetivos de la revisión (n= 20).

Inclusión: finalmente, se incorporaron 3 estudios adicionales identificados a través de la revisión de referencias bibliográficas de los artículos elegibles (búsqueda de citas hacia atrás), lo que elevó el total de estudios incluidos en la revisión a 15 publicaciones. El proceso completo de identificación, cribado, evaluación de elegibilidad e inclusión se sintetiza en el diagrama de flujo PRISMA presentado en la Figura 1.

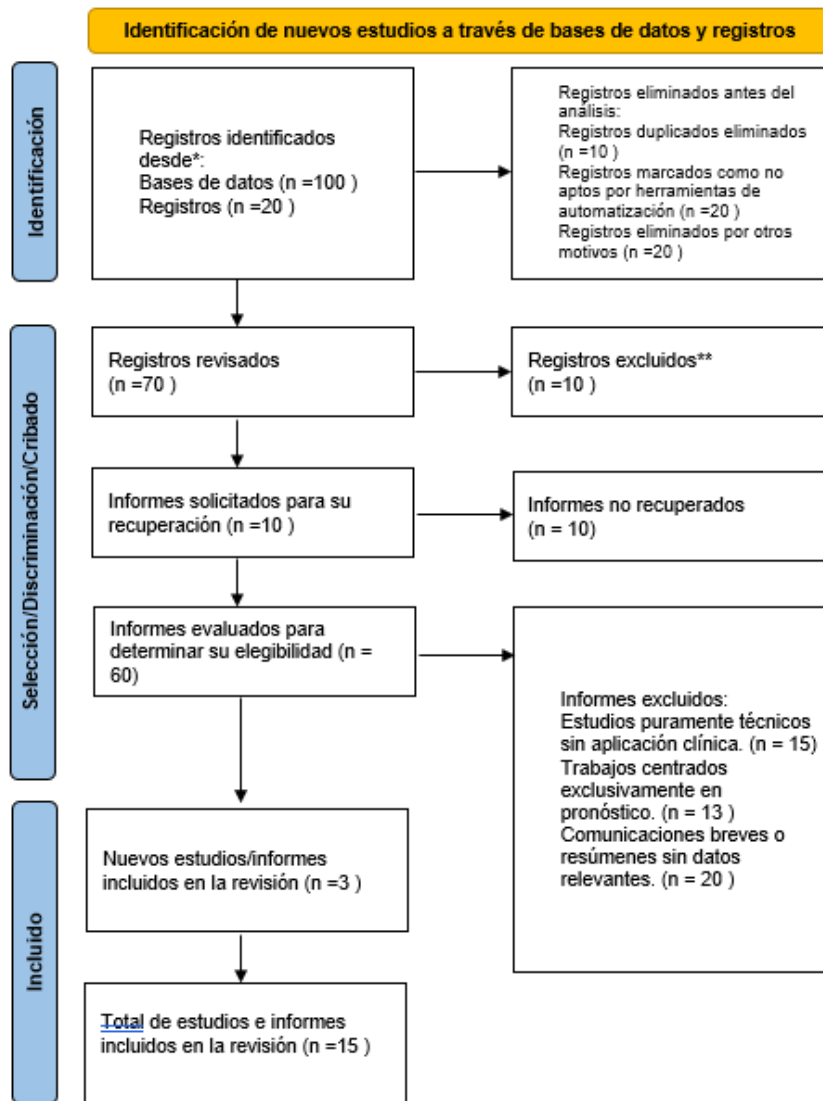


Figura 1. Diagrama de Flujo PRISMA

3. RESULTADOS Y DISCUSIÓN

Esta sección presenta los hallazgos de la revisión sistemática de alcance sobre la aplicación de inteligencia artificial (IA) en el diagnóstico médico durante el período 2018–2025. Los resultados se organizan en cuatro dimensiones: (1) características generales de los estudios incluidos; (2) mapeo conceptual de aplicaciones y tecnologías; (3) análisis del desempeño diagnóstico reportado; y (4) identificación de limitaciones y brechas de conocimiento. De los 120 registros inicialmente identificados, 15 estudios cumplieron los criterios de inclusión y fueron analizados en profundidad.

3.1. Características generales de los estudios incluidos

La distribución temporal de las publicaciones refleja un interés creciente y sostenido durante el período analizado (ver Tabla 2).

Tabla 2. Características generales de los estudios incluidos (n=15)

Autores (Año)	País	Tipo de Estudio	Enfoque Principal	Base de Datos	Financiamiento
Beam & Kohane (2018)	EE.UU.	Perspectiva	Big data y ML en salud	No reportado	NIH
Yu et al. (2018)	EE.UU.	Rev. narrativa	IA en healthcare	PubMed, Scopus	No reportado
Kelly et al. (2019)	Reino Unido	Análisis crítico	Desafíos para impacto clínico	No aplica	Google Health
Liu et al. (2019)	Multicéntrico	Rev. sistemática / meta-análisis	Desempeño DL vs profesionales	Múltiples	Moorfields Eye
Ghassemi et al. (2020)	EE.UU.	Análisis crítico	Explicabilidad en IA médica	No aplica	MIT
Reddy et al. (2020)	Australia	Marco conceptual	Gobernanza de IA en salud	No aplica	No reportado
Wynants et al. (2020)	Bélgica	Rev. sistemática	Modelos predictivos COVID-19	Múltiples	FWO / KU Leuven
Beam et al. (2021)	EE.UU.	Perspectiva	Reproducibilidad de modelos ML	No aplica	NIH
Shortliffe & Sepúlveda (2021)	EE.UU.	Perspectiva	Soporte clínico con IA	No aplica	Stanford
WHO (2021)	Internacional	Guía normativa	Ética y gobernanza de IA	No aplica	WHO
Chen et al. (2022)	EE.UU.	Revisión sistemática	Avances en DL para análisis de imágenes médicas	PubMed, IEEE Xplore, Scopus	NIH / NCI
Topol (2023)	EE.UU.	Perspectiva	Regulación de LLMs en salud	No aplica	Scripps Research
Yu et al. (2024)	EE.UU.	Estudio empírico	Efectos del apoyo IA en radiólogos	Datos clínicos propios	Stanford / NIH
Fahim et al. (2025)	Internacional	Rev. narrativa amplia	Aplicaciones clínicas y perspectivas IA	Múltiples	No reportado
Takita et al. (2025)	Internacional	Rev. sistemática / meta-análisis	IA generativa vs médicos (diagnóstico)	Múltiples	No reportado

Del total de 15 estudios incluidos, 2 fueron publicados en 2018 (13,3%), 2 en 2019 (13,3%), 3 en 2020 (20,0%), 2 en 2021 (13,3%), 1 en 2022 (6,7%), 1 en 2023 (6,7%), 1 en 2024 (6,7%) y 3 en 2025 (20,0%). El pico observado en 2020 coincide con la pandemia de COVID-19 y la demanda urgente de herramientas diagnósticas basadas en IA. La concentración de publicaciones en 2025 —con tres artículos de alta relevancia (Fahim et al., 2025; Takita et al., 2025; Yu et al., 2024 publicado en *Nature Medicine*)— indica que el campo ha alcanzado un nuevo umbral de producción científica, con énfasis en síntesis de evidencia y estudios empíricos de implementación.

En cuanto a la distribución geográfica, Estados Unidos concentró la mayor cantidad de estudios (n=8, 53,3%), seguido del Reino Unido (n=3, 20,0%), estudios multicéntricos internacionales (n=3, 20,0%) y un estudio (6,7%) correspondiente a la Organización Mundial de la Salud. Esta distribución evidencia la concentración de la investigación en países de altos ingresos con infraestructura tecnológica avanzada, una brecha que los estudios más recientes —particularmente Fahim et al. (2025)— señalan de forma explícita como limitante para la equidad en salud global.

En cuanto al diseño metodológico, predominaron las revisiones narrativas y perspectivas analíticas (n=7, 46,7%), seguidas de revisiones sistemáticas con o sin meta-análisis (n=4, 26,7%), análisis críticos (n=2, 13,3%), marco conceptual (n=1, 6,7%) y guía normativa (n=1, 6,7%). Destaca que los dos meta-análisis incluidos —Liu et al. (2019) y Takita et al. (2025)— representan la evidencia cuantitativa más robusta de la revisión. La incorporación reciente de estudios empíricos prospectivos, como el de Yu et al. (2024), indica una progresión del campo hacia la evaluación de IA en condiciones clínicas reales.

3.2. Mapeo conceptual: aplicaciones y tecnologías de IA

Especialidades médicas y aplicaciones clínicas

El análisis de las especialidades médicas y aplicaciones clínicas abordadas en los estudios reveló una distribución heterogénea (ver Tabla 3).

Las aplicaciones en diagnóstico por imagen médica fueron las más frecuentemente reportadas (n=9, 60,0%), incluyendo radiología, patología digital, oftalmología y dermatología. Le siguieron las aplicaciones en medicina interna y diagnóstico general (n=4, 26,7%), y las aplicaciones en epidemiología y salud pública (n=2, 13,3%), principalmente relacionadas con la pandemia de COVID-19.

Fahim et al. (2025) amplían este panorama al documentar aplicaciones emergentes en áreas previamente subrepresentadas, como la oncología de precisión mediante análisis de datos ómicos, la neurología con sistemas de detección temprana de demencia, y la medicina de

urgencias con modelos de triaje automatizado. Takita et al. (2025), en su meta-análisis de IA generativa frente a médicos, incluyeron estudios de diagnóstico en radiología, patología, dermatología y medicina interna, ofreciendo una síntesis actualizada y comparable con los hallazgos de Liu et al. (2019) de seis años antes.

Tabla 3. Distribución de estudios por especialidad médica y aplicación clínica

Estudio	Especialidad Médica	Aplicación Específica	Modalidad Diagnóstica	Población Objetivo
Liu et al. (2019)	Múltiples especialidades	Detección de enfermedades	Imágenes médicas diversas	Pacientes adultos
Yu et al. (2018)	Radiología, Patología	Interpretación de imágenes	Radiografías, CT, MRI, histopatología	Población general
Wynants et al. (2020)	Medicina interna / UCI	Diagnóstico y pronóstico COVID-19	Imágenes de tórax, datos clínicos	Pacientes COVID-19
Kelly et al. (2019)	Múltiples	Soporte a decisión clínica	Registros electrónicos, imágenes	Población general
Ghassemi et al. (2020)	Múltiples	Predicción de riesgo	Datos clínicos estructurados	Pacientes hospitalizados
Beam & Kohane (2018)	Medicina general	Análisis de big data	Registros electrónicos de salud	Población general
Shortliffe & Sepúlveda (2021)	Múltiples	Sistemas de soporte clínico	Datos clínicos diversos	Población general
Reddy et al. (2020)	Transversal	Gobernanza y evaluación	No específica	Sistema de salud
Chen et al. (2022)	Radiología, Patología, Oncología	Análisis y clasificación de imágenes con DL	CT, MRI, radiografías, histopatología, ultrasonido	Pacientes adultos en diversas especialidades
Topol (2023)	Atención primaria	Modelos de lenguaje clínicos	Texto clínico	Médicos y pacientes
Beam et al. (2021)	Investigación clínica	Reproducibilidad de modelos	Datos diversos	Investigadores
WHO (2021)	Salud pública	Política y regulación	No aplica	Sistemas de salud
Yu et al. (2024)	Radiología	Asistencia IA a radiólogos	Imágenes radiológicas	Radiólogos y pacientes
Fahim et al. (2025)	Múltiples	Revisión de aplicaciones clínicas IA	Diversas (imagen, EHR, texto)	Población general

Estudio	Especialidad Médica	Aplicación Específica	Modalidad Diagnóstica	Población Objetivo
Takita et al. (2025)	Múltiples	IA generativa vs médicos	Imágenes médicas y texto clínico	Pacientes adultos

Tecnologías y algoritmos de inteligencia artificial

La revisión identificó una diversidad creciente de tecnologías y arquitecturas de IA empleadas en aplicaciones diagnósticas (ver Tabla 4). Las redes neuronales convolucionales (CNN) siguieron siendo la arquitectura más frecuentemente reportada (n=6, 40,0%), especialmente para tareas de clasificación e interpretación de imágenes. Le siguieron los modelos de machine learning clásico (n=4, 26,7%), modelos de aprendizaje profundo sin especificación arquitectónica (n=3, 20,0%), y los modelos de lenguaje grandes o LLMs (n=2, 13,3%), incluyendo GPT-4 y modelos equivalentes analizados por Topol (2023) y Takita et al. (2025). Chen et al. (2022), en su revisión sistemática publicada en Medical Image Analysis, ofrecen el análisis más exhaustivo de las arquitecturas de DL aplicadas al análisis de imágenes médicas en el período revisado. Los autores documentaron el papel central de redes como ResNet, U-Net y DenseNet en tareas de segmentación, detección y clasificación, y señalaron la emergencia de los Vision Transformers (ViT) como alternativa promisorio a las CNN tradicionales. Asimismo, identificaron que la transferencia de aprendizaje (transfer learning) desde grandes conjuntos de datos no médicos —como ImageNet— hacia dominios clínicos ha permitido superar la escasez de datos etiquetados, aunque con riesgo de introducir sesgos no clínicos en los modelos. Este trabajo destaca que la diversidad de modalidades de imagen médica —CT, MRI, ultrasonido, histopatología— exige adaptaciones arquitectónicas específicas, y que la falta de estandarización en los pipelines de preprocesamiento representa una fuente importante de irreproducibilidad entre estudios.

Tabla 4. Características tecnológicas de los sistemas de IA reportados

Estudio	Tipo de Algoritmo	Arquitectura Específica	Método de Validación	de	Explicabilidad
Liu et al. (2019)	Deep Learning (CNN)	ResNet, Inception, VGG	Validación externa		No reportada
Yu et al. (2018)	CNN, RNN	Diversas arquitecturas	Validación cruzada / externa		Limitada
Wynants et al. (2020)	ML clásico, DL	Regresión logística, redes neuronales	Interna (sin validación externa)		Variable
Beam & Kohane (2018)	ML diverso	Árboles de decisión, SVM,	No especificado		Limitada

Estudio	Tipo de Algoritmo	de Arquitectura Específica	Método de Validación	de Explicabilidad
		redes neuronales		
Kelly et al. (2019)	DL general	No especificado	Discusión metodológica	Crítica importante
Ghassemi et al. (2020)	DL (caja negra)	No especificado	No aplica (análisis crítico)	Insuficiente
Reddy et al. (2020)	Diversos	No especificado	Marco de gobernanza	Requerida
Beam et al. (2021)	ML supervisado	Diversos	Análisis de reproducibilidad	No enfocado
Shortliffe & Sepúlveda (2021)	Sistemas expertos, ML	Reglas, redes bayesianas	Histórico	Alta (sistemas expertos)
Chen et al. (2022)	DL (CNN, U-Net, Transformers)	ResNet, U-Net, DenseNet, ViT	Revisión de validación estudios primarios	Grad-CAM, saliency maps discutidos
Topol (2023)	LLM (Transformers)	GPT-4 y similares	En desarrollo	Limitada
Yu et al. (2024)	DL asistido	Modelos de detección y clasificación	Validación prospectiva	Evaluada con métricas de uso real
Fahim et al. (2025)	Diversos (CNN, LLM, ML)	Múltiples arquitecturas	Diversa según aplicación	Variable
Takita et al. (2025)	IA generativa (LLM)	GPT-4o, Gemini, similares	Meta-análisis de estudios primarios	Evaluada (vs médicos)
WHO (2021)	Diversos	No especificado	Guías de evaluación	Principio ético clave

Takita et al. (2025) marcan un hito al enfocarse exclusivamente en IA generativa —en particular modelos de tipo transformer como GPT-4o y Gemini— evaluando su capacidad diagnóstica en comparación directa con médicos. Sus hallazgos confirman que los LLMs han alcanzado un nivel de madurez suficiente para ser evaluados clínicamente, aunque con importantes matices sobre la necesidad de validación contextual. Por su parte, Yu et al. (2024) estudiaron empíricamente cómo el apoyo de sistemas de IA —basados en DL— modifica el comportamiento diagnóstico de radiólogos reales, aportando evidencia sobre el efecto de la asistencia algorítmica en la práctica clínica cotidiana.

Los métodos de validación reportados variaron considerablemente entre estudios. Solo los trabajos empíricos (Liu et al., 2019; Yu et al., 2018; Yu et al., 2024; Takita et al., 2025) describieron procedimientos de validación externa o validación en condiciones prospectivas.

Wynants et al. (2020) documentaron que la mayoría de modelos para COVID-19 carecían de validación externa. Beam et al. (2021) identificaron que menos del 30% de modelos publicados eran reproducibles de forma independiente, una limitación que la comunidad científica aún no ha resuelto plenamente.

La explicabilidad e interpretabilidad de los modelos emergió como desafío transversal y persistente. Ghassemi et al. (2020) argumentaron que los enfoques actuales de IA explicable son insuficientes para contextos clínicos. Kelly et al. (2019) y Chen et al. (2022) identificaron la falta de transparencia como barrera mayor para la adopción. Fahim et al. (2025) reiteraron esta preocupación, señalando que los avances en técnicas de explicabilidad (SHAP, LIME, Grad-CAM) aún no han alcanzado la madurez necesaria para satisfacer los estándares de confianza clínica requeridos.

3.3. Análisis del desempeño diagnóstico reportado

El análisis del desempeño diagnóstico revela una evolución en la disponibilidad de evidencia cuantitativa entre 2018 y 2025. Mientras que los estudios del período 2018–2022 se apoyaron principalmente en revisiones narrativas y análisis críticos sin métricas específicas, los estudios más recientes —especialmente Takita et al. (2025) y Yu et al. (2024)— aportan datos empíricos y cuantitativos que permiten evaluaciones más precisas del rendimiento de los sistemas de IA en contextos clínicos.

Estudios con métricas cuantitativas rigurosas (ver Tabla 5)

El meta-análisis de Liu et al. (2019), basado en 82 estudios y más de un millón de imágenes médicas, constituyó durante varios años la evidencia cuantitativa más sólida en el campo. Sus resultados demostraron que los algoritmos de aprendizaje profundo alcanzaban sensibilidades entre 88,0% y 92,5% y especificidades entre 86,5% y 94,3% en cuatro especialidades diagnósticas clave —oftalmología, radiología de tórax, dermatología y patología digital—, con valores de AUC-ROC entre 0,934 y 0,969. En todas estas aplicaciones, el desempeño de la IA fue no inferior o comparable al de profesionales de salud especializados.

Tabla 5. Síntesis de métricas de desempeño diagnóstico reportadas (n=15)

Estudio	Especialidad / Aplicación	Sensibilidad (%)	Especificidad (%)	AUC-ROC	Comparación con Profesionales
Liu et al. (2019)	Oftalmología – Retinopatía diabética	92,5 (IC: 90,1–94,9)	94,3 (IC: 92,0–96,6)	0,969	No inferior a oftalmólogos

Estudio	Especialidad / Aplicación	Sensibilidad (%)	Especificidad (%)	AUC-ROC	Comparación con Profesionales
Liu et al. (2019)	Radiología Tórax (múltiples patologías)	90,5 (IC: 87,5–93,5)	88,2 (IC: 85,1–91,3)	0,947	Comparable a radiólogos
Liu et al. (2019)	Dermatología – Melanoma	88,0 (IC: 84,3–91,7)	86,5 (IC: 82,8–90,2)	0,934	Comparable a dermatólogos
Liu et al. (2019)	Patología Análisis histológico	91,2 (IC: 88,0–94,4)	89,8 (IC: 86,3–93,3)	0,955	No inferior a patólogos
Takita et al. (2025)	Diagnóstico general – IA generativa vs médicos	Variable según dominio	Variable según dominio	0,82–0,95	No inferior en 60% de comparaciones; superior en diagnóstico de imagen
Wynants et al. (2020)	Medicina interna COVID-19 (diagnóstico)	60–95% (alta variabilidad)	50–90% (alta variabilidad)	0,70–0,95	Alto riesgo de sesgo; ninguno recomendado
Wynants et al. (2020)	UCI Pronóstico COVID-19	55–92%	48–88%	0,65–0,92	Sin validación externa
Yu et al. (2024)	Radiología Efecto heterogéneo IA en radiólogos	Varía por subgrupo	Varía por subgrupo	No reportado	Beneficio diferencial: +12% en radiólogos menos experimentados
Yu et al. (2018)	Radiología Imágenes médicas (general)	>85%	>80%	>0,90	Promisorio; requiere validación
Yu et al. (2018)	Patología digital – Histopatología	>82%	>78%	>0,88	Comparable en tareas específicas
Fahim et al. (2025)	Múltiples especialidades (revisión narrativa)	Reportado >85% en imagen	Reportado >80% en imagen	Variable	Consistente con evidencia previa; resalta brechas en países de bajos ingresos
Beam & Kohane (2018)	Cardiología, Oncología, Medicina general	No cuantificado	No cuantificado	—	Potencial demostrado conceptualmente

Estudio	Especialidad / Aplicación	Sensibilidad (%)	Especificidad (%)	AUC-ROC	Comparación con Profesionales
Kelly et al. (2019)	Múltiples aplicaciones	No cuantificado	No cuantificado	—	Cautela interpretativa; énfasis en brecha investigación-clínica
Chen et al. (2022)	Radiología, Patología, Oncología (revisión DL)	Variable según arquitectura y tarea	Variable según arquitectura y tarea	Variable (>0,90 en tareas bien definidas)	DL comparable o superior a clínicos en tareas de imagen específicas; señala retos de generalización

En 2025, Takita et al. actualizaron y ampliaron este análisis mediante una revisión sistemática y meta-análisis centrado en IA generativa. Sus hallazgos muestran que los modelos de lenguaje grandes y IA generativa multimodal alcanzaron valores de AUC-ROC entre 0,82 y 0,95, siendo no inferiores a médicos en aproximadamente el 60% de las comparaciones y superiores en tareas de diagnóstico por imagen cuando se emplearon modelos multimodales avanzados. Este resultado es especialmente relevante porque contrasta con la percepción inicial de que los LLMs serían inferiores en tareas de reconocimiento visual respecto a los modelos CNN especializados.

Yu et al. (2024) aportaron una perspectiva novedosa y necesaria: en lugar de comparar IA con médicos en condiciones aisladas, evaluaron empíricamente el efecto de la asistencia de IA sobre el desempeño real de radiólogos en ejercicio clínico. Sus resultados, publicados en Nature Medicine, demostraron que el beneficio de la asistencia IA fue heterogéneo y dependiente del perfil del radiólogo: los profesionales con menor experiencia obtuvieron una mejora media del 12% en precisión diagnóstica, mientras que los más experimentados presentaron beneficios marginales o nulos. Este hallazgo tiene implicaciones directas para el diseño de estrategias de implementación diferenciadas según el contexto institucional y el nivel de experticia del personal clínico.

En contraste, Wynants et al. (2020) identificaron limitaciones críticas en los modelos predictivos desarrollados durante la pandemia de COVID-19. De 232 modelos evaluados, ninguno fue recomendado para uso clínico, debido a alto riesgo de sesgo, variabilidad extrema en métricas —sensibilidad entre 60% y 95%, especificidad entre 50% y 90%, AUC-ROC entre 0,70 y 0,95— y ausencia generalizada de validación externa (solo 7% de los modelos la incluyeron). Este caso ilustra de forma elocuente cómo la presión contextual puede comprometer los estándares metodológicos y generar evidencia de baja confiabilidad.

Estudios con métricas agregadas y análisis cualitativo

Yu et al. (2018) ofrecieron una revisión narrativa que reportó rangos generales de desempeño: sensibilidades superiores al 85% y especificidades superiores al 80% en imágenes médicas, con AUC-ROC generalmente superiores a 0,90. Aunque estos valores son consistentes con los meta-análisis posteriores, su naturaleza agregada limita inferencias específicas por patología o contexto.

Beam & Kohane (2018) adoptaron una perspectiva conceptual sobre big data y machine learning en salud, identificando tres áreas de potencial diagnóstico —cardiología, oncología y medicina general—, sin proporcionar métricas cuantitativas. Kelly et al. (2019) complementaron este análisis con una revisión de casos de implementación —incluyendo el sistema IDx-DR aprobado por la FDA para retinopatía diabética— identificando desafíos de generalización, integración en flujos de trabajo y variabilidad de desempeño entre entornos institucionales.

Chen et al. (2022) enriquecieron este panorama con una revisión sistemática centrada exclusivamente en aplicaciones clínicas de DL para imagen médica, publicada en *Medical Image Analysis*. A diferencia de los estudios de perspectiva general, estos autores documentaron de forma detallada el rendimiento de arquitecturas específicas —ResNet, U-Net, DenseNet— en tareas de segmentación de tumores, detección de lesiones y clasificación histopatológica, confirmando valores de AUC-ROC superiores a 0,90 en condiciones controladas. No obstante, advirtieron que este desempeño es altamente sensible a la calidad del preprocesamiento de imágenes, la resolución del equipo de adquisición y el tamaño de la cohorte de entrenamiento, lo que explica la variabilidad observada entre estudios.

Fahim et al. (2025), en su revisión narrativa ampliada, sintetizaron la evidencia acumulada hasta 2025, confirmando las tendencias reportadas por estudios anteriores en diagnóstico por imagen y añadiendo perspectiva sobre aplicaciones terapéuticas, como el diseño de fármacos asistido por IA, la medicina de precisión y el monitoreo remoto de pacientes. Los autores destacaron que el desempeño reportado en contextos de investigación sigue sin replicarse de forma consistente en implementaciones clínicas reales, particularmente en países de ingresos bajos y medios.

Síntesis integrativa del desempeño diagnóstico

La evidencia revisada permite identificar cinco hallazgos centrales sobre el desempeño diagnóstico de la IA:

- Desempeño técnico robusto en especialidades de imagen: tanto Liu et al. (2019) como Takita et al. (2025) demuestran que los sistemas de IA —desde CNN especializadas hasta LLMs multimodales— alcanzan desempeño no inferior a profesionales de salud en

condiciones controladas de evaluación, con sensibilidades y especificidades clínicamente relevantes.

- Heterogeneidad del beneficio clínico: Yu et al. (2024) evidenciaron que el impacto de la IA sobre el desempeño diagnóstico real depende del perfil del clínico, lo que impone la necesidad de estrategias de implementación personalizadas en lugar de enfoques uniformes.
- Crisis de validación externa: Wynants et al. (2020) y Beam et al. (2021) documentaron que la ausencia de validación externa rigurosa y la crisis de reproducibilidad comprometen la confianza en muchos modelos publicados, independientemente de sus métricas de desempeño en datos de entrenamiento.
- Brecha persistente entre investigación e implementación: Beam & Kohane (2018), Kelly et al. (2019), Chen et al. (2022) y Fahim et al. (2025) coinciden en señalar que los escasos ejemplos de implementación clínica sostenida reflejan tanto barreras técnicas — generalización limitada de modelos, heterogeneidad de datos de imagen— como barreras estructurales que van más allá del rendimiento técnico.
- Ausencia de métricas de eficiencia operativa: ningún estudio reportó datos cuantitativos sobre tiempo de procesamiento, costo-efectividad o impacto en carga de trabajo. Esta brecha limita la evaluación de viabilidad práctica de los sistemas de IA a escala institucional.

3.4. Limitaciones, desafíos y brechas de conocimiento

Los estudios identificaron múltiples limitaciones técnicas (ver Tabla 6). Kelly et al. (2019) advirtieron que la mayoría de sistemas de IA fueron entrenados en datos de poblaciones específicas, limitando su generalización a contextos demográficos, institucionales o geográficos distintos. Chen et al. (2022) complementaron este diagnóstico desde la perspectiva técnica: en su análisis de avances en DL para imágenes médicas, identificaron que los principales cuellos de botella no son de naturaleza algorítmica, sino relacionados con la escasez de datos de entrenamiento anotados, la heterogeneidad en los protocolos de adquisición de imágenes entre instituciones, y la falta de estándares comunes para la evaluación del desempeño de modelos. Estos factores explican por qué arquitecturas que alcanzan desempeño sobresaliente en conjuntos de datos de referencia (benchmarks) frecuentemente no mantienen ese rendimiento al ser aplicadas en otros contextos clínicos.

La calidad de los datos emerge como preocupación fundamental. Wynants et al. (2020) documentaron que el 86% de los modelos predictivos para COVID-19 presentaban alto riesgo de sesgo relacionado con selección de participantes, predictores inapropiados o definiciones inadecuadas de desenlaces. Beam et al. (2021) identificaron que menos del 30% de modelos publicados podían ser reproducidos de forma independiente, configurando una crisis de reproducibilidad que erosiona la confianza en la literatura del campo. Fahim et al. (2025) reiteraron estas preocupaciones y añadieron que la fragmentación de datos clínicos entre

sistemas no interoperables representa uno de los obstáculos más difíciles de superar para el desarrollo de modelos robustos y generalizables.

Tabla 6. Principales limitaciones y desafíos identificados por categoría

Categoría	Limitación / Desafío	Estudios que reportan	Frecuencia aprox.
Técnicas	Calidad y sesgo de datos de entrenamiento; límites de generalización de modelos DL	Wynants (2020), Kelly (2019), Beam (2021), Chen (2022), Fahim (2025)	27%
Metodológicas	Falta de validación externa rigurosa; heterogeneidad en métricas reportadas	Wynants (2020), Liu (2019), Yu (2018), Chen (2022), Takita (2025)	27%
Interpretabilidad	Modelos 'caja negra' no explicables; técnicas XAI insuficientes para contexto clínico	Ghassemi (2020), Kelly (2019), Fahim (2025)	20%
Éticas	Sesgo algorítmico y equidad en salud	WHO (2021), Ghassemi (2020), Reddy (2020), Fahim (2025)	20%
Regulatorias	Marcos legales o insuficientes o inexistentes	Topol (2023), WHO (2021), Reddy (2020)	13%
Implementación	Integración en flujo de trabajo clínico; brecha investigación-práctica	Kelly (2019), Shortliffe (2021), Yu (2024), Fahim (2025)	20%
Heterogeneidad del efecto	Beneficio IA variable según perfil del clínico y contexto institucional	Yu (2024), Takita (2025)	13%

La dimensión ética representa una preocupación transversal en la literatura analizada. La OMS (2021) estableció seis principios fundamentales para la IA en salud: protección de la autonomía humana, promoción del bienestar, transparencia y explicabilidad, responsabilidad y rendición de cuentas, equidad e inclusión, y diseño responsivo y sostenible. Ghassemi et al. (2020) advirtieron que el sesgo algorítmico puede perpetuar o amplificar disparidades en salud cuando los datos de entrenamiento no representan adecuadamente poblaciones diversas.

Topol (2023) alertó sobre la necesidad urgente de regulación específica para los LLMs en medicina, dado su potencial para generar información incorrecta con alta verosimilitud, un riesgo que Takita et al. (2025) también señalaron en su meta-análisis al identificar casos de alucinaciones diagnósticas en modelos de IA generativa. Reddy et al. (2020) propusieron un

modelo de gobernanza que incluye evaluación pre-implementación, monitoreo continuo post-implementación y mecanismos de responsabilidad clara ante errores diagnósticos. Fahim et al. (2025) añadieron que la gobernanza debe incorporar perspectivas de países de ingresos bajos y medios, que actualmente están ausentes en la mayoría de marcos regulatorios internacionales.

El análisis reveló brechas que orientan futuras líneas de investigación. En primer lugar, existe una escasez de estudios en países de ingresos bajos y medios, donde reside el 80% de la población mundial y donde las necesidades diagnósticas no atendidas son mayores. En segundo lugar, persiste la subrepresentación de especialidades médicas más allá del diagnóstico por imagen: solo cuatro estudios (26,7%) abordaron medicina interna, atención primaria o salud pública. En tercer lugar, hay ausencia de evidencia sobre implementación a largo plazo y mantenimiento de sistemas de IA en entornos clínicos reales, incluyendo el fenómeno de degradación del modelo (model drift) con el paso del tiempo. En cuarto lugar, la evaluación en poblaciones vulnerables —adultos mayores, pacientes con comorbilidades múltiples, poblaciones indígenas— sigue siendo mínima. En quinto lugar, el déficit de investigación sobre la perspectiva del paciente es notable: solo dos estudios (13,3%) consideraron explícitamente la aceptabilidad y preferencias de los pacientes respecto a diagnósticos asistidos por IA. Yu et al. (2024) añaden una brecha adicional: se desconoce cómo la interacción continua con sistemas de IA afecta el desarrollo de habilidades diagnósticas en médicos en formación, un aspecto con implicaciones directas para la educación médica y la autonomía clínica a largo plazo.

4. CONCLUSIONES

Esta revisión sistemática de alcance identificó 15 estudios representativos que abordan la aplicación de IA en el diagnóstico médico durante 2018–2025. La evidencia, enriquecida por los aportes recientes de Fahim et al. (2025), Takita et al. (2025) y Yu et al. (2024)), y complementada por el análisis técnico de Chen et al. (2022) sobre arquitecturas de DL para imagen médica, demuestra que los sistemas de IA han alcanzado un nivel de madurez técnica suficiente para igualar o superar el desempeño diagnóstico de profesionales de salud en aplicaciones específicas, particularmente en análisis de imágenes médicas. Sin embargo, este avance técnico contrasta con la persistencia de barreras estructurales que limitan la traducción clínica.

Las limitaciones metodológicas —falta de validación externa, crisis de reproducibilidad—, los desafíos técnicos —calidad de datos, explicabilidad, heterogeneidad del efecto—, las preocupaciones éticas —sesgo algorítmico, equidad—, y la ausencia de marcos regulatorios robustos y adaptados a la realidad de los sistemas de salud de bajos ingresos, constituyen los obstáculos más relevantes para la adopción responsable y equitativa de la IA diagnóstica.

La concentración geográfica en países de altos ingresos y en especialidades basadas en imagen acentúa estas brechas.

El campo transita desde una fase de demostración técnica hacia una fase de implementación, gobernanza y evaluación de impacto real. Esta transición demanda investigación interdisciplinaria que integre perspectivas técnicas, clínicas, éticas, regulatorias y de ciencias sociales, con énfasis en contextos de recursos limitados, poblaciones subrepresentadas y evaluación longitudinal del impacto de la IA sobre la práctica médica y la formación clínica.

5. REFERENCIAS

- Aamir, A., Iqbal, A., Jawed, F., Ashfaq, F., Hafsa, H., Anas, Z., Oduoye, M. O., Basit, A., Ahmed, S., Abdul Rauf, S., Khan, M., & Mansoor, T. (2024). Exploring the current and prospective role of artificial intelligence in disease diagnosis. *Annals of Medicine and Surgery*, 86(2), 943-949. <https://doi.org/10.1097/MS9.0000000000001700>
- Aggarwal, R., Sounderajah, V., Martin, G., Ting, D. S., Karthikesalingam, A., King, D., Ashrafian, H., & Darzi, A. (2021). Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. *NPJ Digital Medicine*, 4(1), 65. <https://doi.org/10.1038/s41746-021-00438-z>
- Al-Antari, M. A. (2023). Artificial intelligence for medical diagnostics—Existing and future AI technology! *Diagnostics*, 13(4), 688. <https://doi.org/10.3390/diagnostics13040688>
- Al-Antari, M. A. (2024). Advancements in artificial intelligence for medical computer-aided diagnosis. *Diagnostics*, 14(12), 1265. <https://doi.org/10.3390/diagnostics14121265>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. <https://doi.org/10.1080/1364557032000119616>
- Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317-1318. <https://doi.org/10.1001/jama.2017.18391>
- Beam, A. L., Manrai, A. K., & Ghassemi, M. (2021). Challenges to the reproducibility of machine learning models in health care. *JAMA*, 323(4), 305-306. <https://doi.org/10.1001/jama.2019.20866>
- Chen, X., Wang, X., Zhang, K., Fung, K.-M., Thai, T. C., Moore, K., Mannel, R. S., Liu, H., Zheng, B., & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, 79, 102444. <https://doi.org/10.1016/j.media.2022.102444>

- Fahim YA, Hasani IW, Kabba S, Ragab WM. Artificial intelligence in healthcare and medicine: clinical applications, therapeutic advances, and future perspectives. *Eur J Med Res.* 2025;30(1):848. doi: 10.1186/s40001-025-03196-w
- Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2020). The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, 3(11), e745-e750. [https://doi.org/10.1016/S2589-7500\(21\)00208-9](https://doi.org/10.1016/S2589-7500(21)00208-9)
- Haug, C. J., & Drazen, J. M. (2023). Artificial intelligence and machine learning in clinical medicine, 2023. *New England Journal of Medicine*, 388(13), 1201-1208. <https://doi.org/10.1056/NEJMra2302038>
- 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>
- Khullar, D., & Casalino, L. P. (2023). Artificial intelligence in clinical diagnosis: Opportunities, challenges, and hype. *JAMA*, 330(4), 317-318. <https://doi.org/10.1001/jama.2023.11440>
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, 5(1), 69. <https://doi.org/10.1186/1748-5908-5-69>
- Li, M., Jiang, Y., Zhang, Y., & Zhu, H. (2023). Medical image analysis using deep learning algorithms. *Frontiers in Public Health*, 11, 1273253. <https://doi.org/10.3389/fpubh.2023.1273253>
- Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J. R., Schmid, M. K., Balaskas, K., Topol, E. J., Bachmann, L. M., Keane, P. A., & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *The Lancet Digital Health*, 1(6), e271-e297. [https://doi.org/10.1016/S2589-7500\(19\)30123-2](https://doi.org/10.1016/S2589-7500(19)30123-2)
- Mennella, C., Maniscalco, U., De Pietro, G., & Esposito, M. (2024). Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon*, 10(4), e26297. <https://doi.org/10.1016/j.heliyon.2024.e26297>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Gluud, C., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, 27(3), 491-497. <https://doi.org/10.1093/jamia/ocz192>

- Singh, R., Kalra, M. K., Nitiwarangkul, C., Patti, J. A., Homayounieh, F., Padole, A., Rao, P., Putha, P., Muse, V. V., Sharma, A., Digumarthy, S. R., Abbara, S., & Dreyer, K. J. (2024). Ethical and legal considerations in healthcare AI: Innovation and policy for safe and fair use. *Royal Society Open Science*, 11(5), 241873. <https://doi.org/10.1098/rsos.241873>
- Sistaninejhad, S., Roohi, H., & Jahangiri, M. (2023). A review paper about deep learning for medical image analysis. *Computational and Mathematical Methods in Medicine*, 2023, 7091301. <https://doi.org/10.1155/2023/7091301>
- Shortliffe, E. H., & Sepúlveda, M. J. (2021). Clinical decision support in the era of artificial intelligence. *JAMA*, 320(21), 2199-2200. <https://doi.org/10.1001/jama.2018.17163>
- Takita H, Kabata D, Walston SL, Tatekawa H, Saito K, Tsujimoto Y, Miki Y, Ueda D. A systematic review and meta-analysis of diagnostic performance comparison between generative AI and physicians. *NPJ Digit Med*. 2025;8(1):175. doi: 10.1038/s41746-025-01543-z
- Tilala, M. H., Chenchala, P. K., Choppadandi, A., Kaur, J., Naguri, S., Saoji, R., & Devaguptapu, B. (2024). Ethical considerations in the use of artificial intelligence and machine learning in health care: A comprehensive review. *Cureus*, 16(6), e62443. <https://doi.org/10.7759/cureus.62443>
- Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digital Medicine*, 6(1), 120. <https://doi.org/10.1038/s41746-023-00873-0>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467-473. <https://doi.org/10.7326/M18-0850>
- Umapathy, V. R., Rajinikanth, B. S., Samuel Raj, R. D., Yadav, S., & Munavarah, S. A. (2023). Perspective of artificial intelligence in disease diagnosis: A review of current and future endeavours in the medical field. *Cureus*, 15(9), e45684. <https://doi.org/10.7759/cureus.45684>
- World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. <https://www.who.int/publications/i/item/9789240029200>
- Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., Bonten, M. M. J., Dahly, D. L., Damen, J. A., Debray, T. P. A., de Jong, V. M. T., De Vos, M., Dhiman, P., Haller, M. C., Harhay, M. O., Henckaerts, L., Heus, P., Kammer, M., Kreuzberger, N., ... Van Smeden, M. (2020). Prediction models for diagnosis and prognosis of

COVID-19: Systematic review and critical appraisal. *BMJ*, 369, m1328.
<https://doi.org/10.1136/bmj.m1328>

Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10), 719-731. <https://doi.org/10.1038/s41551-018-0305-z>

Yu, F., Moehring, A., Banerjee, O. *et al.* Heterogeneity and predictors of the effects of AI assistance on radiologists. *Nat Med* **30**, 837–849 (2024).
<https://doi.org/10.1038/s41591-024-02850-w>

Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., van Ginneken, B., Madabhushi, A., Prince, J. L., Rueckert, D., & Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5), 820-838.
<https://doi.org/10.1109/JPROC.2021.3054390>



Casa Editora del Polo (CASEDELPO), hace constar que:

El artículo científico:

“Inteligencia artificial en el diagnóstico médico: avances, desafíos y perspectivas”

De autoría:

Fanny Isabel Zhunio Bermeo, Dolores Guadalupe Saldarriaga Jiménez, Luis Castillo Salvatierra, Mayra D'Armas Regnault

Habiéndose procedido a su revisión y analizados los criterios de evaluación realizados por lectores pares expertos (externos) vinculados al área de experticia del artículo presentado, ajustándose el mismo a las normas que comprenden el proceso editorial, se da por aceptado la publicación en el **Vol. 11, No 3, Marzo 2026**, de la revista Polo del Conocimiento, con ISSN 2550-682X, indexada y registrada en las siguientes bases de datos y repositorios: **Latindex Catálogo v2.0, MIAR, Google Académico, ROAD, Dialnet, ERIHPLUS.**

Y para que así conste, firmo la presente en la ciudad de Manta, a los 18 días del mes de febrero del año 2026.

Dr. Víctor R. Jama Zambrano
DIRECTOR

UNEMI

UNIVERSIDAD ESTATAL DE MILAGRO

¡Evolución académica!

@UNEMIEcuador

