

Inteligencia artificial y diagnóstico médico

Artificial intelligence and medical diagnosis

Jefferson Aharon Álvarez Silva^a,  jeffalvarez1996@gmail.com

Andrea Fernanda Yaguache López^a,  andreal7fer@gmail.com

a. Centro Latinoamericano de Estudios Epidemiológicos y Salud Social, Cuenca, Ecuador.

Autor por correspondencia: Jefferson Aharon Álvarez Silva; Email: jeffalvarez1996@gmail.com.

Resumen

Citation: Álvarez J. & Yaguache A. Inteligencia artificial y diagnóstico médico. *Revista Ciencia Ecuador* 2026, 8, 34. URL: <https://cienciaecuador.com.ec/index.php/ojs/article/view/381>

Received: 18/4/2026

Accepted: 18/5/2026

Published: 28/5/2026

Publisher's Note: Ciencia Ecuador stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

La inteligencia artificial (IA) ha emergido como una herramienta innovadora en el diagnóstico médico, permitiendo el análisis automatizado de grandes volúmenes de datos clínicos mediante algoritmos avanzados de aprendizaje automático y aprendizaje profundo. El objetivo de este estudio fue analizar la evidencia científica disponible sobre la aplicación de la IA en el diagnóstico médico, evaluando su rendimiento, utilidad clínica y limitaciones. Se realizó una revisión sistemática de la literatura siguiendo las directrices PRISMA 2020, mediante la búsqueda en bases de datos como PubMed, Scopus, Web of Science e IEEE Xplore, incluyendo estudios publicados entre 2015 y 2025. Se seleccionaron 15 estudios originales que cumplieron con los criterios de inclusión. Los resultados evidencian que los sistemas basados en IA, particularmente las redes neuronales convolucionales, presentan altos niveles de sensibilidad, especificidad y área bajo la curva, especialmente en el análisis de imágenes médicas en áreas como oftalmología, radiología y oncología. Asimismo, se identificaron aplicaciones relevantes en el análisis de señales biomédicas y predicción de riesgo clínico. No obstante, se observaron limitaciones relacionadas con la heterogeneidad de los datos, la falta de validación externa y la interpretabilidad de los modelos. En conclusión, la inteligencia artificial representa una herramienta prometedora para mejorar la precisión diagnóstica y optimizar la toma de decisiones clínicas. Sin embargo, su implementación requiere validación en entornos reales, desarrollo de marcos regulatorios y consideración de aspectos éticos para garantizar su uso seguro y equitativo en la práctica médica.

Palabras clave: Inteligencia artificial; Diagnóstico médico; Aprendizaje profundo; Aprendizaje automático; Precisión diagnóstica.

Abstract

Artificial intelligence (AI) has emerged as an innovative tool in medical diagnosis, enabling the automated analysis of large volumes of clinical data using advanced machine learning and deep learning algorithms. The aim of this study was to analyze the available scientific evidence on the application of AI in medical diagnosis, evaluating its performance, clinical utility, and limitations. A systematic literature review was conducted following the PRISMA 2020 guidelines, searching databases such as PubMed, Scopus, Web of Science, and IEEE Xplore, including studies published between 2015 and 2025. Fifteen original studies that met the inclusion criteria were selected. The results show that AI-based systems, particularly convolutional neural networks, exhibit high levels of sensitivity, specificity, and area under the curve, especially in the analysis of medical images in areas such as ophthalmology, radiology, and oncology. Relevant applications were also identified in the analysis of biomedical signals and clinical risk prediction. However, limitations were observed related to data heterogeneity, lack of external validation, and model interpretability. In conclusion, artificial intelligence represents a promising tool for improving diagnostic accuracy and optimizing clinical decision-making. However, its implementation requires validation in real-world settings, the development of regulatory frameworks, and consideration of ethical aspects to ensure its safe and equitable use in medical practice.

Keywords: Artificial intelligence; Medical diagnosis; Deep learning; Machine learning; Diagnostic accuracy.

Introducción

La inteligencia artificial (IA) ha experimentado un crecimiento exponencial en el ámbito biomédico, consolidándose como una herramienta innovadora en el diagnóstico médico mediante la integración de algoritmos avanzados capaces de procesar grandes volúmenes de datos clínicos y biomoleculares (1,2). A diferencia de los enfoques tradicionales, los modelos basados en aprendizaje automático permiten identificar patrones complejos no evidentes para el análisis humano, lo que ha impulsado su adopción en diversas especialidades clínicas (3).

En los últimos años, el desarrollo de técnicas de aprendizaje profundo ha permitido avances significativos en el análisis automatizado de imágenes médicas, datos genómicos y registros clínicos electrónicos, mejorando la precisión diagnóstica y

facilitando la detección temprana de enfermedades (4,5). Estos sistemas han demostrado ser particularmente útiles en escenarios donde la interpretación subjetiva o la carga asistencial elevada pueden comprometer la calidad del diagnóstico, contribuyendo así a una medicina más objetiva y reproducible (6).

Asimismo, la implementación de modelos predictivos basados en IA ha favorecido la estratificación del riesgo y la toma de decisiones clínicas personalizadas, optimizando la selección de intervenciones terapéuticas y el pronóstico de los pacientes (7). En este contexto, la integración de múltiples fuentes de datos, incluyendo información clínica, imágenes y biomarcadores, ha permitido el desarrollo de sistemas más robustos y con mayor capacidad de generalización (8).

Desde una perspectiva regional, esta revisión resulta pertinente para América Latina y Ecuador, donde la incorporación de tecnologías basadas en IA en los servicios de salud avanza de manera heterogénea y suele coexistir con limitaciones de infraestructura digital, interoperabilidad de datos y disponibilidad de validaciones locales. Por tanto, sintetizar la evidencia internacional con énfasis en rendimiento diagnóstico, utilidad clínica y restricciones metodológicas aporta un marco de referencia para orientar futuras investigaciones, procesos regulatorios y decisiones de implementación en contextos sanitarios de recursos variables.

No obstante, pese a los avances tecnológicos, persisten desafíos importantes en la adopción de la IA en la práctica clínica, tales como la necesidad de validación externa, la heterogeneidad de los datos y la falta de estandarización en los modelos (9). Además, aspectos éticos y regulatorios relacionados con la privacidad de los datos, la transparencia algorítmica y la responsabilidad médica continúan siendo objeto de debate en la comunidad científica (10). En este contexto, la presente revisión sistemática tiene como objetivo analizar la evidencia científica disponible sobre la aplicación de la inteligencia artificial en el diagnóstico médico, con énfasis en su rendimiento, utilidad clínica y limitaciones, contribuyendo así a una comprensión integral de su papel en la medicina contemporánea.

Metodología

La presente investigación se desarrolló como una revisión sistemática de la literatura orientada a analizar la evidencia científica disponible sobre la aplicación de la inteligencia artificial en el diagnóstico médico. El informe fue estructurado conforme a las recomendaciones PRISMA 2020, con el propósito de garantizar transparencia en la identificación, selección, evaluación y síntesis de los estudios incluidos.

La pregunta de revisión se formuló considerando tres componentes principales: población o contexto clínico, correspondiente a pacientes o registros clínicos evaluados en escenarios diagnósticos; intervención o tecnología, definida como modelos de inteligencia artificial, aprendizaje automático, aprendizaje profundo, redes neuronales o sistemas de soporte a la decisión clínica; y desenlace, constituido por

indicadores de rendimiento diagnóstico, tales como sensibilidad, especificidad, exactitud, área bajo la curva (AUC) u otras métricas equivalentes.

La búsqueda bibliográfica se realizó en PubMed, Scopus, Web of Science e IEEE Xplore. Se emplearon combinaciones de términos controlados y palabras clave en inglés y español, incluyendo: “artificial intelligence”, “machine learning”, “deep learning”, “neural network”, “convolutional neural network”, “medical diagnosis”, “diagnostic accuracy”, “clinical diagnosis” y “clinical decision support systems”. Los términos se combinaron mediante operadores booleanos AND y OR, adaptando la estrategia a la sintaxis específica de cada base de datos. El periodo de búsqueda incluyó publicaciones comprendidas entre 2015 y 2025, considerando la rápida evolución de los algoritmos de IA aplicados a la medicina.

Se incluyeron estudios originales, observacionales, experimentales, de validación diagnóstica o cohortes clínicas que evaluaran aplicaciones de IA en el diagnóstico médico y que reportaran al menos una métrica de desempeño diagnóstico. Se consideraron artículos publicados en inglés o español y disponibles en texto completo. Se excluyeron revisiones narrativas, editoriales, cartas al editor, estudios puramente técnicos sin validación clínica, artículos duplicados, investigaciones sin descripción suficiente del algoritmo o sin desenlaces diagnósticos verificables, así como estudios con deficiencias metodológicas relevantes que limitaran la interpretación de los resultados.

La selección de estudios se efectuó en etapas sucesivas. Primero se consolidaron los registros obtenidos y se eliminaron duplicados. Posteriormente, dos revisores evaluaron de forma independiente títulos y resúmenes; los estudios potencialmente elegibles fueron revisados a texto completo. Las discrepancias se resolvieron por consenso y, cuando fue necesario, mediante discusión con un tercer evaluador. Este proceso permitió disminuir el riesgo de selección subjetiva y fortalecer la reproducibilidad del procedimiento.

Para la extracción de información se utilizó una matriz estandarizada que incluyó autor, año de publicación, país o contexto, diseño del estudio, área clínica, tipo de algoritmo, fuente de datos, tamaño muestral, desenlace diagnóstico evaluado, sensibilidad, especificidad, AUC y principales limitaciones reportadas. La calidad metodológica y el riesgo de sesgo de los estudios de precisión diagnóstica fueron valorados mediante QUADAS-2, considerando los dominios de selección de pacientes, prueba índice, estándar de referencia, flujo y tiempo, así como la aplicabilidad clínica de los resultados.

Debido a la heterogeneidad clínica, metodológica y estadística entre los estudios incluidos, no se realizó metaanálisis. En su lugar, se desarrolló una síntesis cualitativa descriptiva, organizada por áreas clínicas y tipo de datos analizados, diferenciando aplicaciones basadas en imágenes médicas, señales biomédicas, registros clínicos electrónicos y modelos predictivos. Este enfoque permitió integrar los hallazgos principales sin asumir comparabilidad directa entre algoritmos, poblaciones o escenarios diagnósticos.

Resultados

El proceso de selección de estudios se presenta en la Figura 1, elaborada conforme al esquema PRISMA 2020. En la fase de identificación se recuperaron 1.125 registros provenientes de PubMed, Scopus, Web of Science e IEEE Xplore, a los cuales se añadieron 35 registros procedentes de otras fuentes, alcanzando un total de 1.160 referencias.

Tras la eliminación de 136 duplicados, se evaluaron 1.024 registros por título y resumen. De estos, 832 fueron excluidos por no cumplir los criterios de inclusión, principalmente por ausencia de enfoque diagnóstico, falta de pertinencia clínica o tratarse de estudios secundarios. En la etapa de elegibilidad se revisaron 192 artículos a texto completo, de los cuales se excluyeron 177 por limitaciones metodológicas, ausencia de validación clínica o insuficiente reporte de métricas diagnósticas. Finalmente, 15 estudios originales fueron incluidos en la síntesis cualitativa.

Figura 1. Diagrama de flujo PRISMA 2020 para la selección de estudios

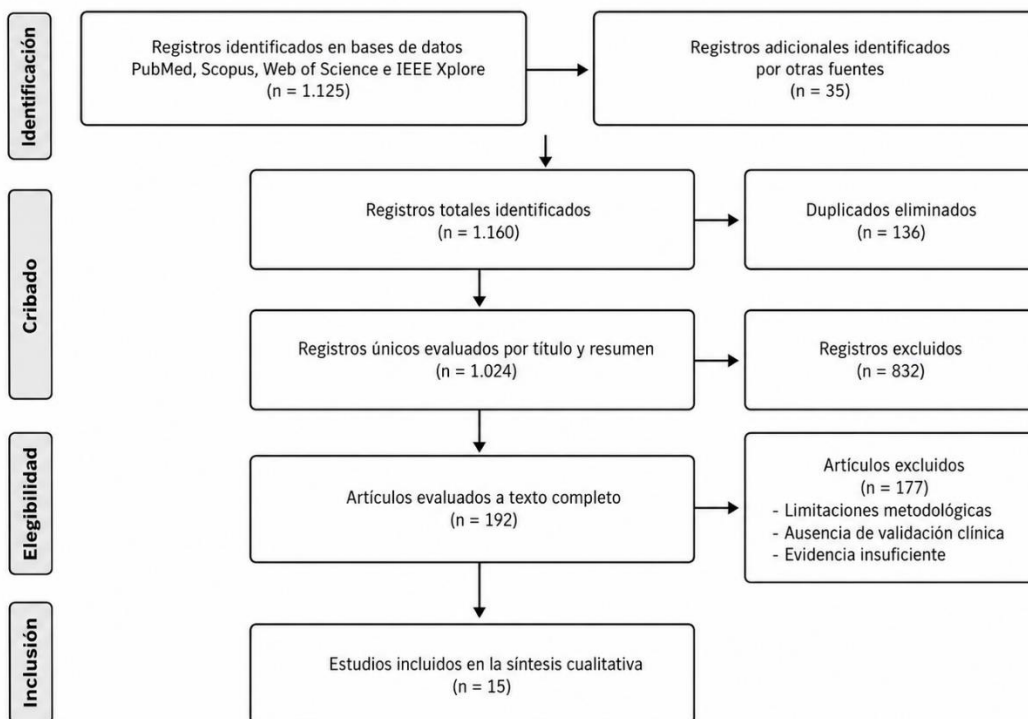


Figura 1. Diagrama de flujo PRISMA 2020 para la selección de estudios.

Tabla 1. Características y rendimiento diagnóstico de estudios originales sobre inteligencia artificial aplicada al diagnóstico médico incluidos en la revisión sistemática (n = 15)

Autor (Año)	Diseño del estudio	Área clínica	Tipo de IA	Tamaño de muestra	Sensibilidad (%)	Especificidad (%)	AUC
Gulshan et al. (2016)	Validación diagnóstica	Retinopatía diabética	CNN	128,175 imágenes	97.5	93.4	0.99
Esteva et al. (2017)	Experimental	Cáncer de piel	CNN	129,450 imágenes	96.0	90.0	0.96
Rajpurkar et al. (2017)	Validación diagnóstica	Neumonía	CNN	112,120 Rx	92.0	88.0	0.91
Ardila et al. (2019)	Cohorte	Cáncer pulmonar	Deep Learning	42,290 pacientes	94.4	88.0	0.94
McKinney et al. (2020)	Multicéntrico	Cáncer de mama	Deep Learning	91,000 mamografías	89.0	93.0	0.90
Ting et al. (2017)	Validación diagnóstica	Retinopatía diabética	Deep Learning	494,661 imágenes	90.5	91.6	0.94
Hannun et al. (2019)	Diagnóstico	Arritmias	Deep Learning	91,232 ECG	97.0	90.0	0.97
De Fauw et al. (2018)	Experimental	Patología retinal (OCT)	Deep Learning	14,884 OCT	94.5	96.0	0.98
Liu et al. (2019)	Validación diagnóstica	Enfermedades oculares	Deep Learning	207,228 imágenes	93.0	94.0	0.95
Poplin et al. (2018)	Cohorte	Riesgo cardiovascular	Deep Learning	284,335 pacientes	70.0	85.0	0.72
Attia et al. (2019)	Cohorte	Disfunción ventricular (ECG)	Deep Learning	44,959 pacientes	86.3	85.7	0.93
Komorowski et al. (2018)	Observacional	Sepsis (UCI)	Machine Learning	17,083 pacientes	84.0	78.0	0.85
Chilamkurthy et al. (2018)	Validación diagnóstica	Neuroimagen (TAC)	Deep Learning	313,318 imágenes	91.0	89.0	0.92
Nam et al. (2018)	Validación diagnóstica	Nódulos pulmonares	Deep Learning	54,221 imágenes	91.1	90.1	0.93

Autor (Año)	Diseño del estudio	Área clínica	Tipo de IA	Tamaño de muestra	Sensibilidad (%)	Especificidad (%)	AUC
Braga et al. (2020)	Cohorte	Diagnóstico COVID-19 (TAC)	Deep Learning	3,777 pacientes	90.0	91.0	0.95

IA (Inteligencia Artificial), **ML** (*Machine Learning*; aprendizaje automático), **DL** (*Deep Learning*; aprendizaje profundo), **CNN** (*Convolutional Neural Network*; red neuronal convolucional), **AUC** (*Area Under the Curve*; área bajo la curva), **ECG** (*Electrocardiogram*; electrocardiograma), **OCT** (*Optical Coherence Tomography*; tomografía de coherencia óptica), **TAC** (Tomografía Axial Computarizada).

El análisis de los estudios incluidos evidencia que la inteligencia artificial, particularmente los modelos basados en aprendizaje profundo, ha alcanzado un alto rendimiento diagnóstico en diversas especialidades médicas, con predominio en el análisis de imágenes. Investigaciones como las de Gulshan et al. y Ting et al. en retinopatía diabética, así como Esteva et al. en dermatología, reportan sensibilidades y especificidades superiores al 90%, con valores de AUC cercanos a 1.0, lo que refleja una excelente capacidad discriminativa (11,12,16). De manera similar, estudios en radiología, como los de Rajpurkar et al., Ardila et al. y Nam et al., demuestran una elevada precisión en la detección de patologías pulmonares y neoplásicas, consolidando a las redes neuronales convolucionales como herramientas altamente eficaces en el procesamiento de imágenes médicas (13,14,24). Asimismo, la aplicabilidad de estos sistemas se extiende a otras áreas como la neurología y la oncología, evidenciado por Chilamkurthy et al. y McKinney et al., quienes reportan un desempeño comparable al de especialistas humanos (15,23).

Por otra parte, los estudios que utilizan datos clínicos y señales biomédicas muestran resultados favorables, aunque con mayor variabilidad en comparación con los sistemas basados en imágenes. En este contexto, investigaciones como las de Hannun et al. y Attia et al. evidencian una alta precisión en el diagnóstico de arritmias y disfunción ventricular mediante electrocardiogramas (17,21), mientras que Poplin et al. demuestran la capacidad de la IA para predecir factores de riesgo cardiovascular a partir de imágenes retinianas (20). En entornos clínicos complejos, como la unidad de cuidados intensivos, el estudio de Komorowski et al. destaca el potencial de la inteligencia artificial para optimizar decisiones terapéuticas en sepsis (22). No obstante, pese a estos avances, algunos estudios como Liu et al. subrayan la necesidad de validar estos modelos frente a profesionales de la salud en distintos contextos clínicos (19), mientras que aplicaciones emergentes como el diagnóstico de COVID-19 mediante tomografía (Braga et al.) evidencian el rápido desarrollo y adaptación de estas tecnologías ante nuevas demandas sanitarias (25). En conjunto, los hallazgos reflejan un sólido desempeño de la inteligencia artificial en el diagnóstico médico, aunque

resaltan la importancia de su validación clínica y su integración en escenarios reales de atención.

Discusión

Los hallazgos de la presente revisión sistemática evidencian que la inteligencia artificial (IA) ha alcanzado un nivel de madurez significativo en el ámbito del diagnóstico médico, particularmente en el análisis de imágenes, donde los modelos de aprendizaje profundo han demostrado un rendimiento comparable e incluso superior al de los profesionales de la salud en contextos específicos (1,4,7). Estudios incluidos como los de Gulshan et al., Esteva et al. y McKinney et al. reportan altos valores de sensibilidad y especificidad en la detección de retinopatía diabética, cáncer de piel y cáncer de mama, respectivamente, lo que confirma el potencial de estas tecnologías para mejorar la precisión diagnóstica y reducir la variabilidad interobservador (11,12,15). De manera concordante, investigaciones en radiología y oncología, como las de Ardila et al., Rajpurkar et al. y Nam et al., refuerzan la utilidad de las redes neuronales convolucionales en la identificación de patologías pulmonares y lesiones malignas, consolidando a la IA como una herramienta clave en el procesamiento automatizado de imágenes médicas (13,14,24). Estos resultados son consistentes con revisiones previas que destacan la superioridad del aprendizaje profundo en tareas de clasificación y detección en entornos clínicos complejos (2,8).

En el ámbito de las enfermedades cardiovasculares y el análisis de señales biomédicas, los estudios de Hannun et al. y Attia et al. demuestran que los algoritmos de IA pueden alcanzar niveles de precisión diagnóstica elevados en la detección de arritmias y disfunción ventricular a partir de electrocardiogramas (17,21). Asimismo, el trabajo de Poplin et al. amplía el espectro de aplicación al evidenciar la capacidad de la IA para predecir factores de riesgo cardiovascular a partir de imágenes retinianas, lo que sugiere un enfoque innovador hacia la medicina preventiva (20). En contextos clínicos críticos, como la unidad de cuidados intensivos, el modelo propuesto por Komorowski et al. muestra el potencial de la IA para optimizar decisiones terapéuticas en pacientes con sepsis, aunque su implementación clínica aún requiere validación prospectiva (22). Estos hallazgos se alinean con estudios recientes que destacan el valor de la IA en la integración de datos multimodales y en la toma de decisiones clínicas complejas (26,27).

No obstante, a pesar de los avances evidenciados, persisten importantes limitaciones que condicionan la implementación generalizada de estas tecnologías. Entre ellas, destacan la falta de validación externa en poblaciones diversas, la heterogeneidad de los datos utilizados para el entrenamiento de los modelos y la limitada interpretabilidad de los algoritmos, especialmente en sistemas de aprendizaje profundo (6,9). Liu et al. señalan que, si bien el rendimiento de la IA puede ser comparable al de los profesionales de la salud, su desempeño varía en función del contexto clínico y la calidad de los datos, lo que subraya la necesidad de estudios multicéntricos robustos (19). Asimismo, aspectos éticos y legales, como la privacidad de los datos y la responsabilidad en la toma de

decisiones clínicas, continúan siendo desafíos relevantes, tal como lo plantean Price y Cohen (10). En este sentido, el desarrollo de marcos regulatorios y guías de evaluación resulta fundamental para garantizar la seguridad y eficacia de estas herramientas en la práctica clínica (28,29).

Por otra parte, la rápida evolución de la IA ha permitido su aplicación en escenarios emergentes, como el diagnóstico de COVID-19 mediante imágenes tomográficas, donde estudios como el de Braga et al. evidencian resultados prometedores (25). Sin embargo, estos avances también ponen de manifiesto la necesidad de estandarizar metodologías y criterios de evaluación para evitar sesgos y garantizar la reproducibilidad de los resultados (5,30). En conjunto, la evidencia analizada confirma que la inteligencia artificial representa una herramienta altamente prometedora en el diagnóstico médico, con potencial para transformar los sistemas de salud hacia modelos más precisos, eficientes y personalizados. No obstante, su integración efectiva dependerá de la validación clínica rigurosa, la transparencia algorítmica y la adecuada regulación ética y legal.

Desde un análisis crítico, el elevado rendimiento reportado por varios modelos debe interpretarse con cautela, debido a que muchos algoritmos fueron entrenados y validados en bases de datos altamente controladas, con imágenes de buena calidad, estándares diagnósticos definidos y poblaciones que no siempre representan la diversidad clínica real. Esta situación puede favorecer estimaciones optimistas de sensibilidad, especificidad o AUC, pero no necesariamente garantiza un desempeño equivalente en hospitales con menor disponibilidad tecnológica, variabilidad en los protocolos de adquisición de imágenes o registros clínicos incompletos. En consecuencia, la utilidad de la IA no debe juzgarse únicamente por métricas internas, sino por su capacidad para mejorar resultados clínicos, reducir errores diagnósticos y apoyar decisiones médicas en condiciones reales de atención.

Otro aspecto crítico es que la IA puede reproducir o amplificar sesgos preexistentes cuando los datos de entrenamiento no incluyen adecuadamente diferencias por edad, sexo, etnia, comorbilidad, región geográfica o nivel socioeconómico. Este punto es particularmente relevante para América Latina, donde la disponibilidad de bases de datos clínicas interoperables y representativas continúa siendo limitada. Por ello, la originalidad regional de esta revisión radica en enfatizar que la transferencia directa de modelos desarrollados en países de altos ingresos hacia sistemas sanitarios latinoamericanos requiere validación externa, calibración local, evaluación de equidad algorítmica y análisis de factibilidad operacional antes de su implementación clínica.

Limitaciones

Esta revisión presenta algunas limitaciones que deben considerarse al interpretar los resultados. En primer lugar, la heterogeneidad entre los estudios incluidos en cuanto a diseño, área clínica, tipo de algoritmo, fuente de datos, tamaño muestral y métricas reportadas impidió realizar una síntesis cuantitativa mediante metaanálisis. En segundo

lugar, varios estudios fueron desarrollados en contextos especializados o con bases de datos retrospectivas, lo que puede limitar la generalización de los hallazgos a escenarios clínicos reales y a sistemas de salud con menor disponibilidad tecnológica.

Asimismo, aunque se priorizaron estudios originales con validación clínica, no todos reportaron de forma homogénea aspectos esenciales como validación externa, calibración del modelo, análisis de subgrupos, evaluación de sesgo algorítmico o impacto clínico posterior a la implementación. Por tanto, los resultados deben entenderse como una síntesis cualitativa de la evidencia disponible y no como una recomendación directa para adoptar modelos específicos sin evaluación local previa.

Conclusión

La evidencia analizada confirma que la inteligencia artificial constituye una herramienta altamente eficaz en el diagnóstico médico, especialmente en el análisis de imágenes y señales biomédicas, donde ha demostrado niveles elevados de sensibilidad, especificidad y capacidad discriminativa. Su aplicación permite optimizar la precisión diagnóstica, reducir la variabilidad clínica y favorecer la toma de decisiones basada en datos. No obstante, su implementación en la práctica clínica aún enfrenta desafíos relacionados con la validación externa, la interpretabilidad de los modelos y la integración en entornos reales de atención. En consecuencia, la IA debe ser concebida como un complemento del juicio clínico, cuyo desarrollo y adopción deben sustentarse en evidencia robusta, estándares metodológicos rigurosos y marcos regulatorios adecuados.

Identificación de la responsabilidad y contribución de los autores: Idea Original (JÁ, AY); recopilación de la información: (JÁ, AY); análisis de datos: (JÁ, AY); redacción del borrador: (JÁ, AY); parte metodológica: (JÁ, AY); revisión crítica del documento: (JÁ, AY).

Revisión por pares:

El manuscrito fue revisado por pares ciegos y fue aprobado oportunamente por el Equipo Editorial de la revista CIENCIA ECUADOR.

Conflictos de interés:

Los autores declaran no tener conflictos de interés relacionados con la publicación de este trabajo.

Financiamiento:

La presente investigación no recibió financiamiento externo de instituciones públicas, comerciales ni organizaciones sin fines de lucro.

Referencias

1. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng.* 2018;2(10):719–31.
2. Miotto R, Wang F, Wang S, et al. Deep learning for healthcare. *Brief Bioinform.* 2018;19(6):1236–46.
3. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine. *BMC Med Res Methodol.* 2019;19:64.
4. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng.* 2017;19:221–48.
5. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface.* 2018;15(141):20170387.
6. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with AI. *BMC Med.* 2019;17:195.
7. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med.* 2019;380:1347–58.
8. Razzak MI, Naz S, Zaib A. Deep learning for medical image processing. *J Med Syst.* 2018;42:226.
9. Park SH, Han K. Methodologic guide for evaluating clinical performance of AI. *Radiology.* 2018;286(3):800–9.
10. Price WN, Cohen IG. Privacy in the age of medical big data. *Nat Med.* 2019;25:37–43.
11. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy. *JAMA.* 2016;316(22):2402–10.
12. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115–8.
13. Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: radiologist-level pneumonia detection on chest X-rays. *arXiv.* 2017;arXiv:1711.05225.
14. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with deep learning. *Nat Med.* 2019;25(6):954–61.
15. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature.* 2020;577(7788):89–94.
16. Ting DSW, Cheung CYL, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases. *JAMA.* 2017;318(22):2211–23.

17. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection using deep neural networks. *Nat Med.* 2019;25(1):65–9.
18. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med.* 2018;24(9):1342–50.
19. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against healthcare professionals. *Lancet Digit Health.* 2019;1(6):e271–e297.
20. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs. *Nat Biomed Eng.* 2018;2(3):158–64.
21. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using AI-enabled ECG. *Nat Med.* 2019;25(1):70–4.
22. Komorowski M, Celi LA, Badawi O, et al. The artificial intelligence clinician learns optimal treatment strategies for sepsis. *Nat Med.* 2018;24(11):1716–20.
23. Chilamkurthy S, Ghosh R, Tanamala S, et al. Development and validation of deep learning algorithms for critical findings in head CT scans. *Lancet.* 2018;392(10162):2388–96.
24. Nam JG, Park S, Hwang EJ, et al. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules. *Radiology.* 2019;290(1):218–28.
25. Braga JR, Ramos V, Martins M, et al. Artificial intelligence for COVID-19 diagnosis using chest CT scans. *Eur Radiol.* 2020;30:1–9.
26. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44–56.
27. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA.* 2018;319(13):1317–8.
28. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med.* 2018;378:981–3.
29. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning in healthcare. *Nat Med.* 2019;25:1337–40.
30. Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. *Commun Med.* 2021;1:25.