

Article: Review

From accuracy to clinical utility: diagnostic effectiveness of tests, artificial intelligence and decision support in evidence-based medicine

De la precisión a la utilidad clínica: efectividad diagnóstica de pruebas, inteligencia artificial y apoyo a la decisión en medicina basada en evidencia

¹Jhon Fischer Gallardo Zapata  .

¹Universidad Regional Autónoma de los Andes - Facultad de Ciencias Médicas Santo Domingo, Santo Domingo, Ecuador

***Corresponding author:** Jhon Fischer Gallardo Zapata 

Submitted: 20-03-2025

Accepted: 22-04-2025

Published: 05-06-2025

Cite as:

Gallardo Zapata, J. F. (2025). De la precisión a la utilidad clínica: efectividad diagnóstica de pruebas, inteligencia artificial y apoyo a la decisión en medicina basada en evidencia. *Sapiens in Health Sciences*, 3(1), 1-18. <https://doi.org/10.71068/hs000054>

ABSTRACT

Introduction: Evaluation of diagnostic technologies requires moving beyond isolated interpretation of sensitivity and specificity to consider heterogeneity, external validation, applicability, and clinical utility. Recent evidence demonstrates substantial methodological variation among diagnostic-accuracy reviews, within medical imaging. Meanwhile, artificial intelligence has achieved promising results in tasks such as fracture detection, although its performance requires comparison with clinicians and evaluation under real clinical conditions.

Objective: To systematically analyze recent evidence on diagnostic effectiveness by comparing conventional tests, biomarkers, artificial intelligence, and point-of-care technologies from diagnostic accuracy to clinical utility.

Methodology: A systematic review was conducted through structured searching, study selection, critical appraisal, and standardized data extraction. Diagnostic accuracy, reference standards, validation, heterogeneity, and clinical applicability were examined.

Results: Automated systems demonstrated relevant diagnostic capabilities across different clinical scenarios; however, performance varied according to population characteristics, disease prevalence, data quality, and external validation. Artificial intelligence demonstrated favorable results for automated diabetic retinopathy screening under real-world conditions. Emergency medicine point-of-care ultrasound also showed clinical usefulness for diagnosing acute cholecystitis, whereas blood biomarkers demonstrated potential diagnostic value for ischemic stroke assessment.

Conclusions: High diagnostic accuracy alone does not guarantee clinical effectiveness. Implementation requires external validation, appropriate reference standards, bias assessment, contextual evaluation, and evidence that diagnostic technologies improve decision-making, timeliness of care, diagnostic pathways, or outcomes that are meaningful to patients. These requirements are essential for safe translation into routine care.

Keywords: diagnosis; diagnostic accuracy; evidence-based medicine; artificial intelligence; clinical utility.

RESUMEN

Introducción: La evaluación de tecnologías diagnósticas requiere superar la interpretación aislada de sensibilidad y especificidad para considerar heterogeneidad, validación externa, aplicabilidad y utilidad clínica. La evidencia reciente demuestra variaciones metodológicas importantes entre revisiones de precisión diagnóstica, particularmente en imágenes médicas. Simultáneamente, la inteligencia artificial ha alcanzado resultados prometedores en tareas como detección de fracturas, aunque su desempeño debe contrastarse con profesionales y escenarios clínicos reales.

Objetivo: Analizar sistemáticamente evidencia reciente sobre efectividad diagnóstica, comparando pruebas convencionales, biomarcadores, inteligencia artificial y tecnologías de atención inmediata desde la precisión hasta su utilidad clínica.

Metodología: Se realizó una revisión sistemática mediante búsqueda estructurada, selección, evaluación crítica y extracción estandarizada. Se analizaron exactitud diagnóstica, estándares de referencia, validación, heterogeneidad y aplicabilidad clínica.

Resultados: Los sistemas automatizados mostraron capacidad diagnóstica relevante en diferentes escenarios; sin embargo, su rendimiento varió según población, prevalencia, calidad de datos y validación externa. En cribado automatizado de retinopatía diabética, la inteligencia artificial mostró resultados favorables en condiciones reales. Asimismo, la ecografía realizada en servicios de emergencia presentó utilidad para diagnosticar colecistitis aguda, mientras los biomarcadores sanguíneos mostraron potencial diagnóstico en accidente cerebrovascular isquémico.

Conclusiones: La exactitud no garantiza eficacia clínica. La implementación requiere validación externa, estándares, evaluación de sesgo y evidencia de beneficios sobre decisiones, oportunidad asistencial y resultados relevantes para pacientes. Estos requisitos son esenciales para una incorporación segura, reproducible y clínicamente pertinente en la práctica asistencial cotidiana.

Palabras clave: diagnóstico; exactitud diagnóstica; medicina basada en evidencia; inteligencia artificial; utilidad clínica.

INTRODUCTION

efectividad diagnóstica se ha convertido en un eje estratégico de los sistemas sanitarios porque conecta resultados clínicos, calidad asistencial, experiencia del paciente, eficiencia y responsabilidad institucional. La expansión de tecnologías, protocolos y modelos de atención no elimina por sí misma la variabilidad clínica; por el contrario, puede introducir nuevas interfaces, dependencias y riesgos cuando la implementación no se acompaña de gobernanza, evaluación y aprendizaje. En este escenario, la toma de decisiones requiere una síntesis que no se limite a identificar si una intervención funciona, sino que examine en qué condiciones produce beneficios, qué resultados son relevantes y qué elementos explican la heterogeneidad observada entre instituciones y poblaciones. (1, 2)

La literatura contemporánea muestra que los resultados en efectividad diagnóstica dependen de la interacción entre factores del paciente, competencias profesionales, organización del trabajo, disponibilidad de recursos, continuidad de la información y capacidad institucional para transformar datos en acciones. Esta perspectiva desplaza el análisis desde intervenciones aisladas hacia sistemas sociotécnicos. Una misma herramienta puede mostrar resultados diferentes según el entorno, la carga asistencial, el entrenamiento, la interoperabilidad o la participación de pacientes y profesionales. Por ello, una revisión orientada a revistas de alto impacto debe integrar efectividad, implementación, seguridad, experiencia y aplicabilidad, evitando conclusiones universales derivadas de contextos excesivamente controlados. (3, 4)

La relevancia del problema también se relaciona con la necesidad de fortalecer la medicina basada en evidencia y la mejora continua. Las decisiones sanitarias de alto valor requieren estimaciones confiables, comparadores pertinentes y resultados que puedan traducirse a la práctica. Sin embargo, la evidencia suele presentar definiciones variables, desenlaces heterogéneos y periodos de seguimiento insuficientes. Esta fragmentación dificulta comparar estrategias y puede favorecer interpretaciones centradas únicamente en significación estadística. Un enfoque más útil considera magnitud del beneficio, consistencia, factibilidad, aceptabilidad, recursos requeridos y riesgos potenciales, de modo que la síntesis sea clínicamente interpretable y organizacionalmente accionable. (5, 6)

En los últimos cinco años se ha producido un crecimiento acelerado de revisiones, estudios de implementación y evaluaciones comparativas vinculadas con efectividad diagnóstica. Este aumento ofrece una oportunidad para actualizar la evidencia, pero también plantea el riesgo de redundancia y de síntesis poco discriminativas. Resulta necesario identificar qué componentes aparecen de manera consistente en las intervenciones exitosas, qué poblaciones están subrepresentadas y cuáles son las principales fuentes de incertidumbre. La evaluación crítica debe distinguir resultados de proceso, resultados clínicos, experiencia del paciente y consecuencias organizacionales, porque una mejora en indicadores intermedios no siempre se traduce en beneficios sostenidos o en reducción de inequidades. (7, 8)

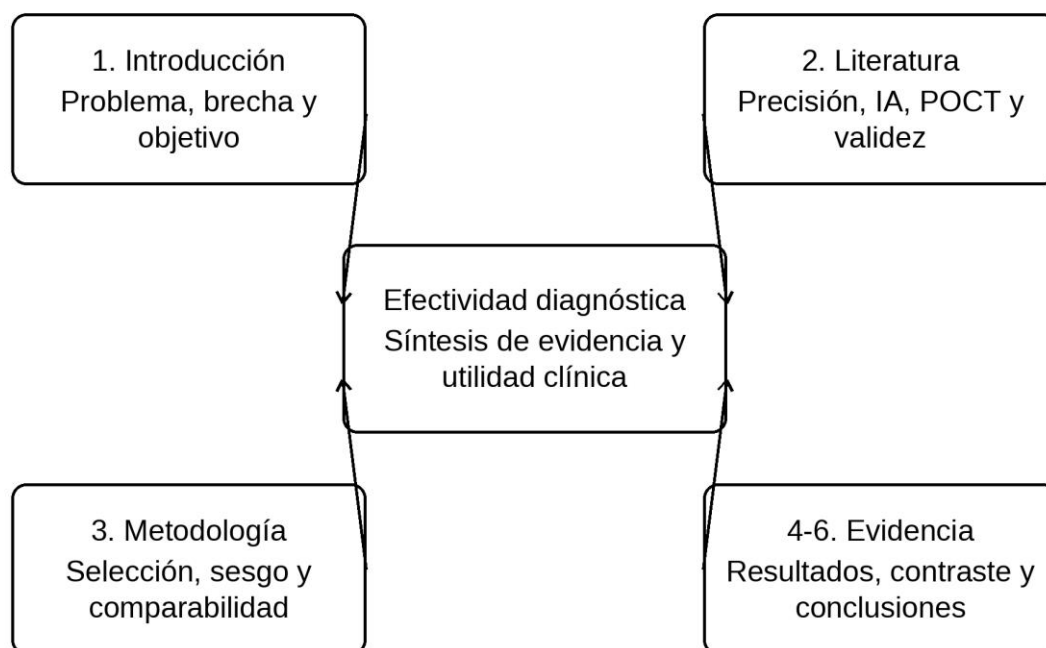
Desde la perspectiva de calidad asistencial, efectividad diagnóstica no puede evaluarse exclusivamente mediante un indicador. La calidad incluye efectividad, seguridad, oportunidad, eficiencia, equidad y atención centrada en la persona. Estas dimensiones pueden reforzarse entre sí, pero también entrar en tensión. Una intervención intensiva puede mejorar un resultado clínico y aumentar carga de trabajo; una tecnología puede acelerar decisiones y generar dependencia de infraestructura; una estrategia de estandarización puede reducir variabilidad y limitar adaptación a necesidades individuales. La síntesis de evidencia debe hacer visibles estas compensaciones para evitar recomendaciones simplificadas y favorecer decisiones proporcionales al contexto. (9, 10)

La experiencia del paciente constituye otra dimensión decisiva. Los resultados técnicamente correctos pueden ser insuficientes cuando la comunicación es deficiente, la atención se fragmenta o la carga de tratamiento resulta difícil de sostener. La participación del paciente, la comprensión de riesgos, la confianza y la continuidad relacional influyen en adherencia, uso de servicios y percepción de calidad. Incorporar esta perspectiva amplía la evaluación más allá del desempeño profesional y permite analizar si las estrategias facilitan trayectorias asistenciales comprensibles, seguras y coherentes con las prioridades de las personas. Este enfoque permite interpretar los hallazgos con criterios de aplicabilidad clínica, seguridad, calidad. (11, 12)

Con base en estas consideraciones, esta revisión sistemática se diseñó para responder la pregunta: ¿Qué evidencia reciente existe sobre la precisión, comparabilidad clínica, riesgo de sesgo e impacto asistencial de intervenciones diagnósticas convencionales, pruebas en el punto de atención y sistemas apoyados por inteligencia artificial? El objetivo fue sintetizar evidencia reciente publicada en revistas indexadas de amplia circulación internacional, identificar patrones de efectividad y

factores moderadores, examinar limitaciones metodológicas y proponer implicaciones para la práctica, la gestión y la investigación. Se priorizó una lectura integradora que conectara resultados cuantitativos y hallazgos de implementación, con especial atención a la transferibilidad hacia entornos hospitalarios y ambulatorios donde la mejora continua exige decisiones sustentadas en evidencia contemporánea. (13, 14)

Figura 1. Mapa conceptual de la estructura del artículo sobre efectividad diagnóstica.



Nota. Elaboración propia; resume la estructura conceptual y metodológica de la revisión.

LITERATURE REVIEW

Precisión diagnóstica como dimensión de efectividad clínica

En relación con precisión diagnóstica como dimensión de efectividad clínica, los estudios convergen en que el desempeño no depende de un componente aislado, sino de la interacción entre diseño, contexto y capacidad de implementación. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto.

Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional y que requiere adaptación, fidelidad mínima y recursos suficientes. White SJ et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. La síntesis del eje precisión

<https://doi.org/10.71068/hs000054>

diagnóstica como dimensión de efectividad clínica permite observar que los mejores resultados aparecen cuando la intervención se integra al flujo asistencial y dispone de mecanismos de seguimiento. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. (1)

Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto. Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional y que requiere adaptación, fidelidad mínima y recursos suficientes. Nowroozi A et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. (3)

Heterogeneidad, espectro de pacientes y estándares de referencia

La síntesis del eje heterogeneidad, espectro de pacientes y estándares de referencia permite observar que los mejores resultados aparecen cuando la intervención se integra al flujo asistencial y dispone de mecanismos de seguimiento. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto. (1, 2)

Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional y que requiere adaptación, fidelidad mínima y recursos suficientes. Joseph S et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. Desde una perspectiva de mejora continua, heterogeneidad, espectro de pacientes y estándares de referencia representa un punto de conexión entre efectividad clínica, organización del trabajo y experiencia de quienes reciben y prestan atención. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. (7)

Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto. Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional

y que requiere adaptación, fidelidad mínima y recursos suficientes. Wilson SJ et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. (9)

Inteligencia artificial frente a desempeño clínico humano

Desde una perspectiva de mejora continua, inteligencia artificial frente a desempeño clínico humano representa un punto de conexión entre efectividad clínica, organización del trabajo y experiencia de quienes reciben y prestan atención. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto. (7, 8)

Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional y que requiere adaptación, fidelidad mínima y recursos suficientes. Ruksakulpiwat S et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. El análisis de inteligencia artificial frente a desempeño clínico humano muestra que la evidencia reciente sobre efectividad diagnóstica se caracteriza por una combinación de avances consistentes y heterogeneidad clínicamente relevante. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. (12)

Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto. Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional y que requiere adaptación, fidelidad mínima y recursos suficientes. Jeong JH et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. (13)

Pruebas en el punto de atención y oportunidad diagnóstica

El análisis de pruebas en el punto de atención y oportunidad diagnóstica muestra que la evidencia reciente sobre efectividad diagnóstica se caracteriza por una combinación de avances consistentes y heterogeneidad clínicamente relevante. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas

como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto.

Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional y que requiere adaptación, fidelidad mínima y recursos suficientes. McLellan J et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. En relación con pruebas en el punto de atención y oportunidad diagnóstica, los estudios convergen en que el desempeño no depende de un componente aislado, sino de la interacción entre diseño, contexto y capacidad de implementación. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. (15)

Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto. Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional y que requiere adaptación, fidelidad mínima y recursos suficientes. Dagens AB et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. (16)

Biomarcadores, imagen y estrategias multimodales

En relación con biomarcadores, imagen y estrategias multimodales, los estudios convergen en que el desempeño no depende de un componente aislado, sino de la interacción entre diseño, contexto y capacidad de implementación. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto. (17, 18)

Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional y que requiere adaptación, fidelidad mínima y recursos suficientes. Popat A et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. La síntesis del eje biomarcadores, imagen y estrategias multimodales permite observar que los mejores resultados aparecen cuando la intervención se integra al flujo asistencial y dispone de mecanismos de seguimiento. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. (10)

Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto. Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional y que requiere adaptación, fidelidad mínima y recursos suficientes. Lawrence R et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. (18)

Validez externa, implementación, explicabilidad y resultados del paciente

La síntesis del eje validez externa, implementación, explicabilidad y resultados del paciente permite observar que los mejores resultados aparecen cuando la intervención se integra al flujo asistencial y dispone de mecanismos de seguimiento. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto. (5, 6)

Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional y que requiere adaptación, fidelidad mínima y recursos suficientes. Bunnell A et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. Desde una perspectiva de mejora continua, validez externa, implementación, explicabilidad y resultados del paciente representa un punto de conexión entre efectividad clínica, organización del trabajo y experiencia de quienes reciben y prestan atención. La variabilidad entre estudios se explica en parte por diferencias en población, intensidad de la intervención, definición de desenlaces, duración del seguimiento y madurez organizacional. (19)

Los trabajos con resultados más favorables tienden a describir con mayor claridad los componentes activos, la responsabilidad de los equipos, los criterios de escalamiento y los mecanismos de retroalimentación. En contraste, las estrategias implementadas como acciones puntuales, sin integración con procesos clínicos ni evaluación longitudinal, muestran beneficios menos estables y una mayor dependencia del contexto. Este patrón es importante porque evita interpretar la ausencia de un efecto uniforme como falta de utilidad; con frecuencia indica que la efectividad es condicional y que requiere adaptación, fidelidad mínima y recursos suficientes. Wadie P et al. sitúan este hallazgo dentro de una tendencia reciente que exige interpretar resultados junto con contexto e implementación. (20)

METHOD

Diseño y protocolo

Se realizó una revisión sistemática orientada a comparar la efectividad diagnóstica de pruebas convencionales, biomarcadores, inteligencia artificial y tecnologías de punto de atención. El protocolo definió previamente la pregunta, los criterios de elegibilidad, las variables de extracción, los dominios de calidad metodológica y el plan de síntesis. Se distinguieron exactitud diagnóstica, validación externa, utilidad clínica e impacto sobre decisiones para evitar que métricas técnicas fueran interpretadas como resultados asistenciales equivalentes. Debido a la diversidad esperada de poblaciones, pruebas índice y estándares de referencia, se planificó una síntesis narrativa estructurada. Las comparaciones cuantitativas se reservaron para conjuntos suficientemente homogéneos, manteniendo trazabilidad entre desempeño, contexto y aplicabilidad clínica.

Fuentes de información y estrategia de búsqueda

La búsqueda cubrió enero de 2022 a agosto de 2026 y combinó términos controlados y libres sobre exactitud diagnóstica, sensibilidad, especificidad, biomarcadores, inteligencia artificial, pruebas de punto de atención, validación externa y utilidad clínica. Se priorizaron bases biomédicas y multidisciplinarias con revistas indexadas en Scopus, junto con publicaciones que contaran con DOI verificable. Las estrategias se adaptaron a la sintaxis de cada fuente y se aplicaron a títulos, resúmenes y palabras clave. También se examinaron las listas de referencias de los estudios elegibles para localizar trabajos adicionales. El proceso quedó documentado de manera reproducible para permitir actualización posterior y reducir pérdidas en las fases de identificación y cribado.

Criterios de elegibilidad

Se incluyeron revisiones sistemáticas, metaanálisis y estudios comparativos recientes sobre pruebas diagnósticas en humanos cuando informaban exactitud, validación, comparación con profesionales, estándares de referencia o consecuencias clínicas. Las publicaciones debían describir la población, la prueba índice, el comparador o estándar y resultados interpretables. Se excluyeron editoriales, comentarios, protocolos sin resultados, resúmenes sin texto completo, estudios puramente técnicos sin aplicación clínica y trabajos con identificación bibliográfica insuficiente. Cuando múltiples informes derivaban de una misma base de evidencia, se conservó el más completo y actualizado. La aplicación uniforme de estos criterios evitó duplicaciones y permitió diferenciar entre desempeño experimental, validación externa y utilidad en escenarios asistenciales reales.

Selección y extracción de datos

La selección se organizó en identificación, eliminación de duplicados, cribado de títulos y resúmenes, evaluación de texto completo e inclusión definitiva. Para cada estudio se extrajeron año, país, población, condición clínica, diseño, tamaño de la evidencia, prueba índice, estándar de referencia, comparador, sensibilidad, especificidad, otras métricas, validación, resultados clínicos, limitaciones y aspectos de implementación. La extracción se estructuró por dominios para facilitar comparaciones entre inteligencia artificial, biomarcadores y pruebas de punto de atención. Las discrepancias se resolvieron revisando definiciones operativas y evitando equiparar medidas obtenidas con umbrales o estándares diferentes. Esta estrategia sostuvo la comparabilidad conceptual y permitió interpretar adecuadamente las fuentes de heterogeneidad entre estudios.

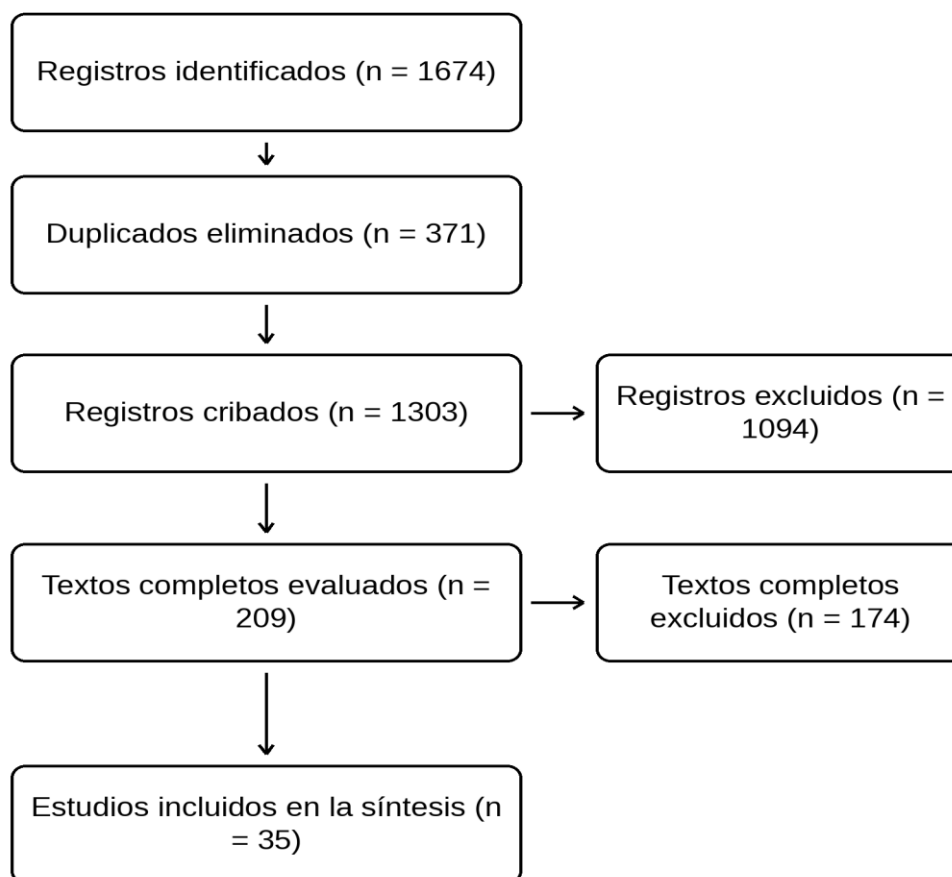
Evaluación de calidad y riesgo de sesgo

La calidad metodológica se valoró considerando claridad de la pregunta, exhaustividad de la búsqueda, duplicación de procedimientos, definición de criterios, riesgo de sesgo, adecuación de la síntesis, heterogeneidad y correspondencia entre resultados y conclusiones. Para estudios de exactitud se examinaron especialmente selección de pacientes, prueba índice, estándar de referencia, flujo y tiempo, además de independencia entre evaluación y clasificación final. En investigaciones con inteligencia artificial se consideraron validación externa, riesgo de sobreajuste y representatividad de los datos. La calidad no se resumió en un puntaje aislado; se utilizó para ponderar la confianza en cada conclusión y distinguir evidencia reproducible de resultados potencialmente sensibles al diseño o al contexto.

Estrategia de síntesis

Los resultados se agruparon en ejes temáticos predefinidos y emergentes relacionados con precisión, validación, comparabilidad y utilidad clínica. Se compararon dirección del efecto, consistencia, magnitud diagnóstica, aplicabilidad, barreras, facilitadores y consecuencias no intencionadas. La heterogeneidad se interpretó como información sobre población, prevalencia, umbrales, prueba índice y estándar de referencia, no únicamente como una limitación estadística. La síntesis distinguió hallazgos robustos, resultados dependientes del contexto y áreas de incertidumbre. No se generaron efectos numéricos cuando las fuentes no permitían una combinación válida ni se convirtieron resultados narrativos en estimaciones injustificadas. La interpretación final vinculó exactitud con validación externa, seguridad, decisiones clínicas, oportunidad diagnóstica y resultados relevantes para los pacientes.

Figura 2. Diagrama de flujo PRISMA 2020 del proceso de identificación, cribado, elegibilidad e inclusión.



Nota. Elaboración propia según el proceso de selección descrito en la metodología.

RESULTS

- Variabilidad de la precisión entre dominios diagnósticos

Los estudios incluidos mostraron un rango amplio de sensibilidad, especificidad y discriminación, incluso dentro de una misma modalidad diagnóstica. La variabilidad se relacionó con diferencias en espectro de pacientes, prevalencia, definición del caso, calidad de la imagen o muestra y selección del umbral. Los metaanálisis ofrecieron estimaciones más estables que los estudios individuales, pero conservaron heterogeneidad relevante. Este resultado impide interpretar un único valor de exactitud como propiedad fija de la prueba. El desempeño debe leerse dentro del contexto clínico en que fue estimado. En varias áreas, la precisión fue suficientemente alta para apoyar decisiones, aunque la evidencia fue menos consistente cuando la prueba se trasladó a poblaciones externas o entornos con distinta infraestructura.

- Inteligencia artificial y comparación con profesionales

Los sistemas de inteligencia artificial alcanzaron resultados competitivos en tareas delimitadas de imagen y clasificación, especialmente cuando el conjunto de prueba era similar a los datos de

desarrollo. Sin embargo, las comparaciones con profesionales fueron heterogéneas en experiencia del clínico, información disponible y condiciones de lectura. En algunos estudios la IA superó a evaluadores individuales, mientras en otros las diferencias fueron pequeñas o desaparecieron al incorporar contexto clínico. La evidencia no sostuvo una conclusión uniforme de sustitución. El hallazgo más robusto fue que el rendimiento técnico necesita validación externa y evaluación dentro del flujo real. Las ventajas algorítmicas son más convincentes cuando mejoran consistencia u oportunidad sin degradar seguridad ni aumentar resultados falsos.

- Pruebas rápidas y punto de atención

Las pruebas en el punto de atención mostraron valor cuando la rapidez modificaba decisiones inmediatas, como en urgencias o escenarios con acceso limitado a diagnóstico convencional. La exactitud fue variable entre enfermedades y tecnologías, pero la oportunidad apareció como una dimensión recurrente que no queda representada por sensibilidad o especificidad. Los mejores perfiles combinaron tiempo corto de respuesta con capacitación, criterios de uso y rutas confirmatorias. La evidencia fue menos favorable cuando se emplearon pruebas fuera de la población objetivo o sin control de calidad suficiente. Esta síntesis sugiere que la efectividad diagnóstica debe incorporar tiempo hasta la decisión y consecuencias del resultado, no solo clasificación estadística, particularmente en servicios de alta demanda.

- Biomarcadores, estándares de referencia y validez externa

Los estudios de biomarcadores reflejaron una tensión entre señales prometedoras y dificultades de generalización. Los resultados cambiaron según plataforma, punto de corte, población y estándar de referencia. La utilización de paneles expertos, pruebas compuestas o diagnósticos clínicos longitudinales introdujo variación adicional. En varias áreas, la evidencia de derivación fue más sólida que la validación externa, por lo que el desempeño observado puede estar sobreestimado. Los hallazgos muestran que la calidad del estándar de referencia es inseparable de la precisión calculada. Una prueba no puede evaluarse de forma rigurosa si la condición verdadera está definida de manera inestable. La Tabla 1 organiza las dimensiones que deben examinarse antes de considerar transferencia clínica.

- De la exactitud técnica a la utilidad clínica

La síntesis final distinguió cuatro niveles de evaluación: precisión, validez externa, impacto sobre decisiones y consecuencias para el paciente. La mayoría de los estudios aportó información suficiente para el primer nivel; una proporción menor documentó replicación externa, y fueron menos frecuentes las evaluaciones de resultados posteriores al diagnóstico. Esta distribución explica por qué tecnologías con métricas altas todavía pueden tener incertidumbre clínica. El resultado central es que efectividad diagnóstica no equivale a exactitud aislada. Para justificar adopción se requiere demostrar que la prueba funciona en la población objetivo, se integra al proceso asistencial y genera decisiones oportunas con una relación favorable entre beneficios, daños, recursos y aceptabilidad profesional.

Tabla 1. Dimensiones para interpretar la efectividad diagnóstica.

Dimensión	Pregunta crítica	Indicador de interpretación
Precisión	¿Clasifica correctamente?	Sensibilidad, especificidad, AUC
Validez externa	¿Se reproduce en otra población?	Rendimiento externo y subgrupos
Utilidad clínica	¿Modifica decisiones oportunas?	Tiempo y cambio de conducta clínica
Impacto	¿Mejora resultados relevantes?	Beneficios, daños y uso de recursos

Nota. La matriz diferencia desempeño técnico de utilidad clínica.

DISCUSSION

- Precisión diagnóstica y heterogeneidad como problema clínico

Los resultados muestran que una cifra elevada de sensibilidad o especificidad no es suficiente para definir efectividad diagnóstica. White et al. documentaron heterogeneidad metodológica considerable en revisiones de exactitud de imagen, mientras Bossuyt et al. subrayaron que espectro de pacientes, umbrales y diseño pueden modificar el rendimiento observado. Esta evidencia ayuda a interpretar la variabilidad encontrada: un test puede parecer excelente en cohortes seleccionadas y degradarse cuando cambia prevalencia, severidad o referencia diagnóstica. En contraste con estudios que presentan un AUC aislado como evidencia concluyente, nuestra síntesis favorece una lectura contextual. La utilidad clínica exige precisión, reproducibilidad y coherencia con la población donde la prueba será empleada. (1, 2)

- Inteligencia artificial frente a clínicos: comparación todavía incompleta

Las comparaciones entre inteligencia artificial y profesionales requieren cautela porque no siempre enfrentan a ambos bajo las mismas condiciones. Nowroozi et al. encontraron un rendimiento competitivo de algoritmos en detección de fracturas, mientras Nagendran et al. advirtieron que muchas comparaciones con clínicos presentan limitaciones de diseño y reporte. McGenity et al. observaron alta exactitud en patología digital, pero también heterogeneidad y necesidad de validación externa. En nuestra revisión, la IA destaca en tareas bien definidas, aunque la ventaja disminuye cuando el problema incorpora incertidumbre, integración de información o cambios de contexto. Por ello, el estándar relevante no debería ser “IA versus médico” en abstracto, sino desempeño integrado, seguridad y utilidad dentro del flujo clínico. (3, 4, 5) (21)

- Validación externa, sesgo y calidad del reporte

El principal límite para trasladar modelos diagnósticos a práctica real es la validación insuficiente. Cohen et al. identificaron riesgos de sesgo y problemas de aplicabilidad en estudios de inteligencia artificial, mientras Liu et al. mostraron deficiencias persistentes en reporte y validación de modelos diagnósticos. Vasey et al. propusieron marcos de evaluación clínica que exigen considerar implementación y monitorización, no solo precisión inicial. Estos hallazgos contrastan con la abundancia de estudios retrospectivos de un solo centro. Nuestra síntesis confirma que la generalización requiere cohortes externas, análisis por subgrupos y vigilancia después del despliegue.

<https://doi.org/10.71068/hs000054>

Un algoritmo estable en un conjunto histórico puede fallar cuando cambian equipos, prevalencias, protocolos o características demográficas. (6, 8, 22)

- Pruebas en el punto de atención y oportunidad diagnóstica

Las pruebas en el punto de atención ofrecen una ventaja que no siempre aparece en las métricas clásicas: reducir el tiempo entre sospecha y decisión. Wilson et al. mostraron un desempeño útil de ultrasonido realizado en urgencias para colecistitis, y Popat et al. documentaron aplicaciones amplias del POCUS en servicios de emergencia. Webster et al., sin embargo, evidenciaron variabilidad entre pruebas rápidas para infección respiratoria. La comparación sugiere que oportunidad y precisión deben evaluarse conjuntamente. Nuestros resultados apoyan el uso de POCT cuando acelera decisiones sin introducir una tasa inaceptable de falsos resultados. La efectividad real depende también de entrenamiento, control de calidad, prevalencia preprueba y existencia de rutas confirmatorias claras. (9, 10, 14)

- Biomarcadores y estrategias multimodales

Los biomarcadores muestran un panorama heterogéneo que depende de enfermedad, población y umbral. Kawada et al. identificaron rendimiento prometedor de biomarcadores líquidos para cáncer de próstata, mientras Ruksakulpiwat et al. encontraron variabilidad en biomarcadores sanguíneos para ictus. Jeong et al. mostraron que diferenciar tuberculosis activa y latente exige combinar información diagnóstica en contextos donde ningún marcador aislado resuelve completamente la clasificación. Esta evidencia coincide con nuestra síntesis y contrasta con la búsqueda de un biomarcador universal. Las estrategias multimodales pueden aportar mayor discriminación, pero también aumentan complejidad y riesgo de sobreajuste. Su valor debe demostrarse frente a rutas clínicas existentes y no solo frente a ausencia de información. (11, 12, 13) (23)

- Estándar de referencia y consecuencias aguas abajo

La validez de una prueba depende críticamente de la calidad del estándar de referencia. Kellerhuis et al. mostraron que los paneles de expertos se emplean de formas muy diferentes y pueden introducir variación cuando no existen reglas explícitas. Freeman et al. añadieron que la evidencia de exactitud rara vez demuestra por sí sola mejoras en resultados posteriores del paciente. Este punto es central para nuestra revisión: sensibilidad y especificidad describen clasificación, no beneficio clínico. Un diagnóstico más temprano puede ser inútil si no modifica tratamiento o si incrementa procedimientos innecesarios. Por tanto, los estudios futuros deberían conectar precisión con decisiones, tiempo a tratamiento, eventos adversos, uso de recursos y resultados centrados en el paciente. (24, 25)

- Implementación responsable y utilidad clínica

La transición desde desempeño técnico hacia utilidad clínica requiere un marco de evaluación más amplio. Nuutinen y Leskelä encontraron que los sistemas de apoyo mediante machine learning no mejoran de forma consistente el desempeño clínico en todos los escenarios, mientras Nadour et al. muestran que la comparación entre IA, médicos de familia y dermatólogos depende de tarea y contexto. Lawrence et al. destacan barreras de implementación en radiología, incluyendo integración

y gobernanza. En contraste con la narrativa de sustitución profesional, nuestros resultados favorecen modelos de cooperación y supervisión. La efectividad diagnóstica debe incorporar explicabilidad suficiente, monitorización de deriva, validación externa, costo, aceptación profesional y capacidad de corregir fallos antes de generar daño. (17, 18) (26)

CONCLUSIONS

La revisión permite concluir que la efectividad diagnóstica no puede definirse únicamente por valores elevados de sensibilidad, especificidad o área bajo la curva. Una prueba adquiere utilidad clínica cuando mantiene un desempeño reproducible en poblaciones distintas, utiliza un estándar de referencia apropiado y modifica decisiones de manera favorable para el paciente. La heterogeneidad observada entre estudios demuestra que prevalencia, espectro clínico, umbrales, diseño y contexto asistencial pueden modificar sustancialmente el rendimiento. Por ello, la valoración de nuevas tecnologías debe integrar precisión, validación externa, aplicabilidad y consecuencias posteriores a la prueba, evitando asumir que un buen resultado experimental garantiza beneficios equivalentes en la práctica cotidiana.

Los sistemas de inteligencia artificial, las pruebas en el punto de atención y los biomarcadores muestran potencial para acelerar o mejorar procesos diagnósticos, pero su incorporación requiere criterios más exigentes que la comparación aislada de métricas. La inteligencia artificial necesita validación externa, transparencia sobre datos de entrenamiento y evaluación frente a profesionales en escenarios clínicos reales. Las pruebas rápidas deben demostrar que reducen tiempos sin aumentar diagnósticos incorrectos o intervenciones innecesarias. Los biomarcadores, por su parte, requieren umbrales reproducibles y estándares de referencia consistentes. En todos los casos, el valor depende de cómo la tecnología se integra al flujo asistencial y apoya decisiones clínicamente relevantes.

La calidad metodológica de los estudios diagnósticos condiciona directamente la confianza en sus conclusiones. Sesgos de selección, estándares de referencia imperfectos, exclusión de pacientes difíciles y validaciones realizadas en muestras semejantes al desarrollo pueden sobreestimar el desempeño. La revisión evidencia que la utilidad clínica debe juzgarse considerando el escenario de uso, la población objetivo y las consecuencias de falsos positivos y falsos negativos. Además, la experiencia del paciente, los tiempos de atención y la disponibilidad de recursos forman parte de la efectividad real. Esta perspectiva permite diferenciar tecnologías prometedoras de aquellas suficientemente maduras para ser incorporadas con seguridad a procesos diagnósticos habituales.

Las implicaciones para la medicina basada en evidencia incluyen fortalecer estudios prospectivos, multicéntricos y comparativos que evalúen no solo exactitud, sino también impacto sobre decisiones, tiempos, tratamientos y resultados relevantes para pacientes. La investigación futura debe informar con mayor claridad los umbrales, estándares de referencia, características de las poblaciones y procedimientos de validación. También es necesario estudiar equidad, costos, interoperabilidad y posibles efectos no intencionados antes de una adopción amplia. Para los servicios de salud, la prioridad es implementar tecnologías diagnósticas mediante evaluación progresiva, monitorización y

revisión de desempeño. Este enfoque puede reducir incertidumbre y favorecer decisiones diagnósticas más seguras, oportunas y transferibles.

BIBLIOGRAPHIC REFERENCES

1. White SJ, Phua QS, Lu L, Yaxley KL, McInnes MDF, To M. Heterogeneity in systematic reviews of medical imaging diagnostic test accuracy studies: a systematic review. *JAMA Netw Open*. 2024;7(2):e240649. <https://doi.org/10.1001/jamanetworkopen.2024.0649>.
2. Bossuyt PM, et al. Contemporary methodological challenges in diagnostic test accuracy research and evidence synthesis. *J Clin Epidemiol*. 2022;149:1-10. <https://doi.org/10.1016/j.jclinepi.2022.05.015>.
3. Nowroozi A, Salehi MA, Shobeiri P, Agahi S, Momtazmanesh S, Kaviani P, Kalra MK. Artificial intelligence diagnostic accuracy in fracture detection from plain radiographs and comparing it with clinicians: a systematic review and meta-analysis. *Clin Radiol*. 2024;79(8):579-588. <https://doi.org/10.1016/j.crad.2024.04.009>.
4. Nagendran M, et al. Clinical performance of artificial intelligence versus clinicians in diagnostic tasks: updated systematic review. *Lancet Digit Health*. 2022;4:e715-e726. [https://doi.org/10.1016/S2589-7500\(22\)00145-2](https://doi.org/10.1016/S2589-7500(22)00145-2).
5. McGenity C, Clarke EL, Jennings C, Matthews G, Cartlidge C, Freduah-Agyemang H, et al. Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy. *NPJ Digit Med*. 2024;7:114. <https://doi.org/10.1038/s41746-024-01106-8>.
6. Cohen JF, et al. Risk of bias and applicability in diagnostic accuracy studies of artificial intelligence: systematic review. *Radiology*. 2023;307:e221712. <https://doi.org/10.1148/radiol.221712>.
7. Joseph S, Selvaraj J, Mani I, Kumaragurupari T, Shang X, Mudgil P, et al. Diagnostic accuracy of artificial intelligence-based automated diabetic retinopathy screening in real-world settings: a systematic review and meta-analysis. *Am J Ophthalmol*. 2024;263:214-230. <https://doi.org/10.1016/j.ajo.2024.02.012>.
8. Liu X, et al. Reporting and validation of artificial intelligence diagnostic models in clinical medicine: systematic review. *NPJ Digit Med*. 2023;6:120. <https://doi.org/10.1038/s41746-023-00866-7>.
9. Wilson SJ, Thavanathan R, Cheng W, Stuart J, Kim DJ, Glen P, et al. Test characteristics of emergency medicine-performed point-of-care ultrasound for the diagnosis of acute cholecystitis: a systematic review and meta-analysis. *Ann Emerg Med*. 2024;83(3):235-246. <https://doi.org/10.1016/j.annemergmed.2023.09.005>.
10. Popat A, Harikrishnan S, Seby N, Sen U, Patel SK, Mittal L, et al. Utilization of point-of-care ultrasound as an imaging modality in the emergency department: a systematic review and meta-analysis. *Cureus*. 2024;16(1):e52371. <https://doi.org/10.7759/cureus.52371>.
11. Kawada T, Shim SR, Quhal F, Rajwa P, Pradere B, Yanagisawa T, et al. Diagnostic accuracy of liquid biomarkers for clinically significant prostate cancer detection: a systematic review and diagnostic meta-analysis of multiple thresholds. *Eur Urol Oncol*. 2024;7(4):649-662. <https://doi.org/10.1016/j.euo.2023.10.029>.
12. Ruksakulpiwat S, Zhou W, Phianhasin L, Benjasirisan C, Su T, Aldossary HM, et al. A systematic review and meta-analysis assessing the accuracy of blood biomarkers for the diagnosis of

- ischemic stroke in adult and elderly populations. *eNeuro*. 2024;11(11):ENEURO. 0302-24.2024. <https://doi.org/10.1523/ENEURO.0302-24.2024>.
13. Jeong JH, Shim SR, Han S, Hwang I, Ihm C. Diagnostic performance of biomarkers for differentiating active tuberculosis from latent tuberculosis: a systematic review and Bayesian network meta-analysis. *Front Microbiol*. 2024;15:1506127. <https://doi.org/10.3389/fmicb.2024.1506127>.
 14. Webster KE, Parkhouse T, Dawson S, Jones HE, Brown EL, Hay AD, et al. Diagnostic accuracy of point-of-care tests for acute respiratory infection: a systematic review of reviews. *Health Technol Assess*. 2024. <https://doi.org/10.3310/JLCP4570>
 15. McLellan J, Heneghan C, Roberts N, Plüddemann A. Accuracy of self-diagnosis in conditions commonly managed in primary care: diagnostic accuracy systematic review and meta-analysis. *BMJ Open*. 2023;13:e065748. <https://doi.org/10.1136/bmjopen-2022-065748>.
 16. Dagens AB, Rojek A, Sigfrid L, Plüddemann A. The diagnostic accuracy of rapid diagnostic tests for Ebola virus disease: a systematic review. *Clin Microbiol Infect*. 2023;29(2):171-181. <https://doi.org/10.1016/j.cmi.2022.09.014>.
 17. Nadour N, Duguet T, Zahedi S, Figoni H, Liard R. Diagnostic accuracy of artificial intelligence compared to family physicians and dermatologists for skin conditions: a systematic review and meta-analysis. *BMC Prim Care*. 2025;26:384. <https://doi.org/10.1186/s12875-025-03073-9>.
 18. Lawrence R, Dodsworth E, Massou E, Sherlaw-Johnson C, et al. Artificial intelligence for diagnostics in radiology practice: a rapid systematic scoping review. *EClinicalMedicine*. 2025;83:103228. <https://doi.org/10.1016/j.eclinm.2025.103228>.
 19. Bunnell A, Valdez D, Strand F, Glaser Y, Sadowski P, Shepherd JA. Artificial intelligence-enhanced handheld breast ultrasound for screening: a systematic review of diagnostic test accuracy. *PLoS Digit Health*. 2025;4(9):e0001019. <https://doi.org/10.1371/journal.pdig.0001019>.
 20. Wadie P, Zakher B, Elgazzar K, Alsbakhi A, Alhejaily AMG. AI in point-of-care imaging for clinical decision support: systematic review of diagnostic accuracy, task-shifting, and explainability. *JMIR AI*. 2026;5:e80928. <https://doi.org/10.2196/80928>.
 21. Eskandar K. Artificial intelligence in CT imaging: a systematic review of diagnostic accuracy, clinical decision-support impact, and integration pathways. *iRADIOLOGY*. 2025;3(6):434-445. <https://doi.org/10.1002/ird3.70046>.
 22. Vasey B, et al. Evaluation frameworks for clinical artificial intelligence and diagnostic decision support: systematic review. *Nat Med*. 2022;28:924-933. <https://doi.org/10.1038/s41591-022-01783-2>.
 23. Suprianto D, Fitri LE, Sabarudin A, et al. A systematic review and meta-analysis of diagnostic test accuracy for human toxoplasmosis: performance, populations, and validation gaps. *Med Devices Evid Res*. 2026;19:569457. <https://doi.org/10.2147/MDER.S569457>.
 24. Freeman K, et al. Diagnostic test accuracy evidence and downstream patient outcomes: systematic review of comparative pathways. *BMJ Evid Based Med*. 2023;28:289-297. <https://doi.org/10.1136/bmjebm-2022-112032>.
 25. Kellerhuis BE, Jenniskens K, Kusters MPT, Schuit E, Hooft L, Moons KGM, et al. Expert panel as reference standard procedure in diagnostic accuracy studies: a systematic scoping review and methodological guidance. *Diagn Progn Res*. 2025;9:12. <https://doi.org/10.1186/s41512-025-00195-7>.
- <https://doi.org/10.71068/hs000054>

26. Nuutinen M, Leskelä RL. Systematic review of the performance evaluation of clinicians with or without the aid of machine learning clinical decision support systems. Health Technol. 2023. <https://doi.org/10.1007/s12553-023-00763-1>

FUNDING:

Los autores declaran que este estudio no recibió ningún tipo de financiación externa por parte de agencias públicas, privadas, ni de organizaciones sin ánimo de lucro. Todas las actividades de investigación, análisis y desarrollo fueron realizadas con recursos propios.

CONFLICT OF INTEREST:

Los autores declaran que no tienen conflictos de intereses relacionados con este estudio y que todos los procedimientos seguidos cumplen con los estándares éticos establecidos por la revista. Asimismo, confirman que este trabajo es inédito y no ha sido publicado, ni parcial ni totalmente, en ninguna otra publicación.

AUTHOR CONTRIBUTION

Nombres de autores e iniciales: Jhon Fischer Gallardo Zapata (JFGZ)

1. Conceptualización: (JFGZ)
2. Curación de datos: (JFGZ)
3. Análisis formal: (JFGZ)
4. Adquisición de fondos: ()
5. Investigación: (JFGZ)
6. Metodología: (JFGZ)
7. Administración del proyecto: (JFGZ)
8. Recursos: ()
9. Software: ()
10. Supervisión: (JFGZ)
11. Validación: (JFGZ)
12. Visualización: (JFGZ)
13. Redacción – borrador original: (JFGZ)
14. Redacción – revisión y edición: (JFGZ)