



INTELIGENCIA ARTIFICIAL PARA EL CRIBADO DE RETINOPATÍA DIABÉTICA: PRECISIÓN DIAGNÓSTICA, IMPLEMENTACIÓN EN ENTORNOS REALES Y DESAFÍOS ÉTICOS

ARTIFICIAL INTELLIGENCE FOR SCREENING FOR DIABETIC RETINOPATHY: DIAGNOSTIC ACCURACY, IMPLEMENTATION IN REAL-WORLD SETTINGS, AND ETHICAL CHALLENGES

Telmo Eduardo Salas Joya¹, Israel Arellano González², Jonathan Alexander Terán Trujillo³,
Leonardo Lanzee Mendoza Jiménez⁴

{elsecretodelavictoria@yahoo.es¹, dr.optiexpress@gmail.com², jalexts@gmail.com³, lic.lanzee.mendoza@hotmail.com⁴}

Fecha de recepción: 25/07/2026 / Fecha de aceptación: 01/09/2026 / Fecha de publicación: 16/09/2025

RESUMEN: La retinopatía diabética (RD) representa la causa principal de ceguera prevenible en las personas adultas en edad laboral, teniendo además una prevalencia en aumento a nivel de la difusión de la diabetes mellitus en el mundo. **Objetivo:** analizar la evidencia científica publicada desde 2016 hasta 2026 sobre la precisión diagnóstica de la inteligencia artificial (IA) en cuanto a cribado de RD, su aplicación en la práctica clínica real y la ética de su aplicación. Se llevó a cabo una revisión bibliográfica de tipo narrativo con carácter sistemático según lineamientos PRISMA 2020, a través de búsqueda en bases de datos indexadas (PubMed/MEDLINE, Scopus, Web of Science y Cochrane Library) de estudios primarios, revisiones sistemáticas y metaanálisis publicados en revistas de alto índice de impacto. Los algoritmos de aprendizaje profundo alcanzan sensibilidades y especificidades superiores al 90 % en la detección de RP referible bajo condiciones de laboratorio rigurosas; sin embargo, estos algoritmos encuentran sus limitaciones en el nivel del desempeño diagnosticado en poblaciones externas y en particular en países de ingresos bajos/módicos. Los estudios de implementación de la IA en el marco de programas nacionales de cribado documentan aumentos en cobertura y reducción en los costos, aunque las brechas de la generalización aún prevalecen asociadas a la representatividad de las bases de entrenamiento. En la ética, encontramos riesgos de sesgo algorítmico, brechas de equidad, problemas de responsabilidad clínica y consentimiento informado. La IA es vista como una herramienta con potencial para aumentar el acceso al cribado de RD, aunque su utilización requiere rigurosos estudios externos

¹Hospital IESS Manta, Ecuador, <https://orcid.org/0009-0009-6767-3393>

²Investigador independiente, Ciudad de México, México, <https://orcid.org/0009-0007-3719-4952>

³Investigador independiente, <https://orcid.org/0009-0006-6488-8370>

⁴Investigador independiente México Cdmx, <https://orcid.org/0009-0005-3322-6916>



que validen su funcionamiento, la creación de marcos regulatorios claros y la implementación de estrategias para mitigar sesgos.

Palabras clave: *Retinopatía diabética, inteligencia artificial, aprendizaje profundo, cribado, ética médica.*

ABSTRACT: Diabetic retinopathy (DR) is the leading cause of preventable blindness among working-age adults, and its prevalence is increasing in line with the global spread of diabetes mellitus. Objective: To analyze the scientific evidence published from 2016 to 2026 regarding the diagnostic accuracy of artificial intelligence (AI) for DR screening, its application in real-world clinical practice, and the ethics of its use. A systematic narrative literature review was conducted according to the PRISMA 2020 guidelines, involving a search of indexed databases (PubMed/MEDLINE, Scopus, Web of Science, and the Cochrane Library) for primary studies, systematic reviews, and meta-analyses published in high-impact journals. Deep learning algorithms achieve sensitivities and specificities exceeding 90% in the detection of referable PR under rigorous laboratory conditions; however, these algorithms face limitations in their diagnostic performance in external populations, particularly in low- and middle-income countries. Studies on the implementation of AI within national screening programs document increases in coverage and reductions in costs, although gaps in generalizability still persist due to issues with the representativeness of training datasets. From an ethical perspective, we identify risks of algorithmic bias, equity gaps, issues of clinical accountability, and informed consent. AI is viewed as a tool with the potential to increase access to RD screening, although its use requires rigorous external studies to validate its performance, the creation of clear regulatory frameworks, and the implementation of strategies to mitigate bias.

Keywords: *Diabetic retinopathy, artificial intelligence, deep learning, screening, medical ethics*

INTRODUCCIÓN

La diabetes mellitus supone una de las emergencias de salud pública más aceleradas del siglo XXI. Para el año 2021, se encontraba estimado que, 537 millones de adultos de 20 a 79 años presentarían esta patología a nivel mundial, cifra que se espera que aumente hasta los 783 millones para el año 2045 (1). Este aumento sostenido de la prevalencia se traduce, automáticamente, en el incremento de las complicaciones, entre las cuales la retinopatía diabética (RD) se erige como una de las principales martilleadores de la función visual y de la calidad de vida de los que la presentan.

La RD es considerada la principal causa de ceguera prevenible en la población adulta en edad laboral. Los estudios de carga global estiman que en el año 2020 había alrededor de 103 millones de personas afectadas por algún grado de RD, cifra que podría llegar a 160 millones hacia el año 2045 de no llevarse a cabo programas de detección precoz (2). El cribado periódico mediante



fotografía de fondo de ojo permite detectar precozmente lesiones retinianas incentivadas por la RD antes de que entren a fase de pérdida visual irreversible; no obstante, la implementación de programas de cribado se ve limitada por importantes restricciones estructurales como la escasez de oftalmólogos y de especialistas en lectura de imágenes, especialmente en zonas rurales y en países de ingresos bajos y medios.

Bajo este contexto, los avances en inteligencia artificial (IA) y en particular, en lo que respecta a aprendizaje profundo (deep learning) han mostrado ser una esperanza para la automatización de la interpretación de imágenes de fondo de ojos. El inicio de la etapa actual fueron los resultados publicados por Gulshan et al. en 2016 (3), donde se entrenó una nueva cartografía con más de 128 000 fotos con el fondo de ojo, obteniendo 97,5% de sensibilidad y 93,4% de especificidad para detectar RD referible, valores comparables a los documentados por oftalmólogos certificados. Al año siguiente Ting et al. validaron un sistema similar en una población multiétnica de diez países, mostrando que el enfoque del sistema era robusto y por tanto generalizable en diferentes contextos poblacionales (4). Estos hallazgos iniciales iniciaron una línea de investigación que está en constante avance y consolidación, constituyendo uno de los lugares más desarrollados en inteligencia artificial médica. Las revisiones de Ting et al. documentan como el aprendizaje profundo ha salido de la RD hacia otras enfermedades oculares, acoplándose poco a poco a la teleoftalmología como una estrategia para acercar el diagnóstico especializado a comunidades con escaso acceso a los servicios de salud visual (5), (6). En entornos con escasos recursos, esa integración cobra una especial relevancia: en la India, Rajalakshmi et al. validaban un sistema de IA acoplado a un dispositivo de fotografía retiniana basado en teléfono inteligente, alcanzando una sensibilidad de más del 95 % en detección de RD amenazante para la visión en una población de pacientes atendidos en un centro terciario de diabetes (7).

El siguiente paso decisivo en la historia de esta tecnología fue la transición desde el ámbito experimental hacia el ámbito clínico autorizado. El sistema IDx-DR fue en 2018 el primer dispositivo de diagnóstico de carácter autónomo y basado en IA que recibió la autorización de la Administración de Alimentos y Medicamentos de los Estados Unidos EPA tras un ensayo pivotal en diez centros de atención primaria y con 900 pacientes que demostraron una sensibilidad y especificidad superiores a los mínimos exigidos por las guías clínicas (8). Metaanálisis a posteriori han confirmado que la exactitud diagnóstica de ese sistema y otros productos comerciales se mantuvo elevada en diferentes escenarios de validación, aunque presentando variabilidad dependiendo de la cámara utilizada, de la calidad de la imagen asociada y de las características de la población evaluada (9), (10). Sin embargo, el paso de la validación controlada a la puesta en marcha en la práctica clínica ha puesto de manifiesto la existencia de una brecha considerable entre el rendimiento teórico de los algoritmos y la realidad en riesgo clínico.

Revisiones sistemáticas específicas de entornos del mundo real demuestran que, aunque la sensibilidad global se mantiene dentro de márgenes aceptables, la especificidad y la tasa de imágenes no revisadas presentan grandes variaciones entre diferentes programas de cribado



(11), algo que ya se había documentado en los primeros estudios de comparación de un software automatizado frente a evaluadores humanos (12).

Esta brecha se intensifica si comparamos un modelo entrenado en poblaciones de alto ingreso cuando se aplica en contextos de menos recursos. La validación clínica realizada en Tailandia, Singapur y distintos países de África subsahariana ha sido una oportunidad de cuantificar tanto las mejoras en cobertura y costo-efectividad, como las limitaciones en la generalizabilidad de los modelos entrenados a partir de bases de datos poco diversas (13), (14). La realización prospectiva del sistema de Google Health, en el transcurso del proceso nacional de cribado de Tailandia, -con el que probó la "implementación" a gran escala, en áreas del sistema nacional de salud (15) así como los estudios recientes en Latinoamérica y en el África subsahariana que muestran que la exactitud interna alcanzada durante la fase de entrenamiento no significa que se apliquen los mismos resultados externos a poblaciones con distintos atributos demográficos y equipos de captura (16), algo que se ha reiterado a modo de preocupación en revisiones de inteligencia artificial y salud ocular global (17).

Las limitaciones técnicas no son ajenas a sus implicaciones éticas. La implementación de los sistemas autónomos de diagnóstico hace surgir interrogantes sobre la responsabilidad clínica ante errores diagnósticos, sobre la "transparencia" de los sistemas "caja negra" o de la necesidad de que exista una forma adecuada del consentimiento informado cuando la decisión de égida no toma en cuenta a un profesional humano (18). A esto hay que añadir la creciente evidencia de que los sesgos contenidos en los conjuntos de características de los conjuntos de entrenamiento relacionados con el origen étnico, el nivel socioeconómico o el tipo de dispositivo de captura, por ejemplo, pueden ser objeto de inequidades diagnósticas que pueden afectar desproporcionadamente a las personas ya subrepresentadas en la investigación biomédica (19), (20). Visiones más recientes de la equidad algorítmica en Medicina sugieren marcos de mitigación para todo el ciclo de vida del producto (desde la curación de datos a la evaluación postmarketing) (21), (22) mientras que las discusiones específicas de la RD han interiorizado la necesidad de realizar un consentimiento informado que comunique explícitamente el carácter autónomo del sistema y los límites de su rendimiento (23).

De este escenario, emerge una tensión no resuelta entre tres dimensiones que rara vez se han abordado de manera conjunta: la exactitud diagnóstica demostrada en ensayos controlados, la efectividad real observada durante su utilización en sistemas de salud muy diversos y las implicaciones éticas de delegar tal decisión clínica a un sistema automatizado. La literatura revisada analiza con frecuencia estas dimensiones de manera aislada, lo que complica el entendimiento de las condiciones que puede requerir una implementación responsable de la IA para RD, y más aún en contextos de escasez de recursos como pueden ser los de América Latina. Por ello, la justificación de la presente revisión es la necesidad de sistematizar y comparar la evidencia científica sobre estas tres dimensiones para poder ofrecer una perspectiva integrada que resulta útil para clínicos, investigadores y responsables de la toma de decisiones en Políticas de Salud ocular.



De este modo, la revisión de la literatura presentada en este trabajo tiene como fin el analizar la evidencia científica existente de entre 2016 y 2026 en torno a la precisión diagnóstica de los sistemas de Inteligencia Artificial aplicada al cribado de la retinopatía diabética, su implementación en entornos clínicos reales y los principales desafíos éticos que surgen en el caso de la IA.

MATERIALES Y MÉTODOS

El presente artículo hace referencia a la revisión bibliográfica narrativa de enfoque sistemático, orientada a sintetizar y comparar la evidencia científica sobre el desempeño diagnóstico, la implementación real y las implicaciones éticas de la inteligencia artificial para el cribado de la retinopatía diabética. El proceso de búsqueda, selección y presentación de la evidencia se estructuró siguiendo los criterios de la declaración PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), que es el marco metodológico de referencia para la realización de las revisiones bibliográficas ya que propone una metodología estandarizada y explícita para identificar, cribado e incluir fuentes.

La unidad de análisis estuvo constituida por artículos científicos originales, revisiones sistemáticas, metaanálisis y estudios de implementación clínica publicados entre enero de 2016 y julio de 2026. Se incluyeron investigaciones que abordaban desarrollo, validez de diagnóstico, implementación en programas de cribado o reflexión ética de sistemas de inteligencia artificial, en concreto de algoritmos de deep learning, aplicados a la detección de retinopatía diabética mediante fotografía del fondo de ojo y otras modalidades de imagen de la retina; se excluyeron cartas al editor sin datos originales, resúmenes de congresos, trabajos no indexables y estudios que se centraban en retinopatías distintas a la diabetes. La búsqueda bibliográfica se llevó a cabo en bases de datos científicas indexadas de renombre internacional: PubMed/MEDLINE, Scopus, Web of Science, Cochrane Library y se complementó con el rastreo de las listas de referencias de los trabajos incluidos. Se utilizaron todas las combinaciones posibles de términos MeSH, de palabras clave en inglés y en español, entre otros "diabetic retinopathy", "artificial intelligence", "deep learning", "screening", "teleophthalmology", "algorithmical bias", "medical ethics".

Para los estudios de precisión diagnóstica se extrajeron, cuando estaban disponibles, los indicadores de sensibilidad, especificidad, área bajo la curva ROC (AUC), valores predictivos positivo y negativo, y odd ratio diagnóstica. Para los estudios de implementación se registraron variables relativas a la cobertura poblacional, a la tasa de imágenes no evaluables, al coste por paciente cribado y al tiempo para el diagnóstico. Para la dimensión ética se sistematizaron las categorías de riesgo propuestas por la literatura (sesgo algorítmico, equidad, responsabilidad clínica y consentimiento informado).

Dada la naturaleza bibliográfica del presente trabajo, no se ha llegado a un metaanálisis cuantitativo propio; el trabajo analítico se ha convertido en un análisis mediante una síntesis narrativa de comparación de los resultados globales, que han sido los que han aportado los



trabajos primarios tal como lo han hecho con las revisiones sistemáticas/metaanálisis integrados, muchas de las cuales han aplicado modelos HSROC y de efectos aleatorios a la hora de combinar las estimaciones de sensibilidad y especificidad. Los resultados se han organizado en tres ejes temáticos precisión diagnóstica, implementación real y desafíos éticos para facilitar la comparación entre estudios con sus respectivas evidencias heterogéneas, todo ello tal como lo tiene planificado el presente trabajo.

Tabla 1. Resumen detallado del diseño metodológico empleado en la presente revisión.

Componente	Descripción
Tipo de estudio	Revisión bibliográfica narrativa con enfoque sistemático, orientada por PRISMA 2020
Período de búsqueda	Enero de 2016 a julio de 2026
Bases de datos	PubMed/MEDLINE, Scopus, Web of Science, Cochrane Library
Idiomas incluidos	Inglés y español
Tipo de fuentes	Estudios de validación diagnóstica, ensayos de implementación, revisiones sistemáticas, metaanálisis y perspectivas éticas
Criterios de inclusión	Publicación en revistas indexadas de alto impacto; IA aplicada a cribado de RD; datos originales o síntesis cuantitativa
Criterios de exclusión	Publicaciones no indexadas, resúmenes de congresos, estudios sobre otras retinopatías no diabéticas
Variables analizadas	Sensibilidad, especificidad, AUC, cobertura, costo, sesgo, equidad y responsabilidad clínica
Tipo de síntesis	Narrativa comparativa por ejes temáticos, sin metaanálisis cuantitativo propio

La forma de diseñar la investigación metodológica justifica la integradora naturaleza del objetivo planteado, ya que al encontrarnos ante un fenómeno que integra a evidencias técnicas de precisión diagnóstica, datos de implementación en sistemas de salud y argumentación ética, un enfoque únicamente cuantitativo podría dejarnos fuera de escena y ser relevantes las dimensiones cualitativas para la toma de decisiones clínicas y regulatorias. El marco PRISMA 2020, aunque diseñado para revisiones sistemáticas con metaanálisis, también aporta un rigor y una trazabilidad en todo el proceso de búsqueda y selección de estudios con el fin de evitar el sesgo de selección, al que nos lleva la revisión estrictamente narrativa. El combinar estudios con diferentes diseños -validación diagnóstica, estudios de cohortes prospectivos, análisis económicos y bioética- ha proporcionado una descripción más completa de la actualidad en el campo, a pesar de tener la limitación de no poder llegar a cálculos de estimaciones combinadas propias, limitación que se apunta más adelante.

RESULTADOS

A continuación, en esta sección, se exponen los hallazgos distribuidos en tres líneas o ejes: la precisión diagnóstica de los sistemas de IA, los resultados de los sistemas de inteligencia artificial en clínicas reales y los principales retos éticos extraídos de la literatura. En cada apartado del eje comenzamos describiendo el hallazgo, seguido de la evidencia de cada hallazgo, sistematizada en tablas y, a continuación, un breve análisis de la evidencia recogida.



Los estudios de validación diagnóstica demuestran de manera recurrente que los algoritmos de aprendizaje profundo son capaces de alcanzar unos niveles de exactitud equivalentes, y hasta superiores en algunos contextos, a los de oftalmólogos entrenados para detectar la retinopatía diabética referible. La Tabla 2 resume los indicadores de rendimiento de los estudios y de los metaanálisis más representativos que hemos podido recuperar en los objetivos de nuestra revisión.

Tabla 2. Indicadores de precisión diagnóstica reportados por estudios seleccionados de sistemas de IA para el cribado de retinopatía diabética.

Estudio / fuente	Año	Sistema o enfoque evaluado	Sensibilidad	Especificidad / AUC
Gulshan et al. (3)	2016	Red neuronal convolucional (EyePACS-1 / Messidor-2)	97,5 %	93,4 % (AUC 0,99)
Ting et al. (5)	2017	Sistema multiétnico (10 países)	90,5 %	91,6 %
Abràmoff et al. (8) (IDx-DR)	2018	Ensayo pivotal, atención primaria (EE. UU.)	87,2 %	90,7 %
Khan et al. (9)	2025	Metaanálisis IDx-DR (múltiples cohortes)	≈ 89 % (agrupada)	≈ 88 % (agrupada)
Uy et al. (11)	2023	Metaanálisis, entornos reales (prospectivos)	Variable (heterogeneidad alta)	Especificidad reducida frente a validación controlada
Mwavu et al. (16)	2026	CNN, validación externa (Latinoamérica / África subsahariana)	Interna: hasta 95,7 %	Externa: caída marcada de AUC (0,51–0,61)

Dada la información que se tiene en la Tabla 2, se aprecia un patrón bastante similar, no sólo en lo que respecta a la exactitud diagnóstica de sistemas de IA en las especificidades de los servicios, condiciones y poblaciones en las que se han entrenado sus modelos (3), (4) sino también en el ensayo clínico fundamental que orientó la autorización regulatoria de IDx-DR (8), los metaanálisis que se llevaron a cabo posteriormente en este sistema (9). Sin embargo, al pivote de estudios prospectivos sobre entornos reales, la especificidad y la proporción de imágenes no evaluables continúan mostrando una heterogeneidad considerable (11), llegando a ser aún más acusada en el estudio de validación externa en Latinoamérica y en África subsahariana; pasando, en efecto, el área bajo la curva ROC de valores cercanos a 0,99 en la validación interna a valores de apenas 0,51 a 0,61 cuando son evaluados sobre una población demográfica y sobre un instrumento de captura totalmente distinto (16). Esta disparidad constituye uno de los hallazgos más relevantes de la revisión presente, ya que implica que la exactitud estimada de los estudios de desarrollo no debería extrapolarse directamente a escenarios clínicos heterogéneos.

Más allá de la precisión diagnóstica en condiciones de validación, un número progresivamente elevado de trabajos ha evidenciado el rendimiento y la factibilidad de la IA al ser implementada



dentro de programas de cribado poblacional. La Tabla 3 presenta un resumen de los hallazgos principales de implementación a los que se ha llegado en los distintos entornos geográficos y recursos.

Tabla 3. Estudios de implementación de sistemas de IA en programas de cribado de retinopatía diabética en entornos reales.

País / programa	Estudio	Diseño	Principal hallazgo
Singapur	Bellemo et al. (2019) (13)	Validación clínica en cohortes africanas	Sensibilidad alta para RD referible; utilidad confirmada como herramienta de apoyo, no de reemplazo total del especialista
Singapur (modelo económico)	Xie et al. (2020) (14)	Análisis de costo-efectividad	Reducción proyectada de costos por paciente cribado frente al cribado manual, con ahorro creciente a mayor escala
Tailandia	Ruamviboonsuk et al. (2022) (15)	Cohorte prospectiva de intervención, 9 sitios de atención primaria	Exactitud del 94,7 % para RD que amenaza la visión; se identificaron barreras logísticas y de flujo de trabajo
Latinoamérica / África subsahariana	Mwavu et al. (2026) (16)	Validación externa multicéntrica	Deterioro sustancial del desempeño al cambiar de población, cámara y prevalencia de enfermedad
Revisión global	Tan et al. (2023) (17)	Perspectiva sobre salud ocular digital global	Oportunidades de ampliación de cobertura condicionadas a infraestructura, conectividad y regulación local

El análisis combinado de estos trabajos permite identificar tres tendencias. En primer lugar, la implementación de IA en programas nacionales de cribado comporta incrementos en coberturas y reducción de costos, como muestra el modelo económico desarrollado en Singapur (14) y la experiencia clínica relatada en cohortes africanas (13). En segundo lugar, el paso de la validación retrospectiva hacia la implementación prospectiva ha puesto al descubierto retos operativos (flujo de pacientes, calidad imagen en condiciones de campo, aceptación del personal de salud) que no aparecen en los estudios de laboratorio, tal como demuestra los resultados de la experiencia tailandesa (15). Como tercer aspecto, y quizás uno de los más críticos, la generalización geográfica de los algoritmos continúa siendo la limitación más importante: el rendimiento que se observa en las respectivas poblaciones latinoamericanas y africanas empeora de manera drástica respecto a los resultados ofrecidos por las cohortes de entrenamiento original (16), algo que coincide con las conclusiones de las revisiones globales sobre inteligencia artificial y salud ocular digital que han señalado cómo la infraestructura tecnológica y regulatoria local determina de forma contundente la probabilidad de éxito en la aceptación (17).



La revisión de la literatura ética ha permitido sistematizar los riesgos y las tensiones normativas presentes en la implementación de sistemas autónomos de diagnóstico en el cribado de la retinopatía diabética, que se plasman en la Tabla 4.

Tabla 4. Principales categorías de desafíos éticos identificados en la literatura sobre IA aplicada al cribado de retinopatía diabética.

Categoría	Descripción del desafío
Responsabilidad clínica	Ambigüedad sobre quién asume la responsabilidad ante un error diagnóstico: el desarrollador del algoritmo, la institución de salud o el profesional que lo supervisa
Sesgo algorítmico	Subrepresentación de determinados grupos étnicos, socioeconómicos o geográficos en las bases de entrenamiento, con riesgo de subdiagnóstico diferencial
Equidad en el acceso	Riesgo de que la tecnología amplíe, en lugar de reducir, las brechas de acceso a diagnóstico entre poblaciones urbanas y rurales o entre países de distinto nivel de ingreso
Consentimiento informado	Necesidad de comunicar con claridad al paciente que el diagnóstico proviene de un sistema autónomo y cuáles son sus límites de desempeño
Transparencia algorítmica	Dificultad para auditar modelos de tipo caja negra y explicar sus decisiones ante pacientes y reguladores
Gobernanza y monitoreo	Necesidad de vigilancia poscomercialización continua para detectar deterioro del desempeño ante cambios poblacionales o tecnológicos

La naturaleza de los hallazgos analizados nos lleva a manifestar que la discusión ética sobre la IA en los casos de RD por parte de la IA no puede desligarse del debate de la IA aplicada a la medicina. Como señalaban Char et al. ya en 2018, la puesta en práctica de algoritmos de aprendizaje automático en la práctica clínica exige afrontar explícitamente las cuestiones de responsabilidad y de transparencia antes de poder disfrutar de todos sus beneficios (18). En el caso concreto de la RD, el hecho de contar con un sistema autónomo y validado para emitir un resultado diagnóstico sin la intervención de un especialista incrementó este debate, propiciando propuestas concretas sobre el contenido mínimo que debería llegar a los pacientes durante el proceso del consentimiento informado (23).

Así mismo, con respecto al sesgo algorítmico, McCradden et al. avisaban de que las soluciones técnicas de equidad como el ajuste de umbrales por subgrupo no resuelven por sí solas los dilemas normativos subyacentes y que podrían introducir nuevas formas de discriminación si se aplican sin un marco ético explícito (19). Esta advertencia coincide con los hallazgos empíricos descritos en el apartado anterior, donde el deterioro del desempeño en poblaciones subrepresentadas no es un simple problema técnico, sino la expresión más concreta de la inequidad estructural de la disponibilidad de datos de entrenamiento (20). Marcos éticos recientes de la equidad de la IA aplicada a la medicina nos plantean la intervención en diferentes momentos del ciclo de vida del producto: curación de datos, validación estratificada por subgrupos y monitoreo continuo como estrategia para reducir estos riesgos (21), (22), recomendación que se puede aplicar directamente a los hallazgos del deterioro del desempeño documentados en los estudios de validación externa revisados en el apartado anterior.



DISCUSIÓN

Los hallazgos de esta revisión avalan, por un lado, que la inteligencia artificial ha llegado a un nivel de madurez técnica significativa para la detección de la retinopatía diabética referible, con valores de sensibilidad y especificidad equiparables o superiores a los de los oftalmólogos entrenados en escenarios de validación controlada (3), (4). Ésta es también la conclusión de la revisión de literatura presentada por Ting et al. (5), quienes relatan una trayectoria de mejora sostenida del rendimiento de los algoritmos de aprendizaje profundo durante los últimos diez años (5), así como la revisión de Grzybowski et al. en el trayecto que va del desarrollo experimental al despliegue clínico de los sistemas (10).

Sin embargo, al contraste de estos resultados con la evidencia de estudios de implementación real, existe una disparidad importante que invita a la reflexión. Pese a que los estudios de desarrollo alcanzan porcentajes de sensibilidad que rondan el 95-97 % (3), los estudios prospectivos de programas de cribado nacional presentan cifras más moderadas y cuentan con proporciones elevadas de imágenes no evaluables (15). Dicho desfase es acorde con lo advertido por Uy et al. en su metaanálisis centrado estrictamente en un entorno del mundo real, donde la heterogeneidad entre los estudios fue considerablemente más amplia que en los basados en cohortes de validación controlada (11). El hallazgo más revelador en este sentido pertenece al estudio de validación externa desarrollado en Latinoamérica y en África subsahariana, en el que arquitecturas de red que alcanzaron un área bajo la curva ROC superior a 0,98 en validación interna se desplomaron hasta zonas cercanas al azar (0,51-0,61) en poblaciones con una composición demográfica, prevalencia de enfermedad y equipo de captura de imágenes bien distintos (16). Además, esta observación es compatible con la literatura reciente en torno a la equidad algorítmica en salud, que pone de manifiesto que una buena gestión del rendimiento técnico en las etapas de desarrollo no es sinónimo de buena equidad o de buena aplicabilidad en la práctica (21).

Aunque los resultados de la revisión de este artículo también matizan el argumento clásico para el que la IA reduce automáticamente los costes del cribado, en contraste con estudios económicos anteriores sobre esta cuestión, en el análisis de coste-efectividad de la investigación desarrollada en Singapur, se obtuvieron ahorros por cribado manual frente al cribado automatizado (14). Sin embargo, este ahorro económico se produce en un sistema de salud con un nivel de ingresos altos y en el que la infraestructura tecnológica está bien consolidada, una circunstancia que no puede considerarse automáticamente como tal en los países de ingresos bajos y medios, donde la carga de las RD no diagnosticadas es de alguna forma mayor a la de los países con altos ingresos (13). Esta observación se ve reflejada, como en la revisión de Tan et al. sobre la salud ocular digital a nivel internacional, en la que la infraestructura de conectividad y los marcos regulatorios territoriales son determinantes tan significativos como la precisión técnica del algoritmo para que su adopción sea exitosa (17).



En el plano ético, la revisión presentada aquí no hace otra cosa que reafirmar la argumentación de que los riesgos de sesgo, equidad y responsabilidad clínica no son elementos periféricos, sino que son esas consideraciones que son centrales en la viabilidad de implantar dicha implementación responsable. Por tanto, una comparación entre los hallazgos técnicos de las presentadas aquí y los de la literatura ética hace aparecer una correlación directa: los mismos factores que indican Chen et al. Identifican como fuentes de sesgo algorítmico en medicina diversidad de adquisición de datos, subrepresentación poblacional, y diversidad entre observadores (21) justo los factores que subyacen al deterioro de desempeño que está documentado empíricamente en los estudios de validación externa de sistemas de RD (16). Esta convergencia entre evidencia técnica y razonamientos normativos sugiere que la solución a la brecha de la generalización no puede limitarse a un ajuste algorítmico y ha de incluir, tal como plantea el marco de la equidad propuesto por Abràmoff et al., una intervención que haya de transitar todo el ciclo del producto, desde la composición de las bases de datos de entrenamiento hasta el seguimiento poscomercialización en cada contexto de despliegue de dicho producto (22).

Cierto es que, por otro lado, hay que reconocer algunas limitaciones inherentes a este análisis. Al ser una revisión bibliográfica de tipo narrativo, de enfoque sistemático, no se hizo un metaanálisis cuantitativo propio y, por tanto, la comparación entre estudios tal como se ha hecho está condicionada por la heterogeneidad de diseños, de poblaciones, y de definiciones de RD referible que son utilizadas en la literatura original. Por otro lado, la mayor parte de la evidencia de implementación proviene de contextos asiático y africano, y la evidencia específica sobre América Latina sigue siendo escasa, lo que limita la extrapolación de los resultados a los sistemas de salud en dicha región. Hacen falta futuras investigaciones primarias con datos poblacionales latinoamericanos que cierren esta brecha en la evidencia.

CONCLUSIONES

La información analizada permite dar respuesta del modo afirmativo que marca el objetivo: los sistemas de inteligencia artificial han conseguido una elevada exactitud diagnóstica del diagnóstico de determinación de retinopatía diabética referible en situaciones con validación controlada, con niveles de sensibilidad y especificidad similares a los médicos humanos. Pero esta exactitud técnica no es la misma que se obtiene en la práctica clínica normal, ya que aspectos relacionados con la diversidad de las poblaciones, con la calidad de la captura de imagen o con la infraestructura existente dan lugar a un rendimiento mucho más variable y, en algunos casos, sensiblemente inferior al encontrado en los estudios de desarrollo.

Con respecto a cómo poner en marcha programas de cribado, lo que estos hallazgos evidencian es que la inteligencia artificial es capaz de incrementar de manera muy significativa la cobertura diagnóstica y, además, de producir eficiencias económicas en sistemas de salud con infraestructura tecnológica robusta pero, por otro lado, que estos beneficios no se transfieren de forma solidaria entre territorios, ni tampoco se garantizan automáticamente por hacerlo con un algoritmo en un contexto determinado. La generalización geográfica y de la población es, por



tanto, el principal cuello de botella técnico que obstaculiza la adopción realmente global de esta tecnología. En cuanto a la parte ética, se concluye que la responsabilidad clínica ante errores diagnósticos, el sesgo algorítmico producidos por bases de entrenamiento no representativas, las brechas de equidad en el acceso y la necesidad de un consentimiento informado adecuado son variables impactantes que no deben dejarse a un lado en cualquier estrategia de implementación responsable.

Por otra parte, el solapamiento entre los factores técnico-susceptibles de explicar el deterioro del rendimiento sobre poblaciones subrepresentadas y el marco ético que advierte sobre la exposición a la inequidad algorítmica reafirma una relación circular entre ambos y un no abordarlos independientemente. En resumen, la inteligencia artificial es una herramienta con un potencial real de transformar el cribado de la retinopatía diabética a nivel poblacional, pero su uso responsable requiere de un esfuerzo constante de validación externa en poblaciones diversas, el desarrollo de marcos regulatorios explícitos que determinen la responsabilidad clínica, y la implementación de estrategias de mitigación de sesgos y monitorización a largo plazo del rendimiento una vez cada sistema esté en uso. Solo bajo estas condiciones es que se podrá llegar a plasmar su potencial para disminuir la carga global de la ceguera prevenible por la diabetes.

REFERENCIAS BIBLIOGRÁFICAS

1. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract.* 2022 Jan;183:109119. doi:10.1016/j.diabres.2021.109119
2. Teo ZL, Tham YC, Yu M, Chee ML, Rim TH, Cheung N, et al. Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045. *Ophthalmology.* 2021 Nov;128(11):1580–91. doi:10.1016/j.ophtha.2021.04.027
3. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA.* 2016 Dec 13;316(22):2402. doi:10.1001/jama.2016.17216
4. Ting DSW, Cheung CYL, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA.* 2017 Dec 12;318(22):2211. doi:10.1001/jama.2017.18152
5. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology.* 2019 Feb;103(2):167–75. doi:10.1136/bjophthalmol-2018-313173
6. Ting DSW, Peng L, Varadarajan A V., Keane PA, Burlina PM, Chiang MF, et al. Deep learning in ophthalmology: The technical and clinical considerations. *Prog Retin Eye Res.* 2019 Sep;72:100759. doi:10.1016/j.preteyeres.2019.04.003



7. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye*. 2018 Jun 9;32(6):1138–44. doi:10.1038/s41433-018-0064-9
8. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. 2018 Aug 28;1(1):39. doi:10.1038/s41746-018-0040-6
9. KHAN Z, GAIDHANE AM, SINGH M, GANESAN S, KAUR M, SHARMA GC, et al. Diagnostic Accuracy of IDX-DR for Detecting Diabetic Retinopathy: A Systematic Review and Meta-Analysis. *Am J Ophthalmol*. 2025 May;273:192–204. doi:10.1016/j.ajo.2025.02.022
10. Grzybowski A, Singhanetr P, Nanegrungsunk O, Ruamviboonsuk P. Artificial Intelligence for Diabetic Retinopathy Screening Using Color Retinal Photographs: From Development to Deployment. *Ophthalmol Ther*. 2023 Jun 2;12(3):1419–37. doi:10.1007/s40123-023-00691-3
11. Uy H, Fielding C, Hohlfeld A, Ochodo E, Opare A, Mukonda E, et al. Diagnostic test accuracy of artificial intelligence in screening for referable diabetic retinopathy in real-world settings: A systematic review and meta-analysis. *PLOS Global Public Health*. 2023 Sep 20;3(9):e0002160. doi:10.1371/journal.pgph.0002160
12. Tufail A, Rudisill C, Egan C, Kapetanakis V V., Salas-Vega S, Owen CG, et al. Automated Diabetic Retinopathy Image Assessment Software. *Ophthalmology*. 2017 Mar;124(3):343–51. doi:10.1016/j.ophtha.2016.11.014
13. Bellemo V, Lim ZW, Lim G, Nguyen QD, Xie Y, Yip MYT, et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. *Lancet Digit Health*. 2019 May;1(1):e35–44. doi:10.1016/S2589-7500(19)30004-4
14. Xie Y, Nguyen QD, Hamzah H, Lim G, Bellemo V, Gunasekeran D V, et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. *Lancet Digit Health*. 2020 May;2(5):e240–9. doi:10.1016/S2589-7500(20)30060-1
15. Ruamviboonsuk P, Tiwari R, Sayres R, Nganthavee V, Hemarat K, Kongprayoon A, et al. Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study. *Lancet Digit Health*. 2022 Apr;4(4):e235–44. doi:10.1016/S2589-7500(22)00017-6
16. Mwavu R, Kaggwa F, Arunga S, Wasswa W. Evaluating the Generalisability of Convolutional Neural Networks for Diabetic Retinopathy Detection in Latin America and Sub-Saharan Africa. *Information*. 2026 Jun 3;17(6):552. doi:10.3390/info17060552
17. Tan TF, Thirunavukarasu AJ, Jin L, Lim J, Poh S, Teo ZL, et al. Artificial intelligence and digital health in global eye health: opportunities and challenges. *Lancet Glob Health*. 2023 Sep;11(9):e1432–43. doi:10.1016/S2214-109X(23)00323-6
18. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *New England Journal of Medicine*. 2018 Mar 15;378(11):981–3. doi:10.1056/NEJMp1714229



19. McCradden MD, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness solutions in health care machine learning. *Lancet Digit Health*. 2020 May;2(5):e221–3. doi:10.1016/S2589-7500(20)30065-0
20. Mhasawade V, Zhao Y, Chunara R. Machine learning and algorithmic fairness in public and population health. *Nat Mach Intell*. 2021 Jul 29;3(8):659–66. doi:10.1038/s42256-021-00373-4
21. Chen RJ, Wang JJ, Williamson DFK, Chen TY, Lipkova J, Lu MY, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat Biomed Eng*. 2023 Jun 28;7(6):719–42. doi:10.1038/s41551-023-01056-8
22. Abràmoff MD, Tarver ME, Loyo-Berrios N, Trujillo S, Char D, Obermeyer Z, et al. Considerations for addressing bias in artificial intelligence for health equity. *NPJ Digit Med*. 2023 Sep 12;6(1):170. doi:10.1038/s41746-023-00913-9
23. Abramoff MD, Mortensen Z, Tava C. Commentary: Diagnosing Diabetic Retinopathy With Artificial Intelligence: What Information Should Be Included to Ensure Ethical Informed Consent? *Front Med (Lausanne)*. 2021 Nov 24;8. doi:10.3389/fmed.2021.765936