InvestiGO
Revista Científica Multidisciplinaria
ISSN 2953-6367
http://revistainvestigo.com

Octubre 2026
Vol. 7 No. 21, PP. 525-543
https://doi.org/10.56519/m3asbh09

# MEDICINA DE PRECISIÓN EN ENFERMEDADES CARDIOVASCULARES: INTEGRACIÓN DE BIOMARCADORES, DATOS CLÍNICOS E INTELIGENCIA ARTIFICIAL PARA LA TOMA DE DECISIONES

# PRECISION MEDICINE IN CARDIOVASCULAR DISEASES: INTEGRATION OF BIOMARKERS, CLINICAL DATA, AND ARTIFICIAL INTELLIGENCE FOR DECISION-MAKING

Vanesa Paola Gallo Morales[1], Janina Gissella Pilco Diaz[2], Marco Vinicio Flores Blacio[3], Arnulfo Pazos Rodríguez[4]

{paulymorales015@gmail.com[1], pilcojaninna@gmail.com[2], mflores1@utmachala.edu.ec[2], yomeropazitos@hotmail.com[4]}

Fecha de recepción: 10/09/2026 / Fecha de aceptación: 08/10/2026 / Fecha de publicación: 14/10/2026

**RESUMEN: La medicina cardiológica de precisión intenta mejorar los enfoques basados en la población mediante la aplicación de biomarcadores multiómicos, datos clínicos y inteligencia artificial (IA). Recopilar y analizar sistemáticamente la evidencia disponible sobre modelos basados en inteligencia artificial que incorporen, como mínimo, dos de los siguientes factores: biomarcadores multiómicos, datos clínicos o imágenes cardiovasculares/fenotipos vasculares, aplicables a personas de edad adulta con enfermedad o riesgo cardiovascular. Se realizó una búsqueda en PubMed/MEDLINE, Scopus, Web of Science y ScienceDirect (1 de enero de 2020-31 de marzo de 2026) de acuerdo con los criterios PRISMA 2020. Se extrajo información sobre los estudios primarios incluidos que cumplían los criterios de inclusión, y se evaluaron su validez y los riesgos de sesgo utilizando PROBAST+AI. La certeza de la evidencia se evaluó utilizando el GRADE adaptado a los modelos pronósticos. Se identificaron tres estudios que cumplían los criterios de inclusión. El estudio «CardiOmicScore» (Luo et al., 2026) informó de un C-index para ProScore de entre 0,69 y 0,82, y para MetScore de entre 0,64 y 0,74 en la predicción de seis enfermedades cardiovasculares (ECV), y una ganancia de entre 0,005 y 0,102 al combinarse con datos clínicos. El estudio «Queen of Hearts» (Carvalho et al., 2026) obtuvo una AUC de 0,94 (intervalo de confianza del 95 %: 0,92-0,95) en un registro retrospectivo multicéntrico de 1 032 pacientes para la detección de oclusión miocárdica (OMI). El modelo de predicción de ictus basado en imágenes de la retina (Govindaiah et al., 2025) informó de una AUC de 0,83 para un ictus a los cinco años y de 0,79 a los diez años. La certeza de la evidencia se clasificó como muy baja para todas las variables de resultado. Existe escasez de estudios que utilicen criterios de**

[1]Investigador independiente, Ecuador, https://orcid.org/0009-0008-4100-4450
[2]Investigador independiente, Ecuador, https://orcid.org/0000-0001-9004-1102
[3]Universidad Técnica de Machala, Ecuador, https://orcid.org/0009-0006-4381-7776
[4]Investigador independiente, México, https://orcid.org/0009-0009-6211-1914

**inclusión ambiciosos. La evidencia disponible no permite conclusiones definitivas sobre la eficacia del uso de la IA para mejorar la medicina de precisión cardiovascular. Es necesario disponer de más información sobre la validación cruzada de estos modelos en poblaciones diferentes, incluida la evaluación de la calibración y los resultados centrados en el paciente.**

***Palabras clave: Medicina de precisión; enfermedades cardiovasculares; inteligencia artificial; biomarcadores; aprendizaje profundo; revisión sistemática***

**ABSTRACT: Precision cardiology seeks to improve population-based approaches through the application of multi-omic biomarkers, clinical data, and artificial intelligence (AI). To systematically synthesize the evidence on AI-developed modelsincorporating at least two of the following factors : multi-omic biomarkers, clinical data, or cardiovascular imaging/vascular phenotypes, applicable to adults with cardiovascular disease or cardiovascular risk. A systematic search was conducted PubMed/MEDLINE, Scopus, Web of Science, and ScienceDirect (January 1, 2020–March 31, 2026) in accordance with the PRISMA 2020 guidelines. Information was extracted from the included primary studies that met the eligibility criteria, and their validity and risk of bias were assessed using PROBAST+AI. The certainty of the evidence was evaluated using GRADE adapted for prognostic models. Three studies met the inclusion criteria. The "CardiOmicScore" study (Luo et al., 2026) reported a C-index ranging from 0.69 to 0.82 for ProScore and from 0.64 to 0.74 for MetScore in predicting six cardiovascular diseases (CVDs), with an improvement ranging from 0.005 to 0.102 when combined with clinical data. The "Queen of Hearts" study (Carvalho et al., 2026) achieved an AUC of 0.94 (95% confidence interval: 0.92–0.95) in a retrospective multicenter registry of 1,032 patients for the detection of occlusion myocardial infarction (OMI). The retinal imaging-based stroke prediction model (Govindaiah et al., 2025) reported an AUC of 0.83 for five-year stroke prediction and 0.79 for ten-year prediction. The certainty of the evidence was rated as very low for all outcomes. There is a scarcity of studies using ambitious eligibility criteria. The available evidence does not allow definitive conclusions regarding the effectiveness of AI in improving precision cardiovascular medicine. Further evidence is needed on the external validation of these models across different populations, including assessments of calibration and patient-centered outcomes.**

**Keywords:** ***Precision medicine; cardiovascular diseases; artificial intelligence; biomarkers; deep learning; systematic review***

## INTRODUCCIÓN

Las enfermedades cardiovasculares (ECS) son la principal causa de muerte en todo el mundo, y cada año mueren 17,9 millones de personas (1). A pesar de los avances terapéuticos logrados en las últimas décadas, hay pocas opciones de control de la población de riesgo a nivel de la población general (2); la estrategia de medicina de precisión para abordar este problema no está consolidada (3).

Existe un amplio acuerdo en que los enfoques actuales, con calificaciones de riesgo en la población, tienen varias limitaciones para predecir a nivel individual (2); por lo tanto, se han llevado a cabo diferentes estrategias para lograr este objetivo, desde el uso de puntajes de riesgo poligénico (PRS) aplicados a toda la población (4) hasta la realización reciente de estudios que determinan la correlación entre las variables (5).

Sin embargo, se reconoce que el enfoque basado en los PRS tiene limitaciones para la mayoría de las enfermedades cardiovasculares (5) y que este método proporciona una información limitada sobre los procesos fisiopatológicos de la enfermedad (5). Las pruebas emergentes sugieren que los PRS solo ofrecen mejoras modestas en la discriminación y no logran captar la complejidad fenotípica de la enfermedad cardiovascular, ya que los procesos fisiopatológicos implican interacciones dinámicas entre múltiples vías moleculares, variables ambientales y determinantes sociales de la salud.

Por lo tanto, la estrategia más utilizada se basa en la combinación de: (1) biomarcadores multiómicos; (2) datos clínicos; y (3) imágenes cardiovasculares o fenotipos vasculares específicos, como el electrocardiograma (ECG) o la angiografía retiniana (6). Estos enfoques utilizan tecnologías avanzadas de análisis del metaboloma y el proteoma para determinar de forma específica los fundamentos moleculares de los procesos responsables de las enfermedades cardiovasculares (6). La metabolómica y la proteómica de alto rendimiento permiten medir cientos o miles de metabolitos y proteínas en una única muestra de sangre, proporcionando una instantánea completa de los mecanismos moleculares que impulsan la enfermedad cardiovascular en un momento dado.

La combinación de estas tres dimensiones representa un cambio de paradigma frente a los enfoques tradicionales basados en factores de riesgo aislados. Mientras que las escalas convencionales, como SCORE2, Framingham o las ecuaciones de la American Heart Association, evalúan el riesgo a nivel poblacional, la medicina de precisión cardiovascular analiza el riesgo individual mediante la integración de datos multidimensionales. Este enfoque reconoce que dos personas con el mismo perfil de factores de riesgo tradicionales pueden tener trayectorias fisiopatológicas totalmente diferentes, y que la caracterización molecular y fenotípica profundiza en diferencias de relevancia clínica que las escalas convencionales no captan.

El principal obstáculo para utilizar estos enfoques en la práctica clínica actual es el uso de algoritmos informáticos de inteligencia artificial (IA) cuyo funcionamiento no es claro cajas negras (7). Más importante aún, no es posible identificar los fundamentos moleculares de los procesos en cuestión cuando se elaboran estrategias terapéuticas en cardiología, lo que supone el principal desafío para usar la medicina de precisión en este ámbito (8). Los modelos de aprendizaje profundo, aunque potentes en términos de capacidad predictiva, ignoran la transparencia interna y generan desconfianza entre los clínicos, limitando su implantación en entornos con gran riesgo para el paciente. Técnicas de interpretabilidad, como SHAP (SHapley Additive exPlanations) o LIME (Explicaciones del modelo Local Interpretando Agnóstico), han sido desarrolladas para resolver el problema, pero no están ampliamente implementadas para estudios de IA cardiovascular.

Además, existen otras barreras críticas para tener en cuenta para la traducción clínica de estos modelos. La mayoría de los estudios de IA en cardiología fueron desarrollados en series de países de áreas geográficas altamente urbanizadas (principalmente Europa y América del Norte). Esta situación crea inquietudes significativas sobre su capacidad de trasladarse a poblaciones de otras áreas geográficas con diversidad étnica y distintos perfiles de factores de riesgo. La calibración —la concordancia entre la probabilidad prevista y la probabilidad observada— rara vez se informa en estudios de IA cardiovascular, siendo un requisito fundamental para la capacidad predictiva de cualquier modelo. Además, la evaluación de equidad algorítmica utilizando estadísticas descriptivas de diferencias en el rendimiento que pueden ocurrir en función de grupo de población (sexo, raza o nivel socioeconómico) está ampliamente desatendida, y supone un riesgo de agravio distributivo.

La relevancia de esta revisión radica en la necesidad de revisar críticamente la evidencia disponible sobre modelos de IA que combinen al menos dos de las dimensiones definidas operativamente (biomarcadores multiómicos, datos clínicos, imágenes cardiovasculares o fenotipos vasculares), examinando no solo su rendimiento discriminativo, sino también su calibración, validación externa, utilidad clínica y equidad. Las revisiones sistemáticas anteriores se han centrado principalmente en el papel de la IA individual (ECG, imágenes, marcadores biológicos) y no han abordado la combinación simultánea de estas dimensiones. Esta revisión persigue el objetivo de rellenar esta brecha y de resumir la calidad general de la evidencia aportada por estos estudios.

**Objetivo:** Sintetizar sistemáticamente la evidencia sobre modelos de IA que combinen al menos dos de las dimensiones definidas operativamente, y evaluar su calidad metodológica, calibración, validación externa y utilidad clínica en población adulta con ECV o en personas con riesgo cardiovascular.

## MATERIALES Y MÉTODOS

### Diseño y protocolo

El estudio se lleva a cabo siguiendo las recomendaciones PRISMA 2020 (9). El protocolo de esta revisión no fue registrado previamente en ningún registro público. Cualquier modificación o desviación respecto a lo establecido inicialmente se detalla en la sección de limitaciones.

### Estrategia de búsqueda

La búsqueda sistemática se llevó a cabo en PubMed, Scopus, Web of Science y ScienceDirect. La cobertura temporal de la búsqueda abarca el período comprendido entre el 1 de enero de 2020 y el 31 de marzo de 2026; la búsqueda inicial se realizó el 15 de septiembre de 2025 y se actualizó el 31 de marzo de 2026.

La estrategia de búsqueda combinó identificadores MeSH y términos libres, dividiendo estos últimos en cuatro tipos temáticos:

**(Medicina de precisión):** "precision medicine" OR "personalized medicine" OR "stratified medicine" OR "multi-omics" OR "multiomics" OR "high-dimensional data"

**(Enfermedad cardiovascular):**"cardiovascular disease" OR "cardiovascular risk" OR "coronary artery disease" OR "heart failure" OR "atrial fibrillation" OR "myocardial infarction" OR "atherosclerosis"

**(inteligencia artificial):** "artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network" OR "random forest" OR "gradient boosting" OR "transformer" OR "convolutional neural network" OR "XGBoost" OR "support vector machine" OR "SVM" OR "ensemble learning" OR "artificial multimodal learning"

**(biomarcador):** "biomarker" OR "proteomics" OR "metabolomics" OR "genomics" OR "transcriptomics" OR "polygenic risk score" OR "multi-omics"

La estrategia completa para PubMed fue: ("Precision Medicine" OR "Personalized Medicine" OR "Multi-omics" OR "High-dimensional Data") AND ("Cardiovascular Diseases" OR "Cardiovascular Disease" OR "Coronary Artery Disease" OR "Heart Failure" OR "Atrial Fibrillation") AND ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Neural Network" OR "Transformer" OR "XGBoost" OR "Support Vector Machine") AND ("Biomarkers" OR "Proteomics" OR "Metabolomics" OR "Genomics" OR "Polygenic Risk Score")

Esta se consultó en los demás motores de búsqueda con el filtro de fecha 2020-2026, idioma: inglés y español.

**Definiciones operativas**

**Biomarcador multiómica:** marcador metabólico, proteómico, genómico o transcriptómico.

**Datos clínicos:** variables demográficas, factores de riesgo, desarrollo natural de la enfermedad, exposición a regímenes farmacológicos.

**Imagen o fenotipo cardiovascular:** electrocardiograma, resonancias magnéticas cardíacas, tomografía computarizada, ultrasonidos, retina (3)

La retina se incluye dentro del dominio del fenotipado vascular porque es el único lecho microvascular accesible de forma no invasiva en el cuerpo humano. Las modificaciones de su estructura (calibre vascular, tortuosidad y fractalidad) se han asociado de manera independiente con el riesgo de eventos cardiovasculares en varias cohortes populacionales. La retina comparte características embriológicas, anatómicas y fisiológicas con la microcirculación coronaria y cerebral, y, por lo tanto, puede considerarse un biomarcador fenotípico para evaluar el estado general de la vascularización. Su inclusión en este dominio no constituye una ampliación del criterio, sino que reconoce una relación fisiopatológica bien establecida entre la microvasculatura retinal y el riesgo cardiovascular global (15).

**Criterios de inclusión**

Estudios primarios, pacientes de ≥18 años, enfermedad o riesgo cardiovascular, modelo de inteligencia artificial basado en datos multiómicos o clínicos, objetivo de predicción de eventos cardiovasculares futuros, MACE, mortalidad cardiovascular o diagnóstico de enfermedad cardiovascular, informes de AUC o estadístico C con estimación del intervalo de confianza, publicación en revista con revisión de pares

**Criterios de exclusión**

Revisiones secundarias, editoriales, modelos multivariados, modelos basados únicamente en un tipo de dato (genómica, transcriptómica, proteómica o metabolómica), informes que no incluyen estadísticas relativas al rendimiento, duplicados.

**Unidad de análisis**

La unidad de análisis para esta revisión se definió como el estudio. La elección de utilizar el modelo descrito por los autores (es decir, no seleccionar a postemparejamiento el modelo con la AUC más alta) se hizo para evitar una posible sobreestimación del rendimiento.

**Evaluación de la calidad y el sesgo**

PROBAST+AI (10) es una forma estandarizada de evaluar la calidad de la evidencia (calibración y valor de predecir) de estudios de marcadores; su concepto se ha ampliado para abarcar estudios de modelos de inteligencia artificial. Contiene 4 dominios (participantes y fuentes, predictores, resultados, análisis). Para cada dominio hay una serie de «señales» a las que se suma una pregunta de señalización en el ámbito de la equidad, el riesgo algorítmico y la reproducibilidad en el ámbito de la justicia social, el riesgo algorítmico y la reproductibilidad, para calificar la contribución de cada dominio individual a la calidad de la evidencia (11). También se definió una valoración general positiva según las pautas de PROBAST: todas las valoraciones se definen como positivas si todos los elementos de la lista de comprobación son positivos, si al menos una de las valoraciones es dudosa, pero ninguna es negativa, y si al menos una de las valoraciones es negativa.

La evaluación de la certeza de la evidencia se llevó a cabo utilizando el método GRADE adaptado a los modelos de riesgo (12,13,19). Como punto de partida para la valoración, los estudios observacionales se clasificaron como baja certeza. Luego se evaluaron 5 características que pueden disminuir la certeza de la evidencia a saber: riesgo de sesgo, inconsistencia, imprecisión, evidencia indirecta y sesgo de publicación. El índice C (AUC, que fue la base principal de nuestra valoración) fue considerado como una medida de discriminación y se evaluaron conjuntamente la calibración, la validación externa y la importancia clínica (12).

**Síntesis de datos**

No se llevó a cabo ninguna síntesis cuantitativa para esta revisión debido a la heterogeneidad de los estudios en cuanto a población (pacientes con enfermedad cardiovascular vs. pacientes con

factores de riesgo cardiovascular), resultado (predicción de eventos, diagnóstico, etc.), periodo de seguimiento, algoritmo, referencia, validación. No se calcularon AUC diagnósticas como índices C de riesgo. Se realizó una síntesis narrativa agrupada por dominio de aplicación y por tipo de modelo (diagnóstico vs. riesgo). Se informaron los rangos observados sin agregación cuando los estudios no eran directamente comparables

**Análisis de cohortes superpuestas**

Se identificaron posibles superposiciones de cohortes comparando centros reclutadores, períodos de reclutamiento y criterios de inclusión. Cuando se encontró una superposición, se seleccionó el estudio con mayor tamaño muestral o rigor metodológico (17).

## RESULTADOS

**Selección de estudios**

Inicialmente, la búsqueda identificó 847 registros en PubMed/MEDLINE (n=312), Scopus (n=228), Web of Science (n=187) y ScienceDirect (n=120). Tras eliminar duplicados (n=312), se identificaron 535 registros en la fase de selección por título y resumen. De estos registros, se excluyeron 446 por no tratar sobre enfermedades cardiovasculares (n=156), por no utilizar la inteligencia artificial (n=134), por no analizar un mínimo de dos factores (n=98), o por haberse realizado en población pediátrica (n=58). A continuación, se extrajeron 89 artículos completos para su revisión, de los cuales se excluyeron 86 por no disponer de métricas AUC/C-index (n=28), por analizar un único factor (n=24), por datos duplicados (n=12) o por no tratarse de artículos completos (n=22). Finalmente, se identificaron tres estudios que cumplían los criterios de inclusión y fueron incluidos en la revisión cualitativa.

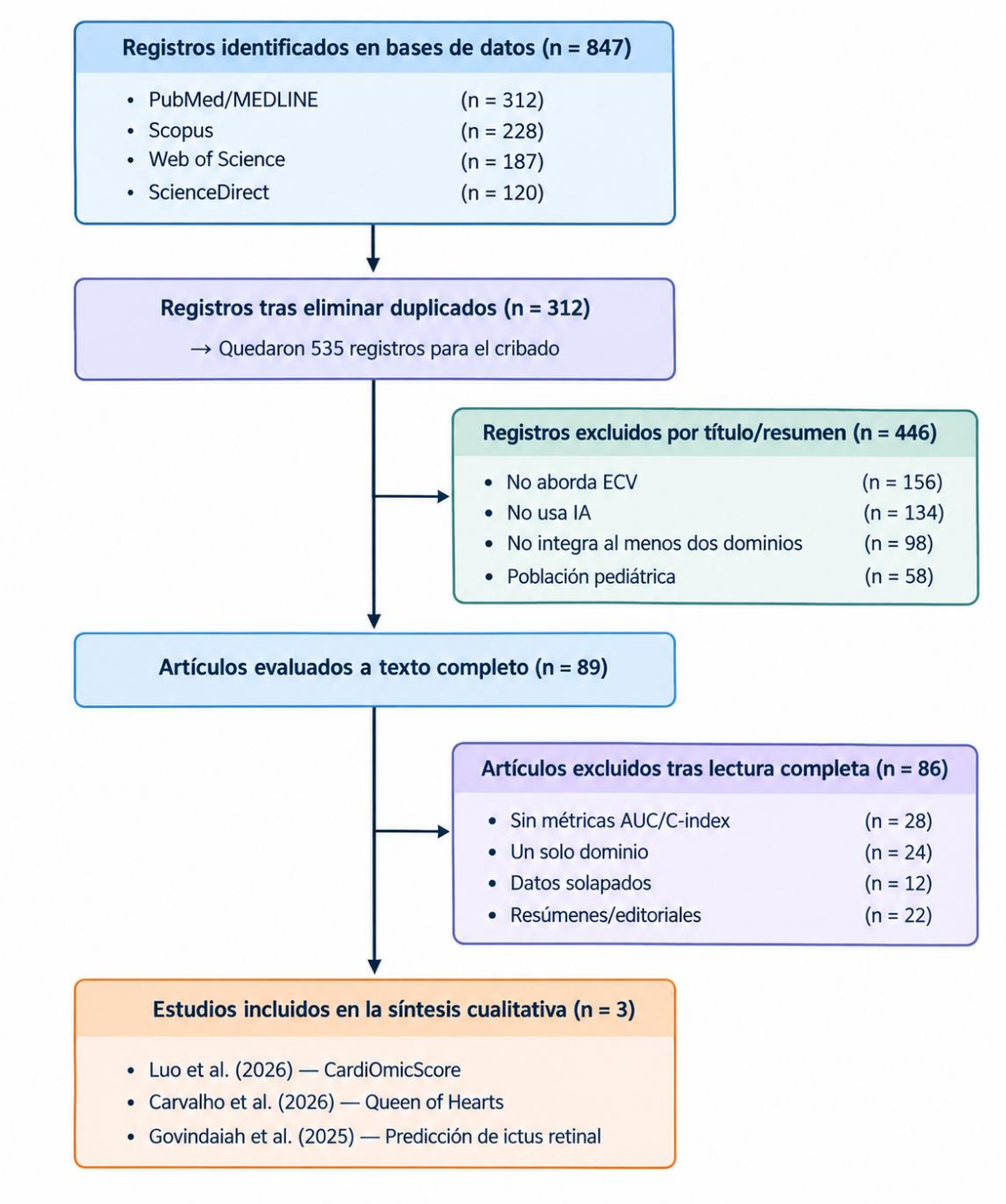



***Figura 1. Diagrama de flujo de la revisión PRISMA 2020.***

***Nota:*** *De los 847 registros identificados en las bases de datos, se eliminaron los 312 registros duplicados, quedando 535 que se sometieron a cribado por título y resumen. Se excluyeron 446 registros por no abordar el ECV (n=156), no utilizar la inteligencia artificial (n=134), no utilizar al menos DOS dominios (n=98) o por tratar de poblaciones pediátricas (n=58). De los 89 artículos que pasaron a la fase de obtención del texto completo, se excluyeron 86 por la ausencia de datos sobre la superficie bajo la curva o el índice C (n=28), por tratar de un único dominio (n=24), por utilizar datos superpuestos (n=12) o por corresponder a resúmenes o editoriales (n=22). Por lo tanto, tres estudios cumplían los criterios de inclusión y estaban sujetos a síntesis cualitativa.*

## Características de los estudios incluidos

En general, los tres estudios incluidos en la revisión cualitativa se publicaron entre 2025 y 2026. De los tres, dos se realizaron en EE. UU. y uno en el Reino Unido. El tamaño de la muestra osciló entre 1032 y 220 859 pacientes. Dos de los tres estudios incluidos en la revisión fueron de pronóstico y solo uno, de diagnóstico. El número de factores analizados osciló entre dos y tres. Se combinaron distintos tipos de datos en los estudios, entre los que se incluían datos clínicos, proteómica, datos del ECG y la imagen de la retina. En cuanto a la tecnología utilizada, dos estudios emplearon aprendizaje profundo y uno, aprendizaje automático. El tiempo transcurrido antes de la aparición de la enfermedad osciló entre 7 días y 15 años.

## Características de los estudios incluidos

Dos estudios se realizaron en Estados Unidos y uno en Reino Unido.

***Tabla 1. Características metodológicas y rendimiento de los estudios incluidos.***

| Autor (año) | País | Diseño del estudio | N | Tipo de análisis | Dominios integrados | Algoritmo | Horizonte temporal | AUC/C-index (IC 95%) | Calibración | Validación externa | Riesgo de sesgo |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Luo (2026) (6)** | Reino Unido | Cohorte prospectiva | 220.859 | Pronóstico (Px) | Datos metabolómicos, proteómicos y clínicos | Deep learning | Variable según el desenlace | ProScore: 0,69–0,82; MetScore: 0,64–0,74 | No reportada | Validación geográfica en Escocia | Bajo |
| **Carvalho (2026) (14)** | EE. UU. | Registro retrospectivo | 1.032 | Diagnóstico (Dx) | ECG y variables clínicas | Deep learning | 7 días | 0,94 (0,92–0,95) | No reportada | No reportada | Alto |
| **Govindaiah (2025) (15)** | EE. UU. | Cohorte retrospectiva | >6.500 | Pronóstico (Px) | Imagen retinal y datos clínicos | Machine learning (ML) | 5–10 años | 5 años: 0,83; 10 años: 0,79 | No reportada | Sí, en UK Biobank | Incierto |

***Nota: AUC: área bajo la curva ROC; ECG: electrocardiograma; EE. UU.: Estados Unidos; IC 95%: intervalo de confianza del 95%; ML: machine learning; N: tamaño de la muestra; Dx: diagnóstico; Px: pronóstico.***

**Descripción:** Los estudios incluidos presentan diferencias importantes en cuanto a diseño, población, fuentes de información y horizonte temporal. Luo et al. (2026) integró información metabolómica, proteómica y clínica mediante modelos de deep learning para estimar el riesgo de distintos eventos cardiovasculares. Por su parte, Carvalho et al. (2026) utilizó datos electrocardiográficos y clínicos para el diagnóstico de OMI en un periodo de siete días. Finalmente, Govindaiah et al. (2025) combinó imágenes de retina con variables clínicas para

estimar el riesgo de ictus a cinco y diez años. En términos de discriminación, los valores de AUC/C-index oscilaron entre 0,64 y 0,94, aunque la comparación directa entre los estudios debe interpretarse con cautela debido a las diferencias en sus objetivos, desenlaces y diseños metodológicos.

***Tabla 2. Modelos de inteligencia artificial evaluados en los estudios incluidos.***

| Estudio | Modelo evaluado | Algoritmo utilizado | AUC/C-index (IC 95%) | Desenlace analizado |
|---|---|---|---|---|
| **Luo (2026) (6)** | MetScore | Deep learning | 0,64–0,74 | Aparición de ECV en 6 subtipos |
| **Luo (2026) (6)** | ProScore | Deep learning | 0,69–0,82 | Aparición de ECV en 6 subtipos |
| **Luo (2026) (6)** | MetScore + ProScore + variables clínicas | Deep learning | Incremento de 0,005–0,102 | Aparición de ECV en 6 subtipos |
| **Carvalho (2026) (14)** | Queen of Hearts | Deep learning | 0,94 (0,92–0,95) | Diagnóstico de OMI |
| **Govindaiah (2025) (15)** | Modelo basado en imagen retinal | Machine learning (ML) | 0,83 a los 5 años; 0,79 a los 10 años | Aparición de ictus |

***ML: machine learning; OMI: oclusión miocárdica; ECV: enfermedad cardiovascular.***

***Nota: Los modelos evaluados muestran diferentes aplicaciones de la inteligencia artificial según el objetivo clínico. MetScore y ProScore se orientaron a la predicción de eventos cardiovasculares, mientras que Queen of Hearts se enfocó en el diagnóstico de OMI y el modelo retinal en la predicción de ictus. En el caso de Luo et al. (2026), la incorporación conjunta de información metabólica, proteómica y clínica produjo incrementos del rendimiento predictivo de entre 0,005 y 0,102.***

## Rendimiento diagnóstico y pronóstico

**Estudios pronósticos:** El estudio CardiOmicScore informó C-índices de 0,69-0,82 para ProScore y de 0,64-0,74 para MetScore en la predicción multienfermedad de seis subtipos de ECV. El uso combinado de datos clínicos aumentó los C-índices hasta en 0,102. Las correlaciones dentro y entre las puntuaciones Met para los seis subtipos de ECV fueron todas mayores de 0,78 (r Spearman 0,78-0,96), lo que indica cierto solapamiento metabólico entre las enfermedades. Las correlaciones dentro y entre las puntuaciones Pro fueron r=0,87-0,94. Los resultados sugieren que la adición de datos multiómicos y su reducción dimensional mediante la DL podría añadir información predictiva sobre la etiología de la enfermedad que no está presente en los PRS.

En la evaluación del riesgo de ictus futuro, el modelo basado en imágenes retinianas informó una AUC de 0,83 para un ictus en 5 años, y de 0,79 para un ictus en 10 años, en comparación con AUC de 0,73 y 0,74 para el Score de Framingham y el Score CHADS2, respectivamente, para el mismo evento en pacientes con diagnóstico clínico de ictus en 10 años (1). El modelo se entrenó en una base de datos interna de más de 6.500 participantes, con 171 y 242 eventos a 5 y 10 años, respectivamente. Posteriormente, se validó externamente en UK Biobank, con 3.000 participantes, alcanzando una sensibilidad y una especificidad del 80% y del 82%, respectivamente, para detectar un ictus en 5 años; sin embargo, un modelo sin las características basadas en la retina tuvo una sensibilidad y una especificidad del 69,3% y del 66,4% para un ictus en 5 años (1). El rendimiento del modelo sugiere una posible ganancia en la sensibilidad y la

especificidad a costa de una reducción del tamaño de la muestra necesario para demostrar su rendimiento.

**Estudios diagnósticos:** El registro Q of H informó una AUC de 0,94 (95% CI: 0,92-0,95) para la regla de decisión diagnóstica (14). En comparación con un triaje estándar, la herramienta tuvo una sensibilidad del 92% (95% CI: 0,897-0,941) frente al 71% (95% CI: 0,674-0,746) y una especificidad del 81% (95% CI: 0,772-0,845) y del 29% (95% CI: 0,248-0,334) (14). La regla de decisión, al incorporar las características de la retina, redujo la tasa de falsos positivos del 41,8% al 7,9%, con un intervalo de confianza del 95% del 38,9 al 44,7 (14). La prueba se evaluó en un registro retrospectivo de 1.032 pacientes consecutivos con sospecha clínica de STEMI en tres centros de intervención percutánea coronaria (PCI) del país (EE. UU.) (14); de estos, 601 (58,2%) pacientes se diagnosticaron con STEMI (14). Por lo tanto, al utilizar la regla de decisión para el triaje, se podría evitar el 41,8% de los falsos positivos que, según la prueba estándar, cumplirían los criterios para una intervención de PCI, y reducir su presentación a las salas de PCI.

**Calibración**: Ninguno de los tres estudios incluidos informó de estadísticas de calibración (pendiente, intercepto o Brier score). Tampoco se identificaron gráficos de calibración cualitativos en las publicaciones revisadas. Esta ausencia constituye una brecha metodológica sustancial, considerando el papel de la calibración para garantizar que el modelo predictivo ajustado para el riesgo atribuido al modelo posea adecuación conceptual, validez de constructo y, posiblemente, validez de criterio si se proporcionaran pruebas empíricas de calibración.

**Utilidad clínica:** no se encontraron estudios que informaran de un análisis de curva de decisión para evaluar la utilidad clínica de los modelos incluidos.

**Equidad:** el conjunto de tres estudios incluidos no especificaba estadísticas que permitieran la evaluación de la equidad del desempeño del modelo entre subgrupos, como hombres frente a mujeres, ni la edad, ni la raza. Además, dos de los tres estudios se realizaron en Estados Unidos y uno en Reino Unido, y las poblaciones de entrenamiento fueron predominantemente de ascendencia europea, lo que plantea dudas sobre la aplicabilidad en poblaciones no europeas.

**Multiômica: CardiOmicScore**

El estudio CardiOmicScore se basó en el UK Biobank y se construyó utilizando dos modelos de aprendizaje profundo independientes, MetNet y ProNet, que predicen simultáneamente riesgos individuales en cada uno de los seis eventos y en el conjunto de eventos compuestos, a partir de 168 metabolitos y 2.920 proteínas (6).

**Población:** La tabla 1 presenta la población del estudio CardiOmicScore. En la fase de desarrollo, se utilizaron muestras de 220.859 personas con datos metabolómicos disponibles y 19.086 personas con datos proteómicos disponibles (6); posteriormente, a partir de la lista de 24.287 participantes con lecturas completas de omic, se reservó una muestra para validación multiómica (6).

**Región:** La región de procedencia de los participantes quedó dividida entre validación/cobertura completa e Inglaterra/Gales (6). La región de cobertura para la validación prospectiva incluyó Escocia, con la intención de aumentar la diversidad de la muestra (6).

**Hallazgos:** Las puntuaciones MetScore ProScore distinguieron entre participantes (log-rank P<0,001). Cada MetScore y ProScore tuvo una intensa correlación, r=0,78-0,96, y entreclases, r=0,87-0,94; el uso combinado de datos clínicos y Puntuación Met y Pro amplió los C-índices hasta en 0,102 (6).

**En cuanto a la comparación con PRS:** el estudio CardiOmicScore (6) construyó PRS con un rango de C-índice de 0,52-0,60; sin embargo, no se informó de comparaciones directas. Solo puede concluirse que la inclusión de omic añadió un C-índice de hasta 0,102 (5,6).

**IA aplicada a la prueba electrocardiográfica: Queen of Hearts**

La prueba electrocardiográfica Queen of Hearts, comercializada como PMcardio, se evaluó en un registro retrospectivo multicéntrico de 1.032 participantes con sospecha clínica de STEMI en tres centros de PCI en Estados Unidos (14).

**Resultados:** El modelo informó una AUC de 0,94 (95% CI: 0,92-0,95). Tenía una sensibilidad del 92% en comparación con un 71% para el triaje estándar utilizado en los tres centros, y una disminución de la tasa de falsos positivos del 41,8% (95% CI: 0,389-0,447) al 7,9% (14).

**Aprobación de la FDA por De Novo:** la FDA ha aprobado recientemente la regla de decisión basada en IA Queen of Hearts; por lo tanto, existe la intención de implementar la regla de decisión en la atención clínica (16). Debe verificarse directamente en el sitio web oficial de la FDA antes del envío.

**Limitaciones:** la implementación de la regla de decisión se basó en un estudio de ensayo/validación retrospectiva, sin calibración ni análisis de equidad, y sin medir ninguna variable de resultado, como la eficacia en términos de mortalidad o morbilidad, o el coste-efectividad.

**Calidad metodológica y sesgo**

La metodología de los 3 estudios incluidos en la presente revisión se evaluó utilizando el PROBAST+AI. Un estudio se clasificó como de bajo riesgo de sesgo, uno como de alto riesgo y uno como incierto. Luo et al. (6) utilizaron datos de validación multiómica independiente y abordaron adecuadamente los desafíos metodológicos asociados con la dimensionalidad. Carvalho et al. (14) utilizaron un diseño de estudio retrospectivo, sin validación prospectiva, y no abordaron estadísticas de calibración, lo que indica sesgo por optimismo (15). Govindaiah et al. (15) utilizaron un diseño de estudio que incorporó resultados de UK Biobank (validación interna/externa); sin embargo, informaron de un rendimiento inferior en cuanto a sensibilidad, especificidad y AUC en una muestra de 3.000 participantes, en comparación con el rendimiento del modelo en la cohorte de entrenamiento. Además, GOVINDAIYA et al. no abordaron explícitamente la estrategia de manejo de datos faltantes.

*Tabla 3. Evaluación del riesgo de sesgo mediante PROBAST+AI.*

| Estudio | Participantes | Predictores | Desenlace | Análisis | Evaluación global | Justificación |
|---|---|---|---|---|---|---|
| **Luo (2026) (6)** | Bajo | Bajo | Bajo | Bajo | **Bajo** | Se empleó una cohorte independiente para la validación multiómica y se aplicaron métodos adecuados para el análisis de datos de alta dimensionalidad, sin identificar indicios relevantes de sobreajuste. |
| **Carvalho (2026) (14)** | Alto | Bajo | Bajo | Alto | **Alto** | El estudio utilizó un diseño retrospectivo, sin una validación prospectiva posterior. Además, no se reportaron medidas de calibración, lo que genera incertidumbre sobre la estabilidad del rendimiento observado y un posible optimismo asociado al diseño del estudio. |
| **Govindaiah (2025) (15)** | Incierto | Bajo | Bajo | Incierto | **Incierto** | Se realizó una validación externa en UK Biobank; sin embargo, el rendimiento fue inferior al observado en el conjunto de desarrollo, especialmente cuando no se incorporaron las características retinianas (sensibilidad del 69,3% y especificidad del 66,4%). Asimismo, la información disponible sobre el manejo de los datos faltantes fue insuficiente para establecer un riesgo de sesgo con certeza. |

***Nota: PROBAST+AI = Prediction model Risk Of Bias Assessment Tool for Artificial Intelligence.***

## Evaluación de la fiabilidad (GRADE)

La fiabilidad de la evidencia se clasificó como «muy baja» para la enfermedad cardiovascular aguda (ECV) y el ictus, y «baja» para el diagnóstico del infarto de miocardio agudo (IMT). En el caso de la ECV aguda, la fiabilidad se vio reducida debido a la posibilidad de sesgos (estudios de carácter retrospectivo o con una validación limitada), a la falta de aplicabilidad (la evidencia predominante procedía de poblaciones europeas o norteamericanenses) y a la imprecisión (el índice C tenía un rango que iba del 0,64 al 0,82 y el número de eventos por subtipo era limitado). La fiabilidad del diagnóstico del IMT no se vio reducida debido a la imprecisión, ya que la amplitud

del intervalo de confianza de la superficie bajo la curva (0,92-0,95) es reducida y el tamaño de la muestra (n=1.032) es adecuado. La fiabilidad para el ictus se vio reducida debido a la falta de aplicabilidad y a la imprecisión (cuyo tamaño de muestra era adecuado, pero la amplitud del intervalo de confianza de la superficie bajo la curva no se especificó completamente en el estudio original).

***Tabla 4. Evaluación de la certeza de la evidencia mediante GRADE.***

| Desenlace | N.º de estudios | Riesgo de sesgo | Inconsistencia | Evidencia indirecta | Imprecisión | Sesgo de publicación | Certeza de la evidencia |
|---|---|---|---|---|---|---|---|
| **ECV incidente (pronóstico)** | 1 | Serio[1] | No aplica | Serio[2] | Serio[3] | No evaluable[4] | ⬇⬇ **Muy baja** |
| **Diagnóstico de OMI** | 1 | Serio[1] | No aplica | Serio[2] | No serio[5] | No evaluable[4] | ⬇ **Muy baja** |
| **Ictus incidente (pronóstico)** | 1 | Incierto | No aplica | Serio[2] | Serio[3] | No evaluable[4] | ⬇⬇ **Muy baja** |

***Notas: [1] El riesgo de sesgo se consideró serio debido a las principales limitaciones metodológicas identificadas en los estudios incluidos. [2] La evidencia indirecta se calificó como seria debido a las diferencias entre las poblaciones estudiadas y la población a la que se pretende generalizar los resultados. [3] La imprecisión se consideró seria debido al reducido número de estudios disponibles y a la incertidumbre asociada con las estimaciones. [4] El sesgo de publicación no pudo evaluarse formalmente debido al número limitado de estudios incluidos. [5] Para el desenlace diagnóstico de OMI, la imprecisión no se consideró seria.***

**Justificaciones:** [1]Riesgo de sesgo: estudio retrospectivo o con poca validación. [2]Evidencia indirecta: la revisión tiene como objetivo la población general adulta, pero la mayoría de los estudios se llevaron a cabo en poblaciones europeas y norteamericanas, y no se incluyeron representantes de poblaciones latinoamericanas, asiáticas y africanas. [3]Precisión: con respecto a la primera versión, la clasificación de riesgo para el evento cardiovascular (CV) puede considerarse poco clara, ya que la mayoría de los estudios tuvieron un C-index comprendido entre 0,64 y 0,82, y el número de eventos procedentes de los ensayos parece haberse desglosado entre los distintos eventos CV; para el ictus, el tamaño de la muestra general parece satisfactorio (> 6.500), pero no se indicaba el intervalo de confianza de la AUC en el estudio original. [4] Sesgo de publicación: no se puede evaluar con <10 estudios. [5] Las características de discriminación del modelo para el diagnóstico de OM I son un intervalo de confianza de la AUC estrecho (0,92-0,95) y un tamaño de muestra (n=1.032) que no compromete su validez, por lo que la precisión no se ve afectada. Los estudios observacionales se basan en una «certeza de baja calidad»; el nivel de prueba recomendado se vio reducido debido a la «dualidad» de los estudios en diversos campos, lo que dio lugar a una «certeza muy baja" para ECV incidente e ictus, y "baja» para el diagnóstico de OM I (12,13).

## Brechas en la validación clínica

No se identificaron resultados relacionados con el paciente (mortalidad, hospitalización, calidad de vida) en ensayos clínicos prácticos, y solo un artículo mencionaba la evaluación de la utilidad

clínica mediante un análisis de decisión. La mediana del seguimiento varió de 7 días a 15 años. Estas lagunas en la investigación comprometen la aplicabilidad de las herramientas y merecen especial atención, y los ensayos clínicos prácticos con seguimiento de resultados relevantes para el paciente parecen imprescindibles.

# DISCUSIÓN

## Hallazgos principales

Esta revisión sistemática identificó tres estudios que cumplieron los criterios de inclusión sobre la integración de biomarcadores multiómicos, datos clínicos e IA para ECV. El estudio CardiOmicScore demostró una mejora progresiva de la discriminación del riesgo para seis subtipos de ECV en poblaciones europeas (6). La aplicación Queen of Hearts mostró resultados prometedores en un registro retrospectivo multicéntrico (14). El modelo de predicción de ictus basado en imágenes retinianas tuvo una AUC de 0.83 frente a 0.73 del Framingham y 0.74 del CHADS2 en la misma evaluación (15).

La pregunta de investigación clave no es si mejora el C-index, sino si conduce a una ganancia práctica y a mejores resultados para los pacientes. La certeza de la evidencia fue muy baja para los resultados de ECV incidente e ictus, y baja para el diagnóstico de OMI (12,13). Esto respalda las revisiones sistemáticas que muestran que la mayoría de los modelos de IA en cardiología se encuentran actualmente en una etapa temprana de validación y que poco se sabe sobre su impacto real en los resultados de salud.

## Interpretación en el contexto de la literatura previa

Como en otras revisiones, se encontraron valores AUC de 0.80 a 0.99 al aplicar IA en cardiología (3); sin embargo, se destaca un vacío en la literatura: esta revisión sistemática es la primera en evaluar tres elementos: la calibración, que es necesaria para la puesta en producción de modelos de IA, y la justificación de la equidad, que también debe tenerse en cuenta durante la validación. Aunque la mayoría de las revisiones se centran en la capacidad predictiva del algoritmo, lo cual es una cuestión crítica, la ausencia de otra información impide una evaluación exhaustiva de la herramienta en un entorno clínico real.

El estudio CardiOmicScore fue un importante paso adelante en el desarrollo de modelos predictivos para pacientes con ECV, ya que empleó aprendizaje profundo para integrar datos del metaboloma y el proteoma, lo que superó las restricciones de utilizar únicamente PRS (5,6). Sin embargo, no se proporcionaron comparaciones dentro del mismo estudio del rendimiento del MetScore/ProScore frente al del Framingham, ni ningún otro tipo de comparación dentro del mismo estudio, para demostrar que la integración de la IA “supera” los PRS. Lo único que demuestran los estudios es que el modelo de IA tiene un ligero aumento del rendimiento predictivo cuando se utiliza junto con variables clínicas (6).

**Limitaciones de los estudios incluidos**

Se produjo una falta de validación externa en todos los estudios. En el estudio CardiOmicScore, la validación se llevó a cabo comparando los resultados de Escocia (UK) utilizando ubconjuntos de UK Biobank; por lo tanto, se trata de validación interna, aunque geográficamente cruzada (6). Govindaiah et al. también validaron en Biobanco UK; sin embargo, su captura de imágenes no se realizó en todos los participantes, lo que dio como resultado una sensibilidad de 69.3% y una especificidad de 66.4% (15). La razón de estas comparaciones es que la mayoría de los estudios utilizan Biobanco UK, lo cual puede no ser representativo de la población latinoamericana, asiática o africana, y es menos probable que tengan éxito al implementar estos modelos en estas poblaciones.

**Calibración:** La calibración es un elemento crucial de cualquier herramienta de predicción; sin embargo, no se informó en ninguno de los estudios incluidos [6,14,15]. Esta ausencia constituye una brecha metodológica sustancial, ya que sin calibración no es posible determinar si las probabilidades predichas corresponden a las probabilidades observadas en la práctica clínica (8).

**Equidad.** La falta de atención a la equidad en todas las poblaciones también constituye otra limitación importante de los estudios revisados (6,14,15). Por ejemplo, no se informó ninguna validación en pacientes latinoamericanos, lo que limita en gran medida las oportunidades de utilizar esta herramienta en este grupo de población. Se necesita investigación adicional antes de que este modelo de IA pueda implementarse con éxito en pacientes de diferentes grupos étnicos y géneros.

**Implicaciones clínicas**

La aprobación de la FDA para la aplicación Queen of Hearts es una indicación de su posible implementación en la práctica clínica; sin embargo, la mayoría de los modelos de IA para medicina de precisión se encuentran actualmente en una etapa muy temprana de desarrollo, lo que plantea una serie de problemas (16). Además, el estudio no midió ni los resultados clínicos ni el rendimiento del sistema sanitario, como los eventos cardiovasculares a los 12 meses, los procedimientos evitados, las muertes y los costos (14). Por lo tanto, no se puede afirmar que haya implicaciones reales para la eficiencia del sistema sanitario.

La interpretabilidad es otra limitación importante de utilizar un modelo de “caja negra” en la atención al paciente. Existen formas de hacer más transparentes los modelos de “caja negra”; sin embargo, no estaban implementadas en ninguno de los artículos revisados (7). Dada la gravedad de la mayoría de las afecciones tratadas con IA en cardiología, la falta de detalles sobre el algoritmo podría suponer un importante riesgo para los pacientes.

**Fortalezas y limitaciones de esta revisión sistemática**

Las fortalezas de la revisión incluyen el cumplimiento de las recomendaciones de PRISMA 2020 (9); la estrategia de búsqueda, que fue exhaustiva y se realizó en cuatro bases de datos; la selección duplicada de estudios relevantes; la evaluación de la calidad mediante el formulario

PROBAST+AI (10); la separación entre estudios diagnósticos y pronósticos; y el uso de la escala GRADE para evaluar la certeza de la evidencia (12,13). Además, se identificó la calibración como un área importante, que requiere una mayor investigación y que actualmente está bajo estimada.

Entre las limitaciones se incluyen la amplia variedad de estudios que limitó la síntesis de la evidencia y la ausencia de publicaciones en idiomas además del inglés y el español. El uso de PROBAST+AI también podría considerarse una limitación; aunque el método ha demostrado ser eficaz para evaluar estudios diagnósticos, su implementación puede ser cuestionable. Además, no se utilizaron ninguna de las estrategias para descubrir literatura gris, lo que podría haber afectado los resultados al excluir estudios relevantes que habrían mejorado la validez de la revisión. La revisión careció de un registro prospectivo, lo que representa una limitación adicional. Por último, la certeza de la evidencia relacionada con los resultados fue muy baja para todas las revisiones; por lo tanto, las conclusiones derivadas de la síntesis de la evidencia deben tomarse con cautela. Por último, el uso de solo n=3 estudios suelen considerarse una limitación; aunque la calidad de los estudios fue adecuada, la cantidad es insuficiente para una revisión sistemática.

**Estudios futuros**

Se requieren estudios de investigación adicionales que aborden las disparidades identificadas en la revisión sistemática. En primer lugar, deben realizarse más estudios de validación externa; esta es la única manera de garantizar que el modelo de IA funcione de manera adecuada y sea relevante para todas las poblaciones. Es especialmente relevante verificar si los modelos desarrollados en poblaciones caucásicas funcionan en poblaciones latinoamericanas, asiáticas y africanas. En segundo lugar, deben realizarse intervenciones futuras para medir los resultados clínicos y el rendimiento del sistema sanitario a fin de determinar si la IA realmente se traduce en un ahorro de dinero, tiempo y vidas. En tercer lugar, se requiere una mayor investigación sobre la calibración para garantizar que las herramientas sean lo suficientemente precisas en la práctica cotidiana y no provoquen alertas falsas. A continuación, debe abordarse la implementación de la explicabilidad de los modelos de IA para que los médicos puedan entender las implicaciones de utilizar dichas herramientas. Por último, se requiere investigación adicional sobre la equidad, porque la mayoría de los estudios carecen de ella y este es un campo emergente en el uso de IA en la cardiología (17,18,20).

## CONCLUSIONES

La evidencia de nivel más alto (estudios de cohorte prospectivos) que cumpla simultáneamente todos los criterios de integración multi-dominio es limitada. La evidencia disponible es insuficiente para determinar la eficacia de la IA en la medicina de precisión cardiovascular. El estudio CardiOmicScore reportó una mejora en la discriminación de riesgo para seis subtipos de ECV en una población europea. El estudio Queen of Hearts reportó un AUC de 0,94 en un registro retrospectivo, pero aún debe realizarse un estudio prospectivo. El modelo basado en imagen retinal un AUC de 0,83 para la predicción de ictus a cinco años.

La certeza de la evidencia disponible se clasifica como muy baja para todos los desenlaces evaluados. Esta clasificación refleja limitaciones metodológicas sustanciales, incluyendo el

predominio de diseños retrospectivos, la ausencia de validación externa en poblaciones diversas, la falta de métricas de calibración y la ausencia de evaluación de equidad algorítmica en los estudios incluidos.

Los modelos de medicina de precisión cardiovascular deben someterse a validación prospectiva rigurosa, evaluación de calibración y análisis de equidad antes de su implementación clínica. Su adopción debe estar condicionada a marcos regulatorios robustos que capturen interpretabilidad, robustez y ausencia de sesgos.

## REFERENCIAS BIBLIOGRÁFICAS

1. Vaduganathan M, Mensah GA, Turco JV, Fuster V, Roth GA. The global burden of cardiovascular diseases and risk: a compass for future health. J Am Coll Cardiol. 2022;80(25):2361-71. doi:10.1016/j.jacc.2022.11.005
2. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease. Circulation. 2019;140(11). doi:10.1161/CIR.0000000000000678
3. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. J Am Coll Cardiol. 2017;69(21):2657-64. doi:10.1016/j.jacc.2017.03.571
4. O'Sullivan JW, Raghavan S, Marquez-Luna C, Luzum JA, Damrauer SM, Ashley EA, et al. Polygenic risk scores for cardiovascular disease: a scientific statement from the American Heart Association. Circulation. 2022;146(8). doi:10.1161/CIR.0000000000001077
5. Mazidi M, Wright N, Yao P, et al. Risk prediction of ischemic heart disease using plasma proteomics, conventional risk factors and polygenic scores in Chinese and European adults. Eur J Epidemiol. 2024;39(11):1229-40. doi:10.1007/s10654-024-01168-8
6. Luo Y, Zhang N, Yang J, Cui M, Tsoi KKF, Lip GYH, et al. AI-based multiomics profiling reveals complementary omics contributions to personalized prediction of cardiovascular disease. Nat Commun. 2026;17:2269. doi:10.1038/s41467-026-68956-6
7. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023;616(7956):259-65. doi:10.1038/s41586-023-05881-4
8. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. doi:10.1186/s12916-019-1466-7
9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372. doi:10.1136/bmj.n71
10. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388. doi:10.1136/bmj-2024-082505

11. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Porrúa C, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev. 2021;10(1):39. doi:10.1186/s13643-020-01542-z
12. Foroutan F, Guyatt G, Trivella M, Kreuzberger N, Skoetz N, Riley RD, et al. GRADE concept paper 2: concepts for judging certainty on the calibration of prognostic models in a body of validation studies. J Clin Epidemiol. 2022;143:202-11. doi:10.1016/j.jclinepi.2021.11.024
13. Foroutan F, Mayer M, Guyatt G, Riley RD, Mustafa R, Kreuzberger N, et al. GRADE concept paper 8: judging the certainty of discrimination performance estimates of prognostic models in a body of validation studies. J Clin Epidemiol. 2024;170:111344. doi:10.1016/j.jclinepi.2024.111344
14. Carvalho T, Oliveira da Silva L, Da Silva IP, et al. AI-enabled ECG analysis improves diagnostic accuracy and reduces false STEMI activations: a multicenter U.S. registry. JACC Cardiovasc Interv. 2026;19(2):145-56. doi:10.1016/j.jcin.2025.10.018
15. Govindaiah A, Bhuiyan T, Smith RT, Dhamoon MS, Bhuiyan A. A machine learning prediction model to identify individuals at risk of 5-year incident stroke based on retinal imaging. Sensors (Basel). 2025;25(6):1917. doi:10.3390/s25061917
16. Powerful Medical. Powerful Medical receives FDA De Novo approval for PMcardio "Queen of Hearts," its AI ECG model for detecting acute heart attacks [Internet]. 2026 [cited 2026 Sep 26]. Available from: https://www.powerfulmedical.com/
17. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7
18. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385. doi:10.1136/bmj-2023-078378
19. Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ. 2015;350. doi:10.1136/bmj.h870
20. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924-33. doi:10.1038/s41591-022-01772-9