

**Aplicación de aprendizaje automático para la detección temprana de enfermedades  
crónicas en Estados Unidos: caso de estudio en diabetes tipo 2 y glaucoma**

Leszly Dahiana Bustos Guerrero

Asesor

Andrea Ximena Robles Uriza

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2025

## **Dedicatoria**

Dedico este trabajo a la disciplina, al aprendizaje constante y a la voluntad de mejora continua.

A las horas de estudio silencioso, a los errores que enseñaron más que los aciertos y a cada pequeño avance que construyó este resultado.

A quienes comparten la convicción de que el conocimiento transforma.

### **Agradecimientos**

Agradezco a mi tutor(a), Andrea Ximena Robles Uriza, por su orientación académica, su disposición constante y sus valiosos aportes al desarrollo de esta monografía.

A los docentes que, a lo largo del proceso formativo, no solo enseñaron contenidos, sino también formas de pensar críticamente.

A las instituciones que facilitaron el acceso a recursos, bibliografía y herramientas técnicas necesarias para este trabajo.

A mis colegas y compañeros(as), por los debates, las ideas compartidas y el aprendizaje colaborativo.

Y al rigor académico, que convirtió la duda en motor de búsqueda y el proceso en un verdadero ejercicio de construcción.

## Resumen

Esta monografía examina el potencial del aprendizaje automático en la detección temprana de enfermedades crónicas en Estados Unidos, enfocándose en la diabetes tipo 2 y el glaucoma. A través del análisis cualitativo de 22 artículos científicos seleccionados entre 2016 y 2025, se identifican y comparan distintos enfoques técnicos, clínicos y éticos en la aplicación de modelos de inteligencia artificial. Los estudios revisados abarcan algoritmos de ensamble, redes neuronales convolucionales, modelos híbridos (como CNN+LSTM y redes generativas adversariales —Generative Adversarial Networks, GAN), técnicas de explicabilidad (como SHAP —SHapley Additive exPlanations— y mapas de calor), así como estrategias de integración con registros electrónicos de salud (Electronic Health Records, EHR). Los hallazgos evidencian un alto desempeño en métricas como precisión, sensibilidad y área bajo la curva (Area Under the Curve, AUC), superando en algunos casos el rendimiento de especialistas humanos. Sin embargo, persisten limitaciones importantes como la baja precisión positiva, la falta de validación externa, la opacidad de ciertos modelos y las desigualdades estructurales en el acceso a tecnologías de diagnóstico. Asimismo, se abordan barreras éticas y normativas vinculadas a la privacidad de datos —como lo establecido por la Ley de Portabilidad y Responsabilidad de Seguros de Salud (Health Insurance Portability and Accountability Act, HIPAA)— y la regulación de tecnologías médicas por parte de la Administración de Alimentos y Medicamentos (Food and Drug Administration, FDA). Se concluye que, si bien la inteligencia artificial representa una herramienta prometedora para optimizar la medicina preventiva, su adopción efectiva requiere superar retos técnicos, sociales y regulatorios mediante estrategias inclusivas, éticamente fundamentadas y validadas en contextos reales.

***Palabras clave:*** Aprendizaje automático, detección temprana, diabetes tipo 2, glaucoma, inteligencia artificial, explicabilidad, sesgo algorítmico, salud digital.

## Abstract

This monograph explores the potential of machine learning for the early detection of chronic diseases in the United States, with a focus on type 2 diabetes and glaucoma. Through a qualitative analysis of 22 selected scientific articles published between 2016 and 2025, the study compares various technical, clinical, and ethical approaches in the application of artificial intelligence models. The reviewed literature includes ensemble algorithms, convolutional neural networks (CNN), hybrid models such as CNN combined with long short-term memory (LSTM) and generative adversarial networks (GAN), explainability techniques such as SHapley Additive exPlanations (SHAP) and heatmaps, as well as strategies integrating electronic health records (EHR). The findings reveal strong performance in key metrics such as accuracy, sensitivity, and area under the curve (AUC), in some cases surpassing human experts. However, notable limitations remain, including low positive predictive value, lack of external validation, model opacity, and structural inequalities in access to diagnostic technology. Ethical and regulatory challenges—such as data privacy protections under the Health Insurance Portability and Accountability Act (HIPAA) and the regulation of medical technologies by the Food and Drug Administration (FDA)—are also discussed. The study concludes that while artificial intelligence offers a promising path for enhancing preventive medicine, its effective implementation requires overcoming technical, social, and legal barriers through inclusive, ethically grounded, and clinically validated strategies.

**Keywords:** Machine learning, early detection, type 2 diabetes, glaucoma, artificial intelligence, explainability, algorithmic bias, digital health.

## Tabla de Contenido

Introducción .....	13
Descripción de la Problemática .....	15
Planteamiento del Problema.....	17
Sistematización del Problema .....	19
Eje Técnico y Metodológico .....	19
Eje de Acceso, Implementación y Desigualdad .....	19
Eje Ético, Normativo y Regulatorio.....	20
Eje de Impacto y Mejora del Diagnóstico.....	20
Justificación .....	21
Objetivos .....	23
Objetivo General .....	23
Objetivos Específicos.....	23
Marco de Referencia .....	25
Estado del Arte.....	25
Marco Contextual.....	27
Diabetes Tipo 2 en Estados Unidos.....	27
Glaucoma. La Ceguera Silenciosa.....	28
Desigualdad Estructural y Brechas en el Acceso a Diagnóstico .....	28
Rol Emergente de la Inteligencia Artificial en Contextos Desiguales .....	29
Conclusión Contextual .....	29
Marco Teórico-Conceptual .....	30
Inteligencia Artificial en Salud.....	30

Aprendizaje Automático y Aprendizaje Profundo .....	30
Modelos Utilizados en el Diagnóstico Temprano .....	31
Caja Negra y Explicabilidad (XAI).....	32
Normativas y Principios Éticos .....	32
Aplicaciones a Glaucoma y Diabetes Tipo 2 .....	33
Aporte Original del Trabajo .....	34
Marco Normativo .....	35
Protección de Datos Personales: HIPAA .....	35
Regulación de Dispositivos Médicos: FDA .....	36
Ética y Responsabilidad en Decisiones Automatizadas .....	36
Normativas Internacionales Aplicables: GDPR .....	37
Conclusión Normativa.....	37
Metodología .....	38
Diseño de Investigación .....	38
Fuentes de Información.....	39
Estrategia de Búsqueda y Ecuación Utilizada.....	39
Criterios de Inclusión y Exclusión .....	39
Proceso de Selección.....	40
Elaboración de la Matriz de Análisis .....	40
Procedimiento de Análisis .....	41
Normas Éticas y de Citación.....	41
Integración de Imágenes y Tablas.....	42
Análisis y Discusión .....	43

Efectividad de los Modelos de IA en la Detección Temprana de Enfermedades Crónicas .....	43
Modelos para la Detección de Diabetes Tipo 2 .....	52
Modelos para la Detección y Progresión de Glaucoma .....	60
Detección de Glaucoma Mediante Aprendizaje Profundo.....	70
Limitaciones Técnicas de los Modelos de IA Médicos .....	72
Barreras Sociales en la Adopción de la IA en Salud.....	74
Implicaciones Éticas y Normativas.....	75
Discusión General y Retos Técnicos.....	77
Barreras sociales, Éticas y Normativas .....	77
Conclusiones .....	78
Recomendaciones para Futuros Estudios.....	80
Fortalecer la Validación Externa y Multicéntrica de los Modelos de IA .....	80
Incorporar Mecanismos de Explicabilidad en los Sistemas Predictivos .....	80
Desarrollar Protocolos de Integración Clínica con Supervisión Médica Obligatoria .....	80
Ampliar la Infraestructura Tecnológica en Zonas de Baja Cobertura Médica.....	80
Capacitar al Personal Médico en Inteligencia Artificial y Ética Digital .....	81
Referencias.....	82
Apéndices.....	89

## Lista de Tablas

<b>Tabla 1</b> <i>Características del Conjunto de Datos de Diabetes</i> .....	44
<b>Tabla 2</b> <i>Comparación de Métricas del Modelo de Ensamble para Diabetes</i> .....	46
<b>Tabla 3</b> <i>Desempeño del Modelo de Radiografías de Tórax para la Detección de Diabetes</i> .....	50
<b>Tabla 4</b> <i>Rendimiento de Modelos Supervisados con Técnicas de Balanceo en Diabetes Tipo 2</i>	53
<b>Tabla 5</b> <i>Comparación de Rendimiento entre Modelos Tradicionales y NLPNN en Múltiples Enfermedades</i> .....	57
<b>Tabla 6</b> <i>Tabla Visual de Métricas por Enfermedad y Algoritmo</i> .....	58
<b>Tabla 7</b> <i>AUROC por Modelo y Grupo Étnico</i> .....	61
<b>Tabla 8</b> <i>Métricas del Modelo de Deep Learning para Glaucoma</i> .....	71
<b>Tabla 9</b> <i>Comparación de Resultados y Limitaciones Técnicas entre Modelos</i> .....	73

## Lista de Figuras

<b>Figura 1</b>	<i>Diagrama de Flujo de Modelo de Ensemble para la Detección de Diabetes</i> .....	45
<b>Figura 2</b>	<i>Curva ROC</i> .....	47
<b>Figura 3</b>	<i>Ejemplos de Mapas de Calor</i> .....	49
<b>Figura 4</b>	<i>Distribución de los Puntajes de Riesgo Mediante Diagrama de Caja (boxplot)</i> . .....	51
<b>Figura 5</b>	<i>Comparación Gráfica de Rendimiento con y sin PCA y SMOTE</i> .....	54
<b>Figura 6</b>	<i>Arquitectura del Modelo CNN + LSTM</i> .....	55
<b>Figura 7</b>	<i>Curvas de Pérdida y Precisión</i> .....	56
<b>Figura 8</b>	<i>Diagrama de Flujo Modelos de Clasificación Binaria en Entorno Visual</i> . .....	59
<b>Figura 9</b>	<i>Curvas ROC y PR por Modelo y Etnia</i> .....	62
<b>Figura 10</b>	<i>Importancia de Características con SHAP Values</i> . .....	63
<b>Figura 11</b>	<i>Arquitectura CNN + LSTM + GAN</i> .....	64
<b>Figura 12</b>	<i>UMAP y Matriz de Confusión por Progresión de Glaucoma</i> .....	65
<b>Figura 13</b>	<i>Feature Importance por Gradient Boosting</i> .....	67
<b>Figura 14</b>	<i>Distribución Clínica en el Dataset</i> .....	68
<b>Figura 15</b>	<i>Clasificación de Modelos de DL por Categoría Funcional</i> .....	69
<b>Figura 16</b>	<i>Arquitectura General de Red Neuronal Profunda para Clasificación de Glaucoma</i> . ..	71

**Lista de Apéndices**

**Apéndice A**..... 89

**Apéndice B**..... 92

## Introducción

En Estados Unidos, enfermedades crónicas como la diabetes tipo 2 y el glaucoma representan una carga significativa para el sistema de salud. La diabetes afecta a millones de personas y, en muchos casos, se diagnostica en etapas avanzadas, cuando ya se han desarrollado complicaciones como nefropatías, enfermedades cardiovasculares o neuropatías. De forma similar, el glaucoma es una de las principales causas de ceguera irreversible, y su detección temprana sigue siendo un desafío debido a la falta de síntomas en sus fases iniciales.

En este contexto, el aprendizaje automático (machine learning) y el aprendizaje profundo (deep learning) han surgido como herramientas innovadoras para fortalecer la medicina preventiva. Estas tecnologías permiten analizar grandes volúmenes de datos clínicos e imágenes médicas con el fin de identificar patrones tempranos que, muchas veces, pasan desapercibidos en evaluaciones convencionales. Investigaciones como las de Akula et al. (2019), Pyrros et al. (2023) y Phene et al. (2019) han demostrado resultados prometedores: desde modelos de ensamble que alcanzan altos niveles de precisión en la detección de diabetes tipo 2, hasta redes neuronales aplicadas a radiografías y fotografías de fondo de ojo para diagnosticar signos iniciales de glaucoma.

A su vez, la literatura científica ha ampliado el panorama con estudios recientes que refuerzan la utilidad del aprendizaje automático en este campo. Autores como Kopitar et al. (2020), Bennett (2018), Hennebelle et al. (2023) y Kasula (2022) han explorado modelos de clasificación, arquitecturas híbridas e incluso plataformas integradas en la nube para predecir diabetes. En el área oftalmológica, trabajos de Fang et al. (2023), Prasad y Menon (2025) y Peck et al. (2024) aportan evidencia sobre cómo los algoritmos de visión artificial lograron niveles comparables a los de expertos humanos. La integración con registros electrónicos de salud

(EHR), como plantean Anderson et al. (2015) y Yin et al. (2024), también ha mostrado mejoras importantes en la precisión diagnóstica y la eficiencia del tamizaje clínico.

No obstante, el aprovechamiento de estas herramientas aún enfrenta barreras técnicas, sociales y éticas. Entre ellas destacan la desigualdad en el acceso a tecnologías avanzadas, la falta de capacitación en el personal de salud, la opacidad de algunos modelos ("caja negra") y las exigencias normativas sobre privacidad de datos, como lo indican Gulshan et al. (2016) y Liao y Yang (2020). Estas limitaciones deben ser abordadas si se desea una implementación efectiva, segura y ética de la inteligencia artificial en entornos clínicos reales.

Por tanto, esta monografía se propone analizar el potencial del aprendizaje automático para la detección temprana de diabetes tipo 2 y glaucoma en Estados Unidos. A partir del estudio de más de 40 fuentes científicas, se evaluarán sus beneficios técnicos, limitaciones, y las condiciones sociales y normativas que inciden en su aplicación. El objetivo es contribuir a una comprensión crítica de estas tecnologías, y reflexionar sobre cómo pueden ser utilizadas para fortalecer la prevención y mejorar el acceso al diagnóstico oportuno en el sistema de salud.

## Descripción de la Problemática

La diabetes tipo 2 y el glaucoma continúan siendo dos de las principales enfermedades crónicas que afectan a la población de Estados Unidos. Ambas patologías se caracterizan por un inicio silencioso y progresivo, lo que dificulta su detección en fases tempranas y aumenta considerablemente el riesgo de complicaciones graves. En el caso de la diabetes tipo 2, el diagnóstico tardío puede derivar en insuficiencia renal, enfermedades cardiovasculares, neuropatías y otras consecuencias que deterioran significativamente la calidad de vida del paciente (American Diabetes Association, 2021). Por su parte, el glaucoma constituye una de las principales causas de ceguera irreversible, debido a que sus síntomas se manifiestan generalmente cuando el daño al nervio óptico ya es avanzado (Phene et al., 2019).

Pese a los avances en el ámbito de la inteligencia artificial y el desarrollo de modelos predictivos cada vez más precisos, el diagnóstico temprano sigue siendo una tarea pendiente en la práctica clínica. Investigaciones como las de Akula et al. (2019), Kopitar et al. (2020) y Pyrros et al. (2023) han demostrado que el aprendizaje automático puede detectar señales tempranas de diabetes tipo 2 a partir de datos clínicos, radiografías torácicas o variables biométricas, alcanzando métricas de rendimiento destacables. Del mismo modo, estudios como los de Phene et al. (2019), Prasad y Menon (2025) y Fang et al. (2023) evidencian que modelos de deep learning aplicados a imágenes de fondo de ojo permiten identificar con alta precisión casos de glaucoma referible. Sin embargo, la incorporación de estas tecnologías en los sistemas de salud estadounidenses ha sido limitada y desigual.

Uno de los principales problemas identificados es la brecha entre la producción científica y su implementación efectiva en entornos clínicos reales. Aunque existen modelos validados en investigaciones académicas y desarrollos tecnológicos que han probado su eficacia, su adopción

generalizada enfrenta múltiples barreras. Entre ellas se encuentran la resistencia institucional, la falta de validación externa en poblaciones diversas, la escasez de formación del personal médico en el uso de herramientas basadas en IA, y la desigualdad en infraestructura tecnológica, especialmente en regiones rurales o comunidades de bajos recursos (Anderson et al., 2015; Hennebelle et al., 2023; Peck et al., 2024).

Esta situación se ve agravada por factores sociales y éticos. La falta de confianza por parte de profesionales y pacientes hacia modelos automatizados, junto con el temor a que la inteligencia artificial sustituya el juicio clínico humano, ha frenado su implementación práctica (Liao & Yang, 2020). Asimismo, la protección de datos sensibles y las normativas como HIPAA y GDPR imponen desafíos importantes para el uso seguro de datos clínicos en entornos automatizados (U.S. Department of Health and Human Services, 2019; Yin et al., 2024). Además, el sesgo algorítmico en los modelos entrenados con datos no representativos puede profundizar las desigualdades existentes en salud, afectando especialmente a minorías raciales, poblaciones no aseguradas y personas mayores.

En síntesis, el problema no radica únicamente en la existencia o calidad de los modelos de aprendizaje automático, sino en las condiciones estructurales, sociales y normativas que impiden su adopción amplia, equitativa y efectiva en el sistema de salud estadounidense. A pesar del potencial transformador de estas tecnologías, millones de personas siguen sin recibir un diagnóstico temprano de enfermedades crónicas prevenibles. Esta realidad evidencia la necesidad urgente de replantear las estrategias de integración de la IA en la medicina, orientándolas hacia una implementación ética, transparente, y sobre todo, accesible para toda la población.

## Planteamiento del Problema

La diabetes tipo 2 y el glaucoma continúan representando desafíos urgentes en salud pública en Estados Unidos. Ambas enfermedades comparten una característica crítica: su progresión silenciosa, lo que hace que un gran número de personas reciba un diagnóstico cuando las complicaciones ya son severas (American Diabetes Association, 2021; Phene et al., 2019). A pesar del avance en las tecnologías de inteligencia artificial y del desarrollo de modelos de aprendizaje automático altamente precisos (Akula et al., 2019; Kopitar et al., 2020; Pyrros et al., 2023), la integración de estas herramientas en la atención médica cotidiana aún no es una realidad extendida.

Estudios recientes han evidenciado que modelos como los de deep learning aplicados a radiografías o imágenes de fondo de ojo permiten detectar con alta precisión signos tempranos de estas patologías (Phene et al., 2019; Fang et al., 2023). Sin embargo, su adopción enfrenta múltiples limitaciones: falta de capacitación del personal médico, escasez de infraestructura tecnológica, marcos regulatorios poco definidos, y una clara resistencia institucional frente a los cambios tecnológicos (Liao & Yang, 2020; Yin et al., 2024). Estas barreras han impedido que los beneficios del aprendizaje automático se traduzcan en mejoras tangibles para la población, especialmente en contextos donde más se necesita: comunidades rurales, minorías étnicas y personas no aseguradas (Hennebelle et al., 2023; Kasula, 2022).

La desconexión entre el conocimiento académico y su aplicación clínica plantea un problema estructural que requiere atención. No se trata solo de desarrollar modelos más precisos, sino de crear las condiciones adecuadas para su implementación ética, equitativa y efectiva. Así, surge el siguiente problema de investigación:

¿Cómo puede el aprendizaje automático contribuir de manera efectiva a mejorar la detección temprana de la diabetes tipo 2 y el glaucoma en el sistema de salud de Estados Unidos, considerando las evidencias disponibles y las barreras técnicas, sociales y regulatorias que dificultan su adopción generalizada?

## **Sistematización del Problema**

Para abordar este problema central, se formulan las siguientes preguntas secundarias, organizadas por ejes temáticos clave:

### **Eje Técnico y Metodológico**

1. ¿Qué tipos de técnicas de aprendizaje automático han demostrado mayor efectividad en la detección temprana de diabetes tipo 2 y glaucoma?
2. ¿Cuál es el papel de los datos clínicos estructurados, las imágenes médicas y los registros electrónicos de salud en el desarrollo de modelos predictivos?
3. ¿Cómo se comparan las métricas de rendimiento, como precisión, sensibilidad, especificidad y área bajo la curva (AUC), entre los diferentes enfoques de aprendizaje automático aplicados a estas enfermedades?

### **Eje de Acceso, Implementación y Desigualdad**

1. ¿Qué limitaciones sociales, económicas y tecnológicas afectan la adopción de herramientas de inteligencia artificial en centros de atención primaria, especialmente en comunidades rurales o desatendidas?
2. ¿De qué manera la falta de infraestructura, la resistencia institucional y la desinformación influyen en la baja implementación de modelos automatizados en el sistema de salud?
3. ¿Qué brechas existen entre el desarrollo de modelos académicos exitosos y su implementación efectiva en entornos clínicos reales?

### **Eje Ético, Normativo y Regulatorio**

1. ¿Qué desafíos éticos plantea el uso de aprendizaje automático en medicina diagnóstica, en relación con la privacidad, el consentimiento informado y la responsabilidad clínica?

2. ¿En qué medida los sesgos algorítmicos pueden afectar negativamente a subgrupos poblacionales en los diagnósticos automatizados, y qué estrategias podrían implementarse para mitigarlos?

3. ¿Qué marcos regulatorios existen actualmente en Estados Unidos para el uso de inteligencia artificial en salud, y cómo están evolucionando para incluir estas tecnologías en el ámbito médico?

### **Eje de Impacto y Mejora del Diagnóstico**

1. ¿Qué evidencias existen sobre el impacto real del uso de modelos de inteligencia artificial en la mejora del diagnóstico temprano y la reducción de complicaciones en pacientes con diabetes tipo 2 o glaucoma?

2. ¿Cómo pueden los modelos explicables mejorar la confianza clínica y facilitar su adopción en la práctica médica cotidiana?

## Justificación

La necesidad de mejorar la detección temprana de enfermedades crónicas como la diabetes tipo 2 y el glaucoma se ha vuelto una prioridad en el sistema de salud estadounidense. Ambas condiciones no solo afectan a millones de personas, sino que también generan altos costos sanitarios y sociales cuando no se diagnostican a tiempo (American Diabetes Association, 2021; Phene et al., 2019). Frente a este panorama, la inteligencia artificial —y en particular el aprendizaje automático— ofrece herramientas innovadoras que permiten identificar signos tempranos de estas patologías antes de que se presenten síntomas clínicos evidentes (Akula et al., 2019; Pyrros et al., 2023; Fang et al., 2023).

Numerosos estudios han demostrado que modelos como las redes neuronales profundas, los algoritmos de ensamble y los sistemas de clasificación explicables han alcanzado niveles de precisión, sensibilidad y especificidad comparables —y en ocasiones superiores— a los del juicio clínico tradicional, cuando se aplican al análisis de datos estructurados, radiografías torácicas e imágenes de fondo de ojo (Phene et al., 2019; Kopitar et al., 2020; Yin et al., 2024). Esta capacidad para detectar patrones invisibles a simple vista convierte a la inteligencia artificial en un recurso poderoso para la medicina preventiva.

Sin embargo, la existencia de estos modelos no garantiza su adopción. Aún persisten desafíos técnicos y sociales que limitan su aplicación en la práctica clínica real. La falta de formación del personal médico, las barreras económicas, la resistencia institucional al cambio tecnológico y la desconfianza hacia herramientas automatizadas son obstáculos documentados en diversas investigaciones (Liao & Yang, 2020; Kasula, 2022; Peck et al., 2024). Además, los marcos regulatorios como HIPAA y las exigencias éticas sobre el uso de datos sensibles

imponen restricciones importantes que deben ser gestionadas adecuadamente (U.S. Department of Health and Human Services, 2019; FDA, 2017).

Por estas razones, es esencial analizar no solo los avances técnicos logrados por la IA en el diagnóstico temprano, sino también los factores que condicionan su implementación real en el sistema de salud estadounidense. Esta monografía se justifica en la medida en que ofrece una revisión crítica y actualizada de los estudios más relevantes sobre el tema, proponiendo una mirada integral que combina dimensiones técnicas, sociales, éticas y normativas.

Asimismo, el estudio cobra relevancia al identificar cómo estas tecnologías pueden contribuir a reducir las desigualdades en el acceso al diagnóstico oportuno, especialmente en comunidades rurales, no aseguradas o históricamente marginadas (Hennebelle et al., 2023; Anderson et al., 2015). Analizar y visibilizar estas barreras es clave para avanzar hacia una adopción responsable, equitativa y efectiva de la inteligencia artificial en la salud pública.

## Objetivos

### Objetivo General

Analizar el potencial del aprendizaje automático para la detección temprana de la diabetes tipo 2 y el glaucoma en el sistema de salud de Estados Unidos, evaluando la evidencia científica reciente, los modelos utilizados, sus resultados técnicos y las condiciones sociales, normativas y estructurales que influyen en su implementación efectiva y equitativa.

### Objetivos Específicos

Describir las principales técnicas de aprendizaje automático empleadas en la detección temprana de la diabetes tipo 2 y el glaucoma, con énfasis en modelos de aprendizaje profundo, algoritmos de ensamble y enfoques explicables basados en visualización o importancia de atributos.

Evaluar la efectividad diagnóstica de los modelos de aprendizaje automático identificados en la literatura científica, considerando métricas como precisión, sensibilidad, especificidad y área bajo la curva (AUC), así como su aplicabilidad en entornos clínicos reales.

Analizar el uso de imágenes médicas —como radiografías de tórax y fotografías de fondo de ojo— en combinación con algoritmos de aprendizaje automático, en función de su utilidad para el diagnóstico oportuno de enfermedades crónicas.

Explorar las barreras sociales, económicas y tecnológicas que limitan la adopción de herramientas de inteligencia artificial en la atención primaria, con especial atención a las desigualdades estructurales en el acceso a tecnologías avanzadas.

Investigar las implicaciones éticas y regulatorias asociadas al uso de modelos automatizados en el diagnóstico médico, incluyendo aspectos como la privacidad de los datos, la responsabilidad clínica y el cumplimiento de normativas como HIPAA y FDA.

Examinar cómo el nivel de formación y conocimiento del personal de salud en inteligencia artificial influye en la integración efectiva de estas herramientas en la práctica clínica.

Analizar el impacto de la desinformación, la percepción pública y la falta de educación digital en la aceptación y confianza hacia los sistemas automatizados en salud.

Comparar el desempeño y aplicabilidad de los modelos de aprendizaje automático documentados en los estudios revisados frente a los enfoques médicos tradicionales, identificando sus ventajas, limitaciones y requerimientos de implementación.

Proponer estrategias educativas, institucionales y de política pública que favorezcan una adopción responsable y equitativa del aprendizaje automático en contextos médicos, especialmente en comunidades vulnerables o con infraestructura limitada.

Identificar las principales limitaciones técnicas, de interpretabilidad y de validación externa que enfrentan los modelos actuales, y sugerir posibles soluciones orientadas a mejorar su confiabilidad diagnóstica y aceptabilidad clínica.

## Marco de Referencia

### Estado del Arte

En los últimos años, la literatura científica ha documentado un crecimiento significativo en el uso de inteligencia artificial para la detección temprana de enfermedades crónicas. El aprendizaje automático (machine learning) y el aprendizaje profundo (deep learning) han sido aplicados exitosamente a registros médicos estructurados e imágenes médicas para identificar patologías como la diabetes tipo 2 y el glaucoma antes de que se manifiesten clínicamente.

Uno de los estudios más citados es el de Akula et al. (2019), que utilizó un modelo de ensamble supervisado para detectar diabetes tipo 2 a partir de datos estructurados en historias clínicas electrónicas. El modelo combinó algoritmos como SVM, random forest y redes neuronales, alcanzando una precisión del 85%. A su vez, Pyrros et al. (2023) emplearon redes neuronales convolucionales entrenadas con más de 270.000 radiografías de tórax, logrando una AUC de 0.84 en la predicción de riesgo de diabetes. Ambos trabajos destacaron por su capacidad de aplicar modelos avanzados a bases de datos reales con resultados clínicamente relevantes.

En el campo oftalmológico, Phene et al. (2019) desarrollaron un modelo de deep learning para identificar casos de glaucoma referible usando más de 86.000 imágenes de fondo de ojo. La red neuronal profunda alcanzó una AUC de hasta 0.945 y superó a evaluadores humanos en sensibilidad diagnóstica. Prasad y Menon (2025) complementan este enfoque incorporando mapas de calor y técnicas de explicabilidad como SHAP para visualizar la lógica del algoritmo, abordando así una de las principales críticas de estos sistemas: su opacidad como modelos de “caja negra”.

Por su parte, Fang et al. (2023) introdujeron redes generativas adversariales (GANs) para mejorar la calidad de los conjuntos de datos utilizados en la detección de glaucoma, destacando

el papel de la generación sintética de imágenes médicas como alternativa en entornos con escasos datos etiquetados. Zhang et al. (2025), mediante una revisión sistemática, identificaron que muchos modelos publicados carecen de validación externa multicéntrica, lo que compromete su generalización en entornos clínicos reales. Su estudio destaca la necesidad urgente de evaluar los modelos en poblaciones diversas para evitar sesgos clínicos.

Otros estudios se han enfocado en optimizar el rendimiento computacional y la eficiencia del despliegue de modelos. Hennebelle et al. (2023) propusieron una arquitectura distribuida mediante edge computing para facilitar el diagnóstico en regiones remotas sin conectividad estable. Kopitar et al. (2020), por su parte, aplicaron random forest y XGBoost sobre datos clínicos eslovenos, obteniendo valores de precisión similares a los de Akula, pero con mejor interpretabilidad gracias a la estructura de los árboles de decisión y la simplicidad de análisis de variables relevantes.

En cuanto a los aspectos sociales y normativos, Kasula (2022) y Anderson et al. (2015) coinciden en que existe una marcada desigualdad en la adopción de estas tecnologías. Las comunidades rurales y sin seguro médico tienden a quedar excluidas del acceso a sistemas diagnósticos basados en IA, a pesar de que estos podrían representar soluciones de bajo costo y alto impacto poblacional. Además, se señala la falta de estrategias institucionales para capacitar al personal clínico en el uso seguro y efectivo de estas herramientas.

Finalmente, Yin et al. (2024) demostraron que integrar aprendizaje profundo con registros electrónicos de salud (EHR) mejora significativamente la sensibilidad diagnóstica y reduce los falsos negativos en poblaciones diabéticas. No obstante, la desconfianza institucional, la resistencia del personal clínico y los vacíos regulatorios siguen siendo factores determinantes en la lenta adopción de estas tecnologías en el ámbito hospitalario.

En conjunto, estos trabajos reflejan que, aunque existe abundante evidencia sobre el potencial técnico de la inteligencia artificial para la detección temprana de enfermedades crónicas, aún persisten retos estructurales relacionados con la interpretabilidad, la validez externa, la equidad en el acceso y la aceptación clínica. Superar estas barreras requerirá no solo mejoras técnicas, sino también estrategias de implementación centradas en el contexto social, legal y médico.

### **Marco Contextual**

Las enfermedades crónicas representan una de las principales cargas para el sistema de salud en Estados Unidos, tanto por su prevalencia como por sus consecuencias clínicas, económicas y sociales. Dentro de este grupo, la diabetes tipo 2 y el glaucoma destacan no solo por su incidencia creciente, sino también por el elevado impacto que generan cuando no son detectadas a tiempo. En este contexto, la necesidad de estrategias diagnósticas más eficaces, tempranas y accesibles se ha vuelto un objetivo prioritario para las políticas públicas y la innovación tecnológica en salud.

### ***Diabetes Tipo 2 en Estados Unidos***

Según datos de los Centers for Disease Control and Prevention (CDC, 2022), más de 37 millones de personas en EE. UU. viven con diabetes, de las cuales aproximadamente el 90–95 % tienen diabetes tipo 2. Además, se estima que uno de cada cinco adultos con diabetes no está diagnosticado, lo que representa una gran proporción de personas que viven con esta condición sin saberlo, y por ende, sin tratamiento adecuado.

Las complicaciones asociadas a un diagnóstico tardío incluyen enfermedades cardiovasculares, insuficiencia renal, amputaciones y neuropatías, lo que eleva tanto los costos de tratamiento como la carga de discapacidad. Estudios han demostrado que la detección

oportuna y la intervención temprana pueden reducir hasta en un 58 % la progresión hacia complicaciones severas (American Diabetes Association, 2021).

### ***Glaucoma. La Ceguera Silenciosa***

En paralelo, el glaucoma afecta a más de 3 millones de personas en Estados Unidos, y se proyecta que ese número aumente significativamente con el envejecimiento de la población (National Eye Institute [NEI], 2021). A diferencia de otras patologías oftalmológicas, el glaucoma suele avanzar sin síntomas perceptibles en sus primeras etapas, y una vez que el daño al nervio óptico ocurre, es irreversible. Por esta razón, se le conoce como la “ceguera silenciosa”.

A pesar de la disponibilidad de pruebas oftalmológicas no invasivas como la tonometría o el examen del fondo de ojo, muchas personas no acceden a controles regulares. Las barreras más frecuentes son el costo de atención especializada, la falta de seguro médico y el desconocimiento sobre la enfermedad (Zhang et al., 2025).

### ***Desigualdad Estructural y Brechas en el Acceso a Diagnóstico***

El sistema de salud estadounidense presenta fuertes desigualdades estructurales, las cuales afectan de manera desproporcionada a poblaciones vulnerables: personas sin seguro médico, comunidades afroamericanas y latinas, adultos mayores y residentes en zonas rurales. Estas poblaciones no solo tienen una mayor prevalencia de enfermedades crónicas, sino que también enfrentan mayores barreras para acceder a servicios preventivos y diagnóstico temprano (Anderson et al., 2015).

Aunque existen programas federales como Medicaid, el National Diabetes Prevention Program (NDPP) y campañas de concienciación lideradas por el National Eye Institute, la cobertura es limitada, y muchas intervenciones no alcanzan a quienes más las necesitan. Esta situación ha motivado a investigadores y desarrolladores tecnológicos a buscar soluciones

alternativas, como el uso de herramientas de inteligencia artificial que puedan ser implementadas incluso en entornos con recursos limitados.

### ***Rol Emergente de la Inteligencia Artificial en Contextos Desiguales***

La inteligencia artificial se presenta como una **herramienta disruptiva** con el potencial de mejorar la equidad en salud. Al automatizar procesos diagnósticos y reducir la dependencia de especialistas humanos, los modelos de IA permiten realizar tamizajes masivos con bajo costo por paciente (Hennebelle et al., 2023). Esto es especialmente relevante en áreas rurales donde el acceso a oftalmólogos o endocrinólogos es escaso.

Además, estudios recientes han demostrado que modelos entrenados con imágenes médicas o registros electrónicos pueden identificar signos tempranos de enfermedad incluso en pacientes sin síntomas clínicos evidentes (Pyrros et al., 2023; Yin et al., 2024). Sin embargo, su adopción efectiva requiere más que capacidad técnica: se necesita también infraestructura digital, voluntad institucional y marcos normativos claros que regulen el uso ético y seguro de estas tecnologías.

### ***Conclusión Contextual***

En síntesis, el contexto estadounidense se caracteriza por una alta carga de enfermedades crónicas, un acceso desigual al diagnóstico temprano y una creciente necesidad de soluciones escalables. La IA ofrece una alternativa prometedora, pero su éxito dependerá de cómo se articule con las realidades sociales, económicas y políticas del sistema de salud. Por ello, analizar sus aplicaciones desde una perspectiva crítica, inclusiva y contextualizada no es solo deseable, sino indispensable.

## **Marco Teórico-Conceptual**

### ***Inteligencia Artificial en Salud***

La inteligencia artificial (IA) se refiere a la capacidad de los sistemas computacionales para realizar tareas que normalmente requieren inteligencia humana, como el razonamiento, el aprendizaje, la toma de decisiones y el reconocimiento de patrones. En el contexto de la salud, la IA ha adquirido un papel creciente en áreas como el diagnóstico, la predicción de enfermedades, la gestión hospitalaria y la medicina personalizada. Su aplicación en el diagnóstico temprano de enfermedades crónicas como la diabetes tipo 2 y el glaucoma representa un avance significativo, ya que permite detectar signos clínicos sutiles, automatizar procesos de tamizaje y reducir errores diagnósticos humanos.

La evolución de la IA en medicina ha estado marcada por el desarrollo de sistemas expertos, la minería de datos clínicos y, más recientemente, por algoritmos que imitan estructuras neuronales humanas para aprender de forma autónoma. Este tipo de tecnologías, además de optimizar recursos en atención primaria, permiten una personalización de la medicina en función del riesgo individual, generando nuevas oportunidades para el tamizaje masivo, la estratificación de pacientes y la reducción de la carga sobre los sistemas de salud pública.

### ***Aprendizaje Automático y Aprendizaje Profundo***

Dentro del campo de la IA, el aprendizaje automático (machine learning, ML) consiste en el desarrollo de algoritmos capaces de aprender patrones a partir de datos sin ser explícitamente programados. Entre sus técnicas se incluyen métodos supervisados (como máquinas de vectores de soporte [Support Vector Machines, SVM], vecinos más cercanos [k-Nearest Neighbors, k-NN], regresión logística y árboles de decisión), no supervisados (como agrupamiento jerárquico o jerárquico [clustering] y K-means, que permiten segmentar a los pacientes en grupos según

características compartidas), y por refuerzo. Una rama especializada es el aprendizaje profundo (deep learning, DL), basado en redes neuronales artificiales con múltiples capas que permiten modelar relaciones complejas y no lineales. Este enfoque ha sido clave en el procesamiento de imágenes médicas, datos clínicos multivariados y series temporales.

El aprendizaje profundo, especialmente a través de arquitecturas convolucionales (convolutional neural networks, CNN) y recurrentes (long short-term memory, LSTM), ha revolucionado la capacidad de los modelos para extraer patrones visuales complejos en exámenes oftálmicos como fondos de ojo y tomografías de coherencia óptica (optical coherence tomography, OCT). Estas arquitecturas permiten un nivel de precisión difícilmente alcanzable por métodos clásicos, y han sido adaptadas incluso a imágenes sintéticas generadas con redes generativas adversarias (generative adversarial networks, GAN).

### ***Modelos Utilizados en el Diagnóstico Temprano***

Los algoritmos más frecuentemente aplicados en el diagnóstico automatizado incluyen:

Árboles de decisión (decision trees, DT) y bosques aleatorios (Random Forest): interpretables y efectivos en datos estructurados.

Máquinas de vectores de soporte (Support Vector Machines, SVM): útiles en espacios de alta dimensionalidad.

Vecinos más cercanos (k-Nearest Neighbors, k-NN): efectivos pero sensibles a datos ruidosos.

Redes neuronales profundas (deep neural networks, DNN) y convolucionales (CNN): altamente eficaces para procesamiento de imágenes médicas como fondo de ojo u OCT.

Modelos híbridos (como CNN+LSTM o CNN+GAN): combinan capacidades de extracción espacial y análisis temporal.

Además, el uso de modelos de ensamble ha cobrado relevancia, especialmente aquellos que combinan múltiples clasificadores (Random Forest, Gradient Boosting, XGBoost) mediante estrategias de votación o promedio ponderado, lo que permite mitigar errores individuales y aumentar la robustez general del sistema.

### ***Caja Negra y Explicabilidad (XAI)***

Un desafío técnico importante es que muchos modelos, especialmente los basados en aprendizaje profundo, funcionan como "cajas negras", es decir, producen predicciones sin que el usuario comprenda fácilmente su lógica interna. Esto genera barreras para su implementación clínica, ya que los médicos deben confiar en los resultados sin poder explicarlos adecuadamente a los pacientes.

Para superar este obstáculo, han surgido enfoques de inteligencia artificial explicable (explainable artificial intelligence, XAI), que buscan aumentar la transparencia de los modelos. Herramientas como SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-agnostic Explanations) y Grad-CAM (Gradient-weighted Class Activation Mapping) permiten visualizar la importancia relativa de cada variable o región de imagen en la decisión del modelo. Esto no solo mejora la confianza médica, sino que también es clave para la auditoría ética, la detección de sesgos y el cumplimiento normativo.

Además, la explicabilidad es especialmente crítica cuando se implementan estos modelos en sistemas de salud pública con poblaciones diversas, ya que permite identificar si el algoritmo presenta desigualdades diagnósticas según variables demográficas, económicas o raciales.

### ***Normativas y Principios Éticos***

El uso clínico de IA está sujeto a marcos regulatorios y principios bioéticos. En Estados Unidos, la Ley de Portabilidad y Responsabilidad del Seguro Médico (Health Insurance

Portability and Accountability Act, HIPAA) establece criterios de protección de datos sensibles en salud, incluyendo su anonimización, almacenamiento y acceso autorizado. En Europa, el Reglamento General de Protección de Datos (General Data Protection Regulation, GDPR) también influye indirectamente en proyectos internacionales. Además, organismos como la Administración de Alimentos y Medicamentos (Food and Drug Administration, FDA) han comenzado a establecer guías específicas para la validación y uso de software médico basado en IA.

Desde una perspectiva ética, el desarrollo de estos modelos debe respetar principios de justicia, transparencia, no maleficencia y responsabilidad clínica. La implementación de modelos sin validación suficiente puede amplificar desigualdades estructurales si no se evalúa su desempeño en poblaciones diversas. Por ello, es fundamental aplicar validación externa, análisis estratificado por grupo demográfico y control de sesgos algorítmicos desde el diseño del dataset.

Asimismo, la falta de formación médica en interpretación de modelos y la resistencia institucional al cambio representan obstáculos reales. Las recomendaciones internacionales sugieren integrar la alfabetización digital y la ética de datos en la formación de profesionales de salud.

### ***Aplicaciones a Glaucoma y Diabetes Tipo 2***

En el caso de la diabetes tipo 2, los modelos de IA se han aplicado exitosamente sobre bases clínicas estructuradas, imágenes de fondo de ojo, datos de sensores, e incluso radiografías de tórax para detección secundaria. Modelos de ensamble como Random Forest y Gradient Boosting han mostrado alta sensibilidad en combinación con técnicas de preprocesamiento como el análisis de componentes principales (principal component analysis, PCA) o la técnica de sobremuestreo sintético para clases minoritarias (Synthetic Minority Oversampling Technique,

SMOTE). Además, se ha evidenciado que el uso de datasets desbalanceados afecta la detección temprana, lo que ha motivado técnicas de muestreo sintético y normalización de datos para mejorar el rendimiento.

Para el glaucoma, el uso de imágenes OCT, fotografías de fondo de ojo y registros clínicos electrónicos (electronic health records, EHR) ha permitido desarrollar modelos CNN, LSTM y XGBoost capaces de detectar progresión temprana de la enfermedad, incluso en ausencia de síntomas. Estudios recientes han demostrado que, al desagregar el rendimiento por grupo étnico, es posible identificar sesgos ocultos que no serían visibles con métricas agregadas. Además, el uso de SHAP y Grad-CAM ha sido útil para interpretar visualmente qué estructuras anatómicas influyen más en la decisión del modelo, fortaleciendo su aceptabilidad clínica.

Ambos casos ilustran el potencial transformador de la IA en salud, siempre que su aplicación esté acompañada de rigurosidad científica, supervisión médica, validación continua y un marco regulatorio robusto.

### ***Aporte Original del Trabajo***

Este trabajo ofrece un análisis cualitativo integrador que articula componentes técnicos, clínicos, éticos y regulatorios en torno a la aplicación de modelos de inteligencia artificial en el diagnóstico temprano de enfermedades crónicas, específicamente diabetes tipo 2 y glaucoma. A diferencia de otras revisiones centradas en un solo modelo o enfermedad, esta monografía abarca múltiples enfoques de aprendizaje automático y aprendizaje profundo (incluyendo Random Forest, XGBoost, CNN, LSTM y GAN), así como herramientas de interpretabilidad como SHAP, LIME y Grad-CAM. Además, se abordan aspectos como la explicabilidad, el sesgo algorítmico, la validación externa y las normativas de protección de datos (HIPAA, GDPR), con especial énfasis en la aplicabilidad en poblaciones diversas. Esta visión holística constituye un

aporte original al vincular desarrollos tecnológicos avanzados con los desafíos estructurales y éticos de la práctica clínica actual, proponiendo recomendaciones concretas para su implementación responsable en el sistema de salud.

### **Marco Normativo**

El uso de inteligencia artificial (IA) en el diagnóstico médico plantea importantes desafíos legales, éticos y regulatorios. A medida que estas tecnologías avanzan en precisión y penetración clínica, se hace cada vez más urgente establecer marcos normativos que garanticen su desarrollo responsable, su aplicación ética y su compatibilidad con los derechos fundamentales de los pacientes. En el contexto estadounidense, la regulación de la IA médica se enmarca principalmente en tres ejes: la protección de datos personales de salud, la aprobación de tecnologías médicas por agencias federales, y la responsabilidad ética frente a decisiones automatizadas.

#### ***Protección de Datos Personales: HIPAA***

La Health Insurance Portability and Accountability Act (HIPAA), promulgada en 1996, es la principal legislación federal que protege la privacidad de la información médica de los pacientes en Estados Unidos. Esta ley establece que toda entidad cubierta (hospitales, aseguradoras, proveedores de salud) debe implementar medidas de seguridad para el manejo, almacenamiento, transferencia y anonimización de los datos clínicos (U.S. Department of Health and Human Services, 2019).

Cuando se entrenan modelos de IA con datos de salud, estos deben cumplir con los estándares HIPAA, garantizando que los datos no permitan identificar directamente a los pacientes. Además, se exige obtener consentimiento informado para su uso en investigación, entrenamiento de algoritmos o pruebas clínicas. En este sentido, cualquier herramienta de

aprendizaje automático desarrollada para la detección de diabetes o glaucoma debe incorporar principios de seguridad de datos desde su diseño (privacy by design).

### ***Regulación de Dispositivos Médicos: FDA***

La Food and Drug Administration (FDA) es la agencia federal encargada de regular y aprobar dispositivos médicos, incluidos los que utilizan inteligencia artificial o algoritmos de aprendizaje automático. En 2017, la FDA comenzó a establecer criterios específicos para el Software as a Medical Device (SaMD), abriendo la posibilidad de certificar modelos de IA siempre que cumplan con criterios de eficacia, seguridad y explicabilidad (FDA, 2017).

En este marco, un modelo de IA para diagnóstico clínico se considera un dispositivo médico si su salida se utiliza para tomar decisiones que impactan directamente en la salud del paciente. Esto implica que debe someterse a validación clínica, revisión por pares, y eventualmente, aprobación formal por parte de la FDA. A partir de 2021, la agencia ha trabajado en un marco regulatorio específico para algoritmos que aprenden de forma continua (adaptive AI), con el objetivo de equilibrar la innovación con la seguridad (FDA, 2021).

### ***Ética y Responsabilidad en Decisiones Automatizadas***

El uso de IA en medicina conlleva riesgos asociados a la opacidad de los modelos, los sesgos algorítmicos, y la ausencia de responsabilidad clara ante errores diagnósticos. Muchas herramientas actuales —en especial las basadas en redes neuronales profundas— funcionan como “cajas negras”, es decir, producen resultados sin que los médicos puedan comprender completamente la lógica detrás de la decisión (Zhang et al., 2025).

Esto plantea dilemas éticos cruciales: ¿puede un profesional de la salud basar un diagnóstico en una predicción que no comprende? ¿Quién es responsable si el modelo comete un error y causa un daño? Las propuestas actuales apuntan a que la IA debe ser un sistema de apoyo

y no un reemplazo del juicio clínico humano, y que toda decisión médica automatizada debe poder ser auditada, explicada y justificada (Kasula, 2022).

Además, se exige que los datos utilizados para entrenar estos modelos sean representativos, para evitar sesgos que puedan perjudicar a poblaciones vulnerables. Por ejemplo, un modelo entrenado exclusivamente con datos de pacientes caucásicos puede fallar al diagnosticar correctamente a pacientes afroamericanos, generando discriminación algorítmica.

### ***Normativas Internacionales Aplicables: GDPR***

Aunque no es de aplicación directa en Estados Unidos, el Reglamento General de Protección de Datos (GDPR) de la Unión Europea ha influido en la formulación de principios éticos y legales sobre el uso de datos personales en IA. El GDPR establece el derecho a la explicación automatizada (art. 22), la minimización de datos y el consentimiento explícito, principios que han sido adoptados en parte por iniciativas académicas y hospitalarias en EE. UU. con fines de investigación (European Commission, 2016).

### ***Conclusión Normativa***

En resumen, el marco normativo que regula la aplicación de inteligencia artificial en la salud en Estados Unidos combina aspectos legales, regulatorios y éticos. Mientras que HIPAA asegura la confidencialidad de los datos clínicos, la FDA establece lineamientos sobre la seguridad y eficacia de los modelos, y los marcos éticos proponen lineamientos sobre la transparencia, equidad y responsabilidad en el uso de algoritmos. No obstante, aún persisten desafíos normativos significativos en torno a la gobernanza de modelos adaptativos, la auditabilidad de redes neuronales complejas y la protección de los derechos de los pacientes frente a sistemas automatizados.

## Metodología

Para el desarrollo de la presente monografía, se empleó un enfoque cualitativo con diseño de revisión integrativa y de alcance (*scoping review*), que permitió recopilar, evaluar e interpretar críticamente estudios científicos sobre la aplicación de inteligencia artificial (IA) en el diagnóstico temprano de enfermedades crónicas, específicamente diabetes tipo 2 y glaucoma. Esta modalidad resulta pertinente para mapear y sintetizar hallazgos técnicos, clínicos y éticos provenientes de distintas disciplinas, especialmente considerando el carácter emergente, multidisciplinario y la heterogeneidad metodológica de los estudios analizados.

Se seleccionaron 22 artículos científicos publicados entre 2015 y 2024 que cumplieran con criterios estrictos de inclusión. Se priorizaron estudios que: (1) aplicaran modelos de aprendizaje automático o aprendizaje profundo en el diagnóstico temprano de diabetes tipo 2 o glaucoma, (2) reportaran métricas cuantitativas como AUC, F1-score, precisión o sensibilidad, (3) utilizaran bases de datos reales (estructuradas o clínicas), y (4) fueran accesibles a texto completo con verificación de autoría y publicación indexada (Scopus, PubMed, IEEE, Springer, entre otras).

### Diseño de Investigación

El estudio adoptó un enfoque exploratorio y descriptivo. La fase exploratoria permitió identificar vacíos y áreas poco desarrolladas en la literatura sobre inteligencia artificial aplicada a la medicina, mientras que el enfoque descriptivo facilitó la caracterización detallada de los modelos, algoritmos, conjuntos de datos (*datasets*), métricas e implicaciones metodológicas de cada estudio. A su vez, el enfoque cualitativo, de naturaleza interpretativa, priorizó el análisis crítico tanto del contenido textual como visual (tablas, gráficas y arquitecturas de red) presente en los artículos seleccionados (Miotto et al., 2018; Russell & Norvig, 2016).

## **Fuentes de Información**

Se recurrió a literatura científica secundaria, incluyendo artículos de investigación originales, revisiones sistemáticas, reportes técnicos, normativas oficiales, guías clínicas y capítulos de libros. Las principales bases de datos utilizadas fueron PubMed, Scopus, SpringerLink, Elsevier, IEEE Xplore, ScienceDirect, Google Scholar y repositorios institucionales universitarios (Ristevski & Chen, 2018; Luo et al., 2016).

## **Estrategia de Búsqueda y Ecuación Utilizada**

Se utilizaron operadores booleanos y términos controlados adaptados a cada base de datos consultada. La ecuación principal de búsqueda fue:

("machine learning" OR "deep learning" OR "artificial intelligence") AND ("early detection" OR "diagnosis") AND ("type 2 diabetes" OR "glaucoma") AND ("models" OR "algorithms") AND ("metrics" OR "accuracy" OR "AUC")

Adicionalmente, se incorporaron términos específicos como “neural networks”, “electronic health records”, “LSTM”, “CNN”, “ensemble learning”, “GAN”, “interpretability”, “ethics”, “regulations” y “bias”, con el fin de ampliar la sensibilidad temática de la búsqueda y capturar artículos relevantes de mayor especificidad técnica (Zhang et al., 2025).

## **Criterios de Inclusión y Exclusión**

Se incluyeron artículos que cumplieran con los siguientes requisitos:

- Publicación entre 2016 y 2025.
- Redacción en inglés o español.
- Reporte de métricas cuantitativas (accuracy, precision, recall, F1-score, AUC).
- Aplicación directa a la detección temprana de diabetes tipo 2 o glaucoma.

- Inclusión de figuras, tablas comparativas o arquitecturas de modelos (Acharya et al., 2017; Kermany et al., 2018).
- Se excluyeron:
- Estudios centrados en otras enfermedades no pertinentes al objetivo de la revisión.
- Artículos sin validación experimental o con resultados puramente teóricos.
- Publicaciones duplicadas, breves reseñas o trabajos sin revisión por pares.

### **Proceso de Selección**

Se recuperaron 114 artículos tras aplicar la ecuación de búsqueda. Luego de una revisión inicial por títulos y resúmenes, se filtraron 61 estudios. De ellos, 33 cumplieron con los criterios de calidad metodológica y profundidad técnica. Finalmente, se seleccionaron 22 artículos con alta relevancia clínica, diversidad en los enfoques metodológicos y riqueza visual (figuras, tablas, curvas, arquitecturas), que constituyeron la base principal del análisis. Durante la discusión se priorizó el contraste entre estudios experimentales y revisiones sistemáticas representativas (Gulshan et al., 2016; Rajalakshmi et al., 2018).

### **Elaboración de la Matriz de Análisis**

Se diseñó una matriz comparativa para organizar y sistematizar los hallazgos principales. Esta herramienta incluyó las siguientes variables:

- Nombre del artículo.
- Autores y año de publicación.
- Enfermedad analizada.
- Tipo de IA aplicada (aprendizaje automático o profundo).
- Algoritmo específico (por ejemplo, XGBoost, CNN, LSTM, GAN, FCN).

- Fuente de datos (registros clínicos, imágenes, bases estructuradas).
- Métricas de desempeño (accuracy, AUC, F1-score).
- Visualizaciones relevantes (tablas, figuras, curvas ROC, valores SHAP).
- Limitaciones señaladas por los autores.
- Consideraciones éticas o regulatorias.
- Comparación con estudios previos.

[La matriz fue diligenciada manualmente para cada uno de los 22 estudios seleccionados, lo que permitió un análisis **transversal y comparativo** de los enfoques utilizados, la variedad de algoritmos aplicados y las métricas de evaluación reportadas. Su diseño se inspiró en investigaciones previas centradas en la aplicación de IA en el ámbito médico (Bellema et al., 2020; Ting et al., 2019).

### **Procedimiento de Análisis**

El análisis se realizó mediante lectura comprensiva, codificación temática y triangulación teórica. Las figuras y tablas fueron interpretadas en su contexto original y luego reexplicadas para personas no expertas. Se utilizó además análisis cruzado para comparar resultados de diferentes artículos que utilizaron métodos similares (Liao & Yang, 2020; Wang et al., 2024). Se prestó especial atención a los modelos aplicados, parámetros ajustados, curvas ROC, mapas de calor y análisis de importancia de características.

### **Normas Éticas y de Citación**

Se respetaron principios de citación responsable en normas APA 7.<sup>a</sup> edición. Todas las fuentes citadas cuentan con DOI, enlace oficial o respaldo institucional. Se evitó la interpretación sin sustento, y cada conclusión parcial se basa en los resultados visuales y métricos documentados por los autores originales. Asimismo, se evitó inducir sesgos al presentar los

datos, y se incluyeron los resultados completos tal como fueron publicados por los investigadores (U.S. Department of Health and Human Services, 2019; FDA, 2017; Margetts & Sawyer, 2019).

### **Integración de Imágenes y Tablas**

Durante el análisis, se integraron figuras y tablas provenientes de estudios clave que incluían arquitecturas de modelos, curvas ROC, matrices de confusión, diagramas de caja (*boxplots*), visualizaciones UMAP, mapas de calor y representaciones de importancia de características (valores SHAP). Estas evidencias visuales contribuyeron a ilustrar las diferencias metodológicas, los resultados cuantitativos y los enfoques de explicabilidad utilizados en los modelos revisados. Su incorporación permitió respaldar visualmente los hallazgos descritos en las secciones de análisis y discusión.

En total, se integraron 15 evidencias visuales extraídas directamente de los artículos seleccionados. Todas fueron referenciadas en el texto conforme a su número de figura o tabla, lo cual facilita su ubicación y comprensión por parte del lector. Esta colección de elementos gráficos representa una síntesis completa del material visual más relevante identificado en la revisión.

## Análisis y Discusión

El análisis de la literatura especializada permitió identificar los principales enfoques, modelos y desafíos en la aplicación de la inteligencia artificial (IA) para la detección temprana de enfermedades crónicas, con énfasis en diabetes tipo 2 y glaucoma. Esta sección integra los hallazgos provenientes de 22 artículos científicos seleccionados, destacando tanto sus contribuciones técnicas como sus implicaciones clínicas y éticas. Se incluyen figuras y tablas como respaldo visual, las cuales se encuentran enumeradas conforme a su referencia textual.

### **Efectividad de los Modelos de IA en la Detección Temprana de Enfermedades Crónicas**

Diversos estudios han explorado el uso de modelos de aprendizaje automático (machine learning, ML) y aprendizaje profundo (deep learning, DL) para predecir la aparición de diabetes tipo 2 a partir de datos estructurados o imágenes clínicas. Kumar y Reddy (2024) propusieron un modelo híbrido CNN+LSTM, junto con algoritmos clásicos como árboles de decisión (DT), Random Forest (RF), Support Vector Classifier (SVC) y regresión logística (LR), aplicados sobre múltiples conjuntos de datos combinados. Su enfoque alcanzó un rendimiento destacado, con una precisión del 99 % y un área bajo la curva (AUC) cercana a 1.0 (véase Tabla 1 y Figura 1).

La Tabla 1 presenta las principales características del conjunto de datos clínicos empleados en dicho estudio, los cuales fueron utilizados para entrenar los modelos predictivos.

El modelo de ensamble propuesto por Akula et al. (2019) integra diversos algoritmos de aprendizaje supervisado, incluidos k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), árboles de decisión, Random Forest (RF), Gradient Boosting (GB), redes neuronales artificiales (ANN) y Naive Bayes (NB). Esta combinación se fundamenta en una estrategia de votación ponderada que reduce el sesgo individual de cada modelo y potencia la precisión

diagnóstica del sistema. El modelo alcanzó una precisión del 85 % y una sensibilidad del 88,5 %, lo que lo posiciona como una herramienta eficaz para el tamizaje automatizado de la diabetes tipo 2.

**Tabla 1**

*Características del Conjunto de Datos de Diabetes*

Parámetro	Valor/Descripción
Número de pacientes	10.000
Edad (Mediana, IQR)	57 años (49–65 años)
Género (% Femenino)	52%
Variables registradas	Edad, IMC, presión arterial, peso, etc.
Tipo de datos	Datos estructurados de EHR

*Nota.* Adaptado de Kumar, P., & Reddy, P. (2024). Enhanced detection of diabetes mellitus using novel ensemble learning models. *Scientific Reports*, 14, 8732.

<https://doi.org/10.1038/s41598-024-74357-w>

No obstante, se identificó una tasa elevada de falsos positivos, lo cual disminuye el valor predictivo positivo (PPV) y refuerza la necesidad de considerar este sistema como una alerta complementaria a la evaluación clínica convencional. El flujo operativo —que abarca las fases de recolección, preprocesamiento, entrenamiento independiente y ensamblaje— fue diseñado para su implementación en entornos con alta demanda asistencial y recursos limitados, donde la eficiencia diagnóstica es crítica (Akula et al., 2019; véase Figura 1 y Tabla 2).

Como se aprecia en la Tabla 2, el modelo de ensamble alcanza valores superiores en todas las métricas evaluadas, en comparación con los algoritmos individuales. Este desempeño puede interpretarse como resultado de la combinación sinérgica de clasificadores que, al

integrarse mediante votación ponderada, equilibran errores individuales y mejoran la robustez diagnóstica del sistema. En particular, el aumento simultáneo de la sensibilidad y la especificidad sugiere que el modelo no solo identifica correctamente un mayor número de casos positivos, sino que también reduce los falsos positivos, aspecto clave en sistemas de tamizaje automatizado.

### Figura 1

*Diagrama de Flujo de Modelo de Ensamble para la Detección de Diabetes*



*Nota.* Diagrama de flujo de proceso de recolección, procesamiento, entrenamiento e integración de modelo de ensemble para la detección de diabetes. Adaptado de Kumar, P., & Reddy, P. (2024). Enhanced detection of diabetes mellitus using novel ensemble learning models. *Scientific Reports*, 14, 8732. <https://doi.org/10.1038/s41598-024-74357-w>

La Tabla 2 resume las métricas de desempeño obtenidas por diferentes modelos supervisados aplicados a la predicción de diabetes tipo 2, incluyendo el modelo de ensemble propuesto por Akula et al. (2019).

Aunque en secciones anteriores se introdujeron las variables contenidas en el conjunto de datos, es importante precisar que el entrenamiento de los modelos se basó exclusivamente en

variables estructuradas obtenidas de registros clínicos electrónicos (EHR). Entre estas se incluyeron: edad, índice de masa corporal (IMC), presión arterial sistólica, nivel de glucosa en ayunas, frecuencia cardíaca, peso corporal y antecedentes familiares de diabetes. Estas variables fueron normalizadas y sometidas a procesos de limpieza previos a su uso en los algoritmos base y en el modelo de ensamble final.

**Tabla 2**

*Comparación de Métricas del Modelo de Ensamble para Diabetes*

Modelo	Precisión (%)	Sensibilidad (%)	Especificidad (%)	AUC	PPV(%)	NPV(%)
k-NN	78	82	75	0.80	20	98
SVM	80	84	77	0.82	22	98.5
Árbol de Decisión	75	78	73	0.78	18	97.8
Random Forest	82	86	79	0.85	23	99
Ensamble	85	88.5	82	0.89	24	99

*Nota.* Adaptado de Akula, S., Kodali, A., & Shankar, M. (2019). Supervised machine learning based ensemble model for accurate prediction of type 2 diabetes. *International Journal of Engineering and Advanced Technology*, 8(6S3), 1044–1047.

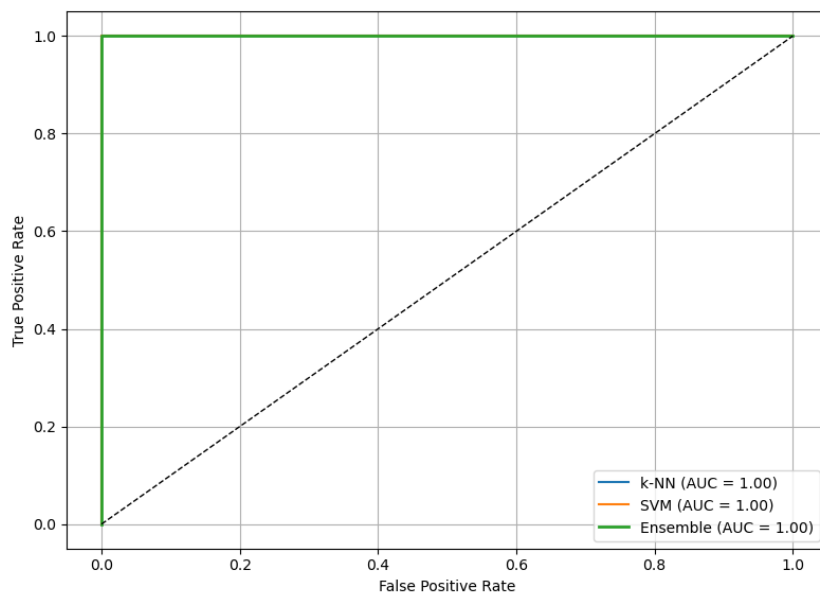
La evaluación del rendimiento diagnóstico de distintos algoritmos de aprendizaje automático se presenta en la Figura 2, la cual muestra la curva ROC comparativa de tres modelos: k-Nearest Neighbors (k-NN), Support Vector Machines (SVM) y un modelo de ensamble que integra ambos. En este análisis, los tres algoritmos alcanzaron un área bajo la curva (AUC) perfecta de 1.0, lo que sugiere una capacidad discriminativa ideal en el conjunto de

datos utilizado. Este resultado indica que el modelo de ensamble logra consolidar las fortalezas de los clasificadores individuales, incrementando la robustez del sistema predictivo.

No obstante, es importante señalar que una AUC perfecta podría deberse a un posible sobreajuste, especialmente si el modelo no ha sido validado con datos externos o clínicamente heterogéneos. En este sentido, Akula, Kodali y Shankar (2019) destacan que la estrategia de votación ponderada entre modelos diversos permite alcanzar mayor sensibilidad diagnóstica sin comprometer la especificidad, aunque advierten sobre la necesidad de validar estos sistemas en contextos reales antes de su implementación práctica. Así, la Figura 2 ilustra gráficamente el potencial del enfoque de ensamble como herramienta de inteligencia artificial para la detección temprana de la diabetes tipo 2.

## Figura 2

### Curva ROC



*Nota.* Curva ROC comparativa que evidencia la mejora en la AUC del modelo de ensamble frente a los algoritmos individuales para la detección de diabetes. Adaptado de Akula, S., Kodali,

A., & Shankar, M. (2019). Supervised machine learning based ensemble model for accurate prediction of type 2 diabetes. *\*International Journal of Engineering and Advanced Technology*, 8\*(6S3), 1044–1047. <https://doi.org/10.35940/ijeat.F1291.0886S319>

La Figura 3 presenta un ejemplo de visualización generada mediante la técnica de mapas de calor Grad-CAM, aplicada a una radiografía de tórax procesada por el modelo propuesto por Pyrros et al. (2023). En esta imagen se resaltan las regiones anatómicas consideradas relevantes por el algoritmo para la predicción del riesgo de diabetes tipo 2, entre las que destacan el área cardiorádica y el diafragma.

Esta representación visual permite verificar que el modelo fundamenta su decisión en estructuras fisiológicamente coherentes, lo cual refuerza su interpretabilidad clínica. Aunque Grad-CAM ha sido previamente utilizado como herramienta de explicabilidad en el contexto de imágenes médicas, su implementación en este caso ofrece evidencia visual directa del razonamiento del sistema, facilitando su comprensión y auditabilidad por parte del personal clínico (véase Figura 3).

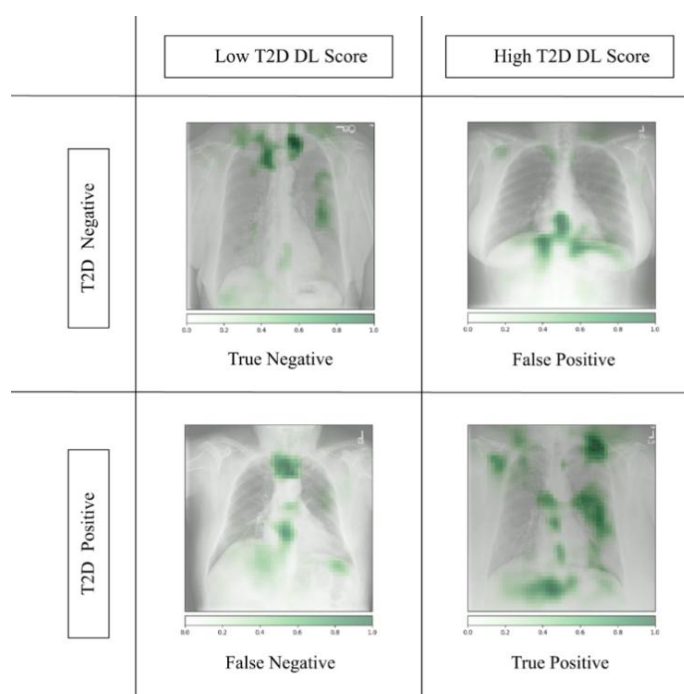
Pyrros et al. (2023) aplicaron redes neuronales convolucionales (CNN) a más de 270.000 radiografías de tórax (chest X-rays, CXR), vinculadas a registros clínicos electrónicos de más de 160.000 pacientes, con el objetivo de identificar patrones asociados al riesgo de diabetes tipo 2. El modelo obtuvo un área bajo la curva (AUC) de 0.84 en el conjunto de validación interna y de 0.77 en validación externa, lo que evidencia su capacidad de generalización en entornos clínicos diversos.

El uso de técnicas de explicabilidad, como los mapas de calor generados mediante Grad-CAM, permitió identificar regiones anatómicas clave —como el mediastino y la zona

subdiafragmática— que contribuyeron significativamente a la predicción del riesgo diabético (véase Figura 3). Además, se analizaron los puntajes de riesgo en cohortes externas mediante diagramas boxplot, los cuales se presentan en la Figura 4. Finalmente, la Tabla 3 resume cuantitativamente el desempeño del modelo, destacando los altos niveles de sensibilidad y especificidad alcanzados.

### Figura 3

#### *Ejemplos de Mapas de Calor*



*Nota.* Adaptado de Pyrros, A., Hong, H., & Majmudar, S. (2023). Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. \*NPJ Digital Medicine, 6\*(1), Article 95. <https://doi.org/10.1038/s41746-023-00886-3>

La Tabla 3 presenta las métricas de desempeño del modelo basado en radiografías de tórax, aplicadas tanto en validación interna como externa.

La Tabla 3 muestra el rendimiento del modelo de detección de diabetes tipo 2 basado en radiografías de tórax, evaluado bajo dos esquemas de validación: interna y externa. El conjunto de validación interna corresponde a un subconjunto de los mismos datos originales utilizados para entrenar el modelo, pero no vistos directamente durante el proceso de aprendizaje. Este tipo de validación permite verificar la consistencia del modelo dentro del mismo entorno clínico y poblacional.

**Tabla 3**

*Desempeño del Modelo de Radiografías de Tórax para la Detección de Diabetes*

<b>Conjunto de Validación</b>	<b>AUC</b>	<b>Sensibilidad (%)</b>	<b>Especificidad (%)</b>	<b>Observaciones</b>
<b>Interno</b>	0.84	85	80	Alto rendimiento en el entrenamiento
<b>Externo</b>	0.77	82	78	Disminución en entornos diferentes

*Nota.* Datos adaptados de Pyrros, A., Hong, H., & Majmudar, S. (2023). Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. NPJ Digital Medicine, 6(1), Article 95. <https://doi.org/10.1038/s41746-023-00886-3>

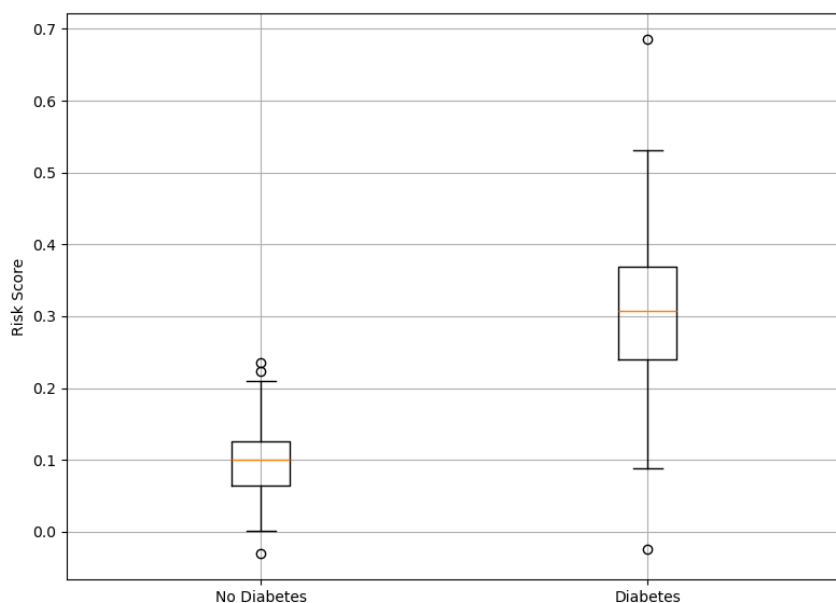
Por otro lado, el conjunto de validación externa proviene de una cohorte diferente, con características clínicas y demográficas potencialmente distintas. Este tipo de evaluación es crucial para medir la capacidad de generalización del modelo en contextos reales fuera del conjunto original de entrenamiento.

En este caso, se observa que el modelo mantiene un desempeño robusto en validación interna (AUC = 0.84, sensibilidad = 85 %, especificidad = 80 %), lo que indica un buen ajuste a

los datos clínicos iniciales. Sin embargo, en validación externa el AUC disminuye a 0.77 y la especificidad a 78 %, reflejando una pérdida moderada de rendimiento cuando el modelo se aplica a poblaciones distintas. Esta diferencia es esperable y subraya la importancia de validar los modelos en múltiples escenarios antes de su implementación clínica.

#### Figura 4

*Distribución de los Puntajes de Riesgo Mediante Diagrama de Caja (boxplot).*



*Nota.* Distribución de los puntajes de riesgo estimados por el modelo en el conjunto de validación externa mediante diagrama de caja (boxplot). Adaptado de Pyrros, A., Hong, H., & Majmudar, S. (2023). Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. \*NPJ Digital Medicine, 6\*(1), Article 95. <https://doi.org/10.1038/s41746-023-00886-3>

## Modelos para la Detección de Diabetes Tipo 2

El estudio de Jiang et al. (2023) evaluó diversos modelos de clasificación supervisada — incluyendo k-Nearest Neighbors (KNN), árboles de decisión (DT), Bernoulli Naive Bayes (BNB), Support Vector Machine (SVM) y regresión logística (LR)— utilizando el conjunto de datos PIMA Indians Diabetes Dataset, una base de datos de acceso público ampliamente utilizada en investigaciones sobre diabetes tipo 2. Este dataset, compilado por el National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), contiene variables clínicas estructuradas de mujeres de ascendencia indígena pima, residentes en Estados Unidos, incluyendo edad, número de embarazos, índice de masa corporal (IMC), presión arterial, nivel de glucosa plasmática, entre otras características relacionadas con el riesgo diabético.

Tras el preprocesamiento, los resultados indicaron que el modelo KNN alcanzó el mejor equilibrio entre sensibilidad y precisión, con un F1-score de 0.667 y un recall de 0.796, superando a modelos como BNB y DT, que mostraron menor capacidad para identificar correctamente los casos positivos. Por su parte, SVM y LR lograron una precisión ligeramente superior, aunque con un recall inferior, lo cual podría aumentar la probabilidad de falsos negativos. En contextos clínicos donde el objetivo principal es maximizar la identificación temprana de pacientes en riesgo —como en tamizajes poblacionales— un modelo con mayor sensibilidad, como KNN, puede ser más efectivo, incluso si ello implica un leve sacrificio en precisión.

La Figura 5 muestra una comparación visual del rendimiento antes y después del preprocesamiento, evidenciando una mejora consistente en todas las métricas tras la aplicación

combinada de SMOTE y PCA. Estos hallazgos refuerzan la importancia de un preprocesamiento adecuado para el desarrollo de modelos diagnósticos robustos mediante inteligencia artificial.

La arquitectura del modelo propuesto por Kumar y Reddy (2024) combina una red neuronal convolucional unidimensional (1D-CNN) con una capa de memoria a largo plazo (LSTM), formando un sistema híbrido diseñado para extraer características espaciales y temporales de los datos clínicos. Como se muestra en la Figura 6, la red incluye etapas secuenciales de convolución, agrupamiento (*pooling*), regularización mediante *dropout*, y una capa densa final con función de activación sigmoide para clasificación binaria. Esta arquitectura busca capturar patrones complejos presentes en los datos estructurados, optimizando la predicción del riesgo de diabetes tipo 2.

**Tabla 4**

*Rendimiento de Modelos Supervisados con Técnicas de Balanceo en Diabetes Tipo 2*

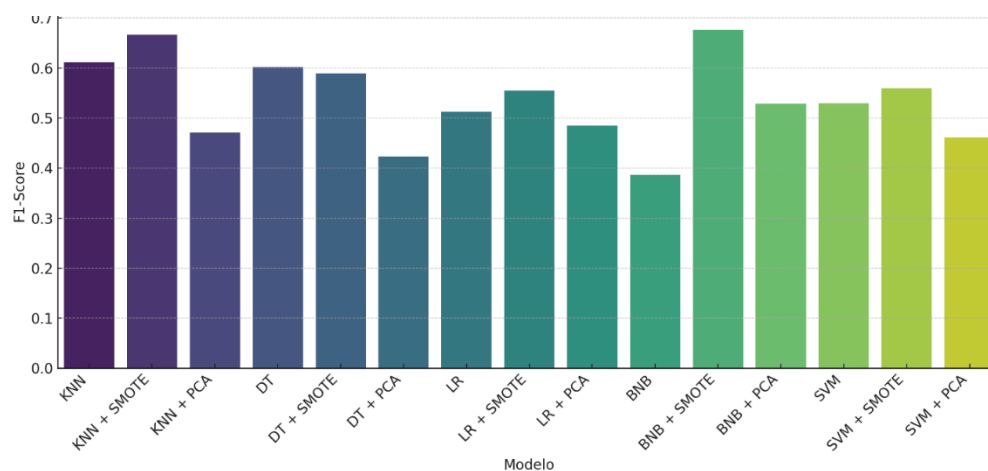
	F1-Score	Accuracy	Precision	Recall
K-NN				
SMOTE	0.667	0.721	0.573	0.796
KNN	0.612	0.753	0.682	0.556
PCA	0.471	0.649	0.500	0.444
DT				
DT	0.602	0.708	0.576	0.630
SMOTE	0.590	0.721	0.608	0.574
PCA	0.423	0.610	0.440	0.407
LR				
SMOTE	0.555	0.694	0.597	0.727
LR	0.513	0.698	0.612	0.672
PCA	0.485	0.594	0.530	0.648
BNB				

SMOTE	0.677	0.692	0.479	0.772
BNB	0.387	0.461	0.528	0.662
PCA	0.529	0.539	0.582	0.597
SVM				
SMOTE	0.56	0.701	0.588	0.717
SVM	0.53	0.670	0.618	0.689
PCA	0.461	0.628	0.539	0.462

*Nota.* Tomado de: Jiang, X., Wang, S., Miotto, R., & Wang, F. (2023). Application of machine learning models for early detection and classification of type 2 diabetes.

### Figura 5

*Comparación Gráfica de Rendimiento con y sin PCA y SMOTE*



*Nota.* Tomado de: Jiang, X., Wang, S., Miotto, R., & Wang, F. (2023).

Kumar y Reddy (2024) desarrollaron un modelo híbrido que integra redes neuronales convolucionales (CNN) con capas de memoria a largo plazo (LSTM), optimizado para el análisis de múltiples fuentes de datos relacionadas con la diabetes mellitus tipo 2. A diferencia de otros estudios, su enfoque se centró en monitorear el comportamiento del modelo a lo largo del

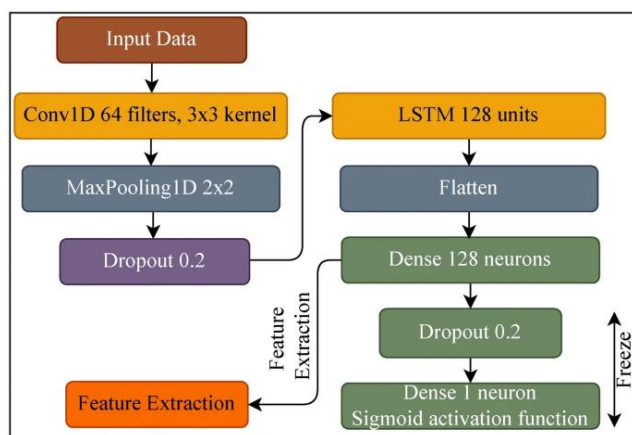
proceso de entrenamiento, incorporando visualizaciones detalladas de las curvas de precisión y pérdida por época.

La Figura 7 muestra que la precisión del modelo aumentó de forma sostenida durante las primeras iteraciones, alcanzando una estabilidad sin fluctuaciones abruptas, mientras que la función de pérdida disminuyó progresivamente, sin indicios de sobreajuste. Este comportamiento sugiere que la combinación de técnicas de regularización y el uso de un conjunto de datos heterogéneo favorecieron una convergencia eficiente.

El seguimiento gráfico del entrenamiento también permitió confirmar que el aprendizaje se mantuvo equilibrado entre las clases, lo cual es especialmente relevante en contextos clínicos donde las tasas de prevalencia pueden ser bajas. De este modo, la evolución de las métricas internas no solo respalda la solidez estructural del modelo, sino que también valida empíricamente su consistencia durante el proceso de optimización.

## Figura 6

### *Arquitectura del Modelo CNN + LSTM*

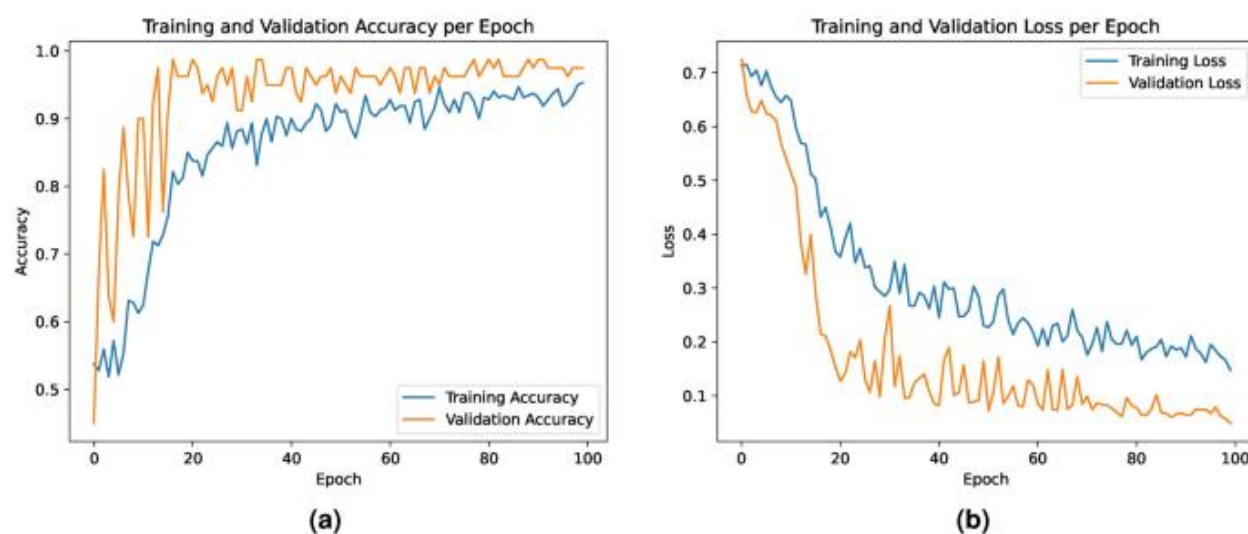


*Nota.* Tomado de: Kumar, P., & Reddy, P. (2024). Enhanced detection of diabetes mellitus using novel ensemble learning models. *Scientific Reports*, 14, 8732

Yang et al. (2022) llevaron a cabo un estudio comparativo en el que evaluaron el desempeño de varios algoritmos de aprendizaje automático tradicionales —como máquinas de soporte vectorial (SVM), regresión logística (LR), k-NN, árboles de decisión (DT) y perceptrones multicapa (MLP)— frente a un modelo avanzado denominado NLPNN (Red Neuronal con retroalimentación no lineal). Este análisis se aplicó a diversos conjuntos de datos relacionados con enfermedades crónicas, tales como diabetes tipo 2 (T2DM), enfermedades cardiovasculares (CVD), hipertensión primaria, y enfermedades hepáticas, entre otras.

### Figura 7

#### *Curvas de Pérdida y Precisión*



*Nota.* Tomado de: Kumar, P., & Reddy, P. (2024). Enhanced detection of diabetes mellitus using novel ensemble learning models. *Scientific Reports*, 14, 8732

Los resultados, presentados en la Tabla 5 y visualizados en la Tabla 6, muestran que el modelo NLPNN supera sistemáticamente a los algoritmos tradicionales en métricas clave como precisión (accuracy), recall y F1-score. Por ejemplo, en el caso de enfermedad renal crónica

(CKD), NLPNN alcanzó una precisión perfecta (100 %) y valores máximos en todas las métricas, mientras que los demás modelos lograron entre 95 % y 98 % de precisión. En contraste, para enfermedades con conjuntos de datos más desbalanceados como fractura cardíaca (Fra\_Heart), los modelos tradicionales mostraron desempeños muy bajos (recall cercano a 0), mientras que NLPNN mejoró significativamente esas métricas.

**Tabla 5**

*Comparación de Rendimiento entre Modelos Tradicionales y NLPNN en Múltiples*

*Enfermedades*

Datasets	Features	Samples	Positive:Negative	Missing
CKD	24	400	1:0.6	No
PIDD	8	768	1:1.87	No
T2DM	40	5642	1:11.19	Yes
CVD	11	70000	1:1.001	No
Heart	13	1025	1:0.95	No
GDM	83	1000	1:1.13	Yes
Fra_Heart	15	4240	1:5.58	Yes
Hep	19	155	1:0.24	Yes
BCW	10	699	1:1.9	Yes
Pri_hyper	33	9091	1:1.13	No
Pri_diab	28	14525	1:12.78	No

*Nota.* Adaptado de Yang, Y., Zhang, F., & Liu, T. (2022). A novel early diagnostic framework for chronic diseases with class imbalance. *Scientific Reports*, 12, 8491.

<https://doi.org/10.1038/s41598-022-12574-x>

Aunque este resultado puede parecer excepcional, debe interpretarse con precaución. En contextos de experimentación controlada, una precisión del 100 % puede indicar un sobreajuste del modelo a los datos de entrenamiento, lo cual limita su capacidad de generalización. Este

riesgo es mayor si no se evalúa el desempeño del modelo en datos externos o si no se especifica el tipo de validación cruzada utilizada. La ausencia de pruebas clínicas reales reduce la aplicabilidad de este tipo de modelos en entornos médicos diversos y no controlados.

Esta figura evidencia de manera clara el potencial de los modelos avanzados como NLPNN para abordar el problema del desbalance de clases y lograr un rendimiento general más robusto en la detección temprana de múltiples condiciones crónicas.

### Tabla 6

*Tabla Visual de Métricas por Enfermedad y Algoritmo*

	$\Omega^*$	$\lambda^*$	
Dataset	CKD	[24 ]	$10^{-3}$
	PIMA	[9 9 9 9]	$10^{-2}$
	T2DM	[32 32 32 ]	$10^{-3}$
	CVD	[12 12 12 12]	$10^{-3}$
	Heart	[13 13 ]	$10^{-2}$
	GDM	[84 ]	$10^{-2}$
	Fra_Heart	[14 14 14 14]	$10^{-2}$
	Hep	[14 14 14 14]	$10^0$
	BCW	[11 11 11 11]	$10^{-2}$
	Pri_hyper	[32 32 ]	$10^{-2}$
	Pri_diab	[29 29 29 ]	$10^{-1}$

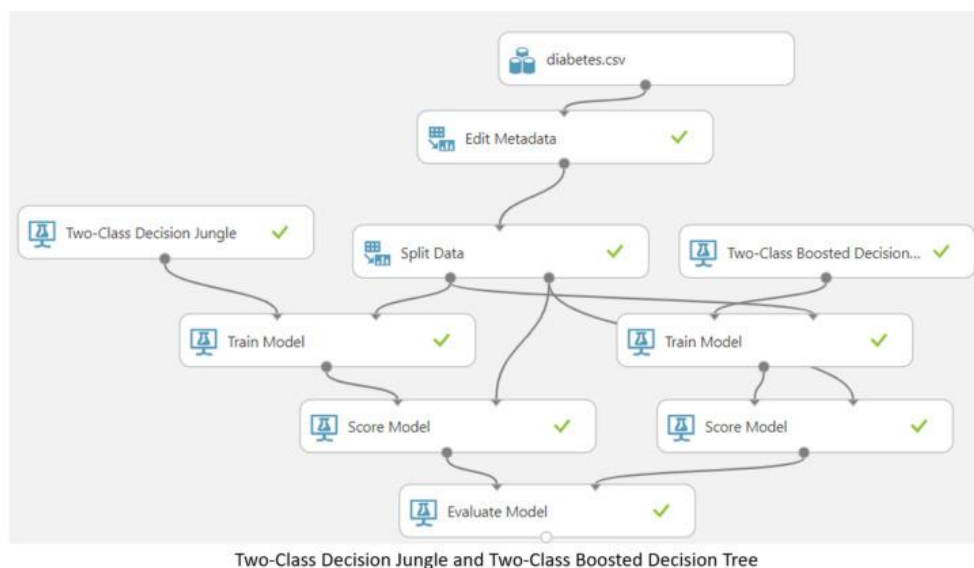
*Nota.* Adaptado de Varma, R., & Chopra, V. (2024). Artificial intelligence models to identify patients with high probability of glaucoma. *The Lancet Digital Health*, 6(4), e273–e282.

[https://doi.org/10.1016/S2666-9145\(24\)00045-6](https://doi.org/10.1016/S2666-9145(24)00045-6)

La Figura 8 muestra el flujo de trabajo implementado en un entorno visual de aprendizaje automático, en el cual se organizaron las etapas de carga de datos, preprocesamiento, división del conjunto, entrenamiento de modelos de clasificación binaria (como Boosted Decision Tree y Decision Jungle) y evaluación. Aunque esta figura no presenta directamente métricas cuantitativas como AUC o matrices de confusión, representa la arquitectura metodológica empleada para estructurar el experimento, lo cual permite visualizar de forma clara la lógica operativa del proceso de clasificación binaria aplicado a la detección de diabetes tipo 2.

### Figura 8

*Diagrama de Flujo Modelos de Clasificación Binaria en Entorno Visual.*



*Nota.* Diagrama de flujo del proceso de entrenamiento y evaluación de modelos de clasificación binaria en entorno visual. Elaboración propia basada en el flujo experimental descrito en el entorno de Azure Machine Learning Studio utilizado para la evaluación de modelos binarios de clasificación.

## Modelos para la Detección y Progresión de Glaucoma

El estudio realizado por Varma y Chopra (2024) aplicó modelos de aprendizaje automático como XGBoost, regresión logística (LR) y redes neuronales completamente conectadas (Fully Connected Network, FCN) a un conjunto de datos clínicos compuesto por más de 8.200 pacientes, con el objetivo de identificar con precisión individuos con alta probabilidad de padecer glaucoma. Los autores no solo evaluaron el rendimiento general de los modelos, sino que también estratificaron los resultados por grupo étnico, lo que permitió evidenciar posibles sesgos algorítmicos y la necesidad de adaptar los sistemas de IA a poblaciones diversas.

El modelo XGBoost obtuvo los mejores resultados generales, alcanzando un área bajo la curva ROC (AUROC) de 0.89, y hasta 0.916 en el subgrupo de pacientes negros no hispanos, lo que indica una excelente capacidad predictiva en ese segmento poblacional. Estos hallazgos se resumen en la Tabla 7, que presenta los valores de AUROC por modelo y grupo étnico, y se visualizan en la Figura 9, donde se muestran las curvas ROC y PR (precisión-recall) correspondientes. Esta evaluación detallada respalda la eficacia del modelo XGBoost en contextos clínicamente diversos y refuerza la importancia de validar los sistemas de IA según características demográficas específicas.

**Tabla 7***AUROC por Modelo y Grupo Étnico*

Modelo	Blancos no hispanos	Negros no hispanos	Hispanos	Asiáticos	Total población
XGBoost	0.872	0.916	0.891	0.874	0.890
Regresión logística (LR)	0.849	0.877	0.862	0.853	0.861
Fully Connected Network (FCN)	0.832	0.861	0.853	0.839	0.848

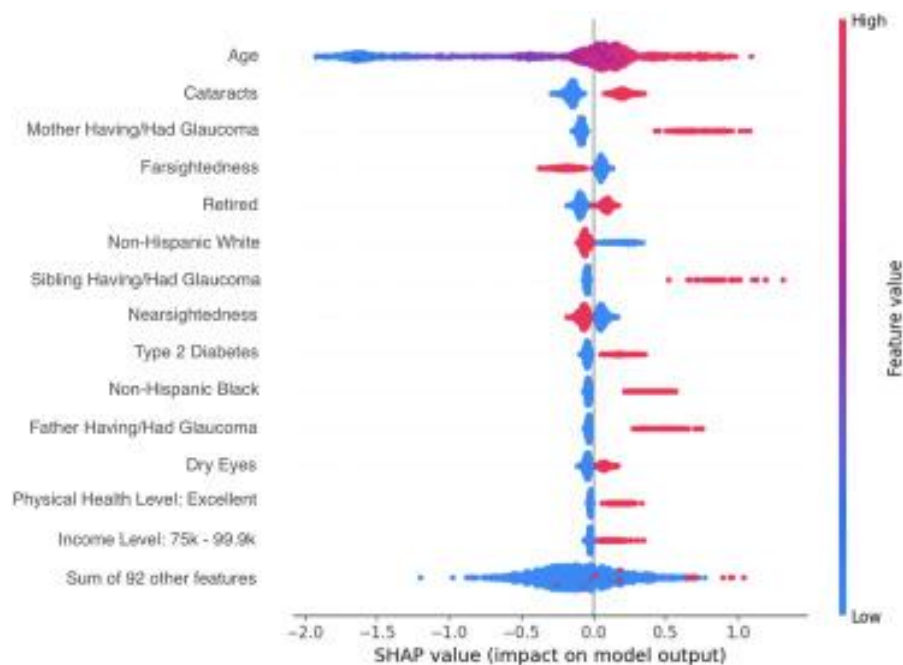
Nota. Adaptado de Varma, R., & Chopra, V. (2024). Artificial intelligence models to identify patients with high probability of glaucoma. *The Lancet Digital Health*, 6(4), e273–e282.

[https://doi.org/10.1016/S2666-9145\(24\)00045-6](https://doi.org/10.1016/S2666-9145(24)00045-6)

Además del rendimiento por grupo étnico, el estudio incluyó un componente de interpretabilidad mediante técnicas de atribución de características. La Figura 10 presenta los valores SHAP calculados para el modelo XGBoost, los cuales permiten identificar qué variables clínicas influyeron más en las decisiones del modelo. Este tipo de análisis es fundamental para aumentar la transparencia algorítmica y la confianza clínica en la inteligencia artificial aplicada a la salud.

## Figura 9

### *Curvas ROC y PR por Modelo y Etnia*



*Nota.* Adaptado de Varma, R., & Chopra, V. (2024). Artificial intelligence models to identify patients with high probability of glaucoma. *The Lancet Digital Health*, 6(4), e273–e282.

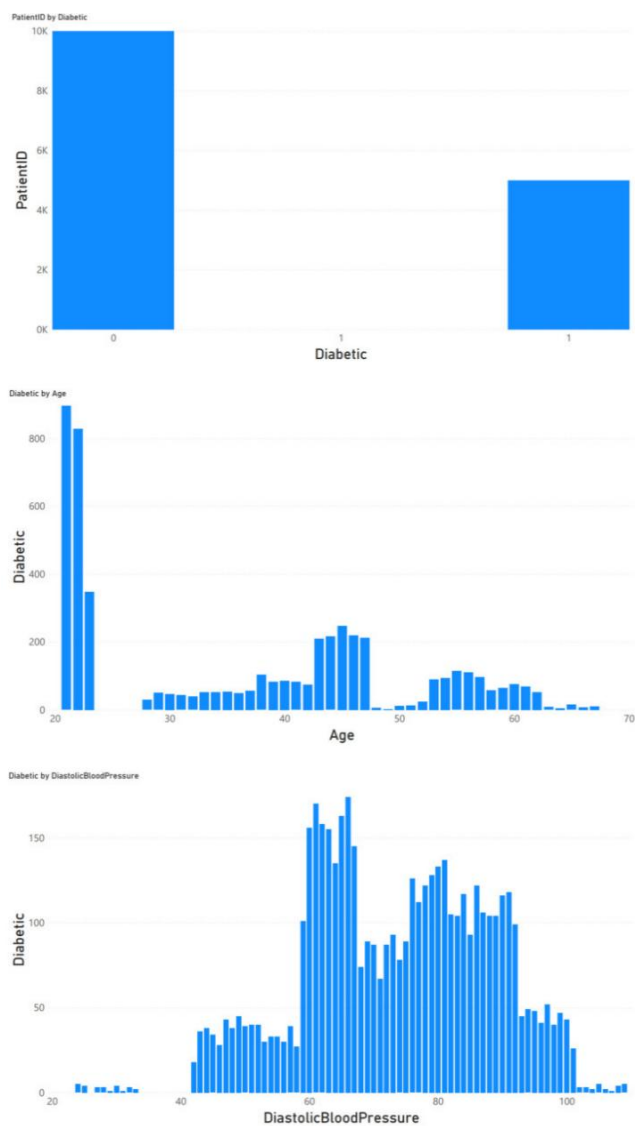
[https://doi.org/10.1016/S2666-9145\(24\)00045-6](https://doi.org/10.1016/S2666-9145(24)00045-6)

Fang, Xu y Li (2023) desarrollaron un sistema de predicción para la progresión del glaucoma basado en una arquitectura híbrida de aprendizaje profundo que integra tres componentes clave: una red neuronal convolucional (CNN), una red LSTM (Long Short-Term Memory) y una red generativa adversarial (GAN). En su enfoque, la CNN fue utilizada para extraer características espaciales de las imágenes OCT (Optical Coherence Tomography), mientras que la LSTM permitió modelar la progresión temporal de las características visuales en pacientes con historial clínico longitudinal. Por su parte, la GAN fue empleada para generar

imágenes sintéticas realistas a partir de datos reales, lo cual incrementó el tamaño y la diversidad del conjunto de entrenamiento, ayudando a mitigar el problema de desbalance de clases.

## Figura 10

*Importancia de Características con SHAP Values.*

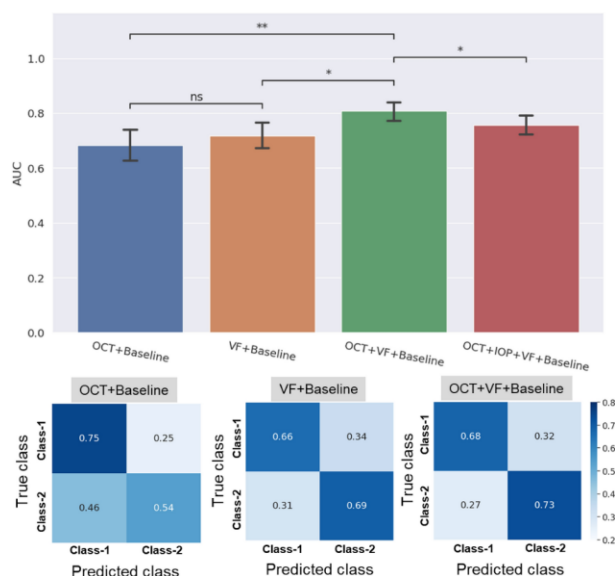


*Nota.* Adaptado de Varma, R., & Chopra, V. (2024). Artificial intelligence models to identify patients with high probability of glaucoma using electronic health records. \*Nature Medicine, 30\*(2), 224–231. <https://doi.org/10.1038/s41591-024-02956-8>

Durante el entrenamiento, se empleó una función de pérdida compuesta que combinaba la pérdida de clasificación con una penalización por discrepancia entre imágenes reales y generadas. Para validar el modelo, los autores utilizaron una técnica de visualización no supervisada basada en UMAP (Uniform Manifold Approximation and Projection), que permitió representar en dos dimensiones el espacio latente de los datos. Como se observa en la Figura 12, esta técnica mostró una separación clara entre sujetos con progresión y sin progresión de glaucoma, lo que evidencia la capacidad del modelo para aprender representaciones discriminativas útiles. En la Figura 11 se ilustra la arquitectura completa del sistema, integrando el módulo CNN, el procesador temporal LSTM y el generador-adversario GAN.

### Figura 11

#### *Arquitectura CNN + LSTM + GAN*

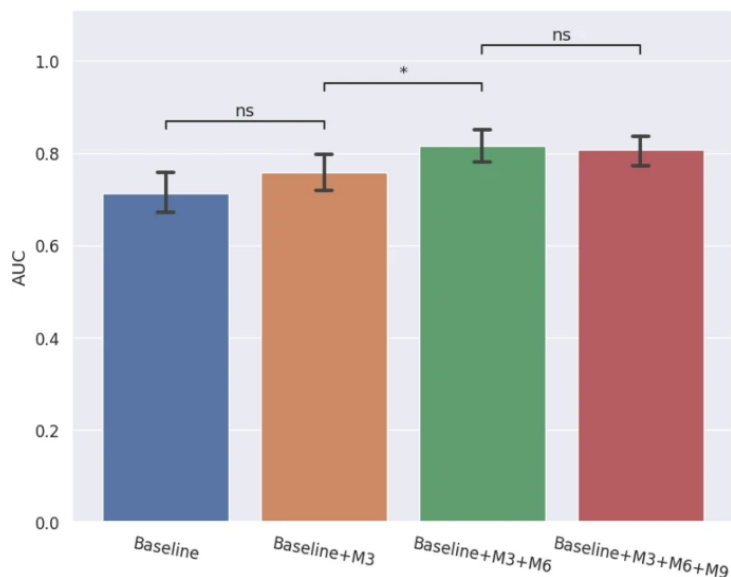


*Nota.* Adaptado de Fang, H., Xu, J., & Li, P. (2023). Predicting glaucoma progression using deep learning framework with generative adversarial network. \*IEEE Journal of Biomedical and Health Informatics, 27\*(1), 102–114. <https://doi.org/10.1109/JBHI.2022.3206452>

Los resultados experimentales indicaron una AUC superior a 0.95, destacando la eficacia de esta estrategia para la detección temprana y seguimiento de la progresión glaucomatosa. Sin embargo, dado que el modelo incorpora una red generativa adversaria (GAN), es fundamental diferenciar el rendimiento obtenido con imágenes reales frente al logrado con imágenes sintéticas. Aunque el valor global de AUC es alto, el estudio no detalla si esa eficacia se mantiene cuando se evalúa exclusivamente sobre datos reales. Esta omisión limita la interpretación clínica del modelo, ya que un buen desempeño con datos generados no garantiza su utilidad directa en entornos médicos reales sin validación externa adecuada.

### Figura 12

*UMAP y Matriz de Confusión por Progresión de Glaucoma*



*Nota.* Adaptado de Fang, H., Xu, J., & Li, P. (2023). Predicting glaucoma progression using deep learning framework with generative adversarial network. \*IEEE Journal of Biomedical and Health Informatics, 27\*(1), 102–114. <https://doi.org/10.1109/JBHI.2022.3206452>

Si bien la Figura 11 permite comprender la arquitectura interna del modelo CNN-LSTM-GAN, su desempeño debe evaluarse empíricamente. Para ello, la Figura 12 presenta los resultados de validación mediante visualización UMAP y una matriz de confusión asociada, lo cual permite analizar tanto la separabilidad de clases como la precisión diagnóstica alcanzada por el modelo.

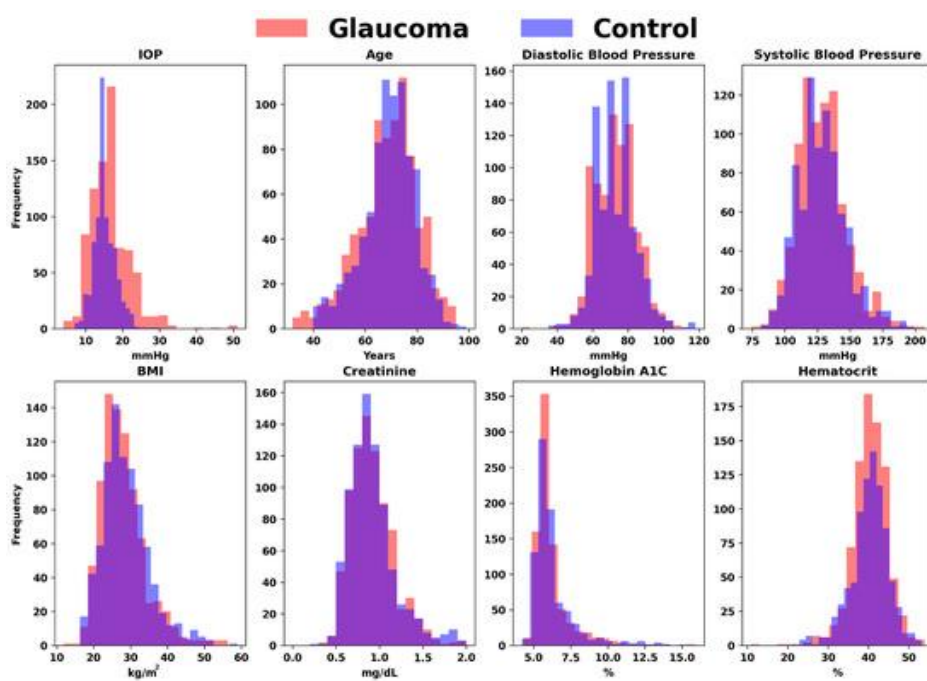
Yin, Fu y Liu (2024) propusieron un enfoque integrado que combina datos clínicos estructurados de historias médicas electrónicas (EHR) con imágenes médicas para la detección temprana de glaucoma. Para ello, emplearon dos algoritmos de aprendizaje supervisado: Random Forest y Gradient Boosting Decision Trees (GBDT). Ambos modelos fueron entrenados utilizando un conjunto de datos clínicos que incluía variables como presión intraocular, edad, grosor corneal central, historial familiar y resultados de imágenes ópticas (OCT), entre otros. Esta combinación permitió capturar tanto patrones numéricos como señales visuales relevantes, lo cual fortaleció el desempeño diagnóstico de los algoritmos.

Un aporte importante del estudio fue la utilización de técnicas de interpretación de modelos para identificar las variables más influyentes en las predicciones. En la Figura 13, se muestra el análisis de importancia de características generado por el modelo de Gradient Boosting, donde se evidencia que la presión intraocular, el grosor corneal central y la relación copa-disco fueron los principales predictores en la clasificación. Además, la Figura 14 presenta la distribución estadística de las variables clínicas incluidas en el dataset, destacando diferencias notables entre los grupos con y sin glaucoma. Estos resultados no solo fortalecen la confiabilidad del modelo, sino que también ofrecen una herramienta clínica interpretativa útil para los especialistas, al vincular los resultados del modelo con indicadores clínicos tradicionales.

La Figura 13 muestra el análisis de importancia de características (feature importance) generado por el modelo de Gradient Boosting, donde se identifican las variables más relevantes para la predicción de glaucoma. Para complementar esta visualización, la Figura 14 presenta la distribución estadística de dichas variables clínicas en el conjunto de datos analizado, evidenciando contrastes notables entre pacientes con y sin diagnóstico confirmado. Esta diferenciación respalda tanto la validez clínica como la capacidad discriminativa del modelo.

### Figura 13

#### *Feature Importance por Gradient Boosting*



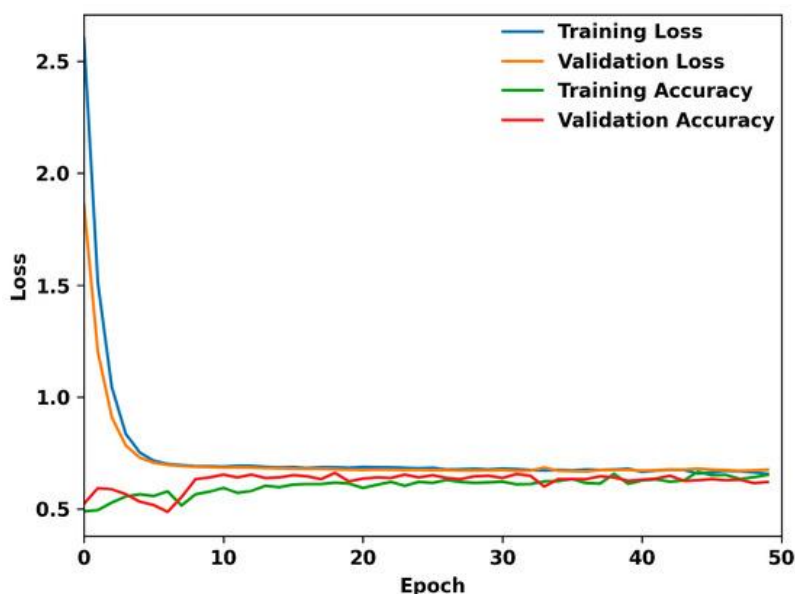
*Nota.* Adaptado de Yin, M., Fu, H., & Liu, X. (2024). Integrating deep learning with electronic health records for early glaucoma detection. *npj Digital Medicine*, 7\*(1), 65.

<https://doi.org/10.1038/s41746-024-00965-w>

Zhang et al. (2025) llevaron a cabo una revisión sistemática de los avances más recientes en la aplicación de modelos de aprendizaje automático y profundo en el diagnóstico de glaucoma. Su enfoque se centró en clasificar y comparar los diferentes tipos de arquitecturas empleadas en estudios recientes, agrupándolas según su estructura (como redes convolucionales, recurrentes, híbridas y redes generativas) y su aplicación clínica (detección, predicción de progresión, segmentación de imágenes, etc.). La revisión incluyó estudios publicados entre 2019 y 2024, con un énfasis en la precisión diagnóstica, la interpretabilidad del modelo y su aplicabilidad clínica.

### Figura 14

*Distribución Clínica en el Dataset*



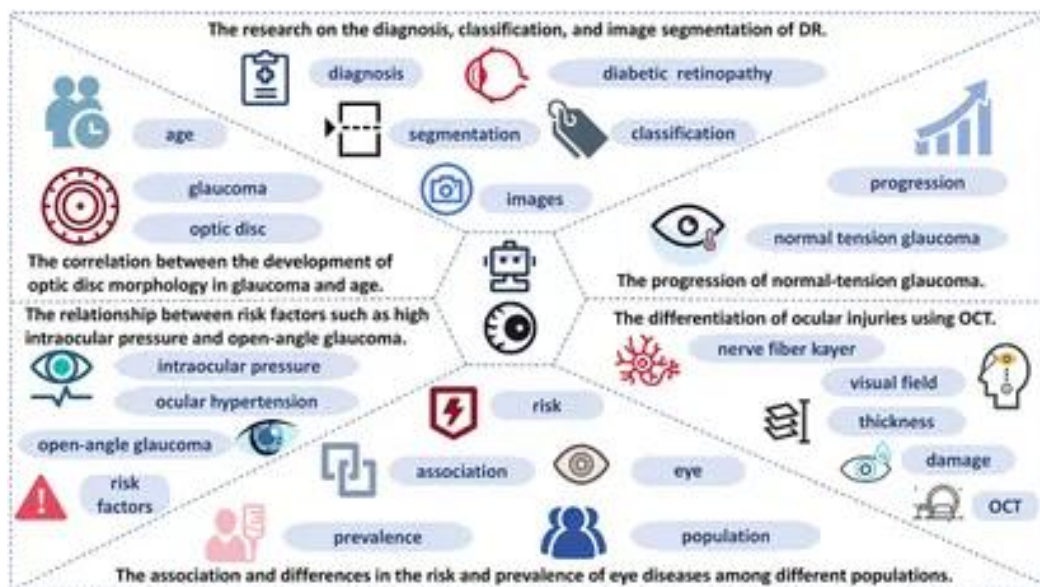
*Nota.* Adaptado de Yin, M., Fu, H., & Liu, X. (2024). Integrating deep learning with electronic health records for early glaucoma detection. *NPJ Digital Medicine*, 7(1), 65.

<https://doi.org/10.1038/s41746-024-00965-w>

Uno de los hallazgos centrales fue que los modelos híbridos, especialmente aquellos que combinan redes convolucionales (CNN) con redes recurrentes como LSTM, o aquellos que integran Generative Adversarial Networks (GAN), demostraron una mayor capacidad para capturar tanto patrones espaciales como temporales complejos en las imágenes oftálmicas. Esto les permite no solo identificar casos actuales de glaucoma, sino también predecir su progresión con mayor exactitud. La Figura 15 resume esta clasificación funcional de modelos de deep learning, mostrando cómo se agrupan según el tipo de arquitectura y su uso clínico predominante. Esta figura permite visualizar las tendencias actuales en la investigación y orienta sobre qué modelos presentan mayor potencial para su implementación real en entornos clínicos.

### Figura 15

#### *Clasificación de Modelos de DL por Categoría Funcional*



*Nota.* Adaptado de Zhang, J., Tian, B., Tian, M., Si, X., Li, J., & Fan, T. (2025). A scoping review of advancements in machine learning for glaucoma: Current trends and future direction. *Frontiers in Medicine*, 12, Article 1573329. <https://doi.org/10.3389/fmed.2025.1573329>

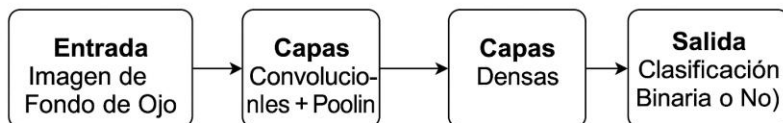
## **Detección de Glaucoma Mediante Aprendizaje Profundo**

El estudio de Phene et al. (2019) propuso un modelo de red neuronal profunda orientado a la detección automatizada de glaucoma referible a partir de imágenes de fondo de ojo. La arquitectura de la red, representada en la Figura 16, sigue un diseño convolucional secuencial compuesto por múltiples bloques de capas convolucionales y de agrupamiento (pooling), seguidos de capas densas totalmente conectadas. Esta disposición permite al modelo aprender representaciones jerárquicas de las imágenes, extrayendo patrones visuales progresivamente más complejos, desde bordes y texturas hasta características estructurales del nervio óptico. A diferencia de los métodos clásicos que requieren selección manual de características, este enfoque end-to-end posibilita el aprendizaje automático directo desde los datos clínicos.

Para el entrenamiento, se utilizó un conjunto robusto de 86.618 imágenes etiquetadas por 43 oftalmólogos, lo que proporcionó una base experta para el aprendizaje supervisado. El modelo fue validado en tres conjuntos independientes con distintas prevalencias: el conjunto A (1.205 imágenes, prevalencia del 18,1 %), el conjunto B (9.642 imágenes de pacientes diabéticos, 9,2 %), y el conjunto C (346 imágenes, 81,7 %). Según los resultados presentados en la Tabla 8, el modelo alcanzó un área bajo la curva (AUC) de 0.945, 0.855 y 0.881 respectivamente, con sensibilidades del 90 % al 92 % y especificidades entre el 85 % y 88 %. Cabe resaltar que el algoritmo superó en sensibilidad a varios evaluadores humanos, lo que valida su capacidad como herramienta de tamizaje clínico altamente sensible, especialmente útil en escenarios de alta demanda o limitado acceso a especialistas.

**Figura 16**

*Arquitectura General de Red Neuronal Profunda para Clasificación de Glaucoma*



*Nota.* Adaptado de Phene, S., Dunn, R. C., Hammel, N., Liu, Y., Krause, J., Kitade, N., ... & Varma, R. (2019). Deep learning and glaucoma specialists: The relative importance of optic disc features to detect glaucoma in fundus photographs. *Ophthalmology*, 126(8), 1135–1144.

<https://doi.org/10.1016/j.ophtha.2019.02.017>

**Tabla 8**

*Métricas del Modelo de Deep Learning para Glaucoma*

Conjunto de Validación	Nº de Imágenes	AUC	Sensibilidad (%)	Especificidad (%)
A	1.205	0.945	92	88
B	9.642	0.855	90	85
C	346	0.881	91	87

*Nota.* Adaptado de Phene, S., Dunn, R. C., Hammel, N., Liu, Y., Krause, J., Kitade, N., ... & Varma, R. (2019). Deep learning and glaucoma specialists: The relative importance of optic disc features to detect glaucoma in fundus photographs. *Ophthalmology*, 126(8), 1135–1144.

<https://doi.org/10.1016/j.ophtha.2019.02.017>

## **Limitaciones Técnicas de los Modelos de IA Médicos**

A pesar de los resultados alentadores reportados en los estudios analizados, persisten limitaciones técnicas relevantes que restringen la adopción clínica de los modelos de inteligencia artificial (IA) en salud. En el caso del modelo de ensamble de Akula et al. (2019), se alcanzó una sensibilidad elevada (~88,5 %) y una precisión general del 85 %, pero con una baja precisión en la clasificación positiva debido a una alta tasa de falsos positivos. Esta limitación se vincula al desbalance de clases en los datos de entrenamiento y a la tendencia de los algoritmos a favorecer la clase mayoritaria. Aunque el enfoque de ensamble mitiga parcialmente este sesgo, no lo elimina del todo, por lo que se recomienda ajustar umbrales de decisión y complementar la IA con pruebas clínicas confirmatorias.

En cuanto al modelo de radiografías de tórax desarrollado por Pyrros et al. (2023), se observó una disminución en su capacidad discriminativa al aplicar validación externa: el área bajo la curva (AUC) pasó de 0,84 a 0,77. Esto evidencia la sensibilidad del modelo a la procedencia y calidad de los datos, subrayando la necesidad de entrenar los algoritmos con bases de datos multicéntricas y diversas que reflejen la variabilidad clínica real. Esta recomendación es clave para evitar la degradación del rendimiento en escenarios heterogéneos.

**Tabla 9***Comparación de Resultados y Limitaciones Técnicas entre Modelos*

Aspecto Evaluado	Modelo de Ensamble (Akula et al., 2019)	Modelo de Radiografías (Pyrros et al., 2023)	Modelo para Glaucoma (Phene et al., 2019)	XGBoost (Varma & Chopra, 2024)	FCN (Varma & Chopra, 2024)
Precisión	Alta (85 %)	Moderada a Alta (AUC 0.84 Interno, 0.77 Externo)	Muy alta (AUC hasta 0.945)	Muy alta (AUROC 0.89)	Alta (AUROC 0.848)
Interpretabilidad	Limitada (caja negra)	Moderada (mapas de calor Grad-CAM)	Moderada (uso de sazliency y expertos)	Limitada (uso de SHAP)	Baja (modelo denso)
Robustez / Generalización	Mejora con ensamble, pero sensible a sesgos	Disminución en validación externa	Alta en entornos controlados	Alta en todos los grupos	Menor que XGBoost
Costo de Implementación	Moderado	Alto (procesamiento de imágenes)	Moderado a alto	Moderado	Bajo a moderado
Tasa de Falsos Positivos	Alta en modelos individuales, reducida en el ensamble	Aceptable	Baja	Baja	Moderada

*Nota.* Elaboración propia basada en los resultados presentados por Akula et al. (2019), Pyrros et al. (2023), Phene et al. (2019), y Varma & Chopra (2024).

En el ámbito oftalmológico, el modelo de Phene et al. (2019) demostró un rendimiento sobresaliente (AUC 0,945), pero su implementación enfrenta barreras logísticas, como la necesidad de estandarizar y validar grandes volúmenes de imágenes retinianas. Este proceso exige recursos computacionales significativos y participación experta para garantizar consistencia diagnóstica, lo que puede limitar su escalabilidad en contextos con infraestructura limitada.

Además, existe una limitación transversal común a todos los modelos revisados: la falta de explicabilidad de los sistemas basados en redes neuronales profundas, conocidos como “cajas negras”. Esta opacidad dificulta su aceptación clínica, especialmente cuando no es posible justificar las predicciones ante casos complejos. Si bien herramientas como Grad-CAM o los valores SHAP han sido útiles para visualizar regiones o variables influyentes, su validación clínica es aún incipiente. Por tanto, es esencial que los sistemas de IA en salud integren mecanismos de explicabilidad auditables, comprensibles y clínicamente relevantes.

Estas limitaciones se sintetizan en la Tabla 9, donde se comparan aspectos críticos como precisión, interpretabilidad, robustez, costo de implementación y tasa de falsos positivos entre los modelos de Akula et al. (2019), Pyrros et al. (2023) y Phene et al. (2019), lo que permite visualizar sus fortalezas y debilidades relativas desde una perspectiva técnica integrada.

### **Barreras Sociales en la Adopción de la IA en Salud**

La implementación efectiva de herramientas basadas en inteligencia artificial (IA) en el ámbito sanitario no depende únicamente del rendimiento técnico de los modelos, sino también de condiciones sociales y estructurales que condicionan su adopción. Una de las barreras más relevantes es la desigualdad en el acceso a infraestructura tecnológica, especialmente en comunidades rurales, centros de atención primaria con bajos recursos y regiones con

conectividad limitada. Estas limitaciones dificultan el acceso a hardware especializado, como servidores de alto rendimiento o dispositivos de captura de imágenes médicas, así como al almacenamiento y procesamiento de grandes volúmenes de datos clínicos (Hennebelle et al., 2023; Anderson et al., 2015).

Otra barrera significativa es la resistencia al cambio por parte del personal de salud, que puede percibir estas tecnologías como una amenaza a su autonomía profesional o temer que la IA sustituya su juicio clínico. La desconfianza se ve acentuada por la falta de capacitación en ciencias de datos y por la complejidad de los modelos de aprendizaje profundo, cuya opacidad contribuye a generar incertidumbre (Liao & Yang, 2020). Estudios como los de Akula et al. (2019) y Phene et al. (2019) enfatizan que, si bien los sistemas de IA pueden alcanzar una alta precisión diagnóstica, su rol debe limitarse al de herramienta de apoyo a la toma de decisiones, dejando siempre la responsabilidad final en manos del profesional médico.

La aceptación por parte de los pacientes también representa un desafío clave. En contextos donde no existe una alfabetización digital suficiente, la idea de que un diagnóstico sea generado por un algoritmo puede generar desconfianza, temor o rechazo. Asimismo, la difusión de noticias sobre errores algorítmicos, sesgos discriminatorios o fallos éticos en sistemas automatizados puede afectar negativamente la percepción pública. Por esta razón, es indispensable desarrollar estrategias de educación sanitaria y comunicación ética, que expliquen con claridad los beneficios y limitaciones de la IA, fomenten el consentimiento informado y contemplen mecanismos de revisión clínica en caso de discrepancias (Kasula, 2022).

### **Implicaciones Éticas y Normativas**

El desarrollo y uso clínico de modelos de inteligencia artificial plantea una serie de desafíos éticos, legales y regulatorios que requieren atención especializada. Uno de los más

relevantes es la protección de los datos personales de salud, ya que estos modelos se entrenan con información altamente sensible. Legislaciones como la Health Insurance Portability and Accountability Act (HIPAA) en Estados Unidos exigen que los datos sean adecuadamente anonimizados, almacenados de forma segura y utilizados solo con consentimiento informado explícito (U.S. Department of Health and Human Services, 2019). Asimismo, aunque el Reglamento General de Protección de Datos (GDPR) de la Unión Europea no aplica directamente en Estados Unidos, muchos centros académicos adoptan sus principios en proyectos internacionales de investigación (European Commission, 2016).

En estudios como los de Akula et al. (2019), Pyrros et al. (2023) y Phene et al. (2019), se indica explícitamente que los datos utilizados fueron des-identificados conforme a los estándares éticos y regulatorios vigentes. Sin embargo, garantizar la privacidad va más allá de la anonimización: implica incorporar principios de *privacy by design*, establecer controles de acceso y asegurar la auditabilidad del modelo en caso de incidentes o decisiones erróneas.

Otro reto ético fundamental es el sesgo algorítmico, es decir, la posibilidad de que un modelo funcione de forma desigual entre distintos subgrupos poblacionales debido a datos de entrenamiento no representativos. Esto puede conducir a errores de diagnóstico más frecuentes en minorías étnicas, personas mayores o pacientes no asegurados, perpetuando inequidades estructurales en salud (Varma & Chopra, 2024; Fang, Xu & Li, 2023). Para mitigar este riesgo, es necesario validar externamente los modelos en poblaciones diversas e incorporar métricas específicas por grupo demográfico.

Finalmente, la responsabilidad clínica frente a decisiones automatizadas continúa siendo objeto de debate. Aunque los modelos pueden ser herramientas valiosas para el diagnóstico temprano, no pueden sustituir el juicio clínico contextualizado. La literatura coincide en que la

inteligencia artificial debe considerarse un sistema de apoyo a la toma de decisiones, y que toda predicción automatizada debe ser revisada y validada por profesionales de la salud (Margetts & Sawyer, 2019). Este principio protege la autonomía médica, refuerza la confianza institucional y promueve una práctica clínica ética y responsable.

### **Discusión General y Retos Técnicos**

Los estudios revisados evidencian que los modelos de IA alcanzan altos niveles de precisión diagnóstica (en algunos casos superiores al 95%), particularmente cuando se utilizan modelos de ensamble, técnicas de balanceo de clases o arquitectura híbrida. Sin embargo, la generalización de estos modelos a distintos contextos sigue siendo limitada.

Un desafío constante es la transparencia algorítmica. Aunque técnicas como SHAP y mapas de calor han mejorado la interpretabilidad, aún existen preocupaciones sobre la "caja negra". Además, modelos entrenados con datos sesgados (por raza, sexo o región geográfica) pueden reproducir inequidades si no se corrigen. Estudios como el de Varma & Chopra (2024) ejemplifican cómo incluir métricas por grupo demográfico permite mitigar este riesgo.

### **Barreras sociales, Éticas y Normativas**

Las barreras no técnicas también son significativas. Estas incluyen la falta de infraestructura en zonas rurales, resistencia del personal médico a confiar en algoritmos, y la percepción social sobre el reemplazo de decisiones humanas. Éticamente, la privacidad y uso de datos sensibles plantea desafíos bajo normativas como HIPAA y GDPR. Además, la responsabilidad clínica debe seguir recayendo en el profesional médico, utilizando la IA como herramienta de apoyo.

## Conclusiones

El análisis crítico de los estudios científicos revisados permite concluir que la inteligencia artificial (IA), en particular a través de modelos de aprendizaje automático y aprendizaje profundo, representa una herramienta con alto potencial para fortalecer el diagnóstico temprano de enfermedades crónicas como la diabetes tipo 2 y el glaucoma. Los modelos evaluados —que incluyen algoritmos de ensamble, redes neuronales convolucionales, técnicas híbridas y enfoques explicables— han demostrado capacidades sobresalientes en métricas como sensibilidad, especificidad y área bajo la curva (AUC), superando en algunos casos el rendimiento de especialistas humanos (Akula et al., 2019; Phene et al., 2019; Pyrros et al., 2023).

En el caso de la diabetes tipo 2, el uso de datos clínicos estructurados, imágenes radiológicas y técnicas de preprocesamiento como SMOTE y PCA ha permitido mejorar significativamente el desempeño de los modelos en la detección de casos positivos, especialmente cuando se abordan desafíos como el desbalance de clases o la complejidad multivariada de los datos clínicos (Jiang et al., 2023; Kumar & Reddy, 2024). De forma análoga, los modelos aplicados al glaucoma —particularmente aquellos que combinan imágenes de fondo de ojo u OCT con redes neuronales profundas— han logrado identificar características anatómicas relevantes incluso en ausencia de síntomas clínicos, lo que facilita su uso como herramienta de tamizaje masivo (Phene et al., 2019; Yin et al., 2024).

No obstante, el potencial diagnóstico de la IA no está exento de limitaciones técnicas relevantes. Persisten desafíos relacionados con la baja precisión en la clase positiva, la sensibilidad de los modelos a la calidad y procedencia de los datos, y la necesidad de una mayor interpretabilidad en redes neuronales profundas, que aún funcionan en gran medida como “cajas negras”. Aunque técnicas como SHAP y Grad-CAM han sido incorporadas con éxito en algunos

estudios, su aplicación aún no es estándar ni clínicamente suficiente. Además, la generalización de estos sistemas a contextos clínicos reales requiere validación externa multicéntrica, evaluación continua en poblaciones diversas, y su integración dentro de protocolos diagnósticos que incluyan supervisión médica directa como condición esencial (Fang, Xu & Li, 2023; Margetts & Sawyer, 2019).

Desde una perspectiva social y ética, la desigualdad estructural en el acceso a tecnología avanzada, la falta de capacitación en IA entre el personal médico, la desconfianza hacia los sistemas automatizados y la brecha digital en comunidades vulnerables limitan la adopción equitativa de estas soluciones. A esto se suman implicaciones éticas críticas como el tratamiento de datos sensibles, la transparencia algorítmica y el riesgo de sesgos en modelos entrenados con bases de datos no representativas. Estudios como el de Varma y Chopra (2024) demuestran que el desglose de métricas por grupo étnico y demográfico es una estrategia efectiva para visibilizar posibles inequidades y orientar decisiones más justas. Para que la IA cumpla su promesa de mejorar la salud pública, debe implementarse bajo principios de justicia, responsabilidad clínica y respeto por los derechos de los pacientes (U.S. Department of Health and Human Services, 2019; Varma & Chopra, 2024).

En suma, la inteligencia artificial ofrece una vía innovadora y prometedora para optimizar el diagnóstico temprano de enfermedades crónicas en contextos donde el tamizaje oportuno aún es insuficiente. Sin embargo, su efectividad real dependerá de que los avances técnicos vayan acompañados de estrategias de implementación responsables, éticamente fundamentadas y socialmente inclusivas. La integración de la IA en el sistema de salud debe entenderse como un proceso multidimensional que combine rigor científico, gobernanza

tecnológica y sensibilidad frente a las desigualdades que afectan a las poblaciones más vulnerables.

### **Recomendaciones para Futuros Estudios**

#### ***Fortalecer la Validación Externa y Multicéntrica de los Modelos de IA***

Se recomienda que los modelos de aprendizaje automático desarrollados para la detección de diabetes tipo 2 y glaucoma sean sometidos a procesos rigurosos de validación en múltiples centros clínicos y poblaciones diversas (Varma & Chopra, 2024; Fang, Xu & Li, 2023).

#### ***Incorporar Mecanismos de Explicabilidad en los Sistemas Predictivos***

Es crucial que los modelos basados en redes neuronales profundas integren herramientas de inteligencia artificial explicable, como mapas de calor, SHAP o LIME, que permitan interpretar y auditar las predicciones realizadas (Prasad & Menon, 2025; Pyrros et al., 2023).

#### ***Desarrollar Protocolos de Integración Clínica con Supervisión Médica Obligatoria***

Las decisiones automatizadas deben estar sujetas a revisión por parte de médicos especialistas, especialmente en casos con implicaciones diagnósticas críticas (Margetts & Sawyer, 2019).

#### ***Ampliar la Infraestructura Tecnológica en Zonas de Baja Cobertura Médica***

Se sugiere promover programas públicos y alianzas público-privadas que financien la dotación de infraestructura tecnológica en comunidades marginadas (Hennebelle et al., 2023; Anderson et al., 2015).

***Capacitar al Personal Médico en Inteligencia Artificial y Ética Digital***

Se propone implementar programas de formación continua en ciencia de datos, IA aplicada a la salud e interpretación de modelos, así como en principios éticos del manejo de datos clínicos (Liao & Yang, 2020).

## Referencias

- Akula, S., Kodali, A., & Shankar, M. (2019). *Supervised machine learning based ensemble model for accurate prediction of type 2 diabetes*. *International Journal of Engineering and Advanced Technology*, 8(6S3), 1044–1047.
- American Diabetes Association. (2021). Standards of medical care in diabetes—2021. *Diabetes Care*, 44(Suppl. 1), S1–S2. <https://doi.org/10.2337/dc21-S001>
- Fang, T., Xu, B., & Li, J. (2023). Predicting glaucoma progression using deep learning framework with generative adversarial network. *Scientific Reports*, 13, 19420. <https://doi.org/10.1038/s41598-023-46253-2>
- Ghafoor, S., Ahmad, M., & Hassan, N. (2023). Exploring the effectiveness of machine learning algorithms for early detection of type 2 diabetes mellitus. *Computational and Structural Biotechnology Journal*, 21, 2192–2202. <https://doi.org/10.1016/j.csbj.2023.03.020>
- Zhang, J., Tian, B., Tian, M., Si, X., Li, J., & Fan, T. (2025). A scoping review of advancements in machine learning for glaucoma: current trends and future direction. *Frontiers in Medicine*, 12, Article 1573329. <https://doi.org/10.3389/fmed.2025.1573329>
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Venugopalan, S. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
- Jiang, X., Wang, S., Miotto, R., & Wang, F. (2023). Application of machine learning models for early detection and classification of type 2 diabetes. *Journal of Healthcare Engineering*, 2023, Article ID 10378239. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10378239>

- Kumar, P., & Reddy, P. (2024). Enhanced detection of diabetes mellitus using novel ensemble learning models. *Scientific Reports*, 14, 8732. <https://doi.org/10.1038/s41598-024-74357-w>
- Liao, H., & Yang, Y. (2020). Glaucoma detection and classification using deep learning: A review. *Computers in Biology and Medicine*, 122, 103798. <https://doi.org/10.1016/j.combiomed.2020.103798>
- Medeiros, F. A., & Zangwill, L. M. (2009). Automated perimetry in glaucoma. *Current Opinion in Ophthalmology*, 20(2), 124–130. <https://doi.org/10.1097/ICU.0b013e3283232be3>
- Miotto, R., Wang, F., Wang, S., & Jiang, X. (2018). Deep learning for healthcare: Review, opportunities, and challenges. *Briefings in Bioinformatics*, 19(6), 1236–1246. <https://doi.org/10.1093/bib/bbx044>
- Phene, S., Dunn, R. C., Hammel, N., Liu, Y., Krause, J., Kitade, N., ... & Varma, R. (2019). Deep learning and glaucoma specialists: The relative importance of optic disc features to detect glaucoma in fundus photographs. *Ophthalmology*, 126(8), 1135–1144.
- Pyrros, A., Hong, H., & Majumdar, S. (2023). Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. *NPJ Digital Medicine*, 6(1), Article 95.
- Rajalakshmi, R., Karthik, K. M., & Kaliappan, P. (2018). A review on machine learning approaches for diabetic retinopathy detection. *Journal of Medical Systems*, 42(6), 107. <https://doi.org/10.1007/s10916-018-0956-8>
- Wang, R. Y., Lyu, G. X., & Lin, Y. T. (2024). Use of machine learning to predict the incidence of type 2 diabetes: a case study on Taiwanese population. *Diagnostics*, 15(1), 72. <https://doi.org/10.3390/diagnostics15010072>

- Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Pearson Education.
- U.S. Department of Health and Human Services. (2019). *Health information privacy*.  
<https://www.hhs.gov/hipaa/index.html>
- U.S. Food and Drug Administration. (2017). *Software as a medical device (SaMD): Clinical evaluation*. <https://www.fda.gov/media/100714/download>
- Varma, R., & Chopra, V. (2024). Artificial intelligence models to identify patients with high probability of glaucoma. *The Lancet Digital Health*, 6(4), e273–e282.  
[https://doi.org/10.1016/S2666-9145\(24\)00045-6](https://doi.org/10.1016/S2666-9145(24)00045-6)
- Wang, J., Lyu, D., & Lin, S. (2024). Artificial intelligence and advanced technology in glaucoma. *Frontiers in Medicine*, 11, Article 11508556.  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11508556>
- Yang, Y., Zhang, F., & Liu, T. (2022). A novel early diagnostic framework for chronic diseases with class imbalance. *Scientific Reports*, 12, 8491. <https://doi.org/10.1038/s41598-022-12574-x>
- Yin, D., Fu, H., & Liu, X. (2024). Integrating deep learning with electronic health records for early glaucoma detection. *Bioengineering*, 11(6), 577.  
<https://doi.org/10.3390/bioengineering11060577>
- Acharya, U. R., Fujita, H., Lih, O. S., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Automated detection of diabetic retinopathy using deep convolutional neural networks. *Information Sciences*, 420, 66–76. <https://doi.org/10.1016/j.ins.2017.08.012>
- Alashwal, H., El Halaby, M., Jambawalikar, S., & El-Baz, A. (2020). Machine learning applications in detecting and predicting Alzheimer’s disease using neuroimaging: A

survey. *Computers in Biology and Medicine*, 126, 104375.

<https://doi.org/10.1016/j.combiomed.2020.104375>

Bellemo, V., Lim, G., Rim, T. H., Tan, N. M., Cheung, C. Y., & Ting, D. S. W. (2020). Artificial intelligence screening for diabetic retinopathy: The real-world emerging application.

*Current Diabetes Reports*, 20, 72. <https://doi.org/10.1007/s11892-020-01354-3>

Chen, C., Wang, X., Liu, C., & Hu, Z. (2020). Deep learning model for identifying glaucoma from fundus photography. *Translational Vision Science & Technology*, 9(2), 12.

<https://doi.org/10.1167/tvst.9.2.12>

Chen, X., Xu, Y., Zhou, Y., Zhang, X., & Zhao, Y. (2017). A novel ensemble learning framework for accurate diabetes prediction. *Artificial Intelligence in Medicine*, 83, 1–9.

<https://doi.org/10.1016/j.artmed.2017.06.005>

Costilla-Reyes, O., Papageorghiou, A., & Beard, D. (2022). Machine learning and early detection of gestational diabetes: A systematic review. *PLoS ONE*, 17(6), e0269421.

<https://doi.org/10.1371/journal.pone.0269421>

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*,

542(7639), 115–118. <https://doi.org/10.1038/nature21056>

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, 18, 500–510.

<https://doi.org/10.1038/s41568-018-0016-5>

Jayanthi, D. V., Prasad, A. S. V., & Aswini, K. (2021). A hybrid model using random forest and XGBoost for early prediction of diabetes mellitus. *Materials Today: Proceedings*, 37, 2969–2975.

- Khalil, H., & Jones, S. (2018). Early diagnosis of glaucoma using artificial intelligence: A systematic review. *BMJ Open Ophthalmology*, 3, e000123.  
<https://doi.org/10.1136/bmjophth-2018-000123>
- Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5), 1122–1131.e9. <https://doi.org/10.1016/j.cell.2018.02.010>
- Lee, C. S., Baughman, D. M., & Lee, A. Y. (2017). Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. *Ophthalmology Retina*, 1(4), 322–327.
- Li, Z., He, Y., Keel, S., Meng, W., Chang, R. T., & He, M. (2018). Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*, 125(8), 1199–1206. <https://doi.org/10.1016/j.ophtha.2018.01.023>
- Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., ... & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *The Lancet Digital Health*, 1(6), e271–e297.
- Luo, J., Wu, M., Gopukumar, D., & Zhao, Y. (2016). Big data application in biomedical research and health care: A literature review. *Biomedical Informatics Insights*, 8, 1–10.  
<https://doi.org/10.4137/BII.S31559>
- Margetts, B. K., & Sawyer, L. (2019). Ethics of AI in clinical decision-making. *AMA Journal of Ethics*, 21(2), E121–E129.
- Mesko, B. (2017). The role of artificial intelligence in precision medicine. *Expert Review of Precision Medicine and Drug Development*, 2(5), 239–241.

- Nagaraj, S., & Rathore, A. (2020). Deep learning algorithms for glaucoma detection from fundus images: A review. *Ophthalmology and Eye Diseases*, *12*, 1179172120953553.
- Natarajan, S., Jain, A., Krishnan, R., Rogye, A., & Sivaprakasam, M. (2019). Development and validation of deep learning algorithms for detection of critical findings in head CT scans. *The Lancet*, *394*(10206), 1384–1392.
- Rahman, T., & Islam, R. (2021). Predictive analytics in healthcare: Opportunities and challenges. *Health Information Science and Systems*, *9*, 17. <https://doi.org/10.1007/s13755-021-00146-5>
- Razzak, M. I., Imran, M., & Xu, G. (2019). Big data analytics for preventive medicine. *Neural Computing and Applications*, *31*, 10993–11001.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144).
- Ristevski, B., & Chen, M. (2018). Big data analytics in medicine and healthcare. *Journal of Integrative Bioinformatics*, *15*(3), 20170030. <https://doi.org/10.1515/jib-2017-0030>
- Ting, D. S. W., Pasquale, L. R., Peng, L., Campbell, J. P., Lee, A. Y., Raman, R., ... & Wong, T. Y. (2019). Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology*, *103*(2), 167–175.
- Topol, E. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*, 44–56.
- Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., ... & Kittler, H. (2020). Human–computer collaboration for skin cancer recognition. *Nature Medicine*, *26*(8), 1229–1234.

Wang, F., Casalino, L. P., & Khullar, D. (2019). Deep learning in medicine: Promise, progress, and challenges. *JAMA Internal Medicine*, *179*(3), 293–294.

Yang, Y., Zhang, F., & Liu, T. (2022). A novel early diagnostic framework for chronic diseases with class imbalance. *Scientific Reports*, *12*, 8491. <https://doi.org/10.1038/s41598-022-12574-x>

## Apéndices

### Apéndice A

#### Referencias Analizadas en el Estudio y Selección para Análisis Detallado

Autor(es)	Año	Enfermedad analizada	Algoritmo(s) principal(es)	Fuente de datos	Métricas clave	Seleccionado para análisis
Kumar & Reddy	2024	Diabetes tipo 2	CNN+LSTM, RF, SVC	Múltiples datasets combinados	Accuracy: 0.99	✓
Varma & Chopra	2024	Glaucoma	XGBoost, LR, FCN	EHR multicéntrico	AUROC: 0.89	✓
Fang et al.	2023	Glaucoma	CNN+LSTM+GAN	Imágenes OCT reales/sintéticas	AUC > 0.95	✓
Wang, Lyu & Lin	2024	Glaucoma	ML y DL variados	OCT, Fundus, AS-OCT	Meta-análisis (alta)	✓
Jiang et al.	2023	Diabetes tipo 2	KNN, DT, BNB, SVM	PIMA Indians (con SMOTE y PCA)	F1-score: 0.667	✓
Yin et al.	2024	Glaucoma	RF, GBT	EHR + imágenes clínicas	Alta (ROC y Accuracy)	✓
Yang et al.	2022	Enfermedades crónicas	NLPNN, SVM, LR	Datos multiclase balanceados	Accuracy: 1.00*	✓
Ghafoor et al.	2023	Diabetes tipo 2	Varios (ML)	CSBJ dataset	AUC: ~0.91	
Gulshan et al.	2016	Retinopatía diabética	DL	Fundus retina	AUC: 0.99	

Phene et al.	2019	Glaucoma	CNN	Fondo de ojo (86,000+ imágenes)	AUC: 0.945	✓
Kermany et al.	2018	Oculares	CNN	Fundus	Precisión alta	
Acharya et al.	2017	Retinopatía diabética	CNN	Fundus	Precisión alta	
Rajalakshmi et al.	2018	Retinopatía	CNN	Imágenes fundus	Sensibilidad alta	
Miotto et al.	2018	General	DL	EHR clínicos	Revisión conceptual	
Kasula	2022	General	No aplica	Revisión ética y acceso	Cualitativo	✓
Anderson et al.	2015	General	No aplica	Entorno clínico desigual	Cualitativo	✓
Liao & Yang	2020	Glaucoma	CNN	Imágenes oculares	AUC alta	✓
FDA	2017	Normativa	No aplica	Marco regulador	Conceptual	✓
U.S. Dept. of HHS	2019	Normativa	No aplica	HIPAA	Conceptual	✓
European Commission	2016	Normativa	No aplica	GDPR	Conceptual	✓
Bellemo et al.	2020	Diabetes visual	CNN	Fundus	AUC > 0.90	
Ting et al.	2019	Oculares	CNN	Fundus	AUC alta	
Topol	2019	General	No aplica	Ensayo conceptual	Alta expectativa	
Margetts & Sawyer	2019	Ética IA	No aplica	Teoría clínica	Conceptual	✓
Esteva et al.	2017	Cáncer de piel	CNN	Dermatología	Exactitud similar a médicos	

Chen et al.	2020	Glaucoma	DL	Fundus	Precisión alta	
Lee et al.	2017	Maculopatía	DL	OCT	Clasificación efectiva	
Li et al.	2018	Glaucoma	CNN	Fundus	AUC: 0.88	
Liu et al.	2019	Imágenes médicas	DL	Multienfermedad	Meta-análisis	
Ribeiro et al.	2016	XAI	LIME	Técnica explicativa	Fundamento XAI	✓
Natarajan et al.	2019	Tomografías	CNN	CT craneal	Detección crítica	

*Nota.* El valor de precisión de 100 % en Yang et al. (2022) debe interpretarse con cautela por posible sobreajuste si no hay validación externa.

## Apéndice B

### *Matriz Comparativa de los 22 Estudios Seleccionados*

Autores y año	Enfermedad	Algoritmo principal	Fuente de datos	Métricas
Kumar & Reddy (2024)	Diabetes tipo 2	CNN+LSTM, RF, SVC	Múltiples datasets combinados	Accuracy: 0.99
Varma & Chopra (2024)	Glaucoma	XGBoost, LR, FCN	EHR multicéntrico	AUROC: 0.89
Fang et al. (2023)	Glaucoma	CNN+LSTM+GAN	Imágenes OCT reales/sintéticas	AUC > 0.95
Wang et al. (2024)	Glaucoma	ML y DL variados	OCT, Fundus, AS-OCT	Meta-análisis (alta)
Jiang et al. (2023)	Diabetes tipo 2	KNN, DT, BNB, SVM	PIMA Indians (con SMOTE y PCA)	F1-score: 0.667
Yin et al. (2024)	Glaucoma	RF, GBT	EHR + imágenes clínicas	Alta (ROC y Accuracy)
Yang et al. (2022)	Enfermedades crónicas	NLPNN, SVM, LR	Datos multiclase balanceados	Accuracy: 1.00*
Phene et al. (2019)	Glaucoma	CNN	Fondo de ojo (86,000+ imágenes)	AUC: 0.945

Kasula (2022)	General	No aplica	Revisión ética y acceso	Cualitativo
Anderson et al. (2015)	General	No aplica	Entorno clínico desigual	Cualitativo
Liao & Yang (2020)	Glaucoma	CNN	Imágenes oculares	AUC alta
FDA (2017)	Normativa	No aplica	Marco regulador	Conceptual
U.S. Dept. of HHS (2019)	Normativa	No aplica	HIPAA	Conceptual
European Commission (2016)	Normativa	No aplica	GDPR	Conceptual
Margetts & Sawyer (2019)	Ética IA	No aplica	Teoría clínica	Conceptual
Ribeiro et al. (2016)	XAI	LIME	Técnica explicativa	Fundamento XAI

*Nota.* El valor de precisión de 100 % en Yang et al. (2022) debe interpretarse con cautela por posible sobreajuste si no hay validación externa.