

Machine Learning y Big Data en la gestión del riesgo en salud

Machine Learning and Big Data in Health Risk Management

Heber Esteban Bermúdez González¹ <https://orcid.org/0009-0000-1598-0927>

Carmen Johana Mosquera Agualimpia¹ <https://orcid.org/0009-0003-8584-4121>

Mónica Liliana Arteaga Sierra^{1*} <https://orcid.org/0000-0002-6433-6086>

¹Centro de Servicios de Salud. Medellín, Colombia.

*Autor para la correspondencia: marteaga@sena.edu.co

RESUMEN

Introducción: La revolución tecnológica ha transformado profundamente la atención médica, al acumular grandes volúmenes de datos clínicos. *Machine Learning* y *Big Data* han surgido como herramientas clave para mejorar la gestión del riesgo y la personalización de tratamientos, al presentar, tanto oportunidades como desafíos en el ámbito de la salud.

Objetivo: Realizar una revisión bibliográfica sobre la utilización de *Machine Learning* y *Big Data* como herramientas clave para mejorar la gestión del riesgo y la personalización de tratamientos.

Métodos: Se realizó una revisión bibliográfica sistemática bajo las pautas de PRISMA, y la utilización de bases de datos académicas y científicas. Se seleccionaron estudios relevantes de los últimos 20 años en inglés y español, con enfoque en la aplicación de ciencia de datos e inteligencia artificial en la salud. La revisión se hizo en dos etapas: selección inicial por títulos y resúmenes, seguida por una lectura detallada de los textos completos.

Resultados: La revisión destacó avances significativos en áreas como el diagnóstico por imágenes y el análisis de datos de dispositivos portátiles, que han mejorado la precisión en la predicción de riesgos de salud. Sin embargo, se

identificaron desafíos éticos y técnicos, como el sesgo en los algoritmos y expectativas desmesuradas sobre las capacidades de estas tecnologías.

Conclusiones: *Machine Learning* y *Big Data* tienen un gran potencial para transformar la atención médica, pero es crucial combinar estas tecnologías con la inteligencia humana para superar los desafíos éticos y técnicos. Esto garantizará una atención médica más personalizada, eficiente y equitativa.

Palabras clave: *Big Data*; desafíos éticos; inteligencia artificial; *Machine Learning*; predicción de riesgos.

ABSTRACT

Introduction: The technological revolution has profoundly transformed healthcare through the accumulation of vast volumes of clinical data. Machine Learning and Big Data have emerged as key tools for improving risk management and treatment personalization, presenting both opportunities and challenges within the healthcare sector.

Objective: To conduct a literature review on the use of Machine Learning and Big Data as key tools for enhancing risk management and treatment personalization.

Methods: A systematic literature review was conducted following PRISMA guidelines, utilizing academic and scientific databases. Relevant studies from the last 20 years, published in English and Spanish, were selected, focusing on the application of data science and artificial intelligence in healthcare. The review was carried out in two stages: an initial selection based on titles and abstracts, followed by a detailed reading of the full texts.

Results: The review highlighted significant advances in areas such as medical imaging and the analysis of data from wearable devices, which have improved the accuracy of health risk prediction. However, ethical and technical challenges were identified, such as algorithmic bias and inflated expectations regarding the capabilities of these technologies.

Conclusions: Machine Learning and Big Data hold great potential to transform healthcare; however, it is crucial to combine these technologies with human intelligence to overcome ethical and technical challenges. This will ensure more personalized, efficient, and equitable healthcare.

Keywords: Big Data; ethical challenges; artificial intelligence; Machine Learning; risk prediction.

Recibido: 01/11/2024

Aceptado: 18/02/2026

Introducción

La gestión del riesgo en salud es esencial para garantizar la prevención, el diagnóstico oportuno y el tratamiento adecuado de enfermedades en la población. En el contexto actual, caracterizado por la revolución de la Industria 4.0 y la creciente relevancia de la ciencia de datos, el aprendizaje automático y la inteligencia artificial (IA) han emergido como herramientas poderosas en la transformación de la atención médica.

Las técnicas de *Machine Learning* y la inteligencia artificial han demostrado ser efectivas en la predicción del riesgo de enfermedades infecciosas en ambientes con altos y bajos recursos. Estas tecnologías son esenciales para proporcionar predicciones de riesgo precisas y detalladas, en especial en regiones donde la carga de enfermedades es alta.⁽¹⁾

La implementación de *Machine Learning* y *Big Data* en la gestión del riesgo en salud ha permitido el desarrollo de modelos predictivos avanzados, análisis en tiempo real de grandes conjuntos de datos, y la identificación de patrones y tendencias que anteriormente eran difíciles de detectar.⁽²⁾ Estas capacidades han llevado a avances significativos en áreas como la predicción de brotes de enfermedades, la identificación de factores de riesgo y la personalización de tratamientos médicos.⁽³⁾

La revolución tecnológica ha impulsado la aplicación de la inteligencia artificial en áreas específicas de la salud, como la salud mental, lo que mejora la relación médico-paciente y proporciona análisis más precisos de los datos.⁽⁴⁾ Además, la combinación de IoT (internet de las cosas) y técnicas de aprendizaje automático promueve una atención de salud más inteligente, y aprovecha la gran cantidad de datos generados por dispositivos portátiles.⁽⁵⁾

En el ámbito del procesamiento de información y datos clínicos, las técnicas de aprendizaje automático han demostrado ser altamente efectivas en aplicaciones de salud. Un ejemplo destacado resulta el uso en la predicción de enfermedades y

personalización de tratamientos médicos. Además, en otros campos como el procesamiento de información, el sistema Vega-MT, desarrollado por la JD Explore Academy, ha alcanzado un rendimiento sobresaliente en tareas complejas. Esto se evidencia en la versatilidad y capacidad de las técnicas de *Machine Learning* para adaptarse a diversos contextos tecnológicos.⁽⁶⁾

En el estudio de datos de salud, la presencia de valores faltantes representa un desafío constante. Sin embargo, recientes avances en algoritmos de aprendizaje, como los modelos de imputación basados en redes neuronales, han demostrado efectividad en el manejo de datos incompletos, lo que es crucial para el análisis de supervivencia y la predicción de resultados clínicos. Estos avances destacan la capacidad de adaptación y la versatilidad de las técnicas de *Machine Learning* para mejorar la gestión del riesgo en salud.⁽⁷⁾

En conclusión, la integración de *Machine Learning* y *Big Data* en la gestión del riesgo en salud representa un cambio paradigmático en la atención médica, lo que permite una mayor precisión en la predicción y prevención de enfermedades. A medida que estas tecnologías evolucionan, el impacto en la personalización del tratamiento y en la toma de decisiones clínicas es cada vez más evidente. Sin embargo, resulta crucial abordar los desafíos éticos y técnicos asociados con su implementación para garantizar que estas herramientas beneficien de manera equitativa y segura a todos los pacientes.

El presente estudio busca destacar estos avances, al tiempo que subraya la importancia de una colaboración interdisciplinaria para maximizar el potencial de estas innovaciones en el ámbito de la salud. Para esto se presentó como objetivo realizar una revisión bibliográfica sobre la utilización de *Machine Learning* y *Big Data* como herramientas clave para mejorar la gestión del riesgo y la personalización de tratamientos.

Métodos

Se realizó una revisión meticulosa y estructurada, diseñada para evaluar y sintetizar de manera eficiente la literatura existente en el campo de la ciencia de datos y la inteligencia artificial aplicada a la gestión del riesgo en salud. El enfoque se alinea con las directrices del Marco de Revisión Sistemática y Metaanálisis (PRISMA), para garantizar así un proceso transparente y riguroso.

La metodología implementada abarcó una estrategia de búsqueda exhaustiva en bases de datos académicas y científicas de renombre, con el objetivo de identificar estudios relevantes que exploraran la aplicación de técnicas de aprendizaje automático y IA en el ámbito de la salud. Se establecieron criterios de inclusión y exclusión claros para asegurar la relevancia y calidad de los estudios seleccionados.

Se incluyó un cuadro resumen, que detalló los métodos utilizados en los artículos seleccionados, el alcance, los datos empleados en las metodologías de aprendizaje automático y las recomendaciones. Este cuadro proporcionó una visión concisa y estructurada de los aspectos clave y los hallazgos principales, lo que facilitó una comprensión rápida y completa de la investigación.

Diseño de la revisión

Esta investigación se centró en una revisión bibliográfica exhaustiva y sistemática, siguiendo las directrices del Marco de Revisión Sistemática y Metaanálisis (PRISMA) para garantizar un enfoque riguroso y transparente en cada etapa del proceso.

Estrategia de búsqueda

La estrategia de búsqueda fue exhaustiva y se centró en la identificación de artículos relevantes en bases de datos académicas y científicas reconocidas como PubMed, IEEE, ScienceDirect, Arxiv, Biometrika, PLoS ONE, Health Information Science and Systems, Nature Medicine, Nature y Google Scholar. Se usaron palabras clave y combinaciones de términos en inglés y español, relacionados con la ciencia de datos, el aprendizaje automático, la inteligencia artificial, la gestión del riesgo en salud y los perfiles epidemiológicos.

Además, se buscaron y seleccionaron revisiones sistemáticas, estudios cuantitativos y cualitativos, y casos de éxito en la aplicación de estas tecnologías en el ámbito de la salud.

Selección de estudios

Los estudios se seleccionaron de acuerdo con los siguientes criterios:

- Inclusión
 - Estudios publicados en los últimos 20 años
 - Artículos que abordaran directamente la aplicación de la ciencia de datos, el aprendizaje automático o la inteligencia artificial en la gestión del riesgo en salud
 - Estudios publicados en inglés o español
- Exclusión
 - Estudios que no abordaran directamente la gestión del riesgo en salud con la ayuda de la ciencia de datos, *Machine Learning* o inteligencia artificial
 - Artículos sin una metodología clara o sin resultados relevantes
 - Revisiones de literatura, opiniones, comentarios y cartas al editor

Proceso de revisión y extracción de datos

El proceso de revisión siguió dos etapas. En la primera se examinaron los títulos y resúmenes de los artículos recuperados para identificar estudios potencialmente relevantes. En la segunda etapa se leyeron los textos completos de estos artículos para determinar su elegibilidad final.

Los datos se extrajeron de los artículos seleccionados e incluyeron información como autores, año de publicación, objetivo del estudio, metodología empleada, resultados principales y limitaciones del estudio. Esta información se organizó en una matriz de análisis.

Los datos extraídos se analizaron de manera crítica para identificar y categorizar los hallazgos clave, patrones emergentes y vacíos en la literatura existente. Se realizó una síntesis narrativa para proporcionar un resumen comprensivo de los resultados, y se identificaron las mejores prácticas y lecciones aprendidas que podrían contribuir al éxito del proyecto propuesto.

Evaluación y calidad de los artículos

Se utilizaron herramientas de evaluación de la calidad reconocidas para evaluar la calidad metodológica de los estudios seleccionados. Esta evaluación permitió el análisis crítico de los estudios y contribuyó a la interpretación de los resultados de la revisión.

Búsqueda y selección de artículos

Siguiendo la metodología establecida, se procedió a la búsqueda y selección de artículos académicos publicados en inglés o español en los últimos 20 años. Los criterios de inclusión fueron aquellos que abordaron directamente el tema de interés. Se excluyeron aquellos estudios con metodología no clara, sin resultados relevantes.

Análisis e integración de la información

En los últimos años, la disponibilidad masiva de datos y los avances en las tecnologías de aprendizaje automático han propiciado un cambio en el enfoque de la investigación en salud, desde la perspectiva tradicional de la epidemiología hacia una metodología basada en el aprendizaje automático (ML).^(8,9) Este cambio ha sido documentado en numerosos estudios que muestran cómo el ML ha permitido desarrollar modelos predictivos más precisos y eficientes en comparación con los enfoques convencionales.

El impacto del ML en la atención médica ha sido potenciado por la acelerada generación de grandes volúmenes de datos de salud, lo que está redefiniendo la manera en que se brinda la atención médica. Sin embargo, el potencial del ML no se limita solo a la creación de modelos predictivos; también ofrece una notable capacidad para profundizar y enriquecer la relación entre médicos y pacientes, lo que fortalece esta interacción esencial mediante nuevas perspectivas tecnológicas.⁽¹⁰⁾

Por ejemplo, *Weng* y otros⁽¹¹⁾ demostraron que el ML puede mejorar la predicción del riesgo cardiovascular, con el uso de datos clínicos rutinarios, con modelos como la regresión logística y los bosques aleatorios, que superan a los modelos tradicionales en términos de precisión. La capacidad del ML para manejar datos complejos y heterogéneos ha sido fundamental en la predicción de complicaciones en enfermedades cardiovasculares, como se observa en estudios como el de *Galimzhanov* y otros,⁽¹²⁾ *Javaid* y otros⁽¹³⁾ y *Bansal* y otros,⁽¹⁴⁾ que han demostrado el uso efectivo del ML en la predicción de resultados clínicos después de intervenciones coronarias y otros procedimientos cardíacos.

La capacidad del ML para manejar datos complejos y heterogéneos también ha sido fundamental en la predicción de complicaciones en enfermedades crónicas. Por ejemplo, *Nicolucci* y otros⁽¹⁵⁾ desarrollaron modelos predictivos para

complicaciones de la diabetes tipo 2, y alcanzaron una alta precisión y sensibilidad en la identificación de riesgos para los pacientes. De manera similar, *Galimzhanov* y otros,⁽¹²⁾ utilizaron ML para predecir los resultados clínicos después de una intervención coronaria percutánea y mostraron que el modelo XGBoost era el más efectivo con un AUC de 0,86.

Además, el ML ha sido aplicado en la gestión de la cadena de suministro de vacunas, como lo demuestra el estudio de *Hu* y otros,⁽⁵⁾ que desarrollaron un sistema inteligente que combinaba *blockchain*, IoT y ML para mejorar la previsión de la demanda y la seguridad en la distribución de vacunas. Este sistema no solo predice la demanda con un error medio del 3 %, sino que realiza análisis de sentimientos con una precisión del 80 %, lo que destaca la versatilidad del ML en diferentes contextos de salud pública.

En el ámbito de la salud mental, *Kuhathasan* y otros⁽⁴⁾ utilizaron ML para analizar cómo el uso agudo de cannabis afecta los síntomas percibidos en diferentes perfiles de usuarios, y encontraron asociaciones claras entre la severidad previa de los síntomas y el cambio percibido, lo que sugiere la necesidad de estudios controlados adicionales.

Finalmente, la revisión bibliográfica agrupa los artículos revisados por año y tema de investigación, lo que resalta el creciente impacto del ML en diversas áreas de la salud:

- 2016-2017: con enfoque en enfermedades específicas. Durante este período, aproximadamente un 20 % de los estudios se centraron en enfermedades específicas como la sepsis y el riesgo cardiovascular, y utilizaron datos clínicos y registros electrónicos de salud para mejorar la precisión en el diagnóstico y la predicción de enfermedades.^(16,17)
- 2018-2019: avances en la detección de cáncer. Cerca del 30 % de los estudios durante estos años se dedicaron a la detección y clasificación de cáncer, especialmente de piel y pulmón, y utilizaron redes neuronales profundas para mejorar la precisión diagnóstica.^(10,18)
- 2020-2023: ampliando el alcance. En este periodo, aproximadamente el 50 % de los estudios expandieron el uso del ML a una gama más amplia de enfermedades, incluyendo la diabetes, las enfermedades cardiovasculares y varios tipos de cáncer, con un enfoque en la predicción de riesgos y resultados clínicos.^(4,14,15,19,20,21)

Estas aplicaciones y sus resultados favorables subrayan la notable utilidad del *Machine Learning* en el campo de la salud y la medicina, lo cual abre nuevas oportunidades de investigación y tratamiento para una amplia gama de condiciones médicas. A pesar de estos avances, es crucial reconocer la importancia de la colaboración entre expertos en ML de diversas áreas para contribuir a que los modelos desarrollados sean clínicamente relevantes y validados en términos de su aplicabilidad en la atención médica y la investigación epidemiológica. Además, la ética y la privacidad de los datos en la aplicación del ML en la salud son aspectos clave que deben considerarse cuidadosamente para proteger a los pacientes y garantizar que se respeten sus derechos.

La tabla resume algunos de los estudios más recientes en los que se ha aplicado ML en diferentes áreas de la salud. Se presentan las metodologías utilizadas en cada estudio, junto con un resumen de los principales alcances y las recomendaciones derivadas de cada investigación. Esta tabla sirve como una referencia consolidada para comprender el impacto y las aplicaciones de ML en la gestión del riesgo en salud.

Tabla - Aplicaciones de *Machine Learning* en la predicción y gestión del riesgo en salud: resumen de estudios recientes

Título-referencia	Metodología/Modelo de <i>Machine Learning</i>	Variables usadas	Alcance	Recomendaciones
<i>Dermatologist-level classification of skin cancer with deep neural networks</i> ⁽²²⁾	Redes neuronales profundas para clasificación de cáncer de piel	Datos clínicos e imágenes médicas	Clasificación de cáncer de piel al nivel de un dermatólogo	No especifica
<i>End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography</i> ⁽¹⁸⁾	Redes neuronales profundas para detección de cáncer de pulmón	Datos clínicos e imágenes médicas	Detección de cáncer de pulmón con alta sensibilidad y especificidad	Propone explorar más técnicas de modelado para evaluar malignidades y recomienda compartir datos y códigos en repositorios públicos para trabajos que usen algoritmos

				o software propios
<i>A Deep Learning Approach to on-Node Sensor Data Analytics for Mobile or Wearable Devices</i> ⁽⁹⁾	Aprendizaje profundo en dispositivos móviles para análisis de datos de sensores, series temporales, métodos de Fourier	Datos de sensores en dispositivos móviles y portátiles	Evalúa la efectividad del sistema propuesto para reconocimiento de actividades humanas en tiempo real, mediante validación cruzada en cinco conjuntos de datos	Destaca los retos en la elección de características para análisis de series temporales y la capacidad de generalización del método propuesto
<i>Can machine-learning improve cardiovascular risk prediction using routine clinical data</i> ⁽¹¹⁾	Predicción de riesgo cardiovascular con <i>Machine Learning</i> : regresión logística, bosques aleatorios, <i>gradient boosting</i> , y redes neuronales	Datos clínicos de rutina, registros electrónicos de salud	Mejora en la predicción de riesgos cardiovasculares en comparación con los modelos convencionales	El estudio insta a explorar más el aprendizaje automático con diversos grandes conjuntos de datos clínicos y en la predicción de otros resultados de enfermedades. Se requerirá investigar la viabilidad y aceptación de estas aplicaciones en la práctica clínica.
<i>Prediction of In-hospital Mortality in Emergency Department Patients with Sepsis: A Local Big Data-Driven, Machine Learning Approach</i> ⁽¹⁶⁾	Predicción de mortalidad en pacientes con sepsis mediante <i>Machine Learning</i>	Datos locales de pacientes con sepsis	Precisión en la predicción de mortalidad en pacientes con sepsis	No especifica

<i>Machine Learning in Medicine</i> ⁽¹⁰⁾	Discute conceptos generales de <i>Machine Learning</i> , incluyendo aprendizaje supervisado	Datos de registros médicos electrónicos, imágenes médicas, datos de sensores y datos genómicos	Discute el potencial del <i>Machine Learning</i> para mejorar la atención médica	Integrar la retroalimentación de los médicos para refinar los modelos de <i>Machine Learning</i>
<i>The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts</i> ⁽¹⁷⁾	No especifica	No especifica	No especifica	Establecer marcos regulatorios y éticos para garantizar la privacidad de los datos y la responsabilidad
<i>Toward advanced diagnosis and management of inherited arrhythmia syndromes: Harnessing the capabilities of artificial intelligence and machine learning</i> ⁽²³⁾	Inteligencia artificial y aprendizaje automático	Datos clínicos: datos de electrocardiograma (ECG). Datos genéticos y no genéticos Datos de riesgo establecidos y no establecidos Datos de resultados de salud y respuesta a la terapia Bases de datos de investigación y biobancos	Mejorar el diagnóstico y manejo de los síndromes de arritmia hereditaria (SAH)	Uso de modelos de IA en la práctica clínica: para identificar pacientes con enfermedades ocultas/subclínicas y mejorar la estratificación de riesgos y la predicción de efectos secundarios y resultados de medicamentos
<i>A machine learning approach for obesity risk prediction</i> ⁽²⁴⁾	k-NN, bosque aleatorio, regresión logística, perceptrón multicapa (MLP), máquina de vectores de soporte (SVM), Bayes ingenuo, <i>boosting</i> adaptativo (ADA <i>boosting</i>), árbol de decisión y clasificador de	Datos de más de 1100 personas de diferentes edades, incluyendo información sobre personas con y sin obesidad. Los datos incluyen variables como altura, peso, género, edad, dieta, actividad física, entre otros	El artículo aborda el problema creciente de la obesidad en Bangladesh y en todo el mundo. Se centra en desarrollar un modelo basado en aprendizaje automático para predecir el riesgo de obesidad	Ampliar la investigación con un conjunto de datos más grande para abarcar un rango más amplio de personas con obesidad baja, media y alta en Bangladesh

	<i>boosting de gradiente</i>			
<i>A machine learning approach for risk factors analysis and survival prediction of Heart Failure patients⁽²⁵⁾</i>	Árbol de Decisión (DT), Regresor de Árbol de Decisión (DTR), Bosque Aleatorio (RF), XGBoost y Algoritmos de <i>boosting</i> de gradiente (GB). El estudio se centró en comparar el rendimiento de estos algoritmos en términos de precisión, <i>recall</i> , medida F y valor de pérdida logarítmica, y encontró que RF proporciona la mayor precisión con un 97,78 %	Se utilizó un conjunto de datos de supervivencia de HF que incluye 299 casos, con 96 resultados de muerte y el resto de supervivientes. Los datos incluyen variables como edad, anemia, nivel de CPK en sangre, diabetes, fracción de eyección (EF), presión arterial alta, plaquetas, creatinina sérica (SC), nivel de sodio en sangre (SS), sexo, tabaquismo y tiempo de seguimiento	Propone el uso de métodos de aprendizaje automático (ML) para analizar factores de riesgo y predecir la supervivencia de pacientes con insuficiencia cardíaca	Ampliar la investigación con un conjunto de datos más grande y emplear tecnologías más actualizadas para construir modelos similares
<i>A breast cancer risk predication and classification model with ensemble learning and big data fusion⁽¹⁹⁾</i>	Técnica mejorada de XGBoost con análisis de grandes datos (I-XGBoost)	Se empleó el conjunto de datos de diagnóstico de cáncer de mama de Wisconsin, que incluye características como edad, anemia, nivel de CPK en sangre, diabetes, fracción de eyección (EF), presión arterial alta, plaquetas, creatinina sérica (SC), nivel de sodio en sangre (SS), sexo, tabaquismo y tiempo de seguimiento	Predicción y clasificación del riesgo de cáncer de mama	Implementar el modelo propuesto en un conjunto de datos en vivo y comparar los resultados con más métodos de aprendizaje automático y aprendizaje profundo

<i>An examination of the hybrid meta-heuristic machine learning algorithms for early diagnosis of type II diabetes using big data feature selection⁽²¹⁾</i>	Algoritmo de optimización por enjambre de partículas (PSO) para reducir cálculos. Algoritmos jerárquicos para optimizar hiperparámetros. El Algoritmo Genético-Máquina de Soporte Vectorial (GA-SVM). Árbol de Decisión (DT), K-Vecinos Más Cercanos (KNN), Redes Neuronales de Múltiples Capas (MLP)	17 características para 520 casos	Algoritmos híbridos para diagnosticar T2DM en etapas tempranas, para evitar pruebas médicas costosas y prolongadas	Enfoque en la diabetes gestacional y estudios longitudinales de pacientes diabéticos para identificar patrones de complicaciones
<i>Machine learning versus regression for prediction of sporadic pancreatic cancer⁽²⁶⁾</i>	<i>Random Survival Forest (RSF), eXtreme Gradient Boosting (XGB), Cox Proportional Hazards Regression (COX)</i>	Edad, dolor abdominal, cambio de peso, alanina aminotransferasa (ALT), HbA1c	Comparación de modelos de ML y regresión para predecir el cáncer de páncreas esporádico. RSF y XGB mostraron mejor discriminación que COX	Se recomienda el uso combinado de los tres modelos para mejorar la sensibilidad de la estrategia de detección
<i>A review of risk prediction models in cardiovascular disease: conventional approach vs. artificial intelligent approach⁽²⁷⁾</i>	Regresiones lineales y logísticas, adaptadas para abordar problemas específicos	No especifica	Los enfoques convencionales involucran las guías de práctica clínica actuales para CVD, que se basan en un "paciente promedio", y utilizan modelos de riesgo como GRACE, TIMI y Framingham	Se propone un enfoque integrado de análisis <i>De Big Data</i> , incluyendo datos ómicos (genómicos, transcriptómicos, proteómicos, metabolómicos) para comprender mejor el desarrollo de CVD y permitir pasos de prevención o

				pronóstico más tempranos
<i>A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data⁽²⁸⁾</i>	Técnicas de aprendizaje automático	No especifica	Análisis comparativo detallado de diversas técnicas de aprendizaje automático (ML) para la predicción de CVD	Se necesitan conjuntos de datos más grandes y representativos para mejorar la precisión de los modelos de ML en CVD
<i>Big Data, Machine Learning, and Artificial Intelligence to Advance Cancer Care: Opportunities and Challenges⁽²⁹⁾</i>	No especifica	No especifica	Explorar cómo los avances rápidos en inteligencia artificial (AI), Big Data y Machine Learning (ML) pueden contribuir a una atención personalizada y eficiente del cáncer	-
<i>Conceptual design of a machine learning-based wearable soft sensor for non-invasive cardiovascular risk assessment⁽³⁰⁾</i>	No especifica	No especifica	No especifica	Las tecnologías de Big Data, AI y ML son clave para mejorar la atención médica, especialmente en personalizar tratamientos. Sin embargo, su adopción enfrenta desafíos como la integración en prácticas clínicas existentes, la aceptación por parte de los profesionales de la salud y

				limitaciones técnicas
<i>Development and validation of machine learning based prediction model for postoperative pain risk after extraction of impacted mandibular third molars⁽³¹⁾</i>	<i>Random Forest (RF), Linear Discriminant Analysis (LDA), Extreme Gradient Boosting (XGBoost), Support Vector Machines (SVM), Gaussian Naive Bayes (Gaussian NB), Logistic Regression (LR), Decision Tree (DT), Adaptive Boosting (AdaBoost), K-Nearest Neighbors (KNN)</i>	Se incluyeron 185 pacientes y 202 casos de datos de terceros molares mandibulares impactados. Se seleccionaron cinco variables para el modelado: calificación del médico, sensibilidad máxima al dolor autoinformada por el paciente, OHI-S-CI, IMC y presión arterial sistólica	Modelo de predicción basado en algoritmos de ML para identificar pacientes con alto riesgo de dolor posoperatorio tras la extracción de terceros molares mandibulares impactados	El modelo de bosque aleatorio mostró el mejor rendimiento
<i>Development of a machine learning-based risk prediction model for cerebral infarction and comparison with nomogram model⁽³²⁾</i>	<i>Extreme Gradient Boosting (XgBoost), Logistic Regression (LR), Support Vector Machine (SVM) y Random Forest (RF). El algoritmo XgBoost resultó el óptimo y se utilizó para desarrollar el modelo de predicción CI-Lab8</i>	Cohorte 1: pacientes con CI y controles sanos de 2017 Cohorte 2: pacientes con CI y sin CI de 2018 a 2020 Cohorte 3: pacientes con CI y sin CI de 2021 Variables clave: fibrinógeno, edad, glucosa, concentración media de hemoglobina eritrocitaria, albúmina, valor absoluto de neutrófilos, tiempo de tromboplastina parcial activada y triglicéridos	Desarrollar un modelo de predicción de riesgo para el infarto cerebral (CI), mediante algoritmos de aprendizaje automático para analizar datos de pruebas rutinarias.	Sugiere la integración de datos clínicos y genéticos para mejorar la predicción, el diagnóstico o el pronóstico de enfermedades en el futuro. Además, se destaca la importancia de utilizar tecnologías de secuenciación genética en aplicaciones clínicas para un análisis más completo y valioso de los riesgos de enfermedades

<i>Diabetes disease detection and classification on Indian demographic and health survey data using machine learning methods⁽³³⁾</i>	Regresión logística, Bayes ingenuo gaussiano, análisis discriminante lineal, clasificador de vectores de soporte, vecino más cercano (<i>k-nearest neighbor</i>), árbol de decisión, <i>boosting</i> de gradiente extremo, análisis de componentes de entropía del kernel y bosque aleatorio (<i>Random Forest</i>)	Se utilizó el conjunto de datos de la Encuesta Demográfica y de Salud de la India de 2016, que incluye datos antropométricos, bioquímicos y clínicos de mujeres entre 19 y 49 años. Se determinaron los factores de riesgo para datos continuos y categóricos	Utilizar métodos de aprendizaje automático para la detección y clasificación de la diabetes, con relevancia clínica, en la población india	El modelo de bosque aleatorio (<i>Random Forest</i>) alcanzó la máxima precisión de clasificación con un 99,84 % en conjuntos de datos desequilibrados y un 96,75 % en conjuntos de datos equilibrados (mediante la técnica de sobremuestreo de minorías sintéticas)
<i>Diabetes risk prediction model based on community follow-up data using machine learning⁽³⁴⁾</i>	Se empleó un clasificador de bosque aleatorio para el modelo de riesgo de diabetes, que seleccionó características óptimas y comparó algoritmos. Se alcanzó una precisión del 91,24 % y un Área Bajo la Curva (AUC) del 97 % en la curva ROC. Además, se diseñó una tarjeta de puntuación de riesgo de diabetes para mejorar la aplicabilidad del modelo en la vida real y clínica	Registros de seguimiento de pacientes con diabetes desde 2016 hasta 2023, incluyendo nueve características como índice de masa corporal (BMI), edad, presión arterial sistólica y diastólica, frecuencia y tiempo de ejercicio, consumo de alimentos básicos, tabaquismo y consumo de alcohol	Desarrollar un modelo de evaluación de riesgo de diabetes	Extraer más variables en investigaciones futuras y considerar el análisis de series temporales de los datos de seguimiento para mejorar la gestión de la salud de los pacientes
<i>Estimation of service length with the machine learning algorithms and neural networks for patients who</i>	Árbol de decisión, bosque aleatorio y red neuronal de perceptrón multicapa	Datos de más de 4000 pacientes que reciben servicios de atención médica domiciliaria	Modelo de estimación que utilizó algoritmos de aprendizaje automático para planificar de manera efectiva y eficiente los	El estudio reconoce la importancia de la calidad de los datos para los algoritmos de aprendizaje automático, ya que los errores,

<i>receiving home health care</i> ⁽³⁵⁾			servicios de atención médica domiciliaria en hospitales	omisiones o etiquetados incorrectos en los datos pueden afectar los resultados. Se encontró que la duración del servicio de atención médica domiciliaria que recibieron los pacientes variaba según su edad y género
<i>From machine learning to deep learning: An advances of the recent data-driven paradigm shift in medicine and healthcare</i> ⁽³⁶⁾	Diferentes modelos de <i>Machine Learning</i> y <i>Deep Learning</i> (ML y DL) aplicados en medicina	Datos clínicos, imágenes médicas, datos de EMR/EHR, <i>Big Data</i> , datos de biosensores, datos genómicos, etc.	Aplicaciones de ML y DL en diagnóstico, tratamiento, análisis de imágenes médicas, medicina personalizada, y más	Se recomienda una mayor investigación en la implementación segura de ML y DL en medicina, incluyendo consideraciones éticas
<i>Machine learning-based paradigm for prediction of neonatal and infant risk of death</i> ⁽³⁷⁾	<i>AdaBoost</i> (ADB), <i>eXtreme Gradient Boosting</i> (XGB), <i>Random Forest</i> (RF), <i>Logistic Regression</i> (LR)	Edad materna, raza materna, edad gestacional, peso al nacer, condición del recién nacido, entre otros	Predicción del riesgo de muerte neonatal e infantil mediante datos de 15,8 millones de nacimientos. Se observó una mejora en la precisión y sensibilidad al balancear los datos con ADASYN	Se recomienda continuar explorando la selección de características y la implementación de técnicas de balanceo de datos para mejorar la predicción
<i>Machine learning-based assessment of preclinical health questionnaires</i> ⁽³⁸⁾	<i>Hopfield neural network</i> (<i>Hopfield detect</i>)	Respuestas de cuestionarios sobre condiciones de salud preclínicas, con enfoque en fumadores embarazadas	Evaluación del estado de salud de mujeres embarazadas fumadoras basado en respuestas de cuestionarios	Se recomienda mejorar la precisión de los datos recolectados y explorar más aplicaciones del modelo en otros

				contextos médicos
<i>Machine learning-based cardiovascular disease prediction⁽³⁹⁾</i>	<i>Logistic Regression, K-Nearest Neighbor (K-NN), Decision Tree, XGBoost (Ensemble classifier)</i>	Edad, sexo, presión arterial, nivel de azúcar, colesterol, entre otros factores de riesgo cardíaco	Predicción de enfermedades cardiovasculares mediante múltiples algoritmos de ML. El modelo XGBoost mostró la mayor precisión	Se recomienda seguir explorando diferentes algoritmos para mejorar la precisión en la detección temprana de enfermedades cardiovasculares
<i>Machine Learning-Based Prediction of Atrial Fibrillation Risk Using Electronic Medical Records in Older Aged Patients⁽⁴⁰⁾</i>	<i>Decision Tree, Support Vector Machine (SVM), Logistic Regression, Random Forest</i>	Edad, género, datos de laboratorio, códigos ICD-9-CM, prescripción de medicamentos	Predicción del riesgo de fibrilación auricular de nuevo inicio en un año, mediante registros médicos electrónicos de pacientes mayores. El modelo <i>Random Forest</i> mostró la mejor precisión con un AUROC de 0,74	Se recomienda implementar el modelo en la práctica clínica diaria para mejorar la identificación y manejo de pacientes en riesgo de fibrilación auricular
<i>Machine learning in cardiology: Clinical application and basic research⁽⁴¹⁾</i>	<i>Supervised learning, Unsupervised learning, Reinforcement learning, Deep learning</i>	Datos de registros médicos electrónicos (EHR/EMR), ECG, imágenes ecocardiográficas, radiografías, MRI, datos genéticos, datos de secuenciación de RNA, entre otros	Aplicación del aprendizaje automático en la clasificación de enfermedades, diagnóstico, predicción de pronóstico, análisis de imágenes y estudios genéticos en cardiología y otras áreas	Se recomienda la combinación de <i>Machine Learning</i> con nuevos avances tecnológicos para mejorar la precisión diagnóstica y desarrollar nuevas estrategias terapéuticas en cardiología

<i>Machine Learning Models and Big Data Tools for Evaluating Kidney Acceptance⁽⁴²⁾</i>	<i>Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, Multilayer Perceptron, Gradient-Boosted Trees</i>	Datos de registros de donantes y candidatos a trasplante de riñón (UNOS), características del donante, características del candidato, historial de decisiones de aceptación o rechazo	Modelos de ML para predecir la aceptación de riñones en un contexto de trasplantes, para optimizar la toma de decisiones y reduciendo el descarte de riñones	Se recomienda la implementación de los modelos en OPOs para priorizar la aceptación de riñones basándose en la probabilidad de aceptación y la validación de los resultados por parte de los profesionales de salud antes de su uso clínico
<i>Machine learning vs. conventional methods for prediction of 30-day readmission following percutaneous mitral edge-to-edge repair⁽⁴³⁾</i>	<i>Logistic Regression (ML-LR), Naïve Bayes (NB), Random Forest (RF), Artificial Neural Network (ANN)</i>	Variables demográficas, características del hospital, comorbilidades cardiovasculares y no cardiovasculares, factores relacionados con el procedimiento y características del hospital	Comparación entre métodos convencionales y ML para predecir la readmisión a 30 días después de una reparación percutánea de borde a borde mitral. Los modelos ML, especialmente NB, mostraron un mejor rendimiento predictivo que los métodos convencionales	Se recomienda la implementación de modelos de ML, como Naïve Bayes, en la práctica clínica para mejorar la predicción de readmisiones y, potencialmente, optimizar los resultados clínicos en pacientes sometidos a reparación mitral percutánea
<i>Management of medical and health big data based on integrated learning-based health care system: A review and comparative análisis⁽⁴⁴⁾</i>	Método de aprendizaje integrado Bagging, Máquina de Aprendizaje Extremo (ELM)	Grandes datos de registros médicos y de salud, incluyendo varios parámetros relacionados con la salud del paciente y el rendimiento del sistema	El estudio presenta un sistema de atención médica (HCS) que integra grandes datos médicos y de salud, mediante un enfoque de aprendizaje integrado, lo que resulta en un mayor volumen de integración,	Se recomienda continuar desarrollando métodos de aprendizaje integrado para manejar de manera más eficiente los grandes datos médicos, especialmente a medida que los datos de salud se vuelven más

			eficiencia y capacidad de almacenamiento en comparación con los sistemas basados en internet	complejos con avances como la tecnología 5G
<i>Medicine 2032: The future of cardiovascular disease prevention with machine learning and digital health technology⁽¹³⁾</i>	Aprendizaje supervisado, aprendizaje no supervisado, aprendizaje profundo, redes neuronales artificiales	Datos de registros médicos electrónicos (EHR), imágenes cardiovasculares, aplicaciones móviles, wearables, sensores, determinantes sociales de la salud	El artículo revisa los avances clave en la aplicación del aprendizaje automático para la prevención de enfermedades cardiovasculares, y destaca cómo pueden impactar la práctica clínica, incluyendo evaluaciones de riesgo personalizadas y manejo médico guiado por directrices	Se recomienda equilibrar la aplicación del aprendizaje automático con consideraciones sobre la privacidad del paciente, la equidad en salud y la interpretabilidad clínica para maximizar su impacto positivo en la atención médica
<i>New perspectives on cancer clinical research in the era of big data and machine learning⁽⁷⁾</i>	Regresión logística, Máquina de soporte vectorial (SVM), Árboles de clasificación y regresión (CART), XGBoost (XGB), bosques aleatorios (RF), K-Nearest Neighbor (KNN)	Datos epidemiológicos de cáncer, información clínica de bases de datos como SEER	El artículo revisa investigaciones recientes basadas en la base de datos SEER y <i>Machine Learning</i> , enfocándose en la predicción de metástasis ganglionar, metástasis a distancia y pronóstico, lo que muestra que los modelos de <i>Machine Learning</i> tienen ventajas sobre los modelos tradicionales en	Se recomienda continuar explorando el uso de algoritmos de <i>Machine Learning</i> en la investigación del cáncer, con especial atención a los desafíos de sobreajuste e interpretabilidad para maximizar su aplicabilidad clínica

			cuanto a adaptabilidad	
<i>Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods⁽⁴⁵⁾</i>	Redes neuronales profundas, algoritmo genético, máquina de soporte vectorial (SVM)	Electrocardiogramas (ECG), fonocardiogramas (PCG), características clínicas de los pacientes	El artículo propone un método novedoso basado en la integración de señales de ECG y PCG para predecir enfermedades cardiovasculares. El modelo muestra un desempeño superior en comparación con los métodos de un solo modelo, y alcanzó un AUC de 0,936	Se recomienda el uso de métodos multi-modales en la predicción de enfermedades cardiovasculares, y se sugiere la recolección de más datos valiosos para mejorar y optimizar el modelo propuesto
<i>Prediction of clinical outcomes after percutaneous coronary intervention: Machine-learning analysis of the National Inpatient Sample⁽¹²⁾</i>	Regresión logística, Máquina de soporte vectorial (SVM), Naive Bayes, bosques aleatorios (RF), XGBoost (XGB)	101 características clínicas, incluyendo edad, sexo, factores de riesgo, comorbilidades y datos del procedimiento de intervención coronaria percutánea (PCI)	El estudio desarrolló un modelo de ML multicategoría para predecir la mortalidad por todas las causas, eventos isquémicos y hemorrágicos en pacientes hospitalizados que se sometieron a PCI. El modelo XGBoost mostró el mejor rendimiento con un AUC de 0,86	Se recomienda la integración de modelos de ML como XGBoost en los sistemas electrónicos de salud para mejorar la predicción de eventos adversos y apoyar la toma de decisiones clínicas en pacientes sometidos a PCI

<i>Prediction of complications of type 2 Diabetes: A Machine learning approach⁽¹⁵⁾</i>	<i>XGBoost (Gradient Boosting Machine), Oversampling con SMOTE, Validación cruzada de diez pliegues</i>	Características demográficas (edad, sexo, etc.), pruebas de laboratorio (presión arterial, índice tobillo-brazo, circunferencia de cintura, entre otros), duración de la diabetes	El estudio desarrolló modelos predictivos para complicaciones de la diabetes tipo 2, incluyendo complicaciones oculares, enfermedades cardiovasculares, cerebrovasculares, nefropatía, y neuropatía diabética. Los modelos mostraron alta precisión y sensibilidad, con AUC superiores a 0,85 para todas las complicaciones	Se recomienda la implementación de estos modelos en la práctica clínica para identificar a los pacientes en mayor riesgo de complicaciones a corto y medio plazo, lo que mejora la calidad de la atención y supera la inercia clínica
<i>Predictive model for early death risk in pediatric hemophagocytic lymphohistiocytosis patients based on machine learning⁽²⁰⁾</i>	Regresión logística, XGBoost, LASSO (Least Absolute Shrinkage and Selection Operator)	Indicadores clínicos clave: tratamiento con inmunoterapia, niveles de colesterol total, bilirrubina total, nitrógeno ureico en sangre (BUN), tiempo de trombina (TT), tiempo de tromboplastina parcial activada (APTT), ratio de histiocitos en médula ósea	El estudio desarrolló y validó un modelo predictivo basado en <i>Machine Learning</i> para evaluar el riesgo de muerte temprana en pacientes pediátricos con linfohistiocitosis hemofagocítica (HLH). El modelo logró un AUC de 0,915 con la regresión logística y 0,889 con XGBoost	Se recomienda el uso del modelo predictivo como una herramienta de soporte para la estratificación del riesgo y la planificación de tratamientos personalizados en pacientes pediátricos con HLH. También se sugiere una validación adicional en otros centros médicos para aumentar la generalizabilidad del modelo

<i>Predictors of perceived symptom change with acute cannabis use for mental health conditions in a naturalistic sample: A machine learning approach⁽⁴⁾</i>	XGBoost, SHapley Additive exPlanations (SHAP)	Datos demográficos (edad, género), ratio CBD/THC, severidad previa de síntomas de ansiedad, depresión, insomnio	El estudio utilizó métodos de <i>Machine Learning</i> para investigar los predictores del cambio de síntomas percibido tras el uso agudo de cannabis en una muestra naturalista grande. Se encontraron asociaciones claras entre la severidad previa de los síntomas, la edad, el género, y el ratio CBD/THC con el cambio de síntomas percibido	Se recomienda continuar con estudios controlados con placebo para determinar la efectividad específica del cannabis en el manejo de síntomas de salud mental en diferentes perfiles de usuarios
<i>Seeing the random forest through the decision trees: Supporting learning health systems from histopathology with machine learning models: Challenges and opportunities⁽⁴⁶⁾</i>	Bosques aleatorios (<i>Random Forest</i>), árboles de decisión, redes neuronales profundas (DNN)	Imágenes histopatológicas, grandes datos de salud, características clínicas y moleculares	El artículo analiza los desafíos y las oportunidades que presenta la integración de modelos de <i>Machine Learning</i> en sistemas de salud basados en el aprendizaje (<i>Learning Health Systems</i>), enfocándose en la histopatología digital y su capacidad para extraer información oculta valiosa para la toma de decisiones clínicas	Se recomienda la adopción de enfoques colaborativos, como el aprendizaje federado, y el desarrollo de métodos innovadores para mejorar la extracción y el análisis de información relevante en imágenes de diapositivas completas. Además, se sugiere una reevaluación conceptual para diferenciar entre la información oculta y la

				relevante para el diagnóstico
<i>Small samples-oriented intrinsically explainable machine learning using Variational Bayesian Logistic Regression: An intensive care unit readmission prediction case for liver transplantation patients⁽⁴⁷⁾</i>	Regresión logística bayesiana variacional, Estrategia de estimación de características salientes (SFE), Aprendizaje secuencial	Duración total de la estancia en UCI (ICUTime), <i>N-terminal pro-brain natriuretic peptide</i> (NTproBNP), tiempo de operación, transfusión intraoperatoria de glóbulos rojos	El estudio desarrolló un modelo de ML intrínsecamente explicable para predecir la probabilidad de readmisión en UCI de pacientes trasplantados de hígado. El modelo OVBLR-SFE logró una precisión del 88,11 % en la predicción de readmisiones, y los factores clínicos identificados como importantes coincidieron con las evaluaciones de expertos clínicos	Se recomienda la implementación de OVBLR-SFE en entornos clínicos, con validaciones adicionales en otros contextos clínicos para confirmar su aplicabilidad generalizada
<i>Understanding the bias in machine learning systems for cardiovascular disease risk assessment: The first of its kind review⁽⁴⁸⁾</i>	Regresión logística, Máquina de soporte vectorial (SVM), bosques aleatorios (RF), Redes neuronales profundas (DNN)	Factores de riesgo demográficos (edad, sexo, índice de masa corporal, etc.), biomarcadores de laboratorio (LBBM), fenotipos de imagen de ultrasonido carotídeo (CUSIP), datos de uso de medicación (MedUSE)	El estudio revisa el sesgo en los sistemas de <i>Machine Learning</i> utilizados para la evaluación del riesgo de enfermedades cardiovasculares, mediante un enfoque sistemático basado en 117 estudios. Se identificaron cinco tipos de estándares de oro en el diseño	Se recomienda mejorar el diseño de los resultados (<i>outcomes</i>) en los sistemas de ML para la evaluación del riesgo cardiovascular, lo que asegura la validación científica y clínica, y la adaptación a grupos multiétnicos en la práctica de IA no vista (<i>unseen AI</i>).

			de ML para la predicción del riesgo de enfermedades coronarias (CAD) y enfermedades cardiovasculares (CVD)	
<i>Use of Machine Learning to Develop Prediction Models for Mortality and Stroke in Patients Undergoing Balloon Aortic Valvuloplasty⁽¹⁴⁾</i>	<i>Machine Learning to Develop Prediction Models for Mortality and Stroke in Patients Undergoing Balloon Aortic Valvuloplasty</i> Regresión logística, bosques aleatorios (Random Forest), LASSO (Least Absolute Shrinkage and Selection Operator)	46 variables candidatas que incluyen características demográficas, comorbilidades basales, características del procedimiento y complicaciones clínicas (choque cardiogénico, soporte circulatorio mecánico, complicaciones vasculares, etc.)	El estudio desarrolló modelos predictivos para mortalidad y accidentes cerebrovasculares en pacientes sometidos a valvuloplastia aórtica con balón (BAV). Los bosques aleatorios mostraron un desempeño superior en la predicción de accidentes cerebrovasculares, mientras que ambos modelos (logística y RF) mostraron un rendimiento comparable en la predicción de la mortalidad	Se recomienda la implementación de los modelos en la práctica clínica para mejorar la estratificación de riesgos y apoyar la toma de decisiones en pacientes sometidos a BAV, especialmente en aquellos con alto riesgo de complicaciones
<i>Using big data searching and machine learning to predict human health risk probability from pesticide site soils in China⁽⁴⁹⁾</i>	Aprendizaje semisupervisado, Algoritmo de propagación de etiquetas (Label Propagation Algorithm, LPA)	Probabilidad de contaminación del suelo, proporciones de la población mayor de 60 y menor de 15 años, contenido de arena en el suelo (30-100 cm), elevación, proporción de tierra cultivable dentro de 1 km del sitio	El estudio desarrolló un modelo de predicción de riesgo para la salud humana (HHR) en sitios pesticidas en China, que utilizó aprendizaje semi-supervisado. El modelo LPA	Se recomienda utilizar el modelo para la gestión y el control del riesgo de salud en sitios contaminados por pesticidas, que priorice la evaluación de los sitios no investigados que presentaron un

			mostró una precisión del 93 % con una tasa de <i>recall</i> del 100 %	alto riesgo predicho
<i>Using machine learning to predict cardiovascular risk using self-reported questionnaires: A right, but a long way to go</i> ⁽⁵⁰⁾	Regresión logística, árboles de decisión, bosques aleatorios (Random Forest), redes neuronales profundas (DNN)	Cuestionarios autoinformados sobre factores de riesgo como hipertensión, colesterol, tabaquismo, diabetes, y uso de medicamentos. No se utilizaron datos de presión arterial o de laboratorio	El estudio presenta un modelo de <i>Machine Learning</i> para predecir el riesgo de mortalidad cardiovascular y hospitalización por cardiopatía isquémica basado únicamente en cuestionarios autoinformados. Aunque omite la necesidad de datos de laboratorio, el modelo demostró un rendimiento similar al de estudios anteriores que sí incluyeron estos datos	Se recomienda llevar a cabo ensayos clínicos controlados para comparar la aceptación general y el rendimiento de esta nueva herramienta en comparación con las herramientas de predicción tradicionales que requieren datos clínicos básicos
<i>Vaccine supply chain management: An intelligent system utilizing blockchain, IoT and machine learning</i> ⁽⁵⁾	Aprendizaje profundo, GRU (<i>Gated Recurrent Unit</i>), BiLSTM (<i>Bidirectional Long Short-Term Memory</i>), análisis de sentimientos	Datos de <i>blockchain</i> , IoT, revisiones de vacunas, predicciones de demanda de vacuna	El estudio desarrolló un sistema inteligente para la gestión de la cadena de suministro de vacunas (VSC) que combina <i>blockchain</i> , IoT y <i>Machine Learning</i> para mejorar la calidad de las vacunas, la previsión de la demanda y la	Se recomienda la implementación de este sistema en la gestión de la cadena de suministro de vacunas para garantizar la seguridad, la eficiencia y la transparencia en la distribución de vacunas, especialmente en el contexto de la

			confianza entre los actores de la cadena de suministro. Se demostró que el sistema puede predecir la demanda de vacunas con un error medio del 3 % y realizar análisis de sentimientos con una precisión del 80 %	pandemia de COVID-19
--	--	--	---	----------------------

La adopción y el interés creciente en las técnicas de *Machine Learning* (ML) tienen el potencial de transformar significativamente la salud, particularmente en aspectos como la prevención, el diagnóstico y el tratamiento de enfermedades, así como en la gestión de riesgos y la emisión de alertas tempranas.

El enfoque de la investigación en salud ha evolucionado hacia una perspectiva centrada en ML, lo que ha catalizado importantes avances en la predicción, la prevención y el manejo de enfermedades. Sin embargo, la efectividad de estos modelos de ML es fuertemente dependiente de la calidad y cantidad de los datos utilizados para su entrenamiento y validación. Por ello, es crucial garantizar que los datos empleados en la investigación y desarrollo de aplicaciones de ML en salud sean precisos, actualizados y representativos.

Asimismo, resulta fundamental continuar trabajando en estrecha colaboración con expertos en epidemiología, estadística, medicina y demás áreas para mejorar la relevancia clínica y la validez de estos modelos. Además, con la constante evolución y mejora de las técnicas de ML, su impacto en el ámbito de la salud probablemente seguirá creciendo. Para que estas innovaciones sean efectivamente beneficiosas y adoptadas, tiene importancia fomentar la colaboración interdisciplinaria entre expertos en ML y profesionales de la salud, así como la comunidad.

Por otro lado, no se pueden ignorar las implicaciones éticas y legales asociadas con el uso de datos de pacientes en la investigación y el desarrollo de modelos de ML. Los investigadores y desarrolladores deben estar al tanto de las regulaciones existentes en torno a la privacidad y protección de datos, y abordar los desafíos

específicos que pueden surgir en el contexto de la atención médica y la investigación en salud.

Conclusiones

Machine Learning es una herramienta poderosa pero no infalible. A través de la colaboración interdisciplinaria, la atención a las consideraciones éticas y legales, y el uso de datos de alta calidad, el ML puede continuar siendo un recurso valioso y efectivo en el campo de la salud.

Referencias bibliográficas

1. Fieggen J, Smith E, Arora L, Segal B. The role of machine learning in HIV risk prediction. *Front. Reprod. Health.* 2022;4. DOI: <https://doi.org/10.3389/frph.2022.1062387>
2. Austin C, Kusumoto F. The application of Big Data in medicine: current implications and future directions. *Journal of Interventional Cardiac Electrophysiology.* 2016;47(1):51-9. DOI: <https://doi.org/10.1007/s10840-016-0104-y>
3. Hamilton AJ, Strauss AT, Martinez DA, Hinson JS, Levin S, Lin G, *et al.* Machine learning and artificial intelligence: applications in healthcare epidemiology. *Antimicrobial Stewardship & Healthcare Epidemiology.* 2021 [acceso 28/01/2022];1(1):e28. Disponible en: <https://www.cambridge.org/core/product/60FFBF2D360CF70446AC5B4A972B299A>
4. Kuhathasan N, Ballester PL, Minuzzi L, MacKillop J, Frey BN. Predictors of perceived symptom change with acute cannabis use for mental health conditions in a naturalistic sample: A machine learning approach. *Compr Psychiatry.* 2023 [acceso 22/02/2024];122:152377. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0010440X23000147>
5. Hu H, Xu J, Liu M, Lim MK. Vaccine supply chain management: An intelligent system utilizing blockchain, IoT and machine learning. *J Bus Res.* 2023 [acceso

22/02/2024];156:113480. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S0148296322009456>

6. Zan C, Peng K, Ding L, Qiu B, Liu B, He S, et al. Vega-MT: The JD Explore Academy Translation System for WMT22. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics; 2022 [acceso 22/02/2024]. Disponible en: <https://github.com/JDEA-NLP/Vega-MT>

7. Li S, Yi H, Leng Q, Wu Y, Mao Y. New perspectives on cancer clinical research in the era of big data and machine learning. Surg Oncol. 2023 [acceso 22/02/2024];102009. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S0960740423001093>

8. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. JAMA. 2018;319(13):1317-8. DOI: <https://doi.org/10.1001/jama.2017.18391>

9. Ravi D, Wong C, Lo B, Yang GZ. A Deep Learning Approach to on-Node Sensor Data Analytics for Mobile or Wearable Devices. IEEE J Biomed Health Inform. 2016;21(1):56-64. DOI: <https://doi.org/0.1109/JBHI.2016.2633287>

10. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. New England Journal of Medicine. 2019 [acceso 22/02/2022];380(14):1347-58. Disponible en: <https://www.nejm.org/doi/abs/10.1056/NEJMr1814259>

11. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One. 2017;12(4):e0174944. DOI: <https://doi.org/10.1371/journal.pone.0174944>

12. Galimzhanov A, Matetic A, Tenekecioglu E, Mamas MA. Prediction of clinical outcomes after percutaneous coronary intervention: Machine-learning analysis of the National Inpatient Sample. Int J Cardiol. 2023 [acceso 22/02/2024];392:131339. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S0167527323012962>

13. Javaid A, Zghyer F, Kim C, Spaulding EM, Isakadze N, Ding J, et al. Medicine 2032: The future of cardiovascular disease prevention with machine learning and digital health technology. Am J Prev Cardiol. 2022 [acceso 22/02/2023];12:100379. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S2666667722000630>

14. Bansal A, Kumar A, Garg C, Kalra A, Puri R, Kapadia SR, et al. Use of Machine Learning to Develop Prediction Models for Mortality and Stroke in Patients Undergoing Balloon Aortic Valvuloplasty. Cardiovascular Revascularization

Medicine. 2022 [acceso 22/02/2024];45:26-34. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1553838922006959>

15. Nicolucci A, Romeo L, Bernardini M, Vespasiani M, Rossi MC, Petrelli M, *et al.* Prediction of complications of type 2 Diabetes: A Machine learning approach. *Diabetes Res Clin Pract.* 2022 [acceso 22/02/2024];190:110013. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0168822722008270>

16. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, *et al.* Prediction of In-hospital Mortality in Emergency Department Patients with Sepsis: A Local Big Data-Driven, Machine Learning Approach. *Academic Emergency Medicine.* 2016;23(3):269-78. DOI: <https://doi.org/10.1111/acem.12876>

17. Mittelstadt BD, Floridi L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Sci Eng Ethics.* 2016;22(2):303-41. DOI: <https://doi.org/10.1007/s11948-015-9652-2>

18. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng LH, *et al.* End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med.* 2019 [acceso 22/02/2022];25:954-61. Disponible en: <https://api.semanticscholar.org/CorpusID:159041422>

19. Jaiswal V, Saurabh P, Lilhore UK, Pathak M, Simaiya S, Dalal S. A breast cancer risk predication and classification model with ensemble learning and big data fusion. *Decision Analytics Journal.* 2023 [acceso 22/02/2024];8:100298. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2772662223001388>

20. Xiao L, Zhang Y, Xu X, Dou Y, Guan X, Guo Y, *et al.* Predictive model for early death risk in pediatric hemophagocytic lymphohistiocytosis patients based on machine learning. *Heliyon.* 2023 [acceso 22/02/2024];9(11):e22202. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2405844023094100>

21. Navazi F, Yuan Y, Archer N. An examination of the hybrid meta-heuristic machine learning algorithms for early diagnosis of type II diabetes using big data feature selection. *Healthcare Analytics.* 2023 [acceso 22/02/2024];4:100227. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2772442523000941>

22. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115-8. DOI: <https://doi.org/10.1038/nature21056>

23. Asatryan B, Bleijendaal H, Wilde AAM. Toward advanced diagnosis and management of inherited arrhythmia syndromes: Harnessing the capabilities of artificial intelligence and machine learning. *Heart Rhythm.* 2023 [acceso

- 22/02/2024];20(10):1399-407. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S1547527123024116>
24. Ferdowsy F, Rahi KSA, Jabiullah MdI, Habib MdT. A machine learning approach for obesity risk prediction. *Current Research in Behavioral Sciences*. 2021 [acceso 22/02/2022];2:100053. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S2666518221000401>
25. Ali MdM, Al-Doori VS, Mirzah N, Hemu AA, Mahmud I, Azam S, *et al.* A machine learning approach for risk factors analysis and survival prediction of Heart Failure patients. *Healthcare Analytics*. 2023 [acceso 22/02/2024];3:100182. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S2772442523000497>
26. Chen W, Zhou B, Jeon CY, Xie F, Lin YC, Butler RK, *et al.* Machine learning versus regression for prediction of sporadic pancreatic cancer. *Pancreatology*. 2023 [acceso 22/02/2024];23(4):396-402. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S1424390323001035>
27. Mohd Faizal AS, Thevarajah TM, Khor SM, Chang SW. A review of risk prediction models in cardiovascular disease: conventional approach vs. artificial intelligent approach. *Comput Methods Programs Biomed*. 2021 [acceso 22/02/2023];207:106190. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S0169260721002649>
28. Azmi J, Arif M, Nafis MT, Alam MA, Tanweer S, Wang G. A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. *Med Eng Phys*. 2022 [acceso 22/02/2024];105:103825. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S1350453322000741>
29. Charalambous A, Dodlek N. Big Data, Machine Learning, and Artificial Intelligence to Advance Cancer Care: Opportunities and Challenges. *Semin Oncol Nurs*. 2023 [acceso 22/02/2024];39(3):151429. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S0749208123000669>
30. Arpaia P, Cuocolo R, Donnarumma F, Esposito A, Moccaldi N, Natalizio A, *et al.* Conceptual design of a machine learning-based wearable soft sensor for non-invasive cardiovascular risk assessment. *Measurement*. 2021 [acceso 22/02/2022];169:108551. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S0263224120310721>
31. Yu D, Liu Z, Zhuang W, Li K, Lu Y. Development and validation of machine learning based prediction model for postoperative pain risk after extraction of impacted mandibular third molars. *Heliyon*. 2023 [acceso 22/02/2024];9(12):e23052.

Disponibile

en:

<https://www.sciencedirect.com/science/article/pii/S240584402310260X>

32. Li X, Wang Y, Xu J. Development of a machine learning-based risk prediction model for cerebral infarction and comparison with nomogram model. *J Affect Disord.* 2022 [acceso 22/02/2024];314:341-8. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0165032722008138>

33. Thotad PN, Bharamagoudar GR, Anami BS. Diabetes disease detection and classification on Indian demographic and health survey data using machine learning methods. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews.* 2023 [acceso 02/02/2024];17(1):102690. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1871402122003071>

34. Jiang L, Xia Z, Zhu R, Gong H, Wang J, Li J, et al. Diabetes risk prediction model based on community follow-up data using machine learning. *Prev Med Rep.* 2023 [acceso 22/02/2024];35:102358. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2211335523002498>

35. Menteş N, Çakmak MA, Kurt ME. Estimation of service length with the machine learning algorithms and neural networks for patients who receiving home health care. *Eval Program Plann.* 2023 [acceso 22/02/2024];100:102324. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0149718923001015>

36. Chakraborty C, Bhattacharya M, Pal S, Lee SS. From machine learning to deep learning: An advances of the recent data-driven paradigm shift in medicine and healthcare. *Curr Res Biotechnol.* 2023 [acceso 22/02/2024];100:164. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2590262823000461>

37. Teji JS, Jain S, Gupta SK, Suri JS. NeoAI 1.0: Machine learning-based paradigm for prediction of neonatal and infant risk of death. *Comput Biol Med.* 2022 [acceso 22/02/2024];147:105639. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0010482522004310>

38. Avram C, Gligor A, Roman D, Soylu A, Nyulas V, Avram L. Machine learning based assessment of preclinical health questionnaires. *Int J Med Inform.* 2023 [acceso 22/02/2024];180:105248. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1386505623002666>

39. Chinnasamy P, Arun Kumar S, Navya V, Lakshmi Priya K, Sruthi Boddu S. Machine learning based cardiovascular disease prediction. *Mater Today Proc.* 2022 [acceso 22/02/2024];64:459-63. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2214785322030851>

40. Kao YT, Huang CY, Fang YA, Liu JC, Chang TH. Machine Learning-Based Prediction of Atrial Fibrillation Risk Using Electronic Medical Records in Older Aged Patients. *Am J Cardiol.* 2023 [acceso 22/02/2024];198:56-63. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0002914923001984>
41. Komuro J, Kusumoto D, Hashimoto H, Yuasa S. Machine learning in cardiology: Clinical application and basic research. *J Cardiol.* 2023 [acceso 22/02/2024];82(2):128-33. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0914508723001041>
42. Ashiku L, Al-Amin Md, Madria S, Dagli C. Machine Learning Models and Big Data Tools for Evaluating Kidney Acceptance. *Procedia Comput Sci.* 2021;185:177-84. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1877050921011005>
43. Sulaiman S, Kawsara A, El Sabbagh A, Mahayni AA, Gulati R, Rihal CS, et al. Machine learning vs. conventional methods for prediction of 30-day readmission following percutaneous mitral edge-to-edge repair. *Cardiovascular Revascularization Medicine.* 2023;56:18-24. Disponible en: <https://www.sciencedirect.com/science/article/pii/S155383892300204X>
44. Ye Y, Shi J, Zhu D, Su L, Huang J, Huang Y. Management of medical and health big data based on integrated learning-based health care system: A review and comparative analysis. *Comput Methods Programs Biomed.* 2021 [acceso 22/02/2024];209:106293. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0169260721003679>
45. Li P, Hu Y, Liu ZP. Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods. *Biomed Signal Process Control.* 2021 [acceso 22/02/2024];66:102474. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1746809421000719>
46. Gonzalez R, Saha A, Campbell CJ V, Nejat P, Lokker C, Norgan AP. Seeing the random forest through the decision trees. Supporting learning health systems from histopathology with machine learning models: Challenges and opportunities. *J Pathol Inform.* 2023 [acceso 22/02/2024];100347. Disponible en: <https://www.sciencedirect.com/science/article/pii/S215335392300161X>
47. Liu J, Wu X, Xie Y, Tang Z, Xie Y, Gong S. Small samples-oriented intrinsically explainable machine learning using Variational Bayesian Logistic Regression: An intensive care unit readmission prediction case for liver transplantation patients. *Expert Syst Appl.* 2024 [acceso 29/12/2024];235:121138. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0957417423016408>

48. Suri JS, Bhagawati M, Paul S, Protogeron A, Sfrikakis PP, Kitas GD, *et al.* Understanding the bias in machine learning systems for cardiovascular disease risk assessment: The first of its kind review. *Comput Biol Med.* 2022 [acceso 22/02/2024];142:105204. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0010482521009987>
49. Wang X, Yu D, Ma L, Lu X, Song J, Lei M. Using big data searching and machine learning to predict human health risk probability from pesticide site soils in China. *J Environ Manage.* 2022 [acceso 22/02/2024];320:115798. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0301479722013718>
50. Shamia D, Zahger D. Using machine learning to predict cardiovascular risk using self-reported questionnaires: A right, but a long way to go. *Int J Cardiol.* 2023 [acceso 22/02/2024];389:131178. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0167527323009853>

Conflicto de intereses

Los autores declaran que no existe conflicto de intereses.