

Comparación de modelos de machine learning para la predicción temprana de diabetes mellitus tipo 2 en entornos de atención primaria: evaluación de precisión y aplicabilidad clínica, revisión sistemática

Jaime Ernesto Pino Cotillo

Asesor

Luis Ángel Anillo Arrieta

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI
Especialización en Ciencia de Datos y Analítica

2025

Nota de Aceptación

Luis Ángel Anillo Arrieta
Director de Trabajo de Grado

Jurado

Jurado

Dedicatoria

Dedico este trabajo de grado a mi esposa, Adriana Lucía Rendón Valencia, por su amor incondicional, su apoyo constante y el impulso decisivo que me dio para atreverme a estudiar aquello que verdaderamente me apasiona.

A mis hijos, quienes con su comprensión y cariño me permitieron tomar parte del tiempo que por derecho les pertenece, para que su padre pudiera seguir creciendo personal y profesionalmente.

A mi hermano Juan Camilo Pino Cotillo, quien me acompañó y apoyó en cada etapa de esta empresa, y que generosamente me brindó los recursos tecnológicos necesarios para llevarla a cabo.

A mi madre, Carmen Elisa Cotillo, ejemplo vivo de trabajo, tenacidad y entrega; gracias a ella descubrí el profundo amor que siento por el conocimiento y la importancia de nunca dejar de aprender.

Con infinito cariño y gratitud.

Agradecimientos

A mi director de Trabajo de Grado, el Doctor Luis Ángel Anillo Arrieta, por su amable guía para lograr la construcción de este trabajo

Resumen

La Diabetes Mellitus tipo 2 (DM2) constituye una patología crónica no transmisible caracterizada por su naturaleza silente y progresiva, que representa una carga significativa para los sistemas de salud debido a sus altos costos de atención y su creciente prevalencia global. Su manifestación clínica suele evidenciarse en estadios intermedios o avanzados de la enfermedad, cuando ya existe compromiso multiorgánico establecido, lo que resalta la importancia crítica de su identificación temprana para mitigar tanto el impacto económico sobre el sistema sanitario, como las consecuencias en la calidad de vida de los pacientes afectados.

Dado el impacto de la DM2, resulta imperativo explorar herramientas innovadoras y de preferencia automatizadas (para ser aplicada en grandes poblaciones), para su detección precoz. Las técnicas de Machine Learning han demostrado resultados prometedores en la identificación temprana de la enfermedad y en la determinación de variables predictivas significativas (Avellán Valdés et al., 2022). En este contexto, surge la siguiente pregunta de investigación:

¿Qué modelo de Machine Learning proporciona el mejor equilibrio entre precisión predictiva y viabilidad de implementación para la detección temprana de Diabetes Mellitus tipo 2 en el contexto de atención primaria en Colombia, considerando las limitaciones de recursos existentes en el sistema de salud nacional?

Palabras claves: Machine Learning, Diabetes Mellitus, Predicción

Abstract

Type 2 Diabetes Mellitus (T2DM) constitutes a chronic non-communicable pathology characterized by its silent and progressive nature, representing a significant burden on healthcare systems due to its high care costs and growing global prevalence. Its clinical manifestation is usually evidenced in intermediate or advanced stages of the disease, when established multi-organ damage already exists, highlighting the critical importance of its early identification to mitigate both the economic impact on the healthcare system and the consequences for the quality of life of affected patients.

Given the impact of T2DM, it is imperative to explore innovative and preferably automated tools (for application in large populations) for its early detection. Machine Learning techniques have shown promising results in the early identification of the disease and in the determination of significant predictive variables (Avellán Valdés et al., 2022). In this context, the following research question arises:

Which Machine Learning model provides the best balance between predictive accuracy and implementation feasibility for the early detection of Type 2 Diabetes Mellitus in the primary care context in Colombia, considering the existing resource limitations in the national health system?

Keywords: Machine Learning, Diabetes Mellitus, Prediction

Tabla de Contenido

Introducción	12
Justificación	14
Objetivos.....	16
Objetivo General	16
Objetivos Específicos.....	16
Marco Teórico.....	17
Definición.....	17
Epidemiología.....	17
Impacto Socioeconómico	18
Factores de Riesgo y Criterios Diagnósticos	19
Escala de Tamizaje Findrisk	20
Utilización de Modelos de Machine Learning para la Predicción de la Diabetes	21
Metodología	23
Método	23
Identificación.....	23
Cribado (Screening)	24
Criterios de Inclusión	24
Criterios de Exclusión.....	24
Elegibilidad	24
Inclusión.....	26
Evaluación del Riesgo de Sesgo	30
Resultados.....	34

Análisis del Rendimiento Predictivo.....	34
Marco de Evaluación para la Selección de Modelos en Atención Primaria	35
Identificación del Enfoque más Adecuado para el Contexto Colombiano	37
Conclusiones.....	38
Recomendaciones	39
Referencias Bibliográficas	41
Apéndices.....	48

Lista de Tablas

Tabla 1 <i>Resultados de la Búsqueda Bibliográfica</i>	23
Tabla 2 <i>Comparación de Métricas de Rendimiento</i>	26
Tabla 3 <i>Modelos con Máxima Precisión</i>	34

Lista de Figuras

Figura 1 <i>Resultados de la Búsqueda Bibliográfica</i>	25
--	----

Lista de Apéndices

Apéndice A *Tabla Sintética de la Revisión Bibliográfica* 48

Apéndice B *Lista de Verificación de Metodología PRISMA* 66

Introducción

La Diabetes Mellitus representa uno de los mayores desafíos para la salud pública global en el siglo XXI. De acuerdo con (Home et al., s. f.) , la prevalencia mundial de diabetes alcanzó 537 millones de personas en 2021, con proyecciones que anticipan un incremento a 783 millones para 2045. En el contexto regional, Centro y Sudamérica registraron 35 millones de casos en 2021, proyectándose un aumento del 50% para alcanzar 57 millones en 2045.

En el contexto colombiano, según la editorial (Editorial La República S.A.S, 2023) , los datos epidemiológicos revelan que la diabetes afecta aproximadamente a 1.474.567 personas, representando una prevalencia del 3% en la población general. Sin embargo, existe una alta probabilidad de subregistro, estimándose que hasta el 50% de los casos permanecen sin diagnosticar. La relevancia clínica de esta patología se evidencia en las estadísticas de mortalidad de 2022, donde la diabetes constituyó la sexta causa de muerte con 7.195 defunciones. Además, su papel como factor de riesgo cardiovascular cobra especial importancia, considerando que las enfermedades cardiovasculares lideraron las causas de mortalidad en Colombia con 41.783 defunciones durante el mismo período (*¿Cuáles fueron las principales causas de muerte Colombia 2022?*, 2023).

Desde la perspectiva económica, el impacto de la diabetes en el sistema de salud colombiano es considerable. Los análisis de costos realizados por (Miksi Ávila, 2022) demuestran que la atención de un paciente diabético en el régimen contributivo supera en 1.3 veces el costo de un paciente no diabético. Esta disparidad se amplifica según la severidad de la enfermedad: los pacientes con severidad alta generan costos incrementales de COP 7.535.909, aquellos con severidad media COP 5.430.048, y los casos de severidad baja representan costos adicionales de COP 1.036.086. La Asociación Colombiana de Empresas de Medicina Integral (Acemi), citada por

(Editorial La República S.A.S, 2023), reporta una inversión anual superior a COP 600.000 millones en el régimen contributivo para la atención de pacientes diabéticos, incluyendo aproximadamente 22 millones de atenciones y medicamentos.

Considerando el significativo impacto epidemiológico y económico de la diabetes, junto con la evidencia que sustenta que su diagnóstico temprano reduce sustancialmente la morbilidad, mortalidad y costos asociados, resulta imperativo explorar herramientas innovadoras para su detección precoz. Las técnicas de Machine Learning han demostrado resultados prometedores en la identificación temprana de la enfermedad y en la determinación de variables predictivas significativas (Avellán Valdés et al., 2022). En este contexto, surge la siguiente pregunta de investigación:

¿Qué modelo de Machine Learning proporciona el mejor equilibrio entre precisión predictiva y viabilidad de implementación para la detección temprana de Diabetes Mellitus tipo 2 en el contexto de atención primaria en Colombia, considerando las limitaciones de recursos existentes en el sistema de salud nacional?

Justificación

La Diabetes Mellitus (DM) representa uno de los mayores desafíos para la salud pública global en el siglo XXI, configurándose como una pandemia de enfermedades no transmisibles con un impacto significativo en los sistemas de salud. De acuerdo con (Home et al., s. f.) la prevalencia mundial de diabetes alcanzó 537 millones de personas en 2021, con proyecciones epidemiológicas que anticipan un incremento sustancial a 783 millones de casos para 2045. En el contexto regional, Centro y Sudamérica registraron 35 millones de casos en 2021, proyectándose un aumento del 50% para alcanzar 57 millones en 2045.

El impacto económico de esta patología es considerable. En Estados Unidos, (Parker et al., 2024) reportan que el costo total de tratamiento de la diabetes en 2022 ascendió a \$412.9 mil millones, distribuidos en \$306.6 mil millones correspondientes a costos médicos directos y \$106.3 mil millones a costos indirectos atribuibles a la enfermedad. Esta carga económica, es equivalente a 1 de cada 4 dólares del presupuesto sanitario nacional.

En el contexto colombiano, según datos del Ministerio de Salud (*En el Día Mundial de la Diabetes, Minsalud promueve acciones prioritarias para evitar el crecimiento de esta enfermedad*, s. f.) la DM constituye la octava causa de mortalidad, con una tasa de mortalidad de 16.5 por 100,000 habitantes en 2022. Las estadísticas de la Cuenta de Alto Costo para el mismo año registraron 159,149 casos nuevos, alcanzando un total de 1,599,368 casos diagnosticados, lo que representa una prevalencia de 3 casos por cada 100 habitantes.

La evidencia presentada caracteriza a la DM como una enfermedad crónica de alta prevalencia, con una tendencia epidemiológica ascendente, que se distingue por su significativa tasa de complicaciones y mortalidad prematura. Estas características epidemiológicas y su considerable carga socioeconómica subrayan la importancia crítica de implementar estrategias efectivas

para su diagnóstico temprano, que permitan la historia natural de la enfermedad y reducir su impacto en la salud pública.

Ante la magnitud de este desafío en salud pública, la comunidad científica internacional ha dirigido esfuerzos significativos hacia la aplicación de tecnologías innovadoras para la detección temprana de la Diabetes Mellitus. Diversos grupos de investigación han explorado la implementación de algoritmos de Machine Learning como herramientas predictivas, obteniendo resultados prometedores en diferentes contextos geográficos y poblacionales. En la India, (Yuk et al., 2022) han desarrollado modelos predictivos aplicables a su población; mientras que en Arabia Saudita, (Khan et al., 2021) han evaluado la efectividad de diversos algoritmos adaptados a las características específicas de su sistema de salud. En el contexto latinoamericano, (Fregoso-Aparicio et al., 2021) han realizado investigaciones similares en México, contribuyendo a la comprensión de la aplicabilidad de estas herramientas en entornos con recursos limitados.

La diversidad de aproximaciones metodológicas y resultados obtenidos en estos estudios fundamenta la necesidad de realizar un análisis comparativo de los diferentes algoritmos implementados. Este análisis permitirá identificar aquellos modelos que ofrezcan el mejor equilibrio entre precisión diagnóstica y viabilidad de implementación, considerando específicamente las características, limitaciones y recursos disponibles en el sistema de salud colombiano.

Objetivos

Objetivo General

Evaluar la eficacia de diversos modelos de Machine Learning para la predicción temprana de Diabetes Mellitus tipo 2 en entornos de atención primaria, mediante un análisis comparativo de su precisión y aplicabilidad clínica, con el fin de identificar el enfoque más adecuado para su implementación en la práctica médica cotidiana en el entorno colombiano.

Objetivos Específicos

Analizar el rendimiento predictivo de diferentes algoritmos de Machine Learning que, utilizando los datos clínicos de los pacientes, ofrecen la mayor precisión en la detección temprana de Diabetes Mellitus tipo 2

Crear un marco de evaluación que combine medidas de rendimiento y factibilidad de implementación, utilizando criterios técnicos y clínicos, y así proporcionar una herramienta de selección de modelos adaptada a las necesidades de la atención primaria.

Proponer una estrategia de implementación para el modelo de Machine Learning más prometedor, considerando las limitaciones de recursos y las necesidades de capacitación del personal, para facilitar su integración efectiva en los flujos de trabajo clínicos existentes en atención primaria en el entorno colombiano.

Marco Teórico

Definición

La diabetes mellitus es definida por la ADA (*standards-of-care-2024.pdf*, s. f.) como un grupo de trastornos del metabolismo de los carbohidratos, caracterizados por dos problemas principales: en el primero la glucosa es subutilizada como fuente de energía y en el segundo hay sobreproducción de la misma, está mediada por alteraciones en gluconeogénesis, proceso metabólico que consiste en la síntesis de glucosa a partir de precursores que no son carbohidratos, como aminoácidos, lactato, glicerol y piruvato) y la glucogenólisis (5.1, 2022) . Estas alteraciones lo que derivan hiperglucemia y afecta otros sistemas corporales, ocasionando complicaciones clasificadas como agudas y crónicas.

Entre las complicaciones agudas se destacan la cetoacidosis diabética y el síndrome hiperglucémico hiperosmolar, mientras que las crónicas se dividen en microvasculares (retinopatía diabética, neuropatía diabética) y macrovasculares (enfermedad cardiovascular, enfermedad cerebrovascular y enfermedad renal crónica (Castro Martínez & e-libro, 2010) las complicaciones crónicas suelen ser de aparición tardía y en muchos casos permanecen asintomáticas, hasta estadios avanzados. Una vez establecidas estas complicaciones requieren atención médica multidisciplinaria permanente.

Epidemiología

Según la Federación Internacional de Diabetes (IDF, por sus siglas en inglés), para el año 2021, la diabetes afectaba a aproximadamente 537 millones de personas en todo el mundo. Se estima que esta cifra incrementará a 738 millones para 2045, impulsada por el envejecimiento de la población, el estilo de vida sedentario, la obesidad y el consumo de alimentos poco saludables (Magliano & Boyko, 2021)

En Colombia, según el Ministerio de Salud, en 2022 se reportaron 1.599.368 casos de diabetes y 159.149 casos nuevos, de los cuales el 59,15% corresponden a mujeres. Los departamentos con mayor prevalencia de esta enfermedad son Valle del Cauca, Atlántico, Bolívar, Risaralda y Meta. *(En el Día Mundial de la Diabetes, Minsalud promueve acciones prioritarias para evitar el crecimiento de esta enfermedad, s. f.)*

Impacto Socioeconómico

La diabetes mellitus representa una carga significativa para los sistemas de salud debido a la necesidad de atención médica continua, el uso de medicamentos crónicos y las complicaciones derivadas, especialmente las macrovasculares. En Estados Unidos, los costos totales de tratamiento en 2022 ascendieron a 412,9 mil millones de dólares, de los cuales 306,6 mil millones corresponden a costos médicos directos y 106,3 mil millones a costos indirectos atribuibles a la enfermedad (Parker et al., 2024). Esto representa el 25% del presupuesto sanitario nacional para este país.

A nivel individual, se estima que un paciente con diabetes incurre en gastos promedio de 19.736 dólares anuales, de los cuales 12.022 están directamente relacionados con la enfermedad, lo que equivale a 2,6 veces más que los gastos de una persona sin diabetes (Parker et al., 2024).

En Colombia, un estudio de 2007 estimó que los costos directos e indirectos de la diabetes ascendieron a 5,7 billones de pesos colombianos (2.708 millones de dólares) (González et al., 2009). Informes recientes emanados de ACEMI, indican que la atención de personas con diabetes en el régimen contributivo requirió una inversión de 600.000 millones de pesos en 22 millones de atenciones, incluyendo medicamentos (Editorial La República S.A.S, 2023). Según (Miksi Ávila, 2022) los costos son 1,3 veces mayores para pacientes diabéticos en comparación con no diabéticos, y varían según la severidad:

- Severidad alta: COP 7.535.909
- Severidad media: COP 5.430.048
- Severidad baja: COP 1.036.086

Factores de Riesgo y Criterios Diagnósticos

Bajo este panorama, la prevención y el diagnóstico temprano de la diabetes son fundamentales, según distintos estudios clínicos recogidos por las autoridades internacionales en diabetes (*standards-of-care-2024.pdf*, s. f.), han encontrado distintos factores de riesgo para padecer la enfermedad, el conjunto de estos factores de riesgo, entre los que se encuentran

1. Sobrepeso u obesidad (Índice de masa corporal $> 25 \text{ kg/m}^2$)
2. Antecedente de diabetes en primer grado de consanguinidad
3. Origen étnico de alto riesgo para diabetes (afrodescendiente, latino, nativo americano)
4. Historial de enfermedad cardiovascular
5. Hipertensión
6. Colesterol HDL $< 35 \text{ mg/dl}$ y/o triglicéridos $> 250 \text{ mg/dl}$
7. Mujeres con síndrome de ovario poliquístico
8. Sedentarismo
9. Condiciones relacionadas con resistencia a la insulina (obesidad severa y acantosis nigricans)

En los pacientes con estas características se propone la realización de exámenes de laboratorio a partir de los 35 años, repitiendo cada 3 años en pacientes con los exámenes iniciales normales, recalando medidas de estilo de vida para corregir los factores de riesgo, en pacientes con exámenes alterados para rango de prediabetes (glucemia en ayunas > 100 y < 126 , glucosa

dos horas después de la comida o espontánea < 160 , hemoglobina glicosilada $> 5.7\%$ y menor de 6.5%), deben ser repetidos anualmente,

Para el diagnóstico de diabetes se proponen los exámenes de laboratorio, en los rangos: glucemia en ayunas > 126 mg/dl, glucemia en cualquier momento > 200 mg/dl o hemoglobina glicosilada $> 6.5\%$

Escala de Tamizaje Findrisk

Conscientes de la presencia de unos claros factores de riesgo para desarrollar diabetes, se han buscado desarrollar herramientas para el tamizaje poblacional el caso más representativo es Findrisk, (Finnish Diabetes Risk Score), es una herramienta desarrollada para evaluar el riesgo de desarrollar diabetes tipo 2 en los próximos 10 años. Fue diseñado originalmente en Finlandia y se ha validado en diversas poblaciones para identificar a personas con riesgo elevado de diabetes no diagnosticada. El cuestionario de Findrisk incluye ocho preguntas que abarcan factores como la edad, el índice de masa corporal (IMC), la circunferencia de la cintura, la actividad física, la dieta, el uso de medicamentos para la hipertensión, los antecedentes familiares de diabetes y los niveles previos de glucosa en sangre.

En Colombia, fue incluida como criterio de tamizaje, en la guía de práctica clínica de diabetes mellitus tipo 2 del año 2015 del Ministerios de Salud, sustentada como recomendación fuerte a favor, con evidencia de calidad moderada, la prueba ostenta una sensibilidad del 74% y una especificidad del 60% , para nuestra población (Mariano Cantillo et al., 2019)

Si bien el Findrisk es una herramienta útil, validada, de uso cotidiano y bajo costo, tiene como desventaja que su aplicación requiere la realización de entrevista por parte de un agente de

salud entrenado para la toma de las mediciones requeridas, consume tiempo de consulta que generalmente es muy limitado y normalmente es requerido por el paciente para otros motivos de consulta.

Es aquí, donde la ciencia de datos, en especial la rama de Machine Learning, adquiere un papel principal, la capacidad de ésta de encontrar patrones específicos en grandes volúmenes de datos y la gran cantidad de datos producidos diariamente en salud, permiten suponer que es posible uso de herramientas de machine learning para la predicción e intervención temprana de la diabetes

Utilización de Modelos de Machine Learning para la Predicción de la Diabetes

La revisión bibliográfica reveló múltiples ejemplos de la aplicación de machine learning para la predicción de diabetes entre los que se encuentran los siguientes:

En el 2021 (Reddy et al., 2021) en la India aplicaron métodos de machine learning (Random Forest, Logistic Regression and XGBoost.) y compararon su rendimiento al aplicarlo sobre la base de datos Pima Indian Diabetes Dataset, que contiene los datos de 768 pacientes, encontrando que la técnica con mayor precisión fue el Random Forest con 0.9065161868651518.

Para el mismo año, (Fregoso-Aparicio et al., 2021) realizaron una revisión sistemática, encontrando el uso de múltiples técnicas de machine learning para la predicción de diabetes, destacan como limitación de su estudio la heterogeneidad de modelos (tamaño de muestra y tipo de población) lo que limitaba su comparación, así como la predicción de complicaciones de la diabetes, más que la predicción de la diabetes por sí misma.

En 2022, (Yuk et al., 2022) utilizaron un dataset que contenía 133.387 instancias recolectadas de 73.767 individuos y 218 variables, fueron obtenidas del datawarehouse clínico y los re-

gistros médicos electrónico en el hospital universitario de Ulsan, Corea del Sur, compararon varios modelos de machine learning encontrando el mejor rendimiento con el modelo SVM (máquina de vectores de soporte)

Para el mismo año (Tasin et al., 2023) utilizaron el PIMA indian Dataset y el RTML private dataset, aplicando las técnicas Decision tree, KNN classifier, Random Forest, Support Vector machine, Logistic regression, adabost, Xgboost, vottinng Classifier, encontrando el mejor rendimiento con el XGBoost, sin embargo ambas bases de datos están limitadas porque ambas bases de datos están principalmente compuestas por mujeres, y la base de datos de RTML, no cuenta con el valor de insulina existente en el PIMA indian dataset, el modelo obtuvo un rendimiento del 81% fue la base de una aplicación móvil para el autodiagnóstico

Para 2023 (Powar et al., 2023) exploraron un total de 34 estudios que utilizaron inteligencia artificial para el diagnóstico de diabetes en la India, encontrando que los factores más importantes para la aparición de la diabetes fueron la edad, el índice de masa corporal, la hipertensión, los antecedentes parentales y el grosor de los pliegues cutáneos.

También en 2023 (Mohsen et al., 2023) realizan una revisión sistemática de 40 estudio sobre la materia, encontrando que los modelos clásicos de aprendizaje automático (ML) dominaron estos estudios, siendo los registros electrónicos de salud (EHR) la modalidad de datos predominante, seguidos por los datos multi-ómicos, mientras que la imagen médica fue la menos utilizada. La mayoría de los estudios emplearon modelos de IA unimodales, y solo diez adoptaron enfoques multimodales. Tanto los modelos unimodales como los multimodales mostraron resultados prometedores, siendo estos últimos superiores.

Metodología

Método

El Método central empleado para la estructuración, búsqueda, selección y análisis de la literatura científica será el marco de la declaración PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Moher et al., 2009)

Esta elección garantiza la transparencia, replicabilidad y calidad del proceso de revisión. PRISMA se utilizará para guiar las siguientes etapas:

Identificación

Durante el mes de septiembre del año 2025, Se realizó la búsqueda Bibliográfica en las bases de datos Pubmed, Web of Sciences, Scopus y Google Scholar, que abarcara resultados desde el 2020 hasta el año 2025, usando la combinación de términos ((type 2 diabetes) AND (machine learning)) AND (primary care), para los cuatro buscadores, arrojando los siguientes resultados:

Tabla 1

Resultados de la Búsqueda Bibliográfica

Base de datos	Resultados
Pubmed	68
Web of Sciences	7
Scopus	95
Scholar Google	170
Total	340

Cribado (Screening)

Aplicación de criterios de inclusión y exclusión a los títulos y resúmenes de los estudios identificados.

Criterios de Inclusión

Mención específica dentro del título o el resumen del artículo de los términos de búsqueda, adicionalmente para Google Scholar se seleccionaron los que hubiesen sido citados más de 50 veces

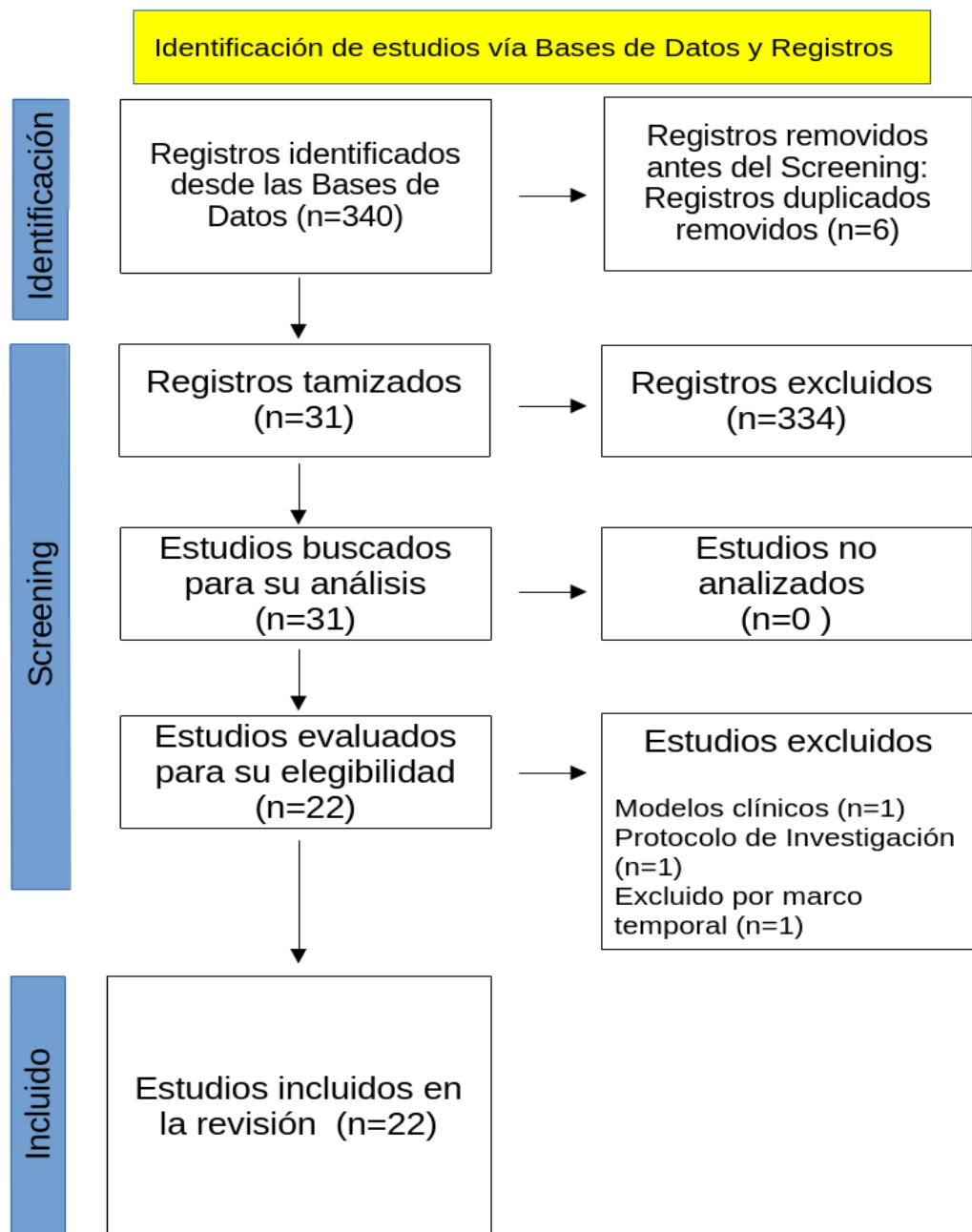
Criterios de Exclusión

Mención específica dentro del título o el resumen del artículo de los siguientes términos o similares: Diabetes tipo I, diabetes gestacional, retinopatía diabética, nefropatía diabética, enfermedad cardiovascular,

Tras la aplicación de estos criterios, quedaron para el análisis 25 artículos.

Elegibilidad

Tras la revisión del material bibliográfico, se descartan para el análisis tres artículos adicionales; Kishi y Fukuma (Kishi & Fukuma, 2023) por tratarse de una revisión de los modelos de predicción clínicos (no utilizan modelos de machine learning para la predicción), De Silva (De Silva et al., 2021), por tratarse de la presentación de un protocolo de investigación y Jaiswal (Jaiswal et al., 2021), ya que su revisión aunque recoge 11 artículos, estos no se encuentran en el marco temporal definido para la revisión sistemática, este proceso fue realizado manualmente por Jaime Ernesto Pino, investigador y se realizó el mismo ejercicio con un análisis automatizado usando la herramienta Notebook LM, se compararon ambos resultados, dejando para el análisis las concordancias en la figura 1 se resume el proceso de revisión bibliográfica.

Figura 1*Resultados de la Búsqueda Bibliográfica*

Inclusión

Se desarrollo una tabla sintética (apéndice 1), en el que se recopilan de cada uno de los textos analizados, el tipo de estudio, los resultados y conclusiones y las métricas utilizadas, el cual sirvió de insumo para construir la tabla 2, que describe y pondera los modelos utilizados, con base a su rendimiento.

Tabla 2

Comparación de Métricas de Rendimiento

Algoritmo / Modelo	Métrica	Valor Reportado	Detalles del Estudio	Referencia
Decision Tree (DT) con Selección de Características (ID3)	Exactitud (Accuracy)	99.9% (Validación LOSO)	Mayor rendimiento reportado mediante selección de características.	(Ahmed et al., 2021)
Ensemble Model (Votación Mayoritaria)	Exactitud (Accuracy)	98.60%	Supera a todos los algoritmos individuales implementados.	(Fazakis et al., 2021)
Random Forest (RF)	Exactitud (Accuracy)	98%	Logrado mediante el uso de selección de características para reducir dimensionalidad.	(Ginting et al., 2023)
Random Forest (RF)	AUC	0.91 (Rango 0.79 a 0.91)	Mejor modelo al combinar características de red de pacientes (Patient Network).	(Haq et al., 2020), (Zhang et al., 2020)
Fusión de Machine Learning (FM) Propuesto	Precisión (Accuracy)	95.2%	Técnica de fusión que utiliza la síntesis de tres algoritmos de ML y sistemas difusos.	(Ravaut et al., 2021)

Algoritmo / Modelo	Métrica	Valor Reportado	Detalles del Estudio	Referencia
Fusión de Machine Learning (FM) Propuesto	Sensibilidad	93.43%	Lograda en la fase de prueba (validación).	(Ravaut et al., 2021)
Fusión de Machine Learning (FM) Propuesto	Especificidad	96.03%	Lograda en la fase de prueba (validación).	(Ravaut et al., 2021)
Decision Tree (DT)	Exactitud (Accuracy)	94.9%	Rendimiento individual en el estudio de fusión.	(Ravaut et al., 2021)
CNN-LSTM (Deep Learning)	Precisión (Accuracy)	91.6%	Modelo de aprendizaje profundo para datos EMR longitudinales.	(S et al., 2024)
CNN-LSTM (Deep Learning)	Sensibilidad	85.7%	Modelo de aprendizaje profundo para datos EMR longitudinales.	(S et al., 2024)
CNN-LSTM (Deep Learning)	Especificidad	95.6%	Modelo de aprendizaje profundo para datos EMR longitudinales.	(S et al., 2024)
CNN-LSTM (Deep Learning)	Puntuación F1	89.2	Modelo de aprendizaje profundo para datos EMR longitudinales.	(S et al., 2024)
Transformer + LinearSVC	AUC	0.90	Superó a modelos clásicos en la base de datos MIMIC-IV.	(Fregoso-Aparicio et al., 2021) (Sarwar et al., 2020)

Algoritmo / Modelo	Métrica	Valor Reportado	Detalles del Estudio	Referencia
NSGA-II + Stacking (Propuesto)	Precisión (Accuracy)	88.18%	Enfoque de apilamiento para selección de características óptimas (Dataset Recolectado).	(Sarwar et al., 2020)
NSGA-II + Stacking (Propuesto)	Sensibilidad	0.88	Enfoque de apilamiento (Dataset Recolectado).	(Sarwar et al., 2020)
NSGA-II + Stacking (Propuesto)	Especificidad	0.94	Enfoque de apilamiento (Dataset Recolectado).	(Sarwar et al., 2020)
Red Neuronal (NN)	Exactitud (Accuracy)	88.6%	NN con dos capas ocultas y 400 épocas (Dataset PIDD).	(Khanam & Foo, 2021)
WeightedVoting LRRFs (Ensemble)	AUC	884	Modelo de votación ponderada propuesto para mejorar la predicción.	(Clarós et al., 2025)
Glmnet (Regresión Lasso)	AUC	0.859 (a T30)	Superó a todos los métodos en AUC en la predicción dicotómica.	(Dong et al., 2022)
Random Forest (RF)	Exactitud (Accuracy)	84%	Tasa de precisión obtenida en un estudio transversal en Indonesia.	(Kopitar et al., 2020)
Random Forest (RF)	Puntuación F1	0.83 (Macro promedio)	Tasa de precisión obtenida en un estudio transversal en Indonesia.	(Kopitar et al., 2020)
Extreme Gradient Boosting (Xgboost)	AUC- ROC	0.822 (Validación)	Modelo sin análisis de laboratorio, para detección de DM y pre-DM.	(Aftab et al., 2021)

Algoritmo / Modelo	Métrica	Valor Reportado	Detalles del Estudio	Referencia
Stochastic Gradient Boosting (SGB)	AUC	0.825 (Máximo)	Modelo desarrollado con datos retrospectivos de Suecia.	(Alkattan et al., 2024)
Regresión Logística (ML Model)	AROC	803	Clasificación a nivel poblacional usando edad, género y códigos ICD-10.	(Lama et al., 2021)
Regresión Logística (ML Model)	Sensibilidad	77.94%	Con un punto de corte ≥ 0.5 .	(Lama et al., 2021)
Regresión Logística (ML Model)	Especificidad	75.13%	Con un punto de corte ≥ 0.5 .	(Lama et al., 2021)
Gradient Boosting Decision Tree (GBDT)	AUC	80.26%	Predicción precisa utilizando solo datos administrativos de salud recopilados rutinariamente.	(Naveed et al., 2023) (Patil et al., 2023)
Gradient Boosting Machine (GBM)	AUC	79% (a 3 años)	Mejor rendimiento usando solo 39 variables autoinformadas.	(Firdous et al., 2022)
Support Vector Machines (SVM)	Exactitud (Accuracy)	78,125%	Rendimiento en el Dataset 1 (PIDD) con todas las características.	(Wändell et al., 2024)

Nota. Ponderación de los métodos de machine learning por valor reportado, según la medida de precisión.

Evaluación del Riesgo de Sesgo

Sesgo de Selección y Generalización Limitada (Población y Muestra) El sesgo de selección es un riesgo predominante debido a la fuerte dependencia de conjuntos de datos específicos y no diversos, lo que restringe gravemente la validez externa de los modelos. Por ejemplo, la base de datos Pima Indians Diabetes Database (PIDD) incluye únicamente mujeres mayores de 21 años y solo tiene dos etiquetas (diabéticas y sanas). Esta limitación en la diversidad de la población afecta la aplicabilidad global de los modelos. Otros conjuntos de datos, como MIMIC-IV v2.2, provienen de un único centro médico académico urbano, limitando la generalización a poblaciones más amplias. También se identificó la limitación en la muestra de SDPP (Programa Preventivo de Diabetes de Estocolmo), que se basó en una cantidad limitada de datos (alrededor de 8000 personas). Además, los participantes en exámenes de salud preventiva (población trabajadora) pueden sesgar la muestra, ya que generalmente se excluye a personas mayores donde la T2DM es más prevalente. (Clarós et al., 2025; Lama et al., 2021; Patil et al., 2023)

Sesgo de Calidad, Tipo de Datos y Manejo de Valores Faltantes La calidad de los datos es crucial, ya que los datos "sucios" (no limpios ni organizados) no pueden alcanzar un rendimiento tan bueno como los datos limpios. Un riesgo común es el de los datos desequilibrados (unbalanced data), donde hay significativamente más observaciones para una clase que para otra, lo que aumenta la probabilidad de clasificación errónea. Por ejemplo, en (Ginting et al., 2023), la población final de DM Tipo 2 representó solo el 4.2% del total de visitas. En cuanto al manejo de la información faltante, tanto (Aftab et al., 2021) (imputación por media) como (Lama et al., 2021) (imputación por mediana) utilizaron técnicas de imputación simples, las cuales pueden introducir sesgos si los datos faltantes no son aleatorios.

Sesgo de Información (Falta de Variables Clave y Autoinforme) Existe un sesgo de información significativo debido a la carencia de factores de riesgo cruciales. El estudio de (Naveed et al., 2023), aunque utilizó datos longitudinales, careció de datos socioeconómicos, historial familiar, hábitos dietéticos, actividad física, patrones de sueño, niveles de estrés psicosocial y datos del microbioma, todos ellos factores que se sabe contribuyen a la diabetes. De manera similar, otros estudios excluyeron antecedentes familiares de diabetes y diabetes gestacional por no estar disponibles en sus respectivos conjuntos de datos. Adicionalmente, la información puede estar sujeta a sesgo de autoinforme en grandes encuestas poblacionales. Por ejemplo, la simplificación de preguntas sobre hábitos alimentarios pudo haber impedido encontrar asociaciones relevantes, como se observó en el estudio de (Zhang et al., 2020)

Sesgo de No Respuesta y Reclutamiento Voluntario El sesgo de no respuesta se evidenció en el estudio de (Alkattan et al., 2024), donde la técnica de reclutamiento voluntario resultó en una alta tasa de no respuesta, alcanzando casi el 75%. Este sesgo puede reducir la generalizabilidad y llevar a una sobreestimación de la precisión del modelo de Machine Learning (ML) propuesto, afectando la capacidad del modelo para identificar con precisión a los pacientes con factores de riesgo.

Sesgo Metodológico en la Comparación y Modelado Se ha identificado sesgo en la comparación de modelos cuando las arquitecturas avanzadas (como Transformer) procesan todos los registros longitudinales sin agregación, mientras que los modelos clásicos analizan un número limitado de características pre-agregadas. Esta diferencia puede sesgar la comparación a favor del modelo avanzado al limitar la capacidad de los modelos clásicos para aprovechar toda la información disponible. Un sesgo similar se observó cuando el modelo de referencia (regresión lineal simple) se construyó con un conjunto fijo limitado de solo siete variables, mientras que los

modelos ML utilizaron un conjunto más amplio de 58 variables. Finalmente, el uso de modelos como el Gradient Boosting Machine (Zhang et al.) introduce el riesgo de sobreajuste (overfitting), aunque se utilicen técnicas de regularización.

Sesgo en la Definición del Resultado y Sesgo Temporal En cuanto a la variable objetivo, (Lama et al., 2021) combinaron prediabetes y Diabetes Tipo 2 (T2D) en una sola categoría de resultado. Esta definición amplia puede influir en la interpretación del modelo específicamente para la T2D. Respecto a la dimensión del tiempo, los modelos tradicionales de ML sufren de sesgo temporal, ya que no consideran el riesgo acumulado con el tiempo, por ejemplo, no distinguen entre un individuo que ha tenido hipertensión por 10 años frente a uno que la ha tenido por 1 año.

Sesgo de Validación Externa y Dependencia de Herramientas Secundarias La falta de validación externa rigurosa es una limitación recurrente. El estudio de (Alkattan et al., 2024) basó su validación en la comparación con la herramienta de evaluación de riesgo de la Asociación Americana de Diabetes (ADA), en lugar de utilizar resultados reales de medidas de diagnóstico de T2DM (como pruebas de glucosa plasmática en ayunas o HbA1c). La herramienta ADA, a su vez, depende de datos autoinformados, lo que perpetúa el sesgo. Además, los autores de (Dong et al., 2022) señalaron la necesidad de una validación adicional en una muestra externa de atención primaria para establecer la validez clínica de sus modelos.

Sesgo de Detección (Causalidad Inversa) En estudios que utilizan datos de registros de Atención Primaria (PHC), la probabilidad de que un paciente sea detectado con diabetes aumenta si visita con frecuencia la atención primaria. Este sesgo de detección implica que el modelo podría estar prediciendo la probabilidad de detección —basada en el uso de servicios de salud— y no la probabilidad real de que la enfermedad se desarrolle en la población general. Este problema

se complementa con el riesgo de inclusión incompleta de pacientes diagnosticados, ya que algunos sujetos podrían haber sido diagnosticados y tratados en clínicas no vinculadas al conjunto de datos, resultando en la pérdida de códigos de diagnóstico pertinentes.

Sesgo de Optimización/Interpretación y Utilidad (Lama et al., 2021). priorizaron la robustez de las explicaciones del modelo (utilizando una medida combinada \$\$\$) sobre la precisión pura (AUC), lo que implicó un sacrificio en la precisión del modelo a favor de interpretaciones robustas. Esto se relaciona con el sesgo de utilidad o implementación (Kopitar et al., 2020), donde la simplicidad y la interpretabilidad de los modelos de regresión más convencionales pueden hacerlos la opción más adecuada para la práctica clínica, a pesar del rendimiento predictivo marginalmente superior de los modelos de ML avanzados.

Resultados

Análisis del Rendimiento Predictivo

Síntesis de la Eficacia y Precisión Predictiva

Tras el análisis de la tabla 2 encontramos que la precisión (Accuracy) y el Área Bajo la Curva (AUC) son las métricas más utilizadas para evaluar el rendimiento. Los estudios demuestran consistentemente que los modelos de Machine Learning superan a los sistemas de puntuación de riesgo convencionales como la escala Findrisk, de uso frecuente en el entorno colombiano

A. Modelos con Máxima Precisión (Accuracy) Varios enfoques han reportado niveles de precisión excepcionalmente altos, aunque a menudo en conjuntos de datos específicos o mediante técnicas avanzadas:

Tabla 3

Modelos con Máxima Precisión

Modelo / Enfoque	Métrica Principal	Rendimiento Destacado	Referencia
Decision Tree (DT) + FS	Accuracy	99.9% (Validación LOSO)	(Ahmed et al., 2021)
Ensemble Model	Exactitud	98.60%	(Fazakis et al., 2021)
Random Forest (RF)	Exactitud	98%	(Ginting et al., 2023)
Fusión de ML (FM)	Precisión	95.2%	(Ravaut et al., 2021)
CNN-LSTM (Deep Learning)	Precisión	91.6%	(S et al., 2024)

Se observa que la precisión más alta la obtienen el algoritmo Decision Tree (DT) con selección de características (99.9%) (Ahmed et al., 2021) y los modelos de Conjunto (Ensemble) (98.60%) (Fazakis et al., 2021) o Random Forest (Ginting et al., 2023). En general, las revisiones sistemáticas indican que los algoritmos basados en árboles (como Random Forest y Decision Tree) muestran consistentemente los mejores rendimientos

B. Modelos con Alto Rendimiento (AUC) El AUC, que mide la capacidad de discriminación, es crucial para la robustez clínica:

Modelos Basados en Redes Neuronales/Transformadores: Los modelos basados en Transformer mostraron un AUC de 0.90 para la DM2, superando significativamente a los modelos clásicos. (Clarós et al., 2025)

Modelos de Conjunto y Árboles: El modelo de votación ponderada (Weighted Voting LRRFs) logró un AUC de 0.884 (Fazakis et al., 2021). El rendimiento de Random Forest ha oscilado entre 0.79 y 0.91 (Fregoso-Aparicio et al., 2021).

Gradient Boosting (GBM/GBDT/XGBoost): Estos modelos también se desempeñaron bien, alcanzando un AUC de 79% a 3 años con datos autoinformados (Ravaut et al., 2021), y 80.26 con datos administrativos rutinarios (S et al., 2024). Además, Xgboost mostró mejor discriminación que la Regresión Logística (AUC-ROC = 0.822) para la detección de DM y pre-DM sin análisis de laboratorio (Dong et al., 2022).

Marco de Evaluación para la Selección de Modelos en Atención Primaria

Para la implementación en el contexto médico colombiano, la aplicabilidad clínica es tan vital como la precisión. Un enfoque es óptimo si requiere datos fácilmente disponibles en el primer nivel de atención y ofrece interpretabilidad para la toma de decisiones.

A. Dependencia de los Datos La viabilidad en la atención primaria depende de la complejidad de los datos de entrada:

Uso de Datos Mínimos o Autoinformados: Los modelos que se basan en datos que no requieren pruebas de laboratorio sofisticadas son altamente aplicables en la atención primaria o en entornos con recursos limitados. (Dong et al., 2022) desarrolló un modelo (Xgboost) sin análisis de laboratorio que logró un AUC-ROC de 0.822 ; (Zhang et al., 2020) usó solo 39 variables autoinformadas (demografía, estilo de vida, historia familiar/médica) y aun así el modelo GBM logró un AUC del 79%

Uso de Datos Administrativos Rutinarios: (Ravaut et al., 2021) demostró que la predicción precisa (AUC 80.26) es posible utilizando solo datos administrativos de salud recopilados rutinariamente; (Alkattan et al., 2024) usó edad, género y códigos ICD-10 para alcanzar un AROC de 0.803

Predictores Clave en Atención primaria: Los factores identificados como más influyentes —como glucosa, edad, Índice de masa corporal, embarazos (Ahmed et al., 2021), historia familiar de Diabetes Mellitus (Ginting et al., 2023), hipertensión y obesidad (Wändell et al., 2024) son información estándar que se registra en la atención primaria.

B. Interpretabilidad y Aceptación Clínica La interpretabilidad es fundamental para la confianza y la adopción clínica. El uso de técnicas como SHAP (explicado por Lama et al. 2021) es crucial para entender qué características (IMC, edad, presión arterial) están impulsando el riesgo en un modelo de Bosque Aleatorio (Lama et al., 2021).

C. Desafío de los Modelos Complejos Aunque el Aprendizaje Profundo (DL) y los modelos basados en Transformer demuestran una precisión superior (AUC de hasta 0.91) (Naveed et al., 2023), (Clarós et al., 2025), (Fregoso-Aparicio et al., 2021), a menudo requieren grandes

volúmenes de datos longitudinales en registros médicos electrónicos (Naveed et al., 2023) o conjuntos de datos muy ordenados, (Fregoso-Aparicio et al., 2021) lo que podría ser un obstáculo para la rápida implementación en todos los establecimiento de la atención primaria en Colombia.

Identificación del Enfoque más Adecuado para el Contexto Colombiano

Basándose en el equilibrio entre la alta precisión demostrada en los estudios y la alta aplicabilidad clínica requerida en entornos de atención primaria con variabilidad de recursos, el enfoque más adecuado para la implementación inicial en Colombia se centra en los Modelos Basados en Árboles (Random Forest o Gradient Boosting) por las siguientes razones:

Rendimiento Robusto y Consistente: Los modelos como Random Forest (RF) y Gradient Boosting Machine (GBM/Xgboost) son los más citados por su capacidad para lograr alta precisión (hasta 98% para RF, (Lu et al., 2022) ; AUC hasta 0.822 para Xgboost sin laboratorios (Dong et al., 2022)) y su eficacia en diversos conjuntos de datos (Fregoso-Aparicio et al., 2021).

Manejo de Datos de Atención Primaria: Estos modelos son eficaces incluso cuando se entrenan con características limitadas (como datos auto informados o administrativos rutinarios) (Ravaut et al., 2021), (S et al., 2024), lo que facilita su integración en los sistemas de Registros Médicos Electrónicos o en la captura de datos en la atención primaria

Interpretación Clínica: La capacidad de estos modelos para integrarse con herramientas de interpretabilidad (como SHAP, utilizado por (Lama et al., 2021)) permite a los médicos de atención primaria entender los factores de riesgo específicos que impulsan la predicción, facilitando intervenciones dirigidas (e.g., confirmar que la edad, el IMC y la historia familiar son los mayores contribuyentes al riesgo) (Ravaut et al., 2021), (Ahmed et al., 2021), (Lama et al., 2021)

Conclusiones

Un enfoque de Machine Learning Interpretable basado en árboles (como Xgboost o Random Forest), que utiliza predominantemente datos de fácil recolección en la atención primaria (demografía, biomarcadores básicos como glucosa e IMC, e historial médico/familiar), ofrece la mejor combinación de precisión (generalmente superior al 80% AUC) y factibilidad operativa para ser implementado en el contexto médico colombiano.

Si la infraestructura de datos madura (posibilitando el uso de registros longitudinales secuenciales), los modelos de Aprendizaje Profundo (CNN-LSTM) o los Ensembles/Fusión podrían adoptarse para alcanzar niveles de precisión cercanos al 95-99% (Fregoso-Aparicio et al., 2021), (Alkattan et al., 2024), (Aftab et al., 2021)

Recomendaciones

Iniciar con Modelos Interpretables y de Alta Viabilidad Operativa; Para lograr una adopción exitosa de algoritmos de machine learning en la Atención Primaria, es fundamental seleccionar modelos que equilibren precisión predictiva con facilidad de implementación. En este sentido, se recomienda comenzar con modelos basados en árboles, como Random Forest o Gradient Boosting (por ejemplo, XGBoost). Estos algoritmos han demostrado un desempeño robusto, alcanzando niveles de exactitud de hasta el 98% en el caso de Random Forest y un área bajo la curva (AUC) de 0.822 para XGBoost, incluso sin requerir datos de laboratorio. A diferencia de enfoques más complejos, como el aprendizaje profundo, que demandan grandes volúmenes de datos longitudinales y una infraestructura robusta, los modelos basados en árboles son más accesibles y ofrecen resultados consistentes en contextos con variabilidad de recursos, lo que los convierte en una opción ideal para una implementación inicial en el sistema de salud colombiano.

Priorizar el Uso de Datos Rutinarios y de Fácil Acceso; La viabilidad de los modelos de machine learning en la Atención Primaria depende en gran medida de la disponibilidad y simplicidad de los datos utilizados para su entrenamiento. Por ello, se sugiere centrar los esfuerzos en la información que ya se recopila de manera rutinaria durante las consultas de atención primaria, como la edad, el índice de masa corporal (IMC), los niveles de glucosa, la historia familiar de diabetes, la hipertensión y la obesidad. Estudios han demostrado que estos datos, comúnmente registrados en los sistemas administrativos de salud, son suficientes para lograr predicciones precisas, con un AUC de hasta 80.26% en modelos que no requieren pruebas de laboratorio sofisticadas. Este enfoque reduce la dependencia de recursos adicionales, como análisis clínicos complejos, y facilita la integración de Machine Learning en entornos con limitaciones logísticas, asegurando una implementación práctica y escalable.

Fomentar la Transparencia y la Confianza mediante Modelos Interpretables; La aceptación de los sistemas de Machine Learning por parte del personal médico depende de su capacidad para comprender las predicciones generadas. Para ello, es esencial que los sistemas incorporen herramientas de aprendizaje automático interpretable, como el uso de técnicas SHAP (SHapley Additive exPlanations) integradas en modelos como Random Forest. Estas herramientas permiten al personal de salud identificar con claridad los factores específicos, como un IMC elevado o la edad del paciente, que contribuyen a una predicción de riesgo. Esta transparencia no solo fortalece la confianza en las decisiones clínicas asistidas por Machine Learning, sino que también facilita intervenciones dirigidas, al proporcionar al médico una comprensión detallada de los factores de riesgo. Garantizar la interpretabilidad del modelo es, por tanto, un pilar fundamental para su adopción efectiva en la Atención Primaria.

Referencias Bibliográficas

- 5.1: *Gluconeogénesis y glucogenólisis*. (2022, octubre 30). LibreTexts Español. https://espanol.libretexts.org/Salud/Ciencias_Biolog%C3%ADa_celular%2C_gen%C3%A9tica_y_bioqu%C3%ADmica_para_estudiantes_precl%C3%ADnicos/05%3A_Combustible_para_M%C3%A1s_Tardes/5.01%3A_Gluconeog%C3%A9nesis_y_glucogen%C3%B3lisis
- Aftab, S., Alanazi, S., Ahmad, M., Khan, M. A., Fatima, A., & Elmitwally, N. S. (2021). Cloud-Based Diabetes Decision Support System Using Machine Learning Fusion. *Computers, Materials and Continua*, 68(1), 1341-1357. <https://doi.org/10.32604/cmc.2021.016814>
- Ahmed, N., Ahammed, R., Islam, Md. M., Uddin, Md. A., Akhter, A., Talukder, Md. A., & Paul, B. K. (2021). Machine learning based diabetes prediction and development of smart web application. *International Journal of Cognitive Computing in Engineering*, 2, 229-241. <https://doi.org/10.1016/j.ijcce.2021.12.001>
- Alkattan, A., Al-Zeer, A., Alsaawi, F., Alyahya, A., Alnasser, R., Alsarhan, R., Almusawi, M., Alabdulaali, D., Mahmoud, N., Al-Jafar, R., Aldayel, F., Hassanein, M., Haji, A., Alsheikh, A., Alfaifi, A., Elkagam, E., Alfridi, A., Alfaleh, A., Alabdulkareem, K., ... Gregg, E. W. (2024). The utility of a machine learning model in identifying people at high risk of type 2 diabetes mellitus. *Expert Review of Endocrinology and Metabolism*, 19(6), 513-522. <https://doi.org/10.1080/17446651.2024.2400706>
- Avellán Valdés, S., Holguín Intriago, C. A., & Cruz Felipe, M. del R. (2022). Predicción de las principales enfermedades que afectan la salud en Ecuador a partir de factores de riesgo. *Serie Científica de la Universidad de las Ciencias Informáticas*, 15(8), 37-50.

- Castro Martínez, M. Guadalupe., & e-libro, Corp. (2010). *Complicaciones crónicas en la diabetes mellitus*. <https://research.ebsco.com/linkprocessor/plink?id=d850df6e-f256-39ed-9e44-eb999a332bbb>
- Clarós, A., Ciudin, A., Muria, J., Llull, L., Mola, J. À., Pons, M., Castán, J., Cruz, J. C., & Simó, R. (2025). A model based on artificial intelligence for the prediction, prevention and patient-centred approach for non-communicable diseases related to metabolic syndrome. *European Journal of Public Health*, 35(4), 642-649. <https://doi.org/10.1093/eurpub/ckaf098>
- ¿Cuáles fueron las principales causas de muerte Colombia 2022? (2023, enero 26). <https://consultorsalud.com/principales-causas-muerte-colombia-2022/>
- De Silva, K., Enticott, J., Barton, C., Forbes, A., Saha, S., & Nikam, R. (2021). Use and performance of machine learning models for type 2 diabetes prediction in clinical and community care settings: Protocol for a systematic review and meta-analysis of predictive modeling studies. *Digital Health*, 7, 20552076211047390. <https://doi.org/10.1177/20552076211047390>
- Dong, W., Tse, T. Y. E., Mak, L. I., Wong, C. K. H., Wan, Y. F. E., Tang, H. M. E., Chin, W. Y., Bedford, L. E., Yu, Y. T. E., Ko, W. K. W., Chao, V. K. D., Tan, C. B. K., & Lam, L. K. C. (2022). Non-laboratory-based risk assessment model for case detection of diabetes mellitus and pre-diabetes in primary care. *Journal of Diabetes Investigation*, 13(8), 1374-1386. <https://doi.org/10.1111/jdi.13790>
- Editorial La República S.A.S. (2023, marzo 11). *Diabetes: Los costos de una pandemia silenciosa*. Diario La República. <https://www.larepublica.co/analisis/katherine-restrepo-erazo-3742780/diabetes-los-costos-de-una-pandemia-silenciosa-3742779>

En el Día Mundial de la Diabetes, Minsalud promueve acciones prioritarias para evitar el crecimiento de esta enfermedad. (s. f.). Recuperado 17 de noviembre de 2024, de

<https://www.minsalud.gov.co/Paginas/dia-mundial-de-la-diabetes.aspx>

Fazakis, N., Kocsis, O., Dritsas, E., Alexiou, S., Fakotakis, N., & Moustakas, K. (2021). Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction. *IEEE Access*, 9, 103737-103757. <https://doi.org/10.1109/ACCESS.2021.3098691>

Firdous, S., Wagai, G. A., & Sharma, K. (2022). A survey on diabetes risk prediction using machine learning approaches. *Journal of Family Medicine and Primary Care*, 11(11), 6929-6934. https://doi.org/10.4103/jfmpe.jfmpe_502_22

Fregoso-Aparicio, L., Noguez, J., Montesinos, L., & García-García, J. A. (2021). Machine learning and deep learning predictive models for type 2 diabetes: A systematic review. *Diabetology & Metabolic Syndrome*, 13(1), 1-22. <https://doi.org/10.1186/s13098-021-00767-9>

Ginting, J. B., Suci, T., Ginting, C. N., & Girsang, E. (2023). Early detection system of risk factors for diabetes mellitus type 2 utilization of machine learning-random forest. *Journal of Family and Community Medicine*, 30(3), 171. https://doi.org/10.4103/jfcm.jfcm_33_23

González, J. C., Walker, J. H., & Einarson, T. R. (2009). Cost-of-illness study of type 2 diabetes mellitus in Colombia. *Pan American Journal of Public Health*, 26(1), 55-63.

Haq, A. U., Li, J. P., Khan, J., Memon, M. H., Nazir, S., Ahmad, S., Khan, G. A., & Ali, A. (2020). Intelligent Machine Learning Approach for Effective Recognition of Diabetes in E-Healthcare Using Clinical Data. *Sensors*, 20(9), 2649.

<https://doi.org/10.3390/s20092649>

- Home, Resources, diabetes, L. with, Acknowledgement, FAQs, Contact, & Policy, P. (s. f.). *IDF Diabetes Atlas 2021 | IDF Diabetes Atlas*. Recuperado 16 de noviembre de 2024, de <https://diabetesatlas.org/atlas/tenth-edition/>
- Jaiswal, V., Negi, A., & Pal, T. (2021). A review on current advances in machine learning based diabetes prediction. *Primary Care Diabetes*, *15*(3), 435-443. <https://doi.org/10.1016/j.pcd.2021.02.005>
- Khan, F. A., Zeb, K., Al-Rakhami, M., Derhab, A., & Bukhari, S. A. C. (2021). Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review. *IEEE Access, Access, IEEE*, *9*, 43711-43735. <https://doi.org/10.1109/ACCESS.2021.3059343>
- Khanam, J. J., & Foo, S. Y. (2021). A comparison of machine learning algorithms for diabetes prediction. *ICT Express*, *7*(4), 432-439. <https://doi.org/10.1016/j.ict.2021.02.004>
- Kishi, A., & Fukuma, S. (2023). Implementation status of prediction models for type 2 diabetes. *Primary Care Diabetes*, *17*(6), 655-657. <https://doi.org/10.1016/j.pcd.2023.09.002>
- Kopitar, L., Kocbek, P., Cilar, L., Sheikh, A., & Stiglic, G. (2020). Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Scientific Reports*, *10*(1), 11981. <https://doi.org/10.1038/s41598-020-68771-z>
- Lama, L., Wilhelmsson, O., Norlander, E., Gustafsson, L., Lager, A., Tynelius, P., Wärvik, L., & Östenson, C.-G. (2021). Machine learning for prediction of diabetes risk in middle-aged Swedish people. *Heliyon*, *7*(7). <https://doi.org/10.1016/j.heliyon.2021.e07419>
- Lu, H., Uddin, S., Hajati, F., Moni, M. A., & Khushi, M. (2022). A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. *Applied Intelligence*, *52*(3), 2411-2422. <https://doi.org/10.1007/s10489-021-02533-w>

- Magliano, D., & Boyko, E. J. (with International Diabetes Federation). (2021). *IDF diabetes atlas* (10th edition). International Diabetes Federation.
- Mariano Cantillo, H. J., Ocampo, D. F., & Cuello Santana, K. L. (2019). Uso del instrumento FINDRISK para identificar el riesgo de prediabetes y diabetes mellitus tipo 2. *Revista Repertorio de Medicina y Cirugía*, 28(3), 157-163. <https://doi.org/10.31260/RepertMedCir.v28.n3.2019.956>
- Miksi Ávila, S. L. (2022). *Estimación de los costos directos de la atención de diabetes mellitus 2 en Colombia*. <http://hdl.handle.net/1992/57362>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Mohsen, F., Al-Absi, H. R. H., Yousri, N. A., El Hajj, N., & Shah, Z. (2023). A scoping review of artificial intelligence-based methods for diabetes risk prediction. *NPJ Digital Medicine*, 6(1), 197. <https://doi.org/10.1038/s41746-023-00933-5>
- Naveed, I., Kaleem, M. F., Keshavjee, K., & Guergachi, A. (2023). Artificial intelligence with temporal features outperforms machine learning in predicting diabetes. *PLOS Digital Health*, 2(10). <https://doi.org/10.1371/journal.pdig.0000354>
- Parker, E. D., Lin, J., Mahoney, T., Ume, N., Yang, G., Gabbay, R. A., ElSayed, N. A., & Banuru, R. R. (2024). Economic Costs of Diabetes in the U.S. in 2022. *Diabetes Care*, 47(1), 26-43. <https://doi.org/10.2337/dci23-0085>
- Patil, R. N., Rawandale, S., Rawandale, N., Rawandale, U., & Patil, S. (2023). An efficient stacking based NSGA-II approach for predicting type 2 diabetes. *International Journal of*

Electrical and Computer Engineering, 13(1), 1015-1023.

<https://doi.org/10.11591/ijece.v13i1.pp1015-1023>

Powar, J. D., Dase, R., & Bhosle, D. (2023). Application of Artificial Intelligence in prediction type 2 diabetes mellitus: A Systematic Review. *Pravara Medical Review*, 15(4), 36-45.

<https://doi.org/10.36848/PMR/2023/00000.51000>

Ravaut, M., Harish, V., Sadeghi, H., Leung, K. K., Volkovs, M., Kornas, K., Watson, T., Poutanen, T., & Rosella, L. C. (2021). Development and Validation of a Machine Learning Model Using Administrative Health Data to Predict Onset of Type 2 Diabetes. *JAMA Network Open*, 4(5), e2111315. <https://doi.org/10.1001/jamanetworkopen.2021.11315>

Reddy, Y. J., David, R., Reddy, Y. K., Vasishta, R., & Ambore, A. (2021). DIABETES DISEASE PREDICTION USING MACHINE LEARNING ENSEMBLE METHOD. *International Journal of Advanced Research in Computer Science*, 12, 125-128.

S, G., Venkata Siva Reddy, R., & Ahmed, M. R. (2024). Exploring the effectiveness of machine learning algorithms for early detection of Type-2 Diabetes Mellitus. *Measurement: Sensors*, 31, 100983. <https://doi.org/10.1016/j.measen.2023.100983>

Sarwar, A., Ali, M., Manhas, J., & Sharma, V. (2020). Diagnosis of diabetes type-II using hybrid machine learning based ensemble model. *International Journal of Information Technology*, 12(2), 419-428. <https://doi.org/10.1007/s41870-018-0270-5>

Standards-of-care-2024.pdf. (s. f.). Recuperado 16 de noviembre de 2024, de https://ada.silver-chair-cdn.com/ada/content_public/journal/care/issue/47/supplement_1/16/standards-of-care-2024.pdf?Expires=1734834101&Signature=ElpiP1eM3I6WC3ohi7StoxNHXPagoPhNGR46YsHf8eqJJNgrinPG4JpM7lolYuycG2eThnan0oAPoVIQk2MDiD-mBmOlxxrmDw2N77EEe-ylHB76afQpAoKmC0LxuQICGhNGTyFXINjiFwk-

TgAdkJ07xYxaqq8B-cl39LYg14IRcSxucBllcW8wIXA-
 btHpuXh4QYRk3~xX66ejDu2YVEZtzG2xq9aTsZ-S-NXNoqX6un8d1IXO0UtN-
 cun~e~~yekxqLg8i~CuZV81d-qHIUNRnEVMZ7A37f4ONtgGyCu4ie-
 sIL22WLv09a~8xIINAJ1Jg8IWm22LvhP3QSBtkAqNw__&Key-Pair-
 Id=APKAIE5G5CRDK6RD3PGA

Tasin, I., Nabil, T. U., Islam, S., & Khan, R. (2023). Diabetes prediction using machine learning and explainable AI techniques. *Healthcare Technology Letters*, *10*(1-2), 1-10.

<https://doi.org/10.1049/htl2.12039>

Wändell, P., Carlsson, A. C., Wierzbicka, M., Sigurdsson, K., Ärnlöv, J., Eriksson, J., Wachtler, C., & Ruge, T. (2024). A machine learning tool for identifying patients with newly diagnosed diabetes in primary care. *Primary Care Diabetes*, *18*(5), 501-505.

<https://doi.org/10.1016/j.pcd.2024.06.010>

Yuk, H., Gim, J., Min, J. K., Yun, J., & Heo, T.-Y. (2022). Artificial Intelligence-based Prediction of Diabetes and Prediabetes Using Health Checkup Data in Korea. *Applied Artificial Intelligence*, *36*(1), 1-25. <https://doi.org/10.1080/08839514.2022.2145644>

Zhang, L., Shang, X., Sreedharan, S., Yan, X., Liu, J., Keel, S., Wu, J., Peng, W., & He, M. (2020). Predicting the Development of Type 2 Diabetes in a Large Australian Cohort Using Machine-Learning Techniques: Longitudinal Survey Study. *JMIR Medical Informatics*, *8*(7), e16850. <https://doi.org/10.2196/16850>

Apéndices

Apéndice A

Tabla Sintética de la Revisión Bibliográfica

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Ahmed et al.	2021	Estudio de desarrollo de modelos de clasificación basados en Aprendizaje Automático (ML) para la detección de diabetes.	El modelo propuesto superó el rendimiento de investigaciones anteriores. El algoritmo SVM superó a otros algoritmos, alcanzando una exactitud del 78.125% en el Dataset 1 con todas las características. Las características más influyentes fueron la glucosa, edad, IMC y embarazos.	Exactitud (Accuracy), Precisión (Precision), Sensibilidad (Recall), Puntuación F1 (F1-score), Curva ROC.	Scholar Google

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Fazakis et al.	2021	Modelado predictivo a largo plazo del riesgo de T2DM, comparando modelos de ML con sistemas de puntuación de riesgo (FINDRISC, Leicester).	El modelo de conjunto WeightedVotingLRRFs es propuesto para mejorar la predicción de la diabetes, con un AUC de 0.884. Los algoritmos de ML mostraron resultados de predicción significativamente mejores que los sistemas de puntuación de riesgo convencionales (FINDRISC, Leicester).	AUC (0.884), Sensibilidad (Recall), Especificidad, Valor Predictivo Positivo (+PV), Valor Predictivo Negativo (-PV).	Scholar Google

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Ginting et al.	2023	Estudio no experimental, diseño cuantitativo, diseño de encuesta transversal y análisis asociativo, enfocado en la detección temprana de factores de riesgo de DM tipo 2.	La tasa de precisión del sistema/modelo obtenida con el algoritmo Random Forest fue del 84%. La historia familiar de DM fue el factor de riesgo más alto (OR = 15.101). El nivel de azúcar en sangre fue el factor de riesgo con la menor Odds Ratio (OR = 0.016, protector).	Exactitud (Accuracy) (84%), Precisión (Precision), Sensibilidad (Recall), Puntuación F1 (F1-score) (Macro promedio de 0.83).	Scholar Google

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Haq et al.	2020	Diseño de un sistema de decisión inteligente basado en ML para la detección de diabetes. Se propuso un algoritmo de selección de características basado en filtros (DT-ID3).	El método propuesto (DT(ID3)+DT) logró una alta precisión de 99% (Hold out), 99.8% (K-fold) y 99.9% (LOSO). El rendimiento es alto comparado con métodos previos. Las características más importantes fueron Concentración de Glucosa en Plasma (GL), Función de Pedigrí de Diabetes (DPF) e Índice de Masa Corporal (BMI).	Exactitud (Accuracy), Especificidad (Specificity), Sensibilidad (Sensitivity), MCC, ROC-AUC, Precisión (Precision), Recall, Puntuación F1 (F1-score).	Scholar Google

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Zhang et al.	2020	Estudio longitudinal retrospectivo de una cohorte. Desarrollo de una herramienta de predicción de riesgo de diabetes T2DM utilizando información autoinformada sin biomarcadores.	El modelo Gradient Boosting Machine (GBM) tuvo el mejor rendimiento (AUC del 79% a 3 años y 75% a 10 años). Todos los modelos de ML identificaron el Índice de Masa Corporal (BMI) como el factor más significativo. La incidencia de T2DM fue del 6.05% durante un seguimiento promedio de 8.8 años.	Área bajo la curva (AUC). Error Cuadrático Medio (RMSE).	Scholar Google

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Ravaut et al.	2021	Modelo analítico de decisión a nivel poblacional para predecir la incidencia de T2DM 5 años antes del inicio.	El modelo GBDT (Gradient Boosting Decision Tree) alcanzó un AUC de prueba de 80.26. La predicción fue precisa utilizando solo datos administrativos de salud recopilados rutinariamente. El 5% de los pacientes de mayor riesgo concentró el 26% del costo anual total de la diabetes.	AUC (80.26), Calibración, Contribución de características (valores de Shapley).	Scholar Google
S et al.	2024	Investigación para evaluar la efectividad de los algoritmos de ML para la identificación temprana de	El algoritmo Random Forest (RF) logró una precisión impresionante del 98%. El uso de selección de características mejoró significativamente la	Exactitud (Accuracy) (98%), Precisión (Precision), Sensibilidad (Recall), Puntuación F1	Scholar Google

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Lu et al.	(Aceptado en 2021).	T2DM. Modelo predictivo basado en una red de pacientes (Patient Network) combinada con técnicas de ML para predecir el riesgo de T2DM.	El modelo Random Forest superó a los demás. El rendimiento del modelo (AUC) osciló entre 0.79 y 0.91. La centralidad de eigenvector, la centralidad de closeness de la red y la edad del paciente fueron las características más importantes.	AUC (0.79 a 0.91), Exactitud (Accuracy), Precisión (Precision), Sensibilidad (Recall), Puntuación F1 (F1-score).	Scholar Google

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Fregoso-Aparicio et al. (Revisión)	2021	Revisión sistemática que analiza 90 estudios sobre modelos predictivos de DM tipo 2.	Los algoritmos basados en árboles (Decision Tree, Random Forest) mostraron los mejores rendimientos. Los modelos entrenados con conjuntos de datos ordenados (tidy datasets) lograron modelos casi perfectos. La Exactitud y el AUC son las métricas más usadas.	Exactitud (Accuracy), Sensibilidad (Sensitivity), Especificidad (Specificity), AUC (ROC), Puntuación F1 (F1-score).	Scholar Google
Sarwar et al.	2020	Estudio experimental. Desarrollo de un sistema experto basado en un modelo de conjunto (ensemble model) para el diagnóstico	La técnica de conjunto (Ensemble) aseguró una exactitud del 98.60%, superando a todos los algoritmos individuales implementados. ANN, Naïve Bayes, SVM y	Porcentaje de clasificados correctamente (Exactitud) y clasificados incorrectamente.	Scholar Google

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
		de DM tipo II.	K-NN fueron los siguientes en exactitud.		
Khanam y Foo	2021	Estudio comparativo y de clasificación utilizando algoritmos ML y Redes Neuronales (NN).	La Red Neuronal (NN) con dos capas ocultas y 400 épocas proporcionó la mayor exactitud: 88.6% (88.57% de exactitud de prueba). LR y SVM mostraron alrededor de 77%–78% de exactitud.	Exactitud (Accuracy), Precisión (Precision), Sensibilidad (Recall), Puntuación F1 (F-measure), Curva ROC.	Scholar Google

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Clarós et al.	2025	<p>Estudio de prueba de concepto.</p> <p>Estudio retrospectivo de evaluación del impacto de un modelo de IA para la predicción, prevención y gestión del Síndrome Metabólico (MetS) y enfermedades no transmisibles (NCDs) relacionadas.</p> <p>Enfoque de dos etapas: preentrenamiento y ajuste fino.</p>	<p>El modelo basado en Transformer superó significativamente a los modelos clásicos, logrando valores de AUC más altos en todas las enfermedades. El AUC para DM Tipo 2 fue 0.90 y para CKD fue 0.91 (Transformer + LinearSVC).</p>	<p>Área bajo la curva (AUC) y Puntuación F1 (F1-score).</p>	Pubmed

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
De Silva et al.	2021 (Fecha de aceptación)	Protocolo para una Revisión Sistemática y Metaanálisis de modelos predictivos de Machine Learning (ML). Se enfoca en la predicción de la diabetes tipo 2 (T2DM) en entornos de atención clínica y comunitaria.	El objetivo es identificar modelos predictivos de ML para T2DM y determinar su rendimiento. No se presentan resultados de un estudio clínico primario.	Habilidad de discriminación (ej. índice c, AUC, AUROC), Calibración (ej. estadística de Hosmer–Lemeshow), y Exactitud de clasificación (ej. sensibilidad, especificidad, valores predictivos).	Pubmed

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Dong et al.	2022	Estudio transversal para desarrollar un modelo de evaluación de riesgo sin análisis de laboratorio. Reportado siguiendo la declaración TRIPOD.	El modelo ML (Xgboost) mostró mejor discriminación y calibración que el LR. Para detección de DM y pre-DM en validación: ML (AUC-ROC = 0.822, AUC-PR = 0.496). La duración del sueño y el tiempo de actividad recreativa vigorosa fueron factores de riesgo modificables significativos.	AUC-ROC, AUC-PR, Sensibilidad, Especificidad, PPV, NPV. Para detección de DM y pre-DM en validación: ML (AUC-ROC = 0.822).	Pubmed

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Kopitar et al.	2020	Estudio retrospectivo de derivación y validación de modelos predictivos de T2DM. Simulación de datos entrantes continuos en lotes de 6 meses (T6, T12, T18, T24, T30).	No se encontró una mejora clínicamente relevante al emplear modelos basados en Machine Learning sobre los modelos de regresión. Con 30 meses de datos (T30), RF tuvo el RMSE más bajo (0.745), seguido de Glmnet (0.747). Glmnet superó a todos los métodos en AUC en T18–T30 (AUC 0.859 en T30).	Error Cuadrático Medio (RMSE) para la predicción de FPGL. AUC (área bajo la curva ROC) y AUPRC (área bajo la curva Precisión-Recall) para la predicción dicotómica. También se observó la Sensibilidad (SENS).	Pubmed

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Aftab et al.	2021	Propuesta de un novedoso sistema de soporte de decisiones basado en la nube (Cloud-Based Decision Support System, CBD-DSS-FM) que utiliza técnicas de fusión de aprendizaje automático para el diagnóstico temprano de diabetes. Incluye capas de entrenamiento y prueba.	El sistema propuesto de Fusión de ML (FM) logró una tasa de precisión del 95.2% en la fase de prueba (validación). Esto superó a los algoritmos de ML aplicados previamente para el diagnóstico de diabetes.	FM (Prueba): Precisión: 95.2%; Tasa de fallos: 4.8%; Especificidad: 96.03%; Sensibilidad: 93.43%. DT (Prueba): Precisión: 94.9%.	Scopus

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Alkattan et al.	2024	Estudio de validación externa que compara un modelo de ML con la evaluación de riesgo de la Asociación Estadounidense de Diabetes (ADA) para identificar personas con alto riesgo de T2DM.	El modelo de ML (Regresión Logística) fue una herramienta adecuada para la clasificación a nivel poblacional para identificar a aquellos con alto riesgo de T2DM o no diagnosticados. El valor AROC fue de 0.803.	AROC: 0.803 (IC 95%: 0.779–0.826). Punto de Corte ≥ 0.5 : Sensibilidad: 77.94%; Especificidad: 75.13%.	Scopus
		Estudio de investigación para predecir el riesgo de	Las características con mayor efecto en el resultado fueron el IMC, la relación	Se utilizó el Área Bajo la Curva ROC (AUC)	
Lama et al.	2021	diabetes/prediabetes utilizando modelos de aprendizaje automático interpretable	cintura-cadera, la edad, la presión arterial sistólica y diastólica y la herencia de diabetes. Valores altos de estas	combinada con una medida de robustez (S) para la optimización de	Scopus

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
		(SHAP) en datos longitudinales con seguimiento a 10 y 20 años.	características aumentaron el riesgo.	hiperparámetros.	
Naveed et al.	2023	Investigación comparativa entre modelos de aprendizaje profundo y modelos de ML base para predecir diabetes, enfocándose en la incorporación de características temporales (riesgo acumulado).	Los modelos de aprendizaje profundo superaron a los modelos de ML de última generación. CNN-LSTM logró la precisión más alta (91.6%). El azúcar en sangre en ayunas (FBS), A1c e IMC son los predictores claves.	CNN-LSTM: Precisión: 91.6%; Sensibilidad: 85.7%; Especificidad: 95.6%; Precisión (Precision): 93.0%; Puntuación F1: 89.2. LSTM: Precisión:	Scopus

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Patil et al.	2023	Estudio para desarrollar un modelo de apilamiento (stacking) basado en el algoritmo evolutivo NSGA-II para la selección de características y la predicción multi-clase de T2DM (incluyendo el estadio prediabético).	El enfoque de apilamiento NSGA-II superó a los métodos de conjunto convencionales. El modelo propuesto detectó la enfermedad en la etapa prediabética. Conjunto de datos recolectados: Precisión: 88.18%.	91.5%. SVM: Precisión: 90%. Dataset Recolectado: Precisión: 88.18%; Especificidad: 0.94; Sensibilidad: 0.88; Puntuación F1: 0.87. Dataset Pima: Precisión: 81.90%.	Scopus

Autor	Año	Diseño (Tipo de Estudio/Fase)	Resultados (Hallazgos Principales/Conclusión)	Métricas de Rendimiento	Fuente
Firdous et al. (2022)	2022	Artículo de revisión/Encuesta (Uso de algoritmos de machine learning para identificar diabetes)	Los algoritmos de clasificación ML como Support Vector Machine (SVM), K-Nearest Neighbor (KNN) y Random Forest (RF) mostraron la mejor precisión para predecir la diabetes en una etapa temprana.	Precisión, valor ROC, F-score, tiempo de cálculo.	Web of Sciences
Wändell et al.	2024	Estudio de casos y controles con datos retrospectivos (Desarrollo de un modelo predictivo para la diabetes de nuevo diagnóstico en PHC).	Los factores predictivos más importantes fueron la hipertensión (NRI 21.4–29.7%) y la obesidad (NRI 4.8–15.2%). Se confirmaron patrones de riesgo conocidos.	AUC: 0.773 a 0.825. Sensibilidad: 0.648 a 0.855. Especificidad: 0.434 a 0.711.	Web of Sciences

Apéndice B*Lista de Verificación de Metodología PRISMA*

Sección/tema	Ítem #	Lista de ítems
TITLE		
Título	1	Portada Pagina 1
RESUMEN		
Resumen	2	Pagina 5
INTRODUCTION		
Justificación	3	Página 13
Objetivos	4	Página 15
MÉTODOS		
Criterios de elegibilidad	5	Página 26 y 27
Fuentes de información	6	Página 26 Identificación
Estrategia de búsqueda	7	Página 26 Identificación
Proceso de selección	8	Página 26 Cribado
Proceso de recopilación de datos	9	Página 26 Cribado
Lista de datos	10a	Página 28-36 Inclusion
	10b	No aplica

Sección/tema	Ítem #	Lista de ítems
Estudio y valoración del riesgo de sesgo	11	Paginas 30-33 Evaluacion del riesgo de sesgo
Medidas de efecto	12	Página 28-36 Inclusion
Métodos de síntesis	13a	No aplica
	13b	No aplica
	13c	No aplica
	13d	No aplica
	13e	No aplica
	13f	No aplica
Informar de la evaluación del sesgo	14	No aplica
Evaluación de la certeza	15	No aplica
RESULTADOS		
Selección de los estudios	16a	Pagina 26 figura 1
	16b	Pagina 26 Elegibilidad
Características del estudio	17	Apendice A Tabla sintetica
Riesgo de sesgo en los estudios	18	Pagina 30-33

Sección/tema	Ítem #	Lista de ítems
Resultados de estudios individuales	19	Tabla 2 pagina 28- 36
Resultados de la síntesis	20a	Paginas 34-36
	20b	No aplica
	20c	No aplica
	20d	No aplica
Reportar sesgos	21	No aplica
Certeza de la evidencia	22	No aplica
DISCUSIÓN		
Discusión	23a	Paginas 34-38
	23b	Paginas 30-33 Evaluación de sesgo
	23c	No aplica
	23d	Paginas 39-41 Conclusiones y recomendaciones
MÁS INFORMACIÓN		
Registro y protocolo	24a	No aplica
	24b	No aplica
	24c	No aplica

Sección/tema	Ítem #	Lista de ítems
Apoyo	25	No aplica
Conflicto de intereses competitivos	26	No aplica
Disponibilidad de datos, código y otros materiales	27	No aplica
