

05. APRENDIZAJE AUTOMÁTICO PARA PREDECIR DIABETES: UNA NUEVA HERRAMIENTA DE PREVENCIÓN.

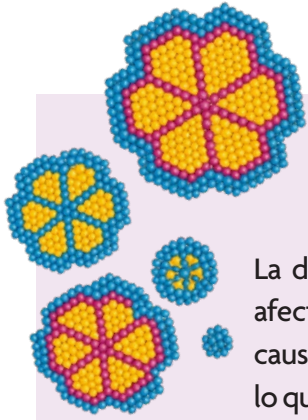
Moisés de Jesús Juárez Álvarez
Estudiante de Ingeniería en
Informática, CUCEI, UDG
ORCID: <https://orcid.org/0009-0007-3571-6729>

Marcos Emmanuel Mariscal Díaz
Estudiante de Licenciatura en Física,
CUCEI, UDG.
ORCID: <https://orcid.org/0009-0003-3353-472X>

Gabriela Etelvina Talavera Ramos
Estudiante del Posgrado en
Cirugía General
ORCID: <https://orcid.org/0009-0003-9263-1052>

Andrea Elizabeth Becerril Talavera
Egresada de la Licenciatura en Ciencia
de Materiales, UDG.
ORCID: <https://orcid.org/0009-0000-0646-1375>





RESUMEN

La diabetes tipo 2 es un problema que se incrementa en la salud pública afectando a millones de personas. En México, representa una de las principales causas de muerte. El diagnóstico tradicional suele requerir múltiples pruebas, lo que puede incrementar costos y retrasar la detección. Este estudio exploró cómo el "aprendizaje automático" (un tipo de inteligencia artificial) identifica los factores de riesgo más importantes para simplificar el diagnóstico. Analizamos datos de pacientes internacionales y los validamos con datos de un hospital en Durango, México. Los resultados identificaron dos pistas clave: la hemoglobina glicosilada y la hipertensión. Los modelos de inteligencia artificial lograron una alta precisión (hasta 94% en datos internacionales y 72% en la población mexicana). Estos hallazgos abren la puerta a protocolos de diagnóstico más simples y accesibles. ¿Podría una "alarma" digital anticipar la diabetes antes de que aparezca? Este estudio explora esa posibilidad.

INTRODUCCIÓN (EL RETO DE LA DETECCIÓN TEMPRANA)

Para entrenar nuestro modelo, usamos dos grupos de datos. Primero, un grupo grande de 1,879 pacientes internacionales con 46 variables clínicas (edad, exámenes de laboratorio, etc.). Luego, para asegurarnos de que los resultados fueran útiles en nuestro contexto, usamos un segundo grupo de 1,199 pacientes mexicanos del Hospital Dr. Santiago Ramón y Cajal en Durango.

El proceso [Figura 1] fue similar a enseñar a un detective:

1. Encontrar las pistas: Le dimos al sistema todos los datos y usamos técnicas computacionales avanzadas (LASSO^[1], Boruta, SelectKBest y Random Forest). Estos métodos funcionan como detectives que revisan todas las pistas y se quedan solo con las más importantes para resolver el misterio: ¿quién tiene mayor riesgo de desarrollar diabetes?

2. Construir el "cerebro" (el modelo): Probamos diferentes tipos de "cerebros" artificiales (Redes Neuronales^[2] y Máquinas de Vectores de Soporte^[3] (SVM)) para ver cuál aprendía mejor a predecir quién podría desarrollar diabetes en el futuro.

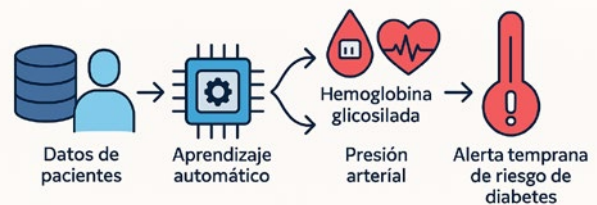


Figura 1. Resumen visual del modelo de aprendizaje automático que identifica factores clave para anticipar el riesgo de diabetes tipo 2.

¹ LASSO, Boruta, SelectKBest, Random Forest: Son los nombres de técnicas estadísticas que se usan para la "selección de características". Ayudan a la computadora a revisar docenas de variables y elegir solo las más importantes o predictivas, eliminando el "ruido".

² Redes Neuronales: Es un modelo inspirado en el cerebro humano. Utiliza "neuronas" artificiales en capas para detectar patrones muy complejos en los datos que otros métodos más simples no pueden ver.

³ Máquinas de Vectores de Soporte (SVM): Es un algoritmo que funciona encontrando la "línea" o frontera óptima que mejor separa dos grupos (por ejemplo, pacientes con diabetes y pacientes sin diabetes).



HALLAZGOS PRINCIPALES: DOS PISTAS SON CLAVE

Los resultados fueron sorprendentes. Sin importar la técnica o la población, dos factores resultaron ser los más importantes, para “predecir” lo que pasaría:

1. La Hemoglobina Glicosilada: Este es un análisis de sangre que muestra el promedio del nivel de azúcar en los últimos tres meses. Es una de las mejores formas de saber si una persona ha mantenido un buen control de su glucosa.
2. La Hipertensión Arterial: Este resultado confirma que existe una relación muy cercana entre la diabetes y la salud del corazón. Tener la presión elevada es un signo importante que también ayuda a predecir complicaciones.

El nivel de glucosa en ayunas [Figura 2] (comúnmente medido en mg/dL, aunque el sistema internacional usa mmol/L) también resultó ser un factor importante. El rendimiento de los modelos fue muy bueno, donde el grupo internacional, el modelo "Random Forest" alcanzó una excelente precisión del 94%. En la población mexicana, las "Redes Neuronales" se adaptaron bien, logrando un 72% de precisión [Figura 3]. Aunque el 72% es más bajo, sigue siendo un resultado muy valioso para un sistema de salud, ya que utiliza menos datos clínicos de entrada.

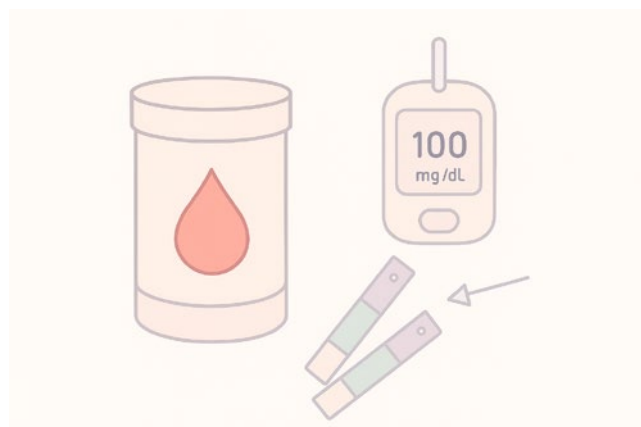


Figura 2. Ejemplo del control de glucosa en sangre usando tiras reactivas y un glucómetro.

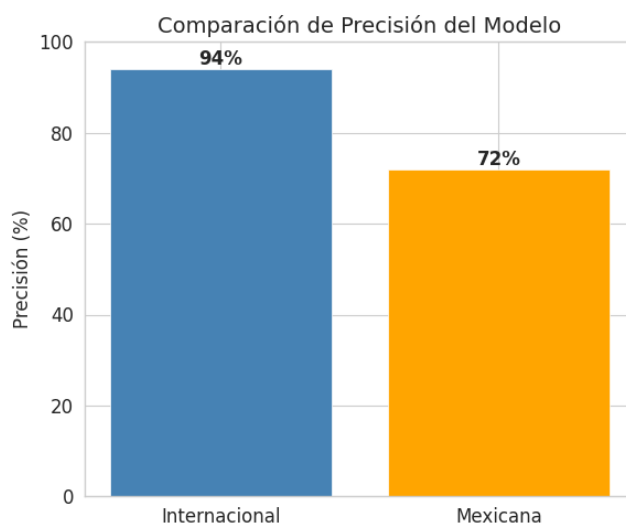


Figura 3. Comparación de precisión del modelo en la población internacional (azul, 94%) y la población mexicana (naranja, 72%). (Elaborada por los autores)

IMPLICACIONES: ¿PARA QUÉ SIRVE ESTO EN LA PRÁCTICA?

Estos hallazgos no son solo números; tienen efectos directamente en la atención médica. Primero, sugieren que podríamos desarrollar protocolos de detección mucho más simples y baratos; en lugar de una serie completa de pruebas, un médico podría enfocarse primero en la hemoglobina glicosilada y revisar la hipertensión del paciente. Segundo, la buena precisión de estos

modelos (nuestra "alarma") facilita la identificación de pacientes en riesgo antes de que se sientan enfermos. Esta detección rápida es la clave de la prevención: permite realizar cambios en el estilo de vida (dieta, ejercicio) o tratamiento antes de que la enfermedad cause daños irreversibles.



CONCLUSIÓN

Este estudio demuestra que la inteligencia artificial es una herramienta poderosa para identificar los factores importantes de la diabetes.

La verdadera promesa de esta tecnología es lo que puede hacer en nuestra vida diaria. Imaginemos su potencial en clínicas rurales en México, donde el acceso a laboratorios y exámenes completos son limitados. Un modelo mejorado, que requiera menos pruebas (como la hemoglobina glicosilada y la presión arterial), podría ser una herramienta de evaluación para que el personal de salud priorice a los pacientes que necesitan atención más urgente.

Aunque se necesita más investigación para mejorar los modelos para poblaciones específicas, estos resultados son un avance hacia una medicina más preventiva, efectiva y económica, enfocándonos en la prevención de la enfermedad en lugar de tratar sus complicaciones.

Referencias Bibliográficas

1. ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, et al. Classification and diagnosis of diabetes: standards of care in diabetes—2023. *Diabetes Care*. 2023;46(Suppl 1):S19–40. doi:10.2337/dc23-S002. PMID: 36507649. [Internet]. [citado 10 Nov 2025]. Disponible en: https://diabetesjournals.org/care/article/46/Supplement_1/S19/148056/2-Classification-and-Diagnosis-of-Diabetes
- 2- Instituto Nacional de Estadística y Geografía (INEGI). Estadísticas de defunciones registradas 2022 [Internet]. México: INEGI; 2023 [citado 10 Nov 2025]. Disponible en: <https://www.inegi.org.mx/programas/mortalidad/>
- 3- International Diabetes Federation. IDF Diabetes Atlas. 10th ed. [Internet]. Brussels (Belgium): IDF; 2021 [citado 10 Nov 2025]. Disponible en: <https://www.diabetesatlas.org>
- 4- Fregoso-Aparicio L, Noguez J, Montesinos L, García-García JA, et al. Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. *Diabetology & Metabolic Syndrome*. 2021;13:148. doi:10.1186/s13098-021-00767-9. [Internet]. [citado 10 Nov 2025]. Disponible en: <https://dmsjournal.biomedcentral.com/articles/10.1186/s13098-021-00767-9>

