

# Modelo predictivo de aprendizaje automático para identificar el estado metabólico en niños mexicanos, utilizando HOMA-IR y la actividad enzimática de la amilasa

Karen E. Villagrana-Bañuelos,<sup>1\*</sup> Carlos E. Galván-Tejada,<sup>1\*</sup> Antonio García-Domínguez,<sup>1</sup> Erika Acosta-Cruz,<sup>2</sup> Miguel A. Vázquez-Moreno<sup>2</sup> y Miguel Cruz-López<sup>3</sup>

<sup>1</sup>Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Zacatecas, Zacatecas; <sup>2</sup>Departamento de Biotecnología, Universidad Autónoma de Coahuila, Saltillo, Coahuila; <sup>3</sup>Hospital de Especialidades "Dr. Bernardo Sepúlveda Gutiérrez", Unidad de Investigación Médica en Bioquímica, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Ciudad de México. México

## Resumen

**Antecedentes:** La obesidad infantil es un problema mundial de salud, por ser un factor de riesgo para desarrollar enfermedades como el síndrome metabólico y diabetes. Actualmente el identificar estas enfermedades ya establecidas, es relativamente fácil para los profesionales de la salud con el apoyo de estudios de laboratorio, la tendencia mundial en cuanto a salud implica actuar antes de que la enfermedad se establezca. **Objetivo:** El objetivo de este estudio es identificar si la actividad de amilasa total es útil para predecir que pacientes desarrollaran síndrome metabólico o diabetes. **Material y métodos:** Utilizando una base de datos con 101 pacientes mexicanos, considerando el valor del modelo de evaluación de insulinoresistencia (HOMA-IR) como variable diagnóstica en tres grupos menor de 2 normal, entre 2 y 5 con riesgo metabólico y mayor de 5 como diabetes, así como el valor de la actividad enzimática de la amilasa. Se utilizó Random forest (RF) como método de aprendizaje automático. **Resultados:** El modelo RF obtuvo los siguientes resultados: AUC 0.7075, especificidad 0.7619, sensibilidad 0.7142 y exactitud 0.7500. **Conclusiones:** Se concluye que es factible con estas variables y RF, contar con un modelo de predicción que contribuya a identificar este tipo de pacientes en el periodo prepatogénico.

**PALABRAS CLAVE:** Diabetes. HOMA-IR. Niños mexicanos. Estado metabólico. Aprendizaje automático.

## Machine learning predictive model to identify metabolic status in Mexican children, using HOMA-IR and amylase enzymatic activity

### Abstract

**Background:** Childhood obesity is a global health problem, as it is a risk factor for developing diseases such as metabolic syndrome and diabetes. At present, identifying these already established diseases is relatively easy for health professionals with the support of laboratory studies. The global trend in health involves acting before the disease is established. **Objectives:** The objective of this study is to identify whether total amylase activity is useful to predict which patients will develop metabolic syndrome or diabetes. **Material and methods:** Using a database with 101 Mexican patients, considering the value of the homeostasis model assessment insulin resistance as a diagnostic variable in three groups < 2 normal, between 2 and 5 with metabolic risk and > 5 as diabetes, as well as the value of the amylase enzymatic activity. Random forest (RF) was used as a machine learning method. **Results:** The RF model obtained the following results: area under the curve 0.7075, specificity

#### \*Correspondencia:

Karen E. Villagrana-Bañuelos  
E-mail: kvillagrana@uaz.edu.mx  
Carlos E. Galván-Tejada  
E-mail: ericgalvan@uaz.edu.mx

Fecha de recepción: 26-11-2024

Fecha de aceptación: 25-06-2025

DOI: 10.24875/GMM.M25001040

Gac Med Mex. 2025;161:565-572

Disponible en PubMed

www.gacetamedicademexico.com

0016-3813/© 2025 Academia Nacional de Medicina de México, A.C. Publicado por Permanyer. Este es un artículo open access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

0.7619, sensitivity 0.7142, and accuracy 0.7500. **Conclusions:** It is concluded that with these variables and RF, it is feasible to have a prediction model that contributes to identifying this type of patients in the prepathogenic period.

**KEYWORDS:** Diabetes. Homeostasis model assessment insulin resistance. Mexican children. Metabolic status. Machine learning.

## Introducción

El *machine learning* es una subdisciplina de la inteligencia artificial que permite a un sistema aprender simulando la inteligencia humana. Existen diversos métodos de *machine learning*, siendo los principales el supervisado y el no supervisado; el primero requiere datos etiquetados y el segundo no. El objetivo de estos modelos es construir una herramienta matemática que, a partir de variables que contienen información relevante sobre una persona, permita calcular la probabilidad de que está presente o no una determinada condición, en este caso, una enfermedad. Estos métodos de *machine learning* se han utilizado en diferentes áreas; no obstante, en medicina han tenido un gran impacto, al identificar relaciones entre variables que a veces resultan difíciles de establecer con los métodos estadísticos tradicionales. Las enfermedades consideradas de etiología multifactorial se ven más favorecidas debido a la complejidad de comprender su origen. A medida que se conoce el papel fisiopatológico, será más sencillo encontrar mecanismos de prevención, tratamiento y limitación del daño; entre estas enfermedades se encuentran el síndrome metabólico y la diabetes. La relevancia de atacar esta enfermedad es tal que existen numerosos trabajos que han combinado *machine learning* con obesidad y diabetes infantil.<sup>1</sup>

En los últimos años, se ha observado un aumento en la prevalencia de la obesidad infantil a nivel mundial. En 2022, la Organización Mundial de la Salud señaló que, entre los 5 y 19 años de edad, había más de 30 millones de personas con sobrepeso y 160 millones de personas con obesidad.<sup>2</sup>

Se sabe que el sobrepeso y la obesidad predisponen, entre muchas otras patologías, al síndrome de insulinoresistencia, la diabetes *mellitus* tipo 2 y las enfermedades cardiovasculares.

La insulinoresistencia, el síndrome metabólico y la diabetes se encuentran estrechamente relacionados y comparten factores comunes: principalmente el sobrepeso y la obesidad, además de la predisposición

genética, el sedentarismo, los hábitos de vida y enfermedades concomitantes, entre otros.

De acuerdo con Ros Pérez y Medina-Gómez,<sup>3</sup> la insulinoresistencia sobreviene cuando la capacidad de esta hormona para llevar a cabo sus funciones se ve reducida, incluida la introducción de moléculas de glucosa en la célula. Durante este periodo, el organismo mantiene funciones compensatorias, como la hipersecreción de la hormona. Este periodo de compensación puede denominarse prediabético; su diagnóstico es difícil en muchos casos, ya que la glucemia permanece dentro de parámetros normales. La compensación alcanza un punto de falla pancreática de las células beta, con la consecuente disminución de la insulina, elevación de la glucemia, una etapa en la que el síndrome metabólico y la diabetes se hacen evidentes.

Por esta razón, algunos autores destacan el índice de evaluación de la insulinoresistencia por modelo de homeostasis (*HOMA-IR*) como una alternativa temprana para el diagnóstico del síndrome de insulinoresistencia. Este se obtiene de la suma de la glucosa en ayunas (mg/dl) y la insulina en ayunas ( $\mu$ U/ml) dividida entre la constante 405.<sup>4</sup>

Asmasy et al.<sup>5</sup> mencionan que el *HOMA-IR* es un valor que puede utilizarse en niños con obesidad para detectar de manera temprana sensibilidad y resistencia a la insulina, con el fin de prevenir o retrasar la aparición de la diabetes.

Existe una diferencia en los parámetros para clasificar a los pacientes según *HOMA-IR*. Algunos autores mencionan claramente que un valor  $\geq 2.5$  se asocia con obesidad y, por lo tanto, con riesgo metabólico. Pelin et al.<sup>6</sup> utilizan valores límite de *HOMA-IR* para resistencia a la insulina:  $\leq 2$  normal, 2–5 tolerancia a la glucosa disminuida y  $> 5$  diabetes.

La amilasa es una enzima que ayuda a metabolizar carbohidratos, especialmente el almidón. De manera indirecta, esta enzima puede utilizarse para evaluar su función en el metabolismo de los carbohidratos, por lo que se ha relacionado con la búsqueda de biomarcadores de diabetes.<sup>7</sup> Existen estudios que han vinculado la actividad enzimática de la amilasa con la obesidad.

Locia-Morales et al señalan que un nivel elevado de amilasa sérica total se relaciona con menores respuestas de glucosa e insulina en la prueba oral de almidón en niños mexicanos con obesidad.<sup>8</sup>

Por ello, el objetivo del presente estudio fue determinar si es posible predecir el estado metabólico a través de la actividad enzimática de la amilasa, utilizando HOMA-IR como variable de salida mediante un método de *machine learning*. El objetivo fue identificar pacientes pediátricos en riesgo metabólico para disponer en el futuro de herramientas cuantitativas que permitan al personal de salud brindar medidas de prevención específicas antes de que se establezca un diagnóstico como la diabetes.

## Material y métodos

Esta sección proporciona una descripción del conjunto de datos utilizado en el estudio; además, se detalla el método de clasificación aplicado y las métricas de evaluación empleadas.

### Descripción de los datos

Los datos utilizados en este estudio fueron proporcionados por la Unidad de Investigación Médica en Bioquímica del Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social. El estudio se realizó de acuerdo con la Declaración de Helsinki y el protocolo fue aprobado por el Comité de Ética del Instituto Mexicano del Seguro Social y la Comisión Nacional de Investigación Científica (R-2016-785-097). Todos los pacientes participantes son de nacionalidad mexicana y firmaron consentimiento informado previo al inicio del estudio.

La base de datos incluyó información de 101 individuos, de los cuales 49 eran varones y 52 mujeres, con un rango de edad de 5 a 12 años. Se evaluaron 20 variables, mostradas en la figura 1, tabla 1 y tabla 2.

### Métodos de clasificación

El *Random Forest* (RF), registrado por Breiman y Cutler, es un método de *machine learning* supervisado, considerado de tipo *ensemble*, ya que utiliza múltiples árboles de decisión que se generan de manera aleatoria e independiente. Se ha utilizado para clasificar con base en la mayoría de votos de cada árbol de decisión y, a su vez, emplea el promedio para realizar regresiones.<sup>9</sup> El modelo puede representarse con la siguiente ecuación:

**Tabla 1. Significado de las variables**

Variable	Descripción
Sexo	0 para Niña, 1 para Niño
Edad	Edad en años
IMC	Índice de masa corporal
IMC-z	Puntaje z del índice de masa corporal
Categoría de peso corporal; 0 para peso normal, 1 para sobrepeso u obesidad	Categoría de peso corporal Presión arterial sistólica (mmHg)
DIABP (mmHg)	Presión arterial diastólica (mmHg)
Glucosa (mg/dl)	Glucosa (mg/dl)
CT (mg/dl)	Colesterol total (mg/dl)
TG (mg/dl)	Triglicéridos (mg/dl)
HDL (mg/dl)	Colesterol de lipoproteínas de alta densidad (mg/dl)
LDL (mg/dl)	Colesterol de lipoproteínas de baja densidad (mg/dl)
Insulina (µg/ml)	Insulina (µg/ml)
HOMA-IR	Evaluación del modelo homeostático de resistencia a la insulina
AMY1 (UI/L)	Actividad de amilasa salival (UI/L)
AMY2 (UI/L)	Actividad de amilasa pancreática (UI/L)
AMYt (UI/L)	Actividad total de amilasa (UI/L)
CNVs-AMY1	Variaciones en el número de copias del gen de la amilasa salival
CNVs-AMY2	Variaciones en el número de copias del gen de la amilasa pancreática

IMC: índice de masa corporal.

$$\hat{f}(x) = \frac{1}{B} \sum_{b=0}^B f_b(x)$$

Donde:

- $\hat{f}(x)$ : predicción agregada o promedio para la entrada  $x$  de todos los árboles del bosque.
- $B$ : n° total de árboles en el bosque (definido habitualmente por el usuario).
- $f_b(x)$ : predicción del árbol de decisión  $b$ -ésimo para la entrada  $x$ .

La base de este método son los árboles de clasificación, en los cuales divisiones binarias dividen recursivamente el árbol en nodos terminales lo más homogéneos posible. RF, según el problema, emplea un n° variable de árboles, y además selecciona de

**Tabla 2.** Medidas de tendencia central y dispersión de las variables

Variable	Media	Mediana	Moda	Desviación estándar
Sexo			1	
Edad	9.38	9.35	8	± 1.602
IMC	21.24	21.2	17	± 4.207
IMC-z			1	
Peso corporal 0 normopeso 1 obesidad			1	
PAD (mmHg)	65.92	63	58	± 11.091
Glucosa (mg/dl)	79.25	80	81	± 7.699
CT (mg/dl)	161.69	160	161	± 28.692
TG (mg/dl)	123	111	81	± 62.66
HDL (mg/dl)	48.31	46	35	± 10.968
LDL (mg/dl)	94.66	91	80	± 20.955
Insulina (µg/ml)	9.14	7	3	± 7.233
HOMA-IR	1.81	1.3	1	± 1.515
AMY1 (UI/l)	40.82	38.5	39	± 19.562
AMY2 (UI/l)	22.31	20.3	19	± 8.551
AMYt (UI/l)	63.13	60	41	± 22.494
CNVs-AMY1	7.04	6	6	± 2.642
CNVs-AMY2	1.98	2	2	± 0.51

IMC: índice de masa corporal; PAD: presión arterial, diastólica; CT: colesterol total; TG: triglicéridos; HDL: lipoproteína de alta densidad; LDL: lipoproteína de baja densidad; HOMA-IR: modelo de homeostasis para la evaluación de resistencia a la insulina; AMY1: amilasa salival; AMY2: amilasa pancreática; AMYt: amilasa total; CNVs-AMY1: número de copias de variaciones del gen de amilasa salival; CNVs-AMY2: número de copias de variaciones del gen de amilasa pancreática.

manera aleatoria un número de variables de entre el total para buscar la mejor división de un nodo, con el fin de descorrelacionar los árboles y reducir la varianza del bosque completo. Según Breiman, esto se logra mediante la pureza utilizando el índice de Gini.<sup>10</sup>

Este método destaca por su solidez a la hora de resolver problemas con grandes o pequeños volúmenes de datos. Entre sus desventajas se menciona la dificultad de interpretación frente a los árboles de decisión, así como su mayor tiempo de entrenamiento.

No obstante, RF ha resultado útil en diversas aplicaciones en salud, debido a su capacidad para manejar datos complejos, identificar variables relevantes e irrelevantes, y controlar mejor el *overfitting*, logrando mayor precisión.<sup>11-14</sup>

## Métricas de evaluación

La curva característica operativa del receptor (curva ROC) muestra gráficamente la relación entre la tasa de verdaderos positivos (sensibilidad o *recall*) y la tasa de falsos positivos (equivalente a 1 menos la especificidad).

Se interpreta que cuanto más cerca se encuentre la curva ROC de la esquina superior izquierda, mejor será el desempeño global del modelo. El área bajo la curva (AUC) resume cuán buena es una prueba independientemente del umbral, siendo 1 la prueba perfecta, mientras que un AUC de 0,5 representa el peor escenario, equivalente a tomar una decisión al azar<sup>15</sup>, como se presenta en la figura 2.

La especificidad alude al porcentaje de personas que obtienen un resultado negativo en una prueba para una enfermedad específica dentro de un grupo de personas que no presentan la enfermedad<sup>16</sup>. Se calcula con la siguiente ecuación 2:

$$\text{Especificidad} = \frac{\text{Verdaderos Negativos}}{\text{Verdaderos Negativos} + \text{Falsos Positivos}}$$

La sensibilidad es la proporción de pruebas verdaderamente positivas entre todos los pacientes con una condición.<sup>15</sup> La ecuación de sensibilidad es la siguiente:<sup>3</sup>

$$\text{Sensibilidad} = \frac{\text{Verdaderos Positivos}}{\text{Verdaderos Positivos} + \text{Falsos Negativos}}$$

La exactitud es la proporción entre las muestras correctamente clasificadas y el número total de muestras en el conjunto de evaluación. En aplicaciones médicas, puede no ser tan conveniente, ya que el modelo podría asignar todas las muestras a la clase prevalente y, de esta manera, alcanzar niveles cercanos a 1, donde la clasificación es perfecta; mientras que 0 representa no clasificar correctamente ninguna muestra.<sup>17</sup> Se calcula con la siguiente ecuación 4:

$$\text{Exactitud} = \frac{\text{Verdaderos Positivos} + \text{Verdaderos Negativos}}{\text{Verdaderos Positivos} + \text{Verdaderos Negativos} + \text{Falsos Positivos} + \text{Falsos Negativos}}$$

Sexo	Nina	Edad	IMC	IMCz	PesoCorp	PesoNormal	PresArt	Diast_mmHg	Gluc_mg	CT_mgdL	TG_mgdL	cHDL_mgdL	cLDL_mgdL	Insulin_ugmt	HOMA1r	AMY1_UIL	AMY2_UIL	AMYr_UIL	CNVs_AMY2
1		6.53	22.9	1.22			1	68	85	180	83	60	105	5.4	1.1	79.2	37.7	116.9	9
0		5.96	19.8	1.94			1	97	81	219	89	65	133	8.7	1.7	30.7	29.3	60	4
0		7.88	20.8	2.16			1	66	80	142	109	45	80	19.1	3.8	31.4	30.6	62	4
0		7.83	23.3	2.15			1	67	80	172	162	53	97	9.8	1.9	62.5	25.5	88	11
1		7.51	24.4	1.28			1	66	69	229	111	67	138	3.5	0.6	33.1	18.8	51.9	8

Figura 1. Muestra de 5 observaciones de la base de datos utilizada.

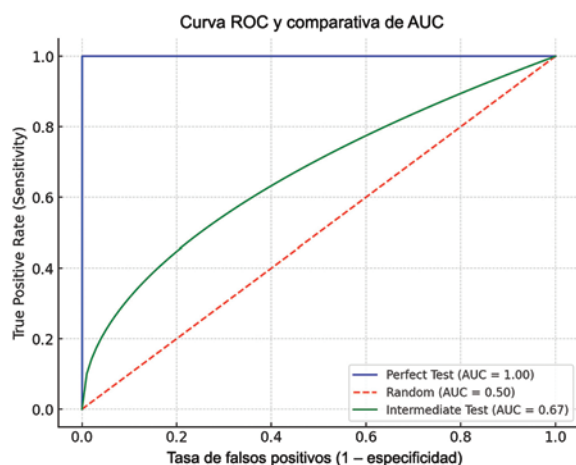


Figura 2. Evaluación del área bajo la curva ROC (AUC). En azul, el valor ideal para clasificar correctamente los casos de pacientes. La línea roja discontinua indica un AUC de 0.5, equivalente a un modelo no óptimo (probabilidad del 50% de error o acierto). En verde, un ejemplo de AUC de 0.67, intermedio entre el ideal y el umbral, representando un valor aceptable para el problema evaluado.

## Experimentación y resultados

Este estudio incluyó a un total de 101 pacientes, de los cuales 49 eran varones y 52 mujeres; la media de edad fue de  $9,3 \pm 1,6$  años, con un mínimo de 5,9 y un máximo de 12,9; 45 pacientes tenían peso normal y 56 obesidad, como se presenta en la figura 3.

Se calculó el HOMA-IR para cada paciente, considerando tres grupos: pacientes con valores  $< 2$  fueron clasificados como normales, valores entre 2-5 como pacientes con riesgo metabólico y  $> 5$  como diabéticos.

La actividad total de amilasa (UI/l) se midió agrupando a los pacientes según los 3 grupos anteriores: pacientes normales o sin riesgo con un promedio de  $66,34 \pm 22,89$ , pacientes con riesgo metabólico con  $55,62 \pm 21,01$  y pacientes con diabetes con  $61,4 \pm 17,84$ , como se muestra en la figura 4.

La base de datos se preparó como parte del pre-procesamiento y posteriormente se dividió en el 70% y en el 30%. Con el 70% de los datos se implementaron algoritmos RF en el paquete R para entrenar un

modelo de clasificación multiclase (clase 0: pacientes normales, clase 1: pacientes en riesgo metabólico y clase 2: pacientes con diabetes), de acuerdo con HOMA-IR. El modelo de *machine learning* RF fue luego sometido a validación, utilizando el 30% restante de los datos como prueba ciega. Se calcularon el AUC, la exactitud, la sensibilidad y la especificidad para determinar si era posible identificar pacientes en riesgo metabólico considerando la actividad enzimática total de la amilasa. Los resultados se muestran en la tabla 3 y la figura 5.

## Discusión

Los modelos de clasificación basados en *machine learning* aplicados a la medicina han aumentado en los últimos años, demostrando su relevancia en la identificación de biomarcadores para diagnóstico, pronóstico y prevención. Uno de los métodos más potentes en problemas de clasificación multiclase es el algoritmo RF, elegido por sus propiedades, que incluyen la capacidad de evitar el *overfitting* (cuando el modelo solo identifica individuos con patrones muy específicos) y su amplia aplicación en problemas médicos con resultados óptimos.<sup>18</sup>

Es destacable que este modelo de clasificación, propuesto con una sola variable, logra una clasificación con un AUC de 0,70. Según el AUC, existe una capacidad moderada de discriminar entre casos y controles. Esto es aceptable porque mejora la toma de decisiones en ausencia de datos cuantitativos, que de otra manera serían producto del azar (probabilidad de 0.5). La sensibilidad de 0.71 indica que el modelo identifica correctamente a 7 de cada 10 pacientes con la condición, lo cual es importante para evitar el sobrediagnóstico. Por su parte, la especificidad de 0.76 muestra que se identifican correctamente 7.6 de cada 10 casos negativos. Finalmente, la exactitud indica que el modelo acierta en el 75% de los casos positivos y negativos combinados. En consecuencia, puede afirmarse que este modelo tiene un poder predictivo clínico moderado a bueno, con un balance adecuado entre sensibilidad y especificidad.

Visión general de la demografía del estudio

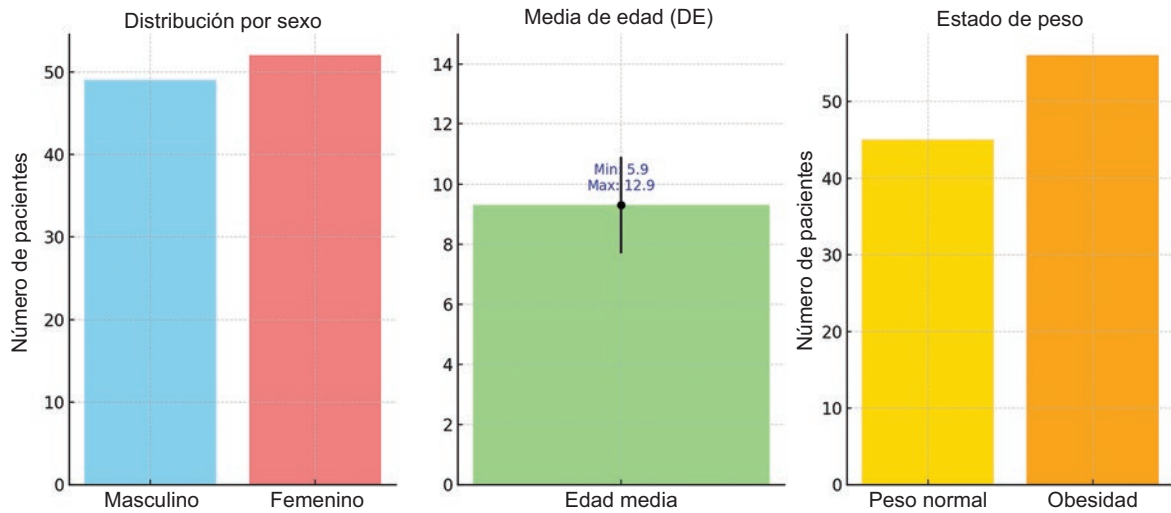


Figura 3. Distribución de variables demográficas descriptivas. Se observa distribución equitativa por sexo, valores de edad media, mínima y máxima, así como la distribución por peso (normal u obesidad).

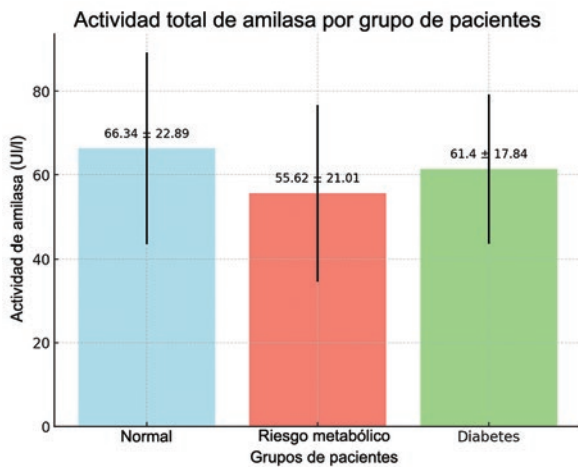


Figura 4. Comparativa de la actividad total de amilasa (U/l) entre los distintos grupos de pacientes: normales, con riesgo metabólico y diabéticos.

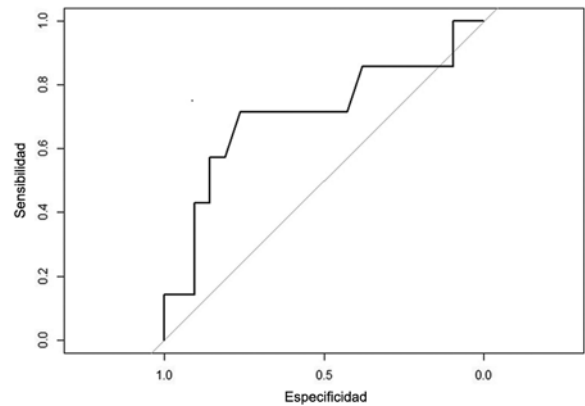


Figura 5. Curva ROC del modelo RF en la prueba ciega.

Se debe mencionar que la validación del modelo se obtuvo en una fase ciega, es decir, el 30% de los datos permanecieron ocultos durante la etapa de entrenamiento. Posteriormente, se solicitó al modelo clasificar sin conocer la variable (HOMA-IR) y se comparó su predicción con los valores reales, calculando aciertos y errores con base en cada métrica. Esta fase ciega es fundamental, ya que se asemeja al uso del modelo en otra población distinta, lo cual ayuda a reducir el *overfitting* y permite su extrapolación.

Tabla 3. Resultados en la etapa ciega del modelo de random forest

Métrica	Valor
AUC	0.7075
Especificidad	0.7619
Sensibilidad	0.7142
Exactitud	0.7500

AUC: área bajo la curva.

Se consideró la posibilidad de explorar datos sintéticos, dado que en muestras pequeñas el riesgo de *overfitting* aumenta. No obstante, el uso de

técnicas de aumento de datos sintéticos en medicina aún se debate, pues aunque la inteligencia artificial puede contribuir a completar datos, eliminar sesgos y validar hipótesis, también plantea riesgos éticos. Debido a que este estudio se centró en población pediátrica, se decidió trabajar solo con datos reales, aunque se reconoce que esta técnica podría explorarse en trabajos futuros.<sup>19</sup>

La variable utilizada para predecir si el estado metabólico de un paciente era normal, de riesgo o diabético fue la actividad enzimática total de la amilasa, lo cual ya se ha asociado a otras investigaciones en este mismo contexto. Un metanálisis halló diferencias significativas en las concentraciones de amilasa en pacientes con sobrepeso, obesidad, síndrome metabólico y diabetes; niveles bajos de amilasa y lipasa séricas se asociaron significativamente con diabetes.<sup>12</sup>

Pérez-Ros et al señalaron la asociación de la amilasa salival con el diagnóstico de diabetes, planteándola como un posible biomarcador para el control glucémico y el manejo clínico.<sup>20</sup>

Otros autores, como Chaudhari y Hansen, en estudios con primates, concluyeron que la dinámica de la amilasa desempeña un papel importante en la progresión metabólica desde un individuo normal hasta síndrome metabólico y diabetes.<sup>21</sup> También se ha descrito que la actividad sérica de la amilasa es un indicador diagnóstico significativo del síndrome metabólico.<sup>22</sup>

La relación entre la actividad enzimática de la amilasa y el riesgo metabólico refuerza la utilidad de combinar técnicas de *machine learning* para proponer un biomarcador con potencial para identificar a pacientes con mayor riesgo de desarrollar síndrome metabólico y diabetes.

Existen pocos trabajos en la literatura<sup>1</sup> que relacionen el uso de *machine learning* para la detección de diabetes en población infantil, por lo que la relevancia de este estudio contribuye al cuerpo de conocimiento en esta área.

## Conclusiones

Considerando la actividad enzimática de la amilasa total, se puede clasificar con el método RF obteniendo un AUC de 0.70, una especificidad de 0.76 y una sensibilidad de 0.71, lo cual permite identificar casos en > 70%. Los verdaderos positivos y negativos se pueden identificar con un 75% de exactitud, lo cual ofrece la oportunidad de detectar en la población infantil mexicana pacientes con riesgo metabólico, como insulinoresistencia o prediabetes.

Como trabajo futuro, pueden incluirse otras variables que aumenten el rendimiento del modelo y el área bajo la curva ROC, además de incrementar el tamaño de la muestra, lo cual permitirá comparar el modelo con otros más robustos, como las redes neuronales.

A pesar de haber utilizado un método para evitar el *overfitting*, se recomienda replicar estudios con la misma metodología y un mayor número de pacientes, con el fin de mejorar la evaluación del modelo.

La relevancia de atender esta enfermedad es tal que múltiples trabajos han combinado *machine learning* con obesidad y diabetes infantil; no obstante, la mayoría se centra en variables como edad, peso y síntomas, que facilitan el diagnóstico clínico. En este estudio, en cambio, se propone una variable con potencial para predecir el riesgo antes de la aparición de la diabetes.

## Financiamiento

Los autores declaran que no recibieron apoyo financiero para esta investigación.

## Conflicto de intereses

Los autores declaran no tener conflicto de intereses.

## Referencias

- Colmenarejo G. Machine learning models to predict childhood and adolescent obesity: a review. *Nutrients*. 2020;12:2466.
- Ferreira AP, Oliveira CE, França NM. Metabolic syndrome and risk factors for cardiovascular disease in obese children: the relationship with insulin resistance (HOMA-IR). *J Pediatr (Rio J)*. 2007;83:21-6.
- Ros Pérez M, Medina-Gómez G. Obesidad, adipogénesis y resistencia a la insulina. *Endocrinol Nutr*. 2011;58:360-9.
- Almeda-Valdés P, Bello-Chavolla OY, Caballeros-Barragán CR, Gómez-Velasco DV, Viveros-Ruiz T, Vargas-Vázquez A, et al. Índices para la evaluación de la resistencia a la insulina en individuos mexicanos sin diabetes. *Gac Med Mex*. 2018;154(Suppl 2):S50-5.
- Asmasary AA, Artati RD, Ganda IJ. Analysis of glycosylated hemoglobin (HbA1c) level and homeostasis model assessment insulin resistance (Homa-IR) value in obese children. *Int J Health Sci Res*. 2023;13:212-6.
- Pelin AM, Balan G, Stefanescu C, Rosca S, Busila C. New criteria in defining the metabolic syndrome in children - an analysis of the relationship between the hepatic enzymes and the insulin resistance, HOMA-IR, glucose tolerance test in the obese children. *Progr Nutr*. 2022;23:e2021316.
- Palomino C, Motta E, Chipayo C, Cornejo R, Paredes R, Borda A, et al. Correlación entre la glucosa salival con la glucosa de ayuno, la hemoglobina glicada y el péptido-C en personas con diabetes *mellitus* tipo 2. *Acta Méd Peru*. 2023;40:7-14.
- Locia-Morales D, Vázquez-Moreno M, González-Dzib R, Domínguez-Hernández C, Pérez-Herrera A, Robles-Ramírez RJ, et al. Association of total and pancreatic serum amylase enzymatic activity with insulin resistance and the glucose and insulin responses to oral starch test in Mexican children. *Pediatr Obes*. 2022;17:e12965.
- Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.
- Chen X, Ishwaran H. Random forests for genomic data analysis. *Genomics*. 2012;99:323-9.
- Su Q, Liu Q, Lau RI, Zhang J, Xu Z, Yeoh YK, et al. Faecal microbiome-based machine learning for multi-class disease diagnosis. *Nat Commun*. 2022;13:6818.
- Ko J, Cho J, Petrov MS. Low serum amylase, lipase, and trypsin as biomarkers of metabolic disorders: a systematic review and meta-analysis. *Diabetes Res Clin Pract*. 2020;159:107974.

13. Faradibah A, Widyawati D, Syahar AU, Jabir SR, Belluano PL. Comparison analysis of random forest classifier, support vector machine, and artificial neural network performance in multiclass brain tumor classification. *Indones J Data Sci.* 2023;4:55-63.
14. Vohra R, Hussain A, Dudyala AK, Pahareeya J, Khan W. Multi-class classification algorithms for the diagnosis of anemia in an outpatient clinical setting. *PLoS One.* 2022;17:e0269685.
15. Erickson BJ, Kitamura F. Magician's corner: 9. Performance metrics for machine learning models. *Radiol Artif Intell.* 2021;3:e200126.
16. Shreffler J, Huecker MR. Diagnostic testing accuracy: Sensitivity, specificity, predictive values and likelihood ratios. In: *Statpearls.* Treasure Island, FL: Statpearls; 2020.
17. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep.* 2022;12:5979.
18. Alaminos-Fernández AF. Árboles De Decisión En R Con Random Forest. Alicante: Limencop; 2022.
19. Resnik DB, Hosseini M, Kim JJ, Epiphaniou G, Maple C. GenAI synthetic data create ethical challenges for scientists. Here's how to address them. *Proc Natl Acad Sci U S A.* 2025;122:e2409182122.
20. Pérez-Ros P, Navarro-Flores E, Julián-Rochina I, Martínez-Arnau FM, Cauli O. Changes in salivary amylase and glucose in diabetes: a scoping review. *Diagnostics (Basel).* 2021;11:453.
21. Chaudhari UK, Hansen BC. Amylase and lipase levels in the metabolic syndrome and type 2 diabetes: a longitudinal study in rhesus monkeys. *Physiol Rep.* 2024;12:e16097.
22. Zhan F, Chen J, Yan H, Wang S, Zhao M, Zhang S, et al. Association of serum amylase activity and the copy number variation of AMY1/2A/2B with metabolic syndrome in Chinese adults. *Diabetes Metab Syndr Obes.* 2021;14:4705-14.