Artículo original/ Original article
ISSN 2953-6553

WarningSugar: Sistema Integral de Predicción de Riesgo de Diabetes en Adultos jóvenes Mediante Machine Learning y Salud Digital

WarningSugar: An Integrated DiabetesRisk Prediction System for Young Adults Using Machine Learning and Digital Health

Diego Mendez [1] 0009-0006-0138-9021, Jonnathan Tigre [2] 0009-0003-6810-0139, Verónica Chimbo[3] 0000-0003-3470-739X
diego.mendez.est@tecazuay.edu.ec , jonnathanf.tigre.est@tecazuay.edu.ec, veronicachimbo@tecazuay.edu.ec
[1 2 3]Instituto Superior Tecnológico del Azuay, Azuay/Cuenca, Ecuador.

DOI 10.36500/atenas.5.012

Resumen

La diabetes tipo 2 es una de las principales causas de muerte en Ecuador y su detección suele ser tardía, porque las pruebas de tamizaje disponibles requieren laboratorio. Este trabajo presenta WarningSugar, un sistema web de estimación del riesgo de disglucemia (prediabetes o diabetes) basado exclusivamente en 12 variables no invasivas. En respuesta a una limitación de la versión previa, cuya variable objetivo se derivaba de los propios predictores, los modelos se reentrenaron con una etiqueta clínica independiente: hemoglobina glucosilada (HbA1c) ≥ 5.7% o diagnóstico médico previo de diabetes, medida en 8 526 adultos de 20 a 69 años de la encuesta NHANES 2015-2018. Se compararon Regresión Logística, SVM, Árbol de Decisión, Random Forest y XGBoost mediante validación cruzada estratificada de 5 pliegues, AUC-ROC, AUC-PR, sensibilidad, especificidad, Brier Score y error de calibración esperado (ECE). Los modelos alcanzaron un AUC-ROC entre 0.785 y 0.816, sin diferencias significativas entre los cuatro mejores. La Regresión Logística obtuvo el mayor AUC-ROC (0.816 ± 0.014; IC95%: 0.798-0.833), la mejor calibración sin ajuste (ECE = 0.018) y un desempeño estable en validación temporal (AUC = 0.806), por lo que se seleccionó para el despliegue. Los umbrales de riesgo (bajo < 25%, moderado 25-64%, alto ≥ 65%) se fijaron para lograr 90% de sensibilidad y 89% de especificidad en sus respectivos extremos. WarningSugar constituye una herramienta de tamizaje preventivo, accesible y reproducible, que requiere validación en población ecuatoriana antes de su uso clínico.

Abstract

Type 2 diabetes is one of the leading causes of death in Ecuador, and its detection is often delayed because available screening tests require laboratory infrastructure. This paper presents WarningSugar, a web-based system that estimates the risk of dysglycemia (prediabetes or diabetes) using only 12 non-invasive variables. To address a limitation of the previous version, whose target variable was derived from the predictors themselves, the models were retrained with an independent clinical label: glycated hemoglobin (HbA1c) ≥ 5.7% or a previous medical diagnosis of diabetes, measured in 8,526 adults aged 20 to 69 from the NHANES 2015-2018 survey. Logistic Regression, SVM, Decision Tree, Random Forest, and XGBoost were compared using stratified 5-fold cross-validation, ROC-AUC, PR-AUC, sensitivity, specificity, Brier score, and expected calibration error (ECE). The models achieved ROC-AUC values between 0.785 and 0.816, with no significant differences among the four best performers. Logistic Regression obtained the highest ROC-AUC (0.816 ± 0.014; 95% CI: 0.798-0.833), the best uncalibrated calibration (ECE = 0.018), and stable performance under temporal validation (AUC = 0.806), and was therefore selected for deployment. Risk thresholds (low < 25%, moderate 25-64%, high ≥ 65%) were set to achieve 90% sensitivity and 89% specificity at their respective ends. WarningSugar is an accessible and reproducible preventive screening tool that requires validation in the Ecuadorian population before clinical use.

Palabras Claves – Diabetes Mellitus Tipo 2, Tamizaje Masivo, Aprendizaje Automático, Telemedicina.
Keywords – Diabetes Mellitus, Type 2, Mass Screening, Machine Learning, Telemedicine.

Recibido: 2026-08-25 Aprobado tras revisión: 2026-10-28

I. INTRODUCCIÓN

La diabetes mellitus tipo 2 es un problema de salud pública creciente en América Latina. En Ecuador es la segunda causa de muerte, con una prevalencia estimada de 6.5% (Instituto Nacional de Estadística y Censos [INEC], 2024), y a escala mundial su prevalencia en adultos se ha duplicado desde 1990 (NCD Risk Factor Collaboration [NCD-RisC], 2024). La ENSANUT-ECU 2012 reportó que el 62.8% de los adultos de 19 a 59 años presentaba sobrepeso u obesidad y que la prevalencia de diabetes aumentaba de forma marcada a partir de los 30 años (Freire et al., 2014); la encuesta STEPS 2018 confirmó la persistencia de la inactividad física, el consumo de alcohol y la adiposidad abdominal en la población de 18 a 69 años (Ministerio de Salud Pública del Ecuador [MSP] et al., 2018).

Los adultos jóvenes son un grupo prioritario: la enfermedad evoluciona durante años sin síntomas, su percepción de riesgo es baja y las pruebas de tamizaje del sistema público (glucosa en ayuno y HbA1c) exigen laboratorio y desplazamiento, barreras que se acentúan en las zonas urbano-marginales y rurales de la Sierra. Un instrumento de autoevaluación no invasivo, accesible desde un teléfono móvil, puede orientar a quienes necesitan una prueba de laboratorio y optimizar los recursos del primer nivel de atención, en línea con el Plan Nacional de Diabetes 2025-2030 (MSP, 2025).

WarningSugar es un sistema web de tamizaje preventivo, no de diagnóstico, que estima la probabilidad de disglucemia a partir de 12 variables que no requieren equipo especializado: edad, sexo, etnia, índice de masa corporal (IMC), circunferencia de cintura, presión arterial sistólica y diastólica, actividad física, consumo de alcohol, tabaquismo, antecedentes familiares de diabetes y diabetes gestacional previa. La versión anterior de este trabajo entrenaba los modelos con una variable objetivo, construida a partir de esos mismos predictores, lo que producía exactitudes artificialmente cercanas al 100%. La presente versión corrige esa limitación evaluando los modelos frente a una medición de laboratorio independiente (HbA1c) en una muestra poblacional real, con métricas de discriminación, calibración y estabilidad del error.

II. MARCO TEÓRICO

Disglucemia y factores de riesgo modificables

La diabetes se diagnostica con glucosa en ayuno ≥ 126 mg/dL o HbA1c ≥ 6.5%, y la prediabetes con

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

glucosa de 100 a 125 mg/dL o HbA1c de 5.7% a 6.4% (American Diabetes Association Professional Practice Committee [ADA], 2024). Este trabajo usa como estándar de referencia la disglucemia (HbA1c ≥ 5.7% o diagnóstico previo de diabetes), porque la prediabetes es la ventana en la que las intervenciones sobre el estilo de vida son más eficaces y porque la diabetes franca es poco frecuente en adultos jóvenes. Los factores de riesgo medibles sin laboratorio incluyen IMC ≥ 25 kg/m², cintura elevada, inactividad física, consumo excesivo de alcohol, tabaquismo, antecedentes familiares, diabetes gestacional previa e hipertensión arterial (ADA, 2024; Organización Mundial de la Salud [OMS], 2024); todos se miden con báscula, tallímetro, cinta métrica o tensiómetro y son modificables.

Evolución de los modelos de predicción y de su margen de error

Los primeros instrumentos de tamizaje no invasivo fueron puntuaciones aditivas derivadas de regresión logística, como el FINDRISC, con AUC superiores a 0.80 para diabetes incidente (Lindström y Tuomilehto, 2003), y la puntuación de la ADA, derivada de NHANES con siete variables (Bang et al., 2009). Ambos estudios documentaron la pérdida de desempeño al aplicarse en poblaciones distintas de la de derivación. El aprendizaje automático elevó la discriminación por encima de 0.90 cuando se incorporan variables de laboratorio, pero los modelos con solo variables no invasivas se sitúan habitualmente entre 0.70 y 0.85 (Dinh et al., 2019), y una revisión de 71 estudios no encontró ventaja del aprendizaje automático sobre la regresión logística cuando la comparación tiene bajo riesgo de sesgo (Christodoulou et al., 2019).

La literatura reciente desplazó el énfasis desde la exactitud puntual hacia tres fuentes de error: la fuga de datos entre predictores y etiqueta, que infla las métricas (Kaufman et al., 2012); la calibración, o concordancia entre la probabilidad predicha y el riesgo observado (Van Calster et al., 2019); y la variabilidad del desempeño entre muestras y en el tiempo (Steyerberg y Vergouwe, 2014). La declaración TRIPOD+AI recoge estas exigencias (Collins et al., 2024) y guía la presente versión. En cuanto a los algoritmos, la regresión logística tiende a producir probabilidades bien calibradas, mientras que los ensambles por bagging (Random Forest) y boosting (XGBoost; Chen y Guestrin, 2016) pueden sesgarlas, lo que se corrige con escalado de Platt o regresión isotónica (Niculescu-Mizil y Caruana, 2005).

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

III. METODOLOGÍA

Diseño, fuente de datos y población

Se realizó un estudio de desarrollo y validación interna de un modelo de predicción diagnóstica, reportado conforme a TRIPOD+AI. Se utilizaron los ciclos 2015-2016 y 2017-2018 de la Encuesta Nacional de Examen de Salud y Nutrición (NHANES) de los Estados Unidos (Centers for Disease Control and Prevention [CDC], 2020), que combina entrevistas estandarizadas, examen físico por personal capacitado y laboratorio centralizado. Se incluyeron los participantes de 20 a 69 años con HbA1c disponible o diagnóstico previo de diabetes (n = 8 526). El análisis se implementó en Python con scikit-learn (Pedregosa et al., 2011) y XGBoost.

NHANES sustituye al conjunto de Kaggle de la versión anterior por dos motivos: allí la variable objetivo se construía con los propios predictores (fuga de datos) y, al usar como etiqueta sus marcadores de laboratorio (glucosa ≥ 126 mg/dL o HbA1c ≥ 6.5%), todos los modelos obtuvieron un AUC-ROC de 0.49 a 0.52, equivalente al azar, lo que indica un origen probablemente sintético.

Variable objetivo y predictores

La variable objetivo se definió como disglucemia: HbA1c ≥ 5.7% o diagnóstico médico previo de diabetes. La HbA1c actúa como estándar de referencia independiente y no se incluye entre los predictores. La prevalencia fue de 41.1% en la muestra y de 11.9% entre los 20 y 29 años; como análisis secundario se evaluó la diabetes (HbA1c ≥ 6.5% o diagnóstico previo; 16.7%). Los 12 predictores, iguales a las preguntas de la aplicación, se describen en la Tabla 1.

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

Tabla 1
Variables predictoras y su origen en NHANES

Variable	Tipo y codificación	Origen NHANES
Edad, IMC, cintura	Numéricas, estandarizadas	RIDAGEYR, BMXBMI, BMXWAIST
Presión sistólica y diastólica	Numéricas, promedio de lecturas (mmHg)	BPXSY1-3, BPXDI1-3
Sexo, etnia, tabaquismo	Nominales, one-hot	RIAGENDR, RIDRETH3, SMQ020/040
Actividad física	Ordinal (baja < moderada < alta)	PAQ605, 620, 650, 665
Consumo de alcohol	Ordinal (nulo < moderado < elevado)	ALQ101/111, ALQ130
Antecedentes familiares; diabetes gestacional	Binarias (0/1; gestacional = 0 en hombres)	MCQ300C, RHQ162

Nota. Los analitos de laboratorio (HbA1c, glucosa, perfil lipídico, GGT y ácido úrico) no se usan como predictores; la HbA1c (LBXGH) solo define la variable objetivo. Elaboración propia.

Preprocesamiento y modelos

El preprocesamiento tuvo cuatro etapas, encapsuladas en un Pipeline ajustado solo con los datos de entrenamiento de cada pliegue: (1) imputación por la mediana (numéricas) y la moda (categóricas); (2) codificación one-hot de las variables nominales sin orden, para que los modelos lineales no supongan un orden inexistente, y ordinal para actividad física y alcohol; (3) estandarización con StandardScaler; y (4) partición estratificada 80/20 (6 820 y 1 706 registros; random_state = 42). Las «cuatro etapas» se refieren a estas transformaciones y no deben confundirse con las 12 variables del formulario. Los faltantes fueron bajos (1.4% a 4.8%) salvo en diabetes gestacional (13.2%) y alcohol (20.2%); para este último se comparó la imputación por la moda con una categoría «desconocido», la exclusión de la variable y el análisis de casos completos (Sterne et al., 2009).

Se entrenaron cinco modelos con complejidad restringida para evitar sobreajuste: Regresión Logística (L2, C = 1.0), SVM con kernel RBF (probabilidades por escalado de Platt), Árbol de Decisión (profundidad 5; 50 observaciones por hoja), Random Forest (300 árboles; 20 por hoja) y XGBoost (300 árboles; profundidad 3; tasa 0.05).

Protocolo de evaluación y selección

Se emplearon validación cruzada estratificada de 5 pliegues (media ± DE e IC95%) de exactitud, sensibilidad, especificidad, AUC-ROC, AUC-PR, más informativo con clases desbalanceadas (Saito y Rehmsmeier, 2015), y Brier Score; evaluación en el conjunto de prueba con IC95% bootstrap del AUC,

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

matrices de confusión y error de calibración esperado (ECE; Guo et al., 2017); curvas de confiabilidad y calibración post-hoc (Platt e isotónica) de Random Forest y XGBoost; y, para el modelo elegido, validación cruzada repetida (5 × 10), validación temporal (entrenamiento 2015-2016, prueba 2017-2018) y análisis por subgrupos. Como comparador se evaluó la regla determinista de la versión anterior (IMC > 25, cintura elevada o actividad física baja). El modelo de despliegue se eligió por discriminación, luego calibración y, ante diferencias no significativas, interpretabilidad. Los umbrales de riesgo se fijaron para lograr cerca de 90% de sensibilidad en el corte inferior y 90% de especificidad en el superior.

Arquitectura del sistema

WarningSugar es una aplicación web en Streamlit con un formulario de 12 pantallas, una por variable (el IMC se calcula a partir del peso y la talla), cada una con una imagen orientadora. Al iniciar se carga el Pipeline serializado, lo que garantiza transformaciones idénticas en entrenamiento y producción. La probabilidad se traduce en tres niveles con código de color: bajo (< 25%), moderado (25% a 64%) y alto (≥ 65%), con recomendaciones preventivas. Cada evaluación se almacena en PostgreSQL (Supabase) y puede exportarse a Excel para el seguimiento longitudinal. Los datos están disponibles en https://wwwn.cdc.gov/nchs/nhanes/ y el código y el modelo en https://github.com/tigreraph/Data_Warning_Sugar.

IV. RESULTADOS

Comparación de modelos

La Tabla 2 presenta los resultados de validación cruzada. El AUC-ROC osciló entre 0.785 y 0.816. La Regresión Logística obtuvo el mayor AUC-ROC (0.816 ± 0.014), el mayor AUC-PR (0.734) y el menor Brier (0.172), aunque sus intervalos se solapan con los de XGBoost, Random Forest y SVM, por lo que las diferencias entre estos cuatro modelos no son relevantes. El Árbol de Decisión tuvo el menor desempeño y la mayor variabilidad (DE de la sensibilidad = 0.061). Todos superaron ampliamente la línea base del AUC-PR, igual a la prevalencia (0.41).

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

Tabla 2

Validación cruzada estratificada de 5 pliegues (media ± DE; IC95% del AUC-ROC)

Modelo	AUC-ROC (IC95%)	AUC-PR	Sensib.	Especif.	Brier
Regresión Logística *	0.816 ± 0.014 (0.798-0.833)	0.734 ± 0.022	0.659 ± 0.020	0.804 ± 0.010	0.172 ± 0.006
SVM (RBF)	0.807 ± 0.016 (0.787-0.827)	0.728 ± 0.028	0.649 ± 0.022	0.798 ± 0.013	0.176 ± 0.008
Árbol de Decisión	0.785 ± 0.018 (0.763-0.807)	0.684 ± 0.019	0.645 ± 0.061	0.777 ± 0.036	0.185 ± 0.007
Random Forest	0.809 ± 0.017 (0.788-0.830)	0.730 ± 0.020	0.640 ± 0.029	0.803 ± 0.011	0.176 ± 0.006
XGBoost	0.813 ± 0.015 (0.795-0.831)	0.732 ± 0.024	0.667 ± 0.021	0.787 ± 0.010	0.174 ± 0.007

Nota. Sensibilidad y especificidad con umbral 0.5. * Modelo seleccionado. Elaboración propia.

En el conjunto de prueba (n = 1 706; 701 casos) los resultados se replicaron (Tabla 3, Figura 1). Con el umbral 0.5, la Regresión Logística identificó 456 de 701 casos (sensibilidad 0.650) y 798 de 1 005 no casos (especificidad 0.794).

Tabla 3

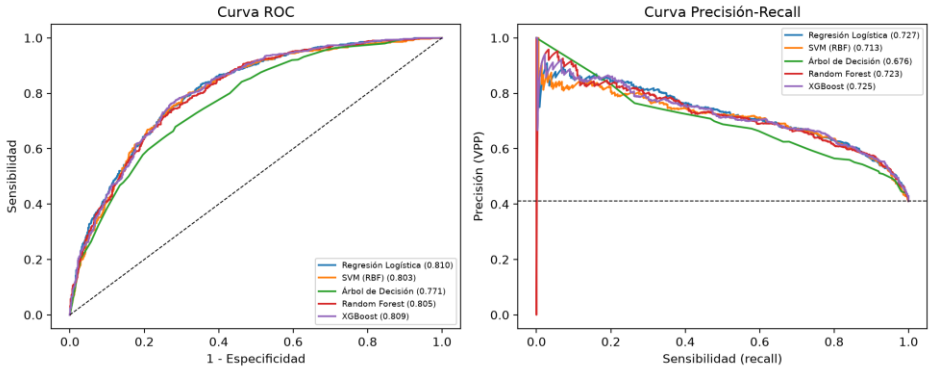
Desempeño en el conjunto de prueba (n = 1 706)

Modelo	AUC-ROC (IC95% bootstrap)	AUC-PR	Brier	ECE	VN / FP / FN / VP
Regresión Logística *	0.810 (0.789-0.830)	0.727	0.175	0.018	798 / 207 / 245 / 456
SVM (RBF)	0.803 (0.782-0.823)	0.713	0.179	0.051	810 / 195 / 259 / 442
Árbol de Decisión	0.771 (0.750-0.793)	0.676	0.191	0.033	794 / 211 / 283 / 418
Random Forest	0.805 (0.785-0.825)	0.723	0.177	0.037	796 / 209 / 244 / 457
XGBoost	0.809 (0.788-0.829)	0.725	0.175	0.033	791 / 214 / 241 / 460

Nota. Matrices de confusión con umbral 0.5 (VN: verdaderos negativos; FP: falsos positivos; FN: falsos negativos; VP: verdaderos positivos). ECE: error de calibración esperado. Elaboración propia.

Figura 1

Curvas ROC (izquierda) y precisión-recall (derecha) en el conjunto de prueba; entre paréntesis, el área bajo la curva



Nota. Elaboración propia.

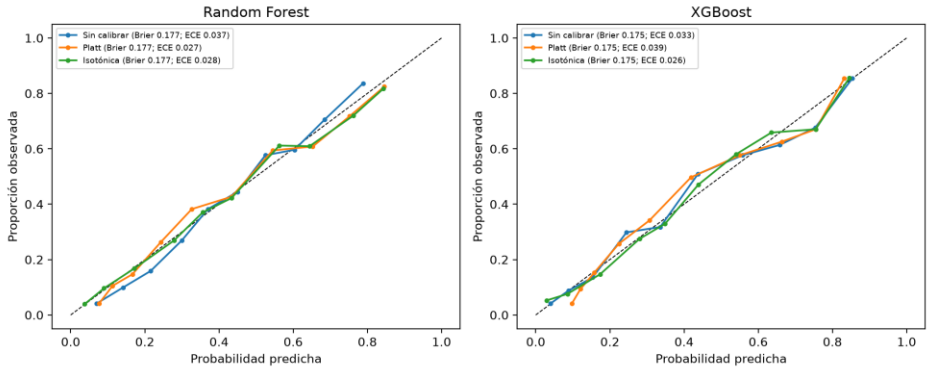
Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

Calibración y selección del modelo de despliegue

La versión anterior sostenía que XGBoost polariza sus probabilidades y que Random Forest las distribuye mejor. Con una etiqueta independiente esa diferencia desaparece (Figura 2): sin calibrar, Random Forest y XGBoost tienen un ECE de 0.037 y 0.033; la regresión isotónica los reduce a 0.028 y 0.026, y el escalado de Platt a 0.027 y 0.039, sin cambios en el Brier (0.177 y 0.175) ni en el AUC. La Regresión Logística, sin ajuste, presenta la mejor calibración (ECE = 0.018).

Figura 2

Curvas de confiabilidad de Random Forest y XGBoost sin calibrar y con calibración de Platt e isotónica



Nota. Elaboración propia.

Por lo tanto, no se sostiene la superioridad de Random Forest ni de XGBoost. Se eligió la Regresión Logística porque obtuvo la mayor discriminación, la mejor calibración sin ajuste y es interpretable. Los predictores más asociados fueron la diabetes gestacional previa (OR = 2.98), la edad (OR = 2.73 por DE), los antecedentes familiares (OR = 1.78), el IMC (OR = 1.56 por DE) y la cintura (OR = 1.21 por DE).

Umbrales de estratificación del riesgo

Con el umbral inferior de 25%, el 89.7% de los casos queda en los niveles moderado o alto y el valor predictivo negativo del nivel bajo es 0.880; con el superior de 65%, el 89.5% de los no casos queda fuera del nivel alto, cuyo valor predictivo positivo es 0.747. La prevalencia observada crece de forma monótona entre niveles (Tabla 4, Figura 3). Los umbrales anteriores (40% y 70%) solo daban una sensibilidad de 0.760 y se reemplazaron.

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

Tabla 4

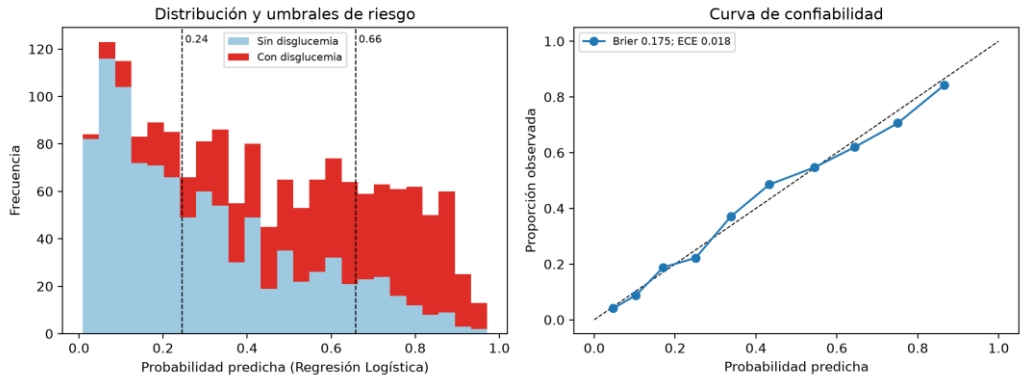
Niveles de riesgo en el conjunto de prueba (Regresión Logística)

Nivel	Probabilidad	Participantes, n (%)	Con disglucemia, n	Prevalencia observada
Bajo	< 25%	598 (35.1%)	72	12.0%
Moderado	25% a 64%	689 (40.4%)	316	45.9%
Alto	≥ 65%	419 (24.6%)	313	74.7%

Nota. Elaboración propia.

Figura 3

Distribución de probabilidades de la Regresión Logística con los umbrales de riesgo (izquierda) y curva de confiabilidad (derecha)



Nota. Elaboración propia.

Estabilidad del error, subgrupos y contraste con el estándar de referencia

En la validación cruzada repetida (50 estimaciones), el AUC-ROC fue 0.815 ± 0.009 (percentiles 2.5-97.5: 0.797-0.829) y el Brier 0.172 ± 0.004. En la validación temporal (entrenamiento 2015-2016, n = 4 390; prueba 2017-2018, n = 4 136) el AUC fue 0.806, el Brier 0.178 y el ECE 0.028, pese a que la prevalencia aumentó de 39.9% a 42.4%. Por subgrupos (Tabla 5), la discriminación fue homogénea entre sexos y grupos étnicos (AUC de 0.789 en afroamericanos a 0.822 en asiáticos; 0.797 en mexicano-americanos y 0.810 en otros hispanos). En el grupo de 20 a 29 años el AUC fue 0.758, en parte por la menor prevalencia y la menor variabilidad de la edad, su predictor más fuerte.

Tabla 5

Desempeño de la Regresión Logística por subgrupos (conjunto de prueba)

Subgrupo	n	Prevalencia	AUC-ROC	Brier
20 a 29 años	307	11.7%	0.758	0.095
30 a 44 años	487	28.1%	0.783	0.162
45 a 69 años	912	57.9%	0.723	0.208

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

Hombres	819	42.9%	0.797	0.182
Mujeres	887	39.5%	0.821	0.168

Nota. Elaboración propia.

Todos los resultados se contrastan con la HbA1c de laboratorio, estándar diagnóstico de la ADA (2024). La regla determinista anterior clasificó como positivos al 81.8% de los participantes, con sensibilidad de 0.904 y especificidad de solo 0.242; con la misma sensibilidad, la Regresión Logística alcanza una especificidad de 0.499, lo que reduce en cerca de un tercio los falsos positivos. Para diabetes (análisis secundario) el AUC fue 0.827. Finalmente, el AUC-ROC apenas varió según el manejo de los faltantes de alcohol (moda, categoría «desconocido» y exclusión: 0.809 ± 0.017; casos completos: 0.804 ± 0.009), por lo que la imputación por la moda no sesga el desempeño.

V. DISCUSIÓN

El hallazgo central es metodológico: las exactitudes de 99% a 100% de la versión anterior reflejaban fuga de datos, no capacidad predictiva. Frente a una medición de laboratorio independiente, la discriminación se sitúa en torno a 0.81, coherente con las puntuaciones no invasivas consolidadas (Bang et al., 2009; Lindström y Tuomilehto, 2003) y con modelos de aprendizaje automático sin laboratorio (Dinh et al., 2019). Es el desempeño esperable de un instrumento que no sustituye la prueba diagnóstica, sino que prioriza a quién realizarla.

Desde la perspectiva biomédica, las asociaciones son plausibles: la edad refleja el deterioro progresivo de la célula beta y de la sensibilidad a la insulina; la diabetes gestacional y los antecedentes familiares, la susceptibilidad genética y metabólica; y el IMC y la cintura, la adiposidad general y visceral, principal determinante de la resistencia a la insulina. La actividad física, el alcohol y el tabaquismo, autorreportados, aportaron poco, probablemente por error de medición. Desde la perspectiva estadística, el modelo es estable frente a las fluctuaciones aleatorias (DE del AUC de 0.009 en 50 particiones), se mantiene entre ciclos de recolección y está bien calibrado; su equivalencia con los ensambles coincide con Christodoulou et al. (2019), por lo que la calibración y la interpretabilidad son criterios de selección más relevantes que diferencias marginales de discriminación.

Los factores externos de la recolección condicionan los resultados: NHANES mide antropometría y presión arterial con protocolos estandarizados, mientras que WarningSugar depende del autorreporte,

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

cuyo mayor error podría reducir el desempeño en uso real; además, la HbA1c puede alterarse por anemia o hemoglobinopatías. Las limitaciones son: el modelo proviene de población estadounidense (aunque incluye 2 472 participantes hispanos con desempeño similar), por lo que requiere validación externa y recalibración en Ecuador; no se aplicaron los pesos muestrales de NHANES; el diseño es transversal, de modo que estima disglucemia presente y no riesgo futuro; y el número de participantes de 20 a 25 años es reducido (n = 916).

VI. CONCLUSIONES

WarningSugar estima la probabilidad de disglucemia a partir de 12 variables no invasivas con una discriminación de 0.81 frente a la HbA1c de laboratorio, comparable a la de las puntuaciones de riesgo de referencia. La corrección de la fuga de datos fue determinante: el desempeño realista es de 0.81 y no de 100%. La Regresión Logística ofrece la mejor combinación de discriminación, calibración e interpretabilidad, sin diferencias significativas frente a Random Forest y XGBoost, y sus umbrales de 25% y 65% permiten una estratificación con 90% de sensibilidad y 89% de especificidad en sus extremos. El modelo es estable entre particiones, ciclos de recolección y subgrupos, aunque algo menos discriminante entre los 20 y 29 años.

Como trabajo futuro se plantea validar y recalibrar el modelo con adultos jóvenes ecuatorianos, idealmente con medición de HbA1c en establecimientos del primer nivel de atención del Azuay, integrarlo con los sistemas del Ministerio de Salud Pública y evaluar su utilidad clínica mediante curvas de decisión.

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

VII. REFERENCIAS BIBLIOGRÁFICAS

American Diabetes Association Professional Practice Committee. (2024). 2. Diagnosis and classification of diabetes: Standards of care in diabetes—2024. *Diabetes Care, 47*(Suppl. 1), S20-S42. https://doi.org/10.2337/dc24-S002

Bang, H., Edwards, A. M., Bomback, A. S., Ballantyne, C. M., Brillon, D., Callahan, M. A., Teutsch, S. M., Mushlin, A. I., & Kern, L. M. (2009). Development and validation of a patient self-assessment score for diabetes risk. *Annals of Internal Medicine, 151*(11), 775-783. https://doi.org/10.7326/0003-4819-151-11-200912010-00005

Centers for Disease Control and Prevention. (2020). *National Health and Nutrition Examination Survey: Data 2015-2016 and 2017-2018* [Conjunto de datos]. National Center for Health Statistics. https://wwwn.cdc.gov/nchs/nhanes/

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. En *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology, 110*, 12-22. https://doi.org/10.1016/j.jclinepi.2019.02.004

Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A.-L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., . . . Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ, 385*, Artículo e078378. https://doi.org/10.1136/bmj-2023-078378

Dinh, A., Miertschin, S., Young, A., & Mohanty, S. D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Medical Informatics and Decision Making, 19*, Artículo 211. https://doi.org/10.1186/s12911-019-0918-5

Freire, W. B., Ramírez-Luzuriaga, M. J., Belmont, P., Mendieta, M. J., Silva-Jaramillo, K., Romero, N., Sáenz, K., Piñeiro, P., Gómez, L. F. y Monge, R. (2014). *Tomo I: Encuesta Nacional de Salud y Nutrición de la población ecuatoriana de cero a 59 años. ENSANUT-ECU 2012*. Ministerio de Salud Pública e Instituto Nacional de Estadística y Censos.

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. En *Proceedings of the 34th International Conference on Machine Learning* (Vol. 70, pp. 1321-1330). PMLR. https://proceedings.mlr.press/v70/guo17a.html

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

Instituto Nacional de Estadística y Censos. (2024). *Estadísticas de causas de muerte 2024*. https://www.ecuadorencifras.gob.ec/

Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data, 6*(4), 1-21. https://doi.org/10.1145/2382577.2382579

Lindström, J., & Tuomilehto, J. (2003). The diabetes risk score: A practical tool to predict type 2 diabetes risk. *Diabetes Care, 26*(3), 725-731. https://doi.org/10.2337/diacare.26.3.725

Ministerio de Salud Pública del Ecuador. (2025). *Estrategia de prevención de enfermedades no transmisibles y Plan Nacional de Diabetes 2025-2030*.

Ministerio de Salud Pública del Ecuador, Instituto Nacional de Estadística y Censos y Organización Panamericana de la Salud. (2018). *Encuesta STEPS Ecuador 2018: Vigilancia de enfermedades no transmisibles y factores de riesgo*. Ministerio de Salud Pública del Ecuador.

NCD Risk Factor Collaboration. (2024). Worldwide trends in diabetes prevalence and treatment from 1990 to 2022: A pooled analysis of 1108 population-representative studies with 141 million participants. *The Lancet, 404*(10467), 2077-2093. https://doi.org/10.1016/S0140-6736(24)02317-1

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. En *Proceedings of the 22nd International Conference on Machine Learning* (pp. 625-632). Association for Computing Machinery. https://doi.org/10.1145/1102351.1102430

Organización Mundial de la Salud. (2024). *Informe mundial sobre la diabetes*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830. https://www.jmlr.org/papers/v12/pedregosa11a.html

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE, 10*(3), Artículo e0118432. https://doi.org/10.1371/journal.pone.0118432

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ, 338*, Artículo b2393. https://doi.org/10.1136/bmj.b2393

Steyerberg, E. W., & Vergouwe, Y. (2014). Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *European Heart Journal, 35*(29), 1925-1931.

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28

https://doi.org/10.1093/eurheartj/ehu207

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. *BMC Medicine, 17*, Artículo 230. https://doi.org/10.1186/s12916-019-1466-7

Recibido: 2026-08-25, Aprobado tras revisión: 2026-10-28