

Predicción de glucosa en niños con DMT1 en ventanas de 30 minutos mediante XGBoost y simulación fisiológica



Colaboración

Olaf Gustavo Sandoval Vázquez; Raquel Ochoa Ornelas; Julio Alberto García Rodríguez, Tecnológico Nacional de México / Instituto Tecnológico de Ciudad Guzmán

Fecha de recepción: 23 de septiembre de 2025

Fecha de aceptación: 28 de septiembre de 2025

RESUMEN: La Diabetes Mellitus Tipo 1 (DMT1) en niños requiere un monitoreo constante debido a la complejidad glucémica de esta población. Este trabajo presenta un modelo predictivo basado en el algoritmo XGBoost, entrenado con datos simulados de 10 pacientes pediátricos virtuales mediante el entorno DMMS.R. El modelo estima niveles futuros de glucosa en ventanas de 30 minutos. Los resultados exhiben un desempeño robusto, con un RMSE de 5.48 mg/dL, MAE de 3.31 mg/dL y un coeficiente de determinación R^2 de 0.9949. Los resultados presentados apoyan la utilidad de la propuesta como una herramienta computacional de apoyo en el tratamiento personalizado de la DMT1 pediátrica.

PALABRAS CLAVE: Diabetes Mellitus Tipo 1, XGBoost, Modelo predictivo, Herramienta computacional.

ABSTRACT: Type 1 Diabetes Mellitus (T1DM) in children requires constant monitoring due to the glycemic complexity of this population. This study presents a predictive model based on the XGBoost algorithm, trained with simulated data from 10 virtual pediatric patients using the DMMS.R environment. The model estimates future glucose levels in 30-minute windows. The results exhibit robust performance with an RMSE of 5.48 mg/dL, MAE of 3.31 mg/dL, and a coefficient of determination R^2 of 0.9949. The results presented support the usefulness of the proposal as a computational tool to support personalized treatment of pediatric T1DM.

KEYWORDS: Type 1 Diabetes Mellitus, XGBoost, Predictive Model, Computational Tool.

INTRODUCCIÓN

La Diabetes Mellitus Tipo 1 (DMT1) se caracteriza por un déficit grave de insulina, la cual es la hormona más importante para la regulación de la glucosa sanguínea. Todas las formas de Diabetes aparecen por una destrucción autoinmune y progresiva de las células beta en el páncreas. Sin embargo, la velocidad de destrucción celular es considerablemente más acelerada en lactantes, niños y adolescentes en comparación con los adultos, lo que agrava el curso clínico y complica su manejo terapéutico. Esta pérdida funcional desencadena desequilibrios metabólicos significativos y en ausencia de una intervención oportuna, puede conducir a complicaciones graves en el corto y largo plazo [1]. En las últimas décadas, el impacto de la DMT1 ha cobrado mayor relevancia como problema de salud pública. Datos recientes estiman que más de 8.4 millones de personas viven con esta condición en el mundo, y se proyecta que la cifra podría superar los 13 millones para el año 2040. Las desigualdades geográficas, económicas y políticas agravan el panorama [2]. En este sentido, se vuelve esencial replantear las estrategias de monitoreo y control de la enfermedad, especialmente en población pediátrica. El avance de tecnologías digitales, el acceso a grandes volúmenes de datos y el uso de modelos de aprendizaje automático han abierto nuevas oportunidades para anticipar alteraciones glucémicas sin recurrir exclusivamente a métodos invasivos. Algoritmos como el XGBoost han demostrado capacidad para detectar patrones complejos en variables como la glucosa, la insulina y la actividad física [3], permitiendo diseñar sistemas predictivos más sensibles a la dinámica fisiológica individual. Esta línea de investigación se dirige hacia propuestas con un enfoque preventivo, personalizado y accesible para el tratamiento de DMT1 [4]. Al adaptar estos modelos a la fisiología infantil, se han identificado características únicas, como una mayor sensibilidad a la insulina, mayor volumen de distribución relativo al peso corporal y una tasa de absorción de glucosa más elevada, lo que exige ajustes específicos en la calibración del sistema [5]. Incluso factores como la actividad física pueden alterar significativamente la respuesta a la insulina, modificando los parámetros fisiológicos [6]. Esto confirma la necesidad de integrar parámetros fisiológicos relevantes en herramientas tecnológica y computacionales orientadas al apoyo clínico.

Esta investigación forma parte de un proyecto interdisciplinario en ingeniería biomédica e inteligencia artificial. El enfoque metodológico combina simulación computacional, diseño experimental y análisis estadístico. La principal contribución es el desarrollo de un modelo predictivo que integra simulación fisiológica y aprendizaje automático para estimar niveles de glucosa con un horizonte de 30 minutos. En la sección de materiales y métodos se describe el proceso de simulación fisiológica y el entrenamiento del modelo XGBoost. En los resultados se presentan las métricas de desempeño obtenidas. Finalmente, en las conclusiones se discute el

potencial como herramienta de apoyo para el monitoreo personalizado de la DMT1 pediátrica.

MATERIALES Y MÉTODOS

La sección de materiales y métodos describe el proceso para desarrollar y evaluar el modelo predictivo propuesto. Se detallan las etapas de simulación fisiológica mediante el entorno DMMS.R (del inglés: Diabetes Mellitus Metabolic Simulator for Research), la construcción de la base de datos, el diseño computacional para el entrenamiento del modelo XGBoost y la evaluación del desempeño mediante métricas especializadas. Esta estructura metodológica permite comprender cómo se obtienen los resultados presentados en este manuscrito.

Simulación de Datos: Entorno DMMS.R

Los datos fueron generados con el simulador DMMS.R distribuido por el grupo Épsilon [7]. El núcleo de esta herramienta computacional fue desarrollado por los trabajos en conjunto de la Universidad de Padova en Italia y la Universidad de Virginia en Estados Unidos de América [5-8]. El DMMS.R reproduce las dinámicas fisiológicas glucosa-insulina de pacientes virtuales adultos, adolescentes o niños DMT1. Además, pueden configurarse una gran diversidad de tratamientos de control glucémico que integren suministro de insulina, escenarios de alimentación y monitorización continua de glucosa en una interfaz de Matlab® y Simulink de MathWorks Inc. [9-12].

Construcción de la Base de Datos

Se simularon perfiles glucémicos correspondientes a 10 sujetos niños virtuales, junto con un perfil promedio, bajo 5 protocolos distintos de ingesta de carbohidratos a lo largo de un día. Las simulaciones incluyeron sólo la administración de insulina en bolos y consumo de comidas. Se excluyeron las configuraciones ejercicio físico para reducir la complejidad inicial. Se obtuvieron 72,000 datos de glucosa correspondientes a un día completo de 1440 minutos por paciente niño y por protocolo.

Las variables consideradas como predictores fueron: tiempo en minutos, carbohidratos totales, carbohidratos por periodo, insulina en bolo, insulina en el cuerpo y glucosa medida en mg/dL como variable objetivo. La Tabla 1 resume los protocolos alimentarios empleados, expresados en cinco horarios de simulación.

Tabla 1. Cantidad de imágenes obtenidas por fuente.

P	B-7	S-12	M-15	S-19	D-23
1	35	10	70	20	10
2	25	5	40	10	5
3	25	5	15	5	15
4	20	5	35	25	5
5	10	10	25	10	5

Fuente: Elaboración propia.

Diseño Computacional y Entrenamiento Del Modelo

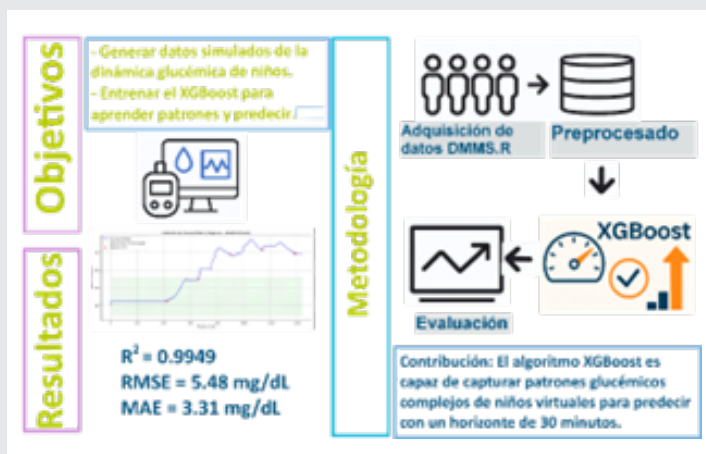


Figura 1. Diagrama general de la propuesta.

Fuente: Elaboración propia.

Los trabajos del presente manuscrito siguen un diseño experimental y computacional estructurado en las siguientes fases: simulación in silico, extracción de variables, preprocesamiento, partición del conjunto de datos para entrenamiento, validación y prueba, entrenamiento del modelo predictivo, evaluación con métricas estándar y visualización de resultados.

La Figura 1 resume de forma esquemática todas las etapas de este estudio. El modelo predictivo fue entrenado con el algoritmo XGBoost, dada su efectividad en problemas de regresión no lineal en contextos médicos [13-14]. La validación cruzada se utilizó para ajustar los hiperparámetros y reducir el riesgo de sobreajuste. El desarrollo se implementó en Python 3.10 con bibliotecas especializadas como XGboost, Pandas, Numpy, TensorFlow y Matplotlib.

Evaluación del modelo

La evaluación del modelo se basa en un enfoque de regresión supervisada. Se estimaron niveles de glucosa con un horizonte de 30 minutos, simulando condiciones de monitoreo en tiempo real. Las métricas utilizadas para la evaluación fueron: Error Absoluto Medio (MAE, del inglés: Mean Absolute Error), Raíz del Error Cuadrático Medio (RMSE, del inglés: Root Mean Square Error), Error Porcentual Absoluto Medio (MAPE, del inglés: Mean Absolute Percentage Error) y Coeficiente de Determinación R².

Se generaron curvas comparativas entre predicciones y datos simulados para diversos escenarios, destacando la capacidad del modelo para adaptarse a la variabilidad interpaciente. Uno de los análisis se enfocó en un día completo de simulación con cinco ingestas, evaluado en el paciente virtual Niño 4.

RESULTADOS

La evaluación del modelo se lleva a cabo empleando los datos en el entorno DMMS.R, específicamente sobre el

paciente virtual Niño 4, bajo cinco protocolos distintos de ingesta de carbohidratos en 24 horas de simulación. Por su relevancia en la dinámica fisiológica se utilizan como variables de entrada: la glucosa proveniente del Monitor Continuo de Glucosa (MCG), la insulina suministrada de manera exógena y la ingesta de carbohidratos.

Para facilitar la interpretación visual de los resultados, este manuscrito presenta únicamente las dinámicas correspondientes al paciente Niño 4. Esta selección permite mostrar con mayor claridad la comparación entre los valores simulados y las predicciones generadas por el modelo XGBoost, evitando la superposición de curvas y la variabilidad interpaciente que dificultarían el análisis gráfico si se incluyeran los diez perfiles simulados.

Desempeño Del Modelo Sobre Datos Simulados

La Figura 2 ilustra la comparación entre los valores de glucosa simulados (línea azul continua) y los valores de predicción por el modelo XGBoost (línea verde continua). El eje horizontal representa el tiempo total de simulación en horas. Iniciando a las 0:00 horas y terminando después de 24 horas. Mientras que el eje vertical muestra la concentración de glucosa en volumen con unidades de mg/dL.



Figura 2: Comparación entre la glucosa simulada y la predicción por el modelo para Niño 4 bajo los 5 protocolos de ingesta.

Fuente: Elaboración propia en entorno Python.

Los indicadores de desempeño se presentan en la Tabla 2. Los resultados reflejan un ajuste altamente preciso, con un coeficiente de determinación R² cercano a 1, lo que indica que el modelo explica prácticamente toda la variabilidad observada en los datos simulados. Tanto el MAE como el RMSE se mantuvieron en rangos bajos, confirmando la capacidad del algoritmo para reproducir el comportamiento glucémico minuto a minuto.

Tabla 2. Métricas de evaluación en entrenamiento de XGBoost.

RMSE	5.48 mg/dL
MAE	3.31 mg/dL
MAPE	43.2%
Coeficiente de determinación (R²)	0.9949

Fuente: Elaboración propia.

Estos resultados demuestran que el modelo no sólo sigue las tendencias generales de la glucosa, sino que también captura con precisión los picos posprandiales y los descensos asociados a la acción de la insulina, reproduciendo la dinámica fisiológica descrita en la literatura [5-12].

Simulación De Un Día Con Cinco Ingestas

Con el fin de evaluar la aplicabilidad práctica del modelo XGBoost en un escenario representativo, se simula un día completo con cinco ingestas de carbohidratos. Estas ingestas son completamente diferentes a las que el modelo XGBoost tuvo en el entrenamiento y se distribuyen en los horarios descritos en la Tabla 3.

Tabla 3. Cantidad y horario de ingestas para un día de simulación en un escenario diferente al del entrenamiento.

Horario (hrs)	Cantidad de CHO (gramos)
07:00	15
11:00	38
15:00	25
19:00	8
23:00	32

Fuente: Elaboración propia.

La Figura 3 presenta la curva de predicción de glucosa para un día hacia el futuro en ventanas de 30 minutos usando el modelo XGBoost. Los puntos color rojo indican el horario y monto en donde se llevó a cabo la ingesta de carbohidratos. El área amarilla corresponde al rango normoglucémico (70-180 mg/dL). El área color azul corresponde a un nivel hipoglucémico (menor a 50 mg/dL). El área color rojo ligero es el área de hiperglucemia (mayor a 180 mg/dL). Después de cada ingesta, el modelo estimó aumentos transitorios en los niveles de glucosa, seguidos por descensos asociados al efecto de la insulina activa.

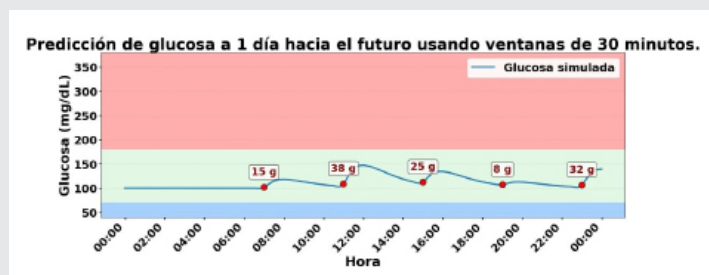


Figura 3: Predicción de glucosa a 1 día hacia el futuro usando ventanas de 30 minutos. Se consideran 5 ingestas con montos distintos a los del entrenamiento.

Fuente: Elaboración propia en entorno Python.

Se pueden identificar claramente que el modelo XGBoost captura la dinámica de los pacientes virtuales ya que en las ingestas de 38 gramos a las 11:00 horas y 32 gramos a las 23:00 horas en cuando el nivel de glucosa se eleva

con mayor pendiente. De hecho, el máximo valor registrado es de 150 mg/dL aproximadamente a las 12:00 horas. No se observaron episodios de hipoglucemia, lo que indica que el modelo evitó predicciones que comprometan la seguridad fisiológica del paciente virtual.

Estos hallazgos refuerzan la utilidad del enfoque basado en XGBoost para anticipar alteraciones glucémicas bajo distintos patrones alimentarios, aportando evidencia de su potencial aplicación en herramientas de apoyo para la toma de decisiones clínicas en diabetes tipo 1.

CONCLUSIONES

El modelo desarrollado que integra simulación fisiológica con el algoritmo XGBoost, logró anticipar con alta precisión la dinámica glucémica minuto a minuto dentro de un día de predicción usando ventanas de 30 minutos en pacientes pediátricos virtuales con DMT1. Los resultados cuantitativos muestran un RMSE de 5.48 mg/dL, un MAE de 3.31 mg/dL y un coeficiente de determinación R2 de 0.9949. Esto demuestra que existe coherencia entre las predicciones y los datos extraídos de las simulaciones fisiológicas con el DMMS.R. Estos valores confirman la capacidad del modelo para reproducir tanto los periodos posprandiales como los descensos mediados por la insulina activa.

A pesar de su desempeño favorable, la naturaleza discreta de XGBoost limita la representación de la fluidez fisiológica de la glucosa. En este sentido, futuras investigaciones podrían incorporar técnicas de suavizado y modelos híbridos como LSTM o Transformers que permitan extender el horizonte de predicción y mejorar la capacidad del sistema frente a la variabilidad interpaciente e intrapaciente.

Este trabajo demuestra un interesante potencial del XGBoost para usarse como un complemento computacional para el personal de salud en la gestión de la DMT1. Sin embargo, las limitaciones de la presente propuesta radican en la aplicación en escenarios clínicos reales. Esto debido a que se necesitan diversas aprobaciones institucionales, de ética y cumplimiento de reglas clínicas. Además, los trabajos futuros deberían incluir nuevas variables fisiológicas que reflejen condiciones más cercanas a la realidad de los pacientes.

AGRADECIMIENTOS

Los autores agradecen al Instituto Tecnológico de Ciudad Guzmán, a través de la División de Estudios de Posgrado por el apoyo brindado para el desarrollo de este manuscrito.

BIBLIOGRAFÍA

[1] National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), "¿Qué es la diabetes?," Abr. 2023. [Online]. Available: <https://www.nidk.nih.gov/health-information/informacion-de-la-salud/diabetes/informacion-general/que-es>.

- [2] I. Ogrotis, T. Koufakis, and K. Kotsa, "Global Burden of Type 1 diabetes: Epidemiology and Public Health Challenges," *Diabetes Therapy*, vol. 14, no. 5, pp. 1007–1019, 2023. doi: <https://doi.org/10.1007/s13300-023-01438-1>.
- [3] P. Xie and J. Xu, "Prediction of diabetes mellitus using XGBoost model," *Applied and Computational Engineering*, vol. 78, pp. 131–141, 2024. doi: <https://doi.org/10.54254/2755-2721/78/20240646>.
- [4] K. Y. Ngiam and I. W. Khor, "Big data and machine learning algorithms for health-care delivery," *The Lancet Oncology*, vol. 20, no. 5, pp. e262–e273, 2019. doi: [https://doi.org/10.1016/S1470-2045\(19\)30149-4](https://doi.org/10.1016/S1470-2045(19)30149-4).
- [5] C. Dalla Man, R. A. Rizza, and C. Cobelli, "Meal simulation model of the glucose-insulin system," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1740–1749, 2007. doi: <https://doi.org/10.1109/TBME.2007.893506>.
- [6] C. Dalla Man, M. Breton, and C. Cobelli, "Physical activity into the meal glucose-insulin model of type 1 diabetes: In silico studies," *Journal of Diabetes Science and Technology*, vol. 3, no. 1, pp. 56–67, 2009. doi: <https://doi.org/10.1177/193229680900300107>.
- [7] The Epsilon Group, *DMMS.R Diabetes Mellitus Metabolic Simulator*, 2024. [Online]. Available: <https://www.epsilon-sim.com/dmmsr>.
- [8] B. P. Kovatchev, M. Breton, C. Dalla Man, and C. Cobelli, "In silico preclinical trials: A proof of concept in closed-loop control of type 1 diabetes," *Journal of Diabetes Science and Technology*, vol. 3, no. 1, pp. 44–55, 2009. doi: <https://doi.org/10.1177/193229680900300106>.
- [9] C. Dalla Man, R. A. Rizza, and C. Cobelli, "The UVA/PADOVA Type 1 Diabetes Simulator: New features," *Journal of Diabetes Science and Technology*, vol. 8, no. 1, pp. 26–34, 2014. doi: <https://doi.org/10.1177/1932296813514502>.
- [10] R. N. Bergman, C. Cobelli, and B. P. Kovatchev, "Quantitative models of insulin action and glucose metabolism," *Annual Review of Physiology*, vol. 83, pp. 55–78, 2021. doi: <https://doi.org/10.1146/annurev-physiol-031820-092853>.
- [11] Ruiz-Velázquez, E., García-Rodríguez, J., Quiroz, G., and Femat, R. (2020). Robust Σ -synthesis: Towards a unified glucose control in adults, adolescents and children with T1DM. *Journal of the Franklin Institute*, 357(14):9633–9653.
- [12] Rios, Y. Y., García-Rodríguez, J., Sanchez, E. N., Alanis, A. Y., Ruiz-Velázquez, E., and Pardo Garcia, A. (2022). Treatment for T1DM patients by a neuro-fuzzy inverse optimal controller including multi-step prediction. *ISA Transactions*, 126:203–212.

