

# Modelado Interpretable de Datos de Glucosa

1° Juan F. Gaitán-Guerrero  
*Departamento de Informática*  
*Universidad de Jaén*  
Jaén, España  
ORCID: 0009-0007-6872-1401

2° José L. López  
*Departamento de Informática*  
*Universidad de Jaén*  
Jaén, España  
ORCID: 0000-0003-2583-8638

3° Macarena Espinilla Estévez  
*Departamento de Informática*  
*Universidad de Jaén*  
Jaén, España  
ORCID: 0000-0003-1118-7782

4° David Díaz-Jiménez  
*Departamento de Informática*  
*Universidad de Jaén*  
Jaén, España  
ORCID: 0000-0003-1791-4258

5° Carmen Martínez-Cruz  
*Departamento de Lenguajes y Sistemas Informáticos*  
*Universidad de Granada*  
Granada, España  
ORCID: 0000-0002-8117-0647

**Abstract**—La diabetes plantea un reto sanitario global que requiere interpretar adecuadamente los datos generados por sensores. Este trabajo propone un sistema basado en lógica difusa (FL) para generar resúmenes lingüísticos de datos de monitorización continua de glucosa (CGM). Integrado en una arquitectura del Internet de las Cosas (IoT) multipaciente, permite describir series temporales (TS) en lenguaje natural, facilitando su comprensión por profesionales, pacientes y familiares. La propuesta se compara con GPT-4, destacando ventajas en interpretabilidad, eficiencia y sostenibilidad.

**Index Terms**—Lógica difusa, Monitorización continua de glucosa, IoT, Resúmenes lingüísticos de series temporales, Generación de lenguaje natural, diabetes, GPT.

## I. INTRODUCCIÓN

La diabetes *mellitus* es una enfermedad metabólica causada por una insuficiencia parcial o total de insulina, afectando a más de 537 millones de personas en el mundo [1]. Pese a que los sistemas sanitarios actuales dotan a sus pacientes con dispositivos inteligentes para la CGM a través de sensores, aun se enfrentan al desafío de ofrecer una atención personalizada, en gran parte por la complejidad y volumen de datos generados.

En este contexto, la FL surge como un enfoque clave para modelar la incertidumbre inherente a los datos de glucosa. Pese a que existen numerosos trabajos en el manejo de datos de diabetes, estos se enfocan en la predicción de valores futuros o en determinar si el paciente padece o no la enfermedad [2], no abordando así técnicas interpretables que permitan llevar a cabo un control adecuado de los pacientes.

Como consecuencia, se propone una arquitectura IoT multipaciente para la recogida de datos en tiempo real, combinada con un sistema basado en FL para generar descripciones lingüísticas interpretables de TS de glucosa. La validación del

sistema se realiza a través de una evaluación del mismo por parte de los usuarios finales, así como una comparativa con modelos de generación de lenguaje natural como GPT-4, en términos de interpretabilidad, eficiencia, contenido semántico y sostenibilidad.

## II. TRABAJOS RELACIONADOS

En el ámbito de la CGM en pacientes con diabetes, numerosos trabajos proponen el uso de dispositivos IoT para recopilar datos biométricos. Sin embargo, muchas de estas propuestas no describen una arquitectura concreta multipaciente ni ofrecen información completa sobre los sensores empleados.

Por otra parte, para afrontar la incertidumbre inherente en los datos monitorizados, se recurre a la FL [3], la cual ha demostrado ser fundamental para generar descripciones lingüísticas de TS (GLiDTS), tal y como se refleja en [4]. Aunque los modelos de lenguaje como GPT-4 han comenzado a utilizarse para resumir datos de CGM [5], estos presentan errores, alucinaciones o subjetividad en descripciones lingüísticas que abarcan varios días, revelando así la necesidad de análisis diarios detallados.

## III. ARQUITECTURA DEL SISTEMA

La arquitectura del sistema combina el sensor comercial Freestyle Libre 3 [6] con una aplicación móvil que recoge los datos en tiempo real; a diferencia del glucómetro tradicional, el dispositivo IoT permite una recopilación completa de los niveles de glucemia sin intervención del usuario, durante un período de 14 días. Para ello, se emplea la aplicación xDrip+ [7], que facilita la lectura continua del sensor y el envío automático de los datos al servidor. Los valores de glucosa, el timestamp y el utcOffset se almacenan y preprocesan en un servidor propio, que además aloja una plataforma web con el motor de generación de resúmenes lingüísticos. La comunicación se realiza mediante una API REST que replica la funcionalidad de Nightscout [8], y adaptada para gestionar múltiples pacientes de forma simultánea.

Este resultado ha sido parcialmente respaldado por el proyecto PID2021-127275OB-I00 y el proyecto PID2021-126363NB-I00 financiados por el MICIU/AEI/10.13039/501100011033 y por “ERDF A way of making Europe”, y por el proyecto PDC2023-145863-I00 financiado por el MICIU/AEI/ 10.13039/501100011033 y por la “European Union NextGenerationEU/PRTR”.

Este resultado es un resumen del trabajo publicado con referencia doi.org/10.1109/JIOT.2024.3419260

#### IV. METODOLOGÍA DE GENERACIÓN DE RESÚMENES LINGÜÍSTICOS

Las TS registradas en tiempo real son convertidas a lenguaje natural de manera interpretable, abordando así las limitaciones encontradas en la literatura.

En primer lugar, los datos con preprocesados, destacando el uso del algoritmo *Moving Average Window* para la imputación de valores perdidos en intervalos de tiempo no superiores a media hora. Posteriormente, se aplica el algoritmo *Ramer Douglas Peucker* [9] para la segmentación de la TS, obteniendo así dos TS (la original y la segmentada). Cada segmento se caracteriza usando funciones de pertenencia difusas asociadas a los niveles de glucosa, la tendencia de los segmentos y, los momentos del día al que pertenece cada uno de los puntos. Así, la TS segmentada permite la caracterización de eventos (i.e.: hipoglucemia, hiperglucemia, picos y mesetas). Por su parte, a partir de la TS original con el etiquetado de niveles de glucosa, los momentos del día y el uso de cuantificadores difusos, se permite la generación de Protoformas de tipo 2 (p.ej.: Por la mañana, la mayoría de los niveles de glucosa fueron altos).

Una vez se genera la descripción de la TS, cada una de las sentencias se evalúa en función del grado de verdad (DoT), el cual debe ser superior o igual a 0.7 para ser admitida en el resumen final. Adicionalmente, un marco de reglas de calidad es utilizado, evitando repeticiones y redundancias, dando así prioridad a eventos significativos de manera progresiva en el tiempo, proporcionando una descripción desde lo más general a lo más específico.

#### V. RESULTADOS Y DISCUSIÓN

El sistema ha sido evaluado sobre un conjunto de datos reales de un paciente con diabetes tipo 1, con más de 200 días monitorizados. Para validar el conjunto de datos recogidos, este fue computado de forma tal que se hizo el cálculo de todas las posibles protoformas que se podrían generar. El resultado fue que el 89% de estas fueron activadas, lo que evidencia una amplia cobertura de los eventos esperados en los datos.

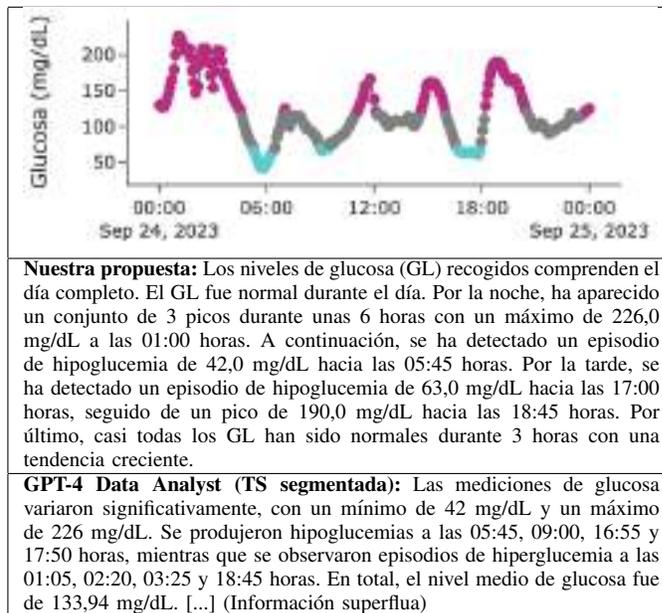
Una encuesta realizada a pacientes, familiares y personal clínico arrojó una puntuación media de usabilidad del 73%, destacando la claridad, brevedad y relevancia de los resúmenes. Comparado con GPT-4 y su *plug-in* Data Analyst (variando el *prompt* del modelo entre la TS original y la TS segmentada), nuestra propuesta se ensalzó en aspectos clave como la precisión, la dinámica temporal, la relevancia del contenido, la claridad narrativa, la escalabilidad del sistema y su sostenibilidad a largo plazo, esto último considerando los Objetivos de Desarrollo Sostenible #3 (Salud y Bienestar) #10 (Reducción de las desigualdades) y #12 (Producción y Consumo Responsables). Igualmente, cabe destacar como el *plug-in* Data Analyst obtuvo la mayor valoración en el escenario en que el *prompt* contenía la TS segmentada.

#### VI. CONCLUSIONES Y TRABAJOS FUTUROS

Este estudio ha presentado una arquitectura completa para la generación de resúmenes lingüísticos de alta calidad a partir

TABLE I

COMPARATIVA DE RESUMEN LINGÜÍSTICO CON GPT-4 DATA ANALYST.



de TS de glucosa, mediante un modelo de representación del conocimiento y una arquitectura de monitorización en tiempo real para múltiples pacientes, basada en sensores continuos de glucosa intersticial. De esta manera, se persigue fomentar una mejora en el bienestar de las personas a la vez que se alivia la carga de los sistemas sanitarios actuales.

El sistema permite visualizar e interpretar los datos de manera comprensible, implicando a expertos y usuarios en su diseño, aplicando FL para transformar patrones relevantes en descripciones en lenguaje natural. Como trabajo futuro, se plantea ampliar el análisis a periodos más largos, con el objetivo de evaluar la evolución del paciente y posibles cambios en sus hábitos alimentarios.

#### REFERENCES

- [1] International Diabetes Federation. (n.d.) What Is Diabetes. [Online]. Available: <https://idf.org/about-diabetes/what-is-diabetes/>
- [2] K. Liu, L. Li, Y. Ma, J. Jiang, Z. Liu, Z. Ye, S. Liu, C. Pu, C. Chen, Y. Wan *et al.*, "Machine Learning Models for Blood Glucose Level Prediction in Patients With Diabetes Mellitus: Systematic Review and Network Meta-Analysis," *JMIR Medical Informatics*, vol. 11, no. 1, p. e47833, 2023.
- [3] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.
- [4] N. Marín and D. Sánchez, "On generating linguistic descriptions of time series," *Fuzzy Sets and Systems*, vol. 285, pp. 6–30, 2016.
- [5] E. Healey, A. Tan, K. Flint, J. Ruiz, and I. Kohane, "Leveraging large language models to analyze continuous glucose monitoring data: A case study," *medRxiv*, pp. 2024–04, 2024.
- [6] Abbot. (n.d.) Freestyle Libre 3 sensor. [Online]. Available: <https://www.freestyle.abbott/es-es/productos/freestylelibre-3.html>
- [7] Nightscout contributors, "Nightscout xDrip+." [Online]. Available: <https://github.com/NightscoutFoundation/xDrip>
- [8] Nightscout contributors, "Nightscout RESTful service framework." [Online]. Available: <https://github.com/cosmonaut/nightscout/blob/develop/swagger.yaml>
- [9] A. Saalfeld, "Topologically consistent line simplification with the douglas-peucker algorithm," *Cartography and Geographic Information Science*, vol. 26, no. 1, pp. 7–18, 1999.