

Clustering y predicción de hipoglucemias nocturnas a partir de monitorización continua de glucosa en pacientes con diabetes mellitus tipo 1

Rafaella Salazar Pérez

Resumen — La monitorización continua de glucosa (MCG) ha evidenciado la gran heterogeneidad de los perfiles glucémicos en pacientes con diabetes mellitus tipo 1, lo que subraya la necesidad de identificar fenotipos y orientar un tratamiento más personalizado. Entre las alteraciones detectadas, las hipoglucemias nocturnas son particularmente relevantes por su elevada frecuencia y asociación con complicaciones graves y deterioro de la calidad de vida. En este estudio se analizó el conjunto de datos T1DiabetesGranada (736 pacientes con FreeStyle Libre 2). Tras armonización y preprocesamiento, se aplicaron técnicas de clustering (K-means, aglomerativo y mezclas gaussianas). El mejor rendimiento se obtuvo con K-means ($k=3$; Silhouette=0,33; DB=1,13), identificando tres fenotipos: equilibrado ($n=383$, 52,0%), hiperglucémico ($n=199$, 27,0%) e hipoglucémico con alta variabilidad ($n=154$, 21,0%). Las variables más discriminantes fueron la carga de hipoglucemias nocturnas (media 6,2% del tiempo), hipoglucemias postprandiales (4,7%) y la glucosa media (152 ± 36 mg/dL). Para la predicción de hipoglucemia nocturna (≥ 1 lectura <70 mg/dL entre 00:00–06:00), se desarrollaron modelos con variables estructuradas (Random Forest, Gradient Boosting, XGBoost) y modelos secuenciales con series CGM (ventanas de 4 h y horizonte de 2 h; LSTM y CNN-1D). En variables estructuradas, Random Forest alcanzó el mejor equilibrio ($F1=0,788$; AUROC=0,902; especificidad=0,85), mientras que XGBoost maximizó la sensibilidad (0,88) y la PR-AUC (0,820). En el modelado secuencial, CNN-1D obtuvo AUROC=0,917 y LSTM AUROC=0,910. En la predicción continua de glucosa, la ANN-MLP superó a la regresión lineal ($RMSE=22,8/32,4/47,5$ mg/dL y Clarke A+B=90,1%/81,0%/66,9% a 30/60/120 min). En conclusión, el clustering permitió caracterizar fenotipos glucémicos diferenciados, mientras que los modelos de machine learning y deep learning lograron predecir hipoglucemias nocturnas con alta precisión, ofreciendo un enfoque prometedor en el manejo personalizado de la diabetes. **Palabras clave:** monitorización continua de glucosa; hipoglucemia nocturna; machine learning; deep learning; clustering; diabetes tipo 1

Abstract— Continuous glucose monitoring (CGM) has revealed the high heterogeneity of glycemic profiles in patients with type 1 diabetes mellitus, underscoring the need to identify phenotypes and guide more personalized treatment strategies. Among the alterations detected, nocturnal hypoglycemia is particularly relevant due to its high frequency and association with severe complications and impaired quality of life. In this study, we analyzed the T1DiabetesGranada dataset (736 patients, using FreeStyle Libre 2). After harmonization and preprocessing, different clustering techniques (K-means, agglomerative, and Gaussian mixtures) were applied. The best performance was achieved with K-means ($k=3$; Silhouette=0.33; DB=1.13), identifying three phenotypes: balanced ($n=383$, 52.0%), hyperglycemic ($n=199$, 27.0%), and hypoglycemic with high variability ($n=154$, 21.0%). The most discriminant variables were nocturnal hypoglycemia burden (mean 6.2% of time), postprandial hypoglycemia (4.7%), and mean glucose (152 ± 36 mg/dL). For nocturnal hypoglycemia prediction (≥ 1 reading <70 mg/dL between 00:00–06:00), we developed models based on structured variables (Random Forest, Gradient Boosting, XGBoost) and sequential CGM models (4-h sliding windows with a 2-h prediction horizon; LSTM and 1D-CNN). In structured variables, Random Forest achieved the best balance ($F1=0.788$; AUROC=0.902; specificity=0.85), while XGBoost maximized sensitivity (0.88) and PR-AUC (0.820). In sequential modeling, 1D-CNN reached AUROC=0.917 and LSTM AUROC=0.910. In continuous glucose prediction, ANN-MLP outperformed linear regression ($RMSE=22.8/32.4/47.5$ mg/dL and Clarke A+B=90.1%/81.0%/66.9% at 30/60/120 min). In conclusion, clustering enabled the characterization of distinct glycemic phenotypes, while machine learning and deep learning models achieved high accuracy in predicting nocturnal hypoglycemia, offering a promising approach for personalized management of type 1 diabetes. **Keywords:** continuous glucose monitoring; nocturnal hypoglycemia; machine learning; deep learning; clustering; type 1 diabetes.

1 INTRODUCCIÓN

La diabetes mellitus tipo 1 (DM1) es una enfermedad crónica autoinmune que se caracteriza por una deficiencia severa en la secreción de insulina, lo que obliga a los pacientes a depender de administración exógena de insulina para mantener sus niveles de glucosa en sangre dentro de un rango objetivo (1,2). La complejidad inherente a esta tarea radica en la gran variabilidad de los patrones de glucosa individuales, condicionada por el estilo de vida, el metabolismo, el sexo y otros factores. (3)

Para un manejo óptimo de la enfermedad y la prevención de complicaciones graves, es fundamental mantener las concentraciones de glucosa en sangre dentro de rangos seguros, evitando tanto la hipoglucemia, niveles bajos de glucosa por debajo de 70mg/dl (3.9mmol/L) en personas con diabetes y evitando también la hiperglucemia mantenida. (niveles altos de glucosa).

- E-mail de contacto: rafaelasp178@gmail.com
- Trabajo tutorizado por: Prof. Javier Serrano
- Curso: 2025

La mayor parte de pacientes está en tratamiento con múltiples dosis de insulina (MDI) pudiendo presentar hipoglucemias nocturnas, definido como niveles de glucosa inferiores a 70mg/dl (1). Uno de los mayores desafíos en el control glucémico de la DM1 es la hipoglucemia nocturna (HN), una complicación peligrosa y frecuente de la terapia con insulina que a menudo pasa desapercibida, ya que los síntomas pueden no ser detectados durante el sueño teniendo dificultades en su reconocimiento, prevención y tratamiento. Los episodios nocturnos se caracterizan por una mayor duración y una recuperación más lenta en comparación con los eventos diurnos, además tienen un impacto significativo en el control glucémico del día siguiente, manifestándose en un peor control glucémico, un aumento en el tiempo por debajo del rango y una mayor variabilidad glucémica (4)

Sus consecuencias abarcan manifestaciones clínicas como alteraciones del sueño, dolores de cabeza matutinos, fatiga, ansiedad y cambios de humor, además de un mayor riesgo de eventos cardiovasculares y mortalidad por todas las causas. Aunque existen soluciones tecnológicas para mitigar la hipoglucemia nocturna en pacientes con bombas de insulina, sería beneficioso contar con sistemas de apoyo tecnológico similares para la prevención en pacientes con DT1 que usan MDI. (5)

El miedo a la hipoglucemia es común y puede llevar a comportamientos compensatorios negativos como reducir la dosis de insulina o comer en exceso para evitar futuros episodios afectando además la calidad de vida de los pacientes (6,7)

El uso extensivo de los monitores continuos de glucosa (CGM) en el manejo de la diabetes tipo 1 (DT1), ha permitido detectar la marcada variabilidad en los patrones de glucosa entre diferentes individuos.

Además, ha generado una gran cantidad de volúmenes de datos que pueden ser aprovechados por la inteligencia artificial (IA) y el aprendizaje automático (AA) para entrenar algoritmos de aprendizaje automático supervisado (ML) y no supervisado (DL) con el fin de predecir eventos adversos antes de que ocurran, tanto hipoglucemias nocturnas como postprandiales (8,9) e hiperglucemias sostenidas. (10), asimismo se está buscando nuevas métricas que reflejen el control glicémico del paciente que representen mejor los diferentes patrones glucémicos.(11,12)

Aunque los modelos de predicción generales a menudo no logran capturar las diferencias individuales en los patrones de glucosa, y el desarrollo de modelos personalizados para cada paciente es poco práctico debido a la gran cantidad de datos continuos de glucosa que se requieren y a la escasez de conjuntos de datos de acceso abierto de alta calidad con suficiente cantidad de pacientes y longitudinalidad de mediciones de glucosa continua para apoyar una experimentación robusta, muchos trabajos contribuyen un poco mas en el desarrollo

de nuevos algoritmos para mejorar el control glucémico mediante un tratamiento personalizado para mejorar la calidad de vida y autonomía en el manejo de su enfermedad.

Este estudio propone un enfoque integral para la caracterización y predicción de hipoglucemias nocturnas en pacientes con Diabetes Mellitus tipo 1 (DM1) utilizando un extenso conjunto de datos de monitorización continua de glucosa (MCG). Se busca identificar subgrupos de pacientes con patrones glucémicos similares (fenotipos) mediante técnicas de clustering no supervisado, y posteriormente desarrollar modelos de predicción de hipoglucemias nocturnas.

2 OBJETIVOS

Caracterizar fenotipos glucémicos distintivos en pacientes con Diabetes Mellitus tipo 1 (DM1) a partir de datos de monitorización continua de glucosa (MCG), incorporando indicadores de control glucémico derivadas de la MCG como el tiempo en rango (TIR), tiempo por debajo (TBR) y por encima del rango (TAR), variabilidad glucémica (CV), amplitud media de las excursiones glucémicas (MAGE), indicadores de riesgo glucémico (GRI) y Kernel density estimate (KDE) con el fin de identificar patrones y explorar la heterogeneidad clínica en el control metabólico.

Desarrollar modelos de aprendizaje automático (ML) y aprendizaje profundo (DL) basados en monitorización continua de glucosa (MCG) incorporando métricas derivadas de la MCG y variables clínico demográficas para la predicción de hipoglucemias nocturnas (HN)

3 ESTADO DEL ARTE

La gestión de la diabetes mellitus tipo 1 (DM1) se ve significativamente impulsada por la monitorización continua de glucosa (MCG), que genera grandes volúmenes de datos susceptibles de ser analizados mediante inteligencia artificial (IA). Los avances recientes en técnicas de clustering para la identificación de patrones glucémicos, modelos de aprendizaje automático (AA) para la predicción de hipoglucemias, y la aplicación de aprendizaje profundo (DL) en la predicción de glucosa y eventos adversos.

3.1 Clustering de patrones glucémicos

El *clustering* aplicado a datos de monitorización continua de glucosa (MCG) ha demostrado ser una herramienta clave para identificar patrones de variabilidad glucémica y estratificar pacientes con diabetes tipo 1 (DM1). El análisis de agrupamiento ha surgido como un enfoque prometedor para caracterizar la heterogeneidad de los pacientes con DT1 y optimizar estrategias terapéuticas personalizadas.

Chan et al. (2023) emplearon el algoritmo Dynamic Time Warping (DTW) para extraer patrones de variabilidad glucémica, identificando seis perfiles distintos a partir de datos de FreeStyle Libre. Estos patrones representaron desde glucemias en rango hasta tendencias hipo o hiperglucémicas, con el objetivo de orientar vías terapéuticas personalizadas.[\(13\)](#)

Contador Pachón (2021) utilizó el método CHAID para clasificar perfiles de glucosa en función de día de la semana y franjas horarias, encontrando diferencias significativas y dependencias específicas por paciente. Sus modelos basados en CHAID demostraron mayor precisión que los modelos estándar, resaltando la necesidad de enfoques individualizados. [\(14\)](#) Cui et al. (2023), retomando el concepto de glucodensidad introducido por Matabuena et al. (2021), aplicaron técnicas basadas en distancia 2-Wasserstein y la media de Fréchet, identificando tres clústeres de pacientes con distintos niveles y variabilidad glucémica, desde un control deficiente hasta un manejo más estable.[\(12,15\)](#)

Kladov et al. (2023, 2024) aplicaron clustering jerárquico a datos nocturnos de MCG, identificando patrones con y sin hipoglucemia nocturna. Sus resultados mostraron diferencias en el inicio, duración y recuperación de las hipoglucemias. Posteriormente, demostraron que el pre-clustering mejoraba la eficacia de modelos predictivos como Random Forest y Gradient Boosted Trees. [\(16\)](#)

Konstantakopoulos et al. (2023) desarrollaron un modelo de clustering personalizado con Mapas Autoorganizados (SOM) para glucemias nocturnas, dividiendo los datos en dos grupos principales (niveles bajos vs. altos). Encontraron que el 58 % de los pacientes presentaba hiperglucemia nocturna, lo que sugiere un riesgo subestimado en el manejo clínico.[\(17\)](#)

En su tesis doctoral, León (2024) evaluó múltiples algoritmos no supervisados (K-means, Gaussian Mixture, SOM, BIRCH y Fuzzy C-means), identificando seis grupos clínicamente validados por expertos. Estos clústeres diferencian pacientes con buen control (algunos con riesgo de hipoglucemia matutina o fenómeno del amanecer) frente a otros con mayor hiperglucemia persistente, ofreciendo una base sólida para tratamientos personalizados. [\(2\)](#)

Los estudios revisados muestran que el *clustering* permite caracterizar patrones glucémicos individuales y colectivos, identificar subtipos de pacientes con DT1 y mejorar la comprensión de la dinámica de la glucosa. En conjunto, estas investigaciones evidencian el potencial de esta técnica para apoyar la personalización de modelos predictivos y el diseño de estrategias terapéuticas más dirigidas en el manejo de la diabetes.

3.2 Predicción de hipoglucemias nocturnas

La predicción de hipoglucemias nocturnas (HN) mediante inteligencia artificial ha sido abordada principalmente con enfoques de *machine learning* (ML) y *deep learning* (DL).

En los trabajos en los que se uso *Machine Learning*, Afentakis et al. (2025) emplearon clasificadores binarios basados en SVM y Random Forest, alcanzando un rendimiento óptimo con SVM (ROC-AUC \approx 0.79; sensibilidad 74–83%; especificidad 57–66%), validado en datos de Dexcom y FreeStyle Libre. [\(1\)](#)

Berikov et al. (2022) mostraron que Random Forest fue el algoritmo más preciso en pacientes hospitalizados, con un ROC-AUC de 0.97 a 15 minutos y 0.942 a 30 minutos, destacando la relevancia de predictores clínicos como HbA1c, proteinuria y dosis de insulina[\(9\)](#). Bertachi et al. (2020) utilizaron SVM y MLP en modelos individualizados, reportando sensibilidades del 78.8% y especificidades del 82.1%, con la estimación de que más del 70% de los eventos podrían evitarse [\(18\)](#). Mosquera-López et al. (2020) obtuvieron un ROC-AUC del 77% con SVR y demostraron que su modelo podría reducir hasta un 77% de eventos sin afectar el tiempo en rango [\(19\)](#).

Otros estudios, como Jensen et al. (2020) [\(20\)](#), Vu et al. (2019) [\(21\)](#) y Parcerisas et al. (2022) [\(22\)](#), confirmaron la utilidad de LDA, Random Forest y modelos poblacionales o personalizados, con ROC-AUC entre 0.77 y 0.84. Más recientemente, Kladov et al. (2024) demostraron que el uso de pre-clustering mejoraba el rendimiento predictivo de RF y GBT en horizontes de 15–30 minutos [\(16\)](#). La predicción mediante aprendizaje profundo (DL) se ha consolidado como una herramienta prometedora para la predicción de hipoglucemias al permitir la extracción automática de características complejas y mejorar la generalización en distintas poblaciones.

Shao et al. validaron un modelo basado en LSTM, demostrando robustez en poblaciones con DT1 y DT2 y entre diferentes etnias. Alcanzaron AUC >97 % para hipoglucemia leve y >99 % para grave, superando a SVM y RF, con una especificidad del 88.4 % y sensibilidad del 90 % para hipoglucemia leve [\(23\)](#). Kozinetz et al. desarrollaron CNN y MLP para predicción nocturna en DT1, confirmando que los modelos de DL pueden trabajar con datos crudos y detectar automáticamente patrones relevantes, aunque sin reportar métricas numéricas detalladas [\(24\)](#). Katsarou et al. propusieron un modelo ensemble (SVR + XGBoost), que redujo las predicciones erróneas al 19 % en horizonte de 30 min, con un RMSE de ~15.5 mg/dL [3]. Otros trabajos con LSTM multitarea alcanzaron hasta 98.1 % de aceptabilidad clínica en Clarke Error Grid (zonas A+B) a 15 min [\(25\)](#).

Muñoz-Organero diseñó un modelo híbrido RNN-LSTM que emulaba procesos fisiológicos, logrando RMSE <10 mg/dL en datos reales y <5 mg/dL en simulados [\(26\)](#). Mujahid exploró modelos generativos (GAN, S2S-GAN) capaces de producir datos sintéticos clínicamente realistas, que mejoraron la precisión de los modelos de predicción y alcanzaron un tiempo en rango (TIR) del 74.7 % en entornos simulados [\(27\)](#).

Ryu et al. aplicaron transformers (PatchTST, TS Mixer, LTSF-N-linear) en pacientes con trastornos de glucógeno, logrando un AUROC de 0.866 a 15 min y un NPV >90 %

para horizontes de 45–60 min (28). Duckworth et al. integraron aprendizaje explicable (XAI) con XGBoost y SHAP, alcanzando rendimientos de vanguardia (AUROC 0.998 para hipoglucemia y 0.989 para hiperglucemia), con horizontes de hasta 60 min (10).

Otros aportes incluyen los de Mohebbi, que mostró la utilidad de MLP, CNN y LSTM en la predicción de glucosa y adherencia terapéutica (~80 % de precisión) (29); Rodríguez León, quien comparó modelos lineales vs. LSTM/CNN, encontrando que los primeros rendían mejor en datos concentrados en un rango reducido (2); y Thomsen et al., que aplicaron CNN con transferencia de aprendizaje (CNN-TL) en DT2, reduciendo la tasa de falsas alarmas al 5 %, aunque con sensibilidad/especificidad algo menores que estudios previos (30).

Finalmente, las revisiones sistemáticas de Felizardo et al. (2021), Mujahid et al. (2021) y Kulzer et al. (2024) subrayan que los enfoques integradores de ML y DL alcanzan sensibilidades entre 71–96%, especificidades entre 62–91% y ROC-AUC de 0.62–0.97. Además, destacan que el rendimiento predictivo tiende a ser mayor en horizontes cortos (15–30 minutos) y que la fusión de información (glucosa, insulina, actividad física y datos clínicos) mejora la discriminación (5,27,31).

4 MATERIAL Y MÉTODOS

4.1 Dataset

El estudio utilizó el dataset T1DiabetesGranada (32), un recurso multimodal, longitudinal y de acceso abierto disponible bajo solicitud en el repositorio Zenodo. Este conjunto de datos comprende 257,780 días de registros obtenidos durante un período de cuatro años en una cohorte de 736 pacientes con diabetes mellitus tipo 1 (DM1) de Granada, España.

La base de datos incluye **mediciones continuas de glucosa (CGM)**, principalmente procedentes del dispositivo FreeStyle Libre 2, que registra los valores de glucemia cada 15 minutos. Además, incorpora **información demográfica y clínica** de los pacientes, lo que permite un análisis integral de los patrones glucémicos y su relación con variables clínicas relevantes.

4.2 Aprobación ética y anonimización:

Antes de la recopilación, se obtuvo la aprobación del Comité de Ética de Investigación Biomédica de la Provincia de Granada. Técnicos independientes del Servicio de Tecnologías de la Información del Hospital Universitario San Cecilio de Granada realizaron un proceso de anonimización para eliminar información confidencial, asignando a cada paciente un identificador único.

4.3 Entorno computacional

El análisis se llevó a cabo en Python 3.11 dentro de Google Colab, utilizando NumPy y Pandas para la gestión de datos, Matplotlib/Seaborn para visualización y Scikit-learn para clustering (K-means, Agglomerative, GMM), evaluado con los índices de Silhouette y Davies-Bouldin.

Para la predicción de hipoglucemias nocturnas, se aplicaron modelos supervisados (Regresión Logística, Random Forest, Gradient Boosting/XGBoost) y redes neuronales (LSTM con TensorFlow/Keras). El desbalance de clases se abordó mediante ponderación de clases y SMOTE. El rendimiento se evaluó con AUROC, AUPRC, sensibilidad, especificidad, PPV, NPV, Brier score, curvas de calibración y el análisis del tiempo de anticipación efectivo entre la predicción y la ocurrencia del evento de hipoglucemia nocturna. Todos los análisis se documentaron en cuadernos de trabajo tipo Jupyter Notebook (.ipynb), preservando la trazabilidad del pipeline.

4.4 Preprocesamiento

Se llevaron a cabo tareas iniciales de limpieza, incluyendo la eliminación de filas duplicadas o con valores faltantes, la recodificación de variables y la estandarización de los formatos de fecha.

Posteriormente, se realizó un preprocesamiento integral de los archivos originales del repositorio (información de pacientes, datos de monitorización continua de glucosa (CGM) y parámetros bioquímicos), armonizando los nombres de las variables y eliminando campos no informativos para el análisis posterior.

Se convirtieron fechas a *datetime*, se calculó la edad al inicio del seguimiento y se codificó el sexo (0 = mujer, 1 = hombre). Para los datos de monitorización continua de glucosa (CGM), el archivo original contenía las series temporales de glucosa intersticial medidas cada 15 minutos mediante sensores FreeStyle Libre durante un período de hasta cuatro años por paciente. Con el fin de reducir la carga computacional y homogeneizar el seguimiento, se seleccionaron únicamente los primeros 100 días de registros desde el inicio de monitorización de cada individuo

A partir de estos registros se calcularon métricas globales y específicas por franjas horarias, incluyendo la glucosa media, desviación estándar, coeficiente de variación, TIR, TBR1/2, TAR1/2, MAGE, GRI, áreas de densidad (hipo/normo/hiperglucemia), frecuencia de hipoglucemias nocturnas y postprandiales, así como la tasa media de cambio de glucosa. Además, los datos fueron estratificados en cinco intervalos horarios (00:00–06:00, 06:00–13:00, 13:00–16:00, 16:00–20:00, 20:00–24:00) para capturar patrones circadianos.

Los parámetros bioquímicos (14 analitos clínicamente relevantes) se agregaron a media por paciente y se pivotaron a formato ancho. Los datasets preprocesados se unificaron por *Patient_ID*; se imputaron faltantes con la mediana y se aplicó winsorización (p1–p99) para atenuar atípicos sin excluir sujetos. El resultado fue un dataset no escalado con 736 pacientes y 68 variables (demográficas, bioquímicas y métricas CGM globales y por franjas), preservando las unidades originales y listo para el escalado y el *clustering* posterior.

4.5 Clustering

Se llevó a cabo un análisis de agrupamiento no supervisado con el objetivo de identificar fenotipos glucémicos diferenciados en pacientes con diabetes mellitus tipo 1, a partir de métricas derivadas de monitorización continua de glucosa (CGM) y parámetros clínicos asociados.

El dataset inicial incluyó 736 pacientes con un total de 68 variables (demográficas, bioquímicas y métricas glucémicas globales y estratificadas por franjas horarias). Con el fin de evaluar la influencia de las características seleccionadas en la calidad de la agrupación, se construyeron tres versiones del dataset:

- **Versión A (15 variables):** métricas globales de CGM de uso clínico habitual (glucosa media, desviación estándar, coeficiente de variación, TIR, TBR1/2, TAR1/2, MAGE, GRI, áreas KDE y recuento de hipoglucemias nocturnas y postprandiales).
- **Versión B (50 variables):** Versión A más métricas estratificadas en cinco franjas horarias (00:00–06:00, 06:00–13:00, 13:00–16:00, 16:00–20:00, 20:00–24:00).
- **Versión C (63 variables):** Versión B más 13 parámetros bioquímicos como hemoglobina glicosilada (HbA1c), perfil lipídico, creatinina, transaminasas, electrolitos, tirotropina (TSH) y ácido úrico.

El preprocesamiento incluyó imputación de valores faltantes mediante la mediana, winsorización (p1–p99) para reducir la influencia de valores extremos, posteriormente se realizó estandarización mediante z-score.

Se evaluaron cuatro algoritmos de clustering: K-means (50 inicializaciones, algoritmo Elkan), Agglomerative Clustering, Gaussian Mixture Models (GMM) y DBSCAN ($\epsilon=0.5, 1.0$ y 1.5 ; $\text{min_samples}=5$). Para los modelos particionales y jerárquicos (K-means, Agglomerativo y GMM) se exploraron valores de k entre 2 y 6.

La calidad de los modelos se determinó mediante métricas internas complementarias: el índice de Silhouette (cohesión intra-clúster) y el índice de Davies–Bouldin (separación inter-clúster). Los resultados se ordenaron según un ranking combinado (Silhouette decreciente + Davies–Bouldin creciente), seleccionando los modelos con mayor consistencia.

4.6 Predicción de hipoglucemias

La predicción de hipoglucemias nocturnas se abordó mediante dos estrategias metodológicas complementarias: **clasificación binaria** a partir de variables estructuradas, y **modelado secuencial/continuo de series temporales de glucosa intersticial (CGM)**. Definiendo la variable objetivo como la ocurrencia de ≥ 1 episodio de glucosa <70 mg/dL en el intervalo comprendido entre las 00:00 y las 06:00 horas

4.6.1 Clasificación binaria a partir de variables estructuradas

Para este enfoque se construyó un dataset estructurado a nivel paciente que integró variables demográficas, bioquímicas y métricas derivadas de CGM en un total de 736 individuos.

La variable dependiente se definió de manera binaria como la presencia (1) o ausencia (0) de al menos un episodio hipoglucémico (<70 mg/dL) en la franja horaria de 00:00 a 06:00 horas. Con el fin de evitar fuga de información, se eliminaron del conjunto predictor las métricas correspondientes a la franja nocturna (00:00 - 06:00), la variable de recuento de hipoglucemias nocturnas, la etiqueta de clúster y todas las métricas globales de control glucémico (media, desviación estándar, coeficiente de variación, TIR, TBR, TAR, MAGE, GRI, áreas KDE y frecuencia de hipoglucemias postprandiales).

Sobre este conjunto se entrenaron algoritmos de clasificación supervisada: Random Forest (300 árboles), Gradient Boosting y XGBoost. El entrenamiento se llevó a cabo mediante validación cruzada estratificada por paciente (*GroupKFold*, $k=5$). Para abordar el desbalance de clases se aplicó SMOTE (Synthetic Minority Oversampling Technique).

La evaluación incluyó métricas de rendimiento discriminativo (AUROC, AUPRC, sensibilidad, especificidad, precisión, F1-score) y de calibración (Brier score, curvas de calibración).

Modelo	Parámetros principales	Validación	Manejo de desbalance	Métricas
Random Forest	n_estimators=300, criterion=Gini, max_depth=None, min_samples_split=2, bootstrap=True, random_state=42	Validación cruzada estratificada por paciente (Group KFold, k=5)	SMOTE en conjunto de entrenamiento	AUROC, AUPRC, Sensibilidad, Especificidad, F1, Brier score curvas de calibración, matriz de confusión
Gradient Boosting	n_estimators=300, learning_rate=0.1, max_depth=3, subsample=1.0, loss=log_loss, random_state=42	GroupKFold (k=5)	SMOTE en conjunto de entrenamiento	AUROC, AUPRC, sensibilidad, especificidad, precisión, F1
XGBoost	objective-binary:logistic, eval_metric=auc, n_estimators=300, eta=0.1, max_depth=6, subsample=0.8, colsample_bytree=0.8, lambda=1, alpha=0, random_state=42	GroupKFold (k=5)	SMOTE + class_weight en entrenamiento	AUROC, AUPRC, F1-score, Brier score, sensibilidad, especificidad, matriz de confusión

Tabla 1. Clasificación binaria a partir de variables estructuradas

4.6.2 Modelado secuencial/continuo de series temporales de CGM

En este segundo enfoque se emplearon directamente los registros crudos de glucosa intersticial obtenidos mediante monitorización continua de glucosa (CGM), muestreados cada 15 minutos. Se construyeron ventanas deslizantes de 4 horas de historia glucémica (16 mediciones), seguidas de un horizonte de predicción de 2 horas (8 mediciones). La etiqueta binaria fue positiva si en dicho horizonte se registraba al menos un valor <70 mg/dL.

Se implementaron dos arquitecturas de *deep learning* en TensorFlow/Keras:

- **LSTM (Long Short-Term Memory):** 64 unidades, capa intermedia de 32 neuronas con activación ReLU, regularización mediante *dropout* (0.3) y salida sigmoide.
 - **CNN-1D:** 32 filtros convolucionales, *max-pooling*, *dropout*, capa densa de 64 neuronas y salida sigmoide.

Ambos modelos fueron entrenados durante 50 épocas (batch_size=64), aplicando *early stopping* con paciencia de cinco épocas y validación en conjunto independiente (20% por paciente).

Complementariamente, se desarrolló un enfoque de predicción continua orientado a estimar los valores futuros de glucosa a 30, 60 y 120 minutos en horario nocturno. Para ello se evaluaron un modelo de regresión lineal y una red neuronal multicapa (ANN-MLP), empleando métricas de error (MSE, RMSE, MAE, MAPE) e incorporando el Clarke Error Grid como indicador clínico de seguridad.

La evaluación incluyó tanto métricas discriminativas (AUROC, AUPRC, sensibilidad, especificidad, F1-score) como métricas de regresión (RMSE, MAE, MAPE, intervalos de confianza bootstrap al 95%). Además, se consideró el tiempo de anticipación efectivo como indicador clínico adicional, definido como el intervalo entre la predicción positiva y la primera lectura <70 mg/dL.

Modelo	Entrada / Arquitectura	Parámetros principales	Regularización y validación	Manejo de desbalance	Métricas
LSTM	Ventana= 16 pasos (4h), Horizonte=120 min; LSTM(64) → Dense(32, ReLU) → Dropout(0.3) → Sigmoid	Optimizer=Adam (lr=0.001); loss=binary_crossentropy; epochs=50; batch=64	EarlyStopping (pacienicia= 5); Valiadación 80/20 por paciente	AUROC, AUPRC, sensibilidad d, F1-score, tiempo de anticipación	AUROC, AUPRC, sensibilidad d, F1-score, tiempo de anticipación
CNN-1 D	Ventana=16 pasos (4h), Horizonte=120 min; Conv1D(32, kernel=3, ReLU) → MaxPooling1D(2) → Dropout(0.3) → Dense(64, ReLU) → Sigmoid	Adam (lr=0.001); loss=binary_crossentropy; epochs=50; batch=64	EarlyStopping (pacienicia= 5); Valiadación 80/20 por paciente	AUROC, AUPRC, sensibilidad d, F1-score, especificidad	AUROC, AUPRC, sensibilidad d, F1-score, especificidad
Regresión lineal	Predicción de glucosa Sin futura a 30, 60, 120 min regularización	Validación 80/20	No aplica MAE, (problema de MAPE, regresión)	RMSE, Clarke Error Grid	RMSE, MAE, Clarke Error Grid
ANN-MLP	Dense(64, ReLU) → Dense(32, ReLU) → Dropout(0.2) → Linear	Adam (lr=0.001); loss=MSE; epochs=100; batch=128	EarlyStopping (pacienicia= 10); Valiadación 80/20	No aplica MAPE, (problema de Clarke regresión)	RMSE, MAE, Clarke Error Grid, IC95% (bootstrap)

Tabla 2. Modelado secuencial/continuo de series temporales de CGM

5 RESULTADOS

Se presentan los hallazgos obtenidos tras la aplicación de técnicas de *clustering* no supervisado y modelos de predicción de hipoglucemias nocturnas en pacientes con diabetes mellitus tipo 1 (DM1).

5.1 Clustering de patrones glucémicos para identificar fenotipos diferenciados.

La aplicación de algoritmos de *clustering* no supervisado sobre el dataset unificado permitió identificar agrupaciones diferenciadas de pacientes con diabetes mellitus tipo 1 en función de sus perfiles glucémicos y métricas derivadas de la monitorización continua de glucosa (CGM).

Se evaluaron diferentes configuraciones de K-means, Gaussian Mixture Models (GMM) y Agglomerative Clustering, con un número de clústeres entre 2 y 6. El índice de Silhouette y el coeficiente de Davies–Bouldin fueron utilizados como métricas de validación interna. (Tabla 3)

Índice / Algoritmo	k = 2	k = 3	k = 4	k = 5	k = 6
Versión A					
Agglomerative	0.301	0.309	0.190	0.203	0.206
GMM	0.288	0.288	0.237	0.231	0.231
K-means	0.308	0.328	0.268	0.245	0.252
Versión B					
Agglomerative	0.230	0.190	0.129	0.129	0.120
GMM	0.165	0.168	0.168	0.174	0.170
K-means	0.224	0.196	0.173	0.179	0.173
Versión C					
Agglomerative	0.162	0.115	0.100	0.098	0.094
GMM	0.155	0.137	0.121	0.130	0.116
K-means	0.183	0.148	0.130	0.133	0.115

Tabla 3. Resultados de los índices de validación interna según algoritmo y número de clústeres ($k = 2-6$). Los valores representan la media de la métrica de Silhouette para cada combinación de algoritmo de clustering (K-means, agglomerative y Gaussian Mixture Models) y número de clústeres ($k = 2-6$), evaluados en tres versiones diferentes de preprocesamiento.

El modelo final correspondió a K-means con $k=3$ aplicado a la Versión A, que alcanzó un silhouette medio de 0,33 y un Davies–Bouldin de 1,13. Este modelo no solo presentó un rendimiento estadísticamente robusto, sino que también permitió una interpretación clínica clara.

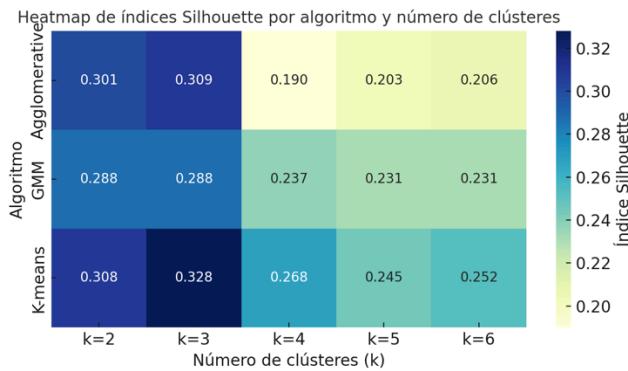


Figura 1. Mapa de calor de indices Silhouette de algoritmos en versión A

Método	Número de clústeres (k)	Índice Silhouette	Índice Davies–Bouldin
K-means	3	0.328	1.131
K-means	4	0.268	1.104
Agglomerative	3	0.309	1.151

Tabla 4. Mejores combinaciones de algoritmos y número de clústeres según índices de validación interna. Aunque K-means con $k=4$ presenta el Davies–Bouldin más bajo, el valor de Silhouette fue considerablemente inferior. El modelo K-means con $k=3$ se seleccionó como la solución óptima de clusterización por su mayor equilibrio entre cohesión interna y separación entre grupos.

Los modelos con mayor número de clústeres tendieron a fragmentar excesivamente los grupos, disminuyendo la cohesión interna.

La representación bidimensional mediante **análisis de componentes principales (PCA)** explicó el 82,7% de la varianza total y evidenció tres conglomerados (Figura 2), reforzando la utilidad clínica de esta estratificación.

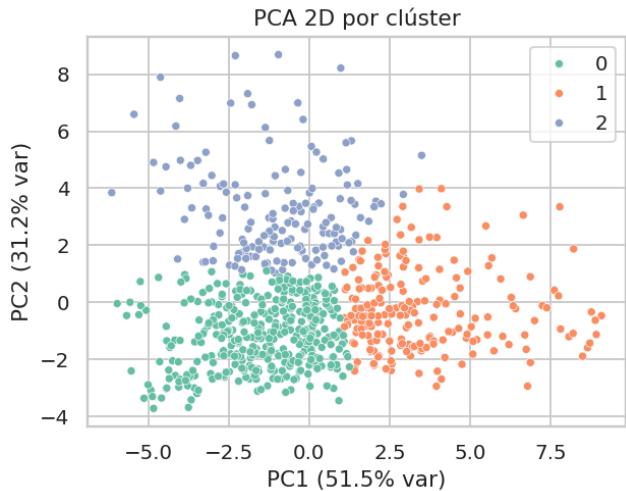


Figura 2. Representación bidimensional mediante análisis de componentes principales (PCA) de los tres clústeres identificados en pacientes con diabetes mellitus tipo 1 ($k=3$)

En conjunto, la clusterización permitió identificar **tres fenotipos clínicamente relevantes** que reflejan perfiles metabólicos diferenciados y podrían orientar la implementación de estrategias terapéuticas personalizadas.

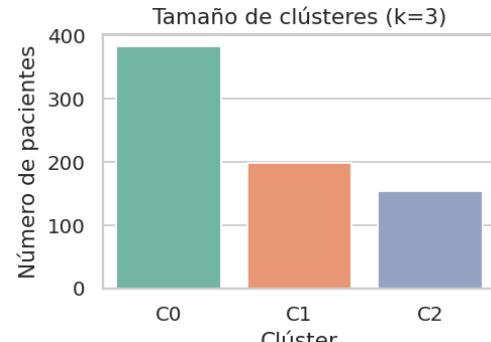


Figura 3. Recuento de pacientes por clúster en el modelo final (KMeans, $k=3$, Versión A). Recuento de pacientes por clúster: C0 = 383, C1 = 199, C2 = 154

La **importancia relativa de las variables** (Tabla 5) mostró que las métricas con mayor peso en la diferenciación interclúster fueron el recuento de hipoglucemias nocturnas (172.3), el recuento de hipoglucemias postprandiales (75.6) y la glucosa media (56.3). Por el contrario, el índice de riesgo glucémico (GRI) y las áreas de densidad KDE mostraron escasa relevancia.

Variable	Distancia entre centroides
Nocturnal_hypo_count	172.27
Postprandial_hypo_count	75.65
mean_glucose	56.34
TIR_%	28.78
TAR1_%	28.75
Ktd_glucose	27.55
TAR2_%	21.49
MAGE	15.63
CV_glucose	10.95
TBR1_%	8.63
TBR2_%	3.20
GRI	1.60
Kde_normo_area	0.28
Kde_hyper_area	0.25
Kde_hypo_area	0.07

Tabla 5. Importancia relativa (distancia entre centroides) de las variables en la diferenciación de clústeres.

El análisis de **valores estandarizados de centroides** (Tabla 6) permitió caracterizar cada clúster:

- **Clúster 0 (n=383, perfil equilibrado):** mejor control glucémico, con TIR% elevado (0.61), mayor tiempo en normoglucemia ($kde_normo_area = 0.61$), glucosa media baja (-0.41) baja variabilidad y mínima frecuencia de hipoglucemias nocturnas o postprandiales y de hiperglucemias. Representa a pacientes con control estable y cercano a los objetivos clínicos.
- **Clúster 1 (n=199, perfil hiperglucémico):** glucosa media (1.23) y TAR elevados TAR1% (1.24) y TAR2% (1.27), TIR reducido (-1.19) y escasa hipoglucemias. Refleja un perfil de **hiperglucemia crónica sostenida**, con bajo riesgo de hipoglucemia, pero mayor riesgo de complicaciones micro y macrovasculares.
- **Clúster 2 (n=154, perfil hipoglucémico/variable):** caracterizado por alta **variabilidad glucémica** ($CV_glucose = 1.06$, $MAGE = 0.28$), valores muy elevados en TBR1% (1.51), TBR2% (1.31), así como mayor frecuencia de hipoglucemias nocturnas (1.27) y postprandiales (1.24). Este grupo representa un perfil de **alto riesgo agudo por hipoglucemias recurrentes**.

Variable	Clúster 0	Clúster 1	Clúster 2
Mean_glucose	-0.41	1.23	-0.58
Std_glucose	-0.62	1.01	0.24
CV_glucose	-0.54	0.23	1.06
GRI	-0.49	1.10	-0.21
MAGE	-0.46	0.67	0.28
TIR_%	0.61	-1.19	0.02
TBR1_%	-0.41	-0.39	1.52
TBR2_%	-0.41	-0.22	1.31
TAR1_%	-0.47	1.24	-0.44
TAR2_%	-0.55	1.28	-0.28
Kde_hypo_area	-0.31	-0.51	1.42
Kde_normo_area	0.61	-1.19	0.02
Kde_hyper_area	-0.45	1.21	-0.44
Nocturnal_hypo_count	-0.32	-0.37	1.27
postpradial_hypo_count	-0.31	-0.36	1.24

Tabla 6. Valores estandarizados de los centroides en cada clúster (k=3)

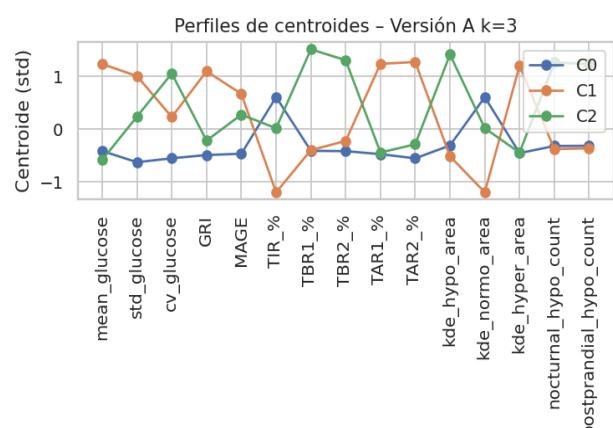


Figura 4. Perfiles de centroides estandarizados (k=3)

La visualización de los **centroides estandarizados** (Figura 4, Figura 5) y los **gráficos radar** (Figura 6) refuerza la existencia de **tres fenotipos glucémicos clínicamente relevantes**.

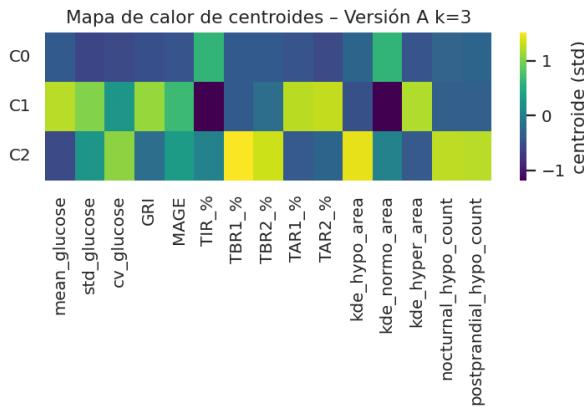


Figura 5. Mapa de calor de centroides estandarizados (Versión A, $k=3$)

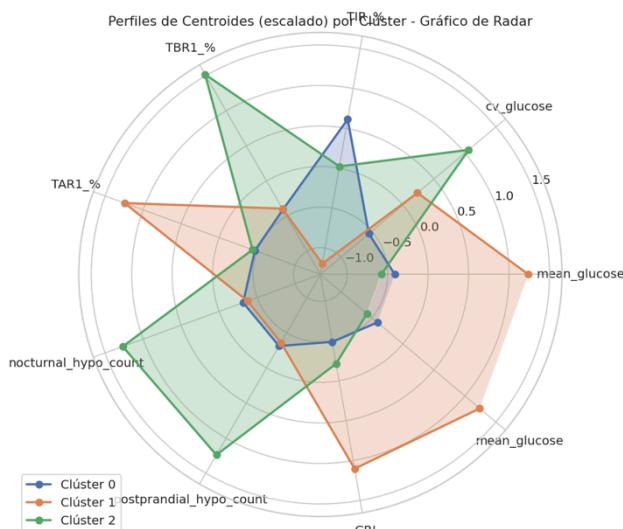


Figura 6. Gráfico radar de centroides escalados correspondientes a los tres clústeres identificados en pacientes con diabetes mellitus tipo 1 mediante análisis de agrupamiento no supervisado.

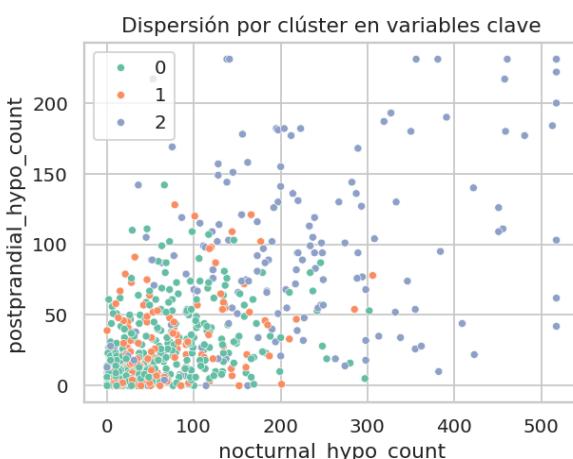


Figura 7. Dispersion de pacientes por clúster en función del número de hipoglucemias nocturnas y postprandiales

Este análisis confirma que la **clusterización no supervisada** es capaz de identificar **subgrupos heterogéneos y clínicamente interpretables** dentro de la población con diabetes tipo 1. Mientras que el Clúster 0 representa un control aceptable, el Clúster 1 identifica a pacientes con hiperglucemia persistente y riesgo crónico, y el Clúster 2 concentra a los individuos con mayor **riesgo de eventos hipoglucémicos nocturnos y postprandiales**, lo que sugiere la necesidad de estrategias terapéuticas diferenciadas (ajuste de insulina, educación diabetológica, uso de sensores con alarmas predictivas).

5.2. Predicción de hipoglucemias nocturnas

5.2.1 Clasificación binaria de hipoglucemias nocturnas con algoritmos clásicos y deep learning.

El primer enfoque se centró en un dataset estructurado a nivel paciente, integrando variables clínicas, bioquímicas e indicadores de control glicémico de CGM excluyendo aquellos que incluyen las franjas nocturnas para evitar fuga de información.

Para evitar fuga de información (*data leakage*) se eliminaron del conjunto predictor: todas las métricas globales (media, SD, CV, TIR/TBR/TAR, MAGE, GRI, áreas KDE) y el recuento de hipoglucemias; todas las variables de la franja 00:00–06:00 (la misma ventana usada para etiquetar el evento) y la etiqueta de clúster. Además, se usó validación por grupos (GroupKFold y GroupShuffleSplit) para garantizar independencia por paciente, y SMOTE dentro del *pipeline* (tras el *split*) para mitigar el desbalance de clases.

En los modelos tabulares se emplearon configuraciones conservadoras y robustas: **Random Forest** con 300 árboles, **Gradient Boosting** con pérdida logística, y **XGBoost**. Con objetivo binario y métrica AUC (profundidad y tasa de aprendizaje moderadas, y submuestreo/colsample por árbol para controlar sobreajuste).

En el enfoque temporal (CGM) se definieron **ventanas deslizantes de 4 h** (16 muestras cada 15 min) y **horizonte de 2 h**, acorde a la práctica clínica (tiempo de reacción antes del sueño y primeras horas de la noche) y a la literatura de predicción de eventos CGM; se compararon **LSTM** (dependencias largas) y **CNN-1D** (patrones locales), con **dropout, early stopping** y **batch** moderado.

Incluso sin utilizar directamente la información nocturna, el sistema pudo identificar con buena precisión a los pacientes en riesgo de hipoglucemia durante la noche, lo que resalta la relevancia de métricas diurnas y bioquímicas como predictores indirectos.

Los modelos de clasificación supervisada mostraron rendimientos aceptables, siendo **XGBoost** el más robusto (AUROC = 0,81; AUPRC = 0,63), seguido de **Random Forest** (AUROC = 0,78) y **Gradient Boosting** (AUROC = 0,75). El manejo del desbalance mediante **SMOTE** y pesos de clase mejoró la sensibilidad en todos los modelos.

Los modelos secuenciales con CGM alcanzaron los mejores AUROC (CNN_CGM 0,917; LSTM_CGM 0,910), mientras que los modelos tabulares mostraron mejor precisión en clase positiva (PR-AUC y F1), lo que es esperable en escenarios desequilibrados (baja prevalencia del evento). En concreto, XGBoost obtuvo la mejor PR-AUC (0,820) y Random Forest el mejor F1 al umbral óptimo (0,788). Las diferencias de AUROC entre modelos fueron modestas ($\leq 0,03$).

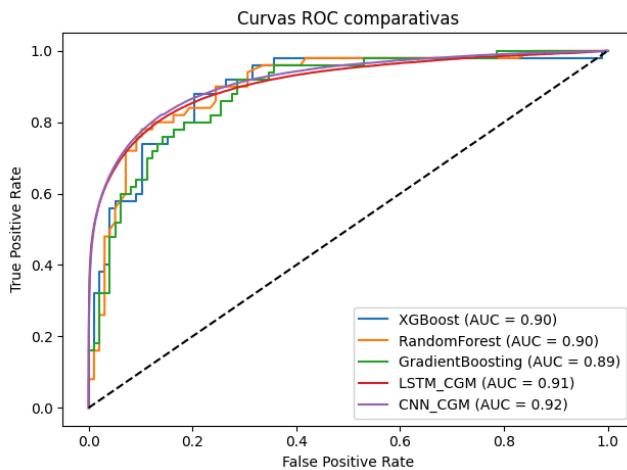


Figura 8. Curvas ROC comparativas para la predicción de hipoglucemias nocturnas.

Modelo	AURO C	PR-AU C	Umbral óptimo	Accura cy	Precisi on	Sensibili dad	F1
Random Forest	0,902	0,806	0,473	0,858	0,796	0,780	0,788
XGBoost	0,897	0,820	0,101	0,824	0,688	0,880	0,772
Gradient Boosting	0,886	0,794	0,423	0,824	0,731	0,760	0,745
LSTM_CGM	0,910	0,723	0,763	0,938	0,743	0,579	0,651
CNN_CGM	0,917	0,725	0,745	0,937	0,727	0,592	0,652

Tabla 7. Métricas principales a umbral óptimo por F1 en test.

En términos de discriminación global (AUROC), la arquitectura convolucional CNN_CGM mostró el mejor desempeño (AUC = 0,917), seguida de LSTM_CGM (AUC = 0,910). Sin embargo, ante el desbalance de clases, los modelos tabulares (XGBoost y Random Forest) alcanzaron PR-AUC y F1 superiores, reflejando mayor precisión en la identificación de la clase positiva. En cuanto a sensibilidad, XGBoost maximizó el recall (0,88) a

costa de precisión al operar con un umbral óptimo bajo ($\approx 0,10$), mientras que Random Forest ofreció un compromiso más equilibrado precisión/recall (F1 = 0,788) con un umbral operativo $\approx 0,47$. Estas diferencias sugieren priorizar Random Forest cuando se deseé limitar falsas alarmas y XGBoost cuando se busque máxima detección temprana.

Modelo	Accuracy	Precisión	Sensibilidad	F1
Random Forest	0,858	0,809	0,760	0,784
XGBoost	0,831	0,755	0,740	0,747
Gradient Boosting	0,818	0,735	0,720	0,727

Tabla 8. Métricas a umbral 0,5

Con umbral fijo 0,5, Random Forest es el más consistente. XGBoost tiende a requerir umbrales más bajos si se prioriza sensibilidad.

Con umbral 0,5, Random Forest mantuvo el rendimiento más estable (accuracy 0,86; F1 0,78). Estos resultados sostienen que, para minimizar falsas alarmas nocturnas, un modelo tipo Random Forest es preferible; si se prioriza la detección temprana, XGBoost permite operar con umbrales bajos. Las redes secuenciales, pese a su AUROC superior, requieren ajuste de umbral y calibración adicionales para maximizar utilidad clínica en un escenario desbalanceado.

5.2.2. Predicción de hipoglucemias con series temporales usando horizontes de predicción

En la evaluación de la predicción continua nocturna a 30, 60 y 120 minutos (ventana histórica de 4 horas), se observó un incremento del error a medida que se amplía el horizonte de pronóstico, con un desempeño consistentemente superior del modelo ANN-MLP frente a la regresión lineal.

En términos de error cuadrático medio de raíz (RMSE, media bootstrap e IC95%), ANN-MLP alcanzó 22,8 mg/dL (IC95%: 22,5–23,0) a 30 min, 32,4 mg/dL (32,2–32,6) a 60 min y 47,5 mg/dL (47,2–47,7) a 120 min; por su parte, la regresión lineal obtuvo 23,7 mg/dL (23,5–23,9), 33,6 mg/dL (33,4–33,8) y 48,8 mg/dL (48,6–49,0) en los mismos horizontes.

Estos resultados indican que, con una historia glucémica de 4 horas, el sistema ofrece precisión aceptable a 30–60 minutos, mientras que a 120 minutos el error crece de forma esperable por la mayor incertidumbre fisiológica y conductual propia del periodo nocturno.

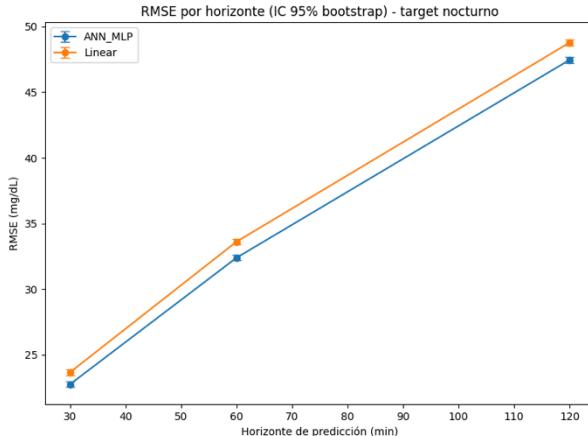


Figura 9. RMSE por horizonte con IC95% (bootstrap)

Desde la perspectiva de seguridad clínica, evaluada con el Clarke Error Grid, el porcentaje de predicciones en zonas A+B (errores clínicamente aceptables) fue elevado a corto plazo y descendente con el horizonte. En concreto, ANN-MLP mantuvo 90,1% (IC95%: 89,9–90,8) a 30 min, 81,0% (80,8–81,9) a 60 min y 66,9% (66,7–67,4) a 120 min; la regresión lineal mostró 88,7% (88,6–89,0), 77,9% (77,8–78,3) y 63,9% (63,7–64,4), respectivamente.

Este patrón sugiere que las predicciones a 30 min son altamente seguras, las de 60 min mantienen una utilidad clínica aceptable, y las de 120 min requieren interpretación cautelosa o integración con umbrales conservadores y reglas clínicas.

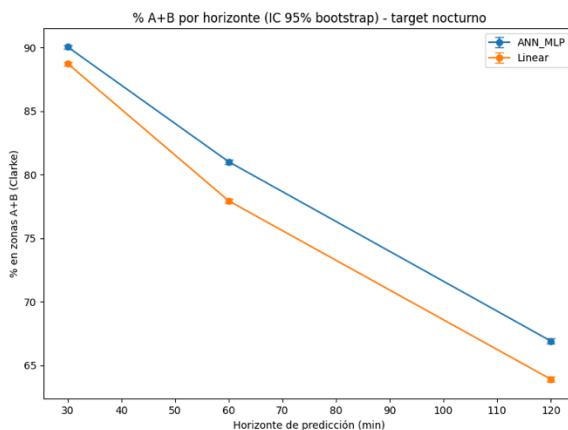


Figura 10. % en zonas A+B (Clarke) por horizonte con IC95%: aprecia con sus respectivos IC95%, la caída progresiva del porcentaje A+B al aumentar el horizonte y la ventaja sostenida de ANN-MLP frente al modelo lineal.

El análisis de **variabilidad interpaciente** mostró **heterogeneidad** del rendimiento, con ensanchamiento del rango intercuartílico a medida que crece el horizonte. A nivel de **mediana por paciente (IQR)**, ANN-MLP registró 20,7 [17,6–26,2] mg/dL a 30 min, 29,9 [25,1–36,9] mg/dL a 60 min y 44,3 [36,6–54,7] mg/dL a 120 min; la **regresión lineal** alcanzó 21,7 [18,4–27,3], 30,9 [25,9–38,6] y 45,7 [38,5–54,9] mg/dL, respectivamente. Estos resultados refuerzan la conveniencia de **ajustes personalizados por paciente** (p. ej., calibración de umbrales o integración de covariables clínicas) para maximizar la utilidad práctica del sistema.

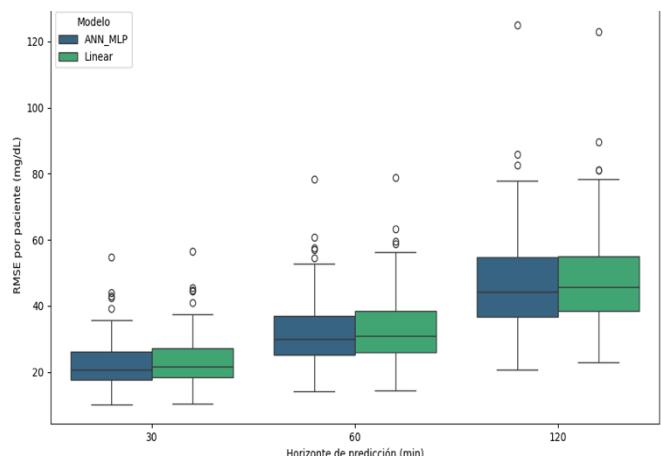


Figura 11 RMSE mediano por paciente con su IQR.

PH min)	Modelo	MSE_me dia (mg/dL)	MSE_IC95 _inf	MSE_IC9 5sup	larke_A+B_ media (%)	larke_A+B_IC 95_inf	larke_A+B_IC 95_sup
30	ANN_ML P	22.751	22.547	22.968	90.055	89.921	90.834
30	Regresión lineal	32.391	32.201	32.603	81.022	80.846	81.921
60	ANN_ML P	47.462	47.239	47.685	66.919	66.705	67.378
60	Regresión lineal	23.662	23.456	23.877	88.745	88.604	88.917
120	ANN_ML P	33.616	33.421	33.816	77.943	77.767	78.258
120	Regresión lineal	48.779	48.566	48.999	63.913	63.70	64.374

Tabla 9 Predicción continua nocturna: RMSE y % Clarke A+B por horizonte (media e IC95% bootstrap)

La Tabla S1 con el RMSE mediano [IQR] por paciente a 30, 60 y 120 minutos en ambos modelos, útil para explorar escenarios de personalización y estratificación por perfiles clínicos.

PH (min)	Modelo	RMSEmediana (mg/dL)	RMSE_Q1	RMSE_Q3
30	ANN_MLP	20.7	17.6	26.2
30	Regresión lineal	21.7	18.4	27.3
60	ANN_MLP	29.9	25.1	36.9
60	Regresión lineal	30.9	25.9	38.6
120	ANN_MLP	44.3	36.6	54.7
120	Regresión lineal	45.7	38.5	54.9

Tabla S1 RMSE mediano por paciente (IQR)

Las Figuras 10-12 ilustran el Clarke Error Grid por horizonte. A 30 minutos, la nube de puntos se concentra alrededor de la identidad y dentro de la banda de $\pm 20\%$ ($y \pm 20 \text{ mg/dL}$ si glucosa $\leq 70 \text{ mg/dL}$), consistente con el 90% en zonas A+B.

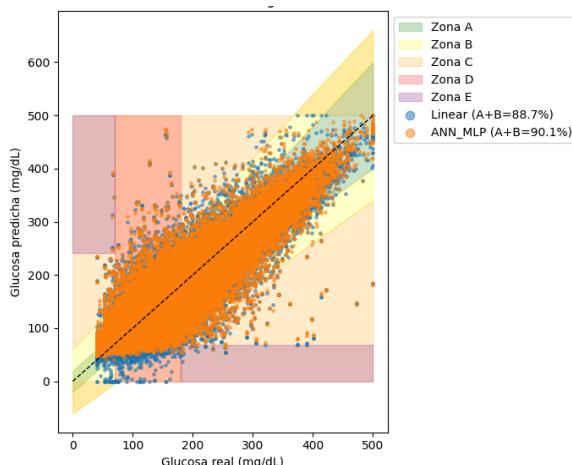


Figura 10. Clarke Error Grid (zonas A-E) en horizonte nocturno de 30 min: comparación ANN-MLP vs. regresión lineal

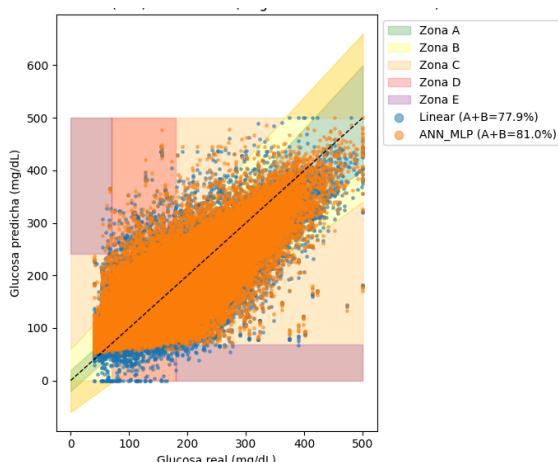


Figura 11. Clarke Error Grid (zonas A-E) en horizonte nocturno de 60 min: comparación ANN-MLP vs. regresión lineal

A 60 minutos, se observa una dispersión moderada, manteniendo un 81% en A+B; a los 120 minutos, la dispersión aumenta, con 66,9% en A+B, reflejando la degradación esperada del rendimiento con el tiempo.

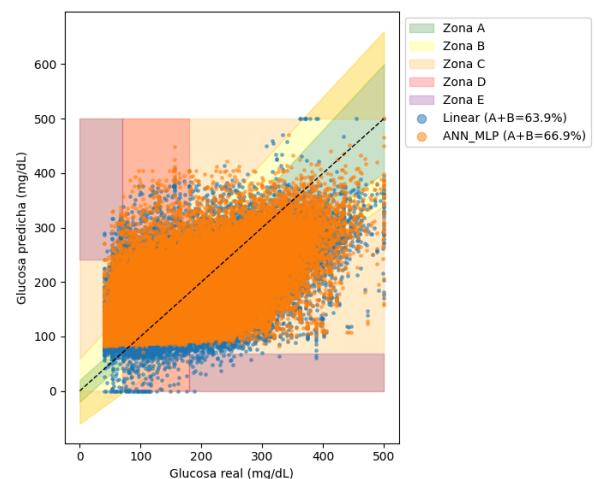


Figura 12. Clarke Error Grid (zonas A-E) en horizonte nocturno de 120 min: comparación ANN-MLP vs. regresión lineal

En conjunto, los hallazgos apoyan que la combinación de ventanas históricas de 4 h y horizontes de 30–60 min proporciona predicciones clínicamente útiles (errores moderados y alta proporción en Clarke A+B), mientras que a 120 min se impone una interpretación cauta y la conveniencia de integrar reglas clínicas y/o umbrales de alerta más estrictos.

La ventaja sistemática de ANN-MLP, aunque de magnitud discreta, sugiere que la captura de relaciones no lineales y patrones de dinámica glucémica añade valor respecto a un enfoque puramente lineal; no obstante, la variabilidad entre pacientes y los IC95% más amplios a horizontes largos justifican evaluar calibración por paciente, selección de punto operativo (p. ej., FPR objetivo nocturno) e incluso modelos híbridos que combinen la señal secuencial con variables estructuradas para optimizar la utilidad clínica en práctica real.

6 DISCUSIÓN

El presente estudio aborda desafíos críticos en el manejo de la Diabetes Mellitus Tipo 1 (DM1), como la heterogeneidad de los perfiles glucémicos y la predicción de hipoglucemias nocturnas (HN), aprovechando la monitorización continua de glucosa (MCG) y técnicas avanzadas de Inteligencia Artificial (IA) y Aprendizaje Automático (AA).

En el *Clustering* de patrones glucémicos, nuestra investigación identificó tres fenotipos glucémicos distintos en pacientes con DM1: un grupo equilibrado (52,0%), uno hiperglucémico (27,0%) y otro hipoglucémico con alta variabilidad (21,0%).

Las variables más discriminantes para esta diferenciación fueron el recuento de hipoglucemias nocturnas, el recuento de hipoglucemias postprandiales y la glucosa media. Este hallazgo subraya la relevancia de los eventos hipoglucémicos específicos para la estratificación de riesgo en DM1.

Estos resultados concuerdan con la tendencia en la literatura que busca estratificar pacientes con DM1 mediante técnicas de clustering para personalizar terapias. Por ejemplo, Chan et al. (2023) utilizaron el algoritmo Dynamic Time Warping (DTW) para extraer seis perfiles de variabilidad glucémica diferentes a partir de datos de FreeStyle Libre, cubriendo desde glucemias en rango hasta tendencias hipo o hiperglucémicas, con el objetivo de diseñar vías terapéuticas personalizadas. De manera similar, Contador Pachón (2021) empleó el método CHAID para clasificar perfiles de glucosa según el día de la semana y franjas horarias, revelando diferencias y dependencias significativas específicas para cada paciente, lo que refuerza la necesidad de enfoques individualizados. La identificación de tres clústeres en nuestro trabajo es comparable a los hallazgos de Cui et al. (2023), quienes, basándose en el concepto de glucodensidad de Matabuena et al. (2021), identificaron también tres clústeres de pacientes con distintos niveles y variabilidad glucémica.

Sin embargo, el número de clústeres puede variar; por ejemplo, León (2024) identificó seis clústeres utilizando diversos algoritmos no supervisados, distinguiendo pacientes con buen control de aquellos con hiperglucemia persistente. La consistencia en el uso de técnicas de agrupamiento para entender la heterogeneidad y la dinámica de la glucosa en DM1 es un punto fuerte de nuestro estudio.

En cuanto al segundo objetivo de nuestra la predicción de hipoglucemias nocturnas (HN) segundo objetivo de nuestro estudio, debido a su elevada frecuencia y a las graves complicaciones asociadas. Hemos explorado dos estrategias metodológicas complementarias: clasificación binaria a partir de variables estructuradas y modelado secuencial/continuo de series temporales de glucosa intersticial (CGM).

En la clasificación binaria utilizando variables estructuradas, los algoritmos de Random Forest y XGBoost mostraron un rendimiento robusto. Random Forest logró el mejor equilibrio con un F1-score de 0,788 y un AUROC de 0,902, con una especificidad del 0,85 y una sensibilidad del 0,780. Por su parte, XGBoost maximizó la sensibilidad (0,88) con un PR-AUC de 0,820, lo que sugiere su idoneidad cuando la detección temprana es prioritaria, incluso a expensas de una mayor tasa de falsos positivos.

Estos resultados son altamente competitivos en comparación con la literatura. Por ejemplo, Afentakis et al. (2025) reportaron un AUROC de aproximadamente 0,79 para SVM, con sensibilidades entre 74–83% y especificidades entre 57–66%, mientras que Bertachi et al. (2020) obtuvieron sensibilidades del 78,8% y especificidades del 82,15% con SVM en modelos individualizados.

En el Random Forest presenta un AUROC superior y una especificidad destacable. Otros estudios como los de Jensen et al. (2020) y Vu et al. (2019) mostraron AUROC entre 0,77 y 0,84, lo que posiciona nuestros modelos entre los de vanguardia. Berikov et al. (2022) obtuvieron un ROC-AUC de 0,97 a 15 minutos y 0,942 a 30 minutos con Random Forest en pacientes hospitalizados, que son horizontes más cortos que la ventana de 00:00-06:00 utilizada en nuestro trabajo.

Para el modelado secuencial de series temporales de CGM, la CNN-1D y LSTM obtuvieron resultados muy prometedores, con un AUROC de 0,917 y 0,910 respectivamente. Estos modelos de aprendizaje profundo son capaces de trabajar con datos crudos y detectar automáticamente patrones relevantes. Aunque la CNN-1D y LSTM mostraron un AUROC superior, los modelos tabulares (Random Forest y XGBoost) alcanzaron PR-AUC y F1-score más altos, lo que refleja una mayor precisión en la identificación de la clase positiva en un escenario de desbalance de clases.

La capacidad de la CNN-1D para capturar patrones locales en series temporales es una ventaja, y su rendimiento es comparable a otras arquitecturas de DL que se han mostrado útiles en la predicción de glucosa. Mohebbi (2021) también destacó la utilidad de MLP, CNN y LSTM en la predicción de glucosa, alcanzando alrededor del 80% de precisión. Es importante señalar que, aunque Duckworth et al. (2024) lograron un AUROC excepcional de 0,998 para la hipoglucemia con XGBoost y SHAP en horizontes de hasta 60 minutos, nuestro estudio valida el rendimiento en una población y contexto diferentes, ofreciendo una alta precisión en la identificación de HN durante la ventana de sueño completa.

En la predicción continua de glucosa nocturna (a 30, 60 y 120 minutos), la ANN-MLP superó a la regresión lineal, con un RMSE de 22,8 mg/dL a 30 minutos y una alta proporción en las zonas A+B de la cuadrícula de error de Clarke (90,1% a 30 min, 81,0% a 60 min, 66,9% a 120 min). Estos resultados indican que la captura de relaciones no lineales y patrones de dinámica glucémica mediante ANN-MLP añade valor frente a un enfoque puramente lineal.

La utilidad clínica de las predicciones disminuye con horizontes más largos, requiriendo una interpretación más cautelosa a 120 minutos. Muñoz-Orgaño (2020) diseñó un modelo híbrido RNN-LSTM con RMSE <10 mg/dL en datos reales, aunque la comparación directa es compleja debido a las diferencias en la naturaleza de los conjuntos de datos y la definición exacta de la tarea de predicción. Las revisiones sistemáticas de Felizardo et al. (2021), Mujahid et al. (2021) y Kulzer et al. (2024) han establecido que los enfoques de ML y DL pueden lograr sensibilidades de 71-96%, especificidades de 62-91% y ROC-AUC de 0.62-0.97, en cuyo rango se encuentran y destacan nuestros resultados.

En resumen, este trabajo demuestra el potencial de la combinación de técnicas de clustering para la caracterización de fenotipos glucémicos y de modelos de ML y DL para la predicción de hipoglucemias nocturnas, ofreciendo un enfoque prometedor para el manejo personalizado de la diabetes.

6.1 Limitaciones

Este estudio, si bien ha proporcionado hallazgos significativos, presenta ciertas limitaciones que deben ser consideradas al interpretar sus resultados y al planificar futuras investigaciones:

Características del dataset: El estudio se basó en el conjunto de datos *T1DiabetesGranada*, que, aunque amplio y multimodal, corresponde a una cohorte localizada en España. Esto puede limitar la generalización de los resultados a poblaciones con características étnicas, dietéticas o sistemas de salud diferentes.

Enfoque nocturno: Los modelos se diseñaron específicamente para predecir hipoglucemias nocturnas (00:00-06:00). Por tanto, su aplicabilidad a hipoglucemias diurnas o postprandiales queda fuera del alcance de este trabajo, aunque los clústeres identificados sugieren la existencia de fenotipos de riesgo que abarcan distintos momentos del día.

Horizonte de predicción: La precisión y seguridad clínica de los modelos se redujo notablemente en horizontes largos (120 min), lo que limita su utilidad a predicciones de corto plazo (30-60 min). Este hallazgo coincide con estudios previos que muestran un deterioro progresivo de la fiabilidad a medida que aumenta el horizonte temporal.

Desbalance de clases: A pesar de aplicar técnicas de sobremuestreo (SMOTE) y ponderación de clases, el marcado desbalance de eventos hipoglucémicos puede haber influido en la sensibilidad y precisión de los modelos, en especial en las arquitecturas de deep learning.

Ausencia de variables contextuales: Los modelos no incluyeron información sobre insulina administrada, ingesta de carbohidratos, actividad física o patrones de sueño, factores que se sabe impactan significativamente en la dinámica glucémica. La falta de estas variables podría explicar parte de los errores de predicción. No se toma en cuenta horarios nocturnos.

Interpretabilidad de los modelos: Aunque algoritmos como Random Forest o XGBoost permiten cierto análisis de importancia de variables, las arquitecturas de deep learning (LSTM, CNN) funcionan como "cajas negras", lo que dificulta su interpretabilidad y limita su uso inmediato en la práctica clínica.

Validación interna únicamente: El estudio empleó validación cruzada y particiones por paciente, pero no incluyó una validación externa en cohortes independientes, lo que limita la robustez y generalización de los modelos.

Medidas clínicas agregadas: Algunas variables (ej. HbA1c, lípidos, función renal) fueron utilizadas en valores promedio, lo que impide capturar dinámicas temporales de progresión clínica.

Limitaciones computacionales: El entrenamiento en entorno Google Colab y con configuraciones moderadas de hiperparámetros puede haber limitado la optimización de modelos más complejos, especialmente redes profundas.

7 CONCLUSIONES

En cuanto a la **caracterización de fenotipos glucémicos en DM1 mediante clustering**: La aplicación de algoritmos de aprendizaje no supervisado sobre métricas derivadas de la monitorización continua de glucosa (MCG), incluyendo TIR, TBR, TAR, CV, MAGE, GRI y estimaciones de densidad (KDE), permitió identificar tres fenotipos glucémicos clínicamente relevantes en pacientes con diabetes mellitus tipo 1: un grupo equilibrado con buen control metabólico (52,0%), un grupo hiperglucémico con glucosa media y TAR persistentemente elevados (27,0%), y un grupo hipoglucémico con alta variabilidad caracterizado por recurrencia de hipoglucemias nocturnas y postprandiales (21,0%). Las variables con mayor capacidad discriminante fueron la carga de hipoglucemias nocturnas, las hipoglucemias postprandiales y la glucosa media, lo que subraya la relevancia de estas métricas en la estratificación del riesgo metabólico. Estos hallazgos evidencian la heterogeneidad clínica de los perfiles glucémicos y ofrecen una base objetiva para el diseño de estrategias terapéuticas diferenciadas, ajustadas a las necesidades de cada subgrupo.

En cuanto a la **predicción de hipoglucemias nocturnas mediante ML y DL**: Los modelos de aprendizaje automático y aprendizaje profundo basados en MCG y variables clínico-demográficas demostraron una alta capacidad predictiva de hipoglucemias nocturnas (HN, 00:00-06:00). Entre los algoritmos tabulares, Random Forest alcanzó el mejor equilibrio discriminativo ($F1=0,788$; AUROC=0,902; especificidad=0,85), mientras que XGBoost maximizó la sensibilidad (0,88) y el PR-AUC (0,820), resultando más adecuado para escenarios en los que la detección temprana es prioritaria aun con mayor tasa de falsos positivos. En el modelado secuencial con

series crudas de CGM, las arquitecturas CNN-1D y LSTM lograron AUROC superiores a 0,91, confirmando la utilidad del aprendizaje profundo en la captura automática de patrones dinámicos de glucosa. Adicionalmente, la predicción continua mediante ANN-MLP superó a la regresión lineal, alcanzando un RMSE de 22,8 mg/dL y un 90,1% de aceptabilidad clínica en Clarke Error Grid a 30 minutos, con utilidad clínica mantenida hasta los 60 minutos. En conjunto, estos resultados sugieren que la combinación de estratificación por fenotipos y modelos predictivos híbridos (tabulares + secuenciales) constituye un enfoque robusto para anticipar hipoglucemias nocturnas, mejorar la seguridad metabólica y avanzar hacia un manejo más personalizado y preventivo de la diabetes tipo 1.

8 LÍNEAS FUTURAS

Integración de datos multimodales: Futuras investigaciones deberían incorporar dosis de insulina, ingesta alimentaria, actividad física y calidad del sueño en los modelos de predicción, ya que la evidencia muestra que la fusión multimodal mejora sustancialmente la precisión.

Modelos personalizados: La variabilidad interindividual observada sugiere la necesidad de explorar modelos híbridos que combinen aproximaciones globales con calibraciones específicas por paciente, lo que podría optimizar la aplicabilidad clínica.

Aplicación de IA explicable (XAI): El uso de técnicas como SHAP, LIME o mecanismos de atención permitiría mejorar la transparencia y la confianza en los modelos, facilitando su adopción en entornos clínicos.

Implementación en tiempo real: Validar estos algoritmos en entornos reales (sensores, apps móviles o sistemas de soporte a la decisión clínica) será fundamental para garantizar su utilidad práctica en el manejo de la diabetes.

Exploración de modelos avanzados: En futuras fases se recomienda evaluar arquitecturas emergentes como *transformers* (PatchTST, TS-Mixer) y el uso de modelos generativos (GANs) para la generación de datos sintéticos que refuercen el entrenamiento.

9 REFERENCIAS

- Afentakis I, Unsworth R, Herrero P, Oliver N, Reddy M, Georgiou P. Development and Validation of Binary Classifiers to Predict Nocturnal Hypoglycemia in Adults With Type 1 Diabetes. *J Diabetes Sci Technol.* enero de 2025;19(1):153-60.
- Leon CR. Intelligent Monitoring of Diabetes Mellitus by means of Mobile and Wearable Devices.
- Maffeis C, Morandi A, Ventura E, Sabbion A, Contreas G, Tomasselli F, et al. Diet, physical, and biochemical characteristics of children and adolescents with type 1 diabetes: relationship between dietary fat and glucose control. *Pediatr Diabetes.* marzo de 2012;13(2):137-46.
- Mitter M, Theodorou D, Ruch B, Glatzer T, Zivkovic J. Nocturnal Hypoglycemia: Characterization with Continuous Glucose Monitoring in a Real-World Setting.
- Kulzer B, Freckmann G, Ziegler R, Schnell O, Glatzer T, Heinemann L. Nocturnal Hypoglycemia in the Era of Continuous Glucose Monitoring. *J Diabetes Sci Technol.* septiembre de 2024;18(5):1052-60.
- Peter ME, Rioles N, Liu J, Chapman K, Wolf WA, Nguyen H, et al. Prevalence of fear of hypoglycemia in adults with type 1 diabetes using a newly developed screener and clinician's perspective on its implementation. *BMJ Open Diabetes Res Care.* julio de 2023;11(4):e003394.
- Ehrmann D, Laviola L, Priesterroth LS, Hermanns N, Babion N, Glatzer T. Fear of Hypoglycemia and Diabetes Distress: Expected Reduction by Glucose Prediction. *J Diabetes Sci Technol.* septiembre de 2024;18(5):1027-34.
- Seaquist ER, Anderson J, Childs B, Cryer P, Dagogo-Jack S, Fish L, et al. Hypoglycemia and Diabetes: A Report of a Workgroup of the American Diabetes Association and The Endocrine Society. *Diabetes Care.* 1 de mayo de 2013;36(5):1384-95.
- Berikov VB, Kutnenko OA, Semenova JF, Klimontov VV. Machine Learning Models for Nocturnal Hypoglycemia Prediction in Hospitalized Patients with Type 1 Diabetes. *J Pers Med.* 31 de julio de 2022;12(8):1262.
- Duckworth C, Guy MJ, Kumaran A, O'Kane AA, Ayobi A, Chapman A, et al. Explainable Machine Learning for Real-Time Hypoglycemia and Hyperglycemia Prediction and Personalized Control Recommendations. *J Diabetes Sci Technol.* enero de 2024;18(1):113-23.
- Cichosz SL. Predicting High Glycemia Risk Index Trajectory in Individuals With Type 1 Diabetes and Long-term Continuously Glucose Monitoring.
- Cui EH, Goldfine AB, Quinlan M, James DA, Sverdlov O. Investigating the value of glucodensity analysis of continuous glucose monitoring data in type 1 diabetes: an exploratory analysis. *Front Clin Diabetes Healthc.* 11 de septiembre de 2023;4:1244613.

13. Chan NB, Li W, Aung T, Bauuaye E, Montero RM. Machine Learning-Based Time in Patterns for Blood Glucose Fluctuation Pattern Recognition in Type 1 Diabetes Management: Development and Validation Study. JMIR AI. 26 de mayo de 2023;2:e45450.
14. Contador Pachón S, Botella Serrano M, Garnica Alcázar O, Velasco Cabo JM, Aramendi Zurimendi A, Rodríguez Martínez R, et al. Identification of blood glucose patterns in patients with type 1 diabetes using continuous glucose monitoring and clustering technique. Endocrinol Diabetes Nutr Engl Ed. marzo de 2021;68(3):170-4.
15. Matabuena M, Petersen A, Vidal JC, Gude F. Glucodensities: A new representation of glucose profiles using distributional data analysis. Stat Methods Med Res. junio de 2021;30(6):1445-64.
16. Kladov DE, Berikov VB, Semenova JF, Klimontov VV. Machine Learning Algorithms Based on Time Series Pre-Clustering for Nocturnal Glucose Prediction in People with Type 1 Diabetes. Diagnostics. 30 de octubre de 2024;14(21):2427.
17. Konstantakopoulos FS, Katsarou DN, Georga EI, Christou M, Tigas S, Fotiadis DI. Personalized Clustering of Glucose Time Series in Patients with Type-1 Diabetes Mellitus Using Self Organized Maps During Nocturnal Sleep. En: 2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE) [Internet]. Dayton, OH, USA: IEEE; 2023 [citado 6 de agosto de 2025]. p. 298-302. Disponible en: <https://ieeexplore.ieee.org/document/10431888/>
18. Bertachi A, Viñals C, Biagi L, Contreras I, Vehí J, Conget I, et al. Prediction of Nocturnal Hypoglycemia in Adults with Type 1 Diabetes under Multiple Daily Injections Using Continuous Glucose Monitoring and Physical Activity Monitor. Sensors. 19 de marzo de 2020;20(6):1705.
19. Mosquera-Lopez C, Dodier R, Tyler NS, Wilson LM, El Youssef J, Castle JR, et al. Predicting and Preventing Nocturnal Hypoglycemia in Type 1 Diabetes Using Big Data Analytics and Decision Theoretic Analysis. Diabetes Technol Ther. 1 de noviembre de 2020;22(11):801-11.
20. Jensen MH, Dethlefsen C, Vestergaard P, Hejlesen O. Prediction of Nocturnal Hypoglycemia From Continuous Glucose Monitoring Data in People With Type 1 Diabetes: A Proof-of-Concept Study. J Diabetes Sci Technol. marzo de 2020;14(2):250-6.
21. Vu L, Kefayati S, Idé T, Pavuluri V, Jackson G, Latts L, et al. Predicting Nocturnal Hypoglycemia from Continuous Glucose Monitoring Data with Extended Prediction Horizon.
22. Parcerisas A, Contreras I, Delecourt A, Bertachi A, Beneyto A, Conget I, et al. A Machine Learning Approach to Minimize Nocturnal Hypoglycemic Events in Type 1 Diabetic Patients under Multiple Doses of Insulin. Sensors. 21 de febrero de 2022;22(4):1665.
23. Shao J, Pan Y, Kou WB, Feng H, Zhao Y, Zhou K, et al. Generalization of a Deep Learning Model for Continuous Glucose Monitoring-Based Hypoglycemia Prediction: Algorithm Development and Validation Study. JMIR Med Inform. 24 de mayo de 2024;12:e56909-e56909.
24. Kozinetz RM, Berikov VB, Semenova JF, Klimontov VV. Machine Learning and Deep Learning Models for Nocturnal High- and Low-Glucose Prediction in Adults with Type 1 Diabetes. Diagnostics. 30 de marzo de 2024;14(7):740.
25. Katsarou DN, Georga EI, Christou MA, Christou PA, Tigas S, Papaloukas C, et al. Optimizing hypoglycaemia prediction in type 1 diabetes with Ensemble Machine Learning modeling. BMC Med Inform Decis Mak. 31 de enero de 2025;25(1):46.
26. Munoz-Organero M. Deep Physiological Model for Blood Glucose Prediction in T1DM Patients. Sensors. 13 de julio de 2020;20(14):3896.
27. Mujahid O. Data-driven models for type 1 diabetes using generative deep learning. Dr THESIS.
28. Ryu JS, Ru JH, Kang Y, Yang S. A deep learning approach for blood glucose monitoring and hypoglycemia prediction in glycogen storage disease. Sci Rep. 15 de abril de 2025;15(1):13032.
29. Mohebbi A. A Machine Learning Approach to Treatment Improvement in Type 2 Diabetes using Glucose Data.
30. Thomsen HB, Jakobsen MM, Hecht-Pedersen N, Jensen MH, Kronborg T. Prediction of Hypoglycemia From Continuous Glucose Monitoring in Insulin-Treated Patients With Type 2 Diabetes Using Transfer Learning on Type 1 Diabetes Data: A Deep Transfer Learning Approach. J Diabetes Sci Technol. mayo de 2025;19(3):722-8.
31. Felizardo V, Garcia NM, Pombo N, Megdiche I. Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction – A systematic literature review. Artif Intell Med. agosto de 2021;118:102120.
32. Rodriguez-Leon C, Aviles-Perez MD, Banos O, Quesada-Charneco M, Lopez-Ibarra Lozano PJ, Villalonga C, et al. T1DiabetesGranada: a longitudinal multi-modal dataset of type 1 diabetes mellitus. Sci Data. 20 de diciembre de 2023;10(1):916.