



**Universidad
Europea**

UNIVERSIDAD EUROPEA DE MADRID

ESCUELA DE ARQUITECTURA, INGENIERÍA Y DISEÑO

**MÁSTER UNIVERSITARIO EN ANÁLISIS DE DATOS MASIVOS
(BIG DATA)**

TRABAJO FIN DE MÁSTER

**PREDICCIÓN MULTIMODAL DE NIVELES DE
GLUCOSA EN SANGRE MEDIANTE REDES
NEURONALES PROFUNDAS A PARTIR DE
SENSORES MCG, ACTIVIDAD FÍSICA E
IMÁGENES DE RETINA**

GEMA DE PABLO LAGUNA

Dirigido por

Dr. David Díaz Vico

CURSO 2024 - 2025

Predicción multimodal de niveles de glucosa en sangre mediante redes neuronales profundas a partir de sensores MCG, actividad física e imágenes de retina



Gema de Pablo Laguna

TÍTULO: PREDICCIÓN MULTIMODAL DE NIVELES DE GLUCOSA EN SANGRE MEDIANTE REDES NEURONALES PROFUNDAS A PARTIR DE SENsoRES MCG, ACTIVIDAD FÍSICA E IMÁGENES DE RETINA

AUTOR: GEMA DE PABLO LAGUNA

TITULACIÓN: MÁSTER UNIVERSITARIO EN ANÁLISIS DE DATOS MASIVOS (BIG DATA)

DIRECTOR DEL PROYECTO: Dr. David Díaz Vico

FECHA: Octubre de 2025

Gema de Pablo Laguna

RESUMEN

Este Trabajo de Fin de Máster (TFM) desarrolla un sistema de predicción multimodal de niveles de glucosa en sangre mediante redes neuronales profundas, integrando información procedente de sensores de monitorización continua de glucosa (MCG), registros de actividad física e imágenes de retina.

El proyecto se apoya en el conjunto de datos *AI-READY*, un recurso internacional de referencia que combina datos clínicos, fisiológicos y visuales de más de mil participantes distribuidos en distintas cohortes según su grado de afectación metabólica.

Se han diseñado y evaluado tres arquitecturas de aprendizaje profundo: una red densa multimodal y dos modelos *Transformer* con distinta configuración de metaparámetros. Los resultados muestran que las versiones basadas en *Transformer* alcanzan el mejor rendimiento global, con un MAE medio de 17,5 mg/dL y un RMSE de 25,9 mg/dL, reduciendo la variancia entre participantes y mejorando la estabilidad del entrenamiento. Estas mejoras son especialmente notables en el horizonte de 5 minutos, lo que las hace especialmente útiles para sistemas de alerta temprana.

En conjunto, el trabajo demuestra la viabilidad del enfoque multimodal para la predicción personalizada de glucosa en pacientes con diabetes tipo 2, y sienta las bases para futuras aplicaciones clínicas basadas en inteligencia artificial.

Palabras clave: aprendizaje profundo, predicción de glucosa, modelos multimodales, *Transformer*, monitorización continua de glucosa, inteligencia artificial en salud.

Gema de Pablo Laguna

ABSTRACT

This Master's Thesis presents the development of a multimodal deep-learning system for blood-glucose prediction, integrating continuous glucose monitoring (CGM) signals, physical-activity data, and retinal images.

The study is based on the *AI-READY* dataset, an international multimodal benchmark containing clinical, physiological, and imaging data from over one thousand participants divided into distinct metabolic cohorts.

Three deep-learning architectures were designed and evaluated: a multimodal dense network and two *Transformer*-based models with different hyperparameter settings. The *Transformer* architectures achieved the best overall performance, with an average MAE of 17.5 mg/dL and RMSE of 25.9 mg/dL, improving inter-participant stability and reducing variance across validation folds. The most significant improvements were observed for the 5-minute horizon, supporting their potential use in early-warning prediction systems.

Overall, this work confirms the feasibility of multimodal integration through deep learning for personalized glucose forecasting in type 2 diabetes and establishes a solid foundation for future clinical applications of artificial intelligence.

Keywords: deep learning, glucose prediction, multimodal models, *Transformer* networks, continuous glucose monitoring, artificial intelligence in healthcare.

ACRÓNIMOS Y ABREVIATURAS

Acrónimo / Abreviatura	Significado
ADA	American Diabetes Association
AI-READI	Artificial Intelligence for Reliable and Equitable Advances in Diabetes Intelligence
BIG DATA	Análisis de Datos Masivos
CGM / MCG	Continuous Glucose Monitoring / Monitorización Continua de Glucosa
CNN	Convolutional Neural Network (Red Neuronal Convolucional)
DICOM	Digital Imaging and Communications in Medicine
DM2 / DMT2	Diabetes Mellitus Tipo 2
FPG	Fasting Plasma Glucose (Glucosa en Plasma en Ayunas)
GPU	Graphics Processing Unit (Unidad de Procesamiento Gráfico)
HbA1c	Hemoglobina Glucosilada (A1c)
HIPAA	Health Insurance Portability and Accountability Act
HRV	Heart Rate Variability (Variabilidad de la Frecuencia Cardíaca)
IMC	Índice de Masa Corporal
LSTM	Long Short-Term Memory (Memoria a Corto y Largo Plazo)
MAE	Mean Absolute Error (Error Medio Absoluto)
MSE	Mean Squared Error (Error Cuadrático Medio)
NIH	National Institutes of Health (Institutos Nacionales de la Salud, EE. UU.)
NoDM	Non-Diabetic Mellitus (Cohorte sin diagnóstico de diabetes)
OGTT	Oral Glucose Tolerance Test (Prueba de Tolerancia Oral a la Glucosa)
ReLU	Rectified Linear Unit (Unidad Lineal Rectificada)
RMSE	Root Mean Square Error (Raíz del Error Cuadrático Medio)
RNN	Recurrent Neural Network (Red Neuronal Recurrente)
SHAP	SHapley Additive exPlanations
TFM	Trabajo Fin de Máster
T2D	Type 2 Diabetes (Diabetes Tipo 2)
FFN	Feed Forward Network
CLS Token	Token de clasificación utilizado en arquitecturas Transformer
Cross-Attention	Mecanismo de atención cruzada en modelos multimodales
Epoch	Época de entrenamiento
Dropout	Técnica de regularización para prevenir sobreajuste

Gema de Pablo Laguna

Acrónimo / Abreviatura	Significado
One-hot Encoding	Codificación binaria utilizada para representar categorías

AGRADECIMIENTOS

A mi tutor David, por su confianza en mí y su involucración en el proyecto. Gracias a su dedicación he podido superar cada desafío que se ha presentado en el camino.

A mi familia, por hacerme creer en mis capacidades y por brindarme el respaldo necesario para alcanzar mis metas.

A mis amigos, por ser mi refugio y fuente de motivación.

A todos vosotros, os doy las gracias de corazón. Este logro es tanto mío como vuestro, y sin vuestra ayuda y apoyo, no habría sido posible.

TABLA RESUMEN

TABLA RESUMEN	
Nombre y apellidos:	Gema de Pablo Laguna
Título del proyecto:	Predicción multimodal de niveles de glucosa en sangre mediante redes neuronales profundas a partir de sensores MCG, actividad física e imágenes de retina
Directores del proyecto:	Dr. David Díaz Vico
El proyecto se ha realizado en colaboración de una empresa o a petición de una empresa:	NO
El proyecto ha implementado un producto: <i>(esta entrada se puede marcar junto a la siguiente)</i>	NO
El proyecto ha consistido en el desarrollo de una investigación o innovación: <i>(esta entrada se puede marcar junto a la anterior)</i>	SI
Objetivo general del proyecto:	Predecir los niveles de glucosa en sangre en diferentes horizontes de predicción mediante redes neuronales profundas a partir de sensores MCG, actividad física e imágenes de retina

Índice general

1. INTRODUCCIÓN Y OBJETIVOS	12
1.1. Introducción	12
1.2. Objetivos del proyecto	12
1.2.1. Objetivos específicos	13
1.3. Beneficios del proyecto	13
2. MARCO TEÓRICO	14
2.1. Diabetes tipo 2	14
2.1.1. Diagnóstico	15
2.1.2. Tratamiento	16
2.1.3. Monitorización de glucosa y administración de insulina	16
2.1.4. Complicaciones asociadas	17
2.2. El papel de la tecnología en la gestión de la enfermedad	18
2.3. Estado del arte	18
2.3.1. AttenGluco: Multimodal Transformer-Based Blood Glucose Forecasting on AI-READI	19
2.3.2. A Comparative Study of Transformer-Based Models for Multi-Horizon Blood Glucose Prediction	19
2.3.3. Hybrid Transformer-LSTM Model for Glucose Prediction	19
2.3.4. Lightweight Sequential Transformers for Blood Glucose Level Prediction	19
2.3.5. A Multimodal Deep Learning Architecture for Predicting Interstitial Glucose Based on Life-Log Data	20
3. CONJUNTO DE DATOS	21
4. METODOLOGÍA	23
4.1. Preparación y preprocesamiento de los datos	24
4.1.1. Series temporales de glucosa y actividad física	24
4.1.2. Imágenes de retina	26
4.2. Arquitecturas de red neuronal	28
4.2.1. Modelo denso multimodal	28
4.2.2. Modelo Transformer multimodal	29
4.2.3. Transformer con ajuste de metapárametros	33
4.2.4. Modelo recurrente (LSTM/GRU)	33
4.3. Evaluación y validación del modelo	33
4.3.1. Estrategia de validación cruzada	33
4.3.2. Métricas de evaluación	34
5. RESULTADOS	35
5.1. Estrategia de evaluación y estructura del análisis	35
5.2. Rendimiento global	35
5.3. Rendimiento por horizonte de predicción	36

5.4. Rendimiento por cohortes de participantes	38
5.5. Comparativa y discusión de resultados	40
6. CONCLUSIONES Y LÍNEAS FUTURAS	43
6.1. Conclusiones	43
6.2. Líneas futuras de trabajo	44
7. REFERENCIAS	46
8. ANEXOS	49
8.1. Aspectos éticos, económicos, sociales y ambientales	49
8.1.1. Introducción	49
8.1.2. Impacto social	49
8.1.3. Impacto económico	49
8.1.4. Impacto medioambiental	49
8.1.5. Impacto ético	50
8.1.6. Conclusiones	50
8.2. Presupuesto económico	51
8.3. Documentación técnica complementaria	52
8.3.1. Descripción general del entorno de trabajo	52
8.3.2. Estructura de notebooks y versiones desarrolladas	52
8.3.3. Conjunto de datos empleado	52
8.3.4. SharePoint	52

Índice de Figuras

1.	Estimación del número total de adultos (20–79 años) con diabetes por país en 2024, incluyendo tanto casos diagnosticados como no diagnosticados [2].	14
2.	Medición de glucosa capilar mediante la utilización de glucómetro [10].	16
3.	Sistema de monitorización continua de glucosa [10].	17
4.	Icono de AI-READI [26].	22
5.	Esquema general del flujo de predicción de glucosa.	23
6.	Esquema del flujo del preprocesamiento de glucosa y de actividad física.	25
7.	Esquema del flujo del preprocesamiento de imágenes DICOM de retina.	27
8.	Esquema de la arquitectura del modelo denso multimodal.	29
9.	Estructura interna del bloque Transformer Encoder[29]	30
10.	Esquema de la arquitectura del modelo Transformer multimodal.	32
11.	Rendimiento global por experimento (media de 5 <i>folds</i>).	36
12.	MAE (mg/dL) por horizonte por experimento (media de 5 <i>folds</i>).	38
13.	RMSE (mg/dL) por horizonte por experimento (media de 5 <i>folds</i>).	38
14.	MAE (mg/dL) por cohortes por experimento (media de 5 <i>folds</i>).	40
15.	RMSE (mg/dL) por cohortes por experimento (media de 5 <i>folds</i>).	40

Índice de Tablas

1.	Criterios de diagnóstico de diabetes tipo 1 y 2 según la ADA.	15
2.	Rendimiento global (media $\pm \sigma$ sobre 5 <i>folds</i>)	35
3.	MAE (mg/dL) por horizonte de predicción (media de 5 <i>folds</i>)	37
4.	RMSE (mg/dL) por horizonte de predicción (media de 5 <i>folds</i>)	37
5.	MAE (mg/dL) por cohorte de participantes (media de 5 <i>folds</i>)	39
6.	RMSE (mg/dL) por cohorte de participantes (media de 5 <i>folds</i>)	39
7.	Presupuesto económico	51

1. INTRODUCCIÓN Y OBJETIVOS

1.1 Introducción

La *diabetes mellitus* es una enfermedad crónica que se presenta cuando el páncreas no secreta suficiente insulina o cuando el organismo no utiliza eficazmente la insulina que produce, provocando niveles elevados de glucosa en sangre. Se clasifica principalmente en diabetes tipo 1, diabetes tipo 2 y diabetes gestacional.

La diabetes tipo 2 representa el 90 % de los casos a nivel mundial. Por tanto, se ha convertido en un grave problema de salud pública debido a su creciente prevalencia, ligada al envejecimiento de la población, el sedentarismo, la alimentación inadecuada y el estrés crónico.

La aparición de la diabetes tipo 2 suele ser mucho menos llamativa con respecto a la diabetes tipo 1 e incluso completamente asintomática durante años. En muchas ocasiones, no es diagnosticada hasta que transcurre un largo período de tiempo, dando lugar a complicaciones como enfermedades cardiovasculares, neuropatías, insuficiencia renal o retinopatía diabética.

Como consecuencia, una gestión eficaz de esta enfermedad requiere el uso de herramientas tecnológicas capaces de monitorizar y anticipar las fluctuaciones de la glucosa en sangre, con el objetivo de optimizar el control glucémico, mejorar la calidad de vida y aumentar la satisfacción de los pacientes.

El monitor continuo de glucosa (MCG) se ha consolidado como una de las herramientas tecnológicas clave en este ámbito, al permitir el seguimiento dinámico y en tiempo real de los niveles de glucosa. La utilización de este tipo de dispositivos, en combinación con información sobre el estilo de vida y la actividad física, genera grandes volúmenes de datos que abren la puerta a modelos predictivos basados en inteligencia artificial. Además, otros tipos de información clínica, como las imágenes de retina, pueden contener señales latentes útiles que complementen la predicción a través de un enfoque multimodal.

Este Trabajo de Fin de Máster (TFM) se desarrolla en el contexto del análisis de datos masivos aplicados a la salud, y tiene como objetivo desarrollar un modelo de predicción de niveles de glucosa en sangre empleando redes neuronales profundas. Para ello, se utilizan datos reales del proyecto *AI-READY*, que incluye registros de glucosa obtenidos mediante sensores MCG, datos de actividad física e imágenes de retina en formato DICOM, entre otros datos.

La integración de estos datos se llevará a cabo mediante una arquitectura de aprendizaje profundo multimodal, combinando los datos temporales, como los niveles de glucosa y la actividad física, junto con la información visual procedente de las imágenes médicas.

1.2 Objetivos del proyecto

El objetivo principal de este Trabajo de Fin de Máster (TFM) es el desarrollo de un modelo de predicción de glucosa basado en redes neuronales profundas, a través de la integración de datos de monitorización continua de glucosa (MCG), registros de actividad física e imágenes médicas de retina mediante un enfoque de aprendizaje multimodal.

1.2.1. Objetivos específicos

Se plantean los siguientes objetivos específicos:

- Estudio en profundidad de la enfermedad de la diabetes, en concreto la diabetes tipo 2, así como la importancia de factores relacionados con el estilo de vida, como el control de ingestas y de actividad física en los pacientes.
- Investigación y exploración de modelos predictivos de glucosa en sangre, especialmente de los que utilizan redes neuronales profundas, con enfoques univariantes o multimodales.
- Análisis del conjunto de datos del proyecto *AI-READY*.
- Procesamiento y limpieza de datos temporales, dando lugar a una correcta unificación y sincronización de las señales de glucosa y actividad física.
- Preprocesamiento e integración de imágenes médicas de retina como fuente visual complementaria dentro del modelo predictivo.
- Diseño e implementación de modelos de redes neuronales profundas orientados a la predicción de glucosa futura en horizontes de 5, 30 y 60 minutos.
- Evaluación del rendimiento del modelo predictivo, mediante métricas como RMSE (Root Mean Square Error) y MAE (Mean Absolute Error).
- Comparación del modelo multimodal con modelos univariantes y bivariantes, para estudiar el impacto real de la integración de imágenes médicas y datos de actividad física en la capacidad predictiva del sistema.
- Interpretación de resultados y análisis crítico del comportamiento del modelo, identificando limitaciones técnicas y líneas de mejora futuras.

1.3 Beneficios del proyecto

Este proyecto demuestra cómo la inteligencia artificial puede aplicarse eficazmente al campo de la salud para mejorar el diagnóstico temprano y la gestión de enfermedades crónicas como la diabetes. La solución propuesta integra datos del mundo real y técnicas avanzadas de *deep learning*, lo que permite anticipar riesgos con mayor precisión y apoyar la toma de decisiones médicas de forma personalizada.

2. MARCO TEÓRICO

2.1 Diabetes tipo 2

La diabetes tipo 2 es una enfermedad crónica que se caracteriza por niveles elevados de glucosa en la sangre, como consecuencia de una producción insuficiente de insulina o de una resistencia de los tejidos a poder utilizarla de manera adecuada [1]. Esta patología constituye aproximadamente el 90 % de los casos de diabetes a nivel mundial y se ha consolidado como una de las enfermedades con mayor impacto sociosanitario, debido a su elevada prevalencia y a las complicaciones crónicas asociadas, que afectan a la salud y el bienestar social de las personas que la padecen [23].

A diferencia de la diabetes tipo 1, la diabetes tipo 2 suele presentar un inicio mucho menos llamativo y, con frecuencia, completamente asintomático. Esta ausencia de síntomas dificulta la identificación del momento exacto en el que comienza la enfermedad. Como resultado, suele transcurrir un largo período de tiempo antes de que se diagnostique, lo que puede dar lugar al desarrollo de complicaciones [2].

La diabetes tipo 2 puede manifestarse a cualquier edad, incluso durante la infancia; sin embargo, su aparición es más frecuente en personas de mediana edad y ancianos. La probabilidad de desarrollar esta enfermedad aumenta significativamente en presencia de determinados factores, como antecedentes familiares, obesidad, sedentarismo, o la coexistencia de otras condiciones de salud como la hipertensión arterial [4].

Según datos de la Federación Internacional de Diabetes (FID), las proyecciones indican que la diabetes tipo 2 podría alcanzar los 852,5 millones en 2050, lo que supondría un incremento del 45 % respecto a los niveles actuales, impulsado por el envejecimiento de la población y el aumento de factores de riesgo como la obesidad, la urbanización y los estilos de vida sedentarios [2].

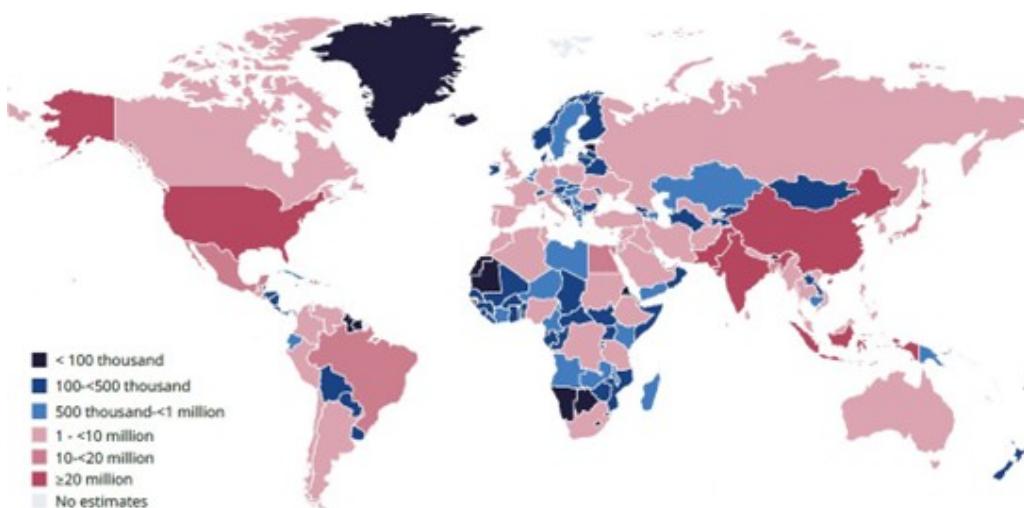


Figura 1. Estimación del número total de adultos (20–79 años) con diabetes por país en 2024, incluyendo tanto casos diagnosticados como no diagnosticados [2].

En la Figura 1 se muestra el número estimado de personas adultas con diabetes por país en 2024, donde destacan especialmente países como China, India y Estados Unidos, con más de 20 millones de casos. No obstante, existe una distribución heterogénea entre los diferentes países del mundo, que refleja tanto la dimensión demográfica como las desigualdades sociales en salud. Además, en muchas regiones con altos niveles de prevalencia, el diagnóstico y tratamiento de la diabetes tipo 2 continúa siendo limitado, lo que agrava su carga sociosanitaria [2].

2.1.1. Diagnóstico

Cuando aparecen síntomas, los profesionales de la salud evalúan a los individuos para determinar si padecen diabetes mellitus tipo 2 (DM2) mediante las siguientes pruebas:

- **Prueba de hemoglobina glucosilada A1c o HbA1c:** análisis de sangre que muestra el nivel promedio de glucosa sanguínea durante los últimos 3 meses [6]. Un nivel de HbA1c de 6,5 % o superior es indicativo de diabetes [7]. Cuanto mayor sea dicho porcentaje, más altos son los niveles de glucosa en sangre [7].
- **Prueba de glucosa plasmática aleatoria:** muestra de sangre en un momento no específico. Independientemente de la última comida, un nivel de glucosa sanguínea aleatorio de 200 mg/dl o superior es indicativo de diabetes. Para confirmar el diagnóstico se requieren pruebas adicionales (p. ej., análisis en ayunas, prueba de tolerancia oral a la glucosa) [6].
- **Examen de glucosa en plasma en ayunas (FPG):** análisis de sangre tras al menos 8 horas de ayuno. Un nivel normal suele estar entre 60 y 100 mg/dl; si en dos análisis diferentes el resultado es 126 mg/dl o mayor, se diagnostica diabetes [6].
- **Prueba de tolerancia oral a la glucosa (OGTT):** si el nivel de glucosa es de 200 mg/dl o superior dos horas después de ingerir una solución de glucosa, se diagnostica diabetes [6].

De acuerdo con la Asociación Americana para la Diabetes (ADA) se establecen los siguientes criterios de diagnóstico (Tabla 1) [6].

Tabla 1. Criterios de diagnóstico de diabetes tipo 1 y 2 según la ADA.

Criterios de diagnóstico			
	Normal	Prediabetes	Diabetes
HbA1c	< 5,7 %	5,7 %-6,4 %	≥ 6,5 %
Prueba de glucosa plasmática aleatoria	-	-	≥ 200 mg/dl
FPG (ayunas)	< 100 mg/dl	100-125 mg/dl	≥ 126 mg/dl
OGTT (2h)	< 140 mg/dl	140-199 mg/dl	≥ 200 mg/dl

2.1.2. Tratamiento

El tratamiento de la diabetes tipo 2 tiene como objetivo principal mantener el nivel de glucosa sanguínea lo más cerca posible del nivel normal, para posponer o evitar complicaciones. Por tanto, el tratamiento busca mantener los niveles de glucosa en sangre en un rango de 80 a 130 mg/dl antes de las comidas y no superar 180 mg/dl dos horas después de la ingesta de alimentos [6]. Para mantener los niveles de glucosa en sangre en un rango adecuado, se deben aplicar las siguientes medidas:

- **Tratamiento con insulina:** las personas con diabetes tipo 2 pueden necesitar tratamiento con insulina de por vida. Lo habitual es requerir varias inyecciones diarias o combinar insulina de acción prolongada e insulina de acción rápida; en algunos casos se sustituye por bombas de insulina programables [6].
- **Control con frecuencia de la glucosa en sangre:** monitorizar niveles en preprandiales o postprandiales, antes de dormir, al despertar y antes de hacer ejercicio o conducir [8].
- **Consumo de alimentos saludables:** basados en alimentos nutritivos, bajos en grasas y ricos en fibra [6].
- **Ejercicio físico regular y mantenimiento de un peso saludable:** el ejercicio aeróbico y de fuerza reduce significativamente los niveles de glucosa en sangre y la HbA1c, con beneficios clínicamente relevantes (reducciones superiores al 0,8 % en HbA1c) [9].

2.1.3. Monitorización de glucosa y administración de insulina

En la actualidad, la utilización de tecnología ha demostrado una mejora significativa en el control glucémico, la calidad de vida, el sueño y la satisfacción global en pacientes con diabetes. Para lograr esto, los pacientes con diabetes tipo 2 deben controlar su diabetes mediante la **monitorización de los niveles de glucosa en sangre**. Entre los métodos más utilizados:

- **Medidor de glucosa capilar (glucómetro):** mide la glucosa capilar a través de la obtención de una gota de sangre proveniente de un pinchazo en el dedo mediante un dispositivo de punción. Una vez tomada la gota de sangre, se recoge con una tira reactiva que se inserta en el glucómetro, el cual la analiza y entrega el resultado [10].



Figura 2. Medición de glucosa capilar mediante la utilización de glucómetro [10].

- **Monitor continuo de glucosa (MCG):** mide continuamente los niveles de glucosa intersticial mediante un sensor insertado bajo la piel, un transmisor y un receptor. El sensor lee el líquido intersticial y envía los resultados al receptor, que muestra niveles y tendencias [10].



Figura 3. Sistema de monitorización continua de glucosa [10].

Esta herramienta representa un avance clave en el tratamiento de la diabetes tipo 2, al generar series temporales densas que reflejan la evolución metabólica del paciente a lo largo del tiempo.

Otra parte del tratamiento es la **administración de insulina**, mediante inyecciones o bombas:

- **Inyecciones:** jeringa con aguja fina para administración subcutánea [8].
- **Bombas de insulina:** dispositivos programables que infunden insulina automáticamente a lo largo del día a través de un conjunto de infusión; un tubo lleva la insulina a una cánula bajo la piel [10].

2.1.4. Complicaciones asociadas

La aparición tardía de síntomas en pacientes con diabetes tipo 2 puede provocar lesiones en otros órganos por el exceso continuado de glucosa y el control inadecuado de la enfermedad [11]. Entre las complicaciones más características:

- **Enfermedades cardiovasculares:** desarrollo de arteriosclerosis y calcificación de paredes arteriales; la coexistencia con otros factores (obesidad, hipertensión, colesterol elevado, tabaquismo) eleva el riesgo de infarto, ictus, cardiopatía o arteriopatía periférica [12].
- **Complicaciones microvasculares:** daño en vasos de menor calibre y microangiopatía diabética; puede causar obstrucción y daño en retina y riñones [12].
- **Nefropatía diabética:** afectación de los vasos sanguíneos renales, pudiendo ocasionar insuficiencia renal [12].
- **Retinopatía diabética:** complicación microvascular en la retina que puede causar pérdida total de visión si afecta a los dos ojos [12].
- **Neuropatía diabética.**
- **Pie diabético:** coexistencia de neuropatía y vasculopatía favorece lesiones isquémicas e infección; causa importante de morbilidad y puede llevar a amputación [13].
- **Disfunción eréctil.**

2.2 El papel de la tecnología en la gestión de la enfermedad

En las últimas décadas, la tecnología en el ámbito sanitario ha crecido de forma exponencial y es clave en el abordaje de enfermedades crónicas como la diabetes. La digitalización de registros médicos, el acceso a sensores biométricos y la monitorización continua han abierto la puerta a una medicina personalizada y basada en datos. Estas herramientas permiten observar el estado actual del paciente, anticipar escenarios futuros y tomar decisiones clínicas más informadas [14, 15].

Aquí entra en juego el *aprendizaje automático (machine learning)*, rama de la inteligencia artificial que permite que un sistema aprenda y mejore de forma autónoma a partir del análisis de grandes cantidades de datos, sin programación explícita. El objetivo es diseñar algoritmos capaces de identificar patrones complejos para realizar predicciones basadas en datos históricos y estimaciones futuras [16].

El aprendizaje automático ya se usa para analizar grandes conjuntos de datos médicos, desde registros de pacientes hasta imágenes e información genómica, identificando patrones que pasan desapercibidos para los clínicos. En DM2, estas detecciones tempranas pueden marcar la diferencia al permitir intervenciones antes de que la enfermedad cause daños significativos [17, 18].

A diferencia de métodos estadísticos tradicionales, los modelos de aprendizaje automático incorporan múltiples fuentes de datos simultáneamente y se adaptan a la evolución temporal del paciente. Entre las técnicas más empleadas:

- **Redes neuronales recurrentes (RNN)**, especialmente LSTM y GRU, para modelar secuencias temporales.
- **Redes convolucionales (CNN)**, para datos visuales como imágenes retinianas.
- **Modelos multimodales**, que combinan distintos tipos de entrada para mejorar la capacidad predictiva.

Estos enfoques han sido validados en estudios recientes. Un caso destacado en DM2 es un estudio con más de 70 000 pacientes y 78 comorbilidades a partir de registros electrónicos, que logró identificar patrones y factores de riesgo con alto rendimiento, mostrando potencial para la detección temprana y la prevención [19].

Además, existen *chatbots* y aplicaciones de salud basadas en IA que proporcionan información, consejos de estilo de vida y recordatorios de medicación o seguimiento de glucosa; también ofrecen un canal accesible para resolver dudas y mejorar la adherencia terapéutica [18].

2.3 Estado del arte

La predicción de los niveles de glucosa en sangre ha evolucionado notablemente en la última década gracias a la aplicación de modelos de aprendizaje profundo y, más recientemente, a la incorporación de arquitecturas basadas en atención y en fusión multimodal de datos. A continuación, se revisan los principales estudios recientes que han marcado la evolución de este campo:

2.3.1. AttenGluco: Multimodal Transformer-Based Blood Glucose Forecasting on AI-READI

Uno de los trabajos más relevantes en el ámbito de la predicción de glucosa es el modelo *AttenGluco*, desarrollado a partir del conjunto de datos AI-READI, un recurso multimodal de referencia internacional. Este estudio propone una arquitectura Transformer específicamente diseñada para integrar información de glucosa continua con variables de actividad física, mediante mecanismos de atención cruzada y atención multi-escala. La atención cruzada permite que el modelo relacione la señal de glucosa con patrones derivados de la actividad, mientras que la multi-escala facilita la detección de tendencias tanto a corto como a largo plazo. Los resultados muestran mejoras significativas frente a modelos convencionales como CNN-LSTM, especialmente en horizontes de predicción de 30 y 60 minutos, donde el Transformer logra una mayor estabilidad en el error medio cuadrático [20].

2.3.2. A Comparative Study of Transformer-Based Models for Multi-Horizon Blood Glucose Prediction

En la misma línea de investigación, Karagöz, Breton y El Fathi llevaron a cabo un estudio comparativo entre distintas variantes de arquitecturas Transformer aplicadas a la predicción de glucosa en sangre en múltiples horizontes temporales. Utilizando los conjuntos de datos DCLP3 y OhioT1DM, los autores evaluaron modelos como Crossformer, PatchTST e iTransformer, analizando el efecto de la longitud de la historia temporal y los esquemas de tokenización empleados. Los resultados evidencian que la tokenización por parches mejora la capacidad del modelo para procesar secuencias largas de forma eficiente, manteniendo un buen equilibrio entre rendimiento y coste computacional. En todos los casos, los Transformers superaron a los modelos recurrentes en horizontes de predicción de hasta cuatro horas [21].

2.3.3. Hybrid Transformer-LSTM Model for Glucose Prediction

Este estudio propone una combinación de un Transformer y una red LSTM con el objetivo de aprovechar la capacidad del primero para modelar relaciones globales y la habilidad del segundo para capturar dependencias locales. El modelo se evaluó en horizontes de hasta 120 minutos, mostrando un desempeño robusto con más del 96 % de las predicciones dentro de las zonas seguras del Clarke Error Grid, lo que lo convierte en un enfoque prometedor para aplicaciones clínicas. Este trabajo destaca por su enfoque híbrido y su validación mediante métricas clínicas, aunque se limita a pacientes con diabetes tipo 1 y no explora la incorporación de señales adicionales como la actividad física o la frecuencia cardíaca [22].

2.3.4. Lightweight Sequential Transformers for Blood Glucose Level Prediction

El trabajo de Barbato introduce una versión optimizada de los Transformers con el propósito de mejorar su eficiencia computacional, permitiendo su implementación en dispositivos por-

tátiles o de bajo consumo. Este estudio demuestra que es posible reducir drásticamente el número de parámetros del modelo sin perder precisión significativa en las predicciones de glucosa. El diseño secuencial propuesto mantiene la capacidad de capturar dependencias temporales de largo alcance con un coste computacional mucho menor, lo que lo convierte en una propuesta relevante desde el punto de vista de la escalabilidad práctica. Sin embargo, al centrarse exclusivamente en señales de glucosa y en pacientes con diabetes tipo 1, el modelo no aborda la complejidad que implica la integración multimodal ni la validación clínica de sus resultados [23].

2.3.5. A Multimodal Deep Learning Architecture for Predicting Interstitial Glucose Based on Life-Log Data

Finalmente, este estudio amplía el enfoque multimodal al integrar datos de glucosa con información contextual procedente de registros de vida diaria o *life-log*, como actividad, sueño o alimentación. La arquitectura propuesta está diseñada para manejar la asincronía y la ausencia parcial de datos entre modalidades, mejorando la robustez de las predicciones frente a entornos reales. Los autores evidencian que el modelo supera en rendimiento a aquellos basados únicamente en MCG, especialmente cuando algunas fuentes de información presentan pérdidas o ruido [24].

3. CONJUNTO DE DATOS

El conjunto de datos utilizado ha sido el de *AI-READI* (Artificial Intelligence for Reliable and Equitable Advances in Diabetes Intelligence), un proyecto de referencia internacional financiado por los *National Institutes of Health* (NIH) de Estados Unidos y desarrollado dentro de la iniciativa Bridge2AI [25]. El programa tiene como finalidad crear infraestructuras de datos abiertas, éticas y equitativas que faciliten la investigación basada en inteligencia artificial en el ámbito sanitario. En particular, *AI-READI* se centra en el estudio de la diabetes tipo 2, proporcionando un recurso multimodal, longitudinal y de alta calidad que permite entrenar y evaluar modelos predictivos orientados a la monitorización, diagnóstico temprano y personalización del tratamiento de esta enfermedad [26].

El conjunto de datos incluye información de más de 1000 participantes, cuidadosamente seleccionados y distribuidos en cuatro cohortes diferenciadas: individuos sin diagnóstico de diabetes (NoDM), personas con prediabetes, pacientes con diabetes tipo 2 tratados con fármacos orales y pacientes con diabetes tipo 2 tratados con insulina. Esta clasificación permite realizar análisis estratificados según el grado de afectación metabólica, lo que aporta un marco sólido para la evaluación de modelos de predicción adaptados a distintos perfiles clínicos.

La recolección de los datos se llevó a cabo en condiciones normales y rutinarias durante un período continuo de dos semanas, combinando dispositivos médicos, sensores portátiles y evaluaciones clínicas. Cada participante fue monitorizado con un sensor *Dexcom G6* de monitorización continua de glucosa (MCG), que proporciona mediciones de glucosa intersticial cada cinco minutos. Esta señal permite representar la dinámica glucémica a lo largo del día y analizar fenómenos como la variabilidad, la respuesta postprandial o la recuperación tras el ejercicio. Las series temporales de glucosa presentan una resolución temporal homogénea y un elevado grado de fiabilidad, lo que las convierte en idóneas para el entrenamiento de modelos de aprendizaje profundo orientados a la predicción en distintos horizontes temporales.

Además de los registros de glucosa, *AI-READI* incorpora información conductual y fisiológica obtenida mediante dispositivos de seguimiento de actividad física. Estos sensores capturan variables como el número de pasos, la intensidad y duración de la actividad, la frecuencia cardíaca y la variabilidad de la frecuencia cardíaca (HRV), indicadores directamente relacionados con el gasto energético y la sensibilidad a la insulina. La integración de estas variables resulta especialmente relevante, dado que la actividad física influye de manera inmediata en los niveles de glucosa y permite dar respuesta a las fluctuaciones observadas en la señal del MCG.

El conjunto de datos también incluye información proveniente de sensores ambientales, encargados de registrar condiciones como temperatura, humedad relativa y calidad del aire. Estas variables permiten analizar la posible influencia del entorno físico en el metabolismo glucémico y en la adherencia al tratamiento. Asimismo, se recogen cuestionarios de salud y estilo de vida diseñados para obtener información sobre los hábitos de sueño, la dieta, el nivel de estrés, la medicación habitual y otros factores conductuales que afectan al control glucémico.

Gema de Pablo Laguna

Además, AI-READI incluye imágenes retinianas en formato DICOM (*Digital Imaging and Communications in Medicine*), obtenidas mediante técnicas de retinografía no invasiva. Estas imágenes constituyen un biomarcador visual de gran valor para la detección de complicaciones microvasculares derivadas de la diabetes, como la retinopatía diabética. Su disponibilidad dentro de un marco de datos multimodal posibilita el desarrollo de modelos que integren información fisiológica, visual y conductual en un mismo *pipeline* de predicción.

De manera complementaria, el conjunto de datos proporciona información clínica y demográfica básica de los participantes, incluyendo edad, sexo, índice de masa corporal (IMC), tipo de tratamiento farmacológico, historial de comorbilidades y variables analíticas de laboratorio.

Desde el punto de vista técnico, AI-READI presenta un formato de datos estructurado y estandarizado, con un etiquetado preciso de las variables y metadatos asociados que cumplen los principios FAIR (*Findable, Accessible, Interoperable, Reusable*) [27]. Además, todos los registros se encuentran anonimizados conforme a las directrices del *Health Insurance Portability and Accountability Act* (HIPAA), lo que asegura la confidencialidad y la protección de la información personal [28].

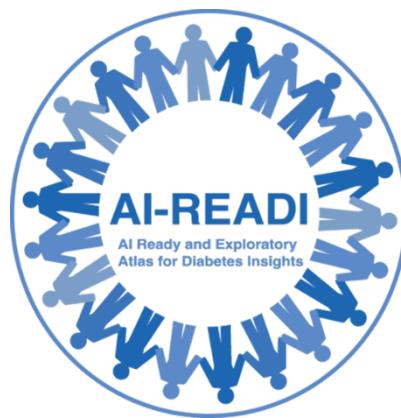


Figura 4. Icono de AI-READI [26].

4. METODOLOGÍA

Este Trabajo de Fin de Máster (TFM) se enmarca en la predicción de niveles de glucosa en pacientes con diabetes mellitus tipo 2 (DMT2) empleando aprendizaje profundo sobre datos continuos de monitorización (MCG), registros de actividad física e imágenes de retina. Cabe destacar que esta combinación de fuentes de datos permite capturar tanto la dinámica temporal del metabolismo glucémico como aspectos estructurales relacionados con la retinopatía diabética, una de las principales complicaciones microvasculares asociadas a esta enfermedad.

La elección de un enfoque basado en aprendizaje automático se justifica por la capacidad de estos algoritmos para identificar patrones no lineales, integrar múltiples tipos de datos y generar predicciones en tiempo real con un elevado nivel de precisión. En concreto, el enfoque multimodal permite fusionar señales heterogéneas mediante el uso de redes neuronales profundas especializadas.

Bajo ese marco, el planteamiento metodológico de este TFM adopta un enfoque experimental y comparativo, en el que se construyeron, entrenaron y evaluaron distintas arquitecturas de redes neuronales profundas, con el objetivo de anticipar la glucosa en horizontes $t+5$, $t+30$ y $t+60$ minutos.

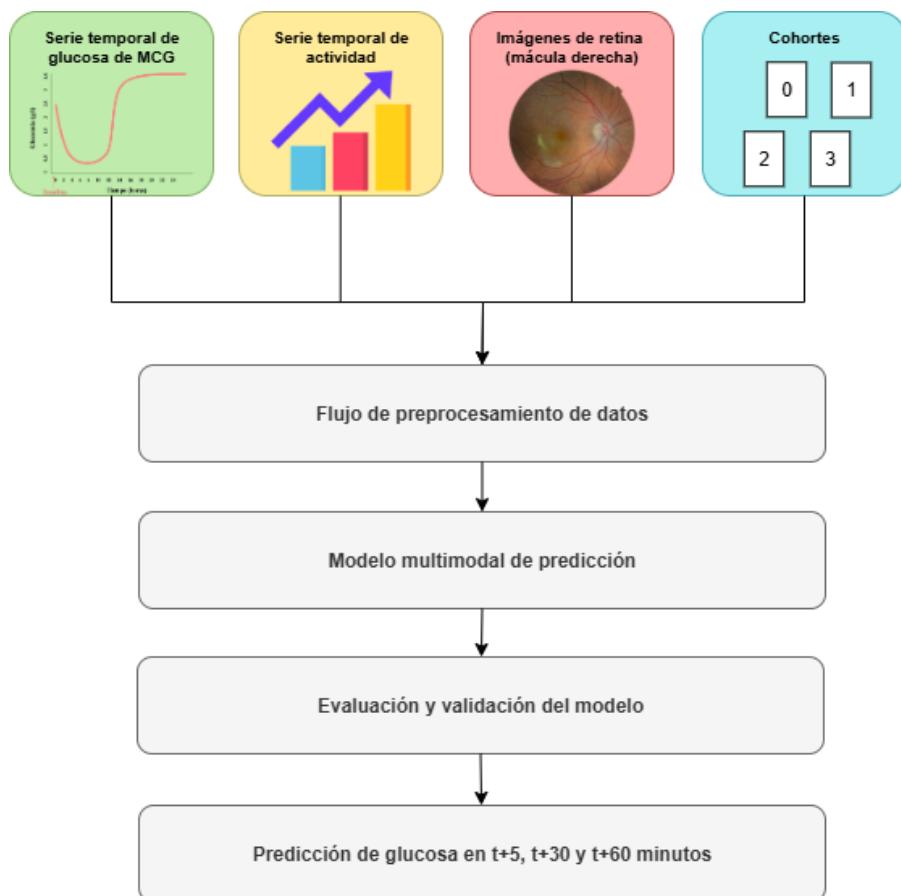


Figura 5. Esquema general del flujo de predicción de glucosa.

En primer lugar, mediante este TFM se busca identificar qué arquitectura ofrece el mejor rendimiento predictivo en términos de error medio absoluto (MAE) y raíz del error cuadrático medio (RMSE). En segundo lugar, se pretende valorar el compromiso entre precisión y complejidad computacional, ya que un modelo más preciso, pero excesivamente costoso en entrenamiento o en predicción puede resultar poco viable para un entorno real de monitorización continua de pacientes.

4.1 Preparación y preprocesamiento de los datos

El conjunto de datos que ha sido utilizado es el del proyecto *AI-READY* (Artificial Intelligence for Reliable and Equitable Advances in Diabetes Intelligence), un proyecto de referencia internacional financiado por los Institutos Nacionales de la Salud (NIH) y desarrollado en colaboración con varias instituciones académicas y médicas de Estados Unidos. Se trata de un conjunto de datos de alta calidad que permite entrenar modelos de inteligencia artificial orientados al tratamiento y predicción de la diabetes tipo 2.

Este conjunto de datos incluye información longitudinal y multimodal de más de mil participantes distribuidos en cuatro cohortes clínicas (NoDM, Prediabetes, T2D-orales y T2D-insulina), monitorizados durante dos semanas mediante sensores Dexcom G6, wearables de actividad y pruebas clínicas complementarias.

4.1.1. Series temporales de glucosa y actividad física

Las señales de glucosa se registraron cada cinco minutos mediante sensores Dexcom G6, proporcionando una serie temporal de alta resolución. Los datos de actividad física, obtenidos con dispositivos de seguimiento, incluyen variables como pasos por minuto, intensidad de movimiento y gasto energético.

Ambas señales se sincronizaron y remuestrearon a una rejilla temporal común de 5 minutos, generando secuencias de 400 instantes previos (equivalentes a unas 33 horas) utilizadas como ventana de entrada para la predicción de la glucosa en los horizontes $t+5$, $t+30$ y $t+60$ minutos.

El preprocesamiento incluyó:

- **Control de calidad:** exclusión de tramos con pérdidas superiores a 20 minutos consecutivos o anomalías abruptas.
- **Interpolación lineal local:** aplicada únicamente a huecos de 1–2 intervalos.
- **Normalización z-score:** parámetros calculados exclusivamente sobre el conjunto de entrenamiento para evitar fuga de información.
- **Codificación de cohortes:** cada participante se representó mediante un vector *one-hot* de 4 dimensiones, identificando el grupo clínico de pertenencia.

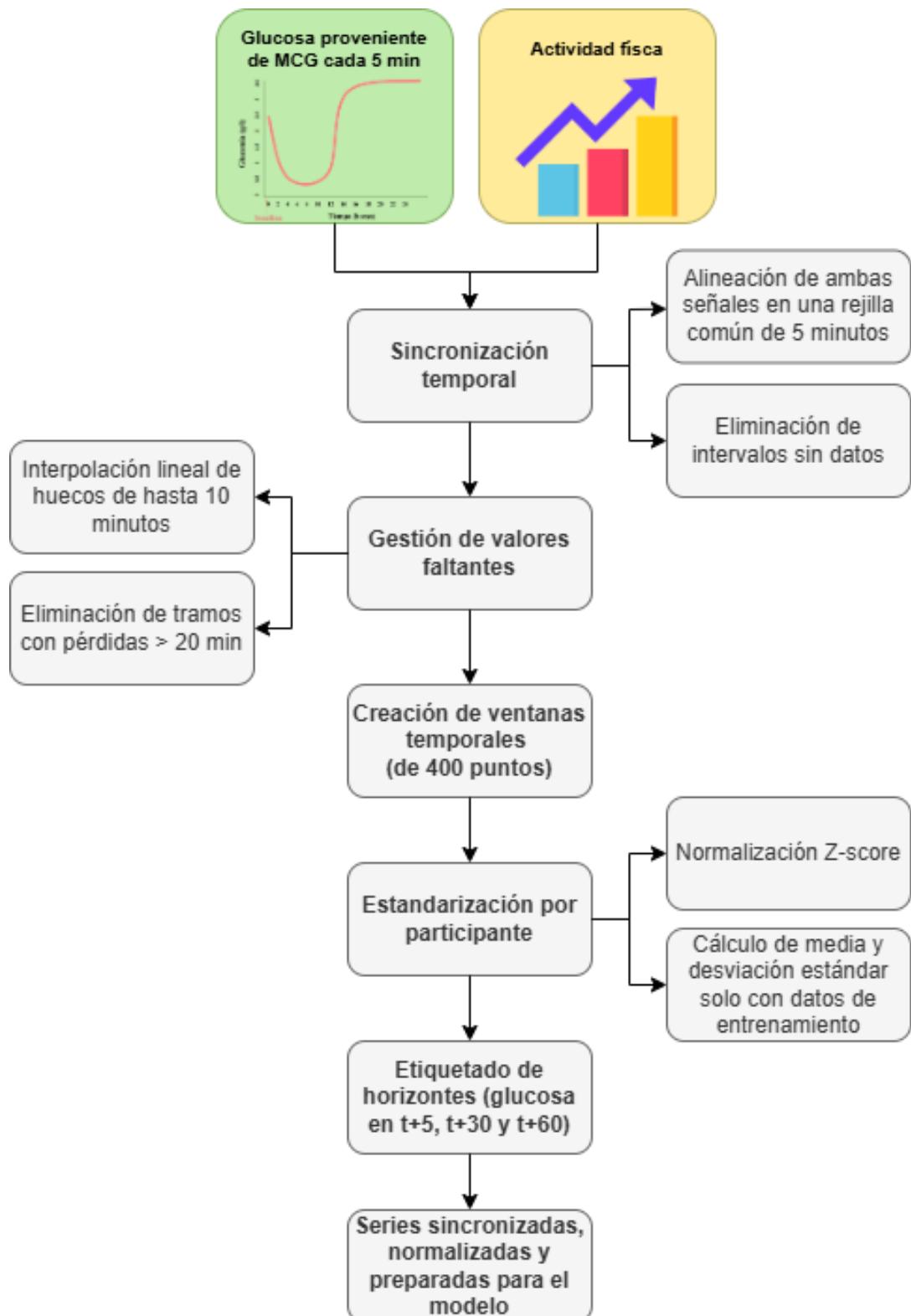


Figura 6. Esquema del flujo del preprocessamiento de glucosa y de actividad física.

4.1.2. Imágenes de retina

Además de las series temporales, el modelo propuesto incorpora retinianas en formato DICOM de la mácula del ojo derecho adquiridas mediante retinografía no midriática estandarizada. Estas imágenes constituyen un biomarcador visual de relevancia para la identificación de microangiopatías y aportan información complementaria a las señales temporales.

El flujo de preprocesamiento de las imágenes incluyó:

- **Lectura de los archivos DICOM:** en esta fase únicamente se seleccionaron las imágenes correspondientes a la región macular, por ser las que aportan mayor valor diagnóstico. Los archivos del conjunto *AI-READY* se encuentran previamente anonimizados, por lo que no fue necesario aplicar un proceso adicional de eliminación de metadatos personales.
- **Aplicación de tabla de correspondencia de intensidades (Modality LUT):** incluida en los metadatos de cada DICOM con el fin de recuperar los valores originales de brillo. En los casos en los que la propiedad Photometric Interpretation presentaba el valor MONOCHROME1, se invirtió la escala de grises para mantener la orientación luminosa coherente entre todas las muestras.
- **Redimensionamiento de las imágenes a 128 × 128 píxeles:** se preserva la proporción anatómica original y se asegura la compatibilidad con la arquitectura convolucional posterior. Tras el ajuste de tamaño, se aplicó una normalización *min–max* por imagen, escalando los valores de intensidad al rango [0,1]. Esta operación se llevó a cabo de manera independiente para cada muestra, con el objetivo de homogeneizar la distribución de intensidades y evitar que las diferencias de exposición o calibración afectaran al aprendizaje del modelo.
- **Implementación de mecanismos de control:** se incluyeron bloques try/except para descartar automáticamente los archivos corruptos o ilegibles, garantizando la integridad del conjunto de datos procesado.
- **Almacenamiento de las imágenes en escala de grises:** se utilizó una única banda, obteniendo tensores de dimensiones 128×128×1, lo que permite optimizar el acceso a disco y acelerar la lectura durante la fase de entrenamiento.
- **Extracción de características:** se introdujeron las imágenes en una rama visual basada en una red convolucional ligera (CNN) integrada dentro del modelo multimodal. Dicha red, construida con capas Conv2D, MaxPooling2D, Flatten y Dense, fue entrenada desde cero junto con las ramas temporales correspondientes a las señales de glucosa y actividad física, generando un vector de representación de la retina que se fusionó posteriormente con las demás modalidades en la etapa de concatenación multimodal.

Todas las transformaciones descritas se aplicaron de manera determinista y se mantuvieron idénticas en los conjuntos de entrenamiento, validación y prueba, con el fin de garantizar la reproducibilidad del proceso experimental.

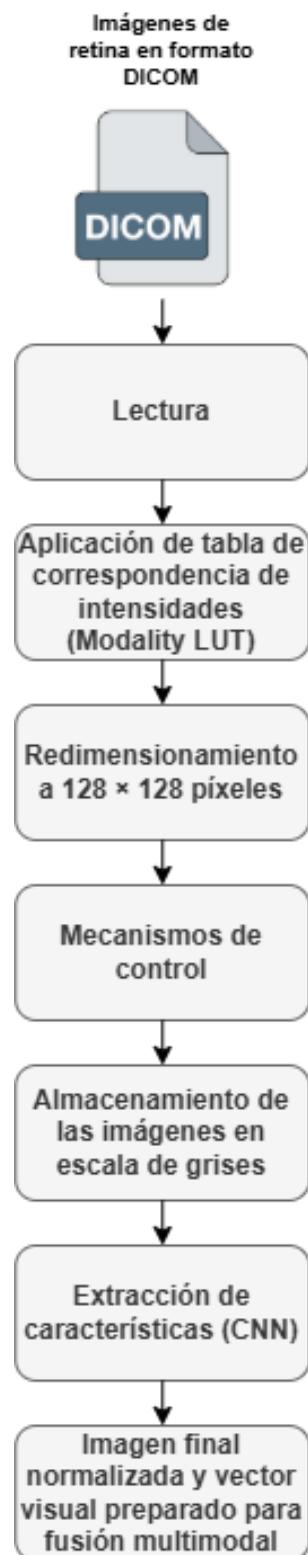


Figura 7. Esquema del flujo del preprocessamiento de imágenes DICOM de retina.

4.2 Arquitecturas de red neuronal

Cada modalidad de entrada se procesó mediante una sub-red específica cuya salida se integró en una capa de fusión multimodal.

4.2.1. Modelo denso multimodal

Este primer modelo constituye la arquitectura base sobre la que se apoyaron los experimentos posteriores. Se diseñó con un enfoque denso completamente conectado, buscando establecer una referencia sólida en términos de precisión, estabilidad y eficiencia computacional.

Se integran cuatro entradas diferenciadas: la serie temporal de glucosa (400 puntos), la serie de pasos (400 puntos), la imagen retiniana de la mácula derecha ($128 \times 128 \times C$) y la codificación *one-hot* de la cohorte clínica (4).

Las ramas temporales aplican una capa densa de 64 unidades con activación ReLU para aprender representaciones compactas de cada señal. La rama asociada a la retina aplica una operación convolucional inicial (16 filtros, 3×3 , ReLU) seguida de *max-pooling* y *flatten*, obteniendo un vector de características de la imagen. La fusión multimodal se realiza mediante concatenación de las cuatro representaciones (glucosa, pasos, retina y cohorte). Sobre este vector combinado se proyecta directamente la capa de salida lineal con tres neuronas, que estima simultáneamente los valores de glucosa en los horizontes $t+5$, $t+30$ y $t+60$ minutos.

El entrenamiento se llevó a cabo con validación *GroupKFold* (cinco particiones por participante), utilizando el optimizador *Adam* y la función de pérdida *MSE*, y empleando generadores por lotes de 32 muestras sin *shuffle* en validación, con el fin de preservar la coherencia temporal de las secuencias. En cada partición del proceso de validación cruzada, el modelo se entrenó durante cinco épocas completas, reentrenándose desde cero en cada iteración del *fold* para evitar la transferencia de información entre conjuntos. En total, el procedimiento supuso 25 ciclos de entrenamiento ($5 \text{ folds} \times 5 \text{ épocas}$). Este número de épocas se fijó de forma experimental atendiendo a las restricciones de capacidad computacional del entorno de ejecución. Aunque un mayor número de iteraciones podría haber permitido una convergencia más precisa del modelo, se optó por este límite para mantener tiempos de entrenamiento razonables en un entorno sin acceso a *GPU* de alto rendimiento. Esta limitación constituye uno de los principales condicionantes del presente estudio, ya que una mayor capacidad de cómputo o un número superior de épocas podrían contribuir a mejorar la estabilidad y la precisión final del modelo.

La evaluación se efectuó a tres niveles: (i) métricas globales agregadas (MAE/RMSE) sobre los tres horizontes, (ii) métricas por horizonte de predicción y (iii) métricas agregadas por cohorte clínica, garantizando independencia entre participantes y evitando cualquier tipo de fuga de información tanto temporal como de sujeto.

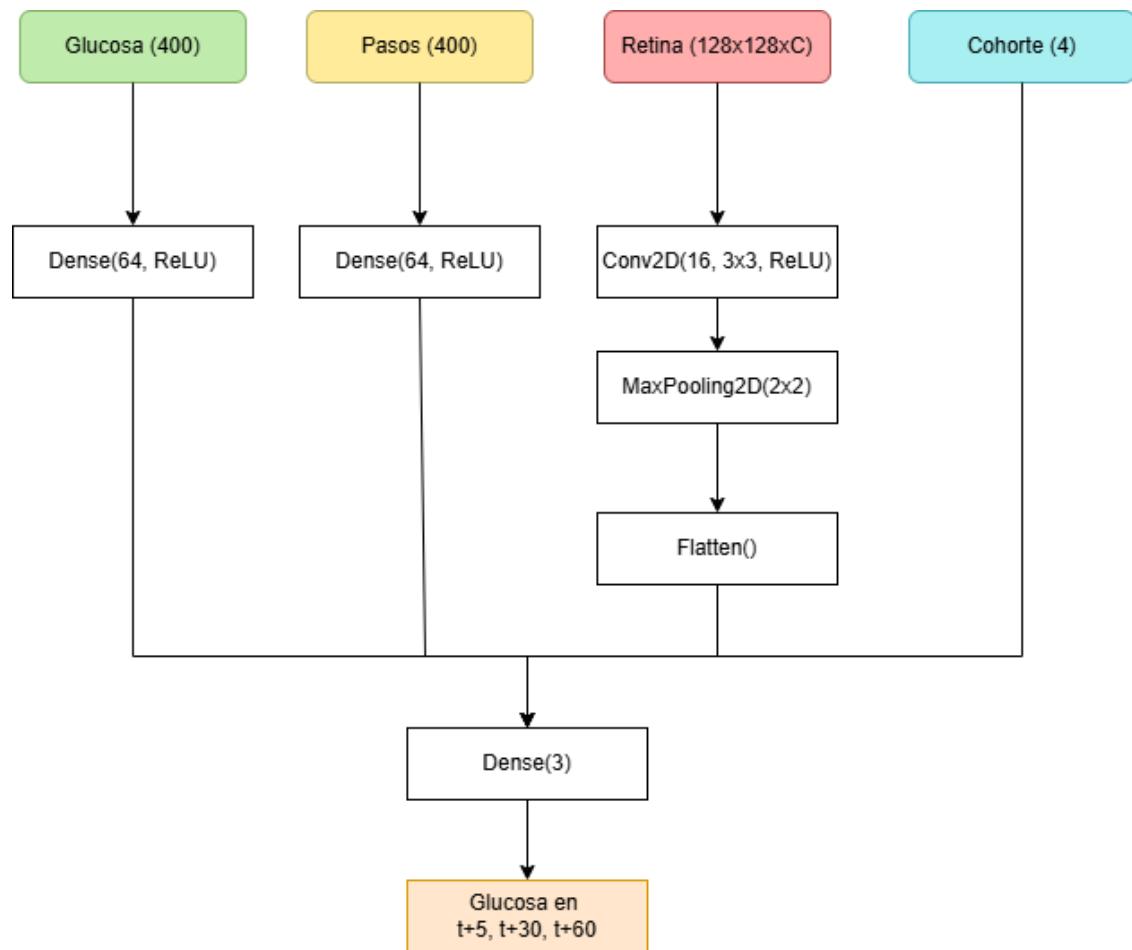


Figura 8. Esquema de la arquitectura del modelo denso multimodal.

4.2.2. Modelo Transformer multimodal

El segundo experimento implementa una arquitectura Transformer multimodal, diseñada para mejorar la representación de las señales temporales y la captura de dependencias de largo alcance entre los valores de glucosa y la actividad física. Este modelo sustituye las ramas densas del diseño anterior por bloques Transformer Encoder, incorporando mecanismos de autoatención que permiten al modelo ponderar dinámicamente la relevancia de cada instante temporal dentro de la secuencia. Este mecanismo de atención se inspira en la arquitectura propuesta por Vaswani en el trabajo “*Attention Is All You Need*”, que introdujo el bloque *Transformer Encoder* como estructura base para el procesamiento secuencial mediante autoatención. Dicho bloque combina operaciones de normalización, proyección multi-cabeza (*Multi-Head Attention*) y capas *Feed-Forward* con conexión residual, permitiendo capturar dependencias de largo alcance dentro de la secuencia temporal. En la Figura 9 se muestra de forma esquemática la estructura interna de este bloque, que constituye el núcleo de la arquitectura utilizada en el presente trabajo [29].

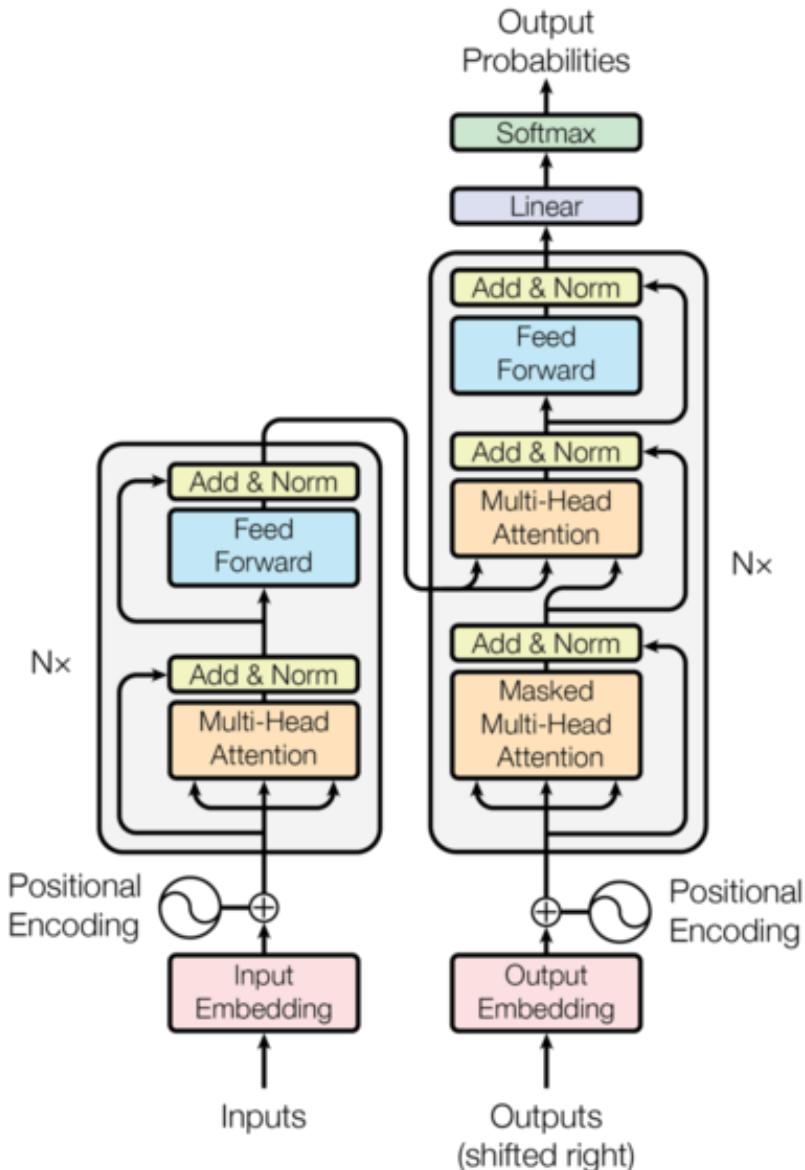


Figura 9. Estructura interna del bloque Transformer Encoder[29]

Las entradas del modelo son las mismas que en la arquitectura base. Además, las ramas asociadas a retina y cohorte mantienen una estructura idéntica a la del modelo denso, mientras que las ramas temporales varían mediante un esquema de codificación secuencial.

Primero, las señales de glucosa y actividad física se transforman mediante *Reshape* ($400 \rightarrow 400 \times 1$) y se concatenan, obteniendo una representación bidimensional de forma (400×2). Esta secuencia se proyecta linealmente a un espacio de 64 dimensiones mediante una capa densa, generando un *embedding* temporal (400×64) sobre el que se añade un *positional embedding* aprendido, necesario para preservar la información del orden temporal.

A continuación, la secuencia resultante se procesa mediante dos bloques Transformer Encoder, cada uno compuesto por:

- Una subcapa de *Layer Normalization* seguida de *Multi-Head Attention* (2 cabezas, *key_dim* = 32) y *Dropout* (0.2), con conexión residual;
- Una segunda subcapa *Feed-Forward* (FFN) con *Dense* (128, ReLU), *Dropout* (0.2), *Dense* (64) y una nueva conexión residual.

Esta estructura permite al modelo combinar información contextual global con relaciones locales entre puntos consecutivos de la serie.

La salida del bloque *Transformer* se resume mediante un vector *CLS readout*, que actúa como un *token* de agregación global de la secuencia y permite condensar la información contextual más relevante en una única representación de 64 dimensiones. Este mecanismo, inspirado en el token [CLS] propuesto por Devlin en el modelo *BERT* [30], resulta especialmente útil en tareas de predicción o clasificación en las que se requiere una salida única representativa de toda la secuencia. Tras esta operación, se aplica una capa *Dropout* (0.2) para reducir el riesgo de sobreajuste. El vector resultante se concatena posteriormente con el *embedding* visual obtenido de la CNN y con el vector de cohorte, conformando la fusión multimodal sobre la que se proyecta la capa de salida lineal encargada de estimar simultáneamente los valores de glucosa en los tres horizontes de predicción $t + 5$, $t + 30$ y $t + 60$ minutos.

El entrenamiento se llevó a cabo utilizando el optimizador *Adam* (tasa de aprendizaje 5×10^{-4}) y la función de pérdida *MSE*, aplicando un esquema de validación *GroupKFold* con cinco particiones por participante, lo que garantiza la independencia entre sujetos. En cada partición del proceso de validación cruzada, el modelo se entrenó durante cinco épocas completas, empleando un tamaño de lote de 32 muestras y manteniendo la secuencia temporal sin *shuffle* en validación. En total, el procedimiento supuso 25 ciclos de entrenamiento ($5 \text{ folds} \times 5 \text{ épocas}$), reentrenando el modelo desde cero en cada partición. El número de épocas se fijó atendiendo a las limitaciones computacionales disponibles, ya que la arquitectura *Transformer* presenta un mayor número de parámetros y, por tanto, un coste de entrenamiento superior al del modelo denso. Para mitigar el riesgo de sobreajuste se aplicó un criterio de *early stopping* basado en el *MAE* de validación, asegurando una convergencia estable dentro del margen de recursos disponible. Esta restricción constituye una de las principales limitaciones del presente estudio, puesto que una mayor capacidad de cómputo permitiría prolongar el entrenamiento y optimizar con mayor profundidad la arquitectura propuesta.

Gema de Pablo Laguna

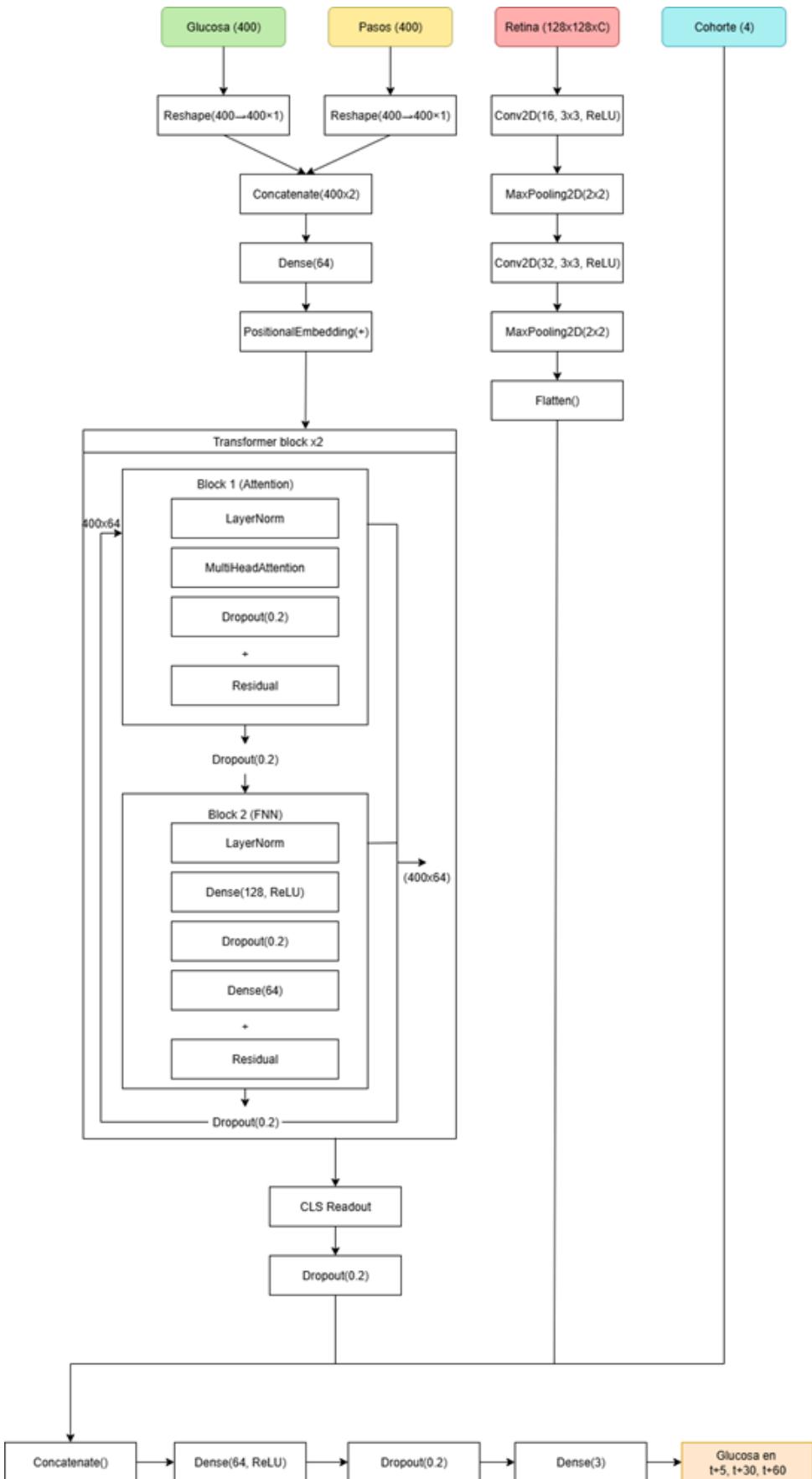


Figura 10. Esquema de la arquitectura del modelo Transformer multimodal.

4.2.3. Transformer con ajuste de metaparámetros

Esta versión se desarrolló con el propósito de analizar la sensibilidad del modelo Transformer multimodal frente a cambios en los metapárametros más relevantes, evaluando si un incremento de la capacidad representativa podía traducirse en una mejora de las métricas predictivas. Para ello, se mantuvo la misma estructura general del modelo anterior, modificando únicamente la configuración interna de la rama Transformer y los parámetros de entrenamiento.

En primer lugar, se introdujo un aumento de capacidad en la atención de la rama temporal. En concreto, se incrementó el número de cabezas de atención de 2 a 8 y se duplicó la dimensionalidad de cada cabeza de 32 a 64. El *embedding* temporal global se mantiene en 64 dimensiones, así como la estructura de los bloques Transformer Encoder.

El resto de la configuración de entrenamiento permanece inalterada respecto a la versión anterior.

4.2.4. Modelo recurrente (LSTM/GRU)

No se implementó una versión recurrente (LSTM/GRU) debido a que la secuencia de 400 puntos temporales abarca un rango suficientemente amplio como para que la red densa o el Transformer capten la dinámica glucémica sin necesidad de recurrencias explícitas.

Además, el enfoque no recurrente simplifica la paralelización en GPU y reduce el tiempo de entrenamiento manteniendo un rendimiento competitivo.

4.3 Evaluación y validación del modelo

En la fase de evaluación se define una estrategia de validación cruzada, un conjunto de métricas cuantitativas y una serie de mecanismos de control para evitar cualquier tipo de fuga de información.

4.3.1. Estrategia de validación cruzada

El proceso de evaluación se basó en una estrategia Group K-Fold con cinco particiones, en la que cada grupo corresponde a un participante distinto del conjunto de datos. De esta manera, se asegura que ningún individuo aparezca simultáneamente en los subconjuntos de entrenamiento y validación, garantizando independencia entre sujetos y evitando fugas derivadas de información personal o temporal.

En cada iteración, cuatro particiones se utilizaron para el entrenamiento del modelo y la restante para su validación. Este proceso se repitió hasta completar las cinco combinaciones posibles, obteniendo así una estimación más robusta y generalizable del rendimiento del modelo. Para evitar sesgos aleatorios, se fijaron semillas reproducibles en todos los experimentos.

4.3.2. Métricas de evaluación

El rendimiento de los modelos se evaluó mediante dos métricas principales ampliamente utilizadas en problemas de regresión en series temporales: Error Medio Absoluto (MAE) y Raíz del Error Cuadrático Medio (RMSE)

Error Medio Absoluto (MAE). El MAE mide la desviación promedio entre los valores predichos (\hat{y}_i) y los valores reales (y_i) de glucosa, ofreciendo una medida intuitiva del error medio cometido.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.1)$$

Raíz del Error Cuadrático Medio (RMSE). El RMSE penaliza de manera más severa los errores grandes y refleja la estabilidad global del modelo, siendo especialmente útil para detectar desviaciones sistemáticas.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.2)$$

Siendo n el número total de observaciones e \bar{y} el valor medio de las observaciones reales.

Estas métricas se calcularon a tres niveles:

- **Global**, considerando todas las predicciones generadas en el conjunto de validación.
- **Por horizonte temporal** ($t+5$, $t+30$ y $t+60$ minutos), para analizar la degradación del rendimiento conforme aumenta la distancia temporal.
- **Por cohorte clínica**, con el fin de estudiar posibles diferencias en la capacidad predictiva del modelo según el grado de afectación metabólica (NoDM, Prediabetes, T2D-orales y T2D-insulina).

5. RESULTADOS

5.1 Estrategia de evaluación y estructura del análisis

El proceso experimental se diseñó con el objetivo de evaluar rigurosamente la capacidad predictiva y la estabilidad de distintas arquitecturas multimodales aplicadas a la predicción de glucosa. Para ello se utilizó un esquema de validación cruzada *GroupKFold* con cinco *folds*, garantizando que los participantes incluidos en cada conjunto de validación fuesen disjuntos respecto a los de entrenamiento. Esta separación estricta evita fugas de información asociadas a características propias de cada sujeto, permitiendo medir con mayor fiabilidad la capacidad de generalización del modelo hacia pacientes no vistos.

En cada iteración del proceso se registraron las métricas **MAE** y **RMSE**, ambas expresadas en mg/dL, de acuerdo con los estándares de evaluación para modelos de predicción de glucosa en sangre. Además de la media global por *fold*, se calcularon resultados desagregados por horizonte de predicción ($t+5$, $t+30$ y $t+60$ minutos) y por cohortes de participantes, lo que permitió analizar el comportamiento de cada modelo tanto en la dimensión temporal, como en distintas subpoblaciones clínicas.

Se comparan tres experimentos principales: el primero, correspondiente al modelo denso multimodal de referencia; el segundo y el tercero, basados en arquitecturas *Transformer* con distinta configuración de metapárametros. El tercer experimento introduce un ajuste más fino de la tasa de aprendizaje y del número de cabezas de atención, con el propósito de mejorar la capacidad de representación sin incrementar la complejidad excesivamente.

5.2 Rendimiento global

El análisis global, que agrupa las predicciones de todos los horizontes y cohortes, constituye el punto de partida para valorar la eficacia general de las tres arquitecturas. Los resultados promedio sobre los cinco *folds* (Tabla 2) muestran que los tres modelos ofrecen un rendimiento robusto, con errores de magnitud moderada, adecuados para la predicción continua de glucosa.

Tabla 2. Rendimiento global (media $\pm \sigma$ sobre 5 folds)

Experimento	MAE medio (mg/dL)	σ (MAE)	RMSE medio (mg/dL)	σ (RMSE)
1	17.892	0.885	26.106	0.602
2	17.518	0.247	25.998	0.348
3	17.547	0.285	25.954	0.345

La red densa multimodal alcanzó un MAE medio de 17,89 mg/dL y un RMSE de 26,10 mg/dL. Mediante las arquitecturas *Transformer* se obtuvieron ligeras mejoras, con MAE de 17,52 y 17,55 mg/dL, y RMSE de 25,99 y 25,95 mg/dL, respectivamente. Aunque las diferencias cuantitativas son reducidas, su consistencia a lo largo de los cinco *folds* es notable. En particular,

la desviación estándar del MAE descendió de 0,89 en la versión densa a 0,25–0,28 en las versiones *Transformer*, lo que indica una mayor estabilidad entre participantes y una disminución de la varianza del error entre sujetos.

Desde un punto de vista metodológico, este resultado sugiere que la introducción de mecanismos de atención y codificación secuencial en el *Transformer* permite representar con mayor precisión los patrones temporales relevantes, reduciendo la sensibilidad del modelo a las particularidades de cada paciente. En términos prácticos, esta mayor estabilidad podría traducirse en una mejor capacidad de generalización en entornos reales, donde los modelos deben enfrentarse a dinámicas metabólicas y hábitos muy diversos.

En la Figura 11 se observa de forma visual el descenso conjunto del MAE y el RMSE en las versiones *Transformer* respecto a la red densa. Aunque el margen absoluto de mejora sea discreto, la consistencia en los cinco *folds* respalda la superioridad técnica del *Transformer* en precisión global y estabilidad de entrenamiento.

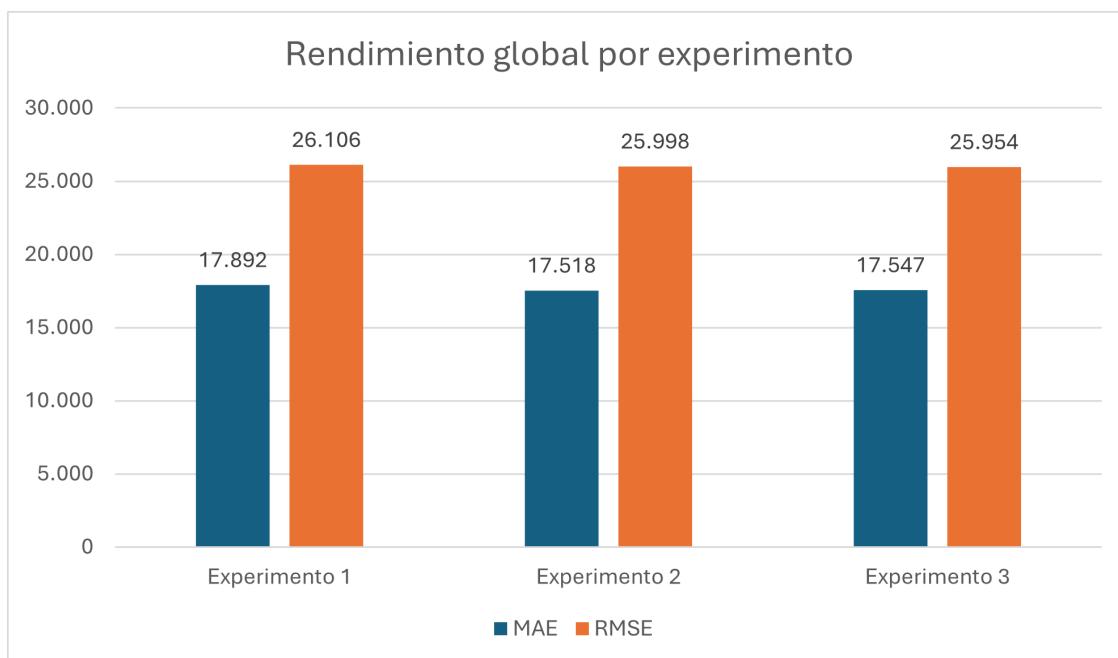


Figura 11. Rendimiento global por experimento (media de 5 folds).

5.3 Rendimiento por horizonte de predicción

Al analizar el error en función del horizonte temporal se confirma la relación esperada entre la distancia temporal de la predicción y la magnitud del error. Cuanto mayor es el intervalo de anticipación, mayor es la incertidumbre inherente y, en consecuencia, los errores aumentan de forma progresiva.

Tabla 3. MAE (mg/dL) por horizonte de predicción (media de 5 folds)

Horizonte	Experimento 1	Experimento 2	Experimento 3
$t+5$ min	13.655	12.669	12.717
$t+30$ min	19.692	19.656	19.697
$t+60$ min	20.330	20.229	20.228

Tabla 4. RMSE (mg/dL) por horizonte de predicción (media de 5 folds)

Horizonte	Experimento 1	Experimento 2	Experimento 3
$t+5$ min	19.354	18.445	18.462
$t+30$ min	28.533	28.633	28.573
$t+60$ min	29.230	29.456	29.387

En el horizonte $t+5$ minutos, los resultados son particularmente positivos para las arquitecturas *Transformer* (experimentos 2 y 3), alcanzando un MAE medio de 12,67 mg/dL y un RMSE de 18,45 mg/dL, y un MAE medio de 12,71 y un RMSE de 18,46 mg/dL, respectivamente. En comparación, el modelo denso multimodal (experimento 1) presenta errores ligeramente superiores (MAE de 13,66 y RMSE de 19,35 mg/dL), lo que supone una mejora relativa cercana al 7 % en MAE y del 5 % en RMSE por parte de los *Transformer*. Esta ganancia refleja la capacidad del mecanismo de autoatención para identificar dependencias de corto plazo en la evolución de la glucosa, aprovechando la información reciente de forma más eficiente que las capas densas tradicionales.

Para el horizonte $t+30$ minutos, las tres versiones presentan resultados prácticamente equivalentes, con MAE en torno a 19,7 mg/dL y RMSE en torno a 28,6 mg/dL, sin diferencias relevantes. En esta escala temporal, la influencia de factores exógenos se intensifica, y las señales fisiológicas previas pierden parte de su capacidad explicativa. Esto explica que, pese a la mayor complejidad arquitectónica del *Transformer*, la mejora sea marginal en horizontes intermedios.

En el horizonte $t+60$ minutos, los errores alcanzan los valores más elevados ($\text{MAE} \approx 20,3$ mg/dL; $\text{RMSE} \approx 29,3$ mg/dL), con diferencias prácticamente nulas entre experimentos. La predicción a una hora vista plantea un desafío considerable, ya que los niveles de glucosa se ven afectados por eventos no siempre observables en los datos disponibles. Aun así, el *Transformer* mantiene su estabilidad, evitando incrementos significativos de error en comparación con el modelo denso.

En conjunto, los resultados muestran que las arquitecturas *Transformer* destacan en el corto plazo, donde las dependencias temporales son más relevantes y los datos multimodales (glucosa, pasos e imágenes de retina) ofrecen información predictiva suficiente. En horizontes más amplios, la ganancia es insignificante, lo que sugiere la necesidad de incorporar variables contextuales adicionales.

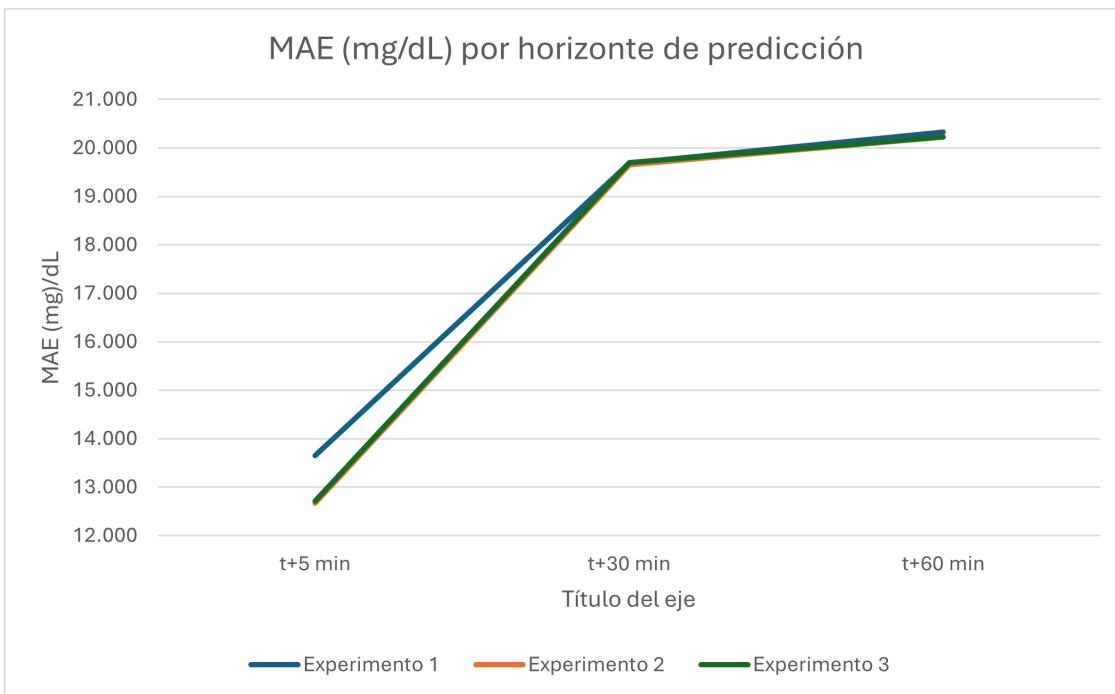


Figura 12. MAE (mg/dL) por horizonte por experimento (media de 5 folds).

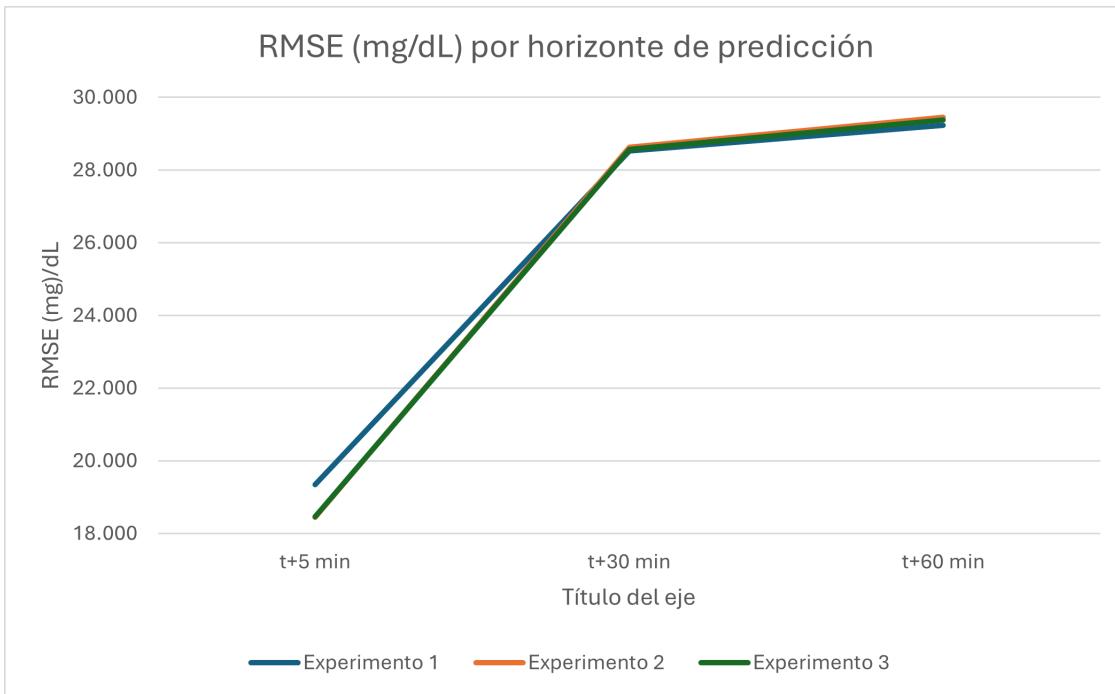


Figura 13. RMSE (mg/dL) por horizonte por experimento (media de 5 folds).

5.4 Rendimiento por cohortes de participantes

El análisis por cohortes permite explorar la influencia de la variabilidad entre individuos en el rendimiento del modelo. Cada cohorte representa un grupo de sujetos con características

clínicas o comportamentales diferenciadas, lo que aporta una dimensión complementaria al análisis temporal.

Los resultados (Tabla 5 y Tabla 6) muestran una tendencia ascendente del error a medida que se avanza hacia cohortes más complejas (cohorte 0: individuos sin diagnóstico de diabetes (NoDM), cohorte 1: personas con prediabetes, cohorte 2: pacientes con diabetes tipo 2 tratados con fármacos orales y cohorte 3: pacientes con diabetes tipo 2 tratados con insulina), con una progresión que refleja la mayor heterogeneidad fisiológica de los sujetos.

Tabla 5. MAE (mg/dL) por cohorte de participantes (media de 5 folds)

Cohorte	Experimento 1	Experimento 2	Experimento 3
0	13.705	13.262	13.310
1	15.413	14.966	15.047
2	20.468	20.223	20.209
3	28.579	28.296	28.242

Tabla 6. RMSE (mg/dL) por cohorte de participantes (media de 5 folds)

Cohorte	Experimento 1	Experimento 2	Experimento 3
0	19.504	19.435	19.328
1	22.133	22.056	22.061
2	28.841	28.856	28.768
3	39.435	39.181	39.095

En las cohortes 0 y 1, las versiones *Transformer* obtienen las mejores métricas, reduciendo el MAE entre un 2% y un 3% respecto a la red densa (0: de 13,70 a 13,26 mg/dL; 1: de 15,41 a 14,97 mg/dL). Estas diferencias, aunque moderadas, confirman que los mecanismos de atención ayudan al modelo a adaptarse mejor a las particularidades individuales de los pacientes con comportamientos más regulares.

En las cohortes de mayor complejidad (2 y 3), las tres versiones convergen hacia un rendimiento prácticamente equivalente. En la cohorte 2, los MAE se sitúan en torno a 20,2 mg/dL y en la cohorte 3 cerca de 28,3 mg/dL, con RMSE entre 28,7 y 39,1 mg/dL, respectivamente. Este patrón sugiere que, cuando los registros presentan alta variabilidad o irregularidad en los patrones de glucosa, el tipo de arquitectura pierde relevancia frente a la imprevisibilidad inherente del proceso. No obstante, incluso en estas cohortes, los modelos *Transformer* mantienen una ligera ventaja en estabilidad entre los *folds*, evitando incrementos abruptos del error en validación.

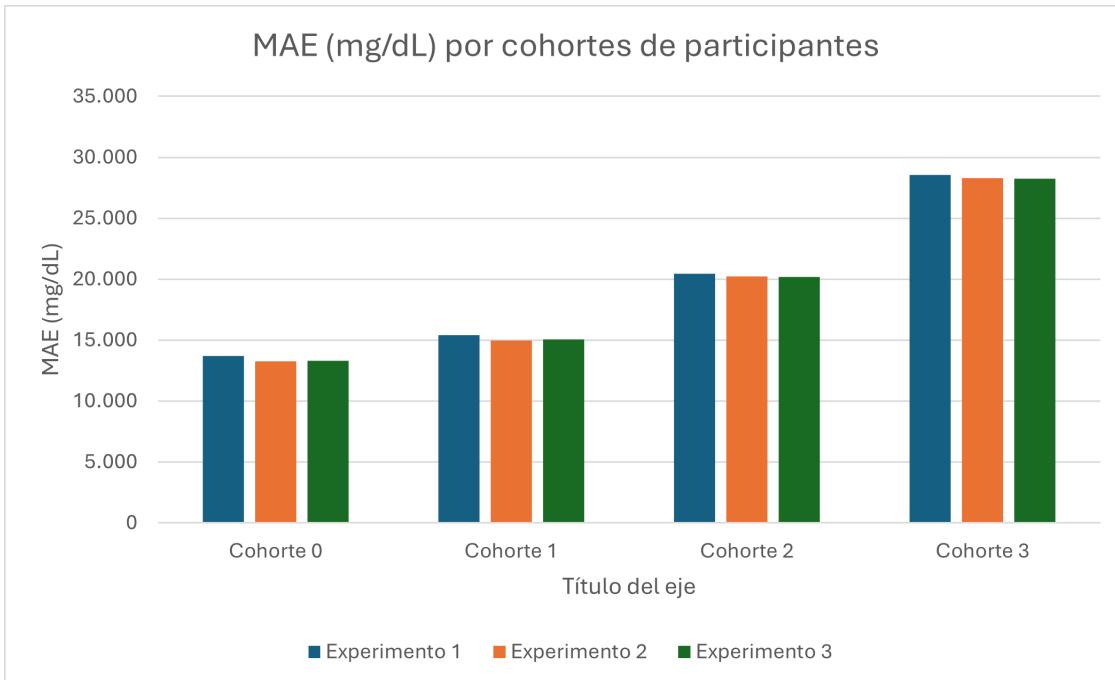


Figura 14. MAE (mg/dL) por cohortes por experimento (media de 5 folds).

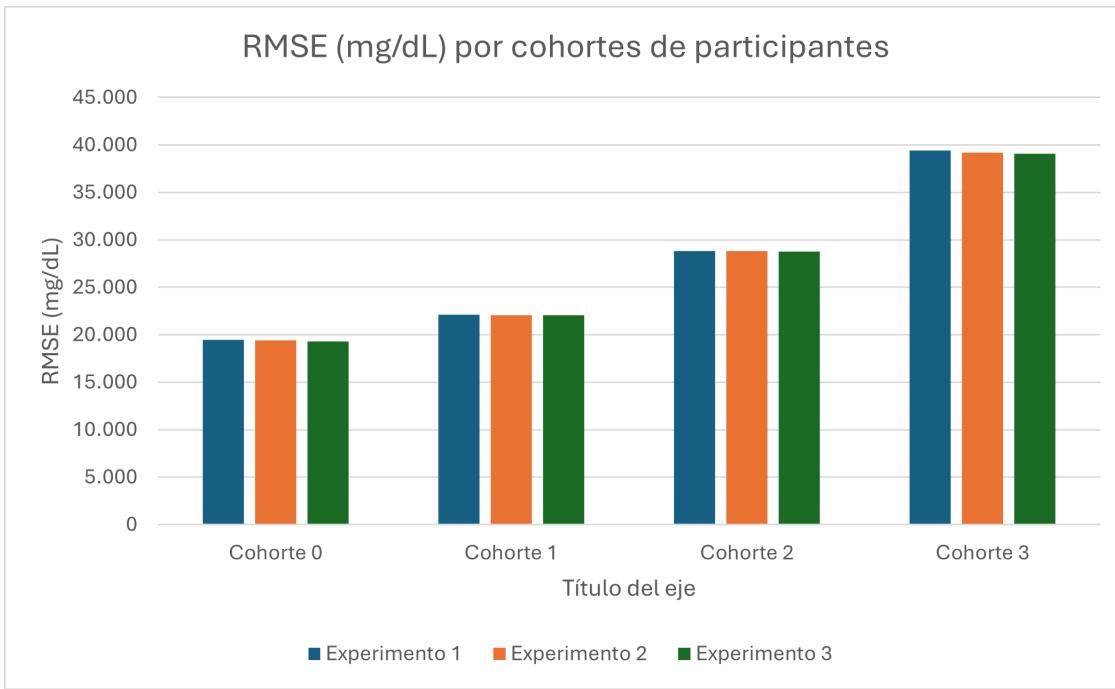


Figura 15. RMSE (mg/dL) por cohortes por experimento (media de 5 folds).

5.5 Comparativa y discusión de resultados

El seguimiento de las curvas de entrenamiento reveló un comportamiento estable en los tres experimentos, con una convergencia clara en torno a las épocas 3 a 5, momento en el que los modelos alcanzaron su menor pérdida de validación. En las versiones *Transformer*, el uso

de *early stopping* permitió restaurar automáticamente los pesos correspondientes al mínimo de validación, evitando sobreajuste y reduciendo el tiempo efectivo de entrenamiento.

Una diferencia importante radica en la estabilidad *inter-fold*. Mientras la red densa presentó una variabilidad más acusada entre *folds* (DE del MAE $\approx 0,89$), las versiones *Transformer* redujeron esta dispersión a valores de 0,25–0,28. Este resultado refleja una mayor robustez del *Transformer* ante cambios en la composición del conjunto de validación, aspecto clave cuando el modelo debe generalizar a nuevos individuos con características no observadas previamente.

Desde el punto de vista computacional, sin embargo, los modelos *Transformer* implican un coste de entrenamiento significativamente mayor. Cada *fold* requirió un tiempo de ejecución varias veces superior al de la versión densa, lo que pone de manifiesto la necesidad de equilibrar el rendimiento y la eficiencia según los objetivos del sistema. En contextos donde la precisión a corto plazo es prioritaria, este sobrecoste puede estar justificado; en entornos de producción con recursos limitados, la arquitectura densa ofrece una alternativa eficiente con un rendimiento muy cercano.

Los resultados obtenidos permiten establecer una conclusión clara sobre el comportamiento relativo de las tres arquitecturas evaluadas. De manera general, las versiones *Transformer* presentan un rendimiento ligeramente superior y más estable que la red densa multimodal. Las mejoras se concentran principalmente en el corto plazo ($t+5$) y en cohortes con dinámica glucémica regular, donde la autoatención permite modelar con mayor precisión las variaciones locales de la serie temporal. En horizontes medios y largos, la ganancia es marginal, lo que sugiere que la información contenida en las señales previas pierde capacidad predictiva con la distancia temporal. Cabe destacar que sería posible obtener resultados notablemente mejores a los recogidos en los experimentos *Transformer*, en el caso de contar con un hardware mucho más potente para poder utilizar muchas más épocas de entrenamiento.

A nivel global, la versión *Transformer* con ajuste de metapárametros logra el menor RMSE promedio (25,95 mg/dL), mientras que la versión *Transformer* sin ajuste de metapárametros obtiene el menor MAE (17,52 mg/dL). Estas diferencias, aunque pequeñas, se mantienen de forma consistente en todos los *folds*, lo que refuerza su validez estadística. Además, la menor desviación entre *folds* evidencia una mejor capacidad de generalización de los modelos *Transformer* frente a los cambios en la población de validación.

En términos de aplicabilidad, los resultados sugieren que, si el objetivo principal es la predicción inmediata de los niveles de glucosa (sistemas de alerta temprana), las arquitecturas *Transformer* constituyen la opción más adecuada por su mayor sensibilidad a las variaciones recientes. En cambio, si se busca un modelo más eficiente, la red densa ofrece un rendimiento competitivo con un coste computacional significativamente inferior.

En definitiva, las tres versiones alcanzan un desempeño adecuado, con errores medios entre 17 y 20 mg/dL, lo que demuestra la viabilidad del enfoque multimodal para la predicción de glucosa en pacientes con diabetes tipo 2. La incorporación de señales fisiológicas, de actividad e imagen refuerza la capacidad del modelo para capturar tanto la dimensión temporal como estructural del fenómeno, consolidando una base sólida para futuros desarrollos en

Gema de Pablo Laguna

predicción personalizada y monitorización inteligente.

6. CONCLUSIONES Y LÍNEAS FUTURAS

6.1 Conclusiones

El presente Trabajo de Fin de Máster ha abordado el desarrollo de un sistema de predicción multimodal de niveles de glucosa en sangre basado en redes neuronales profundas, integrando tres fuentes de información complementarias: series temporales de glucosa obtenidas mediante sensores Dexcom G6, datos de actividad física procedentes de dispositivos portátiles, e imágenes de retina en formato DICOM del conjunto de datos AI-READI.

A diferencia de los enfoques revisados en el estado del arte, que se centran principalmente en la modelización de series temporales univariantes o en la integración de señales fisiológicas de naturaleza homogénea, este estudio incorpora de forma novedosa una fusión multimodal entre datos secuenciales (glucosa y actividad física) y datos espaciales (imágenes retinianas), enriqueciendo la representación global del estado del paciente.

El estudio ha demostrado que la fusión de modalidades heterogéneas mediante aprendizaje profundo permite capturar tanto la dinámica temporal del metabolismo glucémico como la dimensión estructural de los tejidos oculares, enriqueciendo la capacidad predictiva del modelo frente a enfoques univariantes o exclusivamente temporales.

En primer lugar, los resultados obtenidos a lo largo de los experimentos reflejan que los tres modelos propuestos (una red densa multimodal y dos variantes *Transformer* multimodales) alcanzan un rendimiento robusto, con errores medios absolutos (MAE) comprendidos entre 17 y 20 mg/dL y RMSE entre 26 y 29 mg/dL, valores competitivos frente a la literatura reciente en predicción de glucosa. Estos resultados confirman la viabilidad del enfoque multimodal para la monitorización personalizada de pacientes con diabetes tipo 2.

En segundo lugar, la comparación entre arquitecturas evidencia que los modelos *Transformer* ofrecen una mayor estabilidad y generalización respecto al modelo denso, reduciendo la desviación inter-fold del MAE de 0,89 a aproximadamente 0,25. Este comportamiento se debe a los mecanismos de autoatención, que permiten identificar dependencias de corto y medio plazo en las señales temporales y asignar mayor peso a los instantes más relevantes para la predicción.

Las mejoras más notables se observan en el horizonte de 5 minutos, donde los *Transformer* logran una reducción del error en torno al 7 % en MAE y al 5 % en RMSE, lo que refuerza su potencial para aplicaciones de alerta temprana o predicción inmediata de eventos de hipoglucemia o hiperglucemia. Sin embargo, para horizontes de 30 y 60 minutos las diferencias se reducen, lo que sugiere que la información histórica pierde capacidad explicativa con el aumento del horizonte temporal y que la introducción de factores externos sería necesaria para mejorar la precisión a largo plazo.

Por otro lado, el análisis por cohortes clínicas muestra que el rendimiento disminuye progresivamente en función de la complejidad metabólica del grupo: los modelos obtienen mejores resultados en las cohortes de individuos sin diagnóstico de diabetes y prediabéticos, y un error más elevado en pacientes con diabetes tipo 2 tratada con insulina. Este patrón pone de

manifiesto que la heterogeneidad fisiológica y la variabilidad glucémica condicionan la capacidad de generalización del modelo, un aspecto crítico en la predicción personalizada de la enfermedad.

Desde una perspectiva computacional, se ha comprobado que las arquitecturas *Transformer*, pese a su mejor desempeño, requieren mayor coste de entrenamiento, tanto en tiempo como en recursos de GPU, frente a la red densa multimodal, que constituye una alternativa eficiente con una pérdida marginal de precisión. Este equilibrio entre rendimiento y eficiencia resulta esencial en escenarios reales de monitorización continua, donde la latencia y el consumo energético son factores limitantes.

Por tanto, este trabajo confirma que la combinación de datos multimodales y arquitecturas de atención ofrece un camino prometedor para avanzar hacia una medicina predictiva más precisa, personalizada y preventiva. La metodología propuesta constituye un punto de partida sólido para el desarrollo de sistemas clínicos inteligentes de apoyo a la toma de decisiones en pacientes con diabetes tipo 2.

6.2 Líneas futuras de trabajo

A partir de los resultados obtenidos, se identifican diversas líneas de mejora y ampliación que podrían reforzar el valor científico y práctico del modelo propuesto:

1. **Ampliación del conjunto de datos:** incorporar un número mayor de participantes y períodos de seguimiento más largos permitiría mejorar la capacidad de generalización, reducir el sesgo interindividual y entrenar modelos más profundos sin riesgo de sobreajuste.
2. **Integración de nuevas modalidades:** añadir información adicional (registro de sueño o alimentación) permitiría capturar factores contextuales que influyen directamente en la glucosa.
3. **Modelos multimodales avanzados:** explorar arquitecturas híbridas que combinen mecanismos de atención temporal (*Transformer*) con componentes convolucionales o recurrentes, o bien utilizar redes neuronales multimodales fusionadas por atención cruzada (*cross-attention*) que aprendan interacciones entre señales fisiológicas y visuales.
4. **Optimización de metapárametros y entrenamiento:** aumentar el número de épocas, aplicar estrategias de *learning rate scheduling*, *gradient clipping* o *mixed precision training* podría mejorar la convergencia y reducir el tiempo computacional, especialmente en entornos con GPU de mayor capacidad.
5. **Explicabilidad del modelo:** implementar técnicas de interpretabilidad (SHAP, Grad-CAM, *attention maps*) orientadas a identificar qué características temporales o regiones de la retina contribuyen en mayor medida a las predicciones, facilitando su validación clínica.
6. **Evaluación clínica y validación externa:** contrastar el modelo en entornos hospitalarios reales para valorar su aplicabilidad práctica, su robustez frente a ruido y su utilidad

como herramienta de apoyo a la toma de decisiones médicas.

7. **Implementación en tiempo real:** adaptar la arquitectura a un entorno de inferencia continua, integrando el modelo en dispositivos portátiles o aplicaciones móviles para la predicción y alerta temprana de eventos glucémicos críticos.

7. REFERENCIAS

- [1] Mompo, E. E., Espino, D. R., Morató, M. J., Flor, M. V., & Torres, I. V. (s/f). *Diabetes Tipo 2*. Clínic Barcelona. Disponible en: <https://www.clinicbarcelona.org/asistencia/enfermedades/diabetes/diabetes-tipo-2> (Recuperado el 25 de junio de 2025).
- [2] IDF Diabetes Atlas 2025. (2025, febrero 27). *Diabetes Atlas*. Disponible en: <https://diabetesatlas.org/resources/idf-diabetes-atlas-2025/> (Recuperado el 25 de junio de 2025).
- [3] Mediavilla Bravo, J. J. (2002). La diabetes mellitus tipo 2. *Medicina Integral*, 39(1), 25–35. Disponible en: <https://www.elsevier.es/es-revista-medicina-integral-63-articulo-la-diabetes-mellitus-tipo-2-13025480> (Recuperado el 25 de junio de 2025).
- [4] National Institute of Diabetes and Digestive and Kidney Diseases. (s/f). *Diabetes tipo 2*. Disponible en: <https://www.niddk.nih.gov/health-information/informacion-de-la-salud/diabetes/informacion-general/que-es/diabetes-tipo-2> (Recuperado el 25 de junio de 2025).
- [5] Secretaría. (2023, febrero 20). Aplicación de la tecnología avanzada en diabetes: Pros y Contras. *Revista Diabetes*. Disponible en: <https://www.revistadiabetes.org/tecnologia/aplicacion-de-la-tecnologia-avanzada-en-diabetes-pros-y-contras/> (Recuperado el 25 de junio de 2025).
- [6] *Diabetes de tipo 2*. (s/f). MayoClinic.org. Disponible en: <https://www.mayoclinic.org/es/diseases-conditions/type-2-diabetes/diagnosis-treatment/drc-20351199> (Recuperado el 25 de junio de 2025).
- [7] National Institute of Diabetes and Digestive and Kidney Diseases. (s/f). *La prueba de A1C y la diabetes*. Disponible en: <https://www.niddk.nih.gov/health-information/informacion-de-la-salud/pruebas-diagnosticas/prueba-a1c-diabetes> (Recuperado el 25 de junio de 2025).
- [8] Devices & technology. (s/f). Diabetes.org. Disponible en: <https://diabetes.org/about-diabetes/devices-technology> (Recuperado el 25 de junio de 2025).
- [9] CSEDOCS-Servei d'Endocrinologia i Nutrició de l'Hospital Universitari Quirón Dexeus. (s/f). La importancia del ejercicio físico en la diabetes tipo 2. *Endocrino.cat*. Disponible en: <https://www.endocrino.cat/es/blog-endocrinologia.cfm/ID/8290/ESP/la-importancia-ejercicio-fisico-diabetes-tipo-2.htm> (Recuperado el 25 de junio de 2025).
- [10] Wood, J., & Peters, A. P. (2018). *The Type 1 Diabetes Self-Care Manual: A Complete Guide to Type 1 Diabetes Across the Lifespan for People with Diabetes, Parents, and Caregivers*. American Diabetes Association.

-
- [11] Horizonte, C. (s/f). *Cronicidad Horizonte 2025*. Cronicidadhorizonte2025.com. Disponible en: <https://www.cronicidadhorizonte2025.com/ningunpacienteenpausa-patologias-diabetestipo2> (Recuperado el 25 de junio de 2025).
 - [12] SANITAS. (s/f). *Complicaciones de la diabetes mellitus o de tipo 2*. Sanitas. Disponible en: <https://www.sanitas.es/biblioteca-de-salud/enfermedades-y-trastornos/diabetes/diabetes-mellitus-complicaciones> (Recuperado el 25 de junio de 2025).
 - [13] Garrido Calvo, A. M., Cía Blasco, P., & Pinós Laborda, P. J. (2003). El pie diabético. *Medicina Integral*, 41(1), 8–17. Disponible en: <https://www.elsevier.es/es-revista-medicina-integral-63-articulo-el-pie-diabetico-13044043> (Recuperado el 25 de junio de 2025).
 - [14] Heintzman, N. D. (2015). A digital ecosystem of diabetes data and technology: Services, systems, and tools enabled by wearables, sensors, and apps. *Journal of Diabetes Science and Technology*, 10(1), 35–41. Disponible en: <https://doi.org/10.1177/1932296815622453> (Recuperado el 25 de junio de 2025).
 - [15] Klonoff, D. C. (2005). Continuous glucose monitoring roadmap for 21st century diabetes therapy. *Diabetes Care*, 28(5), 1231–1239. Disponible en: <https://doi.org/10.2337/diacare.28.5.1231> (Recuperado el 25 de junio de 2025).
 - [16] Google Cloud. (s/f). *¿Qué es el aprendizaje automático? Tipos y usos*. Disponible en: <https://cloud.google.com/learn/what-is-machine-learning?hl=es-419> (Recuperado el 25 de junio de 2025).
 - [17] Algoritmos de aprendizaje automático revolucionan la predicción de enfermedades. (2024, marzo 11). *Consenso Salud*. Disponible en: <https://www.consensosalud.com.ar/algoritmos-de-aprendizaje-automatico-revolucionan-la-prediccion-de-enfermedades/> (Recuperado el 25 de junio de 2025).
 - [18] Secretaría. (2023, diciembre 11). Inteligencia artificial y diabetes. *Revista Diabetes*. Disponible en: <https://www.revistadiabetes.org/tecnologia/inteligencia-artificial-y-diabetes/> (Recuperado el 25 de junio de 2025).
 - [19] Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, 6, Article 26094. Disponible en: <https://doi.org/10.1038/srep26094> (Recuperado el 25 de junio de 2025).
 - [20] Farahmand, E., Rahimi Azghan, R., Taheri Chatrudi, N., et al. (2025). AttenGluco: Multimodal Transformer-based blood glucose forecasting on the AI-READI dataset. *arXiv preprint arXiv:2502.09919*. Disponible en: <https://arxiv.org/abs/2502.09919> (Recuperado el 7 de julio de 2025).

- [21] Karagöz, M. A., Breton, M. D., & El Fathi, A. (2025). A comparative study of Transformer-based models for multi-horizon blood glucose prediction. *arXiv preprint arXiv:2505.08821*. Disponible en: <https://arxiv.org/abs/2505.08821> (Recuperado el 7 de julio de 2025).
- [22] Bian, Q. X., Guo, M. Y., Huang, T. Y., & Liu, J. C. (2024). A hybrid Transformer-LSTM model applied to glucose prediction. *PLOS ONE*, 19(9), e0310084. Disponible en: <https://doi.org/10.1371/journal.pone.0310084> (Recuperado el 7 de julio de 2025).
- [23] Barbato, M. P., Bragagnolo, J., & D'Angelo, R. (2025). Lightweight sequential Transformers for blood glucose level prediction in type 1 diabetes. *arXiv preprint arXiv:2506.07864*. Disponible en: <https://arxiv.org/abs/2506.07864> (Recuperado el 7 de julio de 2025).
- [24] Haleem, M. S., Al-Sabbagh, A., & Kim, H. (2025). A multimodal deep learning architecture for predicting interstitial glucose based on life-log data. *Scientific Reports*, 15, Article 10721. Disponible en: <https://doi.org/10.1038/s41598-025-07272-3> (Recuperado el 7 de julio de 2025).
- [25] Bridge to Artificial Intelligence (Bridge2AI). (s/f). *Nih.gov*. Disponible en: <https://commonfund.nih.gov/bridge2ai> (Recuperado el 24 de julio de 2025)
- [26] AI-READI Consortium. (2024). *Flagship dataset of type 2 diabetes from the AI-READI project* [Data set]. FAIRhub.
- [27] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. Disponible en: <https://doi.org/10.1038/sdata.2016.18> (Recuperado el 24 de julio de 2025)
- [28] U.S. Department of Health and Human Services. (1996). *Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. Public Law 104-191. Disponible en: <https://www.hhs.gov/hipaa> (Recuperado el 24 de julio de 2025)
- [29] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*. In *Advances in Neural Information Processing Systems* (pp. 5998–6008).
- [30] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. In *Proceedings of NAACL-HLT*, 4171–4186.

8. ANEXOS

8.1 Aspectos éticos, económicos, sociales y ambientales

8.1.1. Introducción

Este Trabajo de Fin de Máster se ha desarrollado dentro del ámbito de la inteligencia artificial aplicada a la salud, explorando cómo los modelos multimodales de aprendizaje profundo pueden contribuir a la predicción y prevención de alteraciones glucémicas en pacientes con diabetes tipo 2 (DMT2). El estudio se apoya en el conjunto de datos *AI-READY*, que combina señales de sensores de monitorización continua de glucosa (*Dexcom G6*), registros de actividad física (*Garmin Vivosmart*) e imágenes de retina en formato DICOM. La finalidad del trabajo es avanzar hacia una medicina predictiva y personalizada, ofreciendo modelos que apoyen la toma de decisiones clínicas en tiempo real. Durante todo el desarrollo del proyecto se ha garantizado el cumplimiento de los principios éticos de investigación, el uso responsable de los datos y la protección de la privacidad de los participantes.

8.1.2. Impacto social

El principal impacto social del proyecto radica en su potencial para mejorar la gestión de la diabetes mediante el uso de tecnologías predictivas. Un sistema capaz de anticipar las variaciones de glucosa podría ayudar a reducir el número de episodios de hipoglucemia e hiperglucemia, disminuir complicaciones asociadas (como la retinopatía diabética) y fomentar una mayor autonomía del paciente. Además, la combinación de datos de sensores portátiles e inteligencia artificial impulsa un modelo de seguimiento remoto y continuo, que facilita la comunicación entre pacientes y profesionales de la salud. A largo plazo, esta línea de investigación podría contribuir a una sanidad más preventiva, eficiente y centrada en la persona.

8.1.3. Impacto económico

Desde el punto de vista económico, los avances en predicción de glucosa mediante inteligencia artificial pueden optimizar los costes sanitarios al evitar ingresos hospitalarios y consultas innecesarias. El uso de herramientas abiertas y gratuitas, como Python, TensorFlow, Keras y PyTorch, reduce considerablemente el coste tecnológico del proyecto y fomenta la reutilización del conocimiento generado. Asimismo, el desarrollo de modelos predictivos aplicables a dispositivos *wearables* crea nuevas oportunidades para el sector tecnológico y biomédico, impulsando la innovación y la transferencia de conocimiento entre universidades y empresas.

8.1.4. Impacto medioambiental

El impacto medioambiental asociado al desarrollo del proyecto es muy limitado, ya que todo el trabajo se ha realizado de forma digital, sin generación de residuos físicos ni necesidad de

desplazamientos o infraestructuras específicas. El entrenamiento de los modelos se llevó a cabo en entornos virtualizados y optimizados, priorizando la eficiencia de cómputo y evitando el uso innecesario de recursos de GPU. Además, se fomentó la reutilización de modelos preentrenados, práctica que disminuye significativamente la huella de carbono derivada del entrenamiento intensivo en inteligencia artificial.

8.1.5. Impacto ético

El uso de datos biomédicos exige un tratamiento riguroso y ético. Los datos empleados provienen de un dataset público y anonimizado, en el que se garantiza la eliminación de información identificable y el cumplimiento del Reglamento General de Protección de Datos (RGPD). En ningún caso se ha tenido acceso a información personal de los participantes, ni se ha realizado reidentificación o uso con fines distintos de los académicos. El modelo se ha diseñado con un enfoque transparente y responsable, evitando sesgos de género o cohorte y priorizando la equidad algorítmica en la predicción. Este enfoque se alinea con los principios de una IA confiable, tal y como promueven la Comisión Europea y la comunidad científica internacional.

8.1.6. Conclusiones

El proyecto contribuye positivamente en los ámbitos social, económico y ético, y presenta un impacto ambiental prácticamente nulo. Los resultados obtenidos demuestran el potencial de la inteligencia artificial para **mejorar la calidad de vida de los pacientes con diabetes**, reduciendo complicaciones y promoviendo una sanidad más eficiente. Por tanto, este trabajo representa un avance hacia una **inteligencia artificial ética, sostenible y orientada al bienestar de las personas**.

8.2 Presupuesto económico

Tabla 7. Presupuesto económico

Concepto	Descripción	Cantidad / Tiempo	Coste unitario (€)	Coste total (€)
Mano de obra (coste directo)	Desarrollo de notebooks (v6, v8, v9), análisis del dataset Al-READI, redacción de la memoria y visualizaciones	430 h	17,00 €/h	7.310,00
Recursos computacionales (coste directo)	Entrenamiento de modelos en GPU (Google Colab Pro / entorno local con GPU RTX)	60 h	0,75 €/h	45,00
Recursos materiales (coste directo)	Ordenador portátil con GPU (8 meses / amortización 4 años)	—	—	175,00
	Disco externo y almacenamiento en nube (SharePoint, Google Drive, FAIRhub)	—	—	40,00
Software y herramientas	Python, TensorFlow, Keras, Scikit-learn, Pydicom, OpenCV (software libre, sin coste)	—	—	0,00
Gastos generales (costes indirectos)	Electricidad, conexión a internet, mantenimiento del equipo y gastos administrativos (10 % sobre costes directos)	—	—	757,00
Supervisión y tutoría académica	Reuniones de seguimiento, revisión técnica y corrección del manuscrito	20 h	25,00 €/h	500,00
SUBTOTAL				8.827,00
IVA (21 %)				1.853,67
TOTAL PRESUPUESTO ESTIMADO				10.680,67

8.3 Documentación técnica complementaria

8.3.1. Descripción general del entorno de trabajo

El desarrollo del proyecto se realizó en un entorno de análisis avanzado basado en Python (versión 3.10), utilizando *Jupyter Notebook* como interfaz principal de ejecución y documentación de los experimentos. El entorno incluyó las siguientes librerías principales:

- **TensorFlow y Keras**: desarrollo y entrenamiento de redes neuronales.
- **Scikit-learn**: cálculo de métricas, validación cruzada y preprocesamiento tabular.
- **Pandas y NumPy**: manipulación de datos.
- **Matplotlib y Seaborn**: visualización de resultados y métricas.
- **Pydicom y OpenCV**: tratamiento de imágenes médicas DICOM.
- **os, gc y random**: gestión del sistema de archivos y control de memoria.

El entrenamiento de los modelos se ejecutó en un entorno local de *Jupyter Notebook*, desplegado mediante *Anaconda Navigator*, utilizando una configuración basada en *CPU*. El equipo de desarrollo disponía de un procesador *Intel Core i5* (4 núcleos, 2.4 GHz), 16 GB de memoria RAM y sistema operativo *Windows 10 Pro* (64 bits). Estas especificaciones condicionaron el número de épocas de entrenamiento y el tamaño de los lotes empleados debido al elevado coste computacional de las arquitecturas de aprendizaje profundo.

8.3.2. Estructura de notebooks y versiones desarrolladas

Durante el desarrollo del TFM se elaboraron tres versiones principales de notebooks, que reflejan la evolución del modelo multimodal:

- Versión Red Densa Multimodal (baseline).
- Versión Modelo Transformer Multimodal.
- Versión Transformer Multimodal con Optimización de Metapárametros.

Cada una de estas versiones fue enlazada al repositorio de datos compartido en SharePoint (Anexo 8.3.4).

8.3.3. Conjunto de datos empleado

El proyecto utilizó el *dataset AI-READI* (*Artificial Intelligence for Reliable and Equitable Advances in Diabetes Intelligence*), accesible a través de la plataforma FAIRhub. El conjunto de datos fue enlazado al repositorio de datos compartido en SharePoint (Anexo 8.3.4).

8.3.4. SharePoint

El conjunto completo está disponible en el espacio compartido de la Universidad Europea de Madrid (SharePoint): https://liveuem-my.sharepoint.com/:f/g/personal/22468215-live_uem_es/EjcY6foT0EBCi7EH87GhreIBvkNMx3y1XIsX5gGBJovHVA?e=i0yLnV