



Universidad Internacional de La Rioja
Escuela Superior de Ingeniería y
Tecnología

Máster Universitario en Inteligencia Artificial

Rendimiento de modelos de aprendizaje profundo en escenarios con pacientes diabéticos reales y simulados

Trabajo fin de estudio presentado por:	José Miguel Pereira Pontón
Tipo de trabajo:	Comparativa de soluciones
Director/a:	Ciro Rodriguez León
Fecha:	20/07/2023

Resumen

En este estudio, se investigó el desempeño de seis modelos de aprendizaje profundo para la predicción de glucosa en pacientes con Diabetes Mellitus Tipo 1, aplicado tanto en casos clínicos como en casos simulados. Se evaluaron diferentes configuraciones de capas RNN, LSTM y GRU, así como combinaciones de estas capas. Se analizó el impacto de emplear diferente tamaño de ventana histórica y diferente horizonte de predicción. Los resultados revelaron que las ventanas históricas de 60 y 90 minutos tuvieron un mejor rendimiento en la predicción, y el horizonte de predicción de 15 minutos fue el más eficiente. La combinación de la capa LSTM y GRU mostró el mejor resultado en términos de precisión, según el análisis del RMSE y del análisis de error de Clarke. Además, se observó que los modelos entrenados con datos reales pudieron alcanzar una ventaja significativa a los modelos entrenados con datos simulados.

Palabras clave: Diabetes Mellitus Tipo 1, Glucosa, Aprendizaje profundo, Predicción, Redes Neuronales Recurrentes.

Abstract

In this study, the performance of six deep learning models for glucose prediction in patients with Type 1 Diabetes Mellitus was investigated, applied to both clinical and simulated cases. Different configurations of RNN, LSTM, and GRU layers, as well as combinations of these layers, were evaluated. The impact of using different window sizes for the historical data and different prediction horizons was analyzed. The results revealed that historical windows of 60 and 90 minutes performed better in the prediction, with a prediction horizon of 15 minutes being the most efficient. The combination of LSTM and GRU layers showed the best result in terms of accuracy, as assessed by RMSE and Clarke error analysis. Furthermore, it was observed that models trained with real data had a significant advantage over models trained with simulated data.

Keywords: Type 1 Diabetes Mellitus, Glucose, Deep Learning, Prediction, Recurrent Neural Network.

Índice de contenidos

Introducción.....	1
1.1. Motivación	3
1.2. Planteamiento del problema	3
1.3. Estructura de la memoria	4
2. Contexto y estado del arte	5
2.1. Diabetes Mellitus	5
2.2. Tipos de diabetes	5
2.2.1. Diabetes Mellitus Tipo 1	5
2.2.2. Diabetes Mellitus Tipo 2	7
2.2.3. Otros tipos de diabetes	7
2.3. Análisis de cuadrícula de errores de Clarke	7
2.4. Inteligencia artificial en la medicina	9
2.5. Aprendizaje profundo	10
2.6. Clasificación de las técnicas de Aprendizaje profundo.....	11
2.7. Arquitecturas de redes de aprendizaje profundo	12
2.7.1. Arquitectura RNN	12
2.7.2. Arquitectura LSTM.....	12
2.7.3. Arquitectura GRU	13
2.7.4. Arquitectura CNN	13
2.8. Simuladores de glucosa	14
2.8.1. UVA/PODOVA Type 1 Diabetes Simulator:.....	15
2.8.2. Bergman Minimal Model.....	15
2.8.3. Cambridge’s Model / Simulator	16

2.9.	Estudios previos	16
3.	Objetivos y metodología de trabajo.....	22
3.1.	Objetivo general.....	22
3.2.	Objetivos específicos	22
3.3.	Metodología del trabajo	22
4.	Planteamiento de la comparativa	24
4.1.	Adquisición de datos.....	24
4.1.1.	Conjunto de información de Naturaleza Real	24
4.1.2.	Conjunto de información de Naturaleza Artificial.....	25
4.2.	Procesamiento de los datos.....	28
4.2.1.	Descripción de selección de datos de pacientes reales	30
4.2.2.	Datos de pacientes simulados	35
4.2.3.	Transformación de los datos	37
5.	Desarrollo de la comparativa	41
5.1.	Visualización de series temporales.....	41
5.1.1.	Pacientes reales.....	41
5.1.2.	Pacientes simulados	41
5.2.	Modelamiento	44
5.2.1.	Modelo 1: Modelo de 1 capa RNN	45
5.2.2.	Modelo 2: Modelo de 1 capa LSTM.....	53
5.2.3.	Modelo 3: Modelo de 2 capas LSTM	58
5.2.4.	Modelo 4: Modelo de 1 capa GRU	66
5.2.5.	Modelo 5: Modelo de 2 capas GRU.....	71
5.2.6.	Modelo 6: Modelo de 1 capa LSTM + 1 capa GRU	78

6. Discusión y análisis de resultados	86
7. Conclusiones y trabajo futuro	89
7.1. Conclusiones	89
7.2. Líneas de trabajo futuro.....	90
Referencias bibliográficas.....	91
Anexo A. Artículo de investigación	95

Índice de figuras

Figura 1: Factores predictivos asociados al riesgo de la DMT1.	6
Figura 2: Gráfica de análisis de cuadrícula de errores de Clarke	8
Figura 3: Aplicaciones de aprendizaje profundo	11
Figura 4: Técnicas de aprendizaje profundo	11
Figura 5: Estructura de la unidad RNN.	12
Figura 6: Estructura de la unidad LSTM.....	13
Figura 7: Estructura de la unidad GRU.	13
Figura 8: Estructura de la red CNN.	14
Figura 9: Publicaciones indexadas en PubMed hasta mayo del 2023 relacionadas con Diabetes Mellitus Tipo 1.	16
Figura 10: Publicaciones indexadas en PubMed relacionadas con la predicción de la Diabetes Mellitus Tipo 1.	17
Figura 11: Metodología CRISP-DM aplicada en el estudio.	23
Figura 12: Condiciones de los pacientes simulados.	28
Figura 13: Selección de conjunto de datos de pacientes reales.	30
Figura 14: Código en Python empleado para la Imputación de datos mediante el promedio del dato anterior y posterior al dato faltante.	31
Figura 15: Código en Python empleado para la generación de los pacientes mediante el simulador.....	35
Figura 16: Tratamiento de datos de pacientes simulados.	36
Figura 17: Modelo de predicción desarrollado, implementado a data real y data sintética. ...	38
Figura 18: Ejemplo de ventana deslizante para predicción de glucosa a 15 minutos con un histórico de 30 minutos.....	38

Figura 19: Código en Python empleado para generar las matrices con diferente tamaño de ventana.....	39
Figura 20: Comportamiento de la ventana deslizante para la generación de los conjuntos de datos.	40
Figura 21: Glucosa vs Tiempo del paciente LIB193410 medida en 1 día. a) Valor muestreado. b) Valor normalizado	42
Figura 22: Glucosa vs Tiempo del paciente PaciecientelD1 simulado medida en 1 día. a) Valor muestreado. b) Valor normalizado.....	43
Figura 23: Código en Python empleado para generar el Modelo 1	46
Figura 24: Mejores respuestas del Modelo 1 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.	48
Figura 25: Gráficas de Clarke para el modelo 1 – Pacientes Reales.....	49
Figura 26: Mejores respuestas del Modelo 1 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.	51
Figura 27: Gráficas de Clarke para el modelo 1 – Pacientes simulados	52
Figura 28: Código en Python empleado para generar el Modelo 2.	53
Figura 29: Mejores respuestas del Modelo 2 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.	54
Figura 30: Gráficas de Clarke para el modelo 2 – Pacientes Reales.....	56
Figura 31: Mejores respuestas del Modelo 2 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.	57
Figura 32: Gráficas de Clarke para el modelo 2 – Pacientes Simulados.....	59
Figura 33: Código en Python empleado para generar el Modelo 3	60
Figura 34: Mejores respuestas del Modelo 3 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.	61
Figura 35: Gráficas de Clarke para el modelo 3 – Pacientes reales.....	62

Figura 36: Mejores respuestas del Modelo 3 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.	64
Figura 37: Gráficas de Clarke para el modelo 3 – Pacientes simulados	65
Figura 38: Código en Python empleado para generar el Modelo 4.	66
Figura 39: Mejores respuestas del Modelo 4 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.	67
Figura 40: Gráficas de Clarke para el modelo 4 – Pacientes reales.....	69
Figura 41: Mejores respuestas del Modelo 4 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.	70
Figura 42: Gráficas de Clarke para el modelo 4 – Pacientes simulados	72
Figura 43: Código en Python empleado para generar el Modelo 5.	73
Figura 44: Mejores respuestas del Modelo 5 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.	74
Figura 45: Gráficas de Clarke para el modelo 5 – Pacientes reales.....	75
Figura 46: Mejores respuestas del Modelo 5 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.	77
Figura 47: Gráficas de Clarke para el modelo 5 – Pacientes Simulados.....	78
Figura 48: Código en Python empleado para generar el Modelo 6.	79
Figura 49: Mejores respuestas del Modelo 6 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.	80
Figura 50: Gráficas de Clarke para el modelo 6 – Pacientes reales.....	82
Figura 51: Mejores respuestas del Modelo 6 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.	83
Figura 52: Gráficas de Clarke para el modelo 6 – Paciente simulados.....	85

Índice de tablas

Tabla 1: Descripción de artículos de revisión de Inteligencia Artificial y DMT1.	18
Tabla 2: Estudios sobre la predicción de glucosa y las arquitecturas de aprendizaje profundo.	20
Tabla 3: Características de los datos	25
Tabla 4: Rangos de Mediciones de la Glucosa en pacientes reales.	25
Tabla 5: Condiciones de los sensores simulados.	26
Tabla 6: Condiciones de las bombas de insulina simuladas.	26
Tabla 7: Métrica de los pacientes que pasaron el filtro de selección.	33
Tabla 8: Información del grupo de pacientes empleados en el conjunto de entrenamiento .	34
Tabla 9: Información del grupo de pacientes empleados en el conjunto de validación	34
Tabla 10: Información del grupo de pacientes empleados en el conjunto de prueba	34
Tabla 11 Información del grupo de pacientes simulados empleados en el conjunto de entrenamiento.....	36
Tabla 12: Información del grupo de pacientes simulados empleados en el conjunto de validación.....	37
Tabla 13: Información del grupo de pacientes simulados empleados en el conjunto de prueba.	37
Tabla 14: Modelos de aprendizaje profundo desarrollados	44
Tabla 15: Métricas del modelo 1 para un horizonte de predicción de 15 minutos en pacientes reales	46
Tabla 16: Métricas del modelo 1 para un horizonte de predicción de 30 minutos en pacientes reales	47
Tabla 17: Métricas del modelo 1 para un horizonte de predicción de 15 minutos en pacientes simulados.....	50

Tabla 18: Métricas del modelo 1 para un horizonte de predicción de 30 minutos en pacientes simulados.....	50
Tabla 19: Métricas del modelo 2 para un horizonte de predicción de 15 minutos en pacientes reales	53
Tabla 20: Métricas del modelo 2 para un horizonte de predicción de 30 minutos en pacientes reales	54
Tabla 21: Métricas del modelo 2 para un horizonte de predicción de 15 minutos en pacientes simulados.....	55
Tabla 22: Métricas del modelo 2 para un horizonte de predicción de 30 minutos en pacientes simulados.....	57
Tabla 23: Métricas del modelo 3 para un horizonte de predicción de 15 minutos en pacientes reales.	60
Tabla 24: Métricas del modelo 3 para un horizonte de predicción de 30 minutos en pacientes reales	60
Tabla 25: Métricas del modelo 3 para un horizonte de predicción de 15 minutos en pacientes simulados.....	63
Tabla 26: Métricas del modelo 3 para un horizonte de predicción de 30 minutos en pacientes simulados.....	63
Tabla 27: Métricas del modelo 4 para un horizonte de predicción de 15 minutos en pacientes reales	66
Tabla 28: Métricas del modelo 4 para un horizonte de predicción de 30 minutos en pacientes reales	67
Tabla 29: Métricas del modelo 4 para un horizonte de predicción de 15 minutos en pacientes simulados.....	68
Tabla 30: Métricas del modelo 4 para un horizonte de predicción de 30 minutos en pacientes simulados.....	70

Tabla 31: Métricas del modelo 5 para un horizonte de predicción de 15 minutos en pacientes reales	73
Tabla 32: Métricas del modelo 5 para un horizonte de predicción de 30 minutos en pacientes reales	73
Tabla 33: Métricas del modelo 5 para un horizonte de predicción de 15 minutos en pacientes simulados.....	76
Tabla 34: Métricas del modelo 5 para un horizonte de predicción de 30 minutos en pacientes simulados.....	76
Tabla 35: Métricas del modelo 6 para un horizonte de predicción de 15 minutos en pacientes reales	79
Tabla 36: Métricas del modelo 6 para un horizonte de predicción de 30 minutos en pacientes reales	80
Tabla 37: Métricas del modelo 6 para un horizonte de predicción de 15 minutos en pacientes simulados.....	81
Tabla 38: Métricas del modelo 6 para un horizonte de predicción de 30 minutos en pacientes simulados.....	83
Tabla 39: Resumen de métrica RMSE y zona clínicamente aceptada de Clarke (Zona A + Zona B) para un horizonte de predicción de 15 minutos en pacientes reales.....	87
Tabla 40: Resumen de métrica RMSE y zona clínicamente aceptada de Clarke (Zona A + Zona B) para un horizonte de predicción de 30 minutos en pacientes reales.....	87
Tabla 41: Resumen de métrica RMSE y zona clínicamente aceptada de Clarke (Zona A + Zona B) para un horizonte de predicción de 15 minutos en pacientes simulados.	88
Tabla 42: Resumen de métrica RMSE y zona clínicamente aceptada de Clarke (Zona A + Zona B) para un horizonte de predicción de 30 minutos en pacientes simulados.	88

Introducción

Se considera a la Diabetes Mellitus o simplemente denominada diabetes como un grupo de enfermedades metabólicas que se encuentran relacionadas por defectos en la secreción de insulina, la acción de la insulina o en su defecto por ambas condicionales (ASSOCIATION, 2010).

La hormona insulina es producida en la célula del islote de Langerhans en el páncreas y se la considera como una hormona esencial. La insulina es de gran importancia para el metabolismo de las grasas y proteínas. Cuando se tiene escasas de producción de insulina, se tiende a tener altos niveles de glucosa (hiperglucemia).

La Federación Internacional de Diabetes, reportó en su décima edición del Atlas de Diabetes de 2021 que se tiene una prevalencia en el incremento global de personas con diabetes confirmada, un total de 537 millones de adultos se encontraban viviendo con diabetes (International Diabetes Federation, 2021). Estiman que para el año 2045 se tenga un total de 784 millones de personas con diabetes a nivel global. Se encontró que se tiene una mayor prevalencia en adultos entre 20 y 79 años.

Se consideran diferentes tipos de diabetes, entre los que se destacan, Diabetes Mellitus Tipo 1 (DMT1) y Diabetes Mellitus Tipo 2 (DMT2). La DMT1 se encuentra causada principalmente por un proceso autoinmune, en el cual, el propio sistema inmune del cuerpo genera un ataque ante las *células beta* que son producidas en el páncreas, lo que produce un déficit en la cantidad de insulina generada. Esta condición puede aparecer en cualquier edad, afectando a niños, jóvenes y adultos. Las personas que padecen de DMT1 necesitan de una o varias dosis de insulina diaria para regular la glucosa en la sangre. Mientras que en el caso de DMT2, es una de los tipos de diabetes más común, se produce una hiperglucemia, esto se encuentra en función a que el cuerpo no es capaz de asimilar correctamente la insulina (Petersmann et al., 2019). Dentro del presente estudio nos enfocaremos en las características y estimación de la glucosa en pacientes con DMT1.

La Inteligencia Artificial (IA) se encuentra innovando y revolucionando diferentes campos de investigación, entre los que encontramos: finanzas, industria, domótica, conducción de automóviles, mercadotecnia, distribución de recursos, reconocimiento facial y medicina, solo por mencionar algunos. La inclusión de la Inteligencia Artificial en el campo de la medicina ha mejorado diferentes aspectos, entre los que encontramos: prevención de enfermedades, diagnósticos de forma anticipada, interpretación de imágenes para el reconocimiento de patrones complejos y tratamientos debido a que la IA se la puede aplicar para predecir reacciones adversas ante diferentes tratamientos y como seguimiento, soporte y monitorización (Ávila-Tomás et al., 2021).

En el caso de la DMT1, la predicción de los niveles de glucosa en sangre es de vital importancia para el cuidado de los pacientes, sin embargo, la parte más complicada es la obtención de la información de los pacientes, debido a que puede ser un proceso costoso y en parte difícil de obtener. Entre los beneficios de la inteligencia artificial en el área de la medicina se encuentra: aliviando a los pacientes, reemplazando datos, separando por categorías de forma más eficiente y aumentando la cantidad de información (Aung et al., 2021). El aumentar la información permite mejorar la calidad, reducir los errores médicos, reducción de los sesgos, así como mejorar el conocimiento médico y una respuesta de intervención más rápida, con mejores pronósticos en los resultados.

Existen diferentes softwares que simulan el comportamiento del cuerpo humano en relación con el metabolismo de la glucosa, estos datos generados artificialmente pueden ser empleados para el entrenamiento de modelos. La principal ventaja del uso de datos sintéticos para la predicción de glucosa es la capacidad de generar grandes cantidades de datos precisos y variados (Pérez, 2014).

Para realizar las predicciones sobre diferentes índoles en el campo de la inteligencia artificial se cuenta con diferentes técnicas y algoritmos. Al momento que se requiere obtener un mejor resultado ante comportamientos complejos, como es el caso de la predicción de la glucosa en la sangre, se ha invertido con un mayor énfasis en modelos de aprendizaje profundo. Se tienen una gama amplia de algoritmos, en este estudio se enfoca en la aplicación de algoritmos RNN, GRU, LSTM y una combinación de capas LSTM y GRU.

1.1. Motivación

El estudio de la DMT1 es de gran importancia debido al número de personas que la padece y la proyección que se tiene acerca del número de personas que pueden llegar a padecerla. Unas de las limitaciones que existe en cuanto a la realización de la predicción sobre la glucosa en la sangre se encuentra relacionado con la falta de un conjunto de datos reales, esto en función de diferentes factores, entre los más importantes encontramos problemas de privacidad o confidencialidad, normas éticas y legales, falta de información almacenada y disponible, el costo que implica el monitoreo y almacenamiento en tiempo real, condiciones y escenarios de las personas que padecen de esta enfermedad.

Lo que se plantea en el presente estudio es realizar una comparativa en el rendimiento de modelos de aprendizaje profundo para predecir la glucosa entre dos grupos de estudio, un grupo de estudio se encuentra relacionado con un conjunto de datos de pacientes reales y un segundo grupo relacionado con un conjunto de datos generados artificialmente.

Al observar el efecto y los resultados que brinden los modelos de IA podría solventar las diferentes limitaciones y desventajas que acarrear la recolección, tratamiento y utilización de información de pacientes con DMT1, según se vea el impacto en los resultados de los pacientes creados In-Sílico. Con lo cual, el emplear estos datos en modelos de entrenamiento en el campo de la investigación permitirá reducir la brecha existente en cuanto a datos, tiempos y resultados.

1.2. Planteamiento del problema

En el presente estudio se comparará el desempeño de arquitecturas de aprendizaje profundo, entrenándola con datos de usuarios con DMT1 y con un conjunto de datos generado artificialmente.

Los modelos creados contemplarán las mismas condiciones de parámetros e hiperparámetros, condiciones de entradas y salidas, las únicas variaciones existentes serán en función a la naturaleza de los datos con los que se entrenarán. Las arquitecturas de aprendizaje profundo

se las entrenarán bajo las dos naturalezas de datos, siendo estos modelos aislados, permitiendo así comparar las predicciones de la glucosa.

1.3. Estructura de la memoria

En el presente estudio dentro del contexto y el estado de arte se considerarán las diferentes temáticas:

- Diabetes como enfermedad
- Tipo de diabetes
- Diabetes Tipo 2
- Inteligencia Artificial en la medicina
- Arquitecturas de aprendizaje profundo
- Generadores de datos sintéticos
- Comparativa de soluciones realizadas previamente

Estos puntos centrales del trabajo darán un mayor realce al estudio final realizado. En el tercer capítulo se describirán los objetivos del estudio, así como la metodología planteada para el desarrollo de las arquitecturas. En el capítulo cuatro del estudio se proyectarán las condiciones que se tomarán para poder generar la comparativa planteada y esta comparativa se la expone en el capítulo capítulo cinco. Mientras que en el capítulo seis, se exponen los resultados del estudio con mayor profundidad y se culmina en el capítulo siete con las conclusiones que se obtuvieron tras la realización del proyecto.

2. Contexto y estado del arte

2.1. Diabetes Mellitus

Diabetes Mellitus se trata de una alteración metabólica, de origen muy diverso, pero conlleva la existencia de alteraciones en la secreción de la insulina, la acción de la hormona o en algunos casos en ambas condiciones. La detección de la enfermedad cuando los síntomas son persistentes y cifras de la glucemia elevada son más rápidas de diagnosticar, mientras que se debe reconocer que existen casos en donde las personas que padecen de la enfermedad son asintomáticas (Conget, 2002).

La hiperglucemia crónica en los casos de pacientes con diabetes se encuentra relacionada con daños a largo plazo, en donde los órganos que más se ven afectados se encuentran: ojos, riñones, corazón, nervios y los vasos sanguíneos (ASSOCIATION, 2010).

2.2. Tipos de diabetes

Los casos de diabetes que se desarrollan, en general, en la población se encuentran divididos en dos categorías etiopatogénicas: Diabetes Mellitus Tipo 1 y Diabetes Mellitus Tipo 2, sin embargo, se conocen diferentes tipos de diabetes, las mismas que se abordarán más adelante.

2.2.1. Diabetes Mellitus Tipo 1

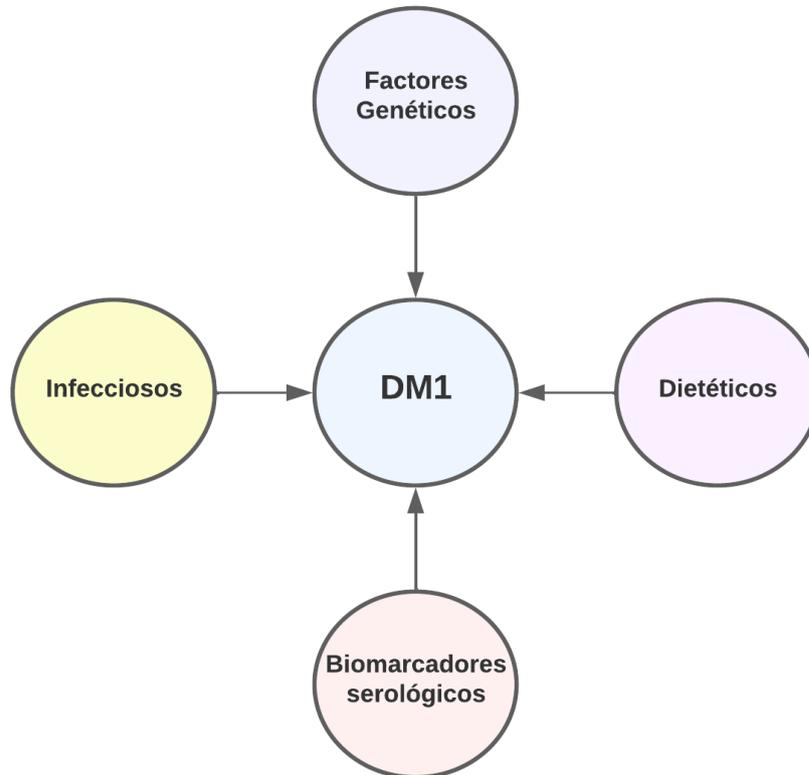
La Diabetes Mellitus Tipo 1 es la enfermedad que abarca cerca del 10% de los casos de diabetes que se registran a nivel mundial, esta enfermedad puede aparecer a cualquier edad, pero principalmente niños y adultos los que la padecen. Existe un riesgo elevado de mortalidad cuando la DMT1 se complica según se expone en (Gillespie, 2006).

Los autores en (Atkinson & Eisenbarth, 2001) indican que la DMT1 se encuentra aumentando rápidamente en regiones específicas y es variable en función de la población de las diferentes etnias. El diagnóstico de la DMT1 se realiza en función de la hiperglucemia.

En el estudio realizado por (Primavera et al., 2020), se hace mención sobre los predictores de riesgo implícitos en DMT1. De los diferentes factores que se han propuesto, los más importantes son los factores genéticos, infecciosos, dietéticos y los factores serológicos. El

enfoque del riesgo no se lo debería de considerar a estos factores de forma aislada, sino en conjunto. Estos factores se los logra apreciar en la Figura 1.

Figura 1: Factores predictivos asociados al riesgo de la DM1.



Nota. Adaptado de *Main predictive factors associated to the risk of T1D*, de Primavera et al, 2000, article. Open Access.

Las estrategias de predicción que se contempla en la actualidad son un factor clave para evitar el desarrollo del proceso autoinmune que se puede generar en las personas con riesgo.

En función del control de DM1, a partir del estudio realizado por (Beck et al., 2019), se aprecia la relevancia de los avances tecnológicos para el manejo de la enfermedad. Se realiza tanto en los medidores de glucosa en la sangre, bombas de insulina, monitores de glucosa continua y de los sistemas que combinan tanto la bomba como el monitor que se encuentran centralizados para realizar la liberación de la insulina. Los sistemas mencionados se encuentran gobernados por algoritmos que permiten la automatización del proceso. Tal como lo

menciona el autor, los sensores se han vuelto cada vez más pequeños, precisos y permiten una gran habilidad para visualizar la información de la glucosa.

2.2.2. Diabetes Mellitus Tipo 2

La Diabetes Mellitus Tipo 2 (DMT2) se trata del tipo de diabetes más común, la misma que afecta un 90% de los casos reportados en el mundo. En la DMT2, las células del cuerpo no responden en su totalidad a la insulina, se tiene una consideración de que se vuelve resistente a la insulina, en donde el resultado que se obtiene es la hiperglucemia. Los síntomas suelen ser similares a las que se tienen en la DMT1, con una diferencia, estos síntomas son menos dramáticos y en ocasiones son asintomáticos. Las causas de la DMT2 no son completamente entendidas, se las suele ligar con el sobrepeso y la obesidad, edad, etnia y en registros de la familia (International Diabetes Federation, 2021).

2.2.3. Otros tipos de diabetes

Existen algunos tipos de diabetes diferentes al tipo 1 y tipo 2, las mismas que se encuentran relacionadas con:

- Defectos Genéticos de las células beta.
- Defectos Genéticos en la acción de insulina.
- Enfermedades exocrinas del páncreas.
- Endocrinopatías.
- Diabetes inducidas por fármacos o debido a productos químicos.
- Por infecciones.
- Forma infrecuente de diabetes autoinmune.
- Por otros síndromes genéticos que se encuentran asociados con la diabetes.
- Diabetes mellitus gestacional (Conget, 2002).

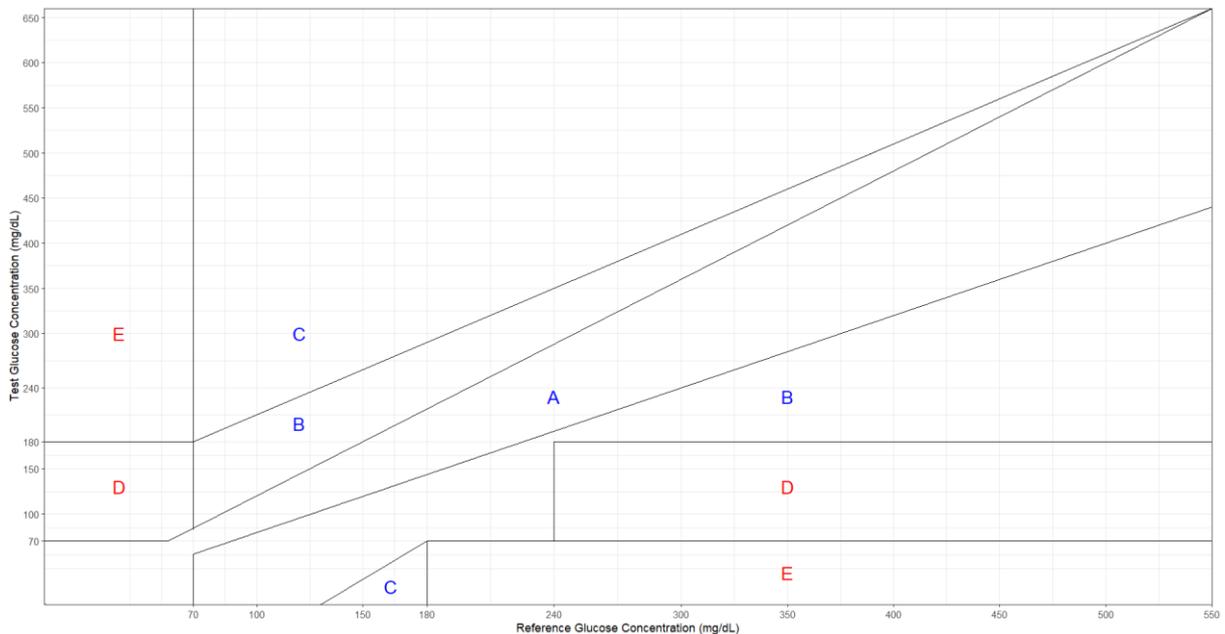
2.3. ANÁLISIS DE CUADRÍCULA DE ERRORES DE CLARKE

El análisis de cuadrícula de errores de Clarke se trata de un acercamiento con enfoque clínico, desarrollado para comparar los dispositivos de monitoreo a partir del valor de referencia y la precisión de sus valores, sin embargo, se lo emplea para realizar la evaluación del resultado que brinda un modelo de predicción (Mayo et al., 2019).

Las regiones de Clarke se tratan de divisiones en un gráfico bidimensional de dispersión, lo que permite determinar la concordancia que existe entre las mediciones reales y las predicciones cuando se encuentra realizando la monitorización de la glucosa en la sangre.

En el presente enfoque se compara la precisión de los diferentes modelos según la región en la que cae, las mismas regiones que puede observar en la Figura 2.

Figura 2: Gráfica de análisis de cuadrícula de errores de Clarke



Nota. Realizado mediante la librería EGA de R.

El significado de la cada una de las regiones las podemos describir de la siguiente forma según el autor (Kopecký et al., 2013).

- Zona A: Se dan medidas clínicamente precisas
- Zona B: Se dan errores benignos, pero no guían hacia un tratamiento inapropiado.
- Zona C: Se generan correcciones innecesarias, es decir, una corrección excedente de niveles aptos de glucosa.
- Zona D: Se generan errores potencialmente peligrosos. Esto se genera debido a una falla en la detección de hipoglucemia o hiperglucemia.
- Zona E: Se generan los errores en el tratamiento.

Las zonas A y B son clínicamente aceptables, mientras que las zonas C, D y E son clínicamente inaceptables.

2.4. INTELIGENCIA ARTIFICIAL EN LA MEDICINA

La inteligencia artificial (Artificial Intelligence – AI) está compuesta por diferentes campos, entre los que se menciona Aprendizaje de Máquinas (Machine Learning – ML), Aprendizaje profundo (Deep Learning – DL) y visión por computadora (Computer Vision). El concepto o uso de los algoritmos de ML se encuentra enfocado en la identificación de patrones para su posterior análisis. Existen problemas que para su correcta identificación e interpretación de patrones se requiere de algoritmos más avanzados y complejos, para lo cual se hace uso del aprendizaje profundo. Por último, en los algoritmos de aprendizaje enfocados en visión por computadora, se le brinda un soporte en la toma de decisiones a partir del reconocimiento de patrones en imágenes y videos. (Kaul et al., 2020).

La expansión e implementación de la Inteligencia Artificial en la Medicina (Artificial Intelligence in Medicine – AIM) de forma directa y personalizada en los diferentes problemas han generado mejores oportunidades a los pacientes, debido a que los modelos generados han permitido mejorar la precisión de los diagnósticos y la eficiencia de los procedimientos médicos.

Uno de los campos que mayor impacto ha generado es el aprendizaje profundo, porque a diferencia de los algoritmos de ML, estos no requieren de manipulación del personal o ingresos de información.

Según los autores (Holzinger et al., 2019) indican que una de las desventajas que existe en la AIM es la explicabilidad. Para poder sobrellevar este impedimento se requiere de tener causabilidad. La aplicación de modelos de AI/ML/DL en el campo de la medicina se trata de uno de los desafíos más grandes que existen.

Los autores de (Chowdhury & Chakraborty, 2017) indicaron que la utilización de la AIM es variada, entre los que se pueden destacar: desarrollo de fármacos, tratamientos y diagnóstico de enfermedades, análisis y monitoreo de salud, consultas digitales, tratamientos médico general y personalizado, entre los más comunes.

2.5. APRENDIZAJE PROFUNDO

El aprendizaje profundo es uno de los subconjuntos del aprendizaje de máquinas, en el cual, los patrones de reconocimiento de información se encuentran relacionados e inspirados en el funcionamiento del cerebro humano. Este tipo de algoritmos se encuentran exentos de un sistema de reglas definidos por los humanos, en contraste hacen uso de información para generar las etiquetas correspondiente (Alzubaidi et al., 2021).

El aprendizaje profundo se encuentra relacionado con el empleo de diferentes números de capas dentro de los algoritmos para generar conocimiento.

Existen dos importantes preguntas que absolver: ¿Cuándo aplicar aprendizaje profundo? y ¿Por qué aplicar aprendizaje profundo?

Las situaciones en las que es apremiante la aplicación de un modelo de aprendizaje profundo son:

- a. Casos en los que los humanos no poseen conocimiento
- b. Casos donde el humano a partir de su expertiz no es capaz de tomar decisiones.
- c. Casos en los que los problemas, así como sus soluciones se encuentran progresando en el tiempo.
- d. Casos en los que se requiere de una personalización en las soluciones.
- e. Casos en los que la información sobre el problema es muy amplia e implica un impedimento en el razonamiento de todas las implicaciones.

Para responder la segunda interrogante, se puede hacer ligeras acotaciones, tal como las menciona (Alzubaidi et al., 2021) en su trabajo:

- a. Se tiene una aproximación de forma universal al aprendizaje
- b. Se tiene robustez en los modelos.
- c. La generalización de las técnicas en cuanto los diferentes tipos de datos.
- d. Es posible generar escalabilidad de los modelos.

Una vez mencionados tanto los casos en los cuales se aplican y las razones de porque aplicar un modelo de aprendizaje profundo, en la Figura 3 se muestran algunas aplicaciones usuales.

Figura 3: Aplicaciones de aprendizaje profundo

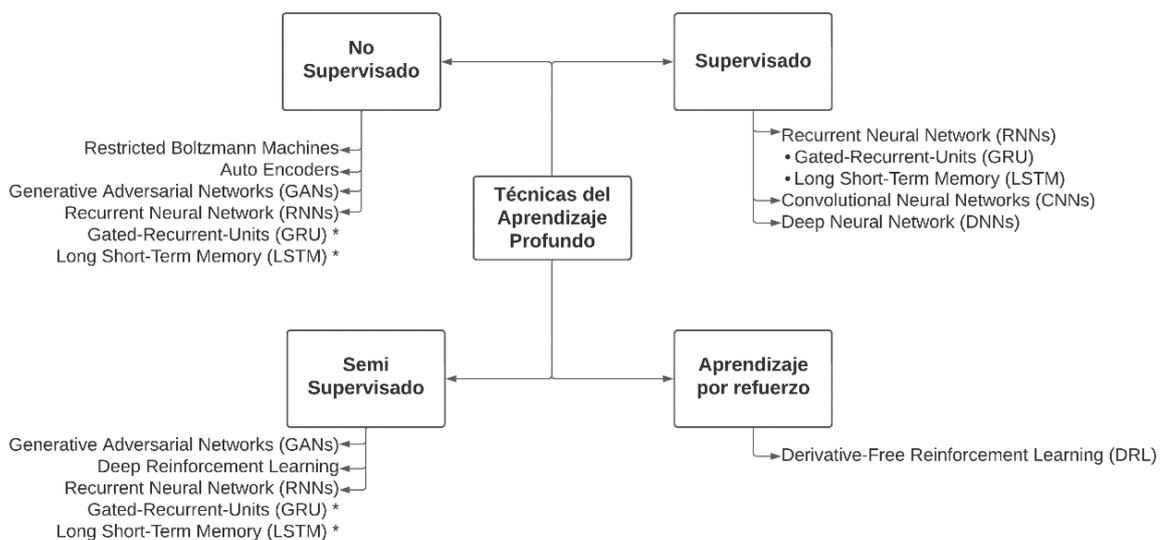


Nota. Adaptado de *Examples of DL applications*, de Alzubaidi et al., Journal. Open Access.

2.6. CLASIFICACIÓN DE LAS TÉCNICAS DE APRENDIZAJE PROFUNDO

Las técnicas del aprendizaje profundo al igual que las técnicas de aprendizaje de máquinas, se clasifican en cuatro grandes grupos, los que son supervisado, no supervisado, semi supervisado y por refuerzo. Para tener una mejor perspectiva se puede referir a la Figura 4.

Figura 4: Técnicas de aprendizaje profundo



2.7. Arquitecturas de redes de aprendizaje profundo

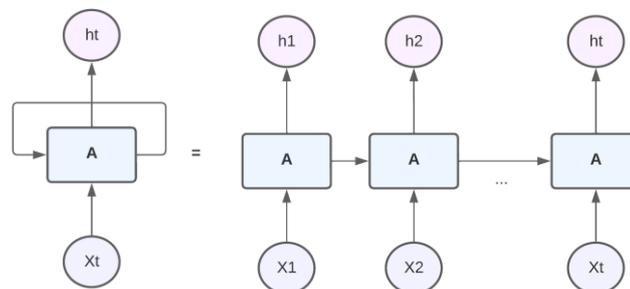
Existen diferentes familias de configuraciones arquitectónicas relacionadas con las redes neuronales, entre las que mencionamos:

- Redes Neuronales Recursivas (RNN),
- RNN
- LSTM
- GRU
- CNN

2.7.1. Arquitectura RNN

En este caso, la arquitectura de la red emplea la información de una de las iteraciones para mejorar la predicción que tiene la salida de la siguiente neurona. Estas redes contemplan un problema, es que a medida que se genera un paso, en cada iteración se almacena la misma información, lo que implica que la relevancia de uno de los pasos previos se vaya perdiendo, este concepto se lo conoce desvanecimiento del gradiente. En la Figura 5 se observa la arquitectura de la red neuronal recurrente.

Figura 5: Estructura de la unidad RNN.



Fuente. Pranoy Radharikrishnan. 2017. Recurrent Neural Network. Medium.

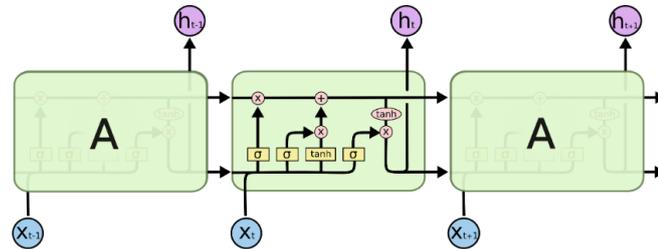
(<https://towardsdatascience.com/introduction-to-recurrent-neural-network-27202c3945f3>)

2.7.2. Arquitectura LSTM

Una red LSTM (Long Short-Term Memory) hace uso de neuronas para determinar qué información es la que se almacena y cuál es la información que se olvida, esta consideración

es en función del estado actual. Las redes LSTM poseen tres puertas de control, puerta de entrada, de salida y de olvido, tal como se aprecia en la Figura 6.

Figura 6: Estructura de la unidad LSTM

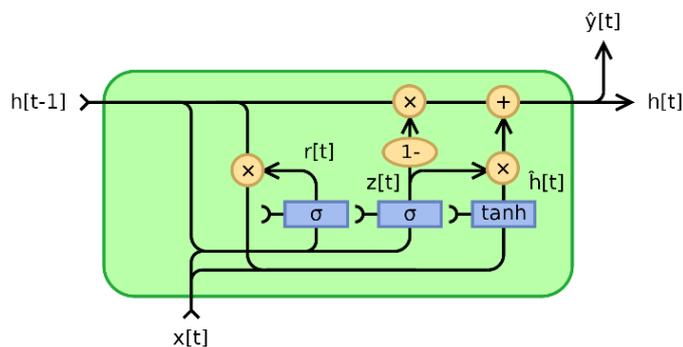


Fuente: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

2.7.3. Arquitectura GRU

Las redes GRU (Gated-Recurrent-Units) se generaron para solventar el problema de la gradiente, su primera aparición realizado por (Cho et al., 2014). Como lo indica el autor (Chung et al., 2014) esta arquitectura permite que cada una de las unidades recurrentes capture adaptativamente las dependencias a diferentes escalas de tiempo. Al igual que las LSTM posee puertas que regulan el flujo de la información, sin embargo, no posee una celda de memoria.

Figura 7: Estructura de la unidad GRU.



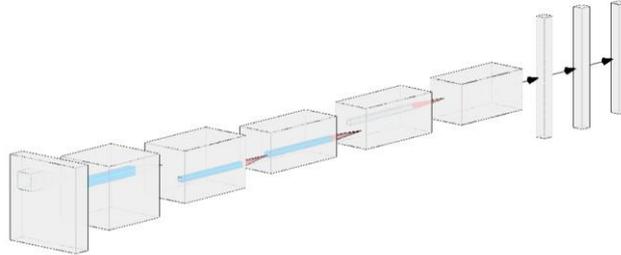
Fuente: <https://atcold.github.io/pytorch-Deep-Learning/es/week06/06-2/>

2.7.4. Arquitectura CNN

Se trata de un tipo especial de redes neuronales artificiales, en donde las operaciones de convolución se tratan de uno de los bloques fundamentales. Se compone especialmente de

unas capas de convolución, seguido de una capa de agrupación, se genera al final capas completamente conectadas (Bezdan & Bačanin Džakula, 2019). Podemos ver un ejemplo de arquitectura CNN en la Figura 8.

Figura 8: Estructura de la red CNN.



Nota. Fue elaborado mediante el uso de <https://alexlenail.me/NN-SVG/AlexNet.html>

2.8. Simuladores de glucosa

Los simuladores del cuerpo humano se tratan de modelos matemáticos que tratan de imitar comportamientos específicos. El tener modelos para la simulación de la glucosa es altamente deseable para efectos de diseñar y evaluar sensores, generar un sistema de apoyo a la toma de decisiones, tal como es mencionado por (Dalla Man et al., 2006). Este enfoque radica en que aplicar un experimento o prueba a un usuario puede ser demasiado difícil, peligroso o puede caer en la parte no ética.

Los simuladores de glucosa son modelos matemáticos que imitan el comportamiento de la glucosa en el cuerpo humano. Estos modelos brindan un gran soporte a los médicos e investigadores en el enfoque de la experimentación, debido a que permiten simular diferentes escenarios, estos escenarios se encuentran relacionados con un entorno seguro y controlado, en el cual, no se llega a poner en riesgo la salud de los pacientes.

Una de las ventajas de los simuladores de glucosa es que se permiten realizar pruebas simuladas para sustituir a los pacientes reales (Rosales, 2011). Esto permite que se puedan explorar diferentes soluciones sin que se ponga en riesgo a los pacientes, debido a que no implica realizar ensayos clínicos, y a su vez se deriva en una reducción de costos al no aplicarlos.

En cuestión de los simuladores de DMT1 existentes, tenemos diferentes opciones, cada uno tiene sus ventajas y desventajas, así como la accesibilidad al mismo, unos son de código abierto. Entre estos simuladores se mencionan:

- UVA/PADOVA Type 1 Diabetes Simulator
- Bergman Minimal Model
- Cambridge DMT1
- McGill DMT1 Simulator

2.8.1. UVA/PODOVA Type 1 Diabetes Simulator:

El simulador UVA/PODOVA, se trata de un modelo de larga escala en el cual se encuentra descrito el comportamiento dinámico de la glucosa, insulina y el glucagón mediante ecuaciones diferenciales. Este simulador fue aprobado en el 2008 por la Administración de Alimentos y Medicamentos - FDA (Vettoretti et al., 2018).

Entre las principales aplicaciones del simulador tenemos dos:

- a. Simular y testear protocolos de estudios clínicos.
- b. Medir el impacto en el manejo y tratamiento de la diabetes (*DMT1MS – The Epsilon Group*, 2017).

El modelo de UVA/PADOVA emula el comportamiento de una población de 30 sujetos simulados, estos grupos se encuentran divididos equitativamente 10 niños, 10 adolescentes y 10 adultos para la versión del 2008 (Dalla Man et al., 2014). Siendo el modelo de UVA/PADOVA de código abierto, en sus versiones para MATLAB y Python.

2.8.2. Bergman Minimal Model

El modelo mínimo de Bergman se basa en la prueba de tolerancia de glucosa intravenosa. Se lo emplea en el campo de la investigación clínica para poder evaluar la resistencia a la insulina y la producción de glucosa. Existe una desventaja, que se encuentra, en que no se toma en consideración la variabilidad interindividual. A diferencia del modelo UVA/PADOVA, este modelo no posee un software ni es de código abierto (Dalla Man et al., 2014).

2.8.3. Cambridge's Model / Simulator

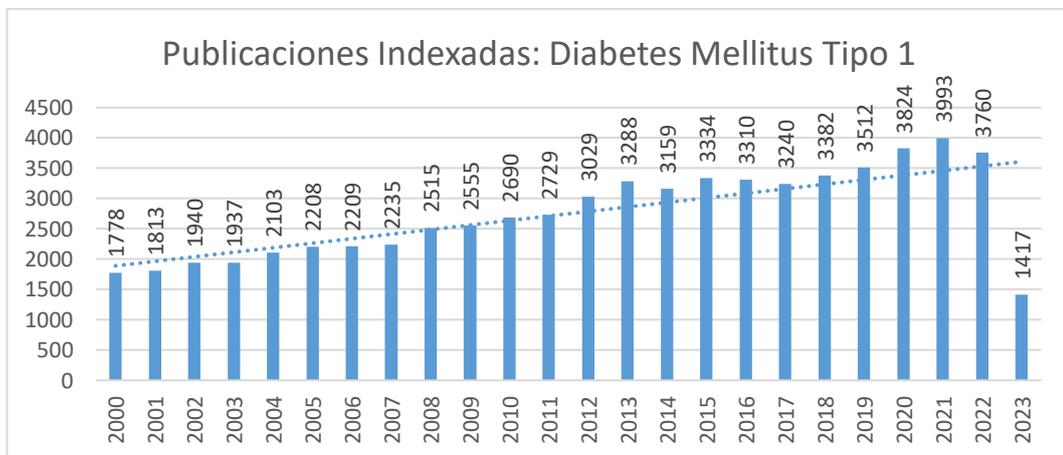
Se basa en el efecto de un páncreas artificial, el mismo que se enfoca en el efecto de la insulina en tres acciones, la distribución, eliminación y producción. El páncreas artificial se generó a partir de la presencia de los dispositivos de monitoreo continuo y los sistemas que permiten la administración de insulina, estos componentes permiten el establecimiento de un circuito cerrado. Si queremos explorar el concepto de un páncreas artificial, este consta de tres elementos, el dispositivo de monitoreo, el algoritmo de cálculo de insulina y una bomba de insulina.

El modelo generado en Cambridge ha permitido realizar diferentes ensayos clínicos, aportando resultados de buen rendimiento, sin embargo, se requiere de tener una mayor prueba clínica antes de realizar una generalización (Colmegna et al., 2015). El modelo de Cambridge al igual que modelo de Bergman no es de código abierto.

2.9. Estudios previos

En los últimos años, el interés en el estudio de la DMT1 y el desarrollo de modelos de predicción de glucosa en pacientes con DMT1 mediante el uso de redes de aprendizaje profundo ha ido en aumento. En la Figura 9, se logra apreciar el número de publicaciones que se encuentran indexadas en PubMed.

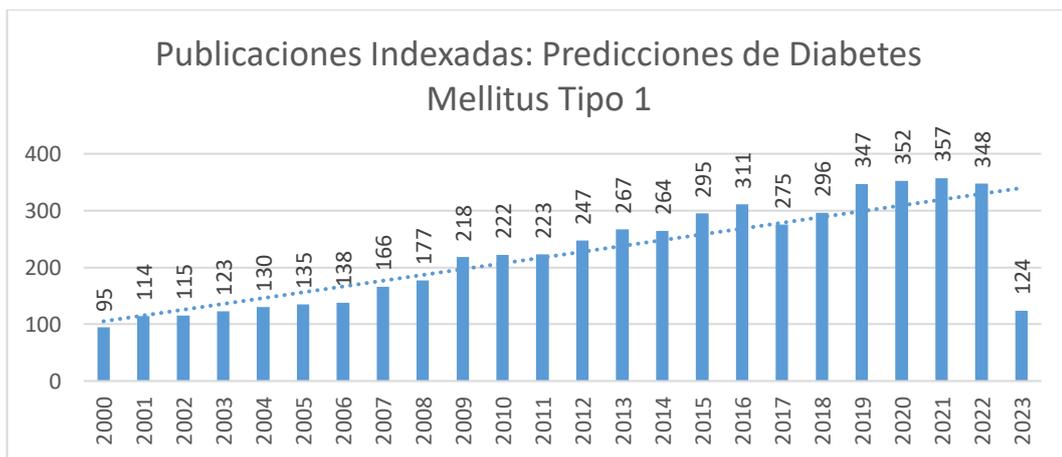
Figura 9: Publicaciones indexadas en PubMed hasta mayo del 2023 relacionadas con Diabetes Mellitus Tipo 1.



Fuente PubMed.

Hasta mayo del 2023 se encuentran 1417 publicaciones relacionadas con esta temática. Mientras que, si revisamos el número de estudios que se encuentran relacionados con la predicción de Glucosa y la DMT1, el número de publicaciones es menor, pero sigue en aumento, esto lo podemos observar en la Figura 10. Estos estudios han brindado soporte para el manejo efectivo de esta enfermedad crónica, generando una visión más precisa e incluso personalizada de los niveles de glucosa. Estudios que tienen el objetivo de mejorar la calidad de vida de los pacientes.

Figura 10: Publicaciones indexadas en PubMed relacionadas con la predicción de la Diabetes Mellitus Tipo 1.



Fuente PubMed.

Los métodos tradicionales del monitoreo de glucosa han sido una pieza clave en el control de la enfermedad, sin embargo, estos modelos tradicionales presentan limitaciones tanto a nivel de precisión por parte de los sensores, así como la capacidad de predicción que se puede tener extrapolando a largo plazo. Mediante el estudio de las diferentes técnicas de aprendizaje profundo, así como su progreso, ha permitido avanzar a pasos agigantados la predicción de la glucosa en los pacientes con DMT1. Las redes de aprendizaje profundo han permitido capturar patrones complejos de los datos de glucosa.

Se presenta una breve descripción de los diferentes estudios que se encuentran relacionados con predicción de glucosa en pacientes con DMT1. En la Tabla 1, se describen los modelos y criterios que se abarcaron dentro de diferentes artículos de revisión.

Tabla 1: Descripción de artículos de revisión de Inteligencia Artificial y DMT1.

Estudio	Estudios Incluidos	Modelos Abarcados	Criterios de evaluación	Autores
Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes	55	<ol style="list-style-type: none"> 1. Feed-forward neural networks (20%) 2. Modelo híbrido entre un modelo psicológico y modelo de ML (19%). 3. Redes neuronales recurrentes (18%). 4. Máquina de soporte vectorial (11%). 5. Técnicas de programación genética (6%). 6. Redes neuronales auto regresivas y redes neuro-fuzzy (5%). 7. Mapas auto organizados (4%) 8. Otros 	<ol style="list-style-type: none"> 1. Métricas de Rendimiento <ol style="list-style-type: none"> 1.1. Criterios de evaluación matemática: RMSE, coeficiente de correlación, FIT, NPE. 1.2. Criterios de evaluación clínica: Análisis de malla de error, tiempo de ganancia promedio, índice J. 2. Horizonte de predicción <p>Entre 15 minutos hasta 2 horas.</p> <p>Alcanzo una mayor predicción entre 15-30 minutos</p> 3. Estudios consideraban tanto información in-silico como real 	(Woldaregay et al., 2019)
Deep Learning for Diabetes: A Systematic Review	40	<ol style="list-style-type: none"> 1. Autoencoder eliminadores de ruido 2. LSTM 3. RBM y RNN 4. CNN y capas completamente conectadas 5. DMLP 	<ol style="list-style-type: none"> 1. Métricas de rendimiento: <ol style="list-style-type: none"> 1.1: Sensitividad, especificidad, AUC, RMSE. 2. Horizonte de predicción: <ol style="list-style-type: none"> 2.1: predicción corto plazo: 30 min 2.2: predicción largo plazo: 60 minutos 2.3: RNN es una herramienta poderosa. 2.4: RNN aumentado por células LSTM es un método ampliamente empleado para la predicción de glucosa, tanto para 30 – 60 minutos. 3: Estudios empleaban bases in silicio, públicas y privadas. 	(Zhu et al., 2021)
Artificial Intelligence for Diabetes Management and Decision Support: Literature Review	141	<ol style="list-style-type: none"> 1. Red neuronal artificial 2. Random Forest 3. Máquina de soporte vectorial 4. Algoritmo de Regresión 5. Algoritmos Evolucionario 6. Árbol de decision 7. Naïve Bayes 8. K- vecinos más cercanos 	<ol style="list-style-type: none"> 1. Métricas de rendimiento: <ol style="list-style-type: none"> 1.1. Sensitividad y especificidad fueron las más empleadas. 2. Horizonte de predicción: <ol style="list-style-type: none"> 2.1. 15,30,45,60,120,150,180. 3. Estudios emplearon data generada in-silico, data real pública y privada. 	(Contreras & Vehi, 2018)

El primer enfoque que se estudió fueron revisiones literarias, debido a que estos estudios abordan diferentes artículos científicos que permiten la asimilación de la temática bajo el concepto brindado por diversos autores. En los diferentes artículos revisados se observó cómo la inteligencia artificial se ha utilizado en el campo de la diabetes para mejorar la gestión y manejo de la enfermedad y proporcionar apoyo en la toma de decisiones clínicas. En donde la inteligencia artificial tiene el potencial de ayudar a los profesionales de la salud y a los pacientes a tomar decisiones más informadas y personalizadas.

Los diferentes autores abordaron puntos comunes, entre los que se considera a la predicción y modelado de la dinámica de la glucosa en sangre como un objetivo clave para la mejora sistemática de la salud y condiciones de los pacientes. Así como el reconocimiento de la importancia de la personalización y adaptación de los modelos a las características individuales de los pacientes.

Se observaron diferentes arquitecturas de redes neuronales, sin embargo, las redes que por mayoría se emplean son las redes recurrentes, en este caso las LSTM. Los conjuntos de datos que se emplearon fueron tanto de pacientes reales, así como de pacientes simulados. Los horizontes de predicción que más se observó su uso es la predicción a 15 minutos, 30 minutos y 60 minutos. La ventana de históricos es variable según los estudios.

Una vez realizada la investigación sobre documentos de revisión sistemática, se propone la revisión individual de diferentes acercamientos de estudio de la predicción de la glucosa sobre pacientes con diabetes, para lo cual se describe en la Tabla 2, algunos artículos científicos que se encuentran relacionados con la predicción de la glucosa.

Tabla 2: Estudios sobre la predicción de glucosa y las arquitecturas de aprendizaje profundo.

Estudio	Orientación	Modelo	Origen de la fuente	Resultados			Autores
A Deep Learning Approach for Blood Glucose Prediction Of Type 1 Diabetes	Predecir los niveles de glucosa para evitar las situaciones de hiperglucemia o hipoglucemia, en pacientes con DMT1	CRNN	Dataset de Universidad de Ohio	HP: 30 minutos			(Freiburghaus et al., 2020)
				RMSE : 17.45	MAE: 11.22		
				HP: 60 minutos			
				RMSE: 33.67	MAE: 23.25		
Developing an Individual Glucose Prediction Model Using Recurrent Neural Network	Predicción de Glucosa personalizada en pacientes hospitalizados con DMT2.	RNN, GRU, LSTM	20 pacientes hospitalizados	RNN	GRU	LSTM	(Kim et al., 2020)
				RMSE	23.44	22.26	
Improving blood glucose level predictability using machine learning	Predicción de la glucosa con una ventana de 30 minutos.	Kernel Ridge Regression (KRR)	11 pacientes con DMT1	KRR con etiqueta especial	KRR con etiqueta Regular		(Marcus et al., 2020)
				RMSE	22.218	23.513	
Long-Term Glucose Forecasting Using a Physiological Model and Deconvolution of the Continuous Glucose Monitoring Signal	Predicción en glucosa de pacientes con DMT1 con diferentes ventanas de predicción.	latent variable-based statistical (LVX), autoregressive exogenous (ARX)	10 pacientes con DMT1	HP: 30 minutos			(Liu et al., 2019)
					LVX	ARX	
				RMSE	12.85	12.84	
				HP: 60 minutos			
					LVX	ARX	
				RMSE	28.00	26.38	
				HP: 90 minutos			
	LVX	ARX					
RMSE	41.04	36.10					
HP: 120 minutos							
	LVX	ARX					
RMSE	48.50	43.03					

Entre los artículos revisados se tiene diferentes acercamientos a la problemática estudiada, de los cuales podemos mencionar regresiones como los modelos ARX y KRR, modelos basados en variables estadísticas LVX, modelos de redes neuronales recursivas como redes RNN, redes LSTM y redes GRU y estudios en los que se aplicaron redes convolucionales CNN. En los estudios destacaron la aplicación de esta técnica de modelamiento para la predicción de glucosa basado en modelos específicos, en el cual los modelos fueron entrenados para ajustarse de mejor forma al comportamiento de cada uno de los pacientes. Para entrenar los diferentes modelos se emplearon diferentes ventanas de historia, en estos estudios, así como

en los estudios previamente analizados, las ventanas de históricos que tienen una mayor aceptación son las de 30 minutos y 45 minutos. La métrica que mayor uso se le da para la revisión del error con los diferentes escenarios es el RMSE. A partir de esta métrica descrita, se logró apreciar que a medida que el horizonte de predicción aumenta, el error también lo hace, generando diferencias considerables al momento de realizar la comparativa. Entre los diferentes horizontes de predicción los de menor error son los de 15 minutos y 30 minutos.

3. Objetivos y metodología de trabajo

3.1. Objetivo general

Comparar el desempeño en la predicción de glucosa a partir de una arquitectura de Aprendizaje Profundo entrenada con datos reales de pacientes con Diabetes Tipo 1 contra datos generados artificialmente mediante un simulador in-silico.

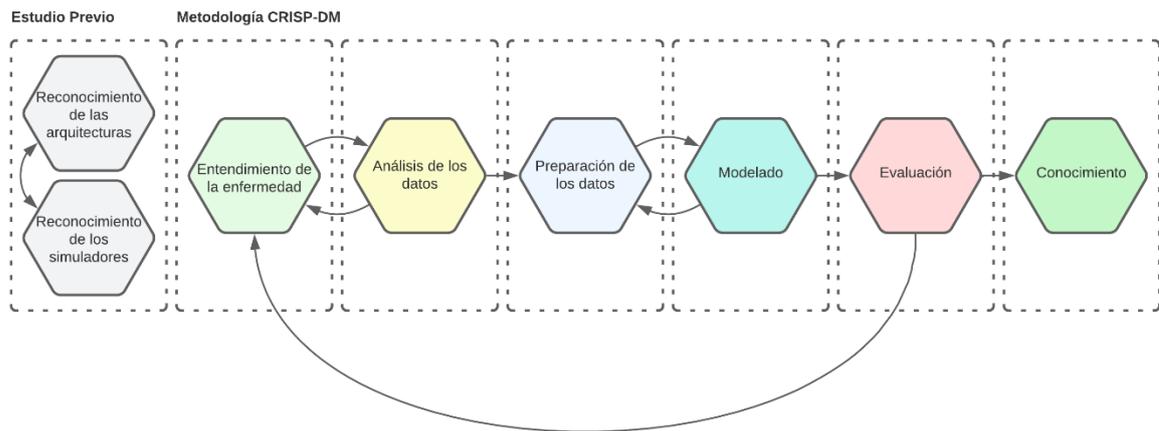
3.2. Objetivos específicos

Para poder cumplir con el objetivo general del trabajo de investigación, se establecen los siguientes objetivos específicos:

1. Investigar los datos relacionados con la enfermedad de Diabetes Mellitus y las diferentes variaciones.
2. Describir el estado del arte de las arquitecturas de aprendizaje profundo orientado hacia la predicción de glucosa.
3. Identificar los diferentes simuladores de glucosa para la generación del conjunto de datos simulados.
4. Experimentar arquitecturas de aprendizaje profundo para el escenario de datos reales y para el escenario de datos artificiales.
5. Evaluar la predicción de los dos escenarios a partir de la arquitectura seleccionada.
6. Comparar los resultados obtenidos por los diferentes modelos, interpretando los valores obtenidos para la obtención de conclusiones.

3.3. Metodología del trabajo

Para la obtención de los resultados del trabajo, se empleó la metodología CRISP-DM, tal como se la muestra en la Figura 11.

Figura 11: Metodología CRISP-DM aplicada en el estudio.

Para cumplir con los objetivos propuestos, se empleó la metodología descrita previamente en la Figura 11, que se desglosa en los siguientes pasos:

1. En fases preliminares, se llevó a cabo un análisis comparativo de las diferentes arquitecturas de aprendizaje profundo y de los simuladores de glucosa disponibles.
2. Se aplicó la metodología CRISP-DM, comenzando por comprender la Diabetes Mellitus Tipo 1 como enfermedad. Esto debido a que es fundamental tener un entendimiento claro de la enfermedad para abordar adecuadamente el problema.
3. En la etapa de análisis de datos, se abordó el desafío de adquirir los datos y comprender sus dimensiones. Se realizó un estudio exhaustivo de los dos conjuntos de datos disponibles.
4. En la fase de preparación de los datos, se realizó el procesamiento necesario para estructurar correctamente los datos de ambos escenarios.
5. En la fase de modelado, se llevaron a cabo los entrenamientos de los diversos modelos utilizando los dos escenarios establecidos.
6. En la etapa final, se evaluaron los modelos entrenados mediante un conjunto de prueba, con el objetivo de generar conocimiento concluyente.

4. Planteamiento de la comparativa

En el presente apartado se presentan las características y condiciones de los conjuntos de datos tanto reales como simulados. Se realizará la explicación del preprocesamiento empleado a ambos conjuntos. Con los datos tratados se explicará el proceso de selección que se empleó para ambos conjuntos de datos. Para finalizar el capítulo se muestran las configuraciones que se emplearon en los modelos. Los parámetros y las métricas empleadas para cuantificar el desempeño se encuentran plasmado dentro del capítulo.

4.1. Adquisición de datos

En el estudio se obtuvo la información desde dos fuentes diferentes:

- Naturaleza clínica – Información Real
- Naturales Artificial – Información Generada

4.1.1. Conjunto de información de Naturaleza Real

Los datos clínicos empleados en el presente estudio fueron obtenidos mediante la toma continua de pacientes con DMT1, los pacientes que participaron en el estudio pertenecen a la provincia de Granada España. Estos datos han sido parte de análisis para la evaluación de la glucemia en diferentes pacientes. Las mediciones de la glucosa fueron capturadas con el sensor FreeStyle Libre 2. Pueden existir variaciones en la información debido a las limitaciones del sensor. Este tipo de sensor se puede usar hasta 14 días, por lo cual se debe de tener consigo otro sensor que se encuentre disponible para poder suplir las lecturas de glucosa en caso de que termine su uso (Abbott, 2020). El conjunto de datos que contemplan las características descritas en la Tabla 3.

Dentro del conjunto de datos se tiene un total de 110 pacientes que abordan en conjunto tres millones de registro. Las mediciones registradas se encuentran descritas en un rango entre $40 \left[\frac{mg}{dL} \right]$ y $500 \left[\frac{mg}{dL} \right]$. Para observar el comportamiento de la glucosa del conjunto de pacientes tenemos la descripción presente en la Tabla 4.

Tabla 3: Características de los datos

Variable	Tipo	Rango
Patient_ID	Char	110 pacientes
Measurment_date	Date	[2018-02-21: 2022-03-21]
Measurement_time	Time	Una ventana de tiempo de aproximadamente 15 minuto. ¹
Measument	Int	[40 : 500] $\left[\frac{mg}{dL}\right]$

Tabla 4: Rangos de Mediciones de la Glucosa en pacientes reales.

Medición estadística	Valor $\left[\frac{mg}{dL}\right]$
Mínimo	40
Primer Cuartil	110
Mediana	148
Promedio	159
Tercer Cuartil	197
Máximo	500

4.1.2. Conjunto de información de Naturaleza Artificial

El simulador empleado en el estudio se trata de UVA/Padova T1D simulator, este simulador se encuentra aprobado por la FDA sobre las simulaciones del nivel de la glucosa.

Este simulador posee diferentes condiciones ajustables, las mismas que constan sensores, bomba de insulina y escenarios.

4.1.2.1. Sensores:

Sensores CGM que eran comercialmente disponibles en la versión del simulador del 2008.

- Dexcom: 3 min de tiempo de muestreo

¹ Los conjuntos de datos presentes en la muestra de datos recopilados no poseen una distribución continua de información. Existen usuarios con datos faltantes en el día y registros que no poseen continuidad.

- GuardianRT: 5 min de tiempo de muestreo
- Navigator: 1 min de tiempo de muestreo

Se muestra en la Tabla 5, los parámetros con los que se encuentran descritos los sensores en el simulador.

Tabla 5: Condiciones de los sensores simulados.

Sensor	PACF	Gsmms	Lambda	Delta	Xi	Tiempo de muestreo	Min	Max
Dexcom	0.7	-0.5444	15.9574	1.6898	-5.47	3.0	39	600
GuardianRT	0.7	-0.5444	15.9574	1.6898	-5.47	5.0	39	600
Navigator	0.7	-0.5444	15.9574	1.6898	-5.47	1.0	32	600

Fuente: Jinyu Xie. Simglucose v0.2.1 (2018) [Online]. Available: <https://github.com/jxx123/simglucose>.

Es importante reconocer que la cantidad de datos que generará el simulador se encontrarán con una separación acorde al sensor empleado, y estarán espaciados según se describe en la Tabla 5, a partir de la columna de tiempo de muestreo.

4.1.2.2. Bombas de insulina

simGlucose tiene dos modelos:

- Cozmo – Deltec Cozmo, descontinuada en 2009
- Insulet – Omnipod, manufacturada por Insulet

En la Tabla 6, se describen los parámetros con los que se simulan las bombas de insulina.

Tabla 6: Condiciones de las bombas de insulina simuladas.

Bomba de insulina	Min_bolus	Max_bolus	Inc_bolus	Min_basal	Max_basal	Inc_basal	Sample_time
Cozmo	0.0	75	0.005	0.0	35	0.05	1
Insulet	0.0	30	0.05	0.0	30	0.05	1

Fuente: Jinyu Xie. Simglucose v0.2.1 (2018) [Online]. Available: <https://github.com/jxx123/simglucose>.

4.1.2.3. Escenarios:

El simulador permite incluir diferentes parámetros para la simulación, tales como son:

- Comidas
 - Calorías ingresadas por horario
 - Aleatorio
- Controlador
 - PIDController(): PID
 - BBController(): Control Bolo Basal

4.1.2.4. Respuesta:

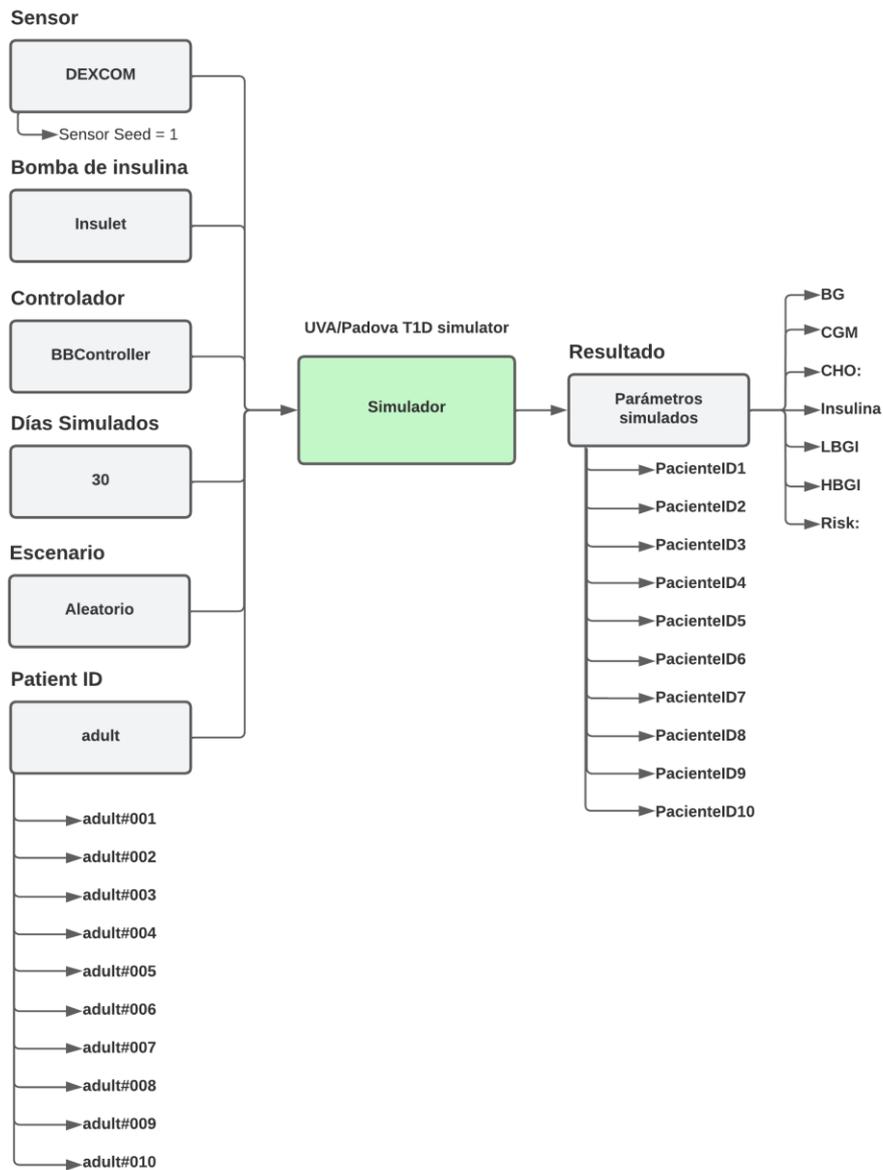
El simulador devuelve un archivo con extensión CSV, el cual contiene 7 valores.

- BG: Glucosa Real en la sangre
- CGM: CGM en la glucosa de la sangre
- CHO: Gramos de carbohidratos (comida)
- Insulina: Insulina liberada en el intervalo
- LBG: Indicador de Glucosa en sangre bajo
- HBGI: Indicador de Glucosa en sangre alto
- Risk: Indicador de riesgo

Se establecen las siguientes condiciones para los 10 pacientes simulados. De cada paciente se simulan 20 días de información. Las condiciones² con las que se simularon los 10 pacientes presentes en el estudio se encuentran descritas en la Figura 12.

² Se establece una condición de escenario “aleatoria” en el cual, la cantidad de comida y horario de ingesta son generados aleatoriamente por el algoritmo

Figura 12: Condiciones de los pacientes simulados.



4.2. PROCESAMIENTO DE LOS DATOS

Una vez descrita la naturaleza de los datos, se procede a establecer criterios para la selección de los datos acorde a la información que se maneja en el presente estudio. Estas características y consideraciones se emplearon en los dos conjuntos de datos. A partir del sensor usado con la muestra de los usuarios reales, se debe de tener en cuenta las siguientes consideraciones:

- **Muestreo constante de 15 minutos:** En función que el sensor utilizado genera muestras de usuarios reales con un muestreo de cada 15, se debe seleccionar aquellos usuarios que tengan una muestra lo más constante posible en este intervalo. Esto implica que se deben descartar las muestras que no cumplan con este muestreo constante, ya que podrían generar inexactitudes en el análisis y la comparativa. Los criterios de selección empleados se los describe con mayor profundidad en el apartado de criterios de selección.
- **Selección de muestras del simulador:** Como el sensor empleado en el simulador muestrea valores cada 3 minutos, sensor Dexcom. Esto implica que se deben seleccionar las muestras simuladas de manera que haya una separación de 15 minutos entre las diferentes muestras. Esto permitirá una mejor comparación entre los escenarios de pacientes reales y simulados.

El diseño del estudio tomó criterios del artículo descrito por (Carrillo-Moreno et al., 2021), los mismo que se los explica a continuación:

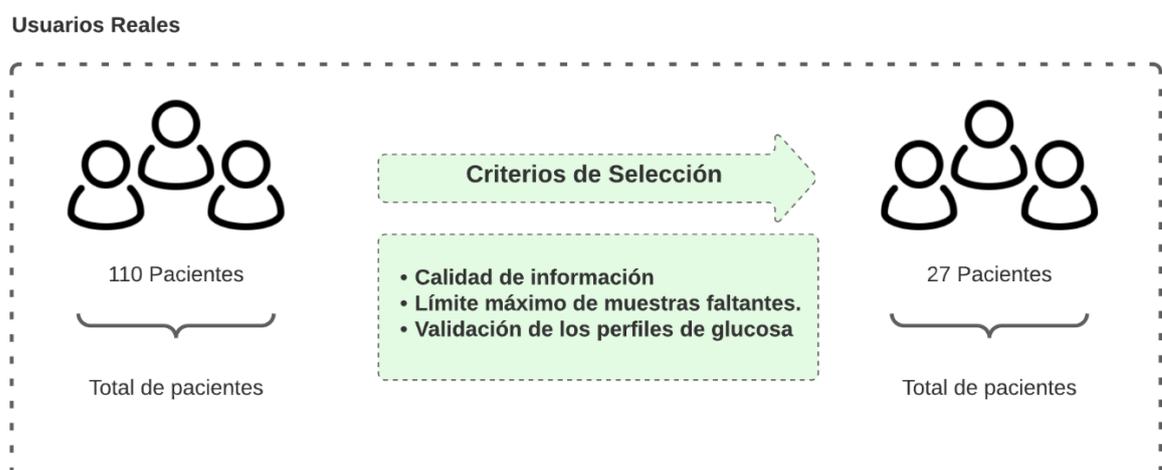
- **Perfil de glucosa:** Se refiere a la cantidad de muestras obtenidas por periodo de 15 min en un día. En este caso, se considera que 1 perfil tiene 96 muestras por día.
- **Imputación de datos faltantes:** Cuando existen datos faltantes en el perfil de glucosa, se imputan utilizando el valor promedio del valor anterior y el valor siguiente en el tiempo. Con la ayuda de esta técnica se pretende completar los datos ausentes y mantener la continuidad en la serie temporal.
- **Normalización de valores de glucosa:** Los valores de glucosa se normalizan para facilitar el aprendizaje adecuado de la red neuronal. Este proceso se realiza mediante un escalamiento de variables min-max.
- **Ventana de histórico (VH):** Es la secuencia de valores de glucosa que se utilizará como entrada para el modelo predictor. Esta ventana contiene los valores históricos de glucosa necesarios para realizar la predicción.
- **Horizonte de predicción (HP):** El horizonte de predicción se refiere al instante en el que se realizará la predicción. Se expresa en unidades de minutos y determina el tiempo en el futuro al que se desea predecir los valores de glucosa.

- **División en subconjuntos:** El conjunto de datos se divide en tres subconjuntos: entrenamiento, validación y prueba. Todos los perfiles de glucosa pertenecientes a un paciente se encontrarán exclusivamente en uno de los subconjuntos.
- **Selección de pacientes y cantidad mínima de perfiles:** Se seleccionan pacientes que tengan al menos 12 perfiles, lo que representa un mínimo de 1152 muestras para utilizar en el conjunto de datos. Entre estos pacientes, se asignan 6 pacientes al conjunto de entrenamiento, 2 pacientes al conjunto de validación y 4 pacientes al conjunto de prueba.
- **Evaluación de modelos:** La evaluación de los modelos predictores se realizará empleando la métrica del error cuadrático medio (RMSE) para medir la diferencia entre los valores reales y las predicciones. Así mismo se emplearán las gráficas de Clarke para evaluar la concordancia entre las mediciones reales y las predicciones realizadas por cada modelo.

4.2.1. Descripción de selección de datos de pacientes reales

De los 110 pacientes obtenidos, se realizó una validación previa bajo el siguiente criterio de selección, tal como se aprecia en la Figura 13. Los criterios de selección y de filtrado se describen a continuación.

Figura 13: Selección de conjunto de datos de pacientes reales.



Criterios de selección:

- a) Calidad de la información: Se requiere un mínimo de 12 perfiles de glucosa consecutivos, cada uno de ellos con aproximadamente 96 muestras. La cantidad de muestras puede variar dependiendo de la hora de inicio de toma del sensor. Es importante que los perfiles de glucosa sean consecutivos para que se consideren una secuencia temporal válida. Se busca un mínimo de 12 días seguidos con alrededor de 96 muestras en cada perfil.
- b) Límite máximo de muestras faltantes: Se establece un límite máximo del 4% de muestras faltantes en la serie temporal. Esto significa que se permite un máximo de 3 pasos de la serie temporal imputados. Los perfiles con un número igual o mayor a 93 muestras se consideran válidos y se incluirán en el estudio. Se escoge un porcentaje de 4% para mantener la integridad de la serie temporal, es decir, que la tasa relativa entre datos reales y datos imputados sea mínima. Esto ayuda a conservar calidad y representatividad.
- c) Validación de los perfiles de glucosa: Se realiza una validación de los perfiles de glucosa por paciente. En caso de que existan datos faltantes en la serie temporal, se realiza una imputación de los datos utilizando técnicas apropiadas, como el promedio del valor anterior y posterior. La Figura 14 proporciona más información sobre el proceso de imputación de datos faltantes.

Figura 14: Código en Python empleado para la Imputación de datos mediante el promedio del dato anterior y posterior al dato faltante.

```
import numpy as np

def completar_valores_faltantes(serie_temporal):
    serie_completada = serie_temporal.copy()

    for i in range(len(serie_completada)):
        if np.isnan(serie_completada[i]):
            if i == 0:
                # Si el primer valor es faltante, establecer el siguiente valor como el promedio
                siguiente_valor = serie_completada[i+1]
                serie_completada[i] = siguiente_valor
            elif i == len(serie_completada) - 1:
                # Si el último valor es faltante, establecer el valor anterior como el promedio
                valor_anterior = serie_completada[i-1]
                serie_completada[i] = valor_anterior
            else:
                # Si el valor faltante está en el medio, calcular el promedio del valor anterior y posterior
                valor_anterior = serie_completada[i-1]
                siguiente_valor = serie_completada[i+1]
                serie_completada[i] = (valor_anterior + siguiente_valor) / 2.0

    return serie_completada
```

Los criterios de filtro y validación aseguran la calidad y la consistencia de los perfiles de glucosa que se emplean en el estudio. La inclusión de perfiles consecutivos, la validación de datos faltantes, así como el límite máximo de muestras faltantes garantizan que se utilicen datos confiables y representativos en el análisis de la serie temporal de glucosa. Al aplicar los criterios anteriores, quedó un total de 27 pacientes, cuyas métricas se pueden apreciar en la Tabla 7.

- d) Selección de usuarios finales: El paso final consistió en considerar únicamente los datos de aquellos pacientes que presentan más de 12 perfiles de glucosa de forma consecutiva. De los 27 pacientes mencionados previamente, se seleccionaron los 12. Estos pacientes que cumplieron con el requisito de tener perfiles de glucosa consecutivos fueron considerados en el estudio. Además de los perfiles de glucosa, se presentan los valores estadísticos de la glucosa para esta muestra seleccionada. La información se la separó según los grupos de datos a emplear dentro del estudio, en la Tabla 8 se presenta la información del conjunto de pacientes empleados en el entrenamiento de los modelos, en la Tabla 9 se presenta la información del conjunto de pacientes que se empleó en la validación y en la Tabla 10 la información del conjunto de pacientes empleados en la prueba de los modelos.

Tabla 7: Métrica de los pacientes que pasaron el filtro de selección.

PatientID	GlucosaMin	GlucosaMáx	GlucosaProm
LIB193269	40	492	156,09
LIB193303	53	438	177,78
LIB193307	40	490	157,64
LIB193310	40	415	146,71
LIB193311	40	500	187,57
LIB193317	45	464	160,90
LIB193320	46	399	145,21
LIB193335	40	419	159,32
LIB193338	52	291	122,43
LIB193346	53	500	273,69
LIB193351	51	500	194,20
LIB193353	51	336	144,98
LIB193356	47	498	171,43
LIB193358	52	500	207,14
LIB193367	40	439	154,65
LIB193370	53	500	189,69
LIB193376	40	409	146,50
LIB193378	52	431	168,12
LIB193379	56	487	177,61
LIB193385	40	304	123,76
LIB193386	40	414	133,71
LIB193406	53	388	164,35
LIB193407	40	500	177,86
LIB193408	40	371	138,66
LIB193410	40	391	126,43
LIB193420	49	500	153,01
LIB193423	54	299	121,80

Tabla 8: Información del grupo de pacientes empleados en el conjunto de entrenamiento

PatientID	Conjunto de datos	#Perfiles	Min	Primer Cuartil	Mediana	Promedio	Tercer Cuartil	Max
LIB193385	Entrenamiento	25	54	125.5	165	170.683055	213	386
LIB193408	Entrenamiento	25	45	125	160	166.447786	203	351
LIB193378	Entrenamiento	25	40	129	175	178.083508	227	363
LIB193310	Entrenamiento	20	56	129	158	164.367928	195	331
LIB193307	Entrenamiento	12	57	131	166	175.455696	209	487
LIB193353	Entrenamiento	12	43	94	129	137.673117	175	328
LIB193338	Entrenamiento	12	40	92	130	138.087066	171	348

Tabla 9: Información del grupo de pacientes empleados en el conjunto de validación

PatientID	Conjunto de datos	#Perfiles	Min	Primer Cuartil	Mediana	Promedio	Tercer Cuartil	Max
LIB193367	Validación	25	40	95	134	138.661106	175	341
LIB193410	Validación	25	59	103	119	123.324584	138	250

Tabla 10: Información del grupo de pacientes empleados en el conjunto de prueba

PatientID	Conjunto de datos	#Perfiles	Min	Primer Cuartil	Mediana	Promedio	Tercer Cuartil	Max
LIB193335	Prueba	25	53	106	145	146.011344	182	272
LIB193379	Prueba	20	55	123	166	176.145359	214	431
LIB193303	Prueba	12	52	106	124	126.512349	144	240

A partir de la Tabla 8, Tabla 9 y Tabla 10 se tienen las cantidades de perfiles por conjunto de entrenamiento:

- Entrenamiento: 131 perfiles que hacen referencia a un total de 12576 registros.
- Validación: 50 perfiles que hacen referencia a un total de 4800 registros.
- Prueba: 61 perfiles que hacen referencia a un total de 5856 registros.

4.2.2. Datos de pacientes simulados

Para generar los datos de los pacientes simulados se empleó el simulador de UVA/Padova que se encuentra desarrollado en el ambiente de Python, a partir de las consideraciones descritas en la Figura 12, se empleó el código descrito en la Figura 15 para su ejecución y obtención de los perfiles a emplear.

Figura 15: Código en Python empleado para la generación de los pacientes mediante el simulador.

```
# Importar las librerías
from simglucose.simulation.env import T1DSimEnv
from simglucose.controller.basal_bolus_ctrller import BBController
from simglucose.sensor.cgm import CGMSensor
from simglucose.actuator.pump import InsulinPump
from simglucose.patient.t1dpatient import T1DPatient
from simglucose.simulation.scenario qen import RandomScenario

from datetime import timedelta
from datetime import datetime

# Se especifica el tiempo en que se empieza la simulación
now = datetime.now()
start_time = datetime.combine(now.date(), datetime.min.time())

# ----- Creación del escenario -----
# Especificación de la carpeta en donde se almacenarán los resultados
path = './results/Random'

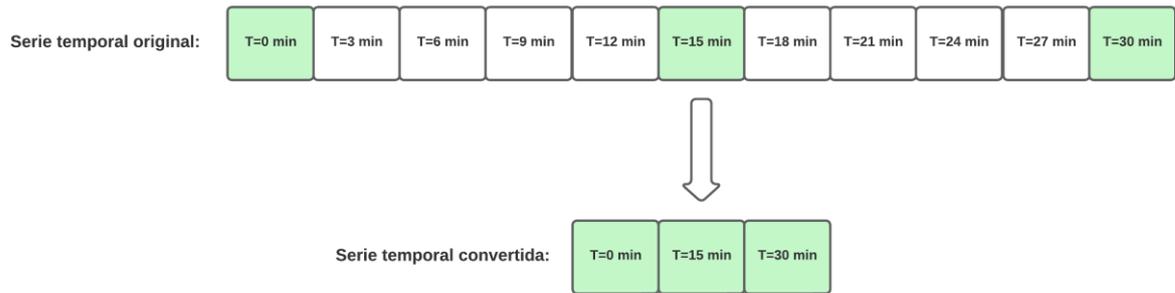
# Creación del ambiente de simulación
patient = T1DPatient.withName('adult#001') # Se modifica con 'adult#001' hasta el 'adult#010'
sensor = CGMSensor.withName('Dexcom', seed=1)
pump = InsulinPump.withName('Insulet')
scenario = RandomScenario(start_time=start_time, seed=1)
env = T1DSimEnv(patient, sensor, pump, scenario)

# Creación del controlador
controller = BBController()

# Creación del objeto de simulación
s1 = SimObj(env, controller, timedelta(days=30), animate=False, path=path)
results1 = sim(s1)
print(results1)
```

En el caso del procesamiento de los datos de los pacientes simulados es más sencillo, debido a que los perfiles que genera el modelo no poseen los problemas observados en el conjunto de pacientes reales, como lo es la falta de datos y la discontinuidad de estos.

La única consideración es seleccionar los datos con una distancia de 15 minutos para mantener la consistencia con el intervalo deseado. Como se menciona en el estado del arte, el simulador genera los perfiles de glucosa utilizando las características del sensor que se seleccione. En el caso del estudio planteado, se seleccionó el sensor Dexcom, el cual proporciona una ventana de 3 minutos entre datos. Se realiza una adecuada selección de los datos tal como se muestra en la Figura 16.

Figura 16: Tratamiento de datos de pacientes simulados.

El proceso descrito previamente implica seleccionar una muestra de cada cinco lecturas proporcionadas por el sensor. Esto asegura que los datos simulados estén en intervalos de 15 minutos y cumplan con la periodicidad requerida para el análisis y la predicción. Los datos de los 10 pacientes simulados se los separó. En la Tabla 11 se presenta la información del conjunto de pacientes empleados en el entrenamiento de los modelos, en la Tabla 12 se presenta la información del conjunto de pacientes que se empleó en la validación y en la Tabla 13 la información del conjunto de pacientes empleados en la prueba de los modelos.

Tabla 11 Información del grupo de pacientes simulados empleados en el conjunto de entrenamiento.

PatientID	#Perfiles	Mínimo	Primer Cuartil	Mediana	Promedio	Tercer Cuartil	Máximo
ID1	22	39.00	74.58	106.62	115.55	150.24	283.08
ID4	22	39.00	87.57	112.69	125.12	147.00	314.89
ID7	22	39.00	91.22	109.84	118.00	140.47	244.11
ID8	22	39.00	76.46	105.11	108.39	137.03	224.94
ID9	22	39.00	60.42	110.71	113.80	156.91	263.95
ID10	21	39.00	80.82	110.81	145.97	181.97	415.97

Tabla 12: Información del grupo de pacientes simulados empleados en el conjunto de validación.

PatientID	#Perfiles	Mínimo	Primer Cuartil	Mediana	Promedio	Tercer Cuartil	Máximo
ID2	25	40.74	83.26	100.50	106.94	127.83	218.87
ID4	25	39.00	87.57	112.69	125.12	147.00	314.89

Tabla 13: Información del grupo de pacientes simulados empleados en el conjunto de prueba.

PatientID	#Perfiles	Mínimo	Primer Cuartil	Mediana	Promedio	Tercer Cuartil	Máximo
ID3	30	39.00	84.23	115.41	125.58	152.69	308.68
ID6	30	39.00	88.48	114.47	123.33	155.70	268.83

Para tener una data en igual proporción con el escenario con los pacientes reales:

- Entrenamiento: 131 perfiles que hacen referencia a un total de 12576 registros.
- Validación: 50 perfiles que hacen referencia a un total de 4800 registros.
- Prueba: 60 perfiles que hacen referencia a un total de 5760 registros.

4.2.3. Transformación de los datos

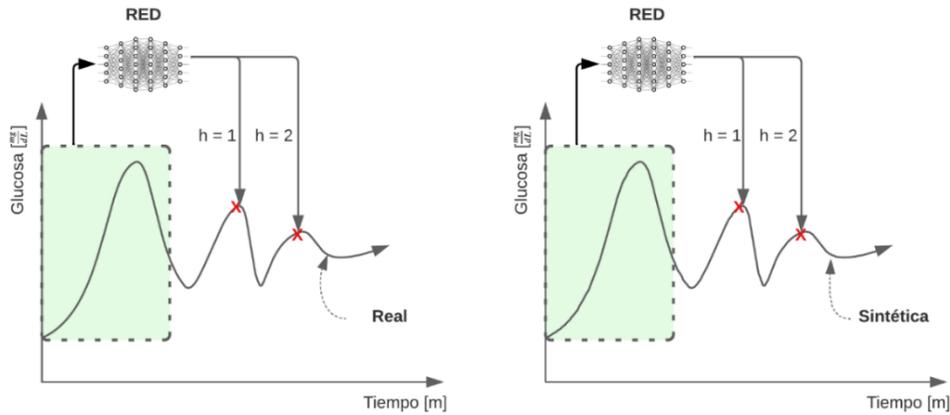
En el presente trabajo se propone utilizar una transformación de datos mediante una ventana deslizante para predecir los niveles de glucosa en el futuro. Esto implica trabajar con una serie temporal de glucosa y utilizar una ventana deslizante que tome los "n" valores históricos para predecir los niveles de glucosa "h" pasos en el futuro.

Para lograr esto, la serie de tiempo se divide en dos matrices: una matriz "X" que contendrá los valores de entrada, es decir, los "n" valores históricos, y una matriz "Y" que contendrá los valores de glucosa "h" pasos en el futuro.

En la Figura 17, se observa que a la "RED" se entrena utilizando elementos de la serie temporal para predecir los valores en dos momentos futuros "h". Dado que el sensor utilizado en el

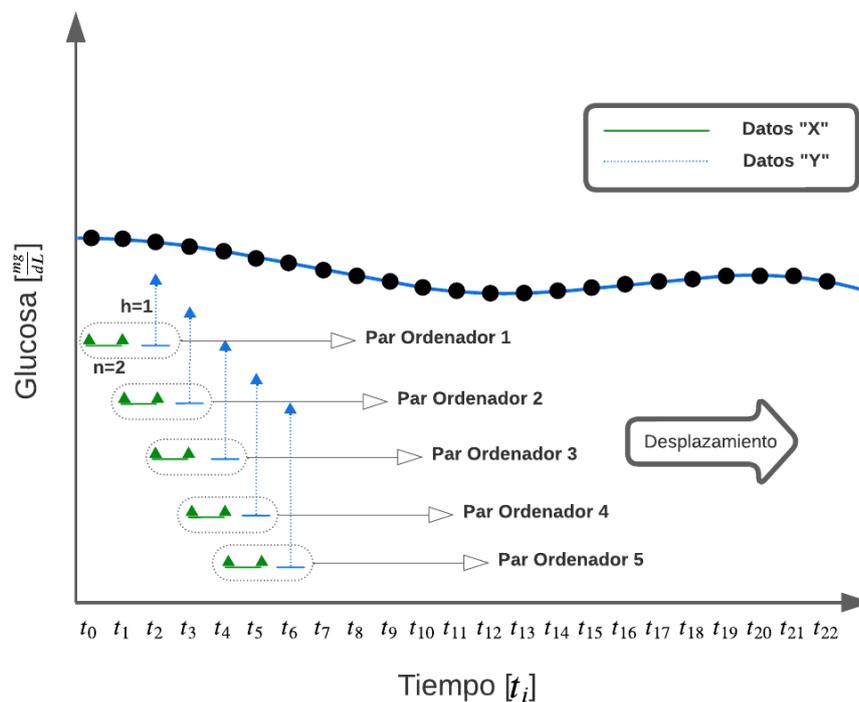
estudio realiza mediciones cada 15 minutos, se eligen dos valores para "h": $h = 1$ y $h = 2$, lo que proporciona una ventana de predicción de 15 y 30 minutos respectivamente.

Figura 17: Modelo de predicción desarrollado, implementado a data real y data sintética.



Se plantea el siguiente ejemplo: El proceso de selección para una serie temporal con una ventana deslizante de 30 minutos de histórico y una predicción de 15 minutos se muestra en la Figura 18. En este caso, se utiliza un valor de "n" igual a 2, lo que equivale a 30 minutos de datos históricos, y un valor de "h" igual a 1, es decir, una predicción de 15 minutos en el futuro.

Figura 18: Ejemplo de ventana deslizante para predicción de glucosa a 15 minutos con un histórico de 30 minutos.



Cada punto en la serie temporal representa una muestra capturada por el sensor. En la Figura 18, los puntos en la línea verde representan los datos que se utilizarán como entrada para entrenar el modelo, mientras que el valor en la línea azul representa el valor que se tomará como la variable objetivo, es decir, el valor a predecir.

Para construir estas ventanas de datos, se utiliza el código descrito en la Figura 19. En este código, se ha creado una función que recibe tres parámetros: la serie de datos, la ventana y el paso. La ventana se refiere al número de elementos históricos "n" que se incluirán en cada ventana, y el paso hace referencia a los "h" pasos utilizados para predecir los valores de glucosa.

Figura 19: Código en Python empleado para generar las matrices con diferente tamaño de ventana.

```
def dividir_serie(serie, ventana, paso):
    X = []
    Y = []

    for i in range(0, len(serie) - ventana):

        if (i+ventana+paso) > len(serie)-1:
            break

        ventana_actual = serie[i:i+ventana]
        siguiente_valor = serie[i+ventana+paso]

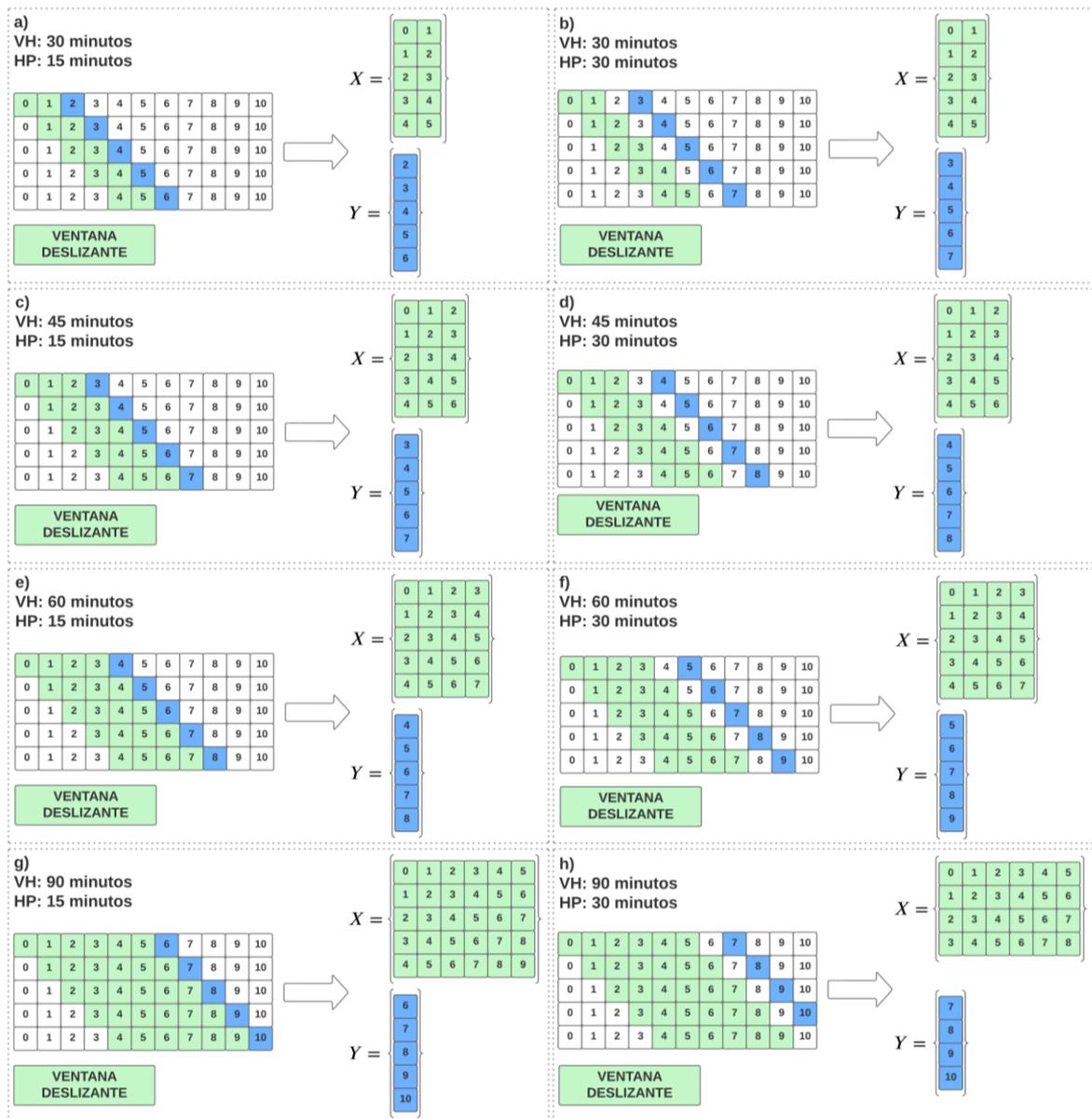
        X.append(ventana_actual)
        Y.append(siguiente_valor)

    X = np.array(X)
    Y = np.array(Y)

    return X, Y
```

En este estudio en particular, se plantea utilizar una VH de 30, 45, 60 y 90 minutos, bajo dos HP de 15 y 30 minutos. Para mostrar la forma de selección de datos se plantea la Figura 20, en donde a partir de los primeros 10 puntos de una serie temporal, se plantean las 8 combinaciones a realizar en el estudio. Cada literal de la Figura muestra la forma de generación de la ventana de histórico, así como del horizonte de predicción.

Figura 20: Comportamiento de la ventana deslizante para la generación de los conjuntos de datos.



Mediante esta función, se generan las matrices "X" y "Y" que serán utilizadas en el estudio. La matriz "X" contendrá los valores históricos de la serie de datos, conformados según la configuración de la ventana, mientras que la matriz "Y" contendrá los valores de glucosa a predecir, correspondientes a los pasos futuros indicados.

Los códigos empleados en el trabajo se encuentran en el siguiente enlace:

<https://github.com/josep5097/TFM-Maestria.git>

5. Desarrollo de la comparativa

En el presente capítulo, se llevará a cabo la comparativa entre los escenarios con pacientes diabéticos reales y simulados. Para realizar esta comparativa, se abordarán las siguientes temáticas:

- Visualización de las series temporales: Se mostrarán gráficamente las series temporales de glucosa tanto para los pacientes reales como para los pacientes simulados. Esta visualización se centrará en un periodo de 1 día.
- Modelos predictores utilizados: Se presentarán los 6 modelos predictores empleados en el estudio. Se describirá la estructura y configuración de cada modelo, así como los resultados y el error obtenido, en este caso, se utilizará la métrica RMSE (Root Mean Square Error) y las gráficas de Clarke.

5.1. VISUALIZACIÓN DE SERIES TEMPORALES

La primera parte del estudio es la parte de la visualización de la serie temporal que se genera a partir de los datos, para lo cual se realiza una muestra de un perfil de un paciente real y un paciente simulado.

5.1.1. Pacientes reales

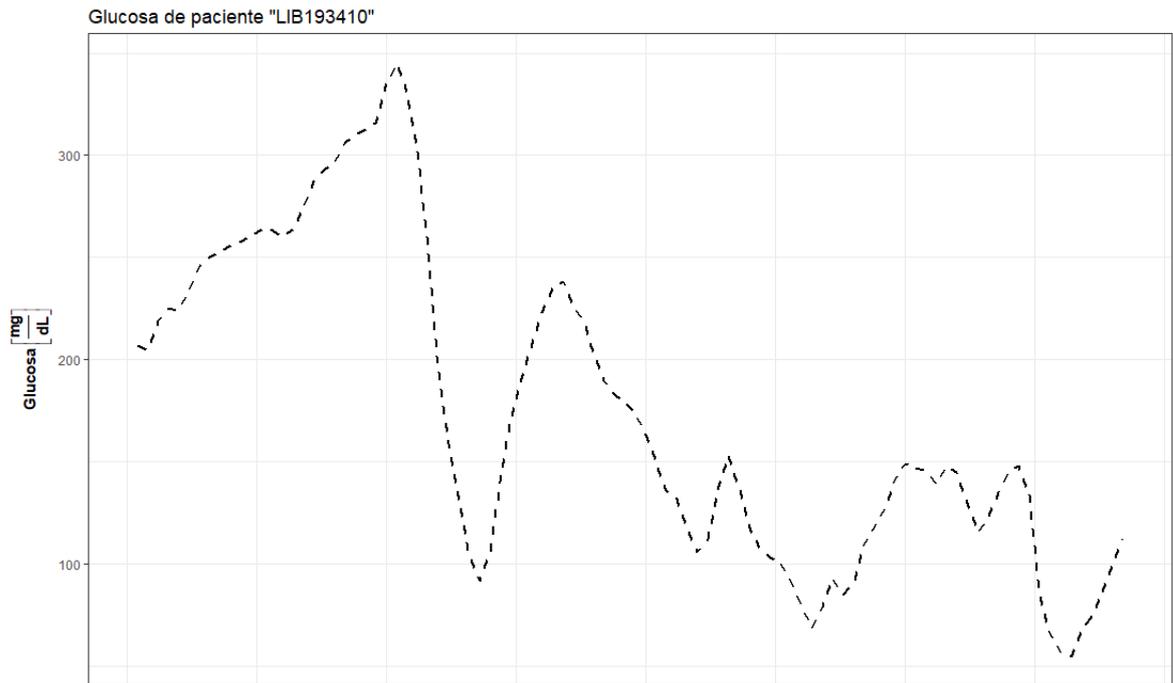
Se toma como ejemplo, uno de los perfiles del Paciente **LIB193410**. Dicho perfil como se mencionó contempla la información 1 día, referirse a la Figura 21 literal a). Se debe de considerar que esta serie temporal se la debe de normalizar para los procesamientos internos del modelo, en la Figura 21 literal b) se observa el perfil de la glucosa normalizada del paciente.

5.1.2. Pacientes simulados

La primera parte de la comparativa parte de la visualización de la serie temporal que se genera a partir de los datos. Se toma como ejemplo, uno de los perfiles del Paciente **ID1**. Dicho perfil como se mencionó contempla la información 1 día, referirse a la Figura 22 literal a). Se debe de considerar que esta serie temporal se la debe de normalizar para los procesamientos internos del modelo, en la Figura 22 literal b) se observa el perfil de la glucosa normalizada del paciente.

Figura 21: Glucosa vs Tiempo del paciente LIB193410 medida en 1 día. a) Valor muestreado. b) Valor normalizado

a)



b)

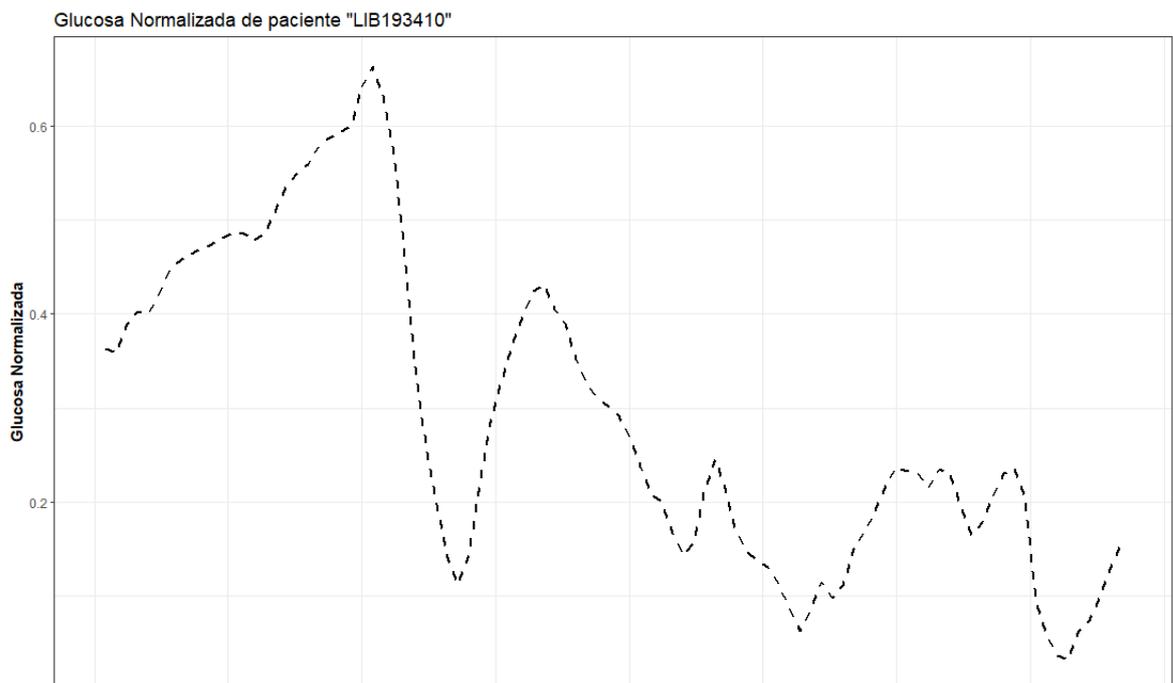
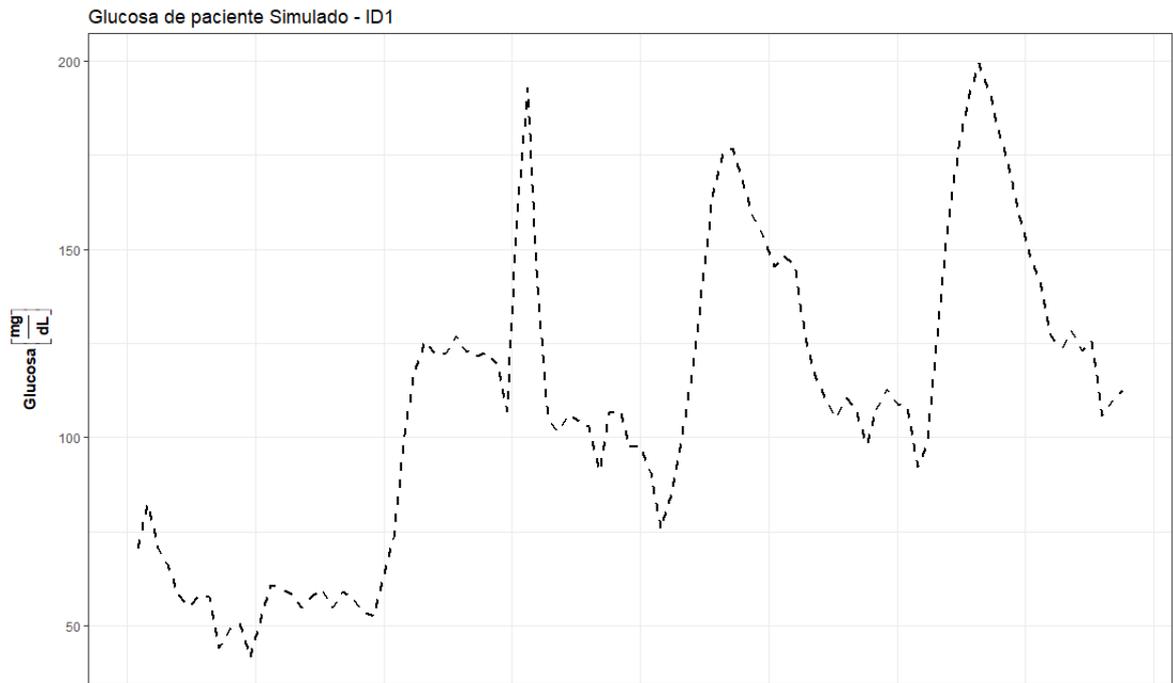
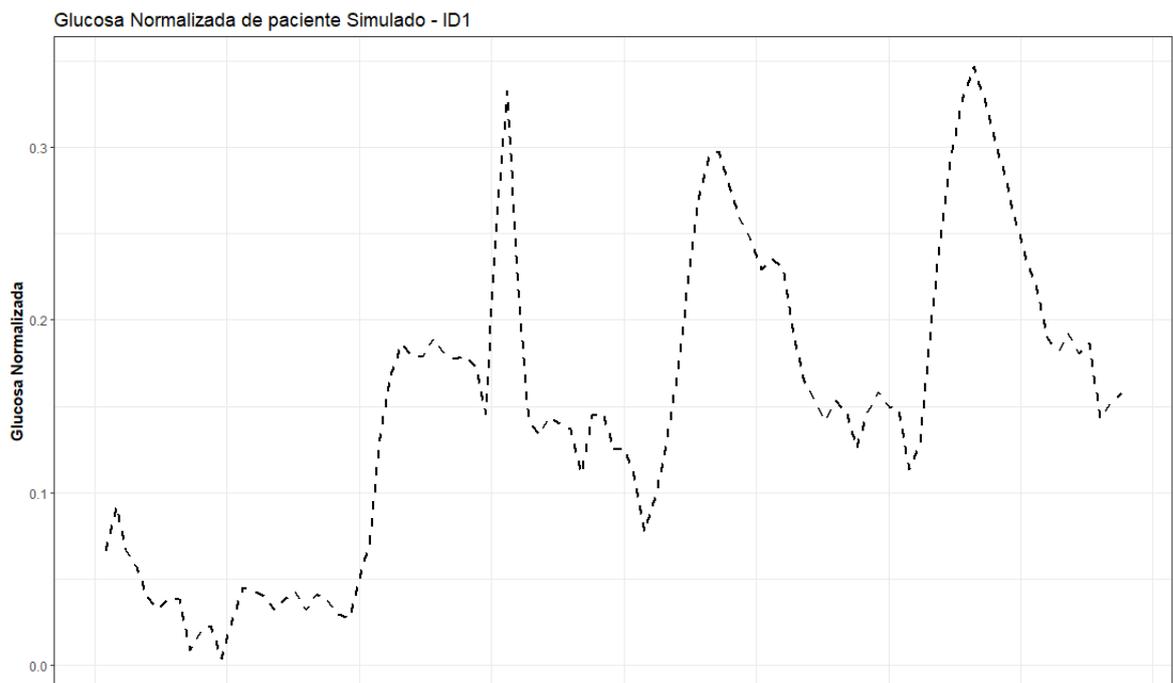


Figura 22: Glucosa vs Tiempo del paciente PaciecienteID1 simulado medida en 1 día. a) Valor muestreado. b) Valor normalizado

a)



b)



5.2. MODELAMIENTO

En el presente apartado se mostrarán los diferentes modelos que se generaron en el estudio. Los modelos se los entrenará en los dos escenarios, tanto para los datos reales como para los datos simulados. Los modelos de aprendizaje profundo desarrollados se los describe en la Tabla 14.

Tabla 14: Modelos de aprendizaje profundo desarrollados

Modelo	Característica
Modelo 1	Modelo de 1 capa RNN: <ul style="list-style-type: none"> • 1 Capa RNN con 64 neuronas • 1 Capa de salida lineal
Modelo 2	Modelo de 1 capa LSTM: <ul style="list-style-type: none"> • 1 Capa LSTM con 64 neuronas • 1 Capa de salida lineal
Modelo 3	Modelo de 2 capas LSTM: <ul style="list-style-type: none"> • 1 Capa LSTM con 128 neuronas • 1 Capa LSTM con 128 neuronas • 1 Capa ReLU con 64 neuronas 1 Capa de salida lineal
Modelo 4	Modelo de 1 capa GRU: <ul style="list-style-type: none"> • 1 Capa GRU con 64 neuronas • 1 Capa de salida lineal
Modelo 5	Modelo de 2 capas GRU: <ul style="list-style-type: none"> • 1 Capa GRU con 128 neuronas • 1 Capa GRU con 128 neuronas • 1 Capa ReLU con 64 neuronas • 1 Capa de salida lineal
Modelo 6	Modelo de 1 capa LSTM + 1 capa GRU <ul style="list-style-type: none"> • 1 Capa LSTM con 64 neuronas • 1 Capa GRU con 64 neuronas • 1 Capa ReLU con 64 neuronas • 1 Capa de salida lineal

En el estudio se están considerando 6 modelos diferentes. Cada modelo será entrenado con 4 ventanas de histórico y 2 horizontes de predicción. Esto genera un total de 48 modelos para cada escenario. Tal como se ha mencionado, el estudio contempla 2 escenarios. En ambos

escenarios se realizan las mismas configuraciones para el entrenamiento de los modelos. Por lo tanto, se tendrá un total de 96 modelos desarrollados en el estudio.

Para compilar los modelos, se siguió el patrón de estudio empleado en (Carrillo-Moreno et al., 2021) para lo cual se hará uso del optimizador ADAM, y la métrica MSE para minimizar la función de pérdida, referirse a la Ecuación 1, debido a que esta métrica simplifica los cálculos de la gradiente.

$$MSE = \frac{1}{N} \sum_{i=1}^N (t_i - y_i)^2 \quad (1)$$

t_i : Valor Esperado (Valor objetivo)

y_i : Predicción realizada

En los diferentes estudios revisados la forma de evaluar el desempeño del modelo predictor es el RMSE, referirse a la Ecuación 2, considerando que a si comparamos con la métrica MSE, esta se encuentra en la misma unidad que la variable evaluada.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (t_i - y_i)^2} \quad (2)$$

t_i : Valor Esperado (Valor objetivo)

y_i : Predicción realizada

Los resultados se los revisará por cada uno de los modelos de forma individual. Esto implica que por cada modelo se mostrará el desempeño para el escenario con datos reales y para datos simulados.

5.2.1. Modelo 1: Modelo de 1 capa RNN

Se empieza el experimento con una red con 1 capa RNN. Para ello generamos un modelo secuencial, y le agregamos una capa SimpleRNN de 64 neuronas. La entrada de la capa inicial se encuentra en función del tamaño de la ventana en la cual se encuentra generado el modelo. A continuación, se lo conecta a una densa de salida con la función de activación lineal. La creación del modelo fue realizada como en se observa en la Figura 23.

Figura 23: Código en Python empleado para generar el Modelo 1

```

from keras.models import Sequential
from keras.layers import SimpleRNN, Dense

# Función para crear el Modelo RNN Simple
def crear_modelo():
    modelo = Sequential()
    modelo.add(SimpleRNN(64, input_shape=(ventana, 1)))
    modelo.add(Dense(1, activation='linear'))

    return modelo

# Crear el modelo LSTM
modelo = crear_modelo()

# Compilar y entrenar el modelo
modelo.compile(optimizer='adam', loss='mse', metrics=['accuracy'])

```

5.2.1.1. Pacientes reales

El modelo 1, generado según se describe en la Figura 23, fue entrenado utilizando los 8 casos de estudio mencionados anteriormente. Durante el proceso de entrenamiento, se utilizó el conjunto de prueba para realizar la predicción. Con el valor real y la predicción se obtuvo el RMSE, se generaron las gráficas que muestran el comportamiento de la serie temporal y se generaron las gráficas de Clarke.

En la Tabla 15, se presenta un resumen de los resultados obtenidos para un horizonte de predicción de 15 minutos. Este resumen incluye el valor del RMSE para cada ventana de histórico empleado y el respectivo porcentaje de las zonas de Clarke. Por otra parte, en la Tabla 16, se presentan las mismas métricas, pero para un horizonte de predicción de 30 minutos.

Tabla 15: Métricas del modelo 1 para un horizonte de predicción de 15 minutos en pacientes reales

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	19.36	88.05%	11.49%	99.54%	0.00%	0.46%	0.00%
VH: 45 minutos	14.04	96.12%	3.75%	99.87%	0.00%	0.13%	0.00%
VH: 60 minutos	12.12	96.81%	3.02%	99.83%	0.00%	0.16%	0.00%
VH: 90 minutos	11.46	97.30%	2.53%	99.83%	0.00%	0.16%	0.00%

Tabla 16: Métricas del modelo 1 para un horizonte de predicción de 30 minutos en pacientes reales

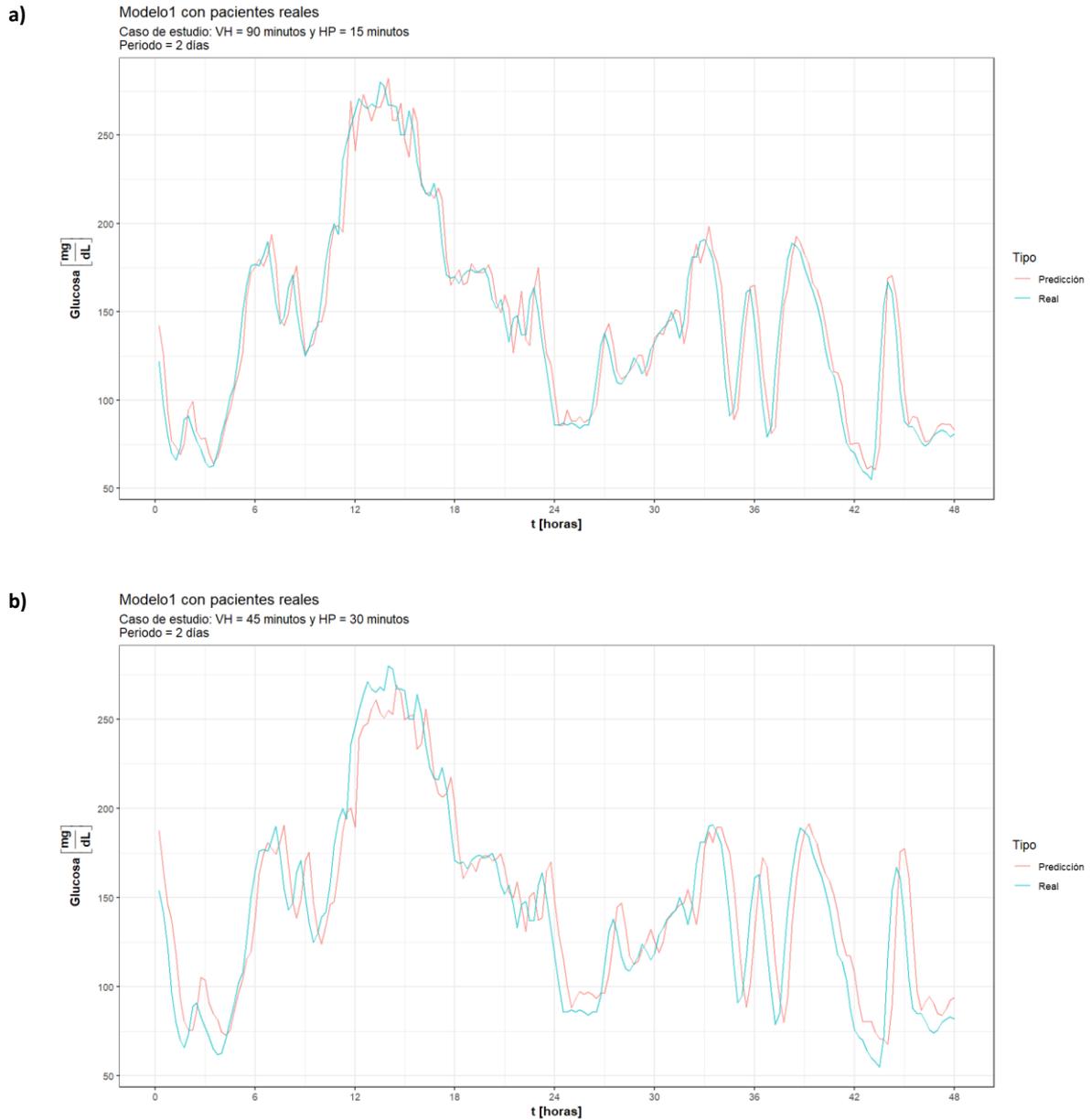
Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	20.52	86.59%	12.13%	98.72%	0.00%	1.28%	0.00%
VH: 45 minutos	19.06	86.81%	10.82%	97.63%	0.00%	2.37%	0.00%
VH: 60 minutos	19.51	84.83%	12.71%	97.54%	0.00%	2.47%	0.00%
VH: 90 minutos	20.31	89.26%	9.20%	98.46%	0.00%	1.55%	0.00%

Durante el análisis, se observó que al aumentar la ventana histórica (VH), el valor del RMSE para un horizonte de predicción (HP) de 15 minutos mejoró significativamente. Lo que implica que al incluir más datos históricos en el modelo permite una mejor predicción. Por otro lado, para un HP de 30 minutos, no se observó un impacto significativo en cuanto al valor del RMSE cuando se varía el tamaño de la VH.

Al analizar el porcentaje de las zonas de Clarke, se encontró que, tanto para un HP de 15 minutos como para un HP de 30 minutos, el porcentaje de las zonas clínicamente aceptadas se ubicó por encima del 97%. Sin embargo, se observó un aumento en el porcentaje de la zona B para un HP de 30 minutos en comparación con un HP de 15 minutos. Al ampliarse la zona B, puede estar representando un riesgo clínico mínimo. El mayor impacto se logró apreciar en la zona D, que llega a tener un máximo de 2.47% para un HP de 30 minutos y un máximo del 0.46% en un HP de 15 minutos.

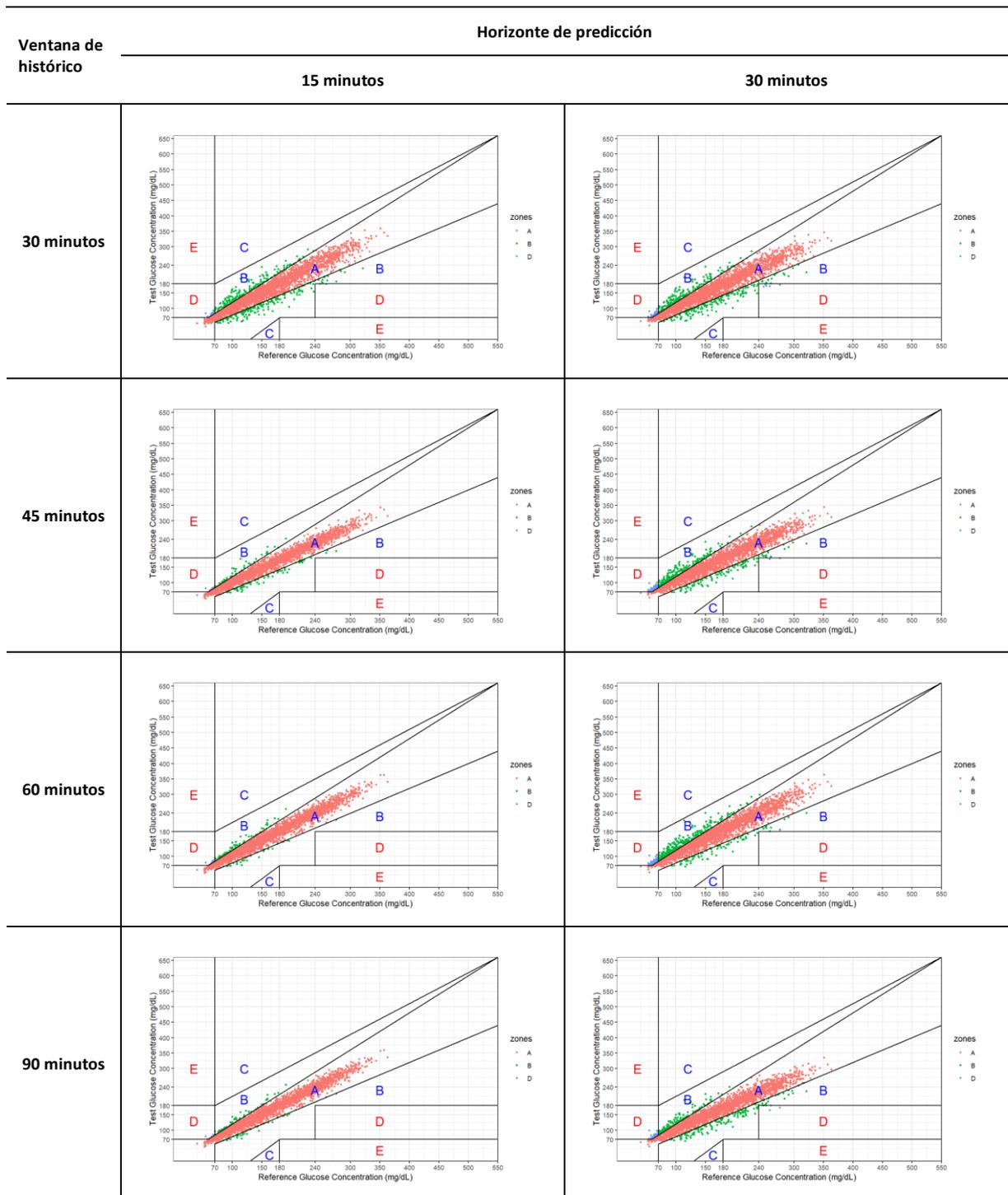
Se observó que a medida que se amplió la VH el RMSE para un HP de 15 fue mejorando significativamente. Sin embargo, para un HP de 30 minutos, la variación de la VH no tuvo un mayor impacto en el valor RMSE. Si analizamos el porcentaje de las zonas de Clarke, el porcentaje de las zonas clínicamente aceptado, tanto para un HP de 15 minutos como para un HP de 30 minutos se encuentra por encima del 97%, sin embargo, el porcentaje que representa la zona B aumenta para un HP de 30. En la Figura 24, se muestran las series de los dos casos de estudio.

Figura 24: Mejores respuestas del Modelo 1 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.



En la Figura 24, se muestra claramente que hay un mayor desfase entre la predicción y el valor real cuando se utiliza el modelo con un HP de 30 minutos en comparación con un HP de 15. En la Figura 25 se muestra la gráfica de Clarke para cada caso de estudio

Figura 25: Gráficas de Clarke para el modelo 1 – Pacientes Reales



5.2.1.2. Pacientes simulados

En el caso del modelo 1 empleando pacientes simulados, también se obtuvieron las métricas descritas previamente. Los resultados para el horizonte de predicción de 15 minutos se presentan en la Tabla 17, y para el horizonte de predicción de 30 minutos en la Tabla 18.

Tabla 17: Métricas del modelo 1 para un horizonte de predicción de 15 minutos en pacientes simulados

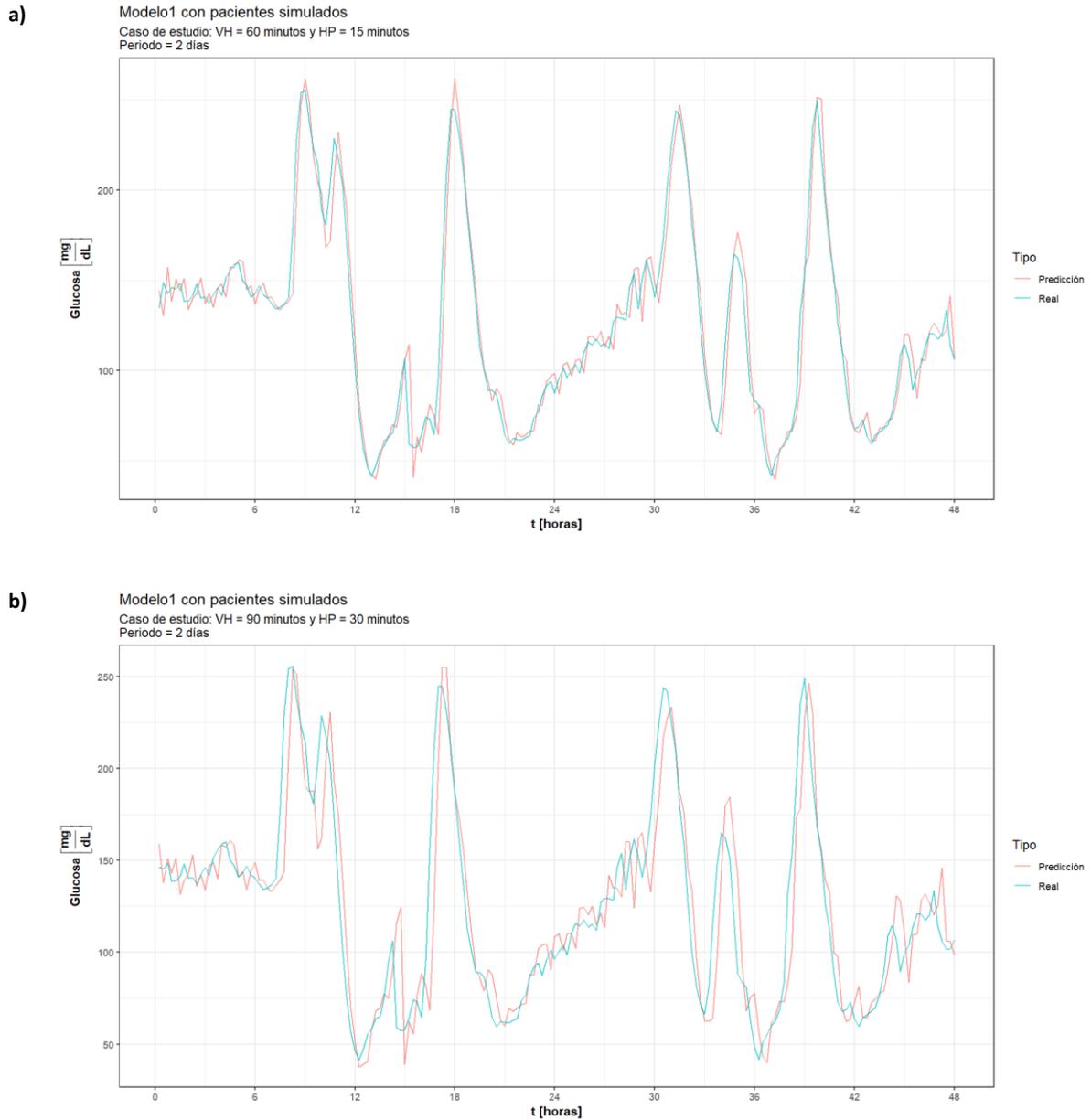
Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	12.74	94.22%	4.64%	98.86%	0.00%	1.14%	0.00%
VH: 45 minutos	12.52	94.59%	4.22%	98.81%	0.00%	1.19%	0.00%
VH: 60 minutos	11.39	95.27%	4.05%	99.32%	0.00%	0.68%	0.00%
VH: 90 minutos	11.70	95.51%	3.84%	99.35%	0.00%	0.65%	0.00%

Tabla 18: Métricas del modelo 1 para un horizonte de predicción de 30 minutos en pacientes simulados.

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	22.74	81.52%	17.06%	98.58%	0.00%	1.40%	0.02%
VH: 45 minutos	21.47	82.86%	14.30%	97.16%	0.04%	2.79%	0.02%
VH: 60 minutos	20.75	82.70%	14.53%	97.23%	0.00%	2.77%	0.00%
VH: 90 minutos	20.23	82.76%	14.65%	97.41%	0.00%	2.58%	0.02%

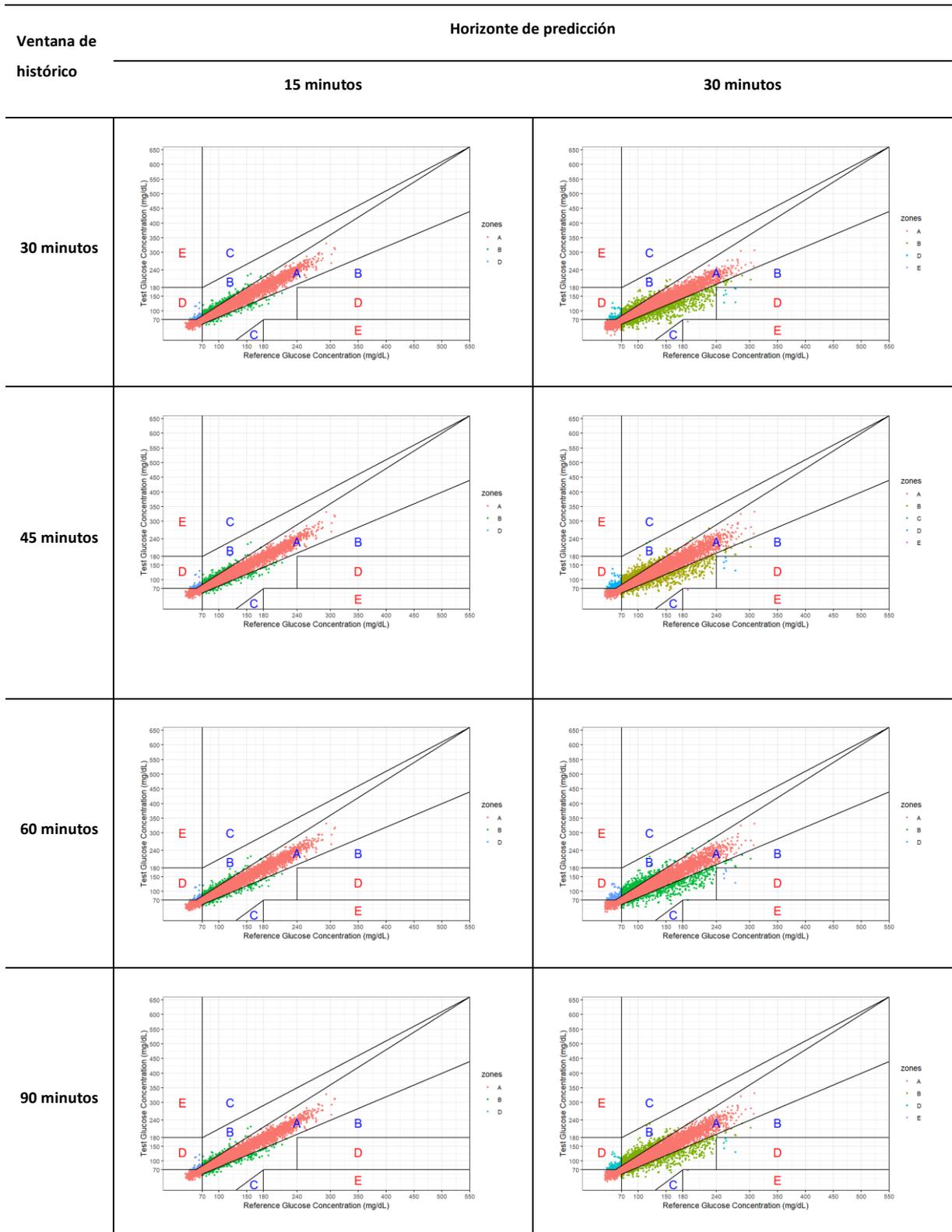
Entre los 8 casos de estudio del presente escenario, se observa que el valor del RMSE mejora para un horizonte de predicción de 15 minutos a medida que se incrementa la ventana histórica. Sin embargo, para contrastar este resultado, es importante considerar los valores que se encuentran en las zonas clínicamente aceptadas de las gráficas de Clarke. Se observa que el porcentaje de puntos en estas zonas es mayor al 98% para un horizonte de predicción de 15 minutos, en donde la mayor influencia se da por la zona A. En el caso del horizonte de predicción de 30 minutos, el porcentaje de puntos en las zonas clínicamente aceptadas sigue siendo superior al 97%, sin embargo, se observa una mayor influencia de la zona B. Las gráficas de las series temporales se las observa en la Figura 26 y las gráficas de Clarke en la Figura 27.

Figura 26: Mejores respuestas del Modelo 1 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.



En la Figura 26 se observa como en el caso del modelo 1 para pacientes reales, un desfase entre la curva real y la curva de predicción para un horizonte de predicción de 30 minutos.

Figura 27: Gráficas de Clarke para el modelo 1 – Pacientes simulados



5.2.2. Modelo 2: Modelo de 1 capa LSTM

Se sigue con las pruebas en una red de 1 capa LSTM. Para ello generamos un modelo secuencial, y le agregamos una capa LSTM de 64 neuronas. Esta capa está diseñada para aprender la dependencia y patrones en los datos secuenciales. La entrada de la capa inicial se encuentra en función del tamaño de la ventana en la cual se encuentra generado el modelo. A continuación, se lo conecta a una densa de salida con la función de activación lineal. La capa se encarga de proporcionar la predicción final a partir de las representaciones aprendidas por la capa LSTM. La creación del modelo fue realizada como en se observa en la Figura 28.

Figura 28: Código en Python empleado para generar el Modelo 2.

```
# Función para crear el modelo LSTM
def crear_modelo():
    modelo = Sequential()
    modelo.add(LSTM(64, input_shape=(ventana, 1)))
    modelo.add(Dense(1, activation='linear'))

    return modelo

# Crear el modelo
modelo = crear_modelo()

# Compilar el modelo
modelo.compile(optimizer='adam', loss='mse', metrics=["accuracy"])
```

5.2.2.1. Pacientes reales

En el caso del modelo 2 empleando pacientes reales, las métricas descritas previamente para el horizonte de predicción de 15 minutos se presentan en la Tabla 19, mientras que los resultados para el horizonte de predicción de 30 minutos se muestran en la Tabla 20.

Tabla 19: Métricas del modelo 2 para un horizonte de predicción de 15 minutos en pacientes reales

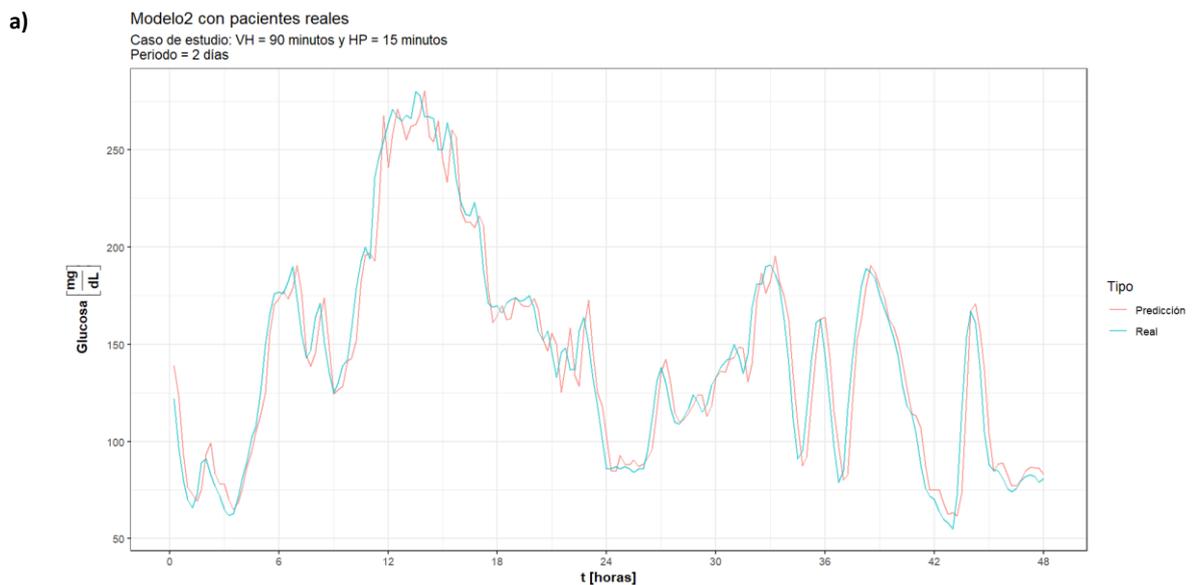
Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	19.19	88.67%	10.31%	98.98%	0.00%	1.02%	0.00%
VH: 45 minutos	12.68	96.29%	3.25%	99.54%	0.00%	0.46%	0.00%
VH: 60 minutos	11.80	96.55%	2.83%	99.38%	0.00%	0.62%	0.00%
VH: 90 minutos	11.78	97.24%	2.60%	99.84%	0.00%	0.16%	0.00%

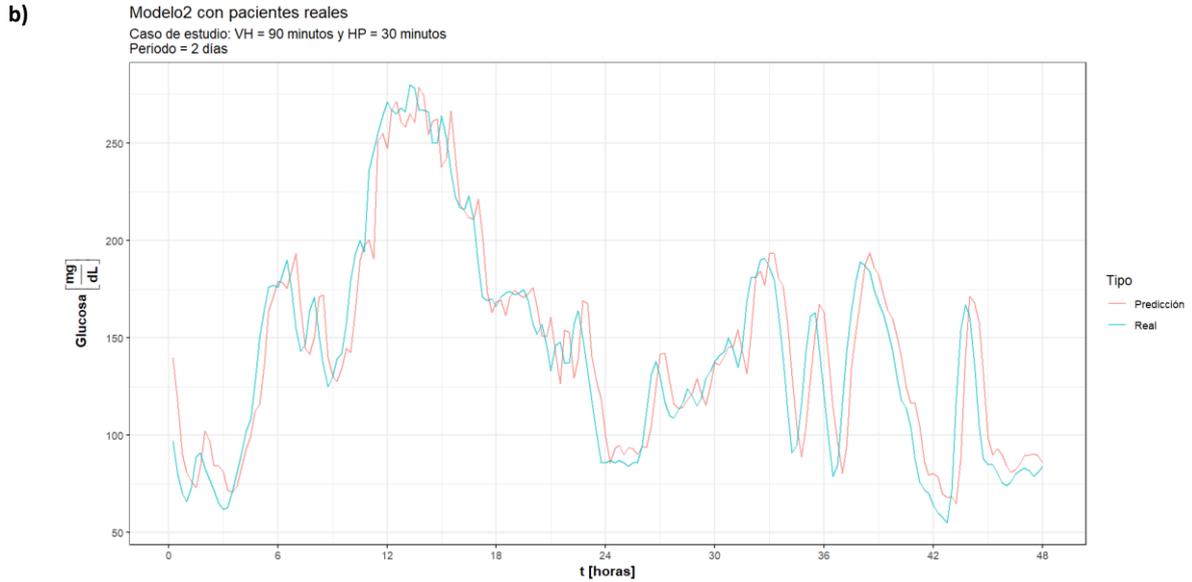
Tabla 20: Métricas del modelo 2 para un horizonte de predicción de 30 minutos en pacientes reales

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	20.39	86.59%	11.41%	98.00%	0.00%	2.01%	0.00%
VH: 45 minutos	19.36	86.09%	11.74%	97.83%	0.00%	2.17%	0.00%
VH: 60 minutos	18.92	85.94%	11.36%	97.30%	0.00%	2.70%	0.00%
VH: 90 minutos	18.57	88.40%	9.95%	98.35%	0.00%	1.65%	0.00%

En el caso de una red LSTM de 1 capa y una red RNN de 1 capa, no se obtuvo una diferencia significativa en el valor del RMSE. Al observar los porcentajes de las zonas de Clarke, se observa que la zona A tiene una ligera influencia mayor que en el caso del Modelo 1. Las gráficas de las series temporales se las observa en la Figura 29 y las gráficas de Clarke en la Figura 30.

Figura 29: Mejores respuestas del Modelo 2 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.





La respuesta obtenida al comportamiento de la serie con el modelo 2 en pacientes reales, Figura 29, no difiere mucho al comportamiento del modelo 1 empleado en pacientes reales, observado previamente en la Figura 24.

5.2.2.2. Pacientes simulados

En el caso del modelo 2 empleando pacientes simulados, las métricas descritas previamente para el horizonte de predicción de 15 minutos se presentan en la Tabla 21, mientras que los resultados para el horizonte de predicción de 30 minutos se muestran en la Tabla 22 .

Tabla 21: Métricas del modelo 2 para un horizonte de predicción de 15 minutos en pacientes simulados

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	12.50	94.96%	4.32%	99.28%	0.00%	0.72%	0.00%
VH: 45 minutos	12.37	94.66%	4.41%	99,07 %	0.00%	1.05%	0.00%
VH: 60 minutos	11.57	94.94%	4.33%	99.27%	0.00%	0.74%	0.00%
VH: 90 minutos	11.33	95.76%	3.77%	99.53%	0.00%	0.47%	0.00%

Figura 30: Gráficas de Clarke para el modelo 2 – Pacientes Reales

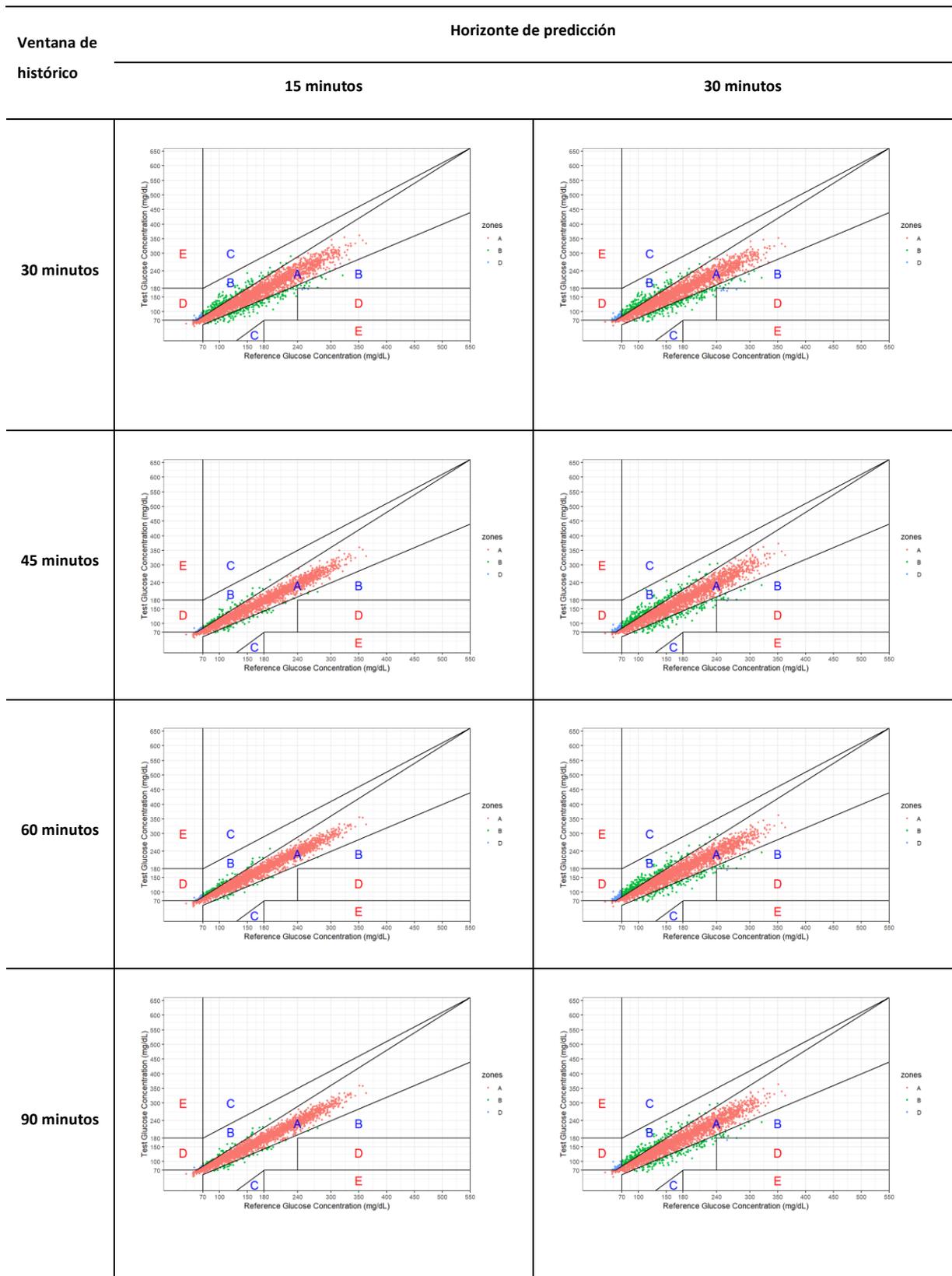
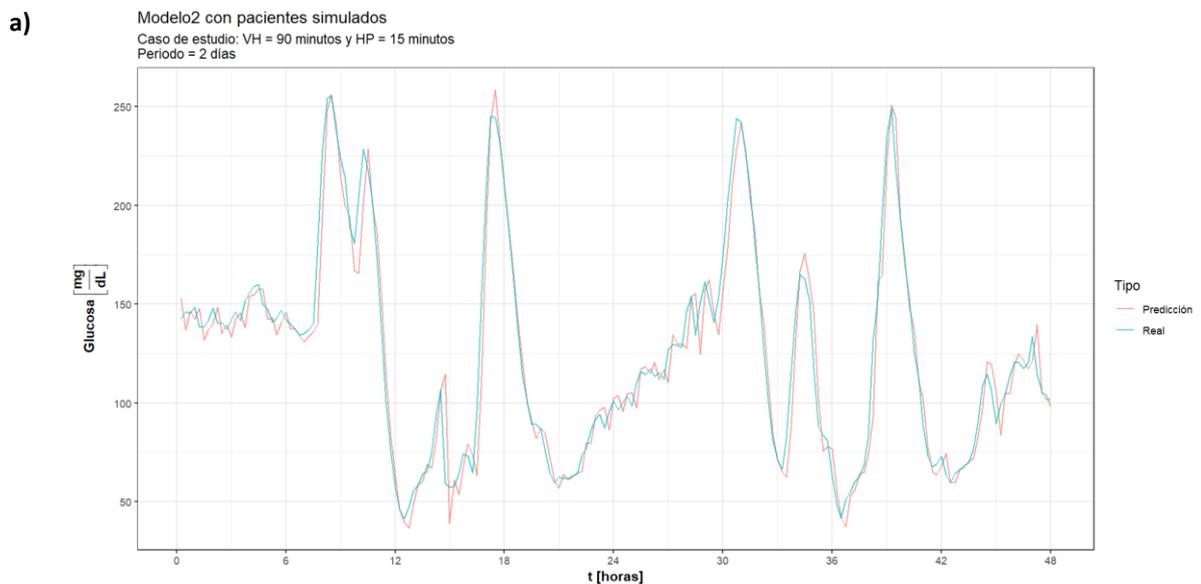


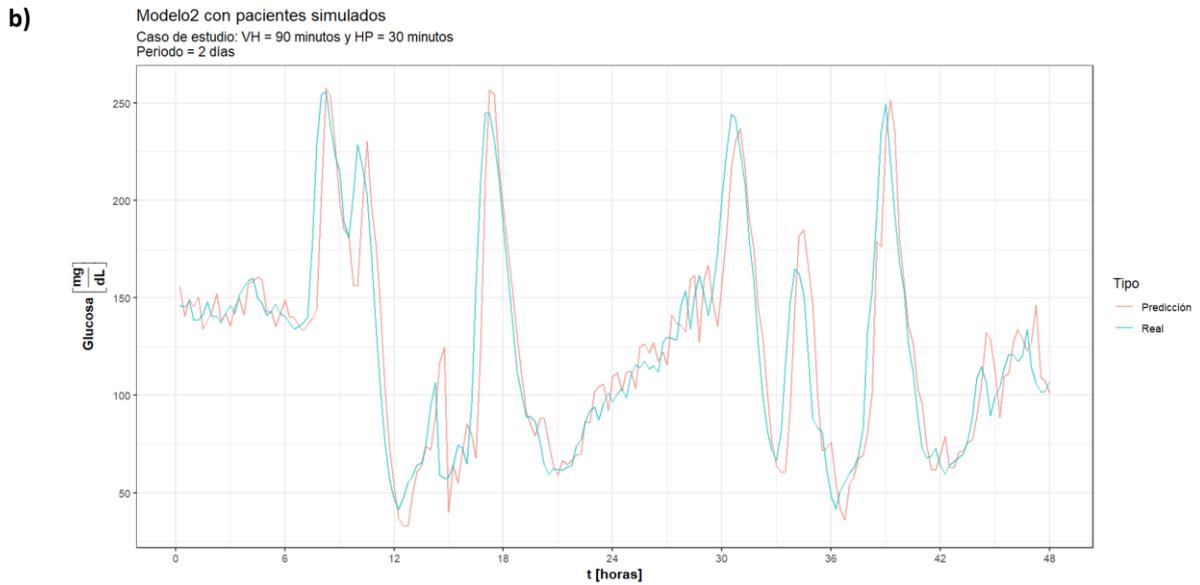
Tabla 22: Métricas del modelo 2 para un horizonte de predicción de 30 minutos en pacientes simulados

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	21.85	82.82%	14.71%	97.53%	0.00%	2.45%	0.02%
VH: 45 minutos	21.29	84.24%	13.90%	98.14%	0.00%	2.15%	0.01%
VH: 60 minutos	20.40	84.14%	13.72%	97.86%	0.03%	2.09%	0.02%
VH: 90 minutos	20.10	83.07%	14.66%	97.73%	0.04%	2.26%	0.01%

Al emplear el modelo 2 en el escenario de pacientes simulado, las zonas clínicamente aceptadas abordan desde el 99% para un HP de 15 minutos y desde un 97% para un HP de 30 minutos. El porcentaje de la zona A para el modelo 2 en el escenario descrito se encuentra por encima del 94% para un HP de 15 a diferencia del 82% para un HP de 30. Las gráficas de las series temporales se las observa en la Figura 31 y las gráficas de Clarke en la Figura 32.

Figura 31: Mejores respuestas del Modelo 2 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.





En la Figura 31, el modelo 2 entrenado con pacientes simulados muestra una mayor precisión en la predicción de la serie en comparación a la predicción obtenida con el modelo 1, Figura 26, tanto para el horizonte de 15 minutos como para el horizonte de 30 minutos. Si comparamos las gráficas obtenidas en ambos modelos, la predicción del modelo 2 se ajusta de mejor manera a la serie real.

5.2.3. Modelo 3: Modelo de 2 capas LSTM

Continuamos el procedimiento con una red LSTM con mayor complejidad que la red anterior. Para ello generamos un modelo secuencial, y le agregamos una capa LSTM de 128 neuronas, generado para procesar secuencias de datos en paralelo. Esta capa devuelve una secuencia completa en lugar de solo la última salida, esto se lo genera a partir de la habilitación del parámetro `return_sequences = True`. A continuación, se le agrega una segunda capa LSTM de 128 neuronas. Esta capa devuelve la última salida de la secuencia. A la salida se le añade una capa completamente conectada de 64 neuronas con una función de activación `ReLU`. A la salida del modelo se le añade la capa de una única neurona con una función de activación lineal. La creación del modelo fue realizada como en se observa en la Figura 33.

Figura 32: Gráficas de Clarke para el modelo 2 – Pacientes Simulados

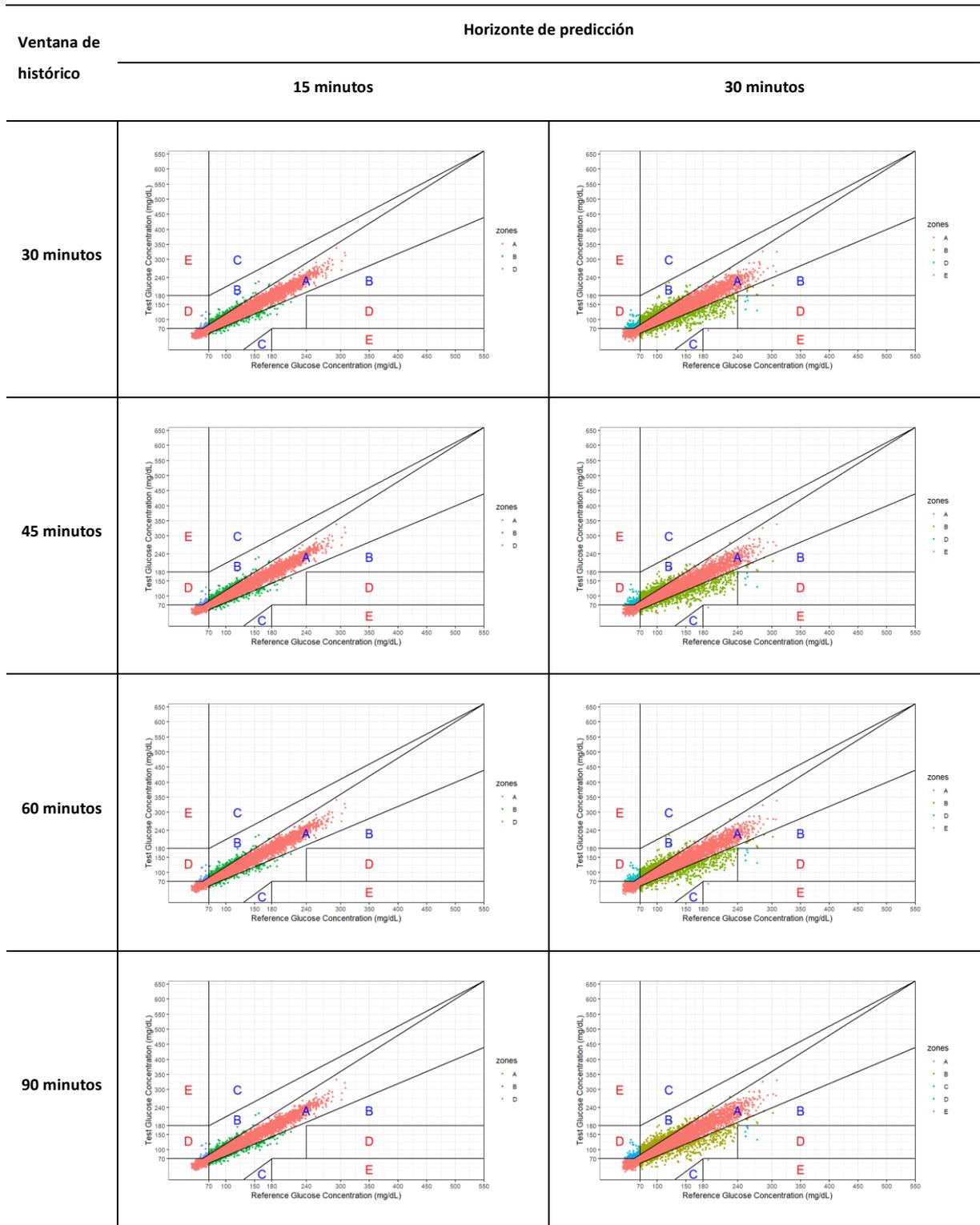


Figura 33: Código en Python empleado para generar el Modelo 3

```
# Función para crear el modelo LSTM 2
def crear_modelo():
    modelo = Sequential()
    modelo.add(LSTM(128, input_shape=(ventana, 1), return_sequences=True))
    modelo.add(LSTM(128))
    modelo.add(Dense(64, activation='relu'))
    modelo.add(Dense(1, activation='linear'))

    return modelo

# Crear el modelo
modelo = crear_modelo()

# Compilar el modelo
modelo.compile(optimizer='adam', loss='mse', metrics=["accuracy"])
```

5.2.3.1. Pacientes reales

En el caso del modelo 3 empleando pacientes simulados, las métricas descritas previamente para el horizonte de predicción de 15 minutos se presentan en la Tabla 23, mientras que los resultados para el horizonte de predicción de 30 minutos se muestran en la Tabla 24.

Tabla 23: Métricas del modelo 3 para un horizonte de predicción de 15 minutos en pacientes reales.

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	11.44	97,60 %	2,27 %	99,87 %	0,00 %	0,13 %	0,00 %
VH: 45 minutos	15.06	93,17 %	3,38 %	96,55 %	0,00 %	3,45 %	0,00 %
VH: 60 minutos	11.17	97,73 %	2,07 %	99,80 %	0,00 %	0,20 %	0,00 %
VH: 90 minutos	10.64	97,56 %	1,74 %	99,30 %	0,00 %	0,69 %	0,00 %

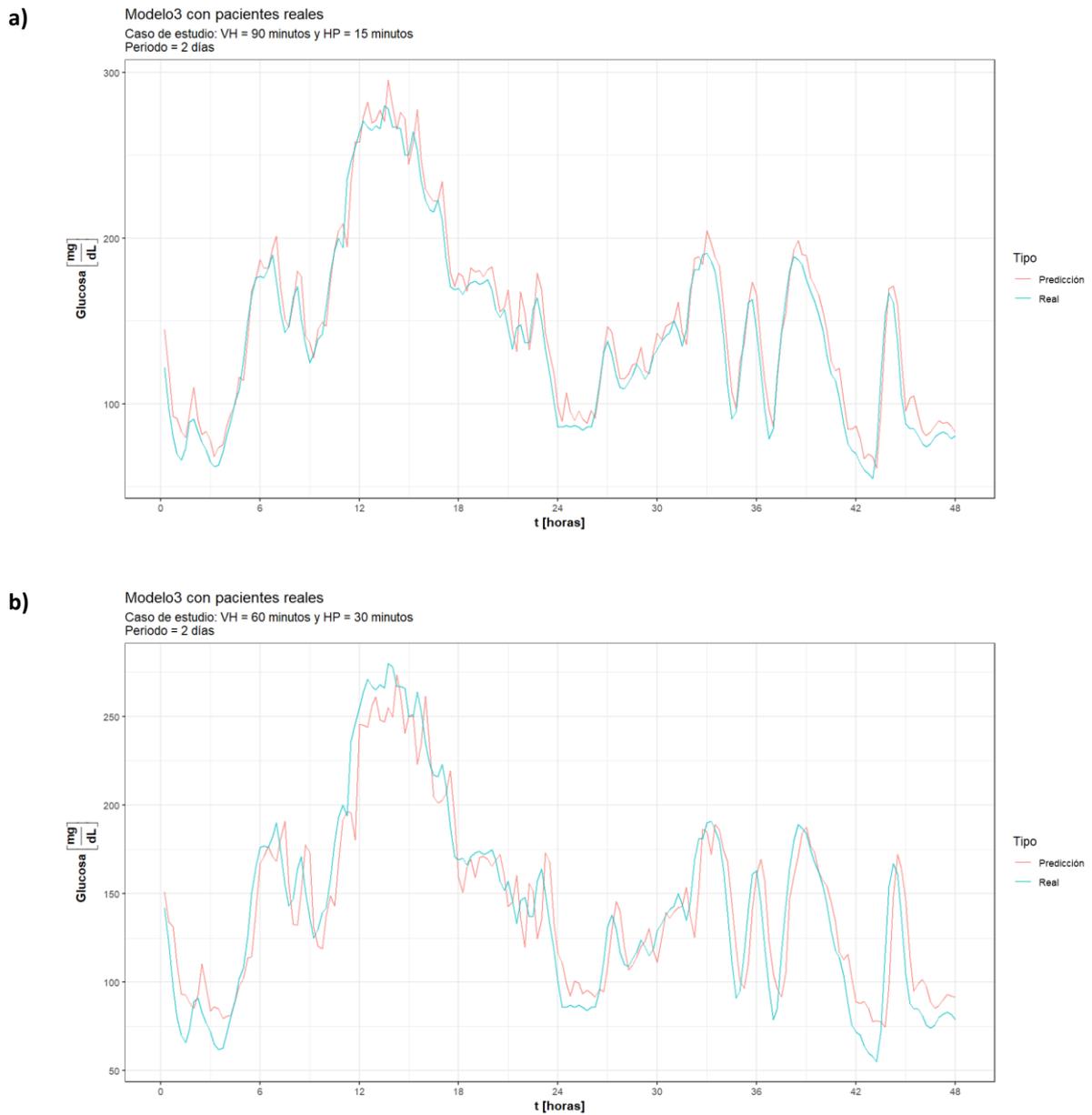
Tabla 24: Métricas del modelo 3 para un horizonte de predicción de 30 minutos en pacientes reales

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	18.82	90,40 %	8,91 %	99,31 %	0,00 %	0,69 %	0,00 %
VH: 45 minutos	17.45	89,44 %	8,65 %	98,09 %	0,00 %	1,91 %	0,00 %
VH: 60 minutos	18.47	87,13 %	8,92 %	96,05 %	0,00 %	3,95 %	0,00 %
VH: 90 minutos	18.76	86,42 %	9,56 %	95,98 %	0,00 %	4,02 %	0,00 %

Al aumentarle la complejidad al modelo mediante el uso de dos capas LSTM consecutivas, se logró una mejora significativa en la métrica de evaluación RMSE y en los porcentajes de las zonas de Clarke para el horizonte de predicción de 15 minutos. Por otra parte, para el

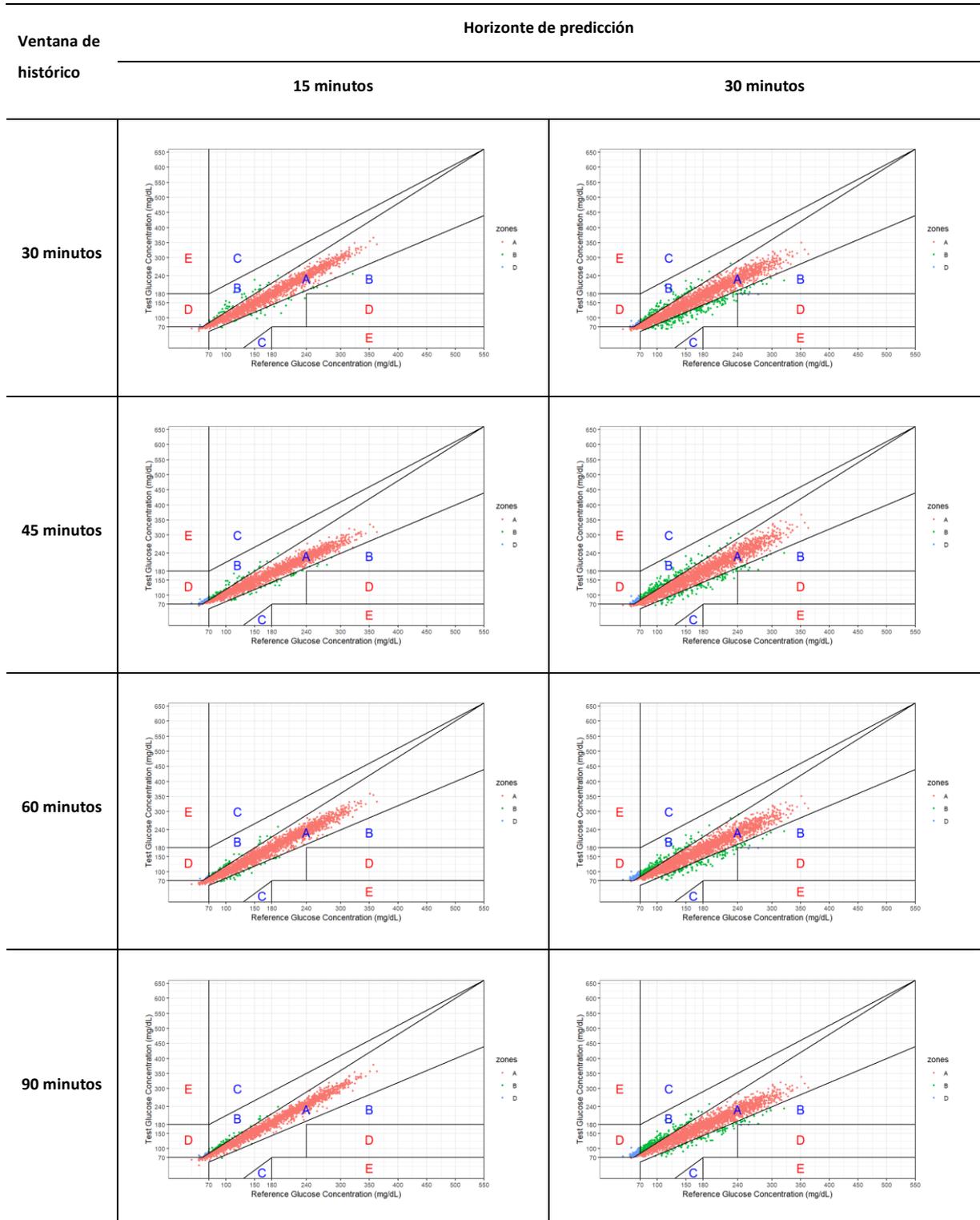
horizonte de predicción de 30 minutos, el porcentaje del error en la zona D aumentó considerablemente para las ventanas históricas de 60 y 90 minutos, llegando a alcanzar un 4%. Las gráficas de las series temporales se las observa en la Figura 34 y las gráficas de Clarke en la Figura 35.

Figura 34: Mejores respuestas del Modelo 3 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.



Con este modelo, se observó en la Figura 34 que, para un horizonte de predicción de 15 minutos, la predicción tiende a tener una ligera sobreestimación, mientras que para un horizonte de predicción de 30 minutos, la predicción tiende a tener una ligera subestimación.

Figura 35: Gráficas de Clarke para el modelo 3 – Pacientes reales



5.2.3.2. Pacientes simulados

En el caso del modelo 3 empleando pacientes simulados, las métricas descritas previamente para el horizonte de predicción de 15 minutos se presentan en la Tabla 25, mientras que los resultados para el horizonte de predicción de 30 minutos se muestran en la Tabla 26.

Tabla 25: Métricas del modelo 3 para un horizonte de predicción de 15 minutos en pacientes simulados

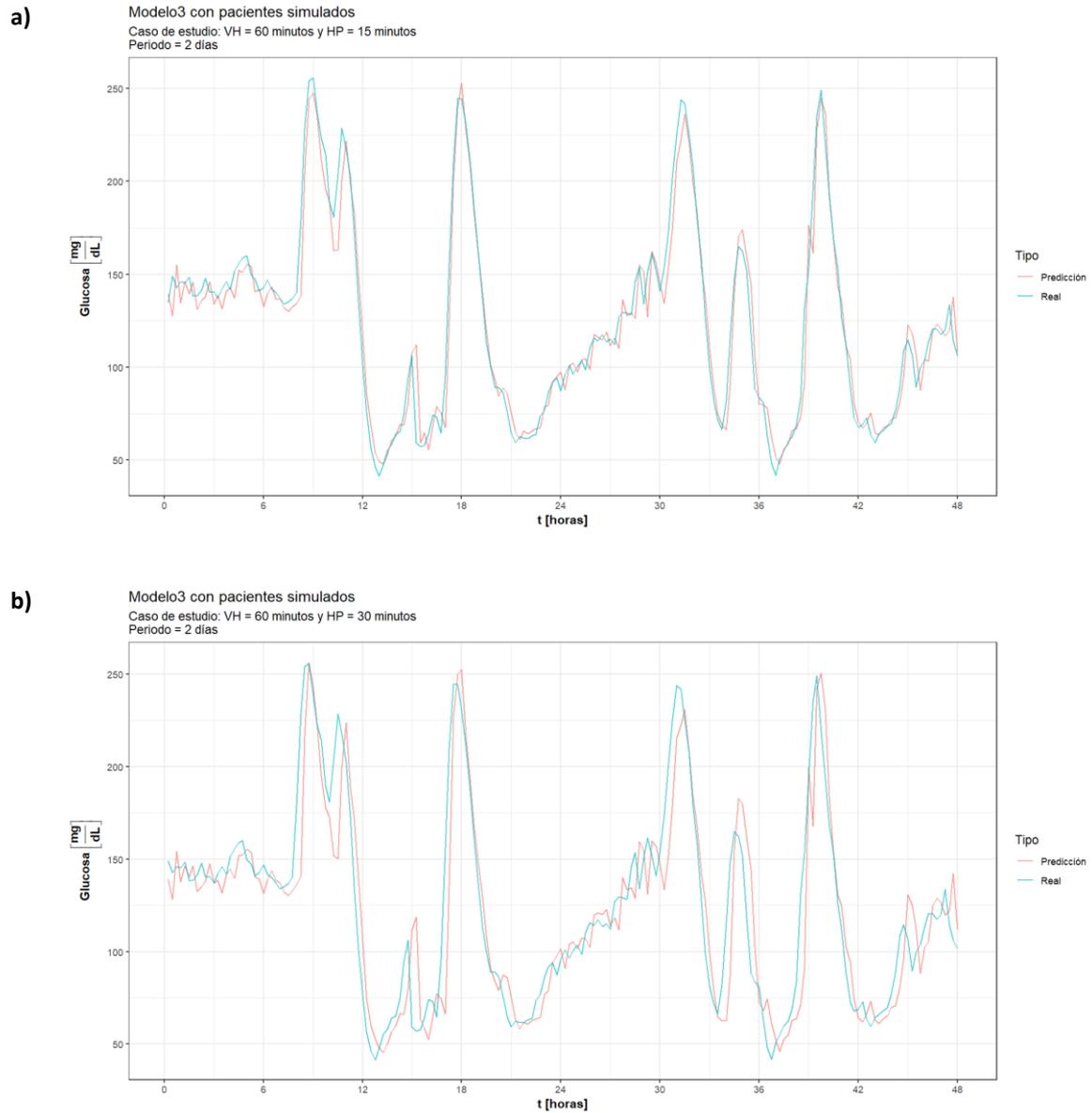
Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	12.24	95.66%	3.47%	99.13%	0.00%	0.88%	0.00%
VH: 45 minutos	11.17	95.04%	3.94%	98.98%	0.00%	1.02%	0.00%
VH: 60 minutos	11.14	96.06%	3.22%	99.28%	0.00%	0.72%	0.00%
VH: 90 minutos	11.32	96.20%	3.28%	99.48%	0.00%	0.53%	0.00%

Tabla 26: Métricas del modelo 3 para un horizonte de predicción de 30 minutos en pacientes simulados

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	21.19	83.39%	14.19%	97.58%	0.00%	2.40%	0.02%
VH: 45 minutos	20.37	83.62%	13.90%	97.52%	0.00%	2.47%	0.02%
VH: 60 minutos	19.75	85.01%	13.36%	98.37%	0.00%	1.61%	0.01%
VH: 90 minutos	19.98	85.46%	13.03%	98.49%	0.00%	1.49%	0.01%

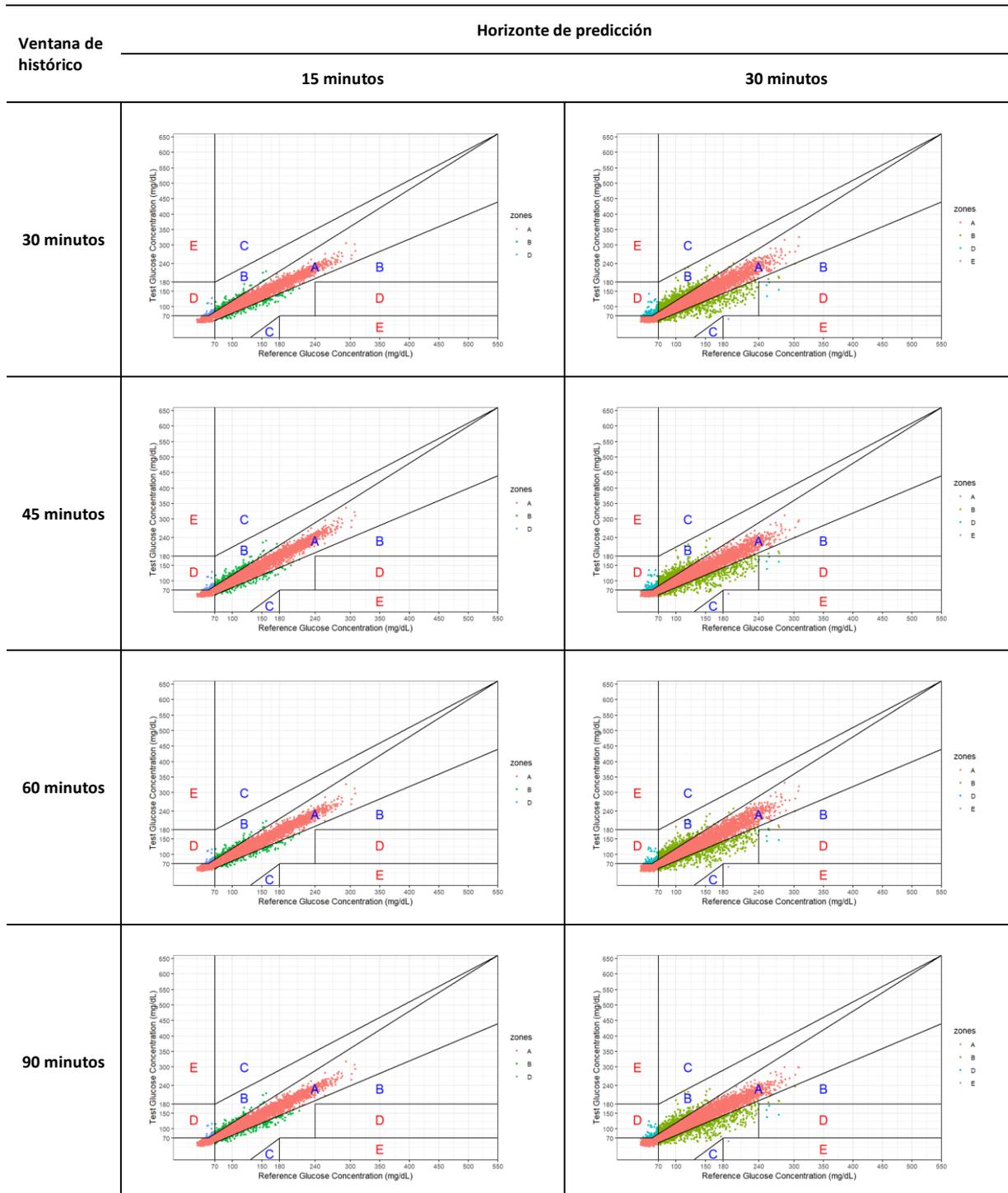
En concordancia con los resultados obtenidos en el escenario de los pacientes reales, en las Tabla 25 y Tabla 26 se observó una ligera mejoría en las métricas al aplicar este modelo con los pacientes simulados. Sin embargo, se distinguió un ligero aumento en el porcentaje del erro en la zona E para el horizonte de predicción de 30 minutos, el mismo que llegó a un valor mínimo de 0.02%. Por otra parte, para un horizonte de predicción de 15 minutos, el porcentaje de las zonas de Clarke se mantuvo sin mostrar cambios significativos. Las gráficas de las series temporales se las observa en la Figura 36 y las gráficas de Clarke en la Figura 37.

Figura 36: Mejores respuestas del Modelo 3 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.



En la Figura 36, se observó que el modelo se ajusta de mejor forma a la curva del valor real cuando se emplea un horizonte de predicción de 15 minutos. Por otra parte, el modelo tiende a generar una subestimación de los valores en los cambios de tendencia. A pesar de estos, el modelo 2 logra tener un mejor ajuste con un horizonte de predicción de 30 minutos.

Figura 37: Gráficas de Clarke para el modelo 3 – Pacientes simulados



5.2.4. Modelo 4: Modelo de 1 capa GRU

El siguiente modelo entrenado y testeado es un modelo GRU Simple. El modelo comienza con una capa GRU, este es una unidad recurrente basada en compuertas con 64 neuronas. Esta capa GRU se trata de una variante de las redes LSTM. La capa GRU recibe como entrada una secuencia de datos con la forma (ventana, 1). En este caso la “ventana” representa la longitud de la ventana deslizante con la que se esté entrenando el modelo. La capa GRU procesa la secuencia de entrada y captura la relación temporal y de dependencia a largo plazo. A continuación, se agrega la capa de salida con una sola neurona. Esta capa tiene una función de activación del tipo lineal. En esta capa de salida se representa la predicción realizada por el modelo. La creación del modelo fue realizada como en se observa en la Figura 38.

Figura 38: Código en Python empleado para generar el Modelo 4.

```
# Función para crear el modelo GRU
def crear_modelo():
    modelo = Sequential()
    modelo.add(GRU(64, input_shape=(ventana, 1)))
    modelo.add(Dense(1, activation='linear'))

    return modelo

# Crear el modelo
modelo = crear_modelo()

# Compilar el modelo
modelo.compile(optimizer='adam', loss='mse', metrics=["accuracy"])
```

5.2.4.1. Pacientes reales

En el caso del modelo 4 empleando pacientes reales, las métricas descritas previamente para el horizonte de predicción de 15 minutos se presentan en la Tabla 27, mientras que los resultados para el horizonte de predicción de 30 minutos se muestran en la Tabla 28.

Tabla 27: Métricas del modelo 4 para un horizonte de predicción de 15 minutos en pacientes reales

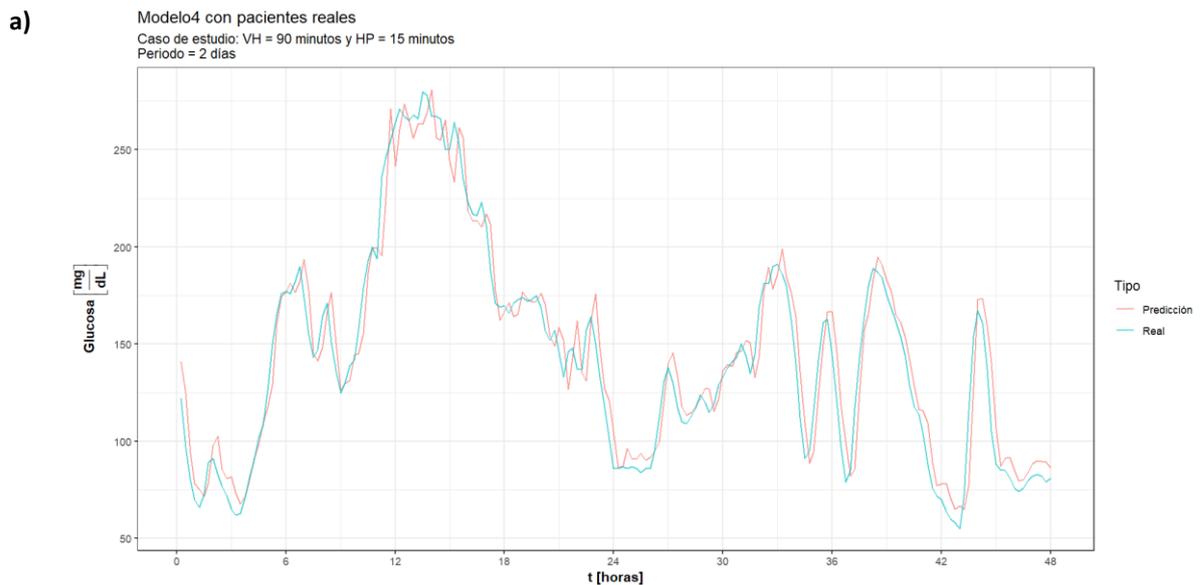
Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	19.29	87.62%	10.93%	98.55%	0.00%	1.44%	0.00%
VH: 45 minutos	13.03	94.84%	3.91%	98.75%	0.00%	1.25%	0.00%
VH: 60 minutos	11.91	97.53%	2.37%	99.90%	0.00%	0.10%	0.00%
VH: 90 minutos	11.53	97.24%	2.27%	99.51%	0.00%	0.49%	0.00%

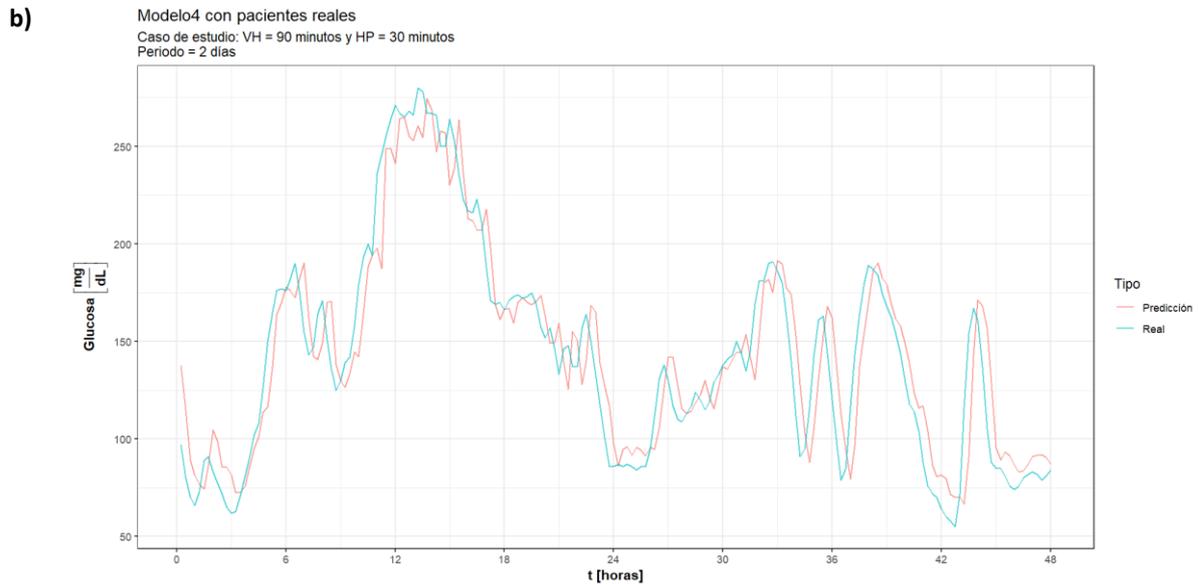
Tabla 28: Métricas del modelo 4 para un horizonte de predicción de 30 minutos en pacientes reales

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	20.34	86.13%	11.77%	97.90%	0.00%	2.10%	0.00%
VH: 45 minutos	19.17	88.75%	9.67%	98.42%	0.00%	1.58%	0.00%
VH: 60 minutos	18.92	86.01%	11.32%	97.33%	0.00%	2.67%	0.00%
VH: 90 minutos	18.46	88.50%	9.46%	97.96%	0.00%	2.04%	0.00%

Para el modelo 4 aplicado en el escenario de pacientes reales, se observa que, para ambos horizontes de predicción, el porcentaje de la zona clínicamente aceptada se ve menos influenciado por los valores en la zona B. El error en la zona D es mínimo para un horizonte de predicción de 15 minutos con una ventana histórica de 60 y 90 minutos. Por otra parte, para el horizonte de predicción de 30 minutos, el error presente en la zona D varía entre el 1% y el 2.67%. Las gráficas de las series temporales se las observa en la Figura 39 y las gráficas de Clarke en la Figura 40.

Figura 39: Mejores respuestas del Modelo 4 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.





En el caso específico de emplear una capa GRU en el modelo, la predicción sigue la tendencia del valor real para un horizonte de predicción de 15 minutos. Sin embargo, al aumentar el horizonte de predicción a 30 minutos, se observa una subestimación en la predicción y un desfase en la predicción, lo que genera una discrepancia entre el valor de la predicción y el valor real, tal como se observa Figura 39.

5.2.4.2. Pacientes simulados

En el caso del modelo 4 empleando pacientes simulados, las métricas descritas previamente para el horizonte de predicción de 15 minutos se presentan en la Tabla 29, mientras que los resultados para el horizonte de predicción de 30 minutos se muestran en la Tabla 30.

Tabla 29: Métricas del modelo 4 para un horizonte de predicción de 15 minutos en pacientes simulados

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	11.92	93.86%	4.95%	98.81%	0.00%	1.19%	0.00%
VH: 45 minutos	11.44	94.80%	4.18%	98.98%	0.00%	1.02%	0.00%
VH: 60 minutos	11.40	95.29%	4.05%	99.34%	0.00%	0.67%	0.00%
VH: 90 minutos	11.27	95.43%	3.98%	99.41%	0.00%	0.60%	0.00%

Figura 40: Gráficas de Clarke para el modelo 4 – Pacientes reales

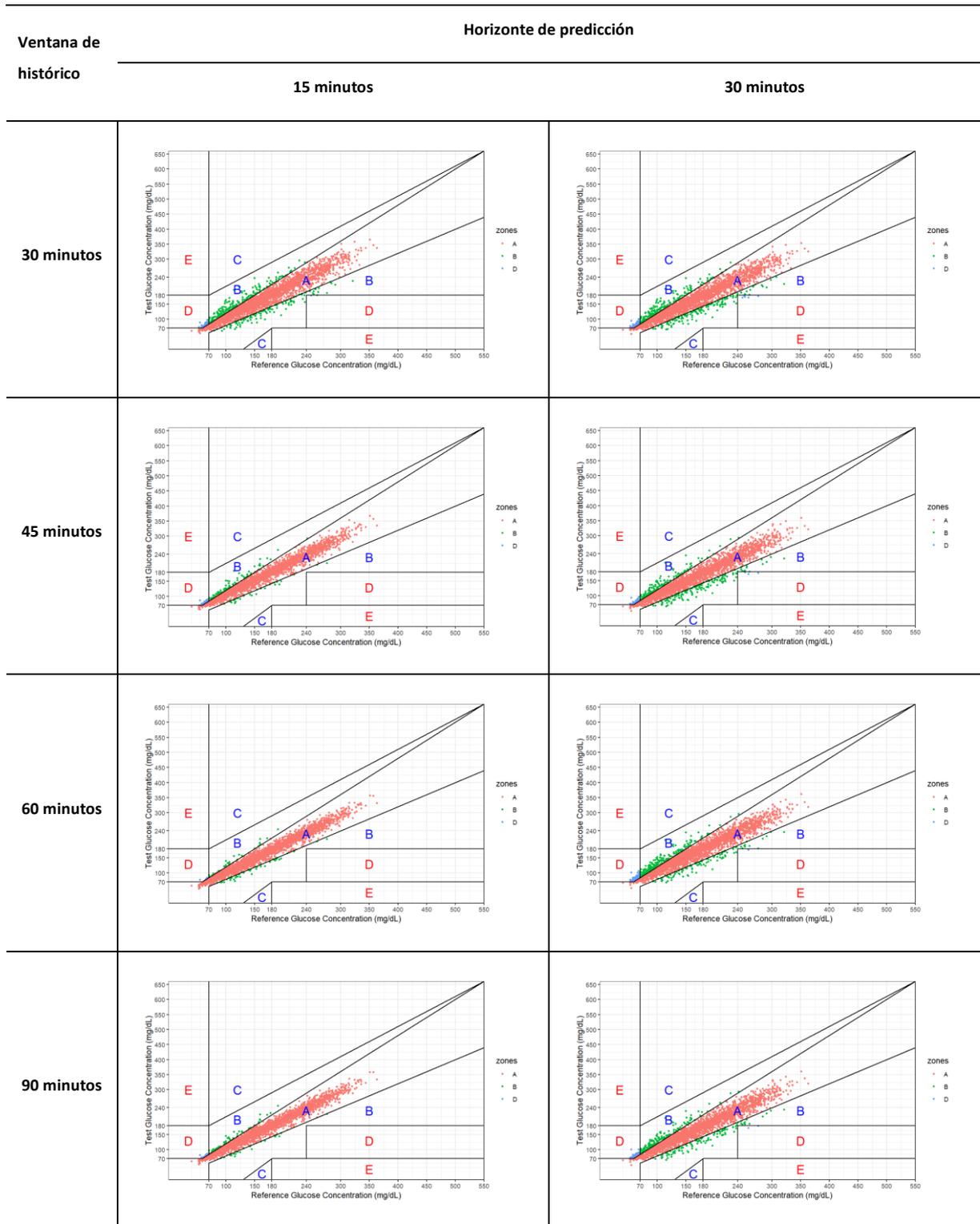
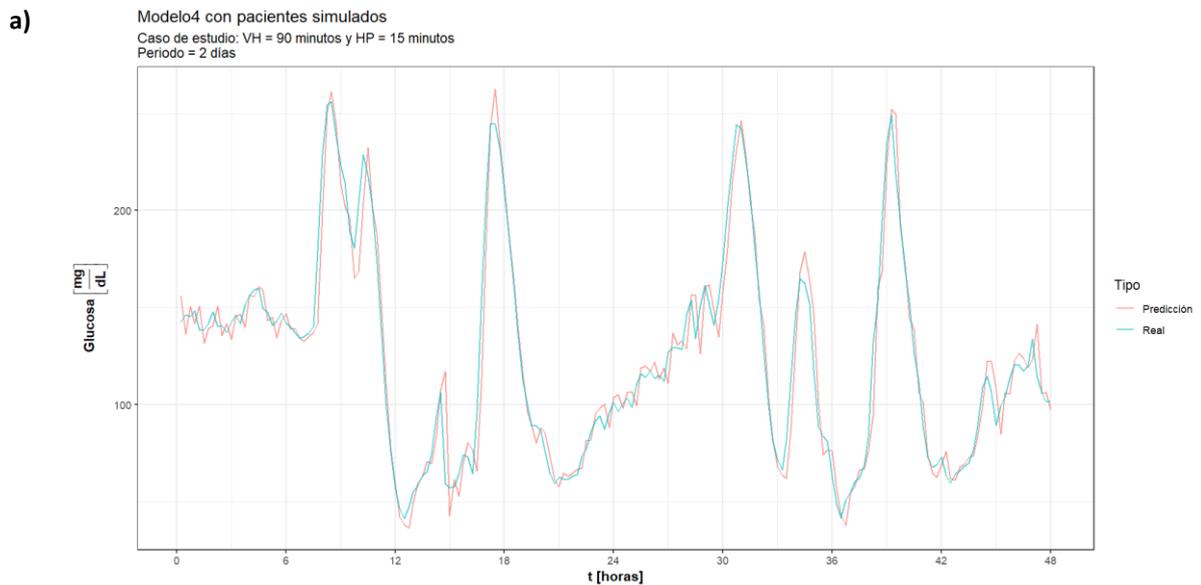


Tabla 30: Métricas del modelo 4 para un horizonte de predicción de 30 minutos en pacientes simulados

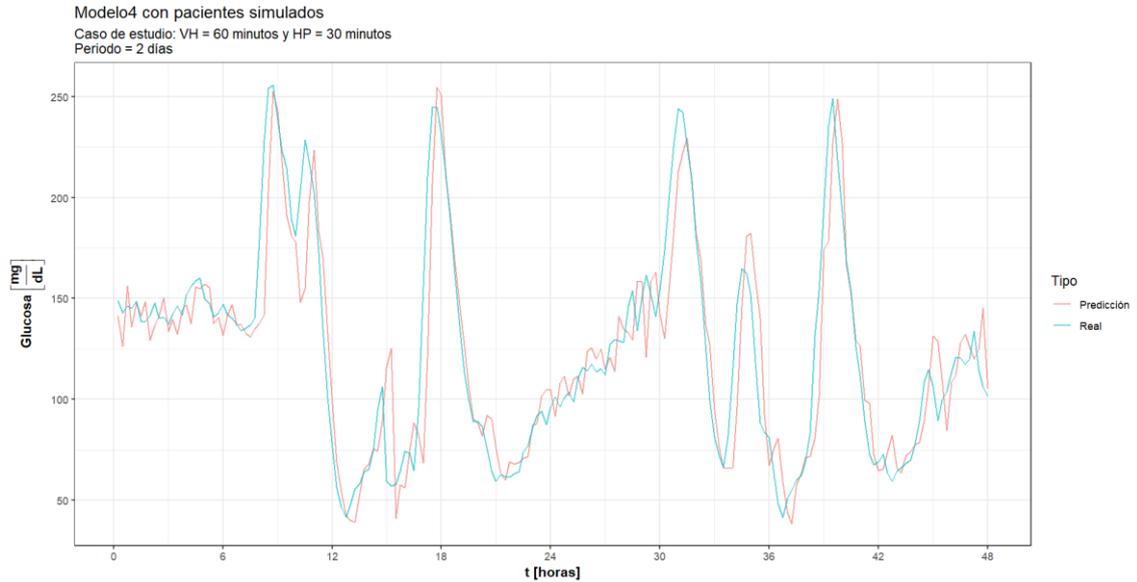
Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	20.66	79.89%	16.52%	96.41%	0.00%	3.59%	0.00%
VH: 45 minutos	21.01	83.79%	14.23%	98.02%	0.00%	1.96%	0.02%
VH: 60 minutos	20.19	83.14%	14.20%	97.34%	2.65%	0.02%	0.02%
VH: 90 minutos	20.41	83.56%	13.96%	97.52%	0.00%	2.46%	0.02%

Tanto el valor RMSE obtenido como el porcentaje de las zonas de Clarke para los diferentes casos de estudio no distan mucho del resultado obtenido con el modelo 2, modelo constituido por 1 capa LSTM. Las gráficas de las series temporales se las observa en la Figura 41 y las gráficas de Clarke en la Figura 42.

Figura 41: Mejores respuestas del Modelo 4 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.



b)



En el caso de los pacientes simulados con un horizonte de predicción de 15 minutos, se observa que la predicción mantiene la tendencia y sigue los valores de la glucosa. Sin embargo, para un horizonte de predicción de 30 minutos, se observa una tendencia a subestimar la predicción de los valores de glucosa.

5.2.5. Modelo 5: Modelo de 2 capas GRU

El siguiente modelo entrenado es un modelo GRU con mayor complejidad. En este modelo se comienza con una capa GRU, una unidad recurrente basada en compuerta de 128 neuronas. Esta capa GRU recibe una secuencia de datos con la forma (ventana, 1), en donde la dimensión de ventana es la longitud de la ventana deslizante. Esta capa tiene configurado el parámetro “return_sequences=True”, lo que implica que a la salida devuelve las secuencias completas. Esta salida se la conecta a una segunda capa GRU con 128 neuronas. Esta segunda capa no devuelve la secuencia. Se le conecta a una capa densa completamente conectada con 64 neuronas, esta capa posee la función de activación “ReLU”. Finalmente, se le añade una capa densa con una activación “Lineal”. A la salida de esta capa se encuentra el valor de la predicción. La creación del modelo fue realizada como en se observa en la Figura 43.

Figura 42: Gráficas de Clarke para el modelo 4 – Pacientes simulados

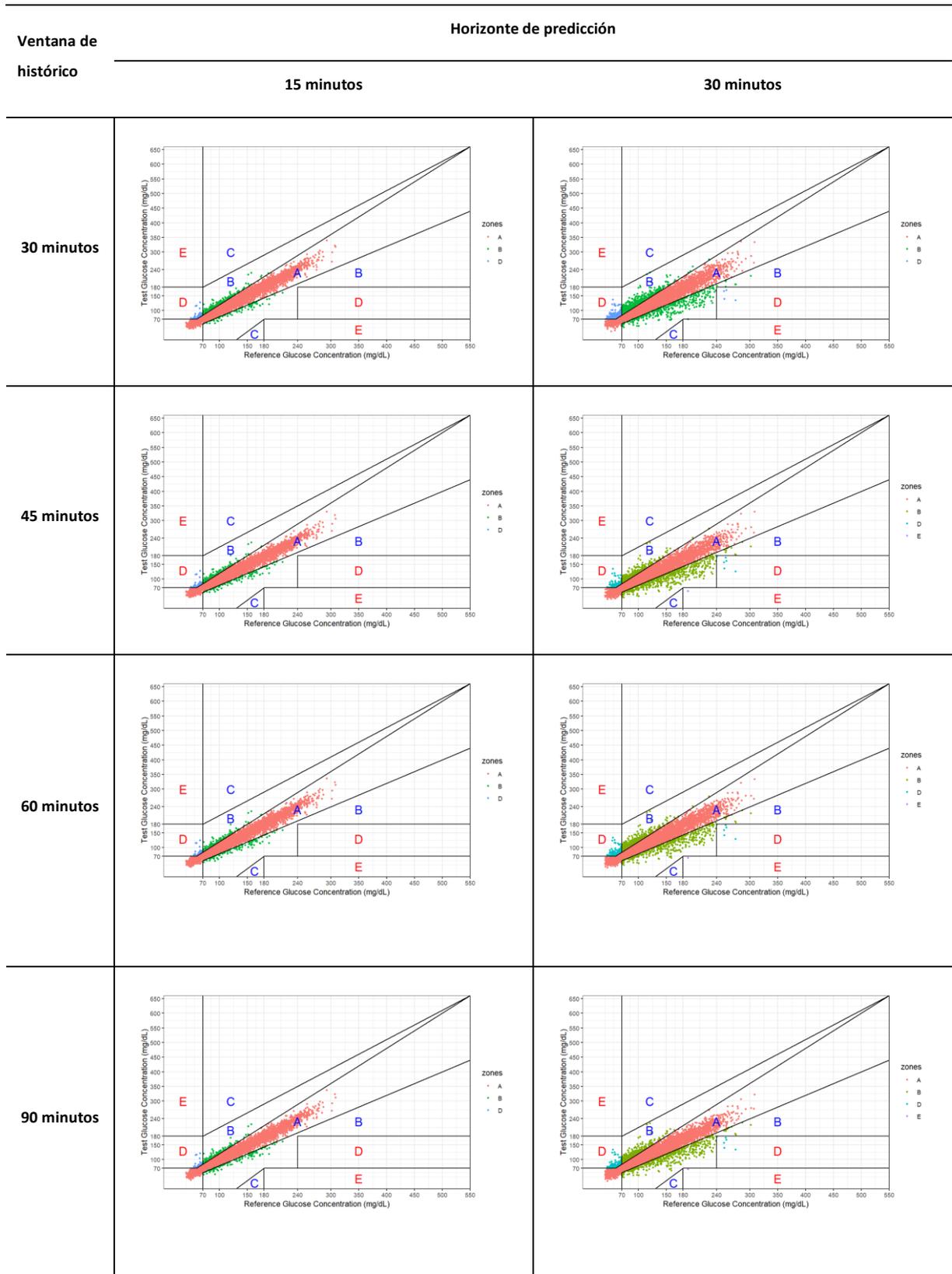


Figura 43: Código en Python empleado para generar el Modelo 5.

```

# Función para crear el modelo GRU 2
def crear_modelo():
    modelo = Sequential()
    modelo.add(GRU(128, input_shape=(ventana, 1), return_sequences=True))
    modelo.add(GRU(128))
    modelo.add(Dense(64, activation='relu'))
    modelo.add(Dense(1, activation='linear'))

    return modelo

# Crear el modelo
modelo = crear_modelo()

# Compilar el modelo
modelo.compile(optimizer='adam', loss='mse', metrics=["accuracy"])

```

5.2.5.1. Pacientes reales

En el caso del modelo 5 empleando pacientes reales, las métricas descritas previamente para el horizonte de predicción de 15 minutos se presentan en la Tabla 31, mientras que los resultados para el horizonte de predicción de 30 minutos se muestran en la Tabla 32.

Tabla 31: Métricas del modelo 5 para un horizonte de predicción de 15 minutos en pacientes reales

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	13.18	96.98%	2.99%	99.97%	0.00%	0.03%	0.00%
VH: 45 minutos	9.16	99.11%	0.85%	99.96%	0.00%	0.03%	0.00%
VH: 60 minutos	8.61	99.08%	0.85%	99.93%	0.00%	0.07%	0.00%
VH: 90 minutos	8.12	99.14%	0.72%	99.86%	0.00%	0.13%	0.00%

Tabla 32: Métricas del modelo 5 para un horizonte de predicción de 30 minutos en pacientes reales

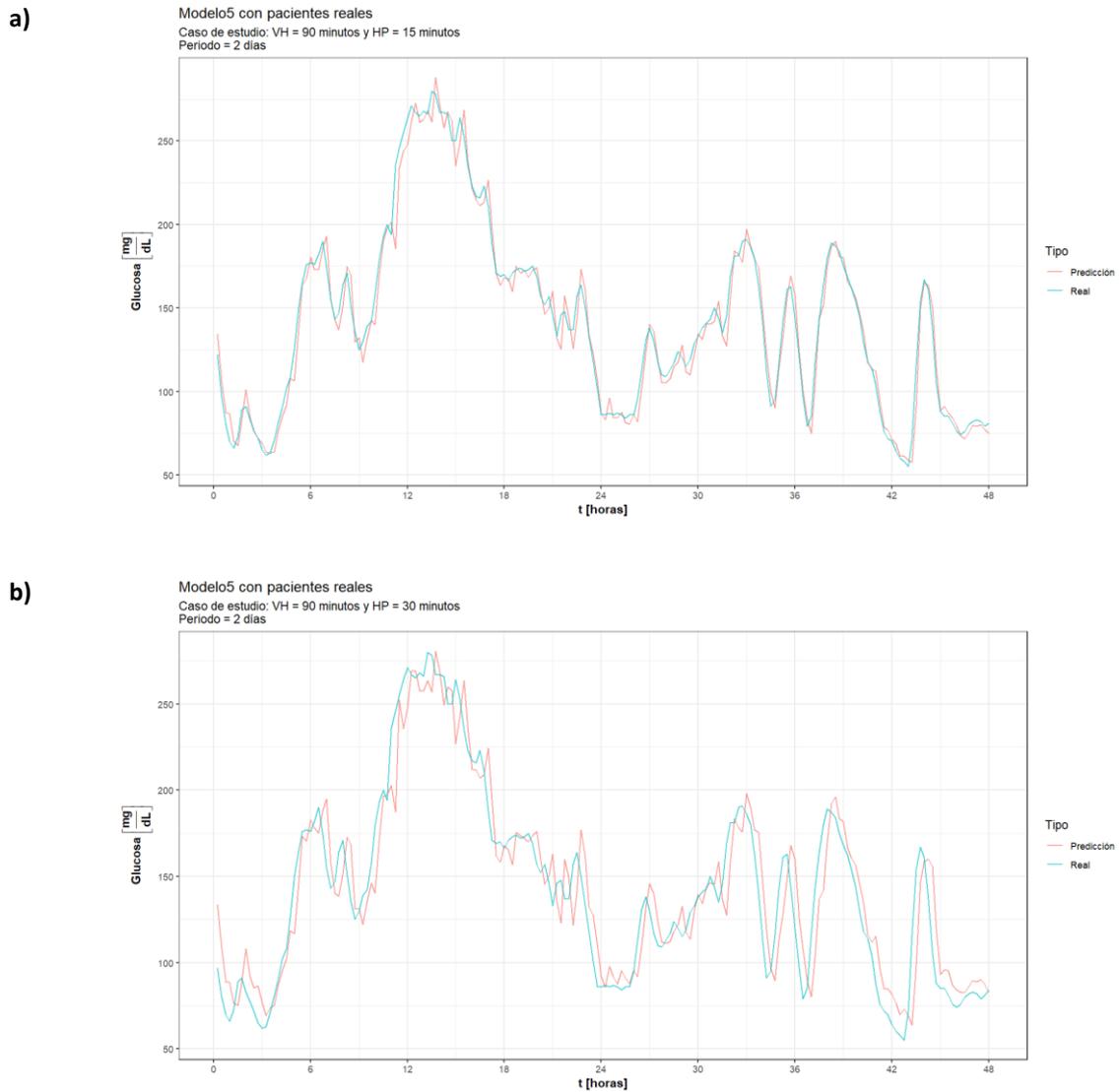
Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	19.22	89.32%	9.24%	98.56%	0.00%	1.45%	0.00%
VH: 45 minutos	18.01	88.32%	9.84%	98.16%	0.00%	1.84%	0.00%
VH: 60 minutos	17.70	90.95%	8.76%	99.71%	0.00%	0.30%	0.00%
VH: 90 minutos	17.53	89.91%	8.34%	98.25%	0.00%	1.75%	0.00%

Al emplear un modelo de 2 capas GRU, y compararlo con el desempeño obtenido con el modelo de 2 capas LSTM, se observa una reducción significativa en el valor del RMSE. Esto implica que tiene una mayor capacidad para predecir que el modelo 3 descrito previamente.

Además, se observa una baja influencia de la zona B en las zonas clínicamente aceptadas en ambos horizontes de predicción, en el cual el mayor porcentaje lo representa la zona A. Incluso

en la zona D, se obtiene un error reducido. Esto permite reconocer que captura y predice mejor los valores en este rango de valores. Las gráficas de las series temporales se las observa en la Figura 44 y las gráficas de Clarke en la Figura 45.

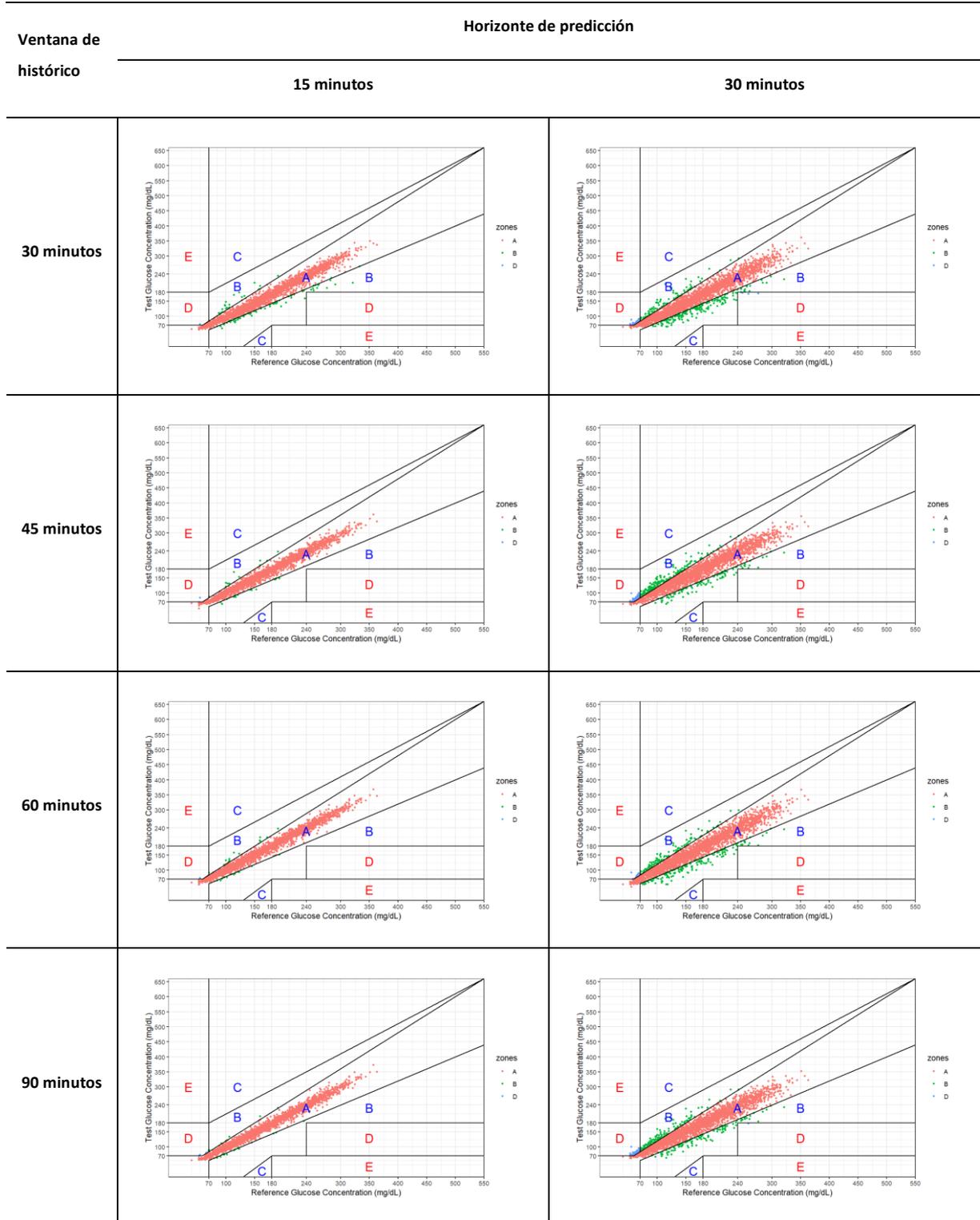
Figura 44: Mejores respuestas del Modelo 5 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.



Al observar las fluctuaciones de la serie temporal de la predicción en la Figura 44, se puede apreciar que, para un horizonte de predicción de 15 minutos, la predicción del modelo muestra un buen ajuste en con los valores reales de la serie temporal. Sin embargo, al considerar un horizonte de predicción de 30 minutos, se observa que la predicción tiende a

subestimar ligeramente los valores de glucosa en comparación con los valores reales. Es importante destacar que, este margen de subestimación es mínimo y la predicción aún se mantiene generando buen ajuste.

Figura 45: Gráficas de Clarke para el modelo 5 – Pacientes reales



5.2.5.2. Pacientes simulados

En el caso del modelo 5 empleando pacientes simulados, las métricas descritas previamente para el horizonte de predicción de 15 minutos se presentan en la Tabla 33, mientras que los resultados para el horizonte de predicción de 30 minutos se muestran en la Tabla 34.

Tabla 33: Métricas del modelo 5 para un horizonte de predicción de 15 minutos en pacientes simulados

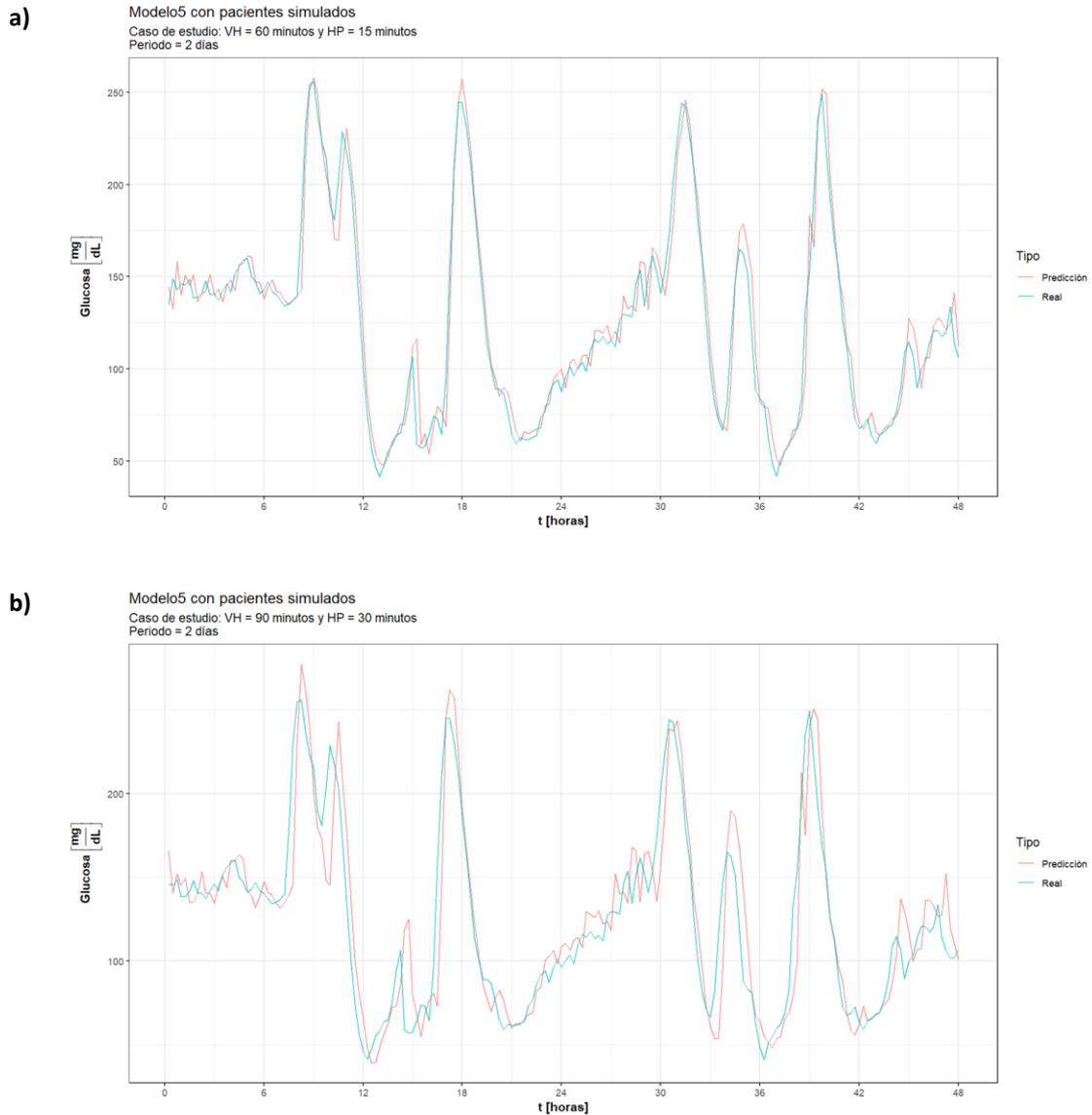
Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	11.74	95.80%	3.71%	99.51%	0.00%	0.49%	0.00%
VH: 45 minutos	11.92	95.46%	4.29%	99.75%	0.00%	0.25%	0.00%
VH: 60 minutos	10.90	95.43%	3.77%	99.20%	0.00%	0.81%	0.00%
VH: 90 minutos	11.62	95.86%	3.89%	99.75%	0.00%	0.25%	0.00%

Tabla 34: Métricas del modelo 5 para un horizonte de predicción de 30 minutos en pacientes simulados

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	21.00	84.10%	13.50%	97.60%	0.00%	2.38%	0.02%
VH: 45 minutos	20.74	84.49%	12.77%	97.26%	0.00%	2.72%	0.02%
VH: 60 minutos	20.19	85.36%	12.66%	98.02%	0.00%	1.96%	0.02%
VH: 90 minutos	19.55	83.79%	14.51%	98.30%	0.00%	1.68%	0.02%

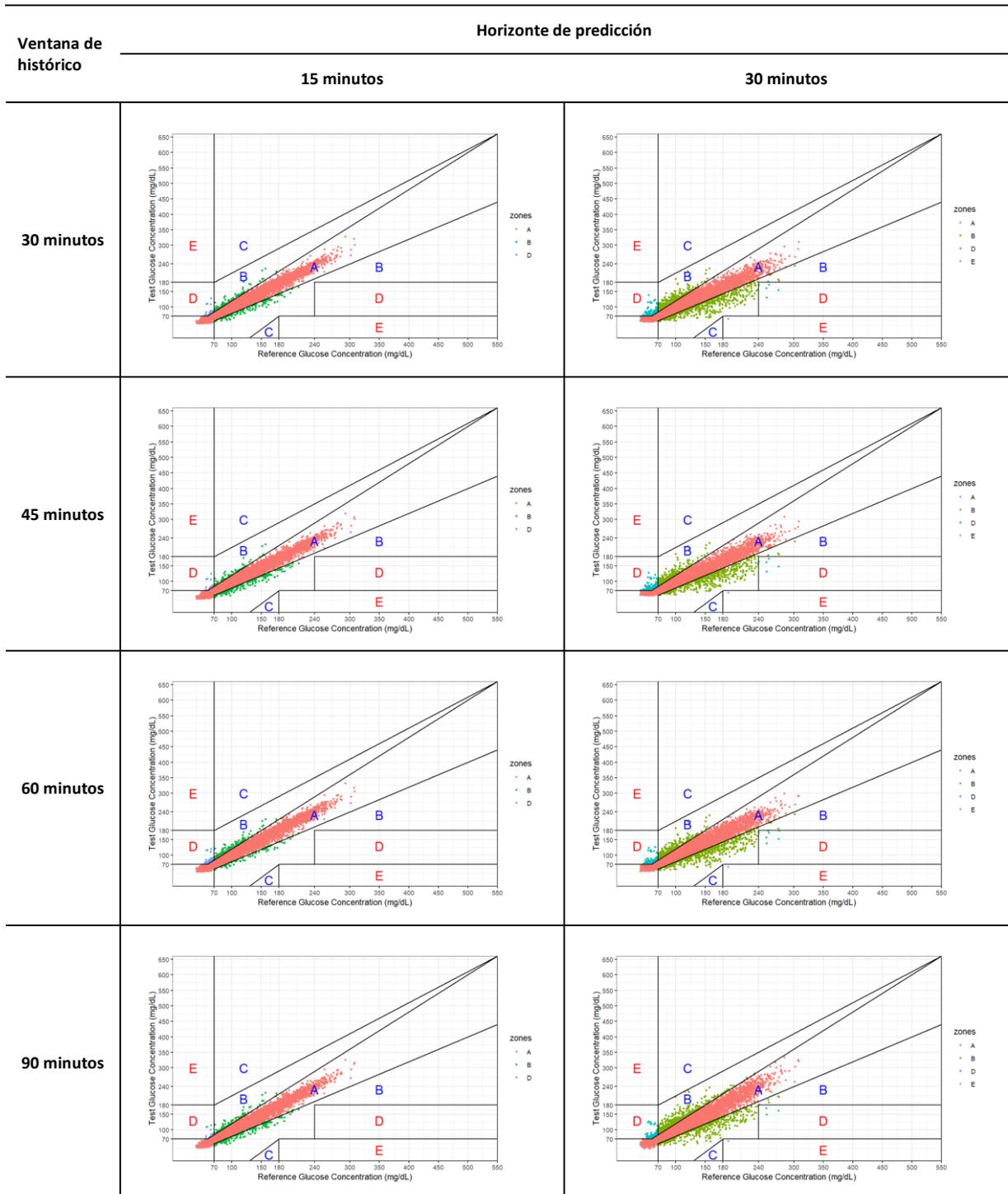
Al comparar el modelo generado con dos capas LSTM y el modelo generado de dos capas GRU, se logró observar en la Tabla 33 que se tiene una reducción de la incidencia de la zona D al emplear el modelo de dos capas GRU con un horizonte de predicción de 15 minutos. Sin embargo, en la Tabla 34, al considerar un horizonte de predicción de 30 minutos, se observa que el error en las zonas D y E se mantiene en proporción en ambos modelos. Las gráficas de las series temporales se las observa en la Figura 46 y las gráficas de Clarke en la Figura 47.

Figura 46: Mejores respuestas del Modelo 5 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.



Al observar la Figura 46, se puede apreciar que el modelo de dos capas GRU muestra un mejor ajuste en comparación con el modelo de dos capas LSTM, Figura 36. Mientras que el modelo de dos capas LSTM subestimaba los valores de glucosa, el modelo de dos capas GRU logra mantener la tendencia y tiende a sobreestimar en un porcentaje mínimo. Es importante destacar que, aunque el modelo de dos capas GRU pueda tener una ligera sobreestimación en algunos casos, este se mantiene en concordancia con los valores reales de glucosa.

Figura 47: Gráficas de Clarke para el modelo 5 – Pacientes Simulados



5.2.6. Modelo 6: Modelo de 1 capa LSTM + 1 capa GRU

El último modelo generado se trata de una arquitectura combinada de una capa LSTM y una capa GRU. El modelo comienza con una capa LSTM constituida por 64 neuronas. Esta capa

recibe una secuencia de datos de la forma (ventana,1), donde predomina el tamaño de la ventana deslizante con la que se entrene el modelo. Esta capa LSTM tiene configurado el parámetro “return_sequences=True”, lo que implica que esta capa a la salida devuelve las secuencias completas. Esto permite que la siguiente capa, la capa GRU opere sobre estas secuencias. Esta capa GRU se conforma por 64 neuronas. Después de la capa GRU, se le agrega una capa densa completamente conectada de 64 neuronas con una activación “ReLU”. Para finalizar el modelo se le agrega la capa de predicción, una capa densa con activación “Lineal” de una sola neurona. La generación del modelo se realizada como en la Figura 48.

Figura 48: Código en Python empleado para generar el Modelo 6.

```
# Función para crear el modelo LSTM + GRU
def crear_modelo():
    modelo = Sequential()
    modelo.add(LSTM(64, input_shape=(ventana, 1), return_sequences=True))
    modelo.add(GRU(64))
    modelo.add(Dense(64, activation='relu'))
    modelo.add(Dense(1, activation='linear'))

    return modelo

# Crear el modelo
modelo = crear_modelo()

# Compilar el modelo
modelo.compile(optimizer='adam', loss='mse', metrics=["accuracy"])
```

5.2.6.1. Pacientes reales

En el caso del modelo 6 empleando pacientes reales, las métricas descritas previamente para el horizonte de predicción de 15 minutos se presentan en la Tabla 35, mientras que los resultados para el horizonte de predicción de 30 minutos se muestran en la Tabla 36.

Tabla 35: Métricas del modelo 6 para un horizonte de predicción de 15 minutos en pacientes reales

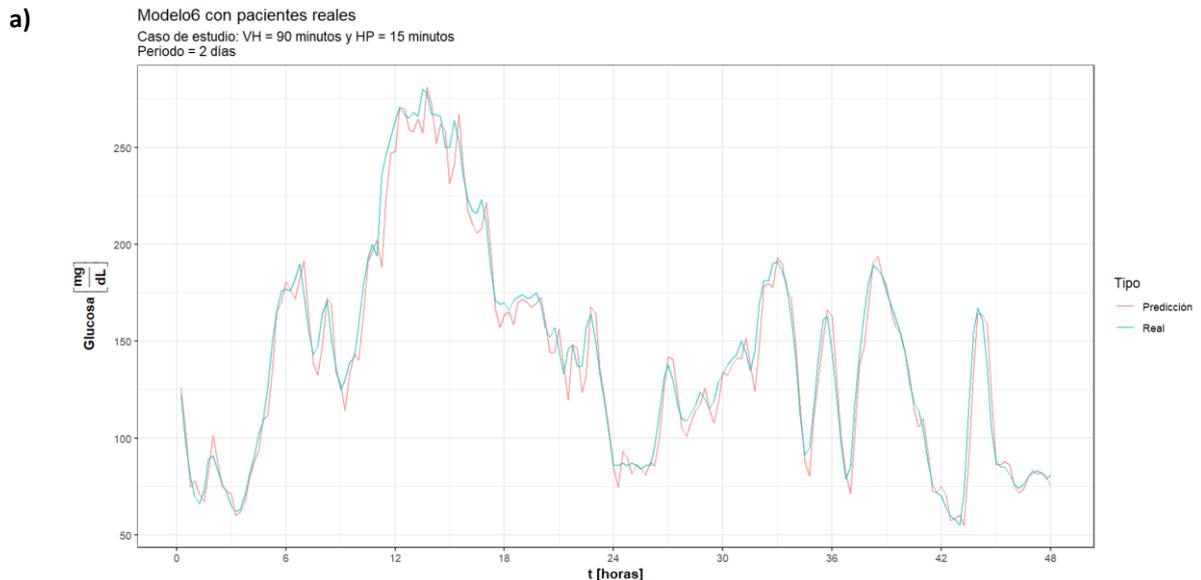
Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	10.6	97.87%	2.07%	99.94%	0.00%	0.07%	0.00%
VH: 45 minutos	12.6	96.02%	2.99%	99.01%	0.00%	0.99%	0.00%
VH: 60 minutos	12.2	97.99%	1.97%	99.96%	0.00%	0.03%	0.00%
VH: 90 minutos	9.22	99.08%	0.86%	99.94%	0.00%	0.07%	0.00%

Tabla 36: Métricas del modelo 6 para un horizonte de predicción de 30 minutos en pacientes reales

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	18.8	89.84%	9.50%	99.34%	0.00%	0.66%	0.00%
VH: 45 minutos	18.1	87.11%	9.21%	96.32%	0.00%	3.68%	0.00%
VH: 60 minutos	18.5	87.62%	10.57%	98.19%	0.00%	1.81%	0.00%
VH: 90 minutos	17.5	89.39%	8.24%	97.63%	0.00%	2.37%	0.00%

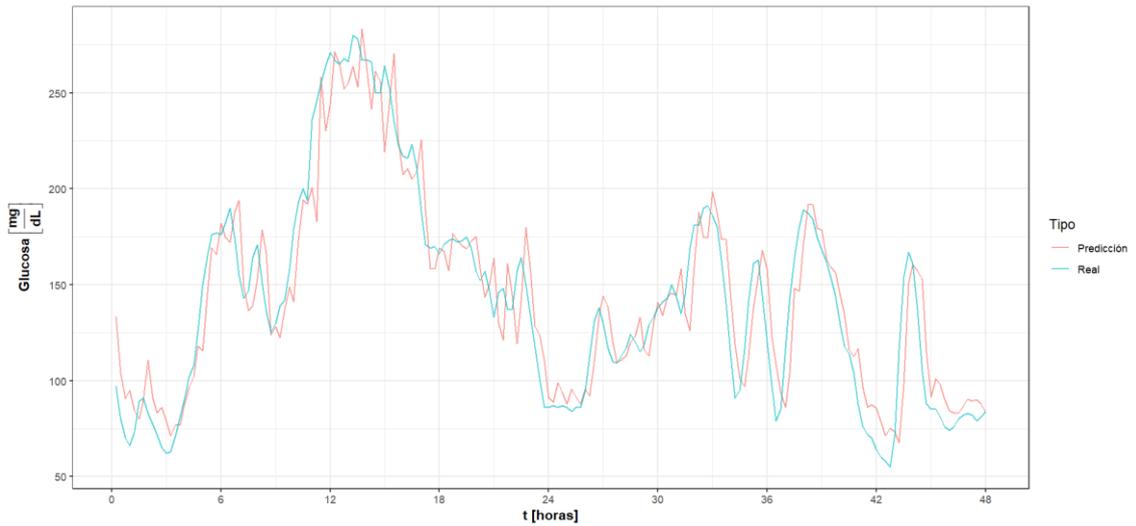
Al emplear el modelo combinando dos capas secuenciales, se logró obtener el mejor resultado en términos de la métrica RMSE tanto para el horizonte de predicción de 15 minutos como para un horizonte de predicción de 30 minutos. Un modelo combinado fue capaz de proporcionar predicciones más precisas en comparación a los modelos individuales. Al analizar el porcentaje de las zonas de Clarke, se observó que, para el horizonte de predicción de 15 minutos, el modelo logra mantener un bajo porcentaje en la zona D por debajo del 1%. Sin embargo, para el horizonte de predicción de 30 minutos, se observa que el porcentaje de la zona D aumentó y se encuentra entre el 1% y el 4%. Las gráficas de las series temporales se las observa en la Figura 49 y las gráficas de Clarke en la Figura 50.

Figura 49: Mejores respuestas del Modelo 6 empleado en pacientes reales. a) HP de 15 minutos. b) HP de 30 minutos.



b)

Modelo6 con pacientes reales
 Caso de estudio: VH = 90 minutos y HP = 30 minutos
 Periodo = 2 días



Al observar la Figura 49, se puede apreciar que, para el horizonte de predicción de 15 minutos, el modelo muestra un buen ajuste, aunque con una ligera subestimación de los valores de glucosa en comparación con los valores reales. Por otro lado, para un horizonte de predicción de 30 minutos, se mantiene la tendencia, pero con ligeras fluctuaciones en la predicción.

5.2.6.2. Pacientes simulados

En el caso del modelo 6 empleando pacientes simulados, las métricas descritas previamente para el horizonte de predicción de 15 minutos se presentan en la Tabla 37, mientras que los resultados para el horizonte de predicción de 30 minutos se muestran en la Tabla 38.

Tabla 37: Métricas del modelo 6 para un horizonte de predicción de 15 minutos en pacientes simulados

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	11.4	95.10%	4.10%	99.20%	0.00%	0.81%	0.00%
VH: 45 minutos	11.6	95.89%	3.82%	99.71%	0.00%	0.30%	0.00%
VH: 60 minutos	11.1	95.31%	3.87%	99.18%	0.00%	0.82%	0.00%
VH: 90 minutos	11.2	96.34%	3.21%	99.55%	0.00%	0.46%	0.00%

Figura 50: Gráficas de Clarke para el modelo 6 – Pacientes reales

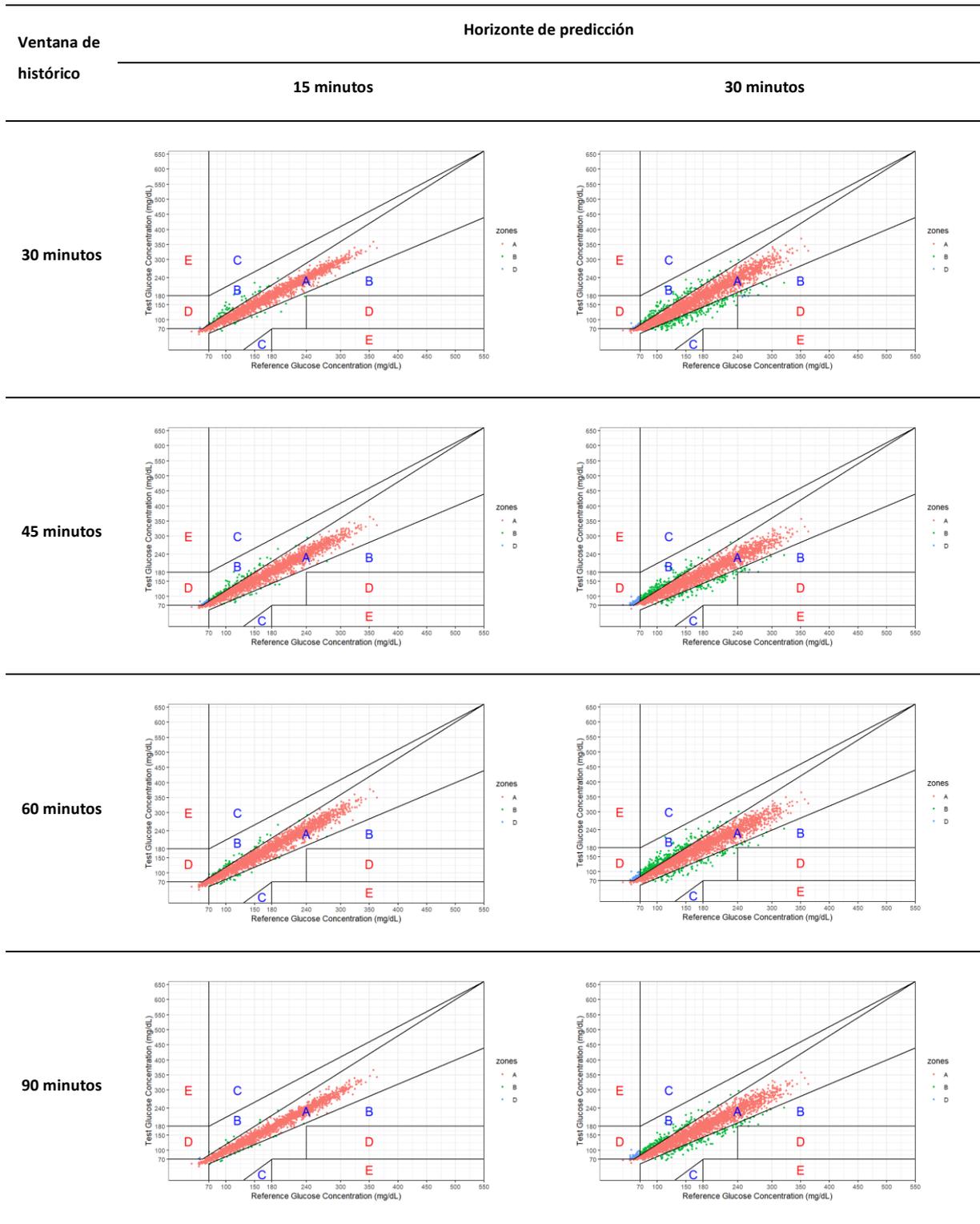
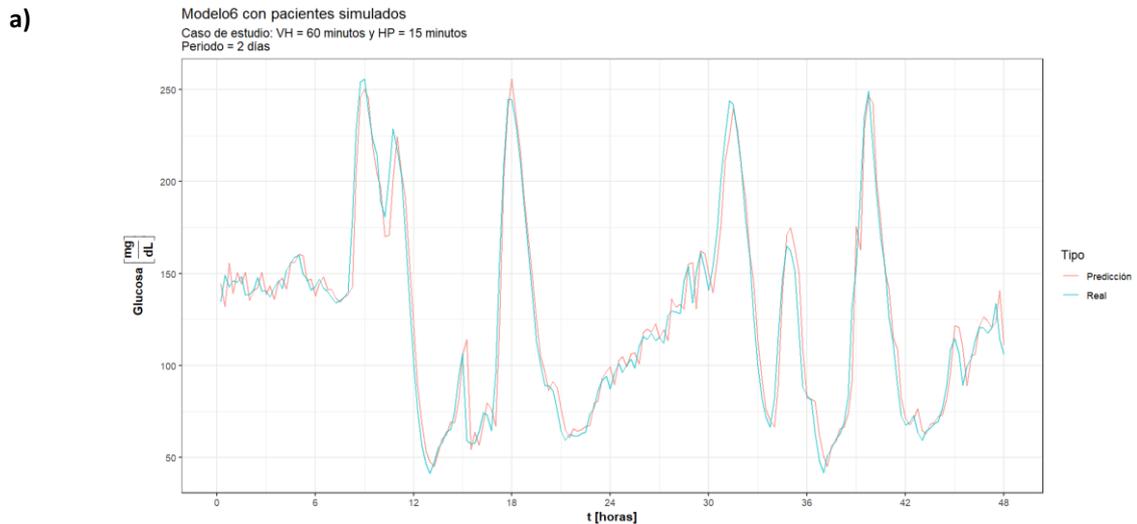


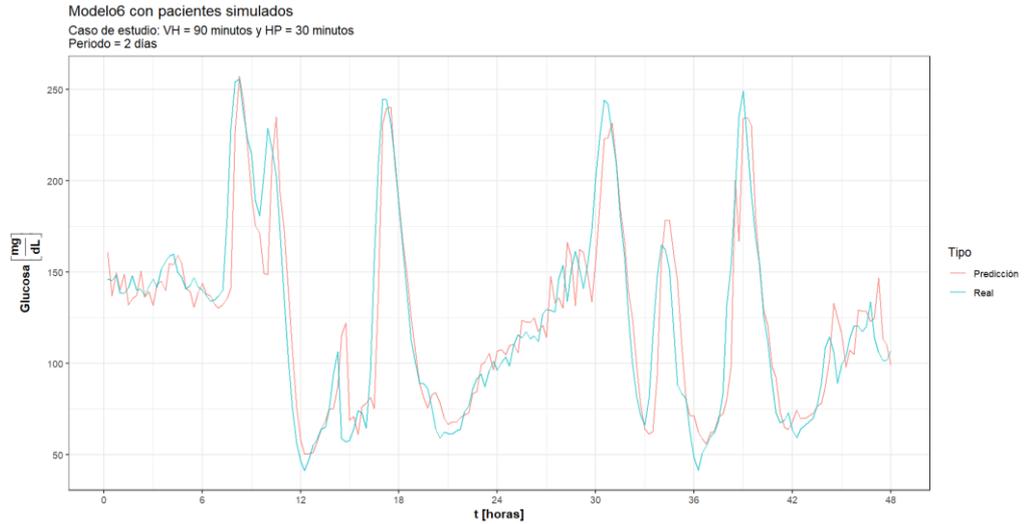
Tabla 38: Métricas del modelo 6 para un horizonte de predicción de 30 minutos en pacientes simulados

Caso de estudio	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
VH: 30 minutos	20.2	82.50%	14.28%	96.78%	0.00%	3.21%	0.02%
VH: 45 minutos	19.8	81.83%	14.86%	96.69%	0.00%	3.29%	0.02%
VH: 60 minutos	19.8	85.19%	12.67%	97.86%	0.04%	2.09%	0.02%
VH: 90 minutos	19.2	84.83%	12.54%	97.37%	0.00%	2.61%	0.02%

Al analizar el valor del RMSE obtenido en el modelo 6 para un horizonte de predicción de 15 minutos, se observa que se mantiene con variaciones pequeñas. Esto indica que el modelo tiene una consistencia en la precisión de las predicciones bajo diferentes ventanas de histórico. Además, al analizar el porcentaje de la zona D, se observa que se mantiene por debajo del 0.80%. Por otro lado, al considerar un horizonte de predicción de 30 minutos, se observa que el porcentaje de las zonas D y E se mantiene en un rango similar al de los modelos de dos capas GRU y de dos capas LSTM. Las gráficas de las series temporales se las observa en la Figura 51 y las gráficas de Clarke en la Figura 52.

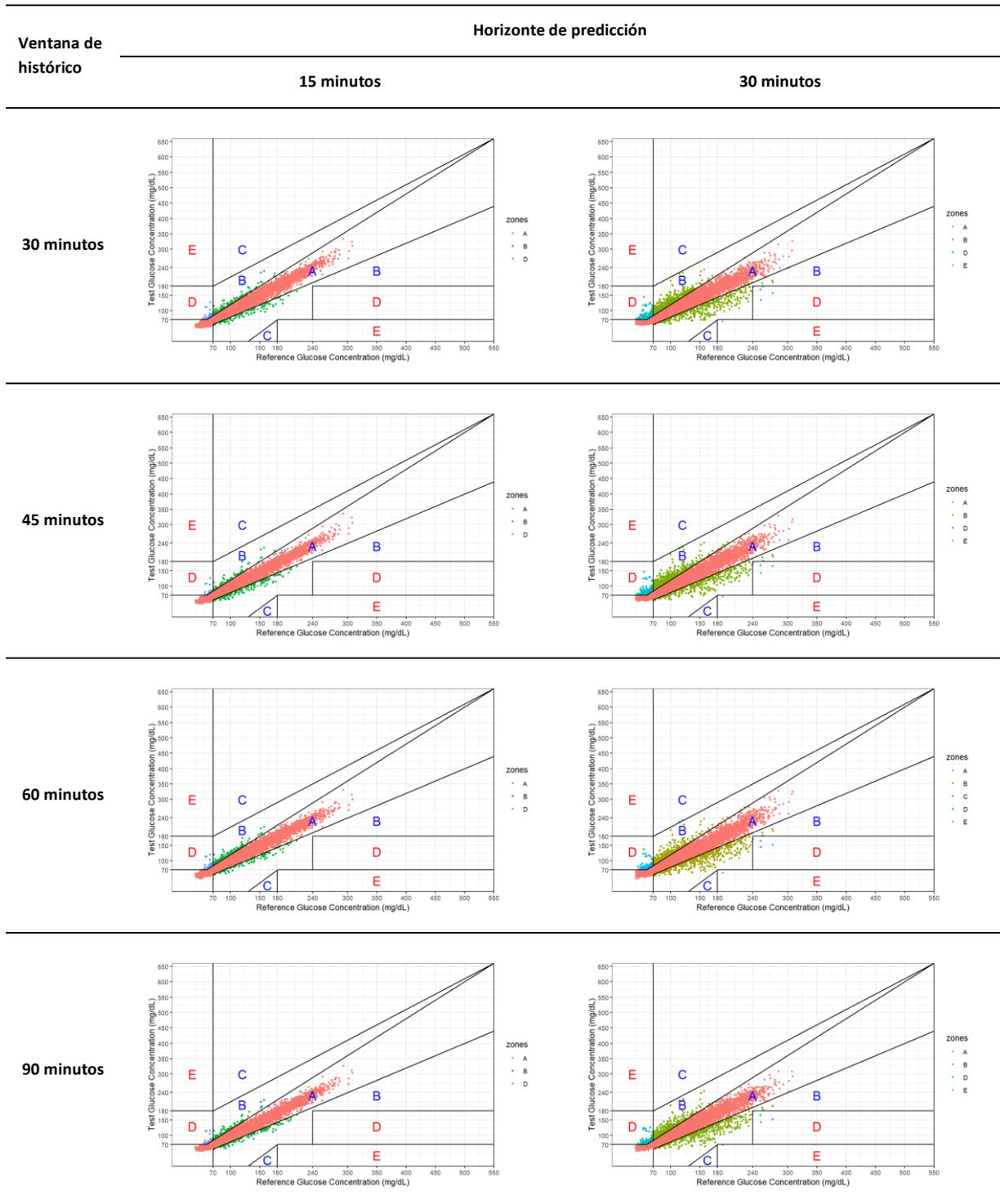
Figura 51: Mejores respuestas del Modelo 6 empleado en pacientes simulados. a) HP de 15 minutos. b) HP de 30 minutos.

b)



Para la comparativa de la serie temporal entre la predicción y el valor real en los pacientes simulados, se observa un mejor ajuste en comparación con modelos anteriores. Se logra seguir de manera más precisa las variaciones en los niveles de glucosa de los pacientes simulados. A diferencia de modelos anteriores, no se observa un desfase significativo entre la predicción y los valores reales para el horizonte de predicción de 30 minutos. Este mejor ajuste se obtiene especialmente para un horizonte de predicción de 15 minutos, lo cual indica que el modelo es capaz de capturar de manera más precisa las fluctuaciones a corto plazo en los niveles de glucosa.

Figura 52: Gráficas de Clarke para el modelo 6 – Paciente simulados



6. Discusión y análisis de resultados

En el presente estudio, se llevó a cabo un análisis exhaustivo de diferentes modelos de predicción de glucosa utilizando datos de pacientes reales y simulados. Se emplearon modelos de redes neuronales recurrentes, específicamente capas, RNN, LSTM y capas GRU, tanto en configuraciones individuales como combinadas, para evaluar su desempeño en la predicción de los niveles de glucosa. Los resultados obtenidos revelaron varias tendencias y patrones.

En primer lugar, se realiza el análisis a los pacientes reales, generando tablas de resumen con el RMSE obtenido en los modelos y el porcentaje de las zonas clínicamente aceptadas (Zona A + Zona B) del análisis de Clarke. En la Tabla 39 se muestran los resultados para un horizonte de predicción de 15 minutos, mientras que la Tabla 40 muestra los resultados para un horizonte de predicción de 30 minutos. Se observó que el tamaño de la ventana histórica tiene un alto impacto en la precisión de la predicción. Como caso puntual, se encontró que los mejores resultados de los modelos en ambos escenarios se encuentran con la ventana histórica de 60 y 90. Esto indica que la inclusión de más información histórica ayuda a que los modelos logren capturar de mejor forma las tendencias y los patrones de la glucosa en la sangre. Mientras que, si analizamos el efecto del horizonte de predicción, se logró una mejora significativa en la métrica del RMSE bajo un horizonte de 15 minutos, es decir que se tiene un mejor desempeño en la predicción a corto plazo, debido a que, para los modelos empleados, el capturar el comportamiento de la serie para predecir 30 minutos en el futuro implica un mayor desafío.

En términos de las zonas de Clarke, se encontró que la mayoría de los modelos lograron un porcentaje alto de zonas clínicamente aceptadas, especialmente en el horizonte de predicción de 15 minutos. Sin embargo, se observó una mayor incidencia de puntos en la zona B para el horizonte de predicción de 30 minutos, lo cual sugiere que la predicción de valores moderadamente altos y bajos de glucosa puede llegar a ser más complicados de predecir en este horizonte de tiempo. Al observar el comportamiento general de las gráficas de Clarke, podemos apreciar que existe una mayor dispersión de los valores en el horizonte de predicción de 30 minutos que en el horizonte de predicción de 15 minutos.

Al comparar el desempeño de los diferentes modelos, el emplear un modelo con una sola capa RNN, no se obtuvo resultados tan prometedores en términos de predicción de la glucosa. Los resultados fueron mejorando con la aplicación de las capas LSTM y capas GRU, a tal punto que, se demostró que se obtuvo un mejor desempeño en términos del RMSE y el análisis de las zonas de Clarke para el modelo que combinó una capa LSTM y una capa GRU.

Tabla 39: Resumen de métrica RMSE y zona clínicamente aceptada de Clarke (Zona A + Zona B) para un horizonte de predicción de 15 minutos en pacientes reales.

VH	Modelo 1		Modelo 2		Modelo 3		Modelo 4		Modelo 5		Modelo 6	
	RMSE	Zona A + B (%)										
30 minutos	19,36	99,54%	19,19	98,98%	11,44	99,87%	19,29	98,55%	13,18	99,97%	10,63	99,94%
45 minutos	14,04	99,87%	12,68	99,54%	15,06	96,55%	13,03	98,75%	9,16	99,96%	12,55	99,01%
60 minutos	12,12	99,83%	11,80	99,38%	11,17	99,80%	11,91	99,90%	8,61	99,93%	12,20	99,96%
90 minutos	11,46	99,83%	11,78	99,84%	10,64	99,30%	11,53	99,51%	8,12	99,86%	9,22	99,94%

Tabla 40: Resumen de métrica RMSE y zona clínicamente aceptada de Clarke (Zona A + Zona B) para un horizonte de predicción de 30 minutos en pacientes reales.

VH	Modelo 1		Modelo 2		Modelo 3		Modelo 4		Modelo 5		Modelo 6	
	RMSE	Zona A + B (%)										
30 minutos	20,52	98,72%	20,39	98,00%	18,82	99,31%	20,34	97,90%	19,22	98,56%	18,83	99,34%
45 minutos	19,06	97,63%	19,36	97,83%	17,45	98,09%	19,17	98,42%	18,01	98,16%	18,08	96,32%
60 minutos	19,51	97,54%	18,92	97,30%	18,47	96,05%	18,92	97,33%	17,70	99,71%	18,54	98,19%
90 minutos	20,31	98,46%	18,57	98,35%	18,76	95,98%	18,46	97,96%	17,53	98,25%	17,47	97,63%

Se completa el análisis con los pacientes simulados, para llevar a cabo el análisis, se generó la Tabla 41 que contempla el resumen de los resultados obtenido para el horizonte de predicción de 15 minutos y la Tabla 42 el resumen de los resultados obtenidos para el horizonte de predicción de 30 minutos. El valor del RMSE no varía mucho entre los diferentes modelos y las ventanas de histórico, sin embargo, con la muestra graficada de la serie temporal, se aprecia que las predicciones se van ajustando de mejor forma con los datos de la prueba. Los modelos tienden a generar una ligera subestimación en la predicción en ambos horizontes de predicción.

La aleatoriedad de los pacientes simulados puede ser punto para analizar, debido a que el comportamiento de la ingesta de alimentos no sigue un patrón definido como es el caso de los pacientes reales. Es importante destacar que este estudio se limita a un conjunto de datos

específico y a configuraciones particulares de modelos de redes neuronales recurrentes. Se requiere la aplicación de datos con mayor diversidad para poder generalizar de mayor forma.

Tabla 41: Resumen de métrica RMSE y zona clínicamente aceptada de Clarke (Zona A + Zona B) para un horizonte de predicción de 15 minutos en pacientes simulados.

VH	Modelo 1		Modelo 2		Modelo 3		Modelo 4		Modelo 5		Modelo 6	
	RMSE	Zona A + B (%)										
30 minutos	12,74	98,86%	12,50	99,28%	12,24	99,13%	11,92	98,81%	11,74	99,51%	11,40	99,20%
45 minutos	12,52	98,81%	12,37	98,95%	11,17	98,98%	11,44	98,98%	11,92	99,75%	11,60	99,71%
60 minutos	11,39	99,32%	11,57	99,27%	11,14	99,28%	11,40	99,34%	10,90	99,20%	11,08	99,18%
90 minutos	11,70	99,35%	11,33	99,53%	11,32	99,48%	11,27	99,41%	11,62	99,75%	11,20	99,55%

Tabla 42: Resumen de métrica RMSE y zona clínicamente aceptada de Clarke (Zona A + Zona B) para un horizonte de predicción de 30 minutos en pacientes simulados.

VH	Modelo 1		Modelo 2		Modelo 3		Modelo 4		Modelo 5		Modelo 6	
	RMSE	Zona A + B (%)										
30 minutos	22,74	98,58%	21,85	97,53%	21,19	97,58%	20,66	96,41%	21,00	97,60%	20,16	96,78%
45 minutos	21,47	97,16%	21,29	98,14%	20,37	97,52%	21,01	98,02%	20,74	97,26%	19,81	96,69%
60 minutos	20,75	97,23%	20,40	97,86%	19,75	98,37%	20,19	97,34%	20,19	98,02%	19,80	97,86%
90 minutos	20,23	97,41%	20,10	97,73%	19,98	98,49%	20,41	97,52%	19,55	98,30%	19,20	97,37%

7. Conclusiones y trabajo futuro

7.1. Conclusiones

Utilizando el simulador UVA/PADOVA Type 1 Diabetes Simulator, se crearon los perfiles de glucosa de 10 pacientes adultos. Este enfoque nos permitió obtener un comportamiento similar al de pacientes reales sin comprometer la privacidad de los pacientes. El simulador proporciona una forma confiable de generar datos que permiten replicar el comportamiento dinámico de glucosa en el cuerpo. Al emplear este simulador se pudo evaluar los modelos de aprendizaje profundo en un entorno controlado.

Después de revisar los datos, se encontró una amplia variedad de combinaciones de arquitecturas de aprendizaje profundo utilizadas para predecir los valores de glucosa en estudios anteriores. Para este estudio, se optó por utilizar variaciones de la arquitectura RNN, teniendo en cuenta las consideraciones necesarias para el preprocesamiento y procesamiento de los datos con el objetivo de obtener los mejores resultados.

Se observó el efecto que tiene la ventana histórica sobre la predicción del nivel de la glucosa en la sangre. Se determinó con la experimentación que, al aumentar el tamaño de la ventana histórica, el valor del RMSE se ve reducido, obteniendo los mejores resultados con ventanas de histórico de 60 minutos y 90 minutos. El tener un mayor número de datos les permitió a los modelos capturar de mejor forma las tendencias y los patrones que poseen este tipo de series temporales. Según el horizonte de predicción que se escoja, la evaluación del modelo se verá influenciada. Se encontró que con un horizonte de predicción de 15 minutos se obtiene mejores resultados en términos del RMSE y del análisis de error de Clarke, mientras que, con un horizonte de 30 minutos, el predecir la glucosa puede tornarse más desafiante para las redes.

El modelo que combinó la capa LSTM y la capa GRU, demostró tener el mejor desempeño en comparación de emplear este tipo de capas de forma individual. Este modelo incluso mejoró el ajuste de la predicción con el valor de la prueba con un horizonte de predicción de 30 minutos en ambos escenarios, tal como se aprecia en las figuras de las series temporales.

En algunos casos al momento de realizar la comparativa de la serie temporal del valor real y la predicción se observan que existían casos de sobreestimación o subestimación. Estas condiciones pueden verse influenciadas por factores como la complejidad de las fluctuaciones en los niveles de glucosa y la capacidad del modelo para capturar estos cambios.

7.2. Líneas de trabajo futuro

Se tiene planteado dos líneas de trabajo futuro. La primera línea de trabajo se encuentra la incorporación de variables clínicas adicionales que permitan enriquecer la capacidad de predicción de los modelos. Debido a que la DMT1 posee diferentes factores de influencia tal como se describió en el estado de arte. Se plantea considerar las variables como la cantidad de ingesta de alimentos, actividad física, medicamentos o inclusive los eventos de estrés. Estas variables implican una mayor complejidad al momento de recolectar la información.

La segunda línea de trabajo planteada es generar la evaluación de los modelos en tiempo real. Esto implica la integración de los modelos generados en sistemas de monitoreo continuo de glucosa. Este enfoque considera dos grandes desafíos, el primero que es la latencia entre que ocurre un evento y se produce la predicción, y el segundo reto presente es la adaptabilidad a cambios en los patrones de la glucosa en la sangre.

Referencias bibliográficas

- Abbott. (2020). *User 's Manual FreeStyle Libre 2*.
https://www.binsons.com/uploads/userfiles/files/documents/products/Libre_2_User_Manual.pdf
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. In *Journal of Big Data* (Vol. 8, Issue 1). Springer International Publishing. <https://doi.org/10.1186/s40537-021-00444-8>
- ASSOCIATION, A. D. (2010). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 33(SUPPL. 1). <https://doi.org/10.2337/dc10-S062>
- Atkinson, M. A., & Eisenbarth, G. S. (2001). Type 1 diabetes: New perspectives on disease pathogenesis and treatment. *Lancet*, 358(9277), 221–229. [https://doi.org/10.1016/S0140-6736\(01\)05415-0](https://doi.org/10.1016/S0140-6736(01)05415-0)
- Aung, Y. Y. M., Wong, D. C. S., & Ting, D. S. W. (2021). The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare. *British Medical Bulletin*, 139(1), 4–15. <https://doi.org/10.1093/bmb/ldab016>
- Ávila-Tomás, J. F., Mayer-Pujadas, M. A., & Quesada-Varela, V. J. (2021). Artificial intelligence and its applications in medicine II: Current importance and practical applications. *Atencion Primaria*, 53(1), 81–88. <https://doi.org/10.1016/j.aprim.2020.04.014>
- Beck, R. W., Bergenstal, R. M., Laffel, L. M., & Pickup, J. C. (2019). Advances in technology for management of type 1 diabetes. *The Lancet*, 394(10205), 1265–1273. [https://doi.org/10.1016/S0140-6736\(19\)31142-0](https://doi.org/10.1016/S0140-6736(19)31142-0)
- Bezdan, T., & Bačaniin Džakula, N. (2019). *Convolutional Neural Network Layers and Architectures*. *July*, 445–451. <https://doi.org/10.15308/sinteza-2019-445-451>
- Carrillo-Moreno, J., Pérez-Gandía, C., Sendra-Arranz, R., García-Sáez, G., Hernando, M. E., & Gutiérrez, Á. (2021). Long short-term memory neural network for glucose prediction. *Neural Computing and Applications*, 33(9), 4191–4203. <https://doi.org/10.1007/s00521-020-05248-0>
- Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder–decoder approaches. *Proceedings of SSST 2014 - 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111.

<https://doi.org/10.3115/v1/w14-4012>

- Chowdhury, S., & Chakraborty, P. pratim. (2017). Universal health coverage - There is more to it than meets the eye. *Journal of Family Medicine and Primary Care*, 6(2), 169–170. <https://doi.org/10.4103/jfmipc.jfmipc>
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. 1–9. <http://arxiv.org/abs/1412.3555>
- Colmegna, P. H., Sánchez Peña, R. S., De Battista, H., Bondía, J., & Basualdo, M. (2015). *Instituto Tecnológico de Buenos Aires Simulation & Control in Type 1 Diabetes*.
- Conget, D. I. (2002). Diagnosis, classification and pathogenesis of diabetes mellitus. *Revista Espanola de Cardiologia*, 55(5), 528–535. [https://doi.org/10.1016/S0300-8932\(02\)76646-3](https://doi.org/10.1016/S0300-8932(02)76646-3)
- Contreras, I., & Vehi, J. (2018). Artificial intelligence for diabetes management and decision support: Literature review. *Journal of Medical Internet Research*, 20(5), 1–21. <https://doi.org/10.2196/10775>
- Dalla Man, C., Micheletto, F., Lv, D., Breton, M., Kovatchev, B., & Cobelli, C. (2014). The UVA/PADOVA type 1 diabetes simulator: New features. *Journal of Diabetes Science and Technology*, 8(1), 26–34. <https://doi.org/10.1177/1932296813514502>
- Dalla Man, C., Rizza, R. A., & Cobelli, C. (2006). Mixed meal simulation model of glucose-insulin system. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 54(10), 307–310. <https://doi.org/10.1109/IEMBS.2006.260810>
- Freiburghaus, J., Rizzotti-Kaddouri, A., & Albertetti, F. (2020). A deep learning approach for blood glucose prediction of type 1 diabetes. *CEUR Workshop Proceedings*, 2675, 131–135.
- Gillespie, K. M. (2006). Type 1 diabetes: Pathogenesis and prevention. *CMAJ. Canadian Medical Association Journal*, 175(2), 165–170. <https://doi.org/10.1503/cmaj.060244>
- Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4), 1–13. <https://doi.org/10.1002/widm.1312>
- International Diabetes Federation. (2021). International Diabetes Federation. In *Diabetes Research and Clinical Practice* (Vol. 102, Issue 2).

<https://doi.org/10.1016/j.diabres.2013.10.013>

- Kaul, V., Enslin, S., & Gross, S. A. (2020). History of artificial intelligence in medicine. *Gastrointestinal Endoscopy*, 92(4), 807–812. <https://doi.org/10.1016/j.gie.2020.06.040>
- Kim, D. Y., Choi, D. S., Kim, J., Chun, S. W., Gil, H. W., Cho, N. J., Kang, A. R., & Woo, J. (2020). Developing an individual glucose prediction model using recurrent neural network. *Sensors (Switzerland)*, 20(22), 1–15. <https://doi.org/10.3390/s20226460>
- Kopecký, P., Mráz, M., Bláha, J., Lindner, J., Svačina, Š., Hovorka, R., & Haluzík, M. (2013). The use of continuous glucose monitoring combined with computer-based eMPC algorithm for tight glucose control in cardiosurgical ICU. *BioMed Research International*, 2013. <https://doi.org/10.1155/2013/186439>
- Liu, C., Vehí, J., Avari, P., Reddy, M., Oliver, N., Georgiou, P., & Herrero, P. (2019). Long-term glucose forecasting using a physiological model and deconvolution of the continuous glucose monitoring signal. *Sensors (Switzerland)*, 19(19), 1–19. <https://doi.org/10.3390/s19194338>
- Marcus, Y., Eldor, R., Yaron, M., Shaklai, S., Ish-Shalom, M., Shefer, G., Stern, N., Golan, N., Dvir, A. Z., Pele, O., & Gonen, M. (2020). Improving blood glucose level predictability using machine learning. *Diabetes/Metabolism Research and Reviews*, 36(8). <https://doi.org/10.1002/dmrr.3348>
- Mayo, M., Chepulis, L., & Paul, R. G. (2019). Glycemic-aware metrics and oversampling techniques for predicting blood glucose levels using machine learning. *PLoS ONE*, 14(12), 1–19. <https://doi.org/10.1371/journal.pone.0225613>
- Pérez, C. (2014). *Propuesta de algoritmos de predicción de glucosa en pacientes diabéticos*. 464. http://oa.upm.es/25604/1/MARIA_DEL_CARMEN_PEREZ_GANDIA.pdf
- Petersmann, A., Müller-Wieland, D., Müller, U. A., Landgraf, R., Nauck, M., Freckmann, G., Heinemann, L., & Schleicher, E. (2019). Definition, Classification and Diagnosis of Diabetes Mellitus. *Experimental and Clinical Endocrinology and Diabetes*, 127(Suppl 1), S1–S7. <https://doi.org/10.1055/a-1018-9078>
- Primavera, M., Giannini, C., & Chiarelli, F. (2020). Prediction and Prevention of Type 1 Diabetes. *Frontiers in Endocrinology*, 11(June), 1–9. <https://doi.org/10.3389/fendo.2020.00248>

- Rosales, N. (2011). Plataforma de simulación de pacientes diabéticos para validación in silico de estrategias de control glucémico. In *Biomédica* (Vol. 31, Issue sup 3.2).
- T1DMS – The Epsilon Group. (2017). <https://tegvirginia.com/software/t1dms-2014/>
- Vettoretti, M., Facchinetti, A., Sparacino, G., & Cobelli, C. (2018). Type-1 diabetes patient decision simulator for in silico testing safety and effectiveness of insulin treatments. *IEEE Transactions on Biomedical Engineering*, 65(6), 1281–1290. <https://doi.org/10.1109/TBME.2017.2746340>
- Woldaregay, A. Z., Årsand, E., Walderhaug, S., Albers, D., Mamykina, L., Botsis, T., & Hartvigsen, G. (2019). Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. *Artificial Intelligence in Medicine*, 98(April 2018), 109–134. <https://doi.org/10.1016/j.artmed.2019.07.007>
- Zhu, T., Li, K., Herrero, P., & Georgiou, P. (2021). Deep Learning for Diabetes: A Systematic Review. *IEEE Journal of Biomedical and Health Informatics*, 25(7), 2744–2757. <https://doi.org/10.1109/JBHI.2020.3040225>

Anexo A. Artículo de investigación

Rendimiento de modelos de aprendizaje profundo en escenarios con pacientes diabéticos reales y simulados

José Miguel Pereira Pontón

Universidad Internacional de la Rioja, Logroño (España)

20 de julio de 2023

RESUMEN

En este estudio, se investigó el desempeño de seis modelos de aprendizaje profundo para la predicción de glucosa en pacientes con Diabetes Mellitus Tipo 1, aplicado tanto en casos clínicos como en casos simulados. Se evaluaron diferentes configuraciones de capas RNN, LSTM y GRU, así como combinaciones de estas capas. Se analizó el impacto de emplear diferente tamaño de ventana histórica y diferente horizonte de predicción. Los resultados revelaron que las ventanas históricas de 60 y 90 minutos tuvieron un mejor rendimiento en la predicción, y el horizonte de predicción de 15 minutos fue el más eficiente. La combinación de la capa LSTM y GRU mostró el mejor resultado en términos de precisión, según el análisis del RMSE y del análisis de error de Clarke. Además, se observó que los modelos entrenados con datos reales pudieron alcanzar una ventaja significativa a los modelos entrenados con datos simulados.

PALABRAS CLAVE

Diabetes Mellitus Tipo 1, Glucosa, Aprendizaje profundo, Predicción, Redes Neuronales Recurrentes.

I. INTRODUCCIÓN

Se considera a la Diabetes Mellitus o simplemente denominada diabetes como un grupo de enfermedades metabólicas que se encuentran relacionadas por defectos en la secreción de insulina, la acción de la insulina o en su defecto por ambas condicionales [1].

La hormona insulina es producida en la célula del islote de Langerhans en el páncreas y se la considera como una hormona esencial. La insulina es de gran importancia para el metabolismo de las grasas y proteínas. Cuando se tiene escasas de producción de insulina, se tiende a tener altos niveles de glucosa (hiperglucemia).

La Federación Internacional de Diabetes, reportó en su décima edición del Atlas de Diabetes de 2021 que se tiene una prevalencia en el incremento global de personas con diabetes confirmada, un total de 537 millones de adultos se encontraban viviendo con diabetes [2]. Estiman que para el año 2045 se tenga un total de 784 millones de personas con diabetes a nivel global. Se encontró que se tiene una mayor prevalencia en adultos entre 20 y 79 años.

Se consideran diferentes tipos de diabetes, entre los que se destacan, Diabetes Mellitus Tipo 1 (DMT1) y Diabetes Mellitus Tipo 2 (DMT2). Las personas que padecen de

DMT1 necesitan de una o varias dosis de insulina diaria para regular la glucosa en la sangre. Mientras que en el caso de DMT2, es uno de los tipos de diabetes más común, se produce una hiperglucemia, esto se encuentra en función a que el cuerpo no es capaz de asimilar correctamente la insulina [3]. Dentro del presente estudio nos enfocaremos en las características y estimación de la glucosa en pacientes con DMT1.

La inclusión de la Inteligencia Artificial en el campo de la medicina ha mejorado diferentes aspectos, entre los que encontramos: prevención de enfermedades y diagnósticos de forma anticipada, interpretación de imágenes para el reconocimiento de patrones complejos, tratamiento debido a que la IA se la puede aplicar para predecir reacciones adversas ante diferentes tratamientos y como seguimiento, soporte y monitorización [4].

En el caso de la DMT1, la predicción de los niveles de glucosa en sangre es de vital importancia para el cuidado de los pacientes, sin embargo, la parte más complicada es la obtención de la información de los pacientes, debido a que puede ser un proceso costoso y en parte difícil de obtener. Entre los beneficios de la inteligencia artificial en el área de la medicina se encuentra: aliviando a los pacientes, reemplazando datos, separando por categorías de forma más eficiente y aumentando la cantidad de información [5]. El aumentar la información permite

mejorar la calidad, reducir los errores médicos, reducción de los sesgos, así como mejorar el conocimiento médico y una respuesta de intervención más rápida, con mejores pronósticos en los resultados.

Existen diferentes softwares que simulan el comportamiento del cuerpo humano en relación con el metabolismo de la glucosa, estos datos generados artificialmente pueden ser empleados para el entrenamiento de modelos. La principal ventaja del uso de datos sintéticos para la predicción de glucosa es la capacidad de generar grandes cantidades de datos precisos y variados [6].

Para realizar las predicciones sobre diferentes índoles en el campo de la inteligencia artificial se cuenta con diferentes técnicas y algoritmos. Al momento que se requiere obtener un mejor resultado ante comportamientos complejos, como es el caso de la predicción de la glucosa en la sangre, se ha invertido con un mayor énfasis en modelos de aprendizaje profundo. Se tienen una gama amplia de algoritmos, en este estudio se enfoca en la aplicación de algoritmos RNN, GRU, LSTM y una combinación de capas LSTM y GRU

II. ESTADO DEL ARTE

Diabetes Mellitus

Diabetes Mellitus se trata de una alteración metabólica, en donde su origen puede ser muy diverso, pero conlleva la existencia de alteraciones en la secreción de la insulina, la acción de la hormona o en algunos casos en ambas condiciones. La detección de la enfermedad cuando los síntomas son persistentes y cifras de la glucemia elevada son más rápidas de diagnosticar, mientras que se debe de reconocer que existen casos en donde las personas que padecen de la enfermedad son asintomáticas [7]. La hiperglucemia crónica en los casos de pacientes con diabetes se encuentra relacionada con daños a largo plazo, en donde los órganos que más se ven afectados se encuentran: ojos, riñones, corazón, nervios y los vasos sanguíneos [1].

Los casos de diabetes que se desarrollan, en general, en la población se encuentran divididos en dos categorías etiopatogénicas: Diabetes Mellitus Tipo 1 y Diabetes Mellitus Tipo 2, sin embargo, se conocen diferentes tipos de diabetes, las mismas que se abordarán más adelante.

Diabetes Mellitus Tipo 1

La Diabetes Mellitus Tipo 1 es la enfermedad que abarca cerca del 10% de los casos de diabetes que se registran a nivel mundial, esta enfermedad puede aparecer a cualquier edad, pero principalmente niños y adultos los que la padecen. Existe un riesgo elevado de mortalidad cuando la DMT1 se complica según se expone en [8]. Los autores en [9] indican que la DMT1 se encuentra aumentando rápidamente en regiones específicas y es variable en función de la población de las diferentes etnias. El diagnóstico de la DMT1 se realiza en función de la hiperglucemia.

En el estudio realizado por [10], se hace mención sobre los predictores de riesgo implícitos en DMT1. De los diferentes factores que se han propuesto, los más importantes son los factores genéticos, infecciosos, dietéticos y los factores serológicos. El enfoque del riesgo no se lo debería de considerar a estos factores de forma aislada, sino en conjunto.

Diabetes Mellitus Tipo 2

La Diabetes Mellitus Tipo 2 (DMT2) se trata del tipo de diabetes más común, la misma que afecta un 90% de los casos reportados en el mundo. En la DMT2, las células del cuerpo no responden en su totalidad a la insulina, se tiene una consideración de que se vuelve resistente a la insulina, en donde el resultado que se obtiene es la hiperglucemia. Los síntomas suelen ser similares a las que se tienen en la DMT1, con una diferencia en la cual, estos síntomas son menos dramáticos y en ocasiones son asintomáticos. Las causas de la DMT2 no son completamente entendidas, se las suele ligar con el sobrepeso y la obesidad, edad, etnia y en registros de la familia [2].

Análisis del error de Clarke

El análisis de cuadrícula de errores de Clarke se trata de un acercamiento con enfoque clínico, desarrollado para comparar los dispositivos de monitoreo a partir del valor de referencia y la precisión de sus valores, sin embargo, se lo emplea para realizar la evaluación del resultado que brinda un modelo de predicción [12]. Las regiones de Clarke se tratan de divisiones en un gráfico bidimensional de dispersión, lo que permite determinar la concordancia que existe entre las mediciones reales y las predicciones cuando se encuentra realizando la monitorización de la glucosa en la sangre. En el presente enfoque se compara la precisión de los diferentes modelos según la región en la que cae, las mismas regiones que puede observar en la Figura 1.

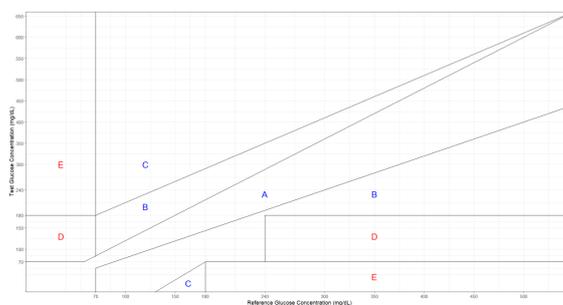


Fig 1. Gráfica de análisis de cuadrícula de errores de Clarke. Realizado mediante la librería EGA de R.

El significado de la cada una de las regiones las podemos describir de la siguiente forma según el autor [13].

- Zona A: Se dan medidas clínicamente precisas
- Zona B: Se dan errores benignos, pero no guían hacia un tratamiento inapropiado.
- Zona C: Se generan correcciones innecesarias, es decir, una corrección excedente de niveles aptos de glucosa.
- Zona D: Se generan errores potencialmente peligrosos. Esto se genera debido a una falla en la detección de hipoglucemia o hiperglucemia.
- Zona E: Se generan los errores en el tratamiento.

Las zonas A y B son clínicamente aceptables, mientras que las zonas C, D y E son clínicamente inaceptables.

Inteligencia Artificial en la medicina

La expansión e implementación de la Inteligencia Artificial en la Medicina [14] de forma directa y personalizada en los diferentes problemas han generado mejores oportunidades a los pacientes, debido a que los

modelos generados han permitido mejorar la precisión de los diagnósticos y la eficiencia de los procedimientos médicos. Uno de los campos que mayor impacto ha generado es el aprendizaje profundo, porque a diferencia de los algoritmos de ML, estos no requieren de manipulación del personal o ingresos de información. Según los autores [15] indican que una de las desventajas que existe en la inteligencia artificial aplicada en la medicina es la explicabilidad. Para poder sobrellevar este impedimento se requiere de tener causabilidad. La aplicación de modelos de AI/ML/DL en el campo de la medicina se trata de uno de los desafíos más grandes que existen.

Aprendizaje profundo

El aprendizaje profundo es uno de los subconjuntos del aprendizaje de máquinas, en el cual, los patrones de reconocimiento de información se encuentran relacionados e inspirados en el funcionamiento del cerebro humano. Este tipo de algoritmos se encuentran exentos de un sistema de reglas definidos por los humanos, en contraste hacen uso de información para generar las etiquetas correspondientes [17]. El aprendizaje profundo se encuentra relacionado con el empleo de diferentes números de capas dentro de los algoritmos para generar conocimiento.

Arquitectura RNN

En este caso, la arquitectura de la red emplea la información de una de las iteraciones para mejorar la predicción que tiene la salida de la siguiente neurona. Estas redes contemplan un problema, es que a medida que se genera un paso, en cada iteración se almacena la misma información, lo que implica que la relevancia de uno de los pasos previos se vaya perdiendo, este concepto se lo conoce desvanecimiento del gradiente.

Arquitectura LSTM

Una red LSTM (Long Short-Term Memory) hace uso de neuronas para determinar qué información es la que se almacena y cuál es la información que se olvida, esta consideración es en función del estado actual. Las redes LSTM poseen tres puertas de control, puerta de entrada, de salida y de olvido.

Arquitectura GRU

Las redes GRU (Gated-Recurrent-Units) se generaron para solventar el problema de la gradiente, su primera aparición realizado por [18]. Como lo indica el autor [19] esta arquitectura permite que cada una de las unidades recurrentes capture adaptativamente las dependencias a diferentes escalas de tiempo. Al igual que las LSTM posee puertas que regulan el flujo de la información, sin embargo, no posee una celda de memoria.

Simuladores de glucosa

Los simuladores del cuerpo humano se tratan de modelos matemáticos que tratan de imitar comportamientos específicos. El tener modelos para la simulación de la glucosa es altamente deseable para efectos de diseñar y evaluar sensores, generar un sistema de apoyo a la toma de decisiones, tal como es mencionado por [21]. Este enfoque radica en que aplicar un experimento o prueba a un usuario puede ser demasiado difícil, peligroso o puede caer en la parte no ética.

Los simuladores de glucosa son modelos matemáticos

que imitan el comportamiento de la glucosa en el cuerpo humano. Estos modelos brindan un gran soporte a los médicos e investigadores en el enfoque de la experimentación, debido a que permiten simular diferentes escenarios, estos escenarios se encuentran relacionados con un entorno seguro y controlado, en el cual, no se llega a poner en riesgo la salud de los pacientes. Una de las ventajas de los simuladores de glucosa es que se permiten realizar pruebas simuladas para sustituir a los pacientes reales [22]. Esto permite que se puedan explorar diferentes soluciones sin que se ponga en riesgo a los pacientes, debido a que no implica realizar ensayos clínicos, y a su vez se deriva en una reducción de costos al no aplicarlos.

En cuestión de los simuladores de DMT1 existentes, tenemos diferentes opciones, cada uno tiene sus ventajas y desventajas, así como la accesibilidad al mismo, unos son de código abierto. Entre estos simuladores se mencionan:

- UVA/PADOVA Type 1 Diabetes Simulator
- Bergman Minimal Model
- Cambridge DMT1
- McGill DMT1 Simulator

UVA/Padova Type 1 Diabetes Simulator

El simulador UVA/PADOVA, se trata de un modelo de larga escala en el cual se encuentra descrito el comportamiento dinámico de la glucosa, insulina y el glucagón mediante ecuaciones diferenciales. Este simulador fue aprobado en el 2008 por la Administración de Alimentos y Medicamentos - FDA [23]. Entre las principales aplicaciones del simulador tenemos dos:

- a) Simular y testear protocolos de estudios clínicos.
- b) Medir el impacto en el manejo y tratamiento de la diabetes (DMT1MS – The Epsilon Group, 2017).

El modelo de UVA/PADOVA emula el comportamiento de una población de 30 sujetos simulados, estos grupos se encuentran divididos equitativamente 10 niños, 10 adolescentes y 10 adultos para la versión del 2008 [25]. Siendo el modelo de UVA/PADOVA de código abierto, en sus versiones para MATLAB y Python.

III. OBJETIVOS Y METODOLOGÍA

El estudio se centra en comparar el desempeño en la predicción de glucosa a partir de una arquitectura de Aprendizaje Profundo entrenada con datos reales de pacientes con Diabetes Tipo 1 contra datos generados artificialmente mediante un simulador in-silico.

Para poder cumplir con el objetivo general del trabajo de investigación, se establecen los siguientes objetivos específicos:

- Investigar los datos relacionados con la enfermedad de Diabetes Mellitus y las diferentes variaciones.
- Describir el estado del arte de las arquitecturas de aprendizaje profundo orientado hacia la predicción de glucosa.
- Identificar los diferentes simuladores de glucosa para la generación del conjunto de datos simulados.
- Experimentar arquitecturas de aprendizaje profundo para el escenario de datos reales y para el escenario de datos artificiales.
- Evaluar la predicción de los dos escenarios a

partir de la arquitectura seleccionada.

- Comparar los resultados obtenidos por los diferentes modelos, interpretando los valores obtenidos para la obtención de conclusiones.

Para la obtención de los resultados del trabajo, se empleó la metodología CRISP-DM, tal como se la muestra en la Figura 2.

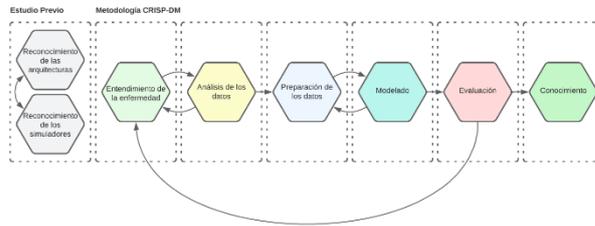


Fig 2. Metodología CRISP-DM aplicada en el estudio.

IV. CONTRIBUCIÓN

En el presente apartado se presentan las características y condiciones de los conjuntos de datos tanto reales como simulados. Se realizará la explicación del preprocesamiento empleado a ambos conjuntos. Con los datos tratados se explicará el proceso de selección que se empleó para ambos conjuntos de datos. Para finalizar el capítulo se muestran las configuraciones que se emplearon en los modelos. Los parámetros y las métricas empleadas para cuantificar el desempeño se encuentran plasmado dentro del capítulo.

Adquisición de datos

En el estudio se obtuvo la información desde dos fuentes diferentes:

- Naturaleza clínica – Información Real
- Naturales Artificial – Información Generada

Conjunto de información de Naturaleza Real

Los datos clínicos empleados en el presente estudio fueron obtenidos mediante la toma continua de pacientes con DMT1, los pacientes que participaron en el estudio pertenecen a la provincia de Granada España. Estos datos han sido parte de análisis para la evaluación de la glucemia en diferentes pacientes. Las mediciones de la glucosa fueron capturadas con el sensor FreeStyle Libre 2. Pueden existir variaciones en la información debido a las limitaciones del sensor. Este tipo de sensor se puede usar hasta 14 días, por lo cual se debe de tener consigo otro sensor que se encuentre disponible para poder suplir las lecturas de glucosa en caso de que termine su uso [34]. El conjunto de datos que contemplan las siguientes características:

- Se contiene un total de 110 pacientes que abordan un total de 3 millones de registros.
- Los registros se encuentran desde 2018-02-21 hasta 2022-03-21.
- Los registros contemplan una ventana de tiempo aproximada de 15 minutos. Sin embargo, los datos presentes en el conjunto de datos no presentan una distribución continua y contienen problemas de datos faltantes y discontinuidad.
- Las mediciones registradas se encuentran

descritas en un rango entre 40 [mg/dL] y 500 [mg/dL]. El promedio de las mediciones del conjunto de datos se encuentra en 159 [mg/dL]

Conjunto de información de Naturaleza Artificial

El simulador empleado en el estudio se trata de UVA/Padova T1D simulator, este simulador se encuentra aprobado por la FDA sobre las simulaciones del nivel de la glucosa. Se establecen las siguientes condiciones para los 10 pacientes simulados. De cada paciente se simulan 20 días de información. Las condiciones con las que se simularon los 10 pacientes presentes en el estudio se describen a continuación:

- Sensor: DEXCOM
- Bomba de Insulina: Insulet
- Controlador: BBController
- Escenario: Aleatorio
- Días Simulados: 30 días
- Pacientes: adult#001, adult#002, adult#003, adult#004, adult#005, adult#006, adult#007, adult#008, adult#009, adult#010

Procesamiento de los datos

Una vez descrita la naturaleza de los datos, se procede a establecer criterios para la selección de los datos acorde a la información que se maneja en el presente estudio. Estas características y consideraciones se emplearon en los dos conjuntos de datos. A partir del sensor usado con la muestra de los usuarios reales, se debe de tener en cuenta las siguientes consideraciones:

- Muestreo constante de 15 minutos: En función que el sensor utilizado genera muestras de usuarios reales con un muestreo de cada 15, se debe seleccionar aquellos usuarios que tengan una muestra lo más constante posible en este intervalo.
- Selección de muestras del simulador: Como el sensor empleado en el simulador muestrea valores cada 3 minutos, sensor Dexcom. Se deben seleccionar las muestras con una separación de 15 minutos entre las diferentes muestras.

El diseño del estudio tomó criterios del artículo descrito por [35], los mismo que se los explica a continuación:

- Perfil de glucosa: Se considera que 1 perfil tiene 96 muestras por día.
- Imputación de datos faltantes: Cuando existen datos faltantes en el perfil de glucosa, se imputan utilizando el valor promedio del valor anterior y el valor siguiente en el tiempo.
- Normalización de valores de glucosa: Los valores de glucosa se normalizan para facilitar el aprendizaje adecuado de la red neuronal, mediante un escalamiento de variables min-max.
- Ventana de histórico (VH): Es la secuencia de valores de glucosa que se utilizará como entrada para el modelo predictor.
- Horizonte de predicción (HP): El horizonte de predicción se refiere al instante en el que se realizará la predicción. La generación de la ventana de histórico y del horizonte de predicción se observa en la Figura 3.



Fig 3. Comportamiento de la ventana deslizante para la generación de los conjuntos de datos.

- División en subconjuntos: El conjunto de datos se divide en tres subconjuntos: entrenamiento, validación y prueba. Todos los perfiles de glucosa pertenecientes a un paciente se encontrarán exclusivamente en uno de los subconjuntos.
- Selección de pacientes y cantidad mínima de perfiles: Se seleccionan pacientes que tengan al menos 12 perfiles, lo que representa un mínimo de 1152 muestras. Entre estos pacientes, se asignan 6 pacientes al conjunto de entrenamiento, 2 al conjunto de validación y 4 al conjunto de prueba.
- Evaluación de modelos: La evaluación de los modelos predictores se realizará empleando la métrica del error cuadrático medio (RMSE). Así mismo se emplearán las gráficas de Clarke para evaluar la concordancia entre las mediciones reales y las predicciones realizadas por cada modelo.

Para el caso de los pacientes reales: De los 110 pacientes obtenidos inicialmente, se realizó una validación previa bajo el siguiente criterio de selección, tal como se aprecia en la Figura 4. Estos criterios de filtro y validación aseguran la calidad y la consistencia de los perfiles de glucosa que se emplean en el estudio.



Fig 4. Selección de conjunto de datos de pacientes reales.

De los 27 pacientes resultantes, se seleccionaron los 12. Estos pacientes que cumplieron con el requisito de tener perfiles de glucosa consecutivos fueron considerados en el estudio.

Se describen las cantidades de perfiles por conjunto:

1. Entrenamiento: 131 perfiles que hacen referencia a un total de 12567 registros.
2. Validación: 50 perfiles que hacen referencia a un total de 4800 registros.
3. Prueba: 61 perfiles que hacen referencia a un total de 5856 registros.

Para el caso de los pacientes simulados: Se ajusta la información a partir de la cantidad de perfiles que se obtuvieron en los conjuntos de datos de los pacientes reales.

V. EVALUACIÓN Y RESULTADOS

Modelo 1: Modelo de 1 capa RNN

Para analizar los resultados obtenido con el modelo 1, se describen la peor y la mejor condición tanto para

pacientes reales como para pacientes simulados. Esta condición se la generó para los dos horizontes de predicción. Estos resultados se los describen en la Tabla I.

Escenario	HP [min]	VH [min]	Condición	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
Reales	15	30	Peor	19.36	88.05%	11.49%	99.54%	0.00%	0.46%	0.00%
Reales	15	90	Mejor	11.46	97.30%	2.53%	99.83%	0.00%	0.16%	0.00%
Reales	30	30	Peor	20.52	86.59%	12.13%	98.72%	0.00%	1.28%	0.00%
Reales	30	45	Mejor	19.06	86.81%	10.82%	97.63%	0.00%	2.37%	0.00%
Simulados	15	30	Peor	12.74	94.22%	4.64%	98.86%	0.00%	1.14%	0.00%
Simulados	15	60	Mejor	11.39	95.27%	4.05%	99.32%	0.00%	0.68%	0.00%
Simulados	30	30	Peor	22.74	81.52%	17.06%	98.58%	0.00%	1.40%	0.02%
Simulados	30	90	Mejor	20.23	82.76%	14.60%	97.41%	0.00%	2.58%	0.02%

Tabla I. Métricas de las mejores y peores condiciones para los horizontes de predicción

Modelo 2: Modelo de 1 capa LSTM

Para analizar los resultados obtenido con el modelo 1, se describen la peor y la mejor condición tanto para pacientes reales como para pacientes simulados. Esta condición se la generó para los dos horizontes de predicción. Estos resultados se los describen en la Tabla II.

Escenario	HP [min]	VH [min]	Condición	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
Reales	15	30	Peor	19.19	88.67%	10.31%	98.98%	0.00%	1.02%	0.00%
Reales	15	90	Mejor	11.78	97.24%	2.60%	99.84%	0.00%	0.16%	0.00%
Reales	30	30	Peor	20.39	86.59%	11.41%	98.00%	0.00%	2.01%	0.00%
Reales	30	90	Mejor	18.57	88.40%	9.95%	98.35%	0.00%	1.65%	0.00%
Simulados	15	30	Peor	12.50	94.96%	4.32%	99.28%	0.00%	0.72%	0.00%
Simulados	15	90	Mejor	11.33	95.76%	3.77%	99.53%	0.00%	0.47%	0.00%
Simulados	30	30	Peor	21.83	82.82%	14.71%	97.53%	0.00%	2.43%	0.02%
Simulados	30	90	Mejor	20.30	83.07%	14.60%	97.79%	0.04%	2.26%	0.02%

Tabla II. Métricas de las mejores y peores condiciones para los horizontes de predicción

Modelo 3: Modelo de 2 capas LSTM

Para analizar los resultados obtenido con el modelo 1, se describen la peor y la mejor condición tanto para pacientes reales como para pacientes simulados. Esta condición se la generó para los dos horizontes de predicción. Estos resultados se los describen en la Tabla III.

Escenario	HP [min]	VH [min]	Condición	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
Reales	15	45	Peor	15.06	93.17%	3.38%	96.55%	0.00%	3.45%	0.00%
Reales	15	90	Mejor	10.64	97.56%	1.74%	99.30%	0.00%	0.69%	0.00%
Reales	30	30	Peor	18.82	90.40%	8.91%	99.31%	0.00%	0.69%	0.00%
Reales	30	60	Mejor	18.47	87.13%	8.92%	96.05%	0.00%	3.95%	0.00%
Simulados	15	30	Peor	12.24	95.66%	3.47%	99.13%	0.00%	0.88%	0.00%
Simulados	15	60	Mejor	11.14	96.06%	3.22%	99.28%	0.00%	0.72%	0.00%
Simulados	30	30	Peor	21.19	83.39%	14.19%	97.58%	0.00%	2.40%	0.02%
Simulados	30	60	Mejor	19.75	85.01%	13.36%	98.37%	0.00%	1.61%	0.02%

Tabla III. Métricas de las mejores y peores condiciones para los horizontes de predicción

Modelo 4: Modelo de 1 capa GRU

Para analizar los resultados obtenido con el modelo 1, se describen la peor y la mejor condición tanto para pacientes reales como para pacientes simulados. Esta condición se la generó para los dos horizontes de predicción.

Estos resultados se los describen en la Tabla IV.

Escenario	HP [min]	VH [min]	Condición	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
Reales	15	30	Peor	19.29	87.62%	10.93%	98.55%	0.00%	1.44%	0.00%
Reales	15	90	Mejor	11.53	97.24%	2.27%	99.51%	0.00%	0.49%	0.00%
Reales	30	30	Peor	20.34	86.13%	11.77%	97.90%	0.00%	2.10%	0.00%
Reales	30	90	Mejor	18.46	88.50%	9.46%	97.96%	0.00%	2.04%	0.00%
Simulados	15	30	Peor	11.92	93.86%	4.95%	98.81%	0.00%	1.19%	0.00%
Simulados	15	90	Mejor	11.27	95.43%	3.98%	99.41%	0.00%	0.60%	0.00%
Simulados	30	30	Peor	20.66	79.89%	16.52%	96.41%	0.00%	3.59%	0.00%
Simulados	30	60	Mejor	20.19	83.14%	14.20%	97.34%	2.65%	0.02%	0.02%

Tabla IV. Métricas de las mejores y peores condiciones para los horizontes de predicción

Modelo 5: Modelo de 2 capas GRU

Para analizar los resultados obtenido con el modelo 1, se describen la peor y la mejor condición tanto para pacientes reales como para pacientes simulados. Esta condición se la generó para los dos horizontes de predicción. Estos resultados se los describen en la Tabla V.

Escenario	HP [min]	VH [min]	Condición	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
Reales	15	30	Peor	13.18	96.98%	2.99%	99.97%	0.00%	0.03%	0.00%
Reales	15	90	Mejor	8.12	99.14%	0.72%	99.86%	0.00%	0.13%	0.00%
Reales	30	30	Peor	19.22	89.32%	9.24%	98.56%	0.00%	1.45%	0.00%
Reales	30	45	Mejor	17.53	89.92%	8.34%	98.25%	0.00%	1.75%	0.00%
Simulados	15	30	Peor	11.74	95.80%	3.73%	99.53%	0.00%	0.49%	0.00%
Simulados	15	60	Mejor	10.90	95.43%	3.77%	99.20%	0.00%	0.81%	0.00%
Simulados	30	30	Peor	21.00	84.10%	13.50%	97.60%	0.00%	2.38%	0.02%
Simulados	30	90	Mejor	19.55	83.79%	14.51%	98.30%	0.00%	1.68%	0.02%

Tabla V. Métricas de las mejores y peores condiciones para los horizontes de predicción

Modelo 6: Modelo de 1 capa LSTM y 1 capa GRU

Para analizar los resultados obtenido con el modelo 1, se describen la peor y la mejor condición tanto para pacientes reales como para pacientes simulados. Esta condición se la generó para los dos horizontes de predicción. Estos resultados se los describen en la Tabla VI.

Escenario	HP [min]	VH [min]	Condición	RMSE	Zona A (%)	Zona B (%)	Zona A + B (%)	Zona C (%)	Zona D (%)	Zona E (%)
Reales	15	45	Peor	12.6	96.02%	2.99%	99.02%	0.00%	0.99%	0.00%
Reales	15	90	Mejor	9.22	99.08%	0.86%	99.94%	0.00%	0.07%	0.00%
Reales	30	30	Peor	18.8	89.84%	9.50%	99.34%	0.00%	0.66%	0.00%
Reales	30	45	Mejor	17.5	89.39%	8.24%	97.63%	0.00%	2.37%	0.00%
Simulados	15	30	Peor	11.4	95.10%	4.10%	99.20%	0.00%	0.81%	0.00%
Simulados	15	60	Mejor	11.1	95.31%	3.87%	99.18%	0.00%	0.82%	0.00%
Simulados	30	30	Peor	20.2	82.50%	14.28%	96.78%	0.00%	3.21%	0.02%
Simulados	30	90	Mejor	19.2	84.83%	12.54%	97.37%	0.00%	2.61%	0.02%

Tabla VI. Métricas de las mejores y peores condiciones para los horizontes de predicción

Para completar el estudio se muestra la gráfica del mejor rendimiento se obtuvo entre todos los modelos para ambos escenarios

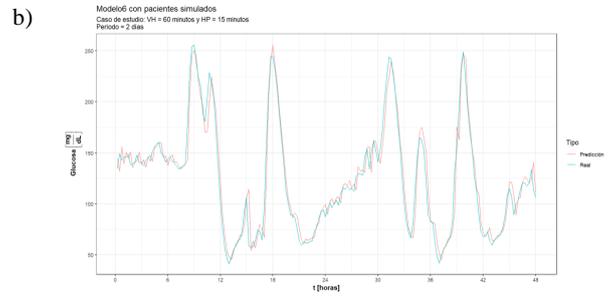
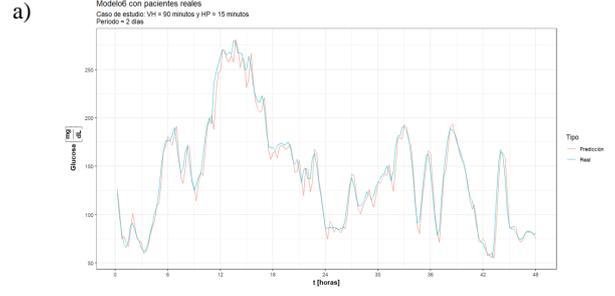


Fig 5. Mejores respuestas del Modelo 6. a) Pacientes reales. b) Pacientes simulados

En la Figura 6 se muestra el comportamiento de la gráfica de Clarke relacionados con los casos de estudio de la Figura 5,

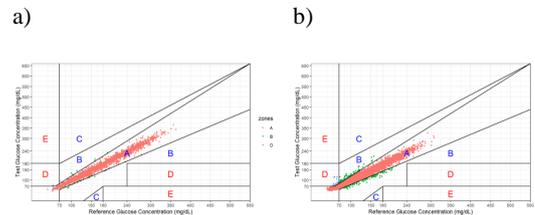


Fig 6. Gráficas de Clarke de las mejores respuestas del Modelo 6. a) Pacientes reales. b) Pacientes simulados.

VI. DISCUSIÓN

En el presente estudio, se llevó a cabo un análisis exhaustivo de diferentes modelos de predicción de glucosa utilizando datos de pacientes reales y simulados. Se emplearon modelos de redes neuronales recurrentes, específicamente capas, RNN, LSTM y capas GRU, tanto en configuraciones individuales como combinadas, para evaluar su desempeño en la predicción de los niveles de glucosa. Los resultados obtenidos revelaron varias tendencias y patrones.

En primer lugar, se observó que el tamaño de la ventana histórica tiene un alto impacto en la precisión de la predicción. Como caso puntual, se encontró que los mejores resultados de los modelos en ambos escenarios se encuentran con la ventana histórica de 60 y 90. Esto indica que la inclusión de más información histórica ayuda a que los modelos logren capturar de mejor forma las tendencias y los patrones de la glucosa en la sangre. Mientras que, si analizamos el efecto del horizonte de predicción, se logró una mejora significativa en la métrica del RMSE bajo un horizonte de 15 minutos, es decir que se tiene un mejor desempeño en la predicción a corto plazo, debido a que, para los modelos empleados, el capturar el

comportamiento de la serie para predecir 30 minutos en el futuro implica un mayor desafío.

En términos de las zonas de Clarke, se encontró que la mayoría de los modelos lograron un porcentaje alto de zonas clínicamente aceptadas, especialmente en el horizonte de predicción de 15 minutos. Sin embargo, se observó una mayor incidencia de puntos en la zona B para el horizonte de predicción de 30 minutos, lo cual sugiere que la predicción de valores moderadamente altos y bajos de glucosa puede llegar a ser más complicados de predecir en este horizonte de tiempo. Al observar el comportamiento general de las gráficas de Clarke, podemos apreciar que existe una mayor dispersión de los valores en el horizonte de predicción de 30 minutos que en el horizonte de predicción de 15 minutos.

Al comparar el desempeño de los diferentes modelos, emplear un modelo con una sola capa RNN, no se obtuvo resultados tan prometedores en términos de predicción de la glucosa. Los resultados fueron mejorando con la aplicación de las capas LSTM y capas GRU, a tal punto que, se demostró que se obtuvo un mejor desempeño en términos del RMSE y el análisis de las zonas de Clarke para el modelo que combinó una capa LSTM y una capa GRU.

En consideración del análisis de los pacientes simulados, el valor del RMSE no variaba mucho entre los diferentes modelos y las ventanas de histórico, sin embargo, con la muestra graficada de la serie temporal, se aprecia que las predicciones se van ajustando de mejor forma con los datos de la prueba. Es importante destacar que este estudio se limita a un conjunto de datos específico y a configuraciones particulares de modelos de redes neuronales recurrentes. Se requiere la aplicación de datos con mayor diversidad para poder generalizar de mayor forma.

VII. CONCLUSIONES

Utilizando el simulador UVA/PADOVA Type 1 Diabetes Simulator, se crearon los perfiles de glucosa de 10 pacientes adultos. Este enfoque nos permitió obtener un comportamiento similar al de pacientes reales sin comprometer la privacidad de los pacientes. El simulador proporciona una forma confiable de generar datos que permiten replicar el comportamiento dinámico de glucosa en el cuerpo. Al emplear este simulador se pudo evaluar los modelos de aprendizaje profundo en un entorno controlado.

Después de revisar los datos, se encontró una amplia variedad de combinaciones de arquitecturas de aprendizaje profundo utilizadas para predecir los valores de glucosa en estudios anteriores. Para este estudio, se optó por utilizar variaciones de la arquitectura RNN, teniendo en cuenta las consideraciones necesarias para el preprocesamiento y procesamiento de los datos con el objetivo de obtener los mejores resultados.

Se observó el efecto que tiene la ventana histórica sobre la predicción del nivel de la glucosa en la sangre. Se determinó con la experimentación que, al aumentar el tamaño de la ventana histórica, el valor del RMSE se ve reducido, obtenido los mejores resultados con ventanas de 60 y 90 minutos. El tener un mayor número de datos le permitió a los modelos capturar de mejor forma las tendencias y los patrones que poseen este tipo de series temporales. Según el horizonte de predicción que se

escoja, la evaluación del modelo se verá influenciada. Se encontró que con un horizonte de predicción de 15 minutos se obtiene mejores resultados en términos del RMSE y del análisis de error de Clarke, mientras que, con un horizonte de 30 minutos, el predecir la glucosa puede tornarse más desafiante para las redes.

El modelo que combinó la capa LSTM y la capa GRU, demostró tener el mejor desempeño en comparación de emplear este tipo de capas de forma individual. Este modelo incluso mejoró el ajuste de la predicción con el valor de la prueba con un horizonte de predicción de 30 minutos en ambos escenarios, tal como se aprecia en las figuras de las series temporales.

Se tiene planteado dos líneas de trabajo futuro. La primera línea de trabajo se encuentra la incorporación de variables clínicas adicionales que permitan enriquecer la capacidad de predicción de los modelos. Debido a que la DMT1 posee diferentes factores de influencia tal como se describió en el estado de arte. Se plantea considerar las variables como la cantidad de ingesta de alimentos, actividad física, medicamentos o inclusive los eventos de estrés. Estas variables implican una mayor complejidad al momento de recolectar la información.

La segunda línea de trabajo planteada es generar la evaluación de los modelos en tiempo real. Esto implica la integración de los modelos generados en sistemas de monitoreo continuo de glucosa. Este enfoque considera dos grandes desafíos, el primero que es la latencia entre que ocurre un evento y se produce la predicción, y el segundo reto presente es la adaptabilidad a cambios en los patrones de la glucosa en la sangre.

REFERENCIAS

- [1] A. D. ASSOCIATION, "Diagnosis and classification of diabetes mellitus," *Diabetes Care*, vol. 33, no. SUPPL. 1, 2010, doi: 10.2337/dc10-S062.
- [2] International Diabetes Federation, *International Diabetes Federation*, vol. 102, no. 2, 2021.
- [3] A. Petersmann et al., "Definition, Classification and Diagnosis of Diabetes Mellitus," *Exp. Clin. Endocrinol. Diabetes*, vol. 127, no. Suppl 1, pp. S1–S7, 2019, doi: 10.1055/a-1018-9078.
- [4] J. F. Ávila-Tomás, M. A. Mayer-Pujadas, and V. J. Quesada-Varela, "Artificial intelligence and its applications in medicine II: Current importance and practical applications," *Aten. Primaria*, vol. 53, no. 1, pp. 81–88, 2021, doi: 10.1016/j.aprim.2020.04.014.
- [5] Y. Y. M. Aung, D. C. S. Wong, and D. S. W. Ting, "The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare," *Br. Med. Bull.*, vol. 139, no. 1, pp. 4–15, 2021, doi: 10.1093/bmb/ldab016.
- [6] C. Pérez, "Propuesta de algoritmos de predicción de glucosa en pacientes diabéticos," p. 464, 2014, [Online]. Available: http://oa.upm.es/25604/1/MARIA_DEL_CARMEN_PEREZ_GANDIA.pdf.
- [7] D. I. Conget, "Diagnosis, classification and pathogenesis of diabetes mellitus," *Rev. Esp. Cardiol.*, vol. 55, no. 5, pp. 528–535, 2002, doi: 10.1016/S0300-8932(02)76646-3.
- [8] K. M. Gillespie, "Type 1 diabetes: Pathogenesis and prevention," *C. Can. Med. Assoc. J.*, vol. 175, no. 2, pp. 165–170, 2006, doi: 10.1503/cmaj.060244.
- [9] M. A. Atkinson and G. S. Eisenbarth, "Type 1 diabetes: New perspectives on disease pathogenesis and treatment," *Lancet*, vol. 358, no. 9277, pp. 221–229, 2001, doi: 10.1016/S0140-6736(01)05415-0.
- [10] M. Primavera, C. Giannini, and F. Chiarelli, "Prediction and Prevention of Type 1 Diabetes," *Front.*

- Endocrinol. (Lausanne)*, vol. 11, no. June, pp. 1–9, 2020, doi: 10.3389/fendo.2020.00248.
- [11] [1] A. D. ASSOCIATION, “Diagnosis and classification of diabetes mellitus,” *Diabetes Care*, vol. 33, no. SUPPL. 1, 2010, doi: 10.2337/dc10-S062.
- [12] [2] International Diabetes Federation, *International Diabetes Federation*, vol. 102, no. 2. 2021.
- [13] [3] A. Petersmann *et al.*, “Definition, Classification and Diagnosis of Diabetes Mellitus,” *Exp. Clin. Endocrinol. Diabetes*, vol. 127, no. Suppl 1, pp. S1–S7, 2019, doi: 10.1055/a-1018-9078.
- [14] [4] J. F. Ávila-Tomás, M. A. Mayer-Pujadas, and V. J. Quesada-Varela, “Artificial intelligence and its applications in medicine II: Current importance and practical applications,” *Aten. Primaria*, vol. 53, no. 1, pp. 81–88, 2021, doi: 10.1016/j.aprim.2020.04.014.
- [15] [5] Y. Y. M. Aung, D. C. S. Wong, and D. S. W. Ting, “The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare,” *Br. Med. Bull.*, vol. 139, no. 1, pp. 4–15, 2021, doi: 10.1093/bmb/ldab016.
- [16] [6] C. Pérez, “Propuesta de algoritmos de predicción de glucosa en pacientes diabéticos,” p. 464, 2014, [Online]. Available: http://oa.upm.es/25604/1/MARIA_DEL_CARMEN_PER_EZ_GANDIA.pdf.
- [17] [7] D. I. Conget, “Diagnosis, classification and pathogenesis of diabetes mellitus,” *Rev. Esp. Cardiol.*, vol. 55, no. 5, pp. 528–535, 2002, doi: 10.1016/S0300-8932(02)76646-3.
- [18] [8] K. M. Gillespie, “Type 1 diabetes: Pathogenesis and prevention,” *C. Can. Med. Assoc. J.*, vol. 175, no. 2, pp. 165–170, 2006, doi: 10.1503/cmaj.060244.
- [19] [9] M. A. Atkinson and G. S. Eisenbarth, “Type 1 diabetes: New perspectives on disease pathogenesis and treatment,” *Lancet*, vol. 358, no. 9277, pp. 221–229, 2001, doi: 10.1016/S0140-6736(01)05415-0.
- [20] [10] M. Primavera, C. Giannini, and F. Chiarelli, “Prediction and Prevention of Type 1 Diabetes,” *Front. Endocrinol. (Lausanne)*, vol. 11, no. June, pp. 1–9, 2020, doi: 10.3389/fendo.2020.00248.
- [21] [11] R. W. Beck, R. M. Bergenstal, L. M. Laffel, and J. C. Pickup, “Advances in technology for management of type 1 diabetes,” *Lancet*, vol. 394, no. 10205, pp. 1265–1273, 2019, doi: 10.1016/S0140-6736(19)31142-0.
- [22] [12] M. Mayo, L. Chepulis, and R. G. Paul, “Glycemic-aware metrics and oversampling techniques for predicting blood glucose levels using machine learning,” *PLoS One*, vol. 14, no. 12, pp. 1–19, 2019, doi: 10.1371/journal.pone.0225613.
- [23] [13] P. Kopecký *et al.*, “The use of continuous glucose monitoring combined with computer-based eMPC algorithm for tight glucose control in cardiosurgical ICU,” *Biomed. Res. Int.*, vol. 2013, 2013, doi: 10.1155/2013/186439.
- [24] [14] V. Kaul, S. Enslin, and S. A. Gross, “History of artificial intelligence in medicine,” *Gastrointest. Endosc.*, vol. 92, no. 4, pp. 807–812, 2020, doi: 10.1016/j.gie.2020.06.040.
- [25] [15] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, “Causability and explainability of artificial intelligence in medicine,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 9, no. 4, pp. 1–13, 2019, doi: 10.1002/widm.1312.
- [26] [16] S. Chowdhury and P. pratim Chakraborty, “Universal health coverage - There is more to it than meets the eye,” *J. Fam. Med. Prim. Care*, vol. 6, no. 2, pp. 169–170, 2017, doi: 10.4103/jfmpc.jfmpc.
- [27] [17] L. Alzubaidi *et al.*, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, vol. 8, no. 1. Springer International Publishing, 2021.
- [28] [18] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder–decoder approaches,” *Proc. SSST 2014 - 8th Work. Syntax. Struct. Stat. Transl.*, pp. 103–111, 2014, doi: 10.3115/v1/w14-4012.
- [29] [19] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” pp. 1–9, 2014, [Online]. Available: <http://arxiv.org/abs/1412.3555>.
- [30] [20] T. Bezdán and N. Bačanin Džakula, “Convolutional Neural Network Layers and Architectures,” no. July, pp. 445–451, 2019, doi: 10.15308/sinteza-2019-445-451.
- [31] [21] C. Dalla Man, R. A. Rizza, and C. Cobelli, “Mixed meal simulation model of glucose-insulin system,” *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, vol. 54, no. 10, pp. 307–310, 2006, doi: 10.1109/IEMBS.2006.260810.
- [32] [22] N. Rosales, “Plataforma de simulación de pacientes diabéticos para validación in silico de estrategias de control glucémico,” 2011.
- [33] [23] M. Vettoretti, A. Facchinetti, G. Sparacino, and C. Cobelli, “Type-1 diabetes patient decision simulator for in silico testing safety and effectiveness of insulin treatments,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 6, pp. 1281–1290, 2018, doi: 10.1109/TBME.2017.2746340.
- [34] [24] “T1DMS – The Epsilon Group,” 2017. <https://tegvirginia.com/software/t1dms-2014/> (accessed Apr. 15, 2023).
- [35] [25] C. Dalla Man, F. Micheletto, D. Lv, M. Breton, B. Kovatchev, and C. Cobelli, “The UVA/PADOVA type 1 diabetes simulator: New features,” *J. Diabetes Sci. Technol.*, vol. 8, no. 1, pp. 26–34, 2014, doi: 10.1177/1932296813514502.
- [36] [26] P. H. Colmegna, R. S. Sánchez Peña, H. De Battista, J. Bondía, and M. Basualdo, “Instituto Tecnológico de Buenos Aires Simulation & Control in Type 1 Diabetes,” 2015.
- [37] [27] A. Z. Woldaregay *et al.*, “Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes,” *Artif. Intell. Med.*, vol. 98, no. April 2018, pp. 109–134, 2019, doi: 10.1016/j.artmed.2019.07.007.
- [38] [28] T. Zhu, K. Li, P. Herrero, and P. Georgiou, “Deep Learning for Diabetes: A Systematic Review,” *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 7, pp. 2744–2757, 2021, doi: 10.1109/JBHI.2020.3040225.
- [39] [29] I. Contreras and J. Vehi, “Artificial intelligence for diabetes management and decision support: Literature review,” *J. Med. Internet Res.*, vol. 20, no. 5, pp. 1–21, 2018, doi: 10.2196/10775.
- [40] [30] J. Freiburghaus, A. Rizzotti-Kaddouri, and F. Albertetti, “A deep learning approach for blood glucose prediction of type 1 diabetes,” *CEUR Workshop Proc.*, vol. 2675, pp. 131–135, 2020.
- [41] [31] D. Y. Kim *et al.*, “Developing an individual glucose prediction model using recurrent neural network,” *Sensors (Switzerland)*, vol. 20, no. 22, pp. 1–15, 2020, doi: 10.3390/s20226460.
- [42] [32] Y. Marcus *et al.*, “Improving blood glucose level predictability using machine learning,” *Diabetes. Metab. Res. Rev.*, vol. 36, no. 8, 2020, doi: 10.1002/dmrr.3348.
- [43] [33] C. Liu *et al.*, “Long-term glucose forecasting using a physiological model and deconvolution of the continuous glucose monitoring signal,” *Sensors (Switzerland)*, vol. 19, no. 19, pp. 1–19, 2019, doi: 10.3390/s19194338.
- [44] [34] Abbott, “User ’ s Manual FreeStyle Libre 2,” 2020, [Online]. Available: <https://www.binsons.com/uploads/userfiles/files/document s/products/Libre 2 User Manual.pdf>.
- [45] [35] J. Carrillo-Moreno, C. Pérez-Gandía, R. Sendra-Arroz, G. García-Sáez, M. E. Hernando, and Á. Gutiérrez, “Long short-term memory neural network for glucose prediction,” *Neural Comput. Appl.*, vol. 33, no. 9, pp. 4191–4203, 2021, doi: 10.1007/s00521-020-05248-0.