



ESCUELA DE INGENIERÍA DE FUENLABRADA

GRADO EN INGENIERÍA BIOMÉDICA

TRABAJO FIN DE GRADO

Generación y Evaluación de Datos Sintéticos de Series Temporales de Glucosa Usando Métodos Basados en Redes Neuronales Generativas

Autor: Blanca Liceras Marcos

Tutor: Cristina Soguero Ruíz

Co-tutor: Cristian David Chuhig Muzo

Curso académico 2022/2023

©2023 Blanca Licerias Marcos

Algunos derechos reservados

Este documento se distribuye bajo la licencia “Atribución 4.0 Internacional” de Creative Commons disponible en: <https://creativecommons.org/licenses/by/4.0/deed.es>

*Let's go invent tomorrow rather than worrying
about what happened yesterday.*

– Steve Jobs

Agradecimientos

Este trabajo supone el cierre de una etapa, una de las más importantes de mi vida, el fin del grado. Es por ello, que en primer lugar, me gustaría agradecer a todos los profesores que han formado parte de este proceso ayudándome a crecer tanto personal, como académicamente. En especial, doy gracias a Cristina Soguero Ruíz y a Cristian David Chushig Muzo por haberme guiado, por su colaboración en este proyecto y por el tiempo dedicado.

Además, me gustaría agradecer a los amigos que me ha dado la universidad. Gracias por compartir penas, alegrías y por el apoyo durante estos años, esta etapa no hubiera sido lo mismo sin vosotros. Estoy muy orgullosa de todos nosotros y de por fin poder decir que somos ingenieros. Del mismo modo, agradecer a mis amigos de toda la vida y a mi familia por su apoyo y por estar ahí siempre que lo necesito y por haberme ayudado a evadirme y divertirme a lo largo de este proceso.

Gracias también a Alex, por haberme acompañado en cada momento de estos años. Por el cariño, por el apoyo y por creer en mí. Gracias por ser parte de mi día a día y haber hecho estos 5 últimos años mejores.

Y gracias, aunque gracias se quede corto, a mis padres. Gracias en primer lugar por haberme dado esta gran oportunidad que tan significativa ha sido tanto a nivel académico como para mi evolución como persona. Gracias por haber estado al pie del cañón, por consolarme en las derrotas y celebrar las victorias. Por el continuo apoyo y la predisposición constante a ayudar. Lo más importante, gracias por el cariño y comprensión. No ha sido fácil pero lo hemos conseguido.

Por último, me gustaría recordar a mis abuelos Paco y Carlos que sé que hoy me acompañan aunque no como me hubiera gustado y sé que estarían enormemente orgullosos de mí.

Resumen

En los últimos años, la prevalencia de la Diabetes Mellitus Tipo 1 (DMT1) ha aumentado considerablemente a nivel mundial. La DMT1 es una enfermedad crónica que destruye las células beta del páncreas, provocando una caída en la producción de insulina y una dependencia de insulina exógena. Los pacientes con DMT1 son más propensos a sufrir episodios de glucemia, los cuales pueden desencadenar en eventos clínicos graves. Los dispositivos de monitorización continua de glucosa (CGM, del inglés *Continuous Glucose Monitoring*) han sido uno de los avances más sobresalientes para la gestión de la DMT1, favoreciendo el control de niveles de glucosa, previniendo eventos adversos y mejorando la calidad de vida.

Los modelos basados en Aprendizaje Automático (ML, del inglés *Machine Learning*) han marcado un hito en la academia e industria debido a su rendimiento en aplicaciones predictivas. Estos modelos pueden ser usados para predecir valores de glucosa en diferentes horizontes temporales, permitiendo así prevenir episodios de glucemia. Uno de los grandes desafíos en la investigación clínica es la disponibilidad de datos de pacientes, pues generalmente estos están estrictamente regulados. Los avances de la investigación clínica se ven limitados por las restricciones de acceso a los datos. Los modelos basados en redes neuronales adversarias (GANs, del inglés *Generative Adversarial Networks*) permiten abordar este problema con la generación de datos sintéticos que capturan la distribución subyacente de los datos originales.

El presente TFG tiene como objetivo evaluar la efectividad y robustez de modelos ML para la generación de series temporales sintéticas con datos de glucosa obtenidas de dispositivos CGM. En concreto, se han considerado los modelos *Conditional Probabilistic AutoRegressive* y *Doppel GANs* (DGAN). Para la experimentación, tres bases de datos con series temporales de glucosa de pacientes con DMT1 fueron usadas. Para la evaluación de calidad de los datos sintéticos, las divergencias *Jensen-Shannon*, *Maximum Mean Discrepancy* (MMD) y la distancia *Dynamic Time Wrapping* (DTW) fueron consideradas. Adicionalmente, tanto las series temporales reales como las sintéticas fueron utilizadas en un escenario de predicción (*forecasting*) en una ventana temporal de 3 horas.

Los resultados experimentales mostraron que DGAN permite generar series sintéticas con alta similitud a series temporales reales de glucosa, obteniendo valores razonables de divergencias (*Jensen-Shannon*, MMD) y DTW. Además, de manera general, las series generadas por DGAN permitieron obtener mejores resultados en el escenario de *forecasting*. El presente TFG contribuye a la evaluación de modelos ML para la generación de series sintéticas, demostrando la buena calidad de estas series. Esto resulta prometedor pues permitiría predecir valores futuros de glucosa y así construir un sistema de alerta temprana de episodios de glucemia.

Índice general

Agradecimientos

Resumen

Índice de figuras **III**

Índice de tablas **XI**

Lista de acrónimos y abreviaturas **XIV**

1. Introducción y objetivos **1**

1.1. Contexto y motivación 1

1.2. Objetivos 3

1.3. Metodología 4

1.4. Estructura de la memoria 5

2. Conceptos previos y descripción de las bases de datos **7**

2.1. Conceptos previos 7

2.2. Análisis descriptivo y preprocesamiento de las bases de datos 11

2.2.1. BD1-SEV 12

2.2.2. BD2-PED 20

2.2.3. BD3-SEN 26

3. Modelos para generación de series temporales sintéticas y predicción de glucosa	33
3.1. Aprendizaje automático	33
3.2. Modelos de generación de series temporales sintéticas	36
3.2.1. <i>Conditional Probabilistic Autorregressive Model</i>	37
3.2.2. <i>Doppel Generative Adversarial Network</i>	40
3.2.3. Métricas para evaluar la calidad de series temporales sintéticas	42
3.3. Modelos de predicción en series temporales	43
3.3.1. Métricas para evaluar predicción en series temporales	46
4. Experimentos y resultados	51
4.1. Software utilizado	51
4.2. Generación de series temporales sintéticas	51
4.2.1. Generación de series temporales sintéticas usando BD1-SEV	52
4.2.2. Generación de series temporales sintéticas usando BD2-PED	56
4.2.3. Generación de series temporales sintéticas usando BD3-SEN	59
4.3. Predicción de niveles de glucosa	63
4.3.1. Predicción de niveles de glucosa usando BD1-SEV	63
4.3.2. Predicción de niveles de glucosa usando BD2-PED	68
4.3.3. Predicción de niveles de glucosa usando BD3-SEN	73
5. Conclusiones y líneas futuras	79
5.1. Conclusiones	79
5.2. Líneas futuras	80
A. Evaluación de series temporales sintéticas	83
B. Evaluación de predicción	87
Bibliografía	107

Índice de figuras

1.1. Diagrama de Gantt de las tareas realizadas para la consecución de este TFG. . .	5
2.1. Evolución de la prevalencia de la diabetes por continente y estimaciones según la FID.	8
2.2. Prevalencia de la diabetes por edad y sexo en el año 2019.	8
2.3. Eventos de hiperglucemia e hipoglucemia severas entre los años 2011 y 2020 en Estados Unidos.	9
2.4. Ejemplo de serie temporal con 4 días de registro de glucosa para 1 paciente y con frecuencia de muestreo de 5 minutos.	11
2.5. Visualización de una serie temporal de glucosa para el mismo paciente de la BD1-SEV durante distintos días.	14
2.6. Visualización de una serie temporal de glucosa durante un día para distintos pacientes de la base de datos BD1-SEV.	14
2.7. Suma de muestras de glucosa de los pacientes de la BD1-SEV recogidas cada día.	15
2.8. Número total de muestras por paciente de la BD1-SEV.	16
2.9. Número total de valores perdidos por paciente de la BD1-SEV.	17
2.10. Resultados de aplicar diferentes técnicas de imputación para una serie temporal de glucosa de BD1-SEV.	19
2.11. Visualización de una serie temporal de glucosa para el mismo paciente de la BD2-PED durante distintos días.	21
2.12. Visualización de una serie temporal de glucosa durante un día para distintos pacientes de la BD2-PED durante distintos días.	21
2.13. Suma de muestras de los pacientes de la BD2-PED recogidas cada día.	22

2.14. Número total de muestras por paciente de la BD2-PED y su media.	23
2.15. Número total de valores perdidos por paciente de la BD2-PED.	24
2.16. Resultados de aplicar diferentes técnicas de imputación para una serie temporal de glucosa de BD2-PED.	25
2.17. Visualización de serie temporales de glucosa para un paciente de la BD3-SEN durante diferentes días.	27
2.18. Visualización de series temporales con datos de glucosa de un día asociadas con distintos pacientes de la BD3-SEN.	28
2.19. Suma de muestras de los pacientes de la base de datos 3 recogidas cada día. . .	29
2.20. Número total de muestras por paciente de la BD3-SEN y su media.	29
2.21. Número total de valores perdidos por paciente de la BD3-SEN.	30
2.22. Resultados de aplicar diferentes técnicas de imputación para una serie temporal de glucosa de BD3-SEN.	31
3.1. Ejemplo de conexión entre capas de una red neuronal con capas ocultas.	35
3.2. Esquema del funcionamiento de las dos partes de las redes GAN, generador y discriminador.	36
3.3. Diagrama de trabajo de la red neuronal CPAR.	38
3.4. Entrenamiento por filas del modelo CPAR.	39
3.5. Entrenamiento de redes neuronales para generación de series temporales sintéticas.	40
3.6. Estructura general de DGAN.	41
3.7. Estructura general de N-BEATS.	44
3.8. Estructura general de N-HiTS.	46
3.9. Series temporales estimadas por distintos <i>stacks</i> y su variación en el número de muestras debido a la interpolación jerárquica.	46
3.10. Distribución de las zonas de riesgo de CEG.	48
3.11. Rangos de aceptación para las mediciones de glucosa en sangre por dispositivos CGM según la ISO 2003 y ISO 2015.	49

4.1. Histogramas de la distribución de datos de glucosa de las series temporales reales de un paciente de la BD1-SEV y series sintéticas generadas con DGAN y CPAR.	52
4.2. Resultados de divergencia de <i>Jensen-Shannon</i> entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD1-SEV y series sintéticas generadas con DGAN y CPAR.	53
4.3. Resultados de MMD entre distribución de datos de glucosa de las series temporales reales de la base de datos BD1-SEV y series sintéticas generadas con DGAN y CPAR.	54
4.4. p_{valor} de test de <i>Kolmogorov-Smirnov</i> entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD1-SEV y series sintéticas generadas con DGAN y CPAR.	55
4.5. Distancia DTW entre series temporales de glucosa reales de la base de datos BD1-SEV y series sintéticas generadas con DGAN y CPAR.	55
4.6. Histogramas de la distribución de datos de glucosa de las series temporales de un paciente reales de la BD2-PED y series sintéticas generadas con DGAN y CPAR.	56
4.7. Resultados de divergencia de <i>Jensen-Shannon</i> entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD2-PED y series sintéticas generadas con DGAN y CPAR.	57
4.8. Resultados de MMD entre distribución de datos de glucosa de las series temporales reales de la base de datos BD2-PED y series sintéticas generadas con DGAN y CPAR.	58
4.9. p_{valor} de test de <i>Kolmogorov-Smirnov</i> entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD2-PED y series sintéticas generadas con DGAN y CPAR.	58
4.10. Distancia DTW entre las series temporales de glucosa reales de la base de datos BD2-PED y las series sintéticas generadas con DGAN y CPAR.	59
4.11. Histogramas de la distribución de datos de glucosa de las series temporales de un paciente reales de la BD3-SEN y series sintéticas generadas con DGAN y CPAR.	60

4.12. Resultados de divergencia de <i>Jensen-Shannon</i> entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD3-SEN y series sintéticas generadas con DGAN y CPAR.	60
4.13. Resultados de MMD entre distribución de datos de glucosa de las series temporales reales de la base de datos BD3-SEN y series sintéticas generadas con DGAN y CPAR.	61
4.14. p_{valor} de test de <i>Kolmogorov-Smirnov</i> entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD3-SEN y series sintéticas generadas con DGAN y CPAR.	62
4.15. Distancia DTW entre las series temporales de glucosa reales de la base de datos BD3-SEN y series sintéticas generadas con DGAN y CPAR.	62
4.16. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD1-SEV real para el decimonoveno paciente con PtId 9 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora.	66
4.17. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD1-SEV sintética generada por el método DAGN para el decimonoveno paciente con PtId 9 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora.	67
4.18. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD1-SEV sintética generada por el método CPAR para el PtId=9 con una ventana de predicción de 12 muestras.	68
4.19. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED real para el tercer paciente con PtId 58 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	71
4.20. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED sintética generada por el método DAGAN para el tercer paciente con PtId 58 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	72
4.21. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED sintética generada por el método CPAR para el tercer paciente con PtId 58 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	73

4.22. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD3-SEN sintética generada por el método CPAR para el segundo paciente con PtId 166 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	76
4.23. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD3-SEN sintética generada por el método DGAN para el segundo paciente con PtId 166 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	77
4.24. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD3-SEN sintética generada por el método CPAR para el segundo paciente con PtId 166 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	78
A.1. Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD1-SEV y series sintéticas generadas con DGAN y CPAR.	83
A.2. Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD1-SEV y series sintéticas generadas con DGAN y CPAR.	84
A.3. Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD2-PED y series sintéticas generadas con DGAN y CPAR.	84
A.4. Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD2-PED y series sintéticas generadas con DGAN y CPAR.	85
A.5. Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD3-SEN y series sintéticas generadas con DGAN y CPAR.	85
A.6. Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD3-SEN y series sintéticas generadas con DGAN y CPAR.	86
B.1. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD1-SEV real para el paciente con PtId 202 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora.	88
B.2. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD1-SEV real para el paciente con PtId 21 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora.	89

B.3. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD1-SEV sintética generada por el método DGAN para el paciente con PtId 202 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora.	90
B.4. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD1-SEV sintética generada por el método DGAN para el paciente con PtId 21 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora.	91
B.5. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD1-SEV sintética generada por el método CPAR para el paciente con PtId 202 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora.	92
B.6. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD1-SEV sintética generada por el método CPAR para el paciente con PtId 21 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora.	93
B.7. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED real para el paciente con PtId 53 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	94
B.8. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED real para el paciente con PtId 86 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	95
B.9. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED sintética generada por el método DAGAN para el paciente con PtId 53 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	96
B.10. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED sintética generada por el método DAGAN para el paciente con PtId 86 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	97

B.11. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED sintética generada por el método CPAR para el paciente con PtId 53 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	98
B.12. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED sintética generada por el método CPAR para el paciente con PtId 86 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	99
B.13. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD3-SEN sintética generada por el método CPAR para el paciente con PtId 151 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	100
B.14. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD3-SEN sintética generada por el método CPAR para el paciente con PtId 57 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	101
B.15. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD3-SEN sintética generada por el método DGAN para el paciente con PtId 151 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	102
B.16. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD3-SEN sintética generada por el método DGAN para el paciente con PtId 57 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	103
B.17. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD3-SEN sintética generada por el método CPAR para el paciente con PtId 151 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	104
B.18. Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD3-SEN 2.2.3 sintética generada por el método CPAR para el paciente con PtId 57 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas.	105

Índice de tablas

2.1.	Descripción de las bases de datos de glucosa empleadas en el desarrollo del TFG.	12
2.2.	Ejemplo de datos de la BD1-SEV.	13
2.3.	Ejemplo de datos de la BD2-PED.	20
2.4.	Ejemplo de datos de la BD3-SEN.	26
2.5.	Recopilación del número de pacientes, número de registros totales muestreados cada 5 minutos después del preprocesamiento para cada base de datos.	32
3.1.	Descripción de las características de las distintas ANNs para la generación de series temporales sintéticas estudiadas.	37
4.1.	Figuras de mérito obtenidas para los distintos pacientes de la BD1-SEV reales, generados con DGAN y generados con CPAR tras realizar la predicción con el modelo N-HiTS.	64
4.2.	Figuras de mérito obtenidas para los distintos pacientes de la BD2-PED reales, generados con DGAN y generados con CPAR tras realizar la predicción con el modelo N-HiTS.	69
4.3.	Errores obtenidos para los distintos pacientes de la BD3-SEN reales, generados con DGAN y generados con CPAR tras realizar la predicción con el modelo N-HiTS.	74

Lista de acrónimos y abreviaturas

AA Aprendizaje Automático

ANN *Artificial Neural Network*

BD1-SEV Base de datos 1: *Severe hypoglycemia in older adults with type 1 diabetes*

BD2-PED Base de datos 2: *The International Diabetes Closed Loop trial: Clinical Acceptance of the Artificial Pancreas in pediatrics*

BD3-SEN Base de datos 3: *Wireless Innovation for Seniors with Diabetes Mellitus*

BGM *Blood Glucose Monitoring*

CDF *Cumulative Distribution Function*

CEG *Clarke Error Grid*

CGM *Continous Glucose Monitoring*

CPAR *Conditional Probabilistic AutoRegressive*

DGAN *Doppel Generative Adversarial Network*

DL *Deep Learning*

DMT1 Diabetes Mellitus Tipo 1

DMT2 Diabetes Mellitus Tipo 2

DM Diabetes Mellitus

DTW *Dynamic Time Warping*

FID Federación Internacional de Diabetes

GAN *Generative Adversarial Network*

IA *Inteligencia Artificial*

KNN *K-Nearest Neighbour*

LSTM *Long Short Term Memory*

MAE *Mean Absolute Error*

MAPE *Mean Absolute Percentage Error*

ML *Machine Learning*

MMD *Maximum Mean Discrepancy*

MRAE *Mean Relative Absolute Error*

MSE *Mean Squared Error*

N-BEATS *Neural Basis Expansion Analysis for interpretable Time Series forecasting*

N-HiTS *Neural Hierarchical Interpolation for Time Series forecasting*

PDF *Probability Density Function*

RF *Random Forest*

TFG *Trabajo Fin de Grado*

Capítulo 1

Introducción y objetivos

Este capítulo presenta el contexto y motivación del desarrollo del presente Trabajo de Fin de Grado (TFG). Además, constituye un preámbulo al problema de generación de series temporales sintéticas usando datos de glucosa y presenta los objetivos, la metodología seguida y una breve estructura de la memoria.

1.1. Contexto y motivación

La Diabetes Mellitus (DM) es una de las enfermedades más graves y comunes a día de hoy que afecta a un alto número de personas a nivel mundial [1]. En el año 2015, alrededor de 5 millones de muertes fueron atribuidas a esta enfermedad [2] y en 2019, se estimó que la DM afectó a 463 millones de personas globalmente [1]. Los dos tipos de DM más prevalentes son la DM Tipo 1 (DMT1) y DM Tipo 2 (DMT2). La DMT1 es una enfermedad autoinmune que afecta mayoritariamente a niños y jóvenes, y que en 2020 afectó aproximadamente a un 9.5% de la población mundial [3]. Por otro lado, la DMT2 es causada por la resistencia a la insulina y está asociada a un 90-95% de los casos de DM [4]. Según varios estudios [2], la mitad de los pacientes que padecen diabetes no lo saben y por ello es más probable que sufran complicaciones debido a esta enfermedad.

Además de que la DM es una gran causa de mortalidad, es también una fuente de morbilidad que da lugar a diversas complicaciones microvasculares y macrovasculares [2]. Las complicaciones microvasculares son aquellas en las que los microvasos cardíacos se ven comprometidos reduciendo su capacidad de contracción [5]. Esto puede dar lugar a una disfunción ventricular, y está asociada a múltiples patologías tales como la retinopatía diabética, nefropatía o neuropatía

tía [5]. Por otro lado, las complicaciones macrovasculares están relacionadas en el proceso de arterosclerosis, que ocasiona un ensanchamiento en las paredes de arterias, aumentando el riesgo de muerte para pacientes diabéticos por enfermedad cardiovascular [5, 6]. Diversos estudios clínicos [6] han demostrado que el riesgo de infarto de miocardio para pacientes diabéticos es el mismo que para pacientes no diabéticos con antecedentes de infarto de miocardio [6]. Además, los pacientes diabéticos tienen un aumento de 150-400 % respecto al resto de la población de riesgo de accidente cerebrovascular, aumentando significativamente el riesgo de demencia y muerte [6].

Unido a esto, los pacientes que padecen DM tienden a sufrir eventos de hipoglucemia e hiperglucemia [7]. La hipoglucemia e hiperglucemia se caracterizan por niveles bajos de glucosa en sangre (< 70 mg/dl) y niveles de glucosa altos (> 180 mg/dl), respectivamente [8]. La hiperglucemia puede ocasionar complicaciones agudas tales como cetoacidosis y coma hiperosmolar, ambas potencialmente mortales [9]. Mientras que las complicaciones de la hipoglucemia incluyen daño neurológico, trauma, eventos cardiovasculares y muerte [7]. Estos episodios de glucemia dependen de diversos factores como el peso, el estrés, la cantidad de insulina administrada, ingesta de calorías, ejercicio realizado entre otras [8,9]. Es muy importante prevenir estos episodios de cara a evitar el desarrollo de complicaciones mayores, además de los riesgos a corto plazo producidos tal como mareos o sensación de debilidad [10]. Para un control eficiente de los niveles de glucemia el uso de dispositivos de monitorización continua de glucosa (CGM, del inglés *Continuous Glucose Monitoring*) es vital. Estos dispositivos surgen a finales de los años 90 [11], y su principal función es medir el nivel de glucemia en sangre con una frecuencia constante de 5 minutos [12]. El correcto uso de estos dispositivos permitiría un control óptimo de niveles de glucosa, previniendo posibles eventos clínicos y mejorando la calidad de vida de individuos diagnosticados con DM.

En los últimos años, los modelos basados en Aprendizaje Automático (ML, del inglés *Machine Learning*) han marcado un hito en la academia e industria debido a sus altas prestaciones en aplicaciones predictivas, en concreto han destacado modelos basados en Redes Neuronales Profundas (en inglés *Deep Neural Networks*) y en Máquinas de vectores de soporte (en inglés *Support Vector Machines*) [13]. En la investigación clínica, los modelos ML se han usado para predicción de hospitalización, estimación de riesgo de mortalidad, identificación de factores de riesgo, entre otros [14]. Referente a la diabetes, diferentes autores han explorado el uso de datos de dispositivos CGM para predecir episodios de hipoglucemia e hiperglucemia en diferentes horizontes temporales (1 hora, 2 horas, 4 horas) [15, 16]. El uso de modelos ML y datos de dispositivos CGM tienen el potencial de permitir predicciones de los valores próximos de glucemia (debido a su muestreo a frecuencia constante) [17]. Esto es prometedor porque

la aplicación de estos modelos puede ayudar a prevenir eventos de hipoglucemia e hiperglucemia [18]. Uno de los grandes desafíos al trabajar con datos clínicos, como los registros de concentración de glucosa en sangre, es la escasa disponibilidad de los mismos. Al ser datos de carácter personal, su uso y disponibilidad están regulados por el *Reglamento General de Protección de Datos* [19]. Los avances de la investigación clínica se ven limitados por el acceso a datos libres para solventar diferentes problemas clínicos (aún sin resolver). En la literatura, diferentes autores han considerado modelos basados en redes neuronales artificiales (ANNs, del inglés *Artificial Neural Networks*) para la generación de series temporales sintéticas con el fin de abordar este problema [20, 21]. Estos modelos buscan crear datos sintéticos lo más parecido posible a los originales, capturando la distribución subyacente de los datos y la dependencia temporal [21].

El presente TFG tiene como objetivo evaluar la efectividad y robustez de modelos basados en ANNs para la generación de series temporales sintéticas con datos de glucosa obtenidos de dispositivos CGM. En concreto, se han considerado los modelos *Conditional Probabilistic AutoRegressive* (CPAR) y *Doppel Generative Adversarial Network* (DGAN). Para el desarrollo del TFG, se han utilizado tres bases de datos con series temporales de glucosa de pacientes con DMT1. Estas bases están constituidas con información de múltiples pacientes diabéticos en diferentes franjas de edad (6-13 años, > 60 años) que tienen implantado un dispositivo CGM, el cual registra y almacena el valor de concentración de glucosa en sangre cada 5 minutos.

1.2. Objetivos

El objetivo principal de este TFG es generar series temporales sintéticas de concentración de glucosa en sangre a partir de las tres bases de datos disponibles con series temporales reales recopiladas a través de dispositivos CGM. Se generan series temporales sintéticas mediante dos métodos y se comparan los resultados obtenidos a través de cada uno de ellos con los datos reales utilizando métricas de evaluación basadas en distribuciones y distancias. Además se busca evaluar el rendimiento de las series temporales sintéticas en tareas predictivas comparado con los resultados obtenidos tras someter las series reales al mismo experimento. Con esto se busca entender si la calidad de los datos generados es o no buena para los problemas de predicción y si estos nuevos datos podrían, por tanto, ayudar a la predicción de valores de glucemia con el fin de controlar eventos de hipo e hiperglucemias. Para la consecución de este objetivo, se han llevado a cabo diferentes objetivos específicos:

- Entrenar modelos basados en ANNs para generación de series temporales sintéticas a

partir de datos de glucosa de pacientes DMT1. En concreto, los modelos CPAR y DGAN, con el objetivo de poder aumentar el número de pacientes disponible en las bases de datos.

- Evaluar la calidad de las muestras sintéticas, proponiendo distintas métricas, para determinar así cuál de los métodos utilizados genera series temporales más similares a las reales.
- Realizar una predicción de los valores de glucosa, en los conjuntos originales y sintéticos con el fin de validar la bondad de las series temporales sintéticas. El modelo de generación de series temporales sintéticas que muestre buenos resultados en las tareas predictivas es el que convendría usar para predecir eventos de hipoglucemia o hiperglucemia.
- Interpretar los resultados obtenidos para determinar los beneficios que se consiguen al aplicar las técnicas de generación de series temporales sintéticas.

1.3. Metodología

Para alcanzar los objetivos descritos anteriormente, se ha llevado a cabo la siguiente metodología:

- Revisar la literatura sobre la prevalencia de la diabetes y el estado actual de la enfermedad a nivel nacional e internacional.
- Buscar y seleccionar bases de datos públicas con mediciones de valores de glucosa obtenidos con dispositivos CGM para pacientes con DMT1.
- Revisar la literatura sobre las ANNs para generación de series temporales sintéticas.
- Revisar la literatura sobre las métricas de calidad para evaluar el rendimiento de modelos de generación de series temporales sintéticas.
- Realizar un análisis descriptivo y exploratorio de las bases de datos para una mejor comprensión de los datos recopilados por los dispositivos CGM visualizando la alta variabilidad que presentan las series temporales de glucosa.
- Realizar un preprocesamiento de las series temporales, para optimizar el rendimiento de los métodos de generación de muestras sintéticas. Para ello, se hace un tratamiento de los valores faltantes para mantener la frecuencia de muestreo constante y se interpolan dichos valores probando diferentes métodos de interpolación.

- Implementar y evaluar diferentes modelos de generación de series temporales sintéticas. Se eligen finalmente los modelos CPAR y DGAN.
- Implementar métricas de calidad para evaluar los datos sintéticos generados.
- Implementar modelos de predicción sobre el conjunto de datos reales y sintéticos, a partir de modelos de regresión basados en ANNs.

En la Figura 1.1 se presenta un diagrama de Gantt que detalla las diferentes tareas realizadas durante el presente TFG. Cabe mencionar que es una representación a modo orientativa, y en todo momento se ha intentado establecer el desarrollo temporal de cada tarea con la mayor exactitud posible.

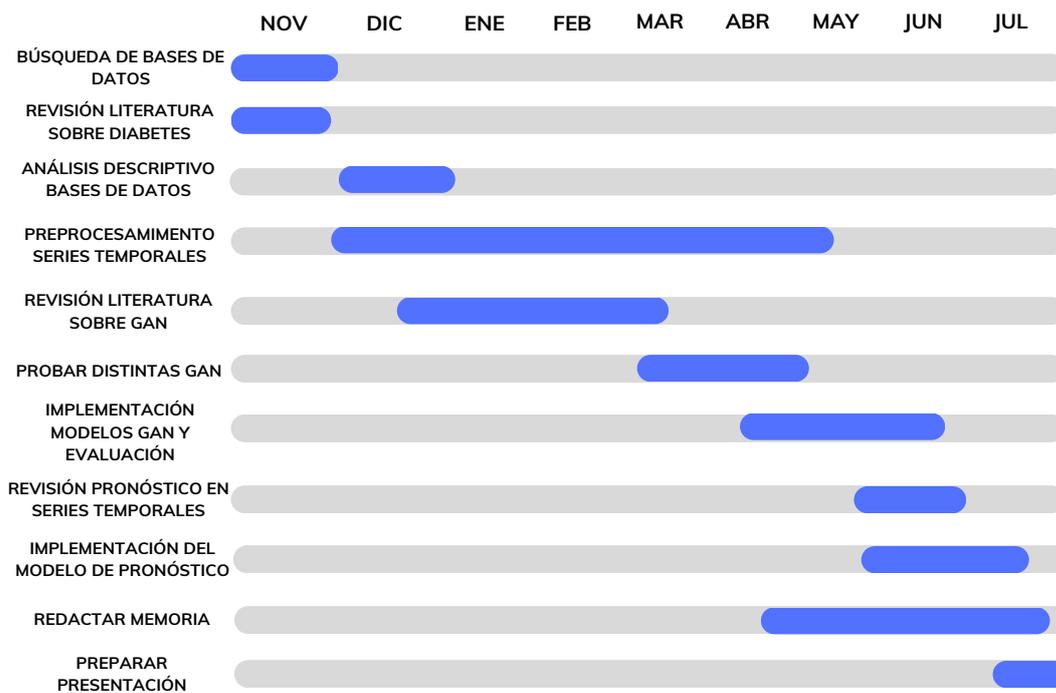


Figura 1.1: Diagrama de Gantt de las tareas realizadas para la consecución de este TFG.

1.4. Estructura de la memoria

A continuación se describe el contenido incluido en cada capítulo de la memoria:

- **Capítulo 1: Introducción y objetivos.** Se trata de un capítulo inicial compuesto por cuatro apartados. En el primero, se introduce el contexto y la motivación del presente trabajo. En el segundo, se presentan los objetivos que se quieren abordar con este estudio y en el tercero se describe la metodología empleada para llevar a cabo dichos objetivos. En el último, se presenta la estructura de la memoria de este TFG dividida por capítulos.
- **Capítulo 2: Conceptos previos y descripción de las bases de datos.** En primer lugar, se introduce la enfermedad de la diabetes así como las series temporales. A continuación, se describen las tres bases de datos seleccionadas de series temporales de glucosa y se explica el preprocesamiento llevado a cabo para cada una de ellas.
- **Capítulo 3: Modelos de generación y predicción de series temporales.** Se describen los conceptos necesarios para la adecuada comprensión de esta memoria. En concreto se realiza una breve introducción del Aprendizaje Automático (AA) y sus métodos y a continuación se explican en detalle los métodos de remuestreo haciendo especial hincapié en los dos métodos utilizados: CPAR y DGAN y los métodos de pronóstico en series temporales explicando detalladamente el modelo empleado en el presente TFG: *Neural Hierarchical Interpolation for Time Series Forecasting* (N-HITS). Por último, se explican las métricas utilizadas para evaluar la generación y el pronóstico de series temporales.
- **Capítulo 4: Experimentos y resultados.** En este capítulo se muestran los resultados obtenidos de la generación de series temporales sintéticas de glucosa evaluados mediante cuatro métricas diferentes: divergencia de Jensen Shannon, distancia de Kolmogorov-Smirnov, *Maximum Mean Discrepancy* y *Dynamic Time Warping*. Además se muestran los resultados obtenidos por el pronóstico de glucosa en series temporales reales y sintéticas (generadas con DGAN y CPAR). Se consideran varias métricas de error para validar este método, así como el test de *Clarke error grid* y la norma ISO 15197:2015.
- **Capítulo 5: Conclusiones y líneas futuras.** En este capítulo se discuten los resultados obtenidos y se sintetizan en forma de conclusión. Además, se comentan posibles líneas futuras de investigación derivadas del presente TFG

Capítulo 2

Conceptos previos y descripción de las bases de datos

Este capítulo consta de dos partes. En la primera parte se explican los conceptos necesarios para comprender este TFG, sobre diabetes y series temporales. En la segunda parte, se describe las bases de datos de series temporales de glucosa consideradas en el TFG, así como una descripción de las etapas de preprocesamiento seguidas y el análisis exploratorio.

2.1. Conceptos previos

La DM es una enfermedad crónica que afecta a millones de personas a nivel mundial, cuya prevalencia aumenta de manera exponencial. Según la Federación Internacional de Diabetes (FID), actualmente viven en el mundo 537 millones de adultos con diabetes [1]. Esto es un 16% más que lo estimado para el año 2019 [1], y se se espera que aumente hasta 629 millones para 2045 [22]. La prevalencia de diabetes varía según país y continente, siendo individuos de países en vías de desarrollo lo más afectados (véase la Figura 2.1). Esta condición es más prevalente en hombres y mujeres de entre 60 y 79 años [1], como se ve en la Figura 2.2, aunque afecta a adultos y niños de todas las edades [1].

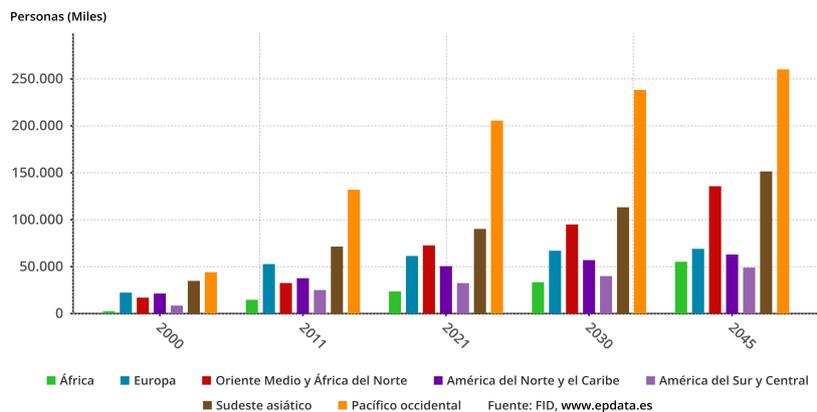


Figura 2.1: Evolución de la prevalencia de la diabetes por continente y estimaciones según la FID. Figura extraída de [23].

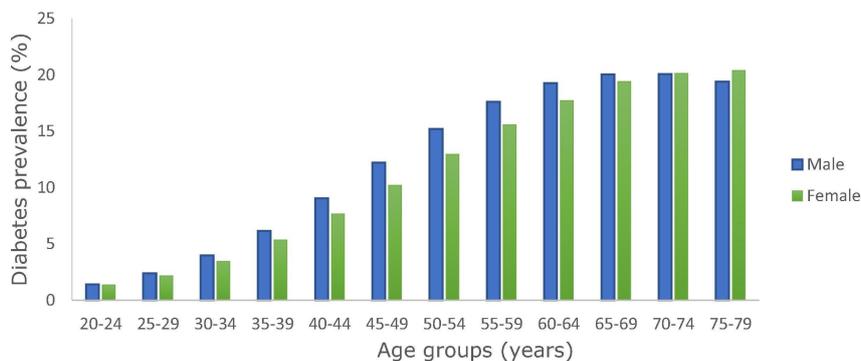


Figura 2.2: Prevalencia de la diabetes por edad y sexo en el año 2019. Figura extraída de [1].

Esta enfermedad se caracteriza por niveles de glucosa en sangre muy elevados y causado porque el páncreas no genera suficiente insulina, o el organismo no utiliza la insulina de forma eficiente [24]. La insulina es una hormona polipeptídica sintetizada por las células beta del páncreas cuya función es permitir el paso de la glucosa en sangre a las células y mantener los niveles de glucemia [24]. Dicha glucosa, es la principal fuente de energía del organismo. Existen dos tipos principales de diabetes, DMT1 y DMT2. La DMT1 es causada por una reacción autoinmune que impide que el cuerpo produzca insulina. Las células del sistema inmune atacan a las células beta del páncreas destruyéndolas y, por tanto, impidiendo la generación de insulina [25]. La DMT1 es más frecuente en niños y jóvenes aunque también afecta a adultos y a población de todas las etnias [26]. La DMT2 es el tipo de diabetes más frecuente, con una incidencia entre 90-95 % [25], y está causada por la resistencia a la insulina. El cuerpo no utiliza la insulina adecuadamente y no logra mantener los niveles de glucosa en sangre dentro del rango

establecido como normal [27]. Las personas que padecen DMT1 necesitan regulación diaria del nivel de glucosa mediante inyecciones de insulina o mediante otros agentes que no están basados en insulina, como la metamorfina [28]. La DMT2 puede aparecer a cualquier edad, pero predomina en las personas mayores, y se puede prevenir con el ejercicio físico y alimentación saludable aunque también pueden requerir de inyecciones de insulina [4]. Además, existe la diabetes gestacional, que puede aparecer en mujeres durante el embarazo y supone un mayor riesgo para el hijo de padecer diabetes y para la madre de desarrollar la enfermedad [4].

La mayoría de los pacientes que padecen diabetes, controlan su concentración de glucosa en sangre midiendo los niveles de la misma con un dispositivo. Esta práctica es muy útil para mantener el azúcar en sangre dentro del rango normal [29]. Los valores establecidos como normales de azúcar en sangre en ayunas son entre 70 y 100 miligramos (mg) de azúcar por decilitro (dL) de sangre según la Organización Mundial de la Salud [30]. En base a estos valores se definen la hipoglucemia e hiperglucemia, de modo que, la hipoglucemia es cuando la glucosa en sangre es menor que 70mg/dL [30], y la hiperglucemia, cuando es mayor a 125 mg/dL [31]. El número de hospitalizaciones por eventos de hipoglucemia e hiperglucemia severa han aumentado en los últimos años en Estados Unidos como se puede ver en la Figura 2.3. Los eventos de hiperglucemia e hipoglucemia pueden llegar a dar lugar a consecuencias clínicas muy graves e incluso mortales entre las que cabe destacar un aumento de la predisposición al desarrollo de una patología cardiovascular, la retinopatía diabética, nefropatía o neuropatía

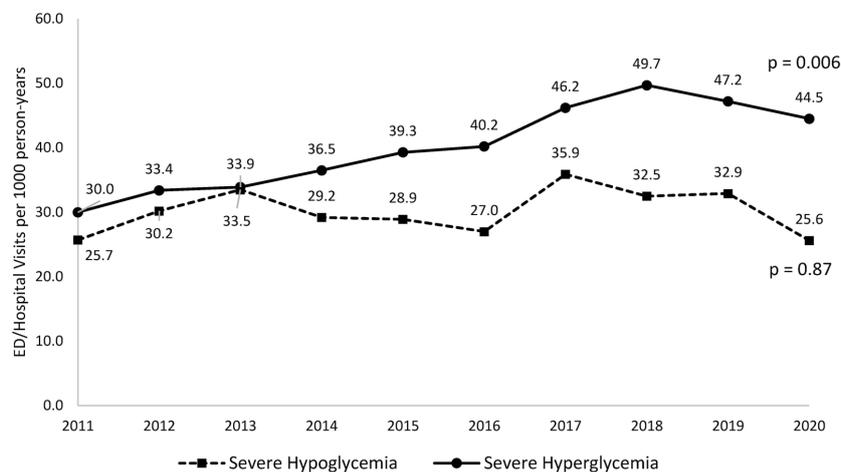


Figura 2.3: Eventos de hiperglucemia e hipoglucemia severas entre los años 2011 y 2020 en Estados Unidos. Figura extraída de [10].

Para obtener el nivel de glucosa en sangre se pueden utilizar dos tipos de medidores: (1) dispositivos de monitorización de glucosa en sangre (BGM, del inglés *Blood Glucose Monito-*

ring); y (2) dispositivos CGM. Por otro lado, los dispositivos BGM son un método más barato para medir la concentración de glucosa en sangre, que consiste en extraer sangre con un pinchazo en una zona del cuerpo, por ejemplo, el dedo. Esta muestra se analiza con un dispositivo externo para evaluar el nivel de glucosa. Con este tipo de dispositivos, la frecuencia de muestreo es variable y no se realizan tantas mediciones diarias [17]. En el año 1999 aparece el primer dispositivo CGM llamado *Medtronic MiniMed*, el cual era capaz de almacenar datos de concentración de glucosa en sangre de hasta 3 días para analizarlos posteriormente [29]. Está formado por un sensor y receptor enzimático electroquímico (insertado bajo la piel) calibrado en base al autocontrol de glucosa en sangre detectado por medidores validados del mercado [29]. El sensor envía una señal cada 10 segundos y se guarda una muestra de glucosa cada 5 minutos [29]. A partir de este dispositivo, surgen los sistemas CGM que se conocen en la actualidad [29]. Los dispositivos CGM utilizan un sensor colocado debajo de la piel que mide la glucosa en sangre realizando pinchazos en la piel sin afectar a los vasos sanguíneos, y transmitiendo así la concentración de glucosa en sangre (medido en mg/dL) a un dispositivo externo como un teléfono móvil o un reloj inteligente [11]. Además, existen dispositivos implantables CGM (que en lugar de ser subcutáneos) que funcionan aplicando radiación electromagnética a través de la piel a los vasos sanguíneos [11], siendo menos invasivos, y más fáciles de quitar [12]. El dispositivo CGM tiene una frecuencia de muestreo constante cada 5 minutos y permite realizar predicciones para ayudar a controlar los niveles de glucosa y prevenir episodios de hipoglucemia e hiperglucemia [11].

Desde el punto de vista de análisis de datos, los datos recopilados por los dispositivos CGM se representan mediante una serie temporal. Una serie temporal es una colección de una o múltiples muestras realizadas de forma secuencial en el tiempo. En las series temporales, cada muestra recogida va ligado al instante temporal en el que se recoge. Por lo que, al trabajar con series temporales se manejan dos variables: la variable a tratar y el instante temporal asociado a la misma [32]. Por ejemplo, en este proyecto, cada valor de glucemia va asociado al instante temporal en el que el dispositivo CGM realiza la medición. Las series temporales tienen una periodicidad establecida, pudiendo ser regular cuando la frecuencia de muestreo es constante, o irregular cuando no lo es. En el caso de este TFG, las muestras se toman cada 5 minutos, ya que los valores de glucosa en sangre están registrados por dispositivos CGM que tienen 5 minutos como frecuencia de recogida de muestra. En la Figura 2.4 se muestra un ejemplo de serie temporal de glucosa extraída de los datos empleados en este TFG.

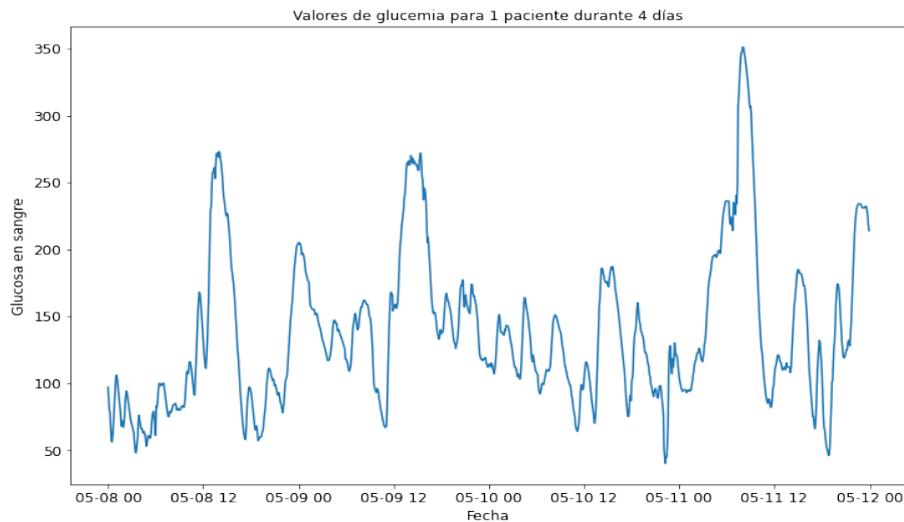


Figura 2.4: Ejemplo de serie temporal con 4 días de registro de glucosa para 1 paciente y con frecuencia de muestreo de 5 minutos.

2.2. Análisis descriptivo y preprocesamiento de las bases de datos

A rasgos generales todas las series temporales consideradas son muy similares. Todas ellas tienen registros de múltiples pacientes, identificados por un identificador de paciente (en inglés *Patient ID* (PtID)), una variable secuencia que indica la fecha-hora de recogida del valor de glucosa, y una última variable que indica el valor de glucosa en sangre registrado. Las bases de datos son públicas y de libre acceso y fueron extraídas de la página oficial de *Jaeb Center for Health Research* [33]. En la Tabla 2.1 se describe brevemente las bases de datos consideradas. Como se observa hay diferencia entre el número de pacientes disponible en cada base de datos, siendo la primera la de mayor número de pacientes y la última la de menor. También hay grandes diferencias en cuanto al número de registros, esto se debe a que la duración de los estudios es variable.

Nombre de base de datos	Identificador	# pacientes	Rango de edad	# registros
Severe hypoglycemia in older adults with type 1 diabetes	BD1-SEV	200	>60 años	655,755
The International Diabetes Closed Loop trial: Clinical Acceptance of the Artificial Pancreas in pediatrics	BD2-PED	93	6-13 años	3,780,007
Wireless Innovation for Seniors with Diabetes Mellitus	BD3-SEN	139	>60 años	6,745,473

Tabla 2.1: Descripción de las bases de datos de glucosa empleadas en el desarrollo del TFG.

Para poder trabajar con los datos, es necesario realizar un procesamiento previo para mejorar el rendimiento de los algoritmos utilizados. En concreto, se requiere tratar los valores perdidos (en inglés *missing data*), ya que, al trabajar con datos médicos, el preprocesamiento cobra gran importancia debido a la calidad de los mismos. A continuación se describen las bases de datos y se muestran el análisis descriptivo y el preprocesamiento.

Las bases de datos consideradas se definirán de la siguiente forma a lo largo de todo el TFG:

- Base de datos 1: *Severe hypoglycemia in older adults with type 1 diabetes* (**BD1-SEV**)
- Base de datos 2: *The International Diabetes Closed Loop trial: Clinical Acceptance of the Artificial Pancreas in pediatrics* (**BD2-PED**)
- Base de datos 3: *Wireless Innovation for Seniors with Diabetes Mellitus* (**BD3-SEN**)

2.2.1. BD1-SEV

Esta base surge de un estudio científico sobre la ocurrencia de eventos de hipoglucemia severa en adultos mayores con DMT1 [34]. Los episodios de hipoglucemia severa son aquellos en los que el paciente requiere de asistencia sanitaria debido a un nivel de glucosa en sangre extremadamente bajo que pueden generar arritmias o patologías cardíacas [34]. El estudio contó con 200 pacientes, 100 tipo caso y 100 tipo control. Los pacientes tipo caso son aquellos que han tenido al menos un evento de hipoglucemia severa en el último año, mientras, los tipo control son aquellos que no han sufrido ningún evento de hipoglucemia severa en los últimos 3 años. Los sujetos considerados en el estudio son adultos mayores de 60 años que padecen DMT1 con necesidad de insulina desde hace, al menos, 20 años. Los datos empleados para este estudio suponen una fracción de datos filtrada obtenida de *T1D Exchange* [35]. *T1D Exchange* consiste en una organización en la que se recogen datos de pacientes que padecen DMT1, y que almacena datos sobre el progreso de su enfermedad hasta datos extraídos de los dispositivos

CGM. Estos datos de monitorización continua recogidos cada 5 minutos son los recogidos en esta base de datos.

En primer lugar, se realiza un análisis descriptivo de la base de datos BD1-SEV. Para ello se comienza por estudiar las características con las que se cuenta. Como se ve en el ejemplo de la Tabla 2.2, en esta base de datos se tienen 4 columnas. La primera columna (PtID) con el identificador de paciente, es una característica numérica y hay 200 identificadores distintos aleatorios y únicos y que identifican a cada paciente. La segunda columna (*DeviceDaysFromEnroll*) es numérica también y hace referencia al día en el que se realiza el registro. Esta columna se transforma a formato fecha siendo el día 0, el primer día, 01-01-2023. La tercera columna (*DeviceTm*) indica la hora, minuto y segundo los que se recoge la muestra. La última (*Glucose*) indica el valor de glucosa obtenido por el dispositivo CGM.

PtID	<i>DeviceDaysFromEnroll</i>	<i>DeviceTm</i>	<i>Glucose</i>
199	1	00:19:30	172
199	1	00:49:30	194
199	1	00:54:30	184
199	1	00:59:30	174

Tabla 2.2: Ejemplo de datos de la BD1-SEV.

Después de conocer las características, se procede a visualizar las variabilidad de las series temporales. Se observa en la Figura 2.5 la variación de glucosa de diversos pacientes durante 1 día. Se puede ver, como cada serie temporal tiene una evolución temporal única y distintos valores máximos y mínimos. A continuación, se hace lo mismo pero con distintos días. Para visualizar la variación las series temporales de glucosa a lo largo de los días, esto se representa la Figura 2.6. En este caso se puede ver como para el mismo paciente, la evolución temporal tiene distinta forma en los distintos días.

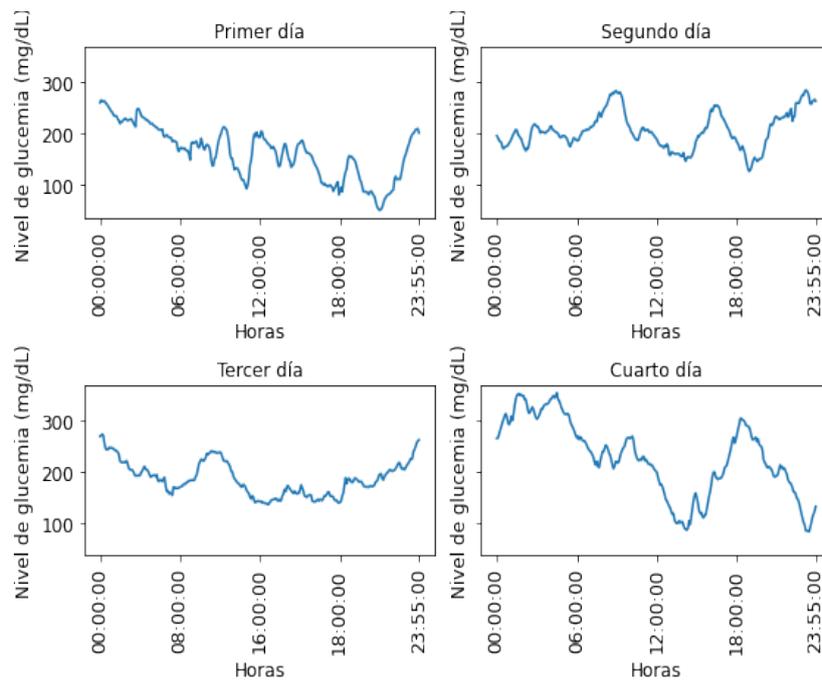


Figura 2.5: Visualización de una serie temporal de glucosa para el mismo paciente de la BD1-SEV durante distintos días.

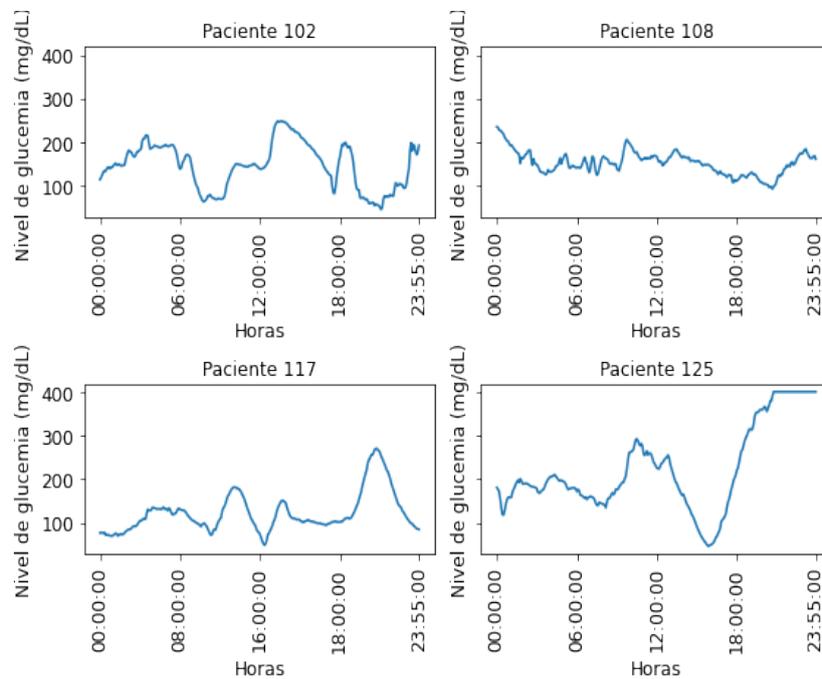


Figura 2.6: Visualización de una serie temporal de glucosa durante un día para distintos pacientes de la base de datos BD1-SEV.

A continuación, se explica el preprocesamiento llevado a cabo para las distintas series temporales (asociadas a los distintos pacientes de la base de datos). En primer lugar, se busca identificar *missing data*. Para ello, se evalúa la distribución de los datos a lo largo del tiempo representando en un histograma la cantidad de datos diarios, es decir, la suma de muestras recogidas de los pacientes por cada día. De este modo, como se ve en la Figura 2.7 (a), se ve como el número de datos de los que se dispone es muy variable entre los distintos días, considerando todos los pacientes. Esto se debe a que los distintos pacientes se unen al estudio en distintos momentos, por lo que no siempre se dispone de los mismos pacientes. Para abordar este problema, se ha decidido hacer una selección de un rango de fechas en las cuales el número de muestras de glucosa sea homogéneo y así mitigar el problema de *missing data*. En esta base de datos se reduce a 6 días de 80 disponibles inicialmente, del día 1 al 6 ambos incluidos. Al hacer esta selección de días, se disminuye el número de pacientes quedando 166 pacientes de los 200 originales. Por tanto, se trabaja con una base de datos reducida cuya distribución en cuanto a número de muestras diarias queda como la que se muestra en la Figura 2.7 (b). Se puede observar como la distribución de las muestras a lo largo de los días es más homogénea y, por ello, menos *missing data*.

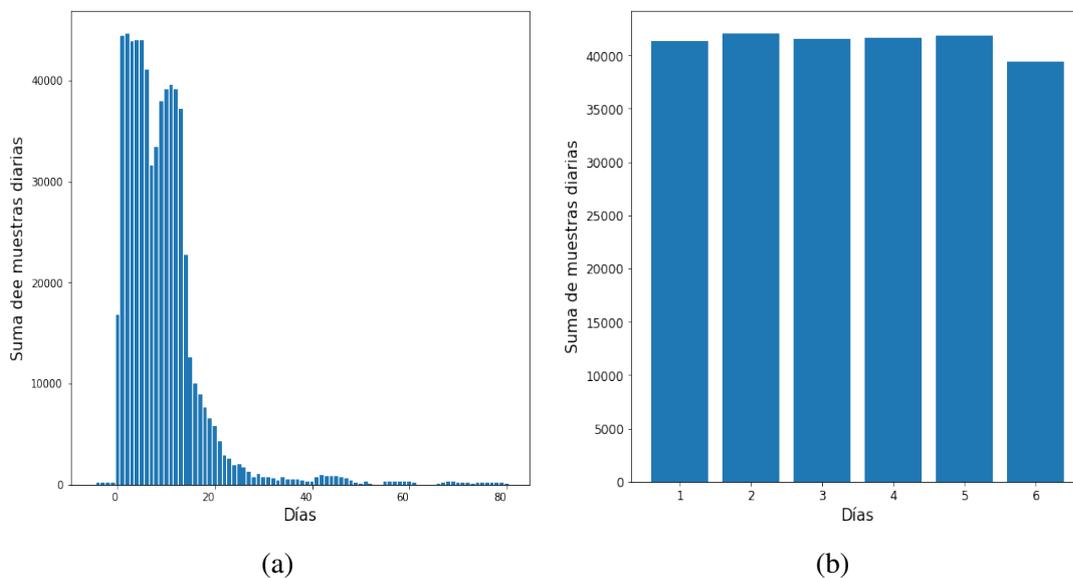


Figura 2.7: Suma de muestras de glucosa de los pacientes de la BD1-SEV recogidas cada día. (a) Previas a la selección de fechas; y (b) después de seleccionar 6 días.

Una vez seleccionado el periodo más homogéneo, se estudia el número de muestras totales disponible para cada uno de los pacientes. En la Figura 2.8 se observa la alta variabilidad entre los distintos pacientes, y se decide eliminar aquellos pacientes cuyo número de muestras

sea menor a la media (1,494 muestras). Como se explicó previamente, esta estrategia permite reducir el número de valores perdidos en las series temporales, y con ello, mejorar la calidad de los datos usados para el entrenamiento de los modelos de generación de series temporales y de predicción de glucosa. Tras esta reducción, el número de pacientes restante es 110 pacientes.

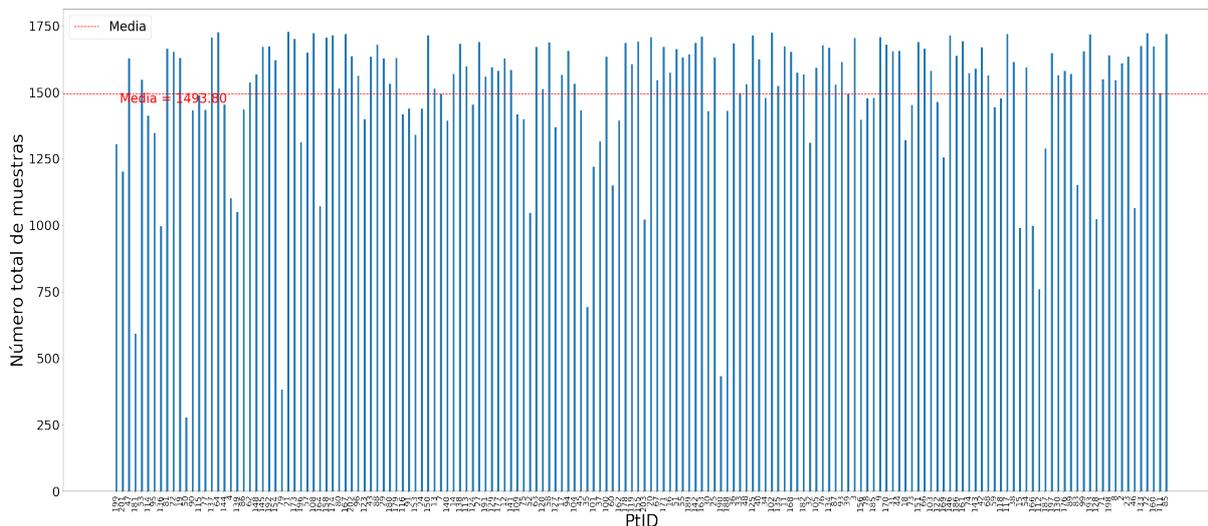


Figura 2.8: Número total de muestras por paciente de la BD1-SEV.

A continuación, se estudia la frecuencia de muestreo. Los dispositivos CGM registran una muestra cada 5 minutos, es decir, cada 300 segundos. Procedemos a comprobar que, para cada paciente, la diferencia de tiempo entre una muestra y la inmediatamente posterior sea siempre 300 segundos. Así es como se determina el alto número de valores perdidos, ya que hay un gran número de casos en los que la frecuencia de muestreo es mayor a 300 segundos. En este caso, como se dispone de un alto número de pacientes (110), se considera un umbral de 1815 segundos (30 minutos), de modo que, cualquier paciente que tenga al menos en una ocasión una diferencia de tiempo entre una muestra y la inmediatamente posterior mayor a 30 minutos, será eliminado. De este modo, de los 200 pacientes iniciales nos quedamos con 14 pacientes y sus respectivos valores de concentración de glucosa.

Una vez se tienen los 14 pacientes finales, se procede al remuestreo de la serie temporal para conseguir una frecuencia de muestreo constante de 5 minutos. Se procede a rellenar con valores nulos (en inglés, *Not A Number* (NaN)) aquellos instantes en los que no hay una muestra de glucosa. Para ello se crea una columna que contiene la fecha y la hora y se agrupan los datos según el identificador de paciente (PtID). Una vez agrupado, se aplica un etapa de remuestreo, mediante el cual se consigue que la diferencia de tiempo entre muestras sea siempre 300 segundos. A continuación se muestra en la Figura 2.9 el número de valores perdidos por paciente.

Se puede observar como hay una gran diferencia entre el número de NaN entre los distintos pacientes, siendo el paciente con PtID=202 el que mayor número de *missing data* presenta y el PtID=21 el que menos. Esto afectará al rendimiento de algoritmos ML, ya que aquellas series temporales con mayor cantidad de valores perdidos tendrán peores prestaciones predictivas.

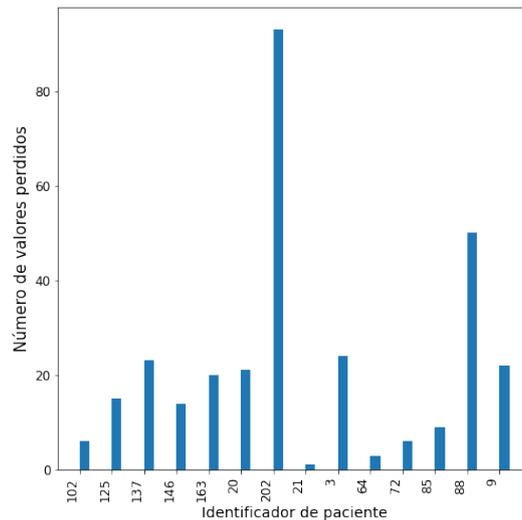


Figura 2.9: Número total de valores perdidos por paciente de la BD1-SEV.

Muchos modelos de ML necesitan datos completos para funcionar, es decir, que los datos nos presenten *missing data*. En concreto el método de predicción empleado en este TFG no permite *missing data* y no se puede llevar a cabo el entrenamiento a no ser que se tenga la serie temporal completa. Además, la falta de datos hace que sea muy difícil identificar patrones temporales, en concreto cuando los datos faltan durante un periodo de tiempo largo [36]. En este TFG se aborda este problema mediante la imputación [37], un proceso por el cual se sustituyen los valores nulos por un valor. En estudios anteriores se han encontrado con este problema a la hora de realizar predicciones con datos obtenidos de dispositivos CGM, ya que por diversas razones como la calibración del sensor, la frecuencia de muestreo de los dispositivos CGM puede variar, y lo han solventado mediante imputación [37–39].

En la Figura 2.10 (a) se puede ver un ejemplo de una serie temporal durante dos días previamente a la imputación. Para visualizar mejor los valores perdidos de glucosa, se han sustituido los NaN por un valor atípico (en este caso, -50). Para tratar esos valores NaN, en este TFG se proponen varios métodos de imputación: imputación por *Random Forest* (RF) (véase la Figura 2.10 (b)), lineal (véase la Figura 2.10 (c)), *K-Nearest Neighbours* (KNN) (véase la Figura 2.10 (d)), polinómica de grado 2 (véase la Figura 2.10 (e)) y polinómica de grado 3 (véase la Figura 2.10 (f)), en las que, siendo los puntos rojos se corresponden con los datos impu-

tados, se ve como afectan los distintos tipos de imputación a los datos dependiendo del método de imputación utilizado. Finalmente, se decide utilizar la imputación por *Random Forest* (RF), un modelo basado en ML que puede resultar muy útil cuando la cantidad de valores nulos es elevada y hay estudios que afirman que ofrece mejor precisión que otros modelos [40]. Este método de aprendizaje máquina permite abordar la no linealidad a diferencia de otros métodos de imputación basados en estadísticos [41].

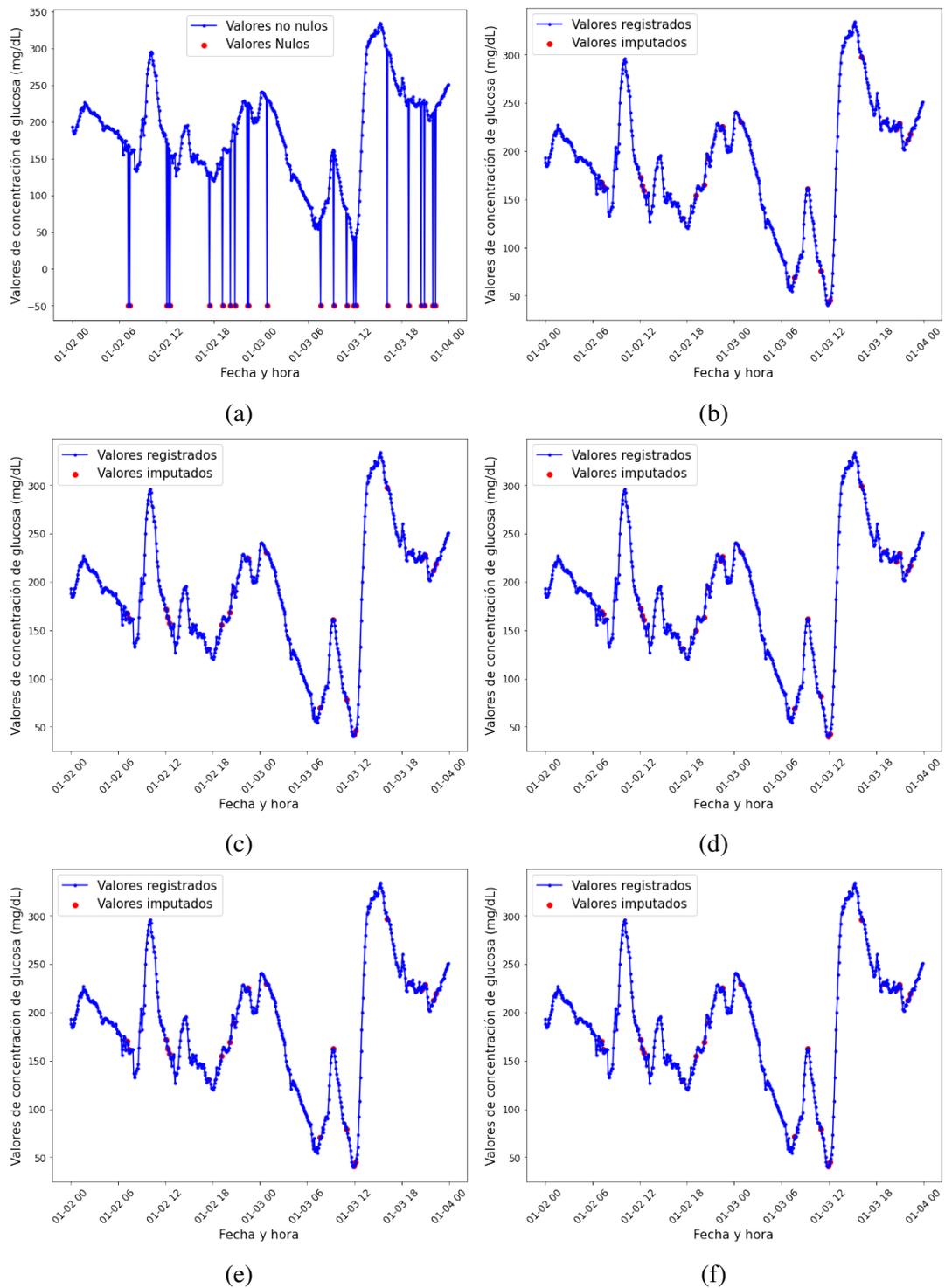


Figura 2.10: Resultados de aplicar diferentes técnicas de imputación para una serie temporal de glucosa de BD1-SEV. (a) Serie temporal sin imputación; imputación basada en: (b) RF; (c) método lineal; (d) KNN; (e) método polinómico de grado 2; (f) método polinómico de grado 3. En el eje x se muestra 'Fecha y hora' en formato: Mes-Día Hora.

2.2.2. BD2-PED

La segunda base de datos empleada en este TFG, se utilizó para un ensayo clínico para la aceptación del páncreas artificial en pediatría publicado en abril de 2019 y realizado en Estados Unidos [42]. Al ser un estudio pediátrico, los pacientes son niños en un rango de edad de 6 a 13 años. Todos los pacientes del estudio padecen DMT1 y se dividen en dos grupos: tipo caso y tipo control. Los pacientes tipo caso utilizan la bomba de insulina que se está probando (*t:slim X2 with Control-IQ System*) y los controles no hacen uso de esta. Cabe mencionar que todos los pacientes usan dispositivos CGM para medir el nivel de glucemia en sangre, obteniendo muestras cada 5 minutos. El estudio consta de dos fases [42]: (1) la primera fase es la principal del estudio y tiene una duración de 16 semanas; (2) la segunda fase es una fase de extensión que dura 12 semanas. En la segunda fase el grupo control pasa a utilizar la bomba de insulina y se ofrece a todos los sujetos participar en ella, aunque, no todos participan [42].

En primer lugar, se estudian las características presentes en la base de datos. En la Tabla 2.3 se muestran un ejemplo de los registros de la base de datos BD2-PED. Como se puede observar la base de datos consta de 3 columnas. La primera columna (PtID) es una característica numérica y única que permite identificar a cada paciente. La segunda columna (*DataDtTm*) indica el momento exacto en el que se recoge la muestra en (año-mes-día hora:minuto:segundo). La tercera columna (CGM) indica el valor de glucosa recogido en cada caso.

PtID	<i>DataDtTm</i>	CGM
199	2018-11-17 23:43:39	221
199	2018-11-17 23:48:38	222
199	2018-11-17 23:53:39	225
199	2018-11-17 23:58:39	226

Tabla 2.3: Ejemplo de datos de la BD2-PED.

Una vez se conocen las características de la serie temporal, se procede a visualizar la variación de los niveles de glucosa en sangre teniendo en cuenta dos escenarios: (1) distintos días para un mismo paciente (véase Figura 2.11); y (2) mismo día para varios pacientes (véase Figura 2.12). En ambos casos se puede ver como cada serie temporal de glucosa es completamente diferente.

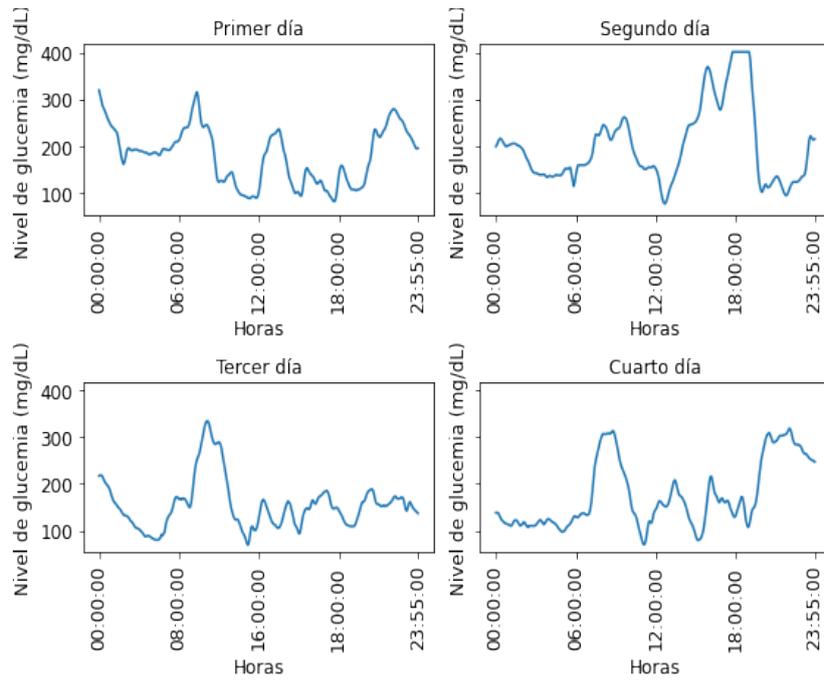


Figura 2.11: Visualización de una serie temporal de glucosa para el mismo paciente de la BD2-PED durante distintos días.

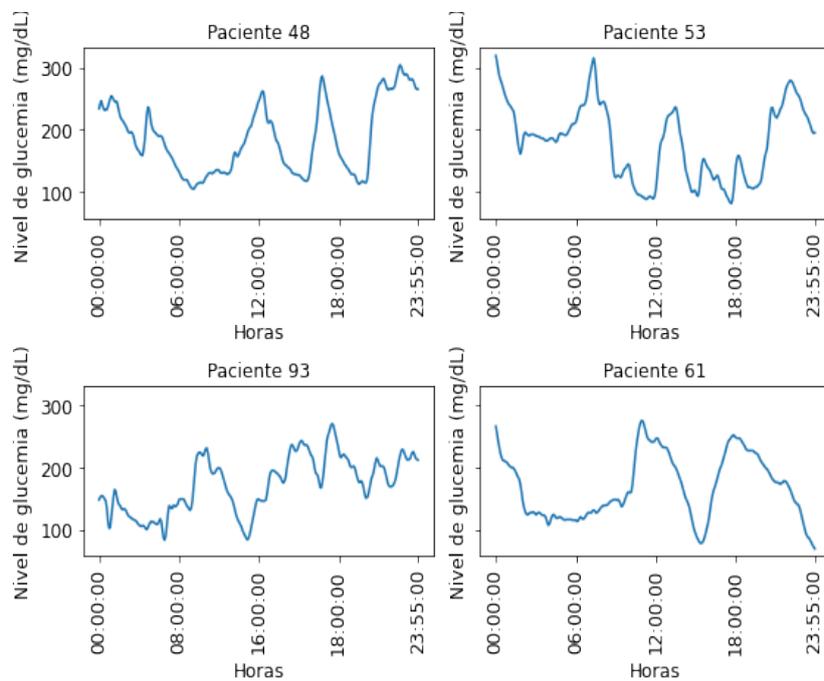


Figura 2.12: Visualización de una serie temporal de glucosa durante un día para distintos pacientes de la BD2-PED durante distintos días.

A continuación, se busca identificar el periodo más homogéneo en cuanto a distribución de datos de glucosa y elegir intervalos de tiempo donde el número de valores nulos sea menor. En Figura 2.13 (a) se representa un histograma que muestra la suma total de muestras de glucosa que hay registradas por cada día, teniendo en cuenta todos los pacientes. Se observa una gran diferencia en el número de datos de glucosa por día, con intervalos de tiempo donde muy poco registros (por ejemplo, entre [2018-07, 2018-09], y [2019-09, 2020-03]). Esto se debe a que no se tiene siempre los mismos pacientes, ya que se incorporan al estudio en distintos momentos, y al alto número de valores perdidos. Para abordar este problema, se ha decidido hacer una selección de un periodo en el que el número de muestras es más homogéneo. Así pues, se escogen 45 días de datos [2019-03-10, 2019-04-24] y al hacer esta reducción se pasa a tener 22 pacientes. Se empieza a trabajar con esta base de datos reducida en la cual el número de muestras diarias es menos variable como se ve en la Figura 2.13 (b).

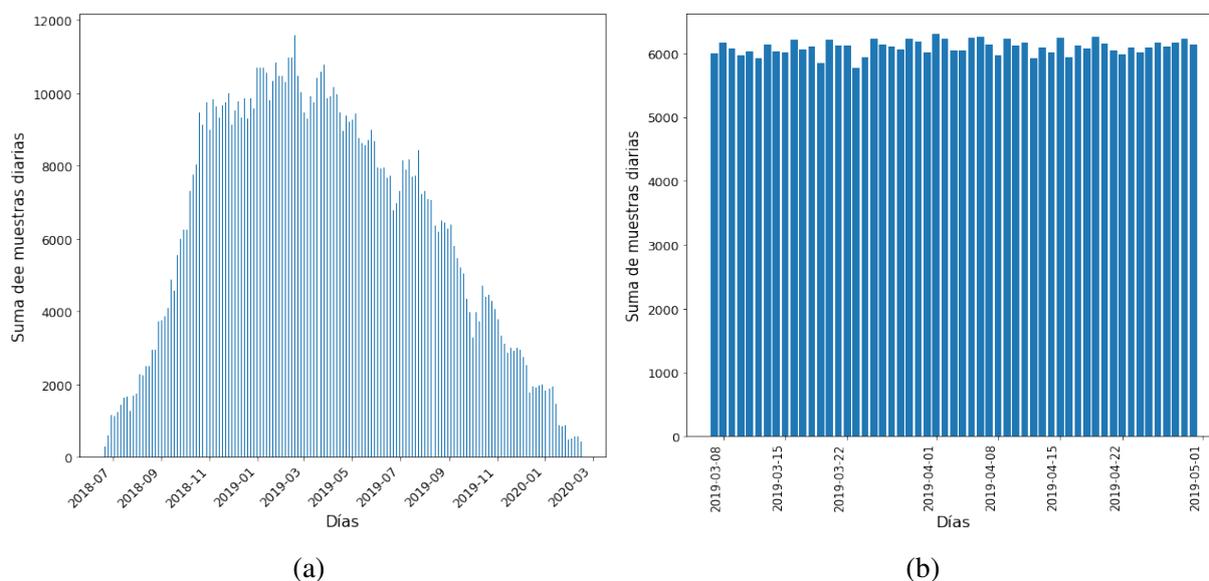


Figura 2.13: Suma de muestras de los pacientes de la BD2-PED recogidas cada día. (a) Previas a la selección de fechas; (b) Después de seleccionar 42 días.

Una vez seleccionado este periodo más homogéneo, se estudia el número de muestras totales que se tiene de cada uno de los pacientes. Para ello se visualiza la Figura 2.14, en la que se ve por cada PtID el número total de registros disponibles. A continuación, para intentar reducir el número de valores nulos se decide eliminar todos los pacientes con un número de datos inferior a la media (15,236) eliminando así 7 pacientes.

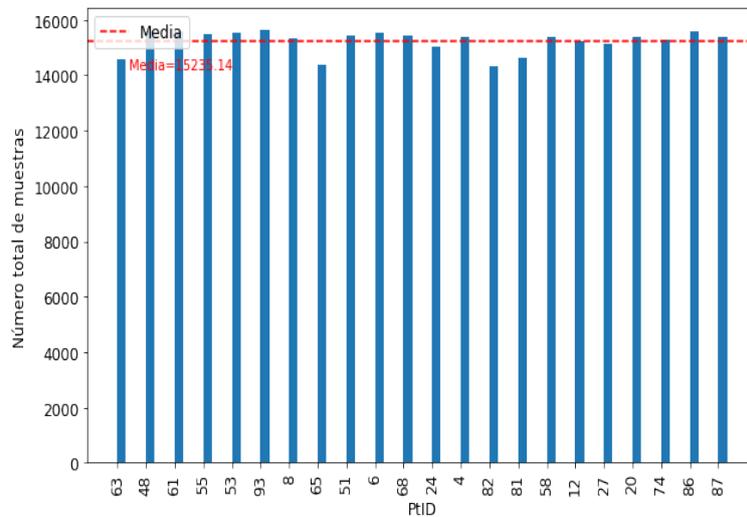


Figura 2.14: Número total de muestras por paciente de la BD2-PED y su media.

A continuación, se estudia la frecuencia de muestreo de las distintas series temporales y se busca que esta sea siempre 5 minutos (300 segundos). El primer paso es comprobar que, para cada paciente, la diferencia de tiempo entre una muestra y la siguiente sea siempre 5 minutos, permitiendo identificar la cantidad de valores perdidos. Cada vez que el periodo sea distinto a 300 segundos faltarán datos, por ejemplo, si la diferencia entre una muestra y la siguiente es de 600 segundos falta una muestra. Siguiendo este enfoque, se eliminarán pacientes con un alto número de valores nulos, con ayuda del gráfico mostrado en la Figura 2.15. En concreto se eliminan aquellos que presentan más de 500 valores perdidos o largos periodos sin muestras, quedando finalmente 6 pacientes.

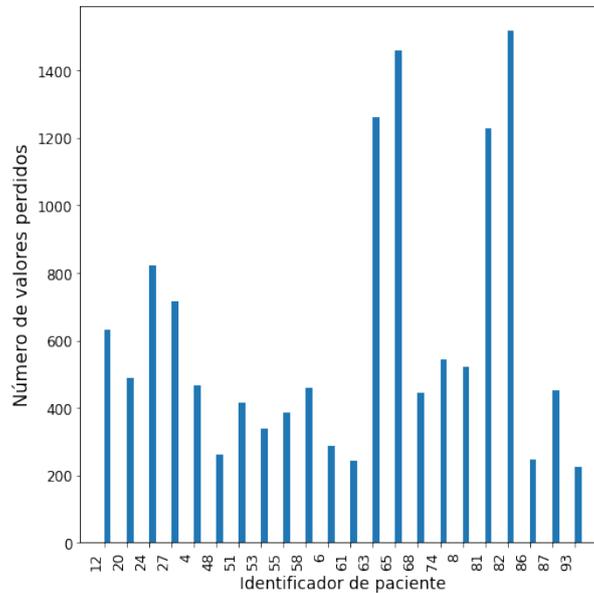


Figura 2.15: Número total de valores perdidos por paciente de la BD2-PED.

Una vez se tienen los 6 pacientes finales, se procede al remuestreo de la serie temporal. Para conseguir una frecuencia de muestreo constante de 5 minutos, se introduce un valor NaN en aquellos instantes en los que no hay muestra. A continuación, se aplican diferentes técnicas de imputación para sustituir valores nulos. En la Figura 2.16 (a) se puede ver un ejemplo de como se distribuyen los valores NaN a lo largo del tiempo durante dos días. Para solucionar el problema de *missing data*, se prueban varios métodos de imputación. En concreto se evalúan los métodos: RF, imputación lineal, imputación por KNN, imputación polinómica de grado 2, imputación polinómica de grado, como se puede ver en la Figura 2.16. Finalmente se imputan los datos mediante RF [41]. El resultado son series temporales como las representadas en la Figura 2.16 (b), con los valores NaN imputados mediante RF.

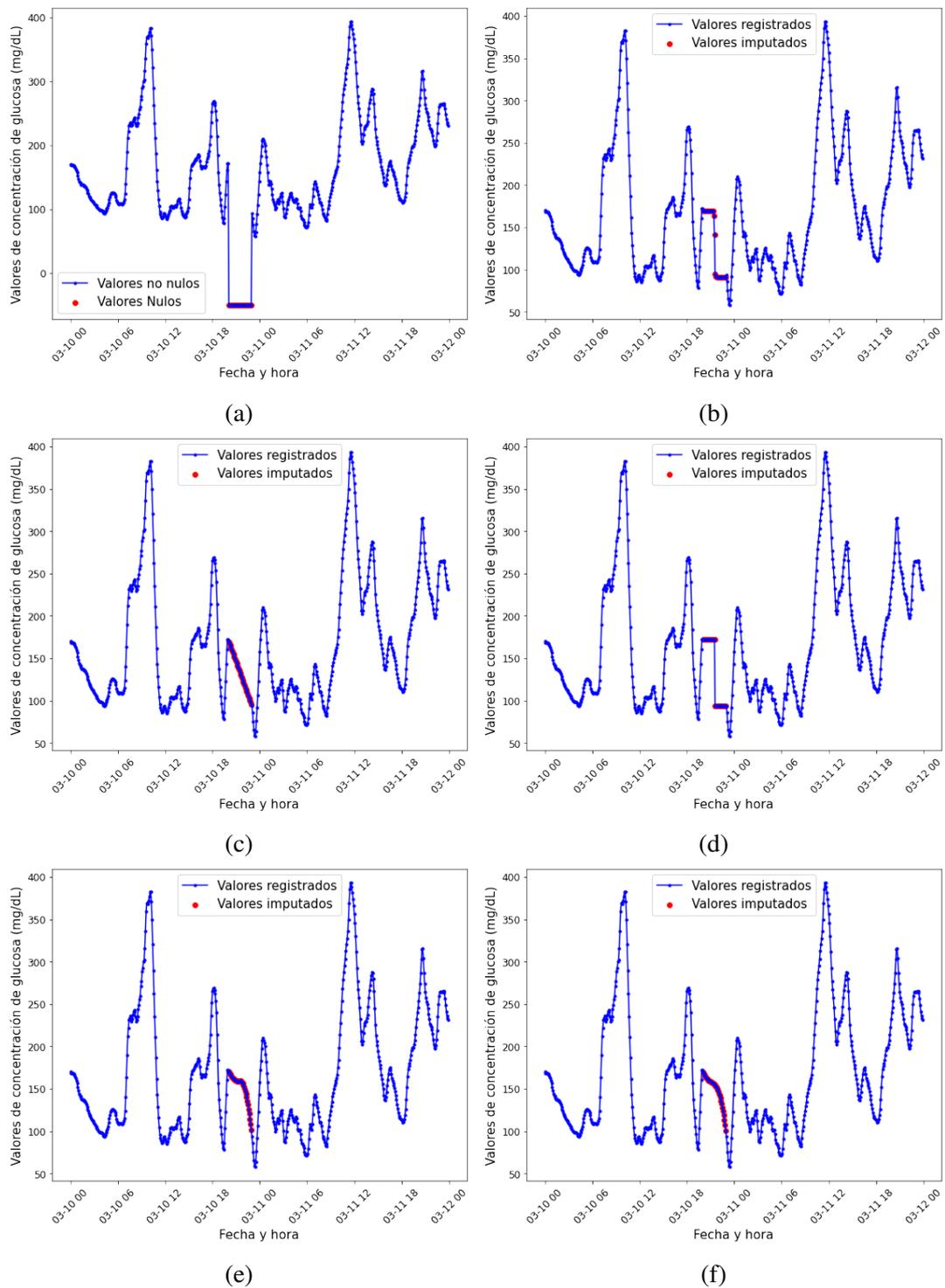


Figura 2.16: Resultados de aplicar diferentes técnicas de imputación para una serie temporal de glucosa de BD2-PED. (a) Serie temporal sin imputación; imputación basada en: (b) RF; (c) método lineal; (d) KNN; (e) método polinómico de grado 2; (f) método polinómico de grado 3. En el eje x se muestra 'Fecha y hora' en formato: Mes-Día Hora.

2.2.3. BD3-SEN

La última base de datos estudiada en este TFG fue obtenida de *T1D Exchange* [35], y se utilizó en un estudio para prevenir la hipoglucemia en adultos mayores [43]. La población de estudio está formada por personas mayores de 60 años con DMT1 diagnosticada antes de los 40 años de edad, que necesiten inyecciones de insulina diaria, no obesos y con familia de primer grado (hermanos, padres, hijos) que padezcan DMT1. Se cuenta con un total de 139 pacientes que reúnen las características anteriores. Los individuos del estudio llevan un dispositivo subcutáneo CGM que mide el nivel de glucosa en sangre en mg/dL, recopilando una muestra de sangre cada 5 minutos. Esto no se cumple cuando se extrae el dispositivo para llevar a cabo el mantenimiento y calibración del mismo. Se recogen datos durante 26 semanas que dura el estudio más una fase consecutiva de extensión que alarga el estudio hasta alcanzar 52 semanas [43].

En primer lugar, se realiza un análisis descriptivo empezando por estudiar las características presentes en la base de datos, como se puede ver en la Tabla 2.4, en esta serie temporal se tienen 5 columnas. La primera columna (*PtID*) con el identificador de paciente, es una característica numérica única para cada paciente. La segunda columna es *DeviceDtTm*, esta columna indica el momento exacto en el que se recoge la muestra en: año, mes, día, hora, minuto, segundo, milisegundo. La tercera (*Value*) indica el valor de glucosa recogido en cada caso. La cuarta columna (*Visit*) indica el tipo de visita, este ensayo tiene varias fases que se pueden identificar según la semana en la que se hizo la visita. La última columna (*TrtGroup*) indica que tratamiento sigue el paciente, en este caso si el medidor de glucosa es continuo o no. En este TFG solo se toman aquellas muestras medidas por el dispositivo CGM, eliminando todas aquellas filas en las que la columna *TrtGroup* sea BGM.

<i>PtID</i>	<i>DeviceDtTm</i>	<i>Value</i>	<i>Visit</i>	<i>TrtGroup</i>
63	20FEB2000:13:39:15.000	99.00	Randomization\n	BGM
63	20FEB2000:13:44:15.000	108.00	Randomization\n	BGM
98	13JAN2001:23:35:54.000	97.00	26 week\n	CGM
98	13JAN2001:23:40:53.000	96.00	26 week\n	CGM

Tabla 2.4: Ejemplo de datos de la BD3-SEN.

Una vez examinadas las características de la serie temporal, se procede a visualizar de manera gráfica la variación de los niveles de glucosa en sangre. Para ello se analiza la variación de datos de glucosa de un paciente en diferentes días, se puede apreciar en la Figura 2.17 como

los valores de glucemia para el tercer día se encuentran en un rango menor y más estables que en los otros tres días. Se evalúa también la variación de la concentración de glucosa en sangre entre distintos pacientes durante un mismo día, como se muestra en la Figura 2.18 en la cual se aprecia como los valores máximos y mínimos de concentración de glucosa en sangre ocurren en distintas horas del día para cada paciente.

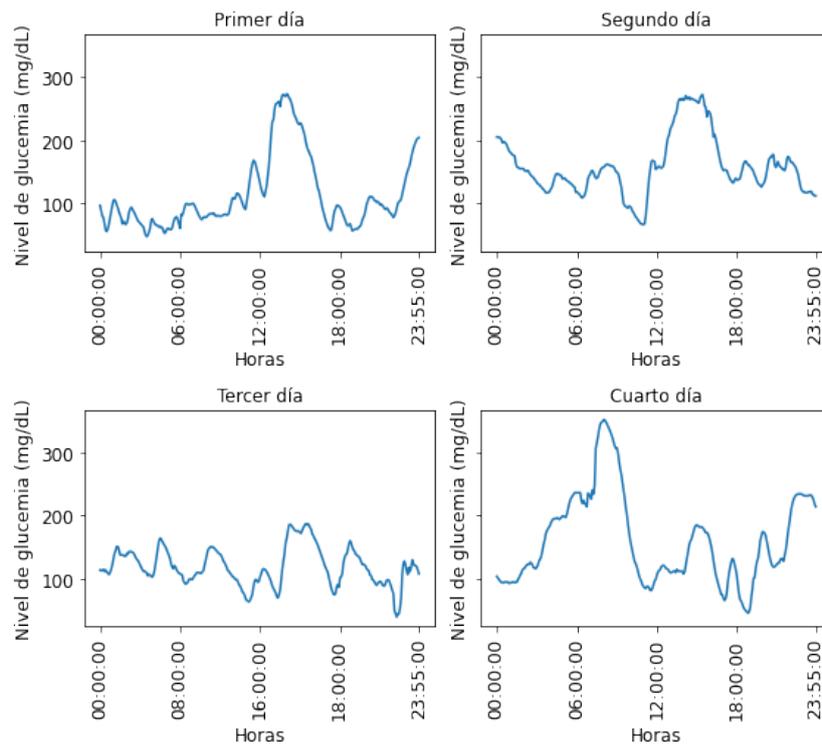


Figura 2.17: Visualización de serie temporales de glucosa para un paciente de la BD3-SEN durante diferentes días.

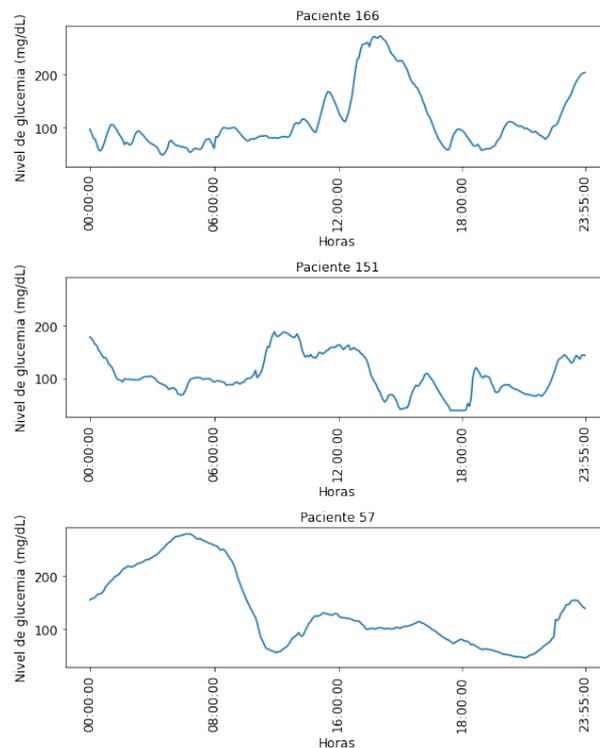


Figura 2.18: Visualización de series temporales con datos de glucosa de un día asociadas con distintos pacientes de la BD3-SEN.

A continuación, se desarrolla el preprocesado llevado a cabo en las series temporales de esta base de datos. En primer lugar, se busca identificar el periodo más homogéneo en cuanto a distribución de datos de glucosa para intentar así, minimizar el número de valores perdidos. Primero, se empieza por eliminar todos aquellos datos que tengan en la columna *Visit* las etiquetas: *39 week* o *52 week*, ya que pertenecen a la fase de extensión del estudio [43] en la cual no participan todos los pacientes. En la Figura 2.19 (a) se representa un histograma que muestra la suma total de muestras que hay por cada día. Se ve como hay una gran diferencia en el número de datos dependiendo del día, ya que no se tiene siempre los mismos pacientes y hay un alto número de valores nulos. Para abordar este problema, se ha decidido hacer una selección de un periodo limitado de un mes en el que el número de muestras es más homogéneo. En este caso, se eligen los días comprendidos entre el 8 de mayo y el 8 de junio del 2000, ambos inclusive. Esta reducción causa a su vez una disminución del número de pacientes, quedándonos con 11 pacientes. Se empieza a trabajar con esta base de datos reducida en la cual el número de muestras diarias es menos variable como se ve en la Figura 2.19 (b).

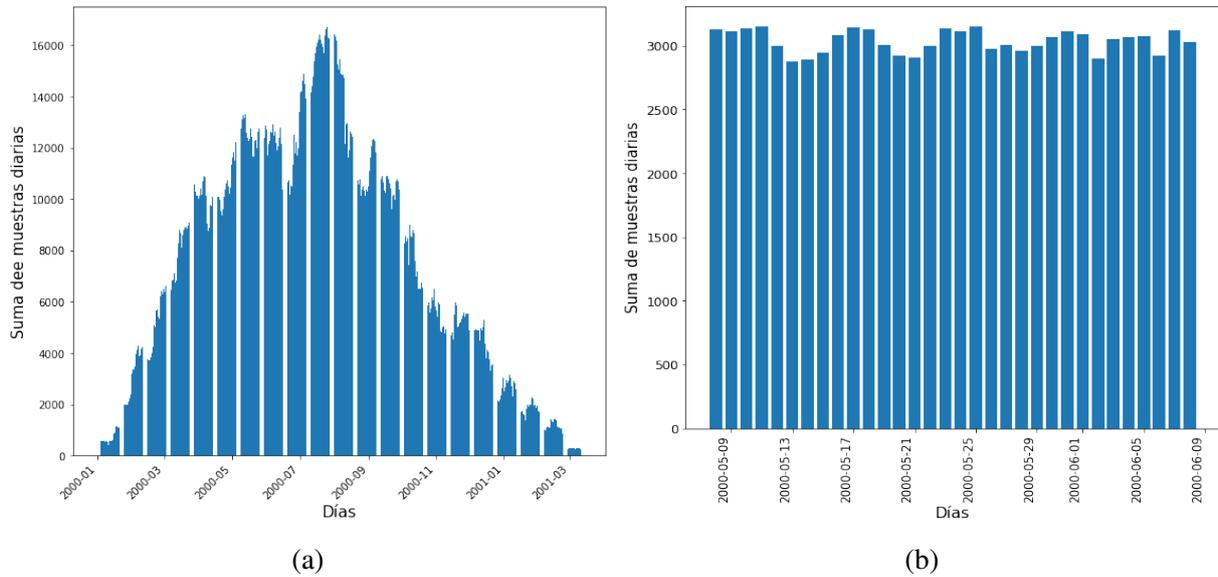


Figura 2.19: Suma de muestras de los pacientes de la base de datos 3 recogidas cada día. (a) Previas a la selección de fechas; (b) Después de seleccionar 32 días.

Una vez seleccionado el periodo más homogéneo, se estudia el número de muestras totales disponibles para cada uno de los pacientes. Para ello se visualiza, como se ve en la Figura 2.20. A continuación, para intentar reducir el número de valores nulos se decide eliminar todos los pacientes con un número de datos inferior a la media (8838 muestras) eliminando así 3 pacientes.

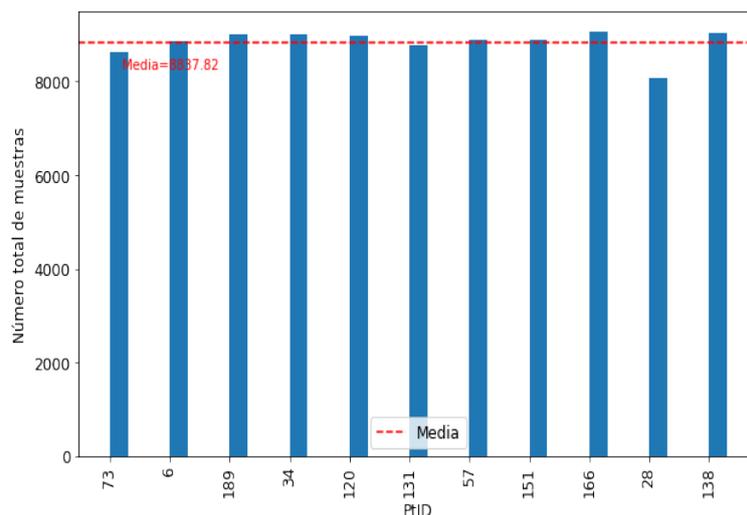


Figura 2.20: Número total de muestras por paciente de la BD3-SEN y su media.

A continuación, se estudia la frecuencia de muestreo y se busca que esta sea siempre 5 mi-

nutos. Por ello, el primer paso es comprobar que, para cada paciente, la diferencia de tiempo entre una muestra y la siguiente sea siempre 300s exactos. En este caso, se eliminan varios sujetos con un alto número de valores nulos, con ayuda del gráfico que se muestra en la Figura 2.21 quedando finalmente 3 pacientes. Además, los pacientes eliminados tenían largos periodos sin muestras. Una vez se tienen los 3 pacientes finales, se realiza al remuestreo de la serie temporal. Para conseguir una frecuencia de muestreo constante de 5 minutos, se introduce un valor NaN en aquellos instantes en los que no hay muestra.

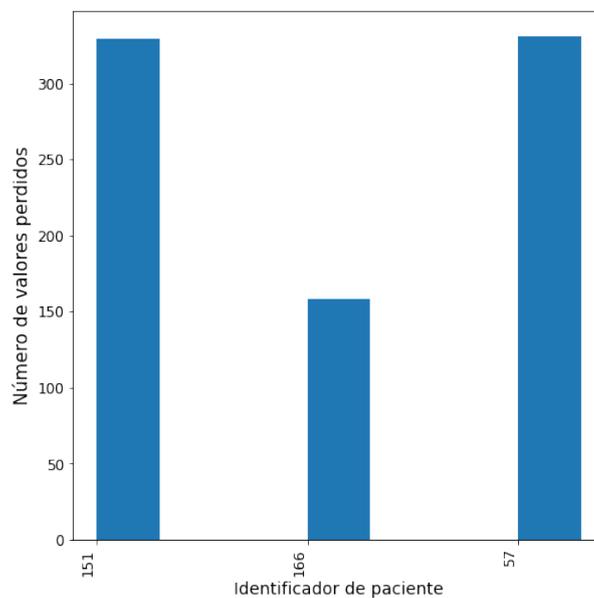


Figura 2.21: Número total de valores perdidos por paciente de la BD3-SEN.

Finalmente, se realiza la imputación, es decir, se sustituyen los valores nulos introducidos por valores de glucosa. Se prueban varios métodos de imputación, como se puede ver en la Figura 2.22, además se puede apreciar como afecta cada uno de los métodos de imputación a la evolución de la serie temporal y con respecto al escenario inicial con *missing data* (véase la Figura 2.22 (a)). Finalmente se imputan los datos mediante RF [41], obteniendo series temporales resultantes como las representadas en la Figura 2.22 (b).

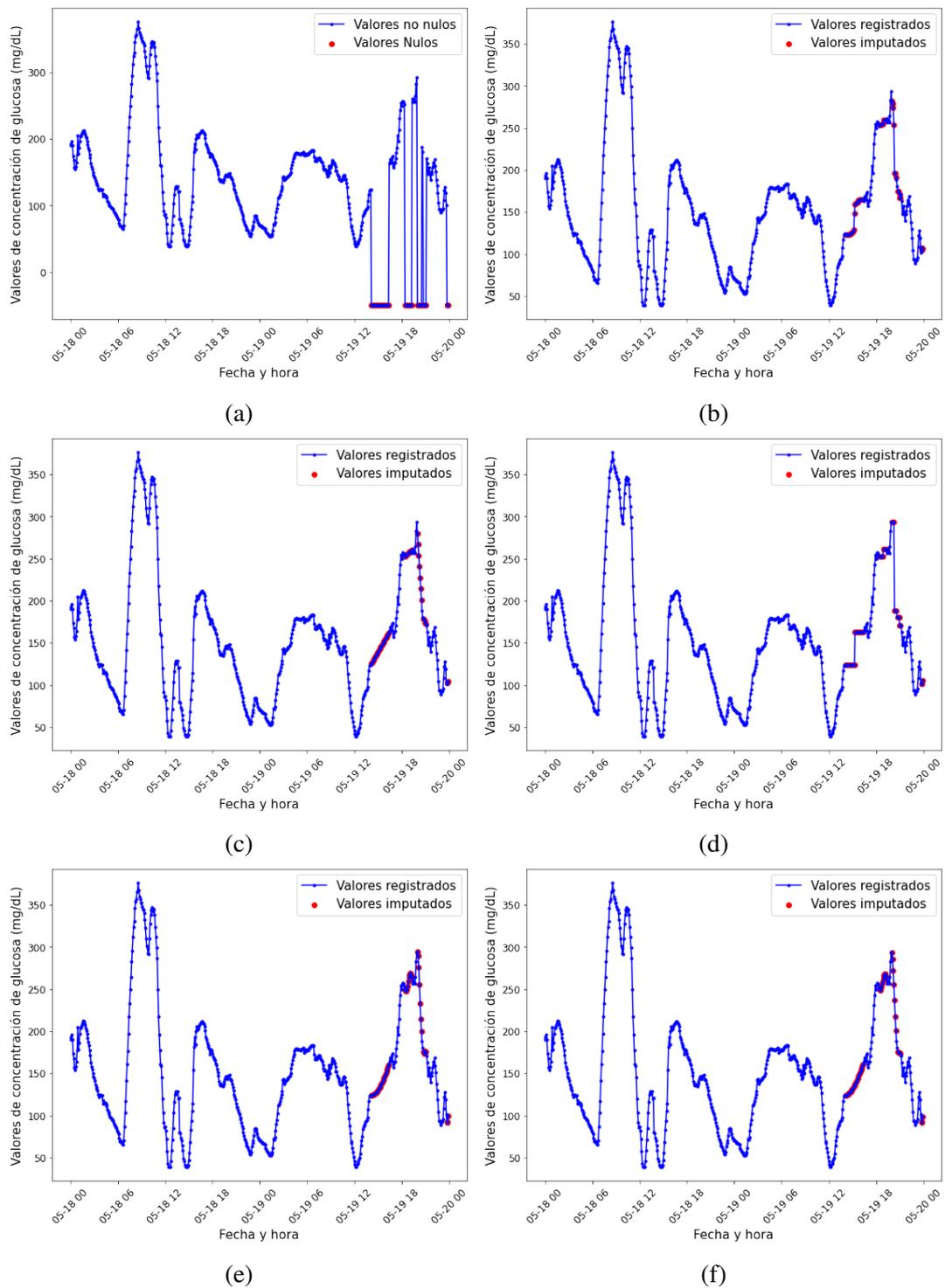


Figura 2.22: Resultados de aplicar diferentes técnicas de imputación para una serie temporal de glucosa de BD3-SEN. (a) Serie temporal sin imputación; imputación basada en: (b) RF; (c) método lineal; (d) KNN; (e) método polinómico de grado 2; (f) método polinómico de grado 3. En el eje x se muestra 'Fecha y hora' en formato: Mes-Día Hora.

Después de la etapa de preprocesamiento aplicado en las tres bases de datos, se tiene en todas las mismas características: valor medido de glucosa en sangre, identificador de paciente y instante temporal en el que se toma la muestra. A pesar de ello, las 3 bases de datos son muy diferentes en cuanto a número de pacientes, número de días de registro y número total de registros, como se resume en la Tabla 2.5. Estos tres aspectos serán muy determinantes a la hora de entrenar los algoritmos.

Identificador base de datos	# pacientes	# días de registro	# total de registros
BD1-SEV	14	6	24,192
BD2-PED	6	45	77,760
BD3-SEN	3	32	27,648

Tabla 2.5: Recopilación del número de pacientes, número de registros totales muestreados cada 5 minutos después del preprocesamiento para cada base de datos.

Capítulo 3

Modelos para generación de series temporales sintéticas y predicción de glucosa

En este capítulo se introducen una serie de conceptos fundamentales que ayudan a entender el marco técnico sobre el que se desarrolla este TFG. En primer lugar se hace una breve introducción sobre el AA, en el que se basan los métodos empleados en este TFG. A continuación se introducen los modelos de generación de series temporales sintéticas considerados, así como los modelos para predicción en series temporales. Finalmente, se detallan las métricas de calidad utilizadas para validar los experimentos llevados a cabo en los dos puntos anteriores.

3.1. Aprendizaje automático

La Inteligencia Artificial (IA) surge en los años 50 y es una rama de la informática que busca crear sistemas informáticos con capacidades similares a las de los seres humanos que demanden inteligencia. Puesto que, uno de los requisitos básicos del comportamiento inteligente es el aprendizaje, surge a continuación el ML como un subconjunto de la IA [44]. Consiste en un proceso de aprendizaje de un ordenador, mediante la utilización de modelos matemáticos de datos. Es decir, aprender de patrones mediante algoritmos, en lugar de hacerlo mediante instrucciones directas [44]. Los modelos ML pueden abordar distintos tipos de tareas, dependiendo de los datos de los que se dispone y del problema a resolver. La salida requerida es lo que determinará el algoritmo a emplear. Los algoritmos de ML se pueden clasificar en dos grupos, según si el

ajuste del modelo es supervisado o no supervisado [45].

- **Aprendizaje supervisado.** Se basa en entrenar una muestra de datos que tiene asignada previamente la clasificación correcta (etiqueta). Es decir, cada observación tiene asignada su etiqueta, el entrenamiento de un método de aprendizaje supervisado utiliza datos etiquetados [44]. Dentro del aprendizaje supervisado, se puede distinguir entre algoritmos de clasificación y de regresión según el tipo de salida (etiqueta). En la clasificación se busca para predecir valores de salida categóricos y la regresión para predecir valores de salida continuos. En el caso de este TFG, se utilizan métodos de regresión ya que la variable que se quiere predecir es el nivel de glucemia en sangre, siendo esta continua y numérica [44].
- **Aprendizaje no supervisado.** El algoritmo de aprendizaje no se calcula el error para evaluar la solución, ya que no se conoce el etiquetado previo y se necesita la variable de salida real para calcular el error. Busca patrones pasados para entrenar el modelo y determinar los resultados de la variable de salida [46].

El aprendizaje profundo (DL, del inglés *Deep Learning*) es una forma especializada de ML que usa redes neuronales artificiales (ANNs, del inglés *Artificial Neural Networks*) que surgen como modelos capaces de aprender por sí mismos ajustándose en función de un error que se calcula en cada etapa del entrenamiento. Cada neurona artificial es una unidad de procesamiento y las neuronas están conectadas entre ellas enviando señales de distinta magnitud en función de los pesos que se van ajustando constantemente durante el entrenamiento de la red [47]. Por lo general, una red neuronal está compuesta por varias capas. En primer lugar, la *capa de entrada*, es la que recibe la matriz de datos de entrada. A continuación las *capas ocultas*, están entre la capa de entrada y la *capa de salida*. Una ANN, puede no tener capas ocultas pero, las redes de DL se caracterizan por tener varias. Estas capas utilizan la combinación de funciones no lineales para determinar la relación entre las características de la matriz de entrada y la salida [48]. Por último, la *capa de salida*, que devuelve el resultado final, se puede ver un ejemplo de una red de DL en la Figura 3.1. Además, las ANN utilizan funciones de pérdida que miden la divergencia entre los valores predichos por la red y los valores reales utilizados en el aprendizaje, de modo que en el entrenamiento se busca minimizar la función de pérdida [49]. En este TFG se utilizarán modelos de DL para la generación de series temporales sintéticas y para la predicción en series temporales.

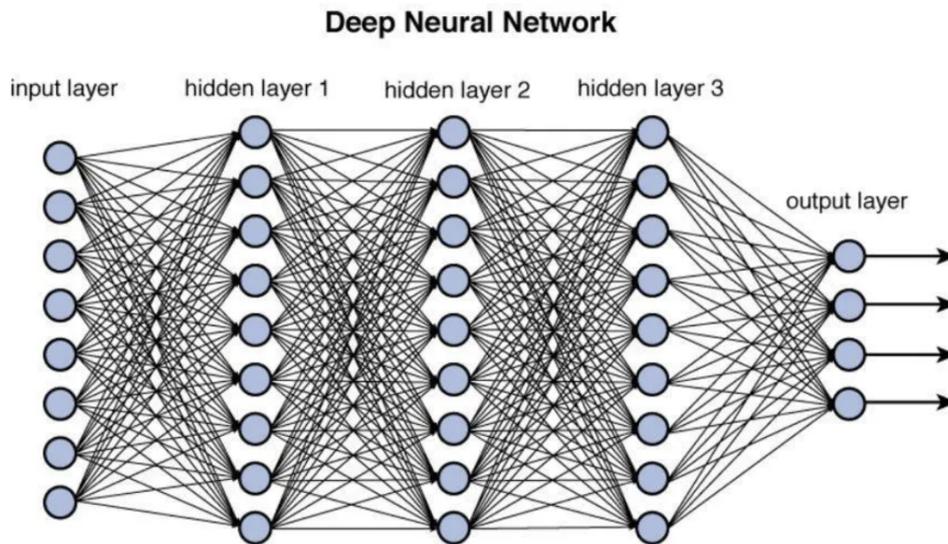


Figura 3.1: Ejemplo de conexión entre capas de una red neuronal con capas ocultas. Figura extraída de [50].

Con el comienzo de la digitalización de datos clínicos y el auge del AA, se comienzan a usar estos algoritmos en conjuntos de datos médicos para diversas aplicaciones. Actualmente, supone una herramienta indispensable en el mundo de la salud, sobre todo desde que gracias a la digitalización, es posible recopilar y almacenar grandes cantidades de datos en formato digital [51]. Los hospitales emiten datos de manera constante, estos datos se almacenan y se comparten en sistemas de información sanitaria. Además de ser almacenados, gracias a la IA y a las técnicas de análisis de datos los datos pueden ser analizados y procesados, esto es especialmente útil para el diagnóstico y el pronóstico [51]. El diagnóstico busca determinar a partir de los datos de un paciente que le ocurre en ese preciso momento, mientras que, el pronóstico busca determinar que le puede ocurrir en el futuro. Una de las grandes tareas llevada a cabo por el AA, con lo que a datos clínicos respecta, es determinar la relación de las características (datos clínicos del paciente que se recopilan) y la salida, en este caso si el paciente esta/estará enfermo o sano. De este modo, se puede saber qué característica está mas asociada con la aparición de una enfermedad concreta [52].

El problema a la hora de trabajar con datos médicos es la calidad de los mismos, ya que, a menudo se enfrentan con problemas de ruido, *missing data*, valores atípicos (en inglés, *outliers*), desbalanceo entre clases (pacientes sanos y enfermos) o escasa cantidad de datos. El rendimiento de los modelos de ML depende de la calidad de los datos, por lo que, la variabilidad de los datos clínicos complica el entrenamiento de los algoritmos, ya que hay que hacer una búsqueda exhaustiva de hiperparámetros para obtener el mejor resultado [53].

3.2. Modelos de generación de series temporales sintéticas

En la actualidad se han mejorado los modelos de generación de datos sintéticos, ya que, anteriormente se generaba un número limitado de datos [54]. Pero, entrenando modelos generativos se consigue tomar cualquier conjunto de datos y generalizarlo para producir otro conjunto [54]. Un ejemplo de estos métodos mejorados son las redes generativas antagónicas (GAN, del inglés *Generative Adversarial Networks*, que han demostrado gran eficacia a la hora de generar datos sintéticos realistas [55]. La estructura de una GAN incluye dos capas, generador y discriminador. El generador crea datos a partir de ruido aleatorio, mientras que, el discriminador tiene como entrada los datos reales y los creados por el generador y debe ser capaz de distinguir entre datos reales y falsos [56], como se puede ver en la Figura 3.2. Una vez el discriminador ha identificado si la muestra enviada por generador es real o no, emite una respuesta categorizando la muestra en real o sintética. Esta respuesta le llega tanto al discriminador como al generador. El generador usa esta información para intentar crear muestras parecidas a las reales y poder engañar al discriminador. Es decir, funcionan con una capa generadora (G) y una discriminadora (D) que se enfrentan, el generador trata de minimizar la función mientras que el discriminador trata de maximizarla. Dicha ecuación se corresponde con tratar de minimizar o maximizar una función basada en las distribuciones, como por ejemplo, la divergencia de Jensen-Shanon [57].

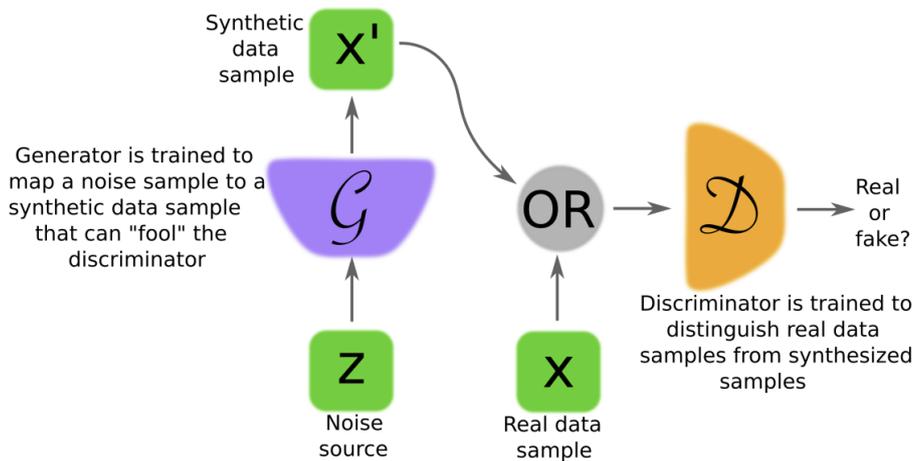


Figura 3.2: Esquema del funcionamiento de las dos partes de las redes GAN, generador y discriminador. Figura extraída de [56].

El éxito de los modelos GAN se debe a que pueden aprender la distribución de bases de datos reales, gracias a ello, las GAN se pueden aplicar en imágenes, series temporales, señales

de audio, señales ECG y datos tabulares [58]. Además, hay varios estudios que utilizan las GAN para generación de datos clínicos, tanto es así que hay una red GAN específica para este propósito: *HealthGen*, creada para generar datos de historia clínica electrónica sintéticos [59]. Se estudia la literatura sobre diversas GAN y se evalúan las distintas características de cada una de ellas, como se muestra en la Tabla 3.1. Finalmente, se decide utilizar para el desarrollo de este TFG dos métodos distintos para generar datos sintéticos: CPAR y DGAN, ya que su implementación se adaptaba bien con los datos disponibles y el tiempo de ejecución de los modelos y coste computacional son bajos.

Red	Muestreo	Valores perdidos	Multivariable	Datos mixtos	Datos clínicos	# muestras	Aplicación
RGAN [55]	Constante	✓	✓	✓	UCI	17,000	Predicción
Time GAN [60]	Irregular	NA	✓	✓	Cáncer	NE	Predicción
Wave GAN [61]	Continuo	NA	×	×	×	NE	Audio
Doppel GAN [62]	Constante	NA	✓	✓	×	Variable	Predicción
COT-GAN [63]	NE	NA	✓	✓	ECG	43	Clasificación
Health Gen [59]	Irregular	✓	✓	✓	✓	>4,000	Clasificación
EHRM GAN [64]	Regular	Imputación	✓	✓	UCI	140,000	Predicción
Time VAE [65]	Constante	NA	✓	×	×	>4,000	Predicción
TTS CGAN [66]	Irregular	NA	✓	×	✓	NE	Clasificación
PSA GAN [67]	Constante	Imputación	✓	×	×	<1,000	NE
CPAR [68]	Constante	NA	✓	✓	×	<25,780	NE

Tabla 3.1: Descripción de las características de las distintas ANNs para la generación de series temporales sintéticas estudiadas. (NA: No Aplica; NE: No Especifica). En negrita se resaltan las redes usadas en este TFG.

3.2.1. *Conditional Probabilistic Autorregressive Model*

El primer modelo basado en ANN que se utiliza en este TFG para la generación de series temporales sintéticas es CPAR (implementada en la librería Python SDV [69]). CPAR consiste en un modelo autorregresivo probabilístico que utiliza una función de pérdida personalizada y permite entradas condicionales [68]. Los modelos autorregresivos asumen que cada observación futura (Y_t) es una combinación de un componente aleatorio ε y la combinación lineal de las observaciones pasadas (Y_{t-1}) [70]. Siendo α el coeficiente de auto regresión un modelo auto regresivo se define por la siguiente función: $Y_t = \alpha Y_{t-1} + \varepsilon_t$.

Una de las grandes ventajas de esta implementación, es que permite varios tipos distintos de datos secuenciales: datos relacionados, multiseuenciales y datos contextuales. Los datos relacionados son aquellos en los que, cada fila de la secuencia tiene dependencia de las anteriores. Se consideran datos multiseuenciales cuando en un mismo conjunto de datos hay múltiples secuencias independientes unas de otras. Los datos contextuales son aquellos que son constantes dentro de una misma secuencia [68]. En concreto en este caso se utilizan múltiples series tem-

porales, es decir, se consideran los valores de glucosa de los distintos pacientes como una serie temporal independiente, ya que, cada paciente tiene un identificador único en la característica *PtID* y esta implementación de CPAR permite imponer *PtID* como identificador de cada serie.

La implementación de CPAR sigue un esquema como mostrado en la Figura 3.3. Comienza con una etapa de preprocesamiento, seguida del modelado y por último postprocesamiento [68]. Al tratarse de una ANN, es muy importante el preprocesamiento ya que los datos de entrada a la red deben normalizarse de tal modo que sean fácilmente tratables por los procesos matemáticos y estadísticos que forman la red. En el caso de este TFG, se tiene una única variable de carácter continuo, el nivel de glucosa en sangre medido. Es por ello que el preprocesado aplicado es la normalización, que consiste en escalar los datos para que estos se conviertan en proporciones comparables. Existen varios métodos de normalización, en esta implementación se utiliza *z-score* para normalizar datos continuos. Siendo x un valor de glucosa, μ la media de los valores de glucosa y σ su desviación estándar, la normalización *z-score* se define como: $Z(x) = \frac{x-\mu}{\sigma}$. El postprocesamiento consiste en invertir el preprocesamiento para que los datos sintéticos resultantes tengan el mismo formato que los reales antes de entrar al modelo.

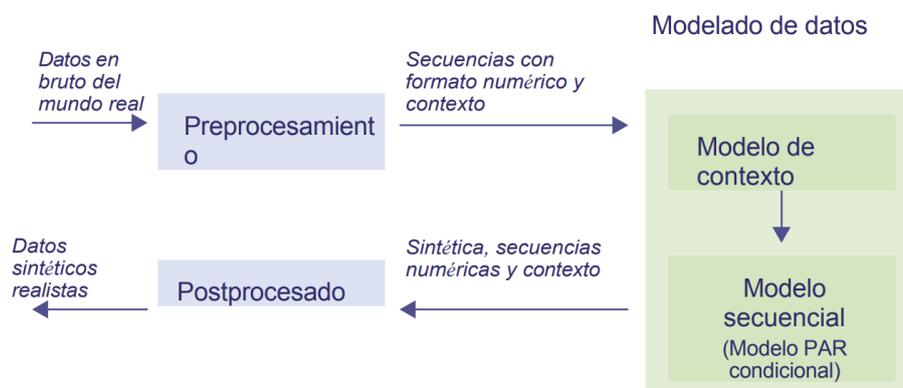


Figura 3.3: Diagrama de trabajo de la red neuronal CPAR. Figura extraída de [68].

El modelado se realiza con CPAR que es un modelo secuencial que busca encontrar las dependencias entre las filas de la serie temporal y sintetizarlas [68]. Para ello, hay que introducir en el modelo la serie temporal y el contexto (en inglés, *metadata*). En el *metadata* se describe cada variable y se clasifica según el tipo de característica que es (categórica, continua, identificador secuencial, identificador de serie, etc.). El entrenamiento se lleva a cabo fila por fila dentro de cada secuencia (S) siendo una secuencia la serie temporal asociada a un identificador de secuencia específico, en este caso el identificador de paciente. Una vez se ha generado una fila nueva, sintética, para la primera fila, esta pasa a formar parte del historial y se ajustan

los parámetros para proceder a repetir el proceso con la fila siguiente como se muestra en la Figura 3.4.

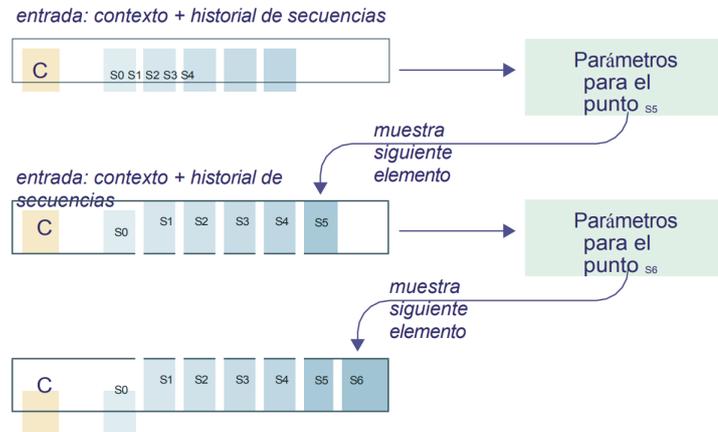


Figura 3.4: Entrenamiento por filas del modelo CPAR. Figura extraída de [68].

Al tratarse de un modelo basado en ANN, durante el entrenamiento se estiman los parámetros para crear cada secuencia π y la red se actualiza con las estimaciones. Además, entrena en base a una función de pérdida que representa la similitud de los parámetros de salida con el valor real, en este caso, se evalúa con una función de pérdida propia, que consiste en la suma de todas las pérdidas individuales de cada secuencia definida como:

$$Loss(L) = \sum_i \sum_t \sum_{j=0}^{k-1} L(S_t^{(i)}, \pi_{(t,j)}^{(i)}) \quad [68].$$

Donde:

- i : cada una de las secuencias S , es decir, pacientes.
- t : cada una de las filas de la secuencia, cada instante temporal.
- j : todos los parámetros para cada fila $\pi_{(t,j)}^{(i)}$, donde cada parámetro $\pi_{(t,j)}^{(i)}$ describe la densidad de probabilidad para todas las dimensiones.

En concreto, para el caso de las variables continuas como la glucosa en sangre en este TFG, la función de pérdida (L), donde: μ y σ son la media y desviación estándar, respectivamente, de una Gaussiana, y m es la probabilidad de que el valor sea un valor perdido. Se define como [68]:

$$L(x; \mu, \sigma, m) = -(\log(f_{\mu, \sigma^2}(x)) + \log(1 - m)) \text{ cuando } x \text{ existe}$$

$$L(x; \mu, \sigma, m) = -\log(m) \text{ cuando } x \text{ es un valor perdido}$$

3.2.2. *Doppel Generative Adversarial Network*

El segundo modelo basado en ANN que se emplea para la generación de series temporales sintéticas es DGAN. Este modelo surge para mejorar la fidelidad de los datos generados con métodos GAN tradicionales. Para ello, DGAN opta por realizar la generación de *metadata*, siendo *metadata* la información sobre los datos, en cada instante de tiempo de manera independiente a los datos empleando un generador y un discriminador independientes para el *metadata*. Por otro lado, genera muestras por bloques de un determinado tamaño que se auto-ajusta en el entrenamiento del modelo. Otras GAN generan bloques individuales, como se puede ver en la Figura 3.5, esto mejora el rendimiento computacional del modelo además de prevenir que este memorice, mejorando así la generalización. Por último, se generan máximos y mínimos aleatorios y una serie normalizada que posteriormente se escala al rango real, tratando de evitar el modo colapso [62].

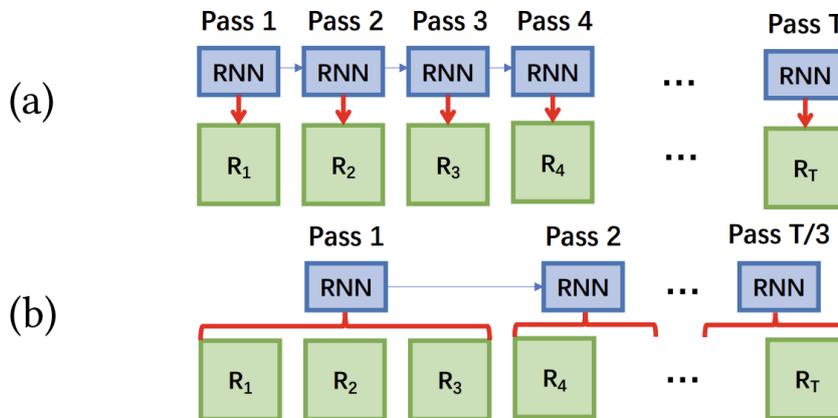


Figura 3.5: Entrenamiento de redes neuronales para generación de series temporales sintéticas. (a) Empleando GAN tradicionales utilizando un instante de tiempo cada vez; (b) Empleando DGAN entrenando en bloques de 3 instantes de tiempo simultáneamente. Figura extraída de [62].

Las GAN originales no logran capturar correlaciones temporales a largo plazo, ya que están basadas en el perceptrón multicapa que no es robusto para series temporales [71]. DGAN utiliza una ANN recurrente específicamente diseñada para tratar con series temporales, específicamente el modelo *Long Short Term Memory (LSTM)* que surge para abordar las limitaciones

de memoria de las redes neuronales recurrentes y permitir memoria a corto y a largo plazo teniendo en cuenta datos relevantes del pasado [71]. El problema del modo colapso consiste la generación de muestras sintéticas homogéneas por parte de los modelos cuando las muestras reales no son homogéneas. Para solucionar este problema y crear muestras diversas cuando los datos de entrenamiento no son homogéneos, en el modelo DGAN se opta por normalizar cada señal de serie temporal individualmente. Se almacenan los mínimos y los máximos como *metadata* y se utiliza el *metadata* para entrenar, además, se vuelve a utilizar en el post procesamiento para re-escalar los datos al rango real [62]. La última tarea que aborda DGAN es la generación independiente de *metadata*, es decir, se divide en dos tareas cada una con un generador: (1) para generar *metadata*; (2) para generar los datos sintéticos condicionados al *metadata*. En concreto, el generador de *metadata* es el estandarizado en las GANs primitivas, basado en el perceptrón multicapa. Como tener dos generadores independientes no resolvía el problema, finalmente se decide introducir un discriminador auxiliar [62]. Se puede consultar la Figura 3.6 para visualizar la estructura completa del modelo DGAN.

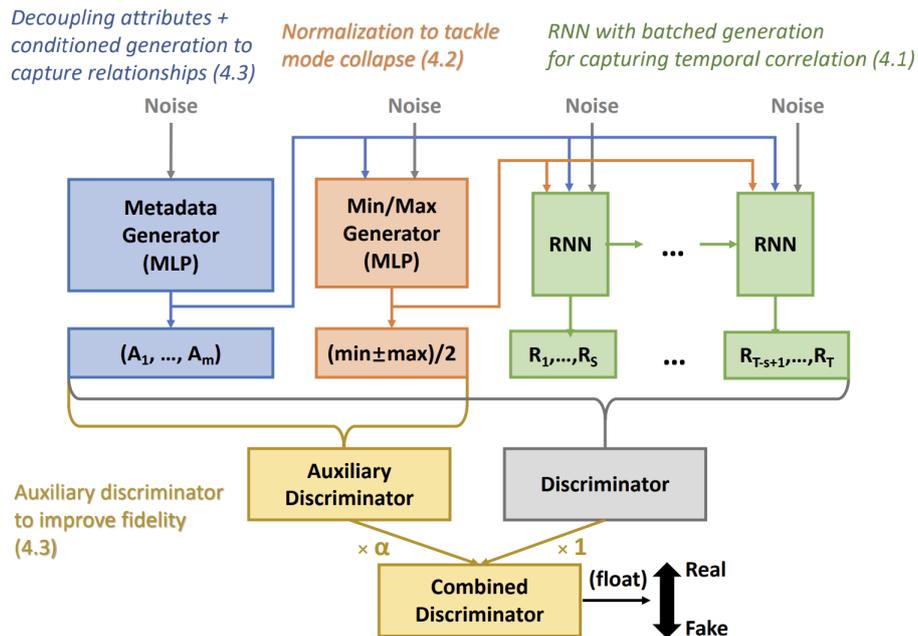


Figura 3.6: Estructura general de DGAN. Figura extraída de [62].

3.2.3. Métricas para evaluar la calidad de series temporales sintéticas

En esta sección se introducen las métricas empleadas en este TFG para evaluar la calidad de las series temporales sintéticas generadas con los modelos CPAR y DGAN. Hay diversas métricas utilizadas para determinar la bondad de una estimación puntual, la mayor parte de ellas se basan en comparar la función de densidad de probabilidad (PDF, del inglés Probability Density Function) de los datos reales con los datos sintéticos generados. En concreto, hemos considerado las siguientes:

- **Divergencia de Jensen-Shannon.** La divergencia de *Jensen-Shannon* mide la similitud entre dos PDFs P y Q , y se basa en la divergencia de *Kullback-Leibler* [72]. La divergencia de *Kullback-Leibler* está definida como: $\sum_i p_i \log \frac{P_i}{Q_i}$ [72]. La divergencia de *Jensen-Shannon* (D_{JS}), es simétrica y está acotada entre 0 y 1 ($0 \leq D_{JS}(P, Q) \leq 1$). Está formulada como [72]:

$$D_{JS}(P, Q) = \frac{1}{2} \left[S \left(P, \frac{P+Q}{2} \right) + S \left(Q, \frac{P+Q}{2} \right) \right]$$

- **Maximum Mean Discrepancy (MMD).** Es un test estadístico que se utiliza para determinar si los datos reales y sintéticos provienen de la misma distribución estadística [73]. Partiendo de que la hipótesis nula (H_0) sugiere que los datos sintéticos siguen la misma distribución que los reales y la hipótesis alternativa (H_1) sugiere lo contrario. Este test comprueba si las distribuciones de probabilidad de los datos reales y sintéticos se parecen en base a la diferencia entre los valores promedio de las muestras. Cuando este valor es grande, se asume que vienen de distribuciones diferentes. Para calcular el estadístico MMD, se define un *kernel* a partir del cual se determina la divergencia entre las distribuciones, en el caso de este TFG se utiliza un *kernel* basado en la distribución Gaussiana [73]. El MMD calcula como:

$$MMD = [F, p, q] = \mathbb{E}_{x, x'} [k(x, x')] - 2\mathbb{E}_{x, y} [k(x, y)] + \mathbb{E}_{y, y'} [k(y, y')]$$

Donde x y x' son variables aleatorias independientes de la distribución p e y y y' son variables aleatorias independientes de la distribución q [73]

- **Distancia de Kolmogorov-Smirnov.** Se trata de un test estadístico no paramétrico [74]. En el, se mide la divergencia entre la función de distribución acumulada (CDF, del inglés *Cumulative Distribution Function*) de los datos reales, que se utiliza de referencia, con

respecto a la CDF de los sintéticos [74]. Siendo la hipótesis nula (H_0) que los datos sintéticos siguen la misma distribución que los reales y la hipótesis alternativa (H_1) que no siguen dicha distribución [74]. Aplicando la formula del test, que viene dada por: $KS = D\sqrt{n}$. Donde: D es la distancia máxima entre las funciones de distribución acumuladas sintética y real, y n es el número total de muestras [74]. Finalmente, se calcula el p_{valor} asociado al test de modo que: si $p_{valor} < 0,05$ se acepta H_0 y es correcto asumir que las dos distribuciones son muy similares, mientras que si $p_{valor} > 0,05$, se acepta H_1 .

- **Dynamic Time Warping (DTW)**. Se trata de una métrica muy popular para comparar series temporales ya que mide tanto la distancia como la deformación que se produce, en este caso, en la serie sintética sobre la real [75]. La fórmula se define como:

$$DTW(X, Y) = \min_{\pi} \sqrt{\sum_{(i,j) \in \pi} d(X_i, Y_j)^2}$$

donde $d(X_i, Y_j)$ es la distancia calculada entre los elementos X_i e Y_j que forman la serie temporal real y sintética, respectivamente y π es un parámetro que minimiza la distancia entre las secuencias [76].

3.3. Modelos de predicción en series temporales

Los modelos de predicción en series temporales son aquellos que predicen los valores futuros de un valor objetivo asociado a un instante temporal, en el caso de este TFG, se predicen valores de glucemia en sangre en distintas ventanas temporales con muestras tomadas cada 5 minutos. Uno de los métodos de predicción más sonados en la literatura es *AutorRegressive Integrated Moving Average* que es un modelo lineal estadístico que se basa en la hipótesis de medias móviles y utiliza la variación estacional del pasado para pronosticar el futuro [70]. No obstante, existen modelos de predicción que presentan buenos resultados basados en DL, estos modelos aprenden las relaciones predictivas mediante capas no lineales. Para lograr estas relaciones, se construyen representaciones intermedias de las características a lo largo del tiempo mediante la codificación de información histórica relevante [77]. Pese a los grandes avances con las ANN en los modelos predictivos, el pronóstico a largo plazo sigue siendo un desafío.

N-BEATS por sus siglas en inglés *Neural Basis Expansion Analysis for interpretable Time Series Forecasting* (implementado en la librería Python *darts* [78]), es un método de pronóstico novedoso basado en ANN. Consiste en, como dice su nombre, la expansión base que es un

método para aumentar los datos y modelar relaciones no lineales [78]. Hay diversos métodos de expansión, por ejemplo: polinómica, logarítmica, etc. Que hacen que los datos del modelo se ajusten a la función utilizada. En el caso de N-BEATS se, busca la mejor expansión base para los datos con los que se entrena el modelo y se realizan las predicciones. La arquitectura de este modelo se muestra en la Figura 3.7. En primer lugar, la serie temporal del conjunto se divide en dos periodos: retrospectivo y prospectivo. El retrospectivo es el conjunto con el que se entrena el modelo y son datos de instantes de tiempo pasados, mientras que, el prospectivo son instantes de tiempo futuros y se utiliza para evaluar el rendimiento del modelo durante el entrenamiento [78]. El tamaño del periodo retrospectivo es proporcional al del periodo prospectivo (n) y la longitud de ambos periodos se define en los hiperparámetros del modelo. Además, se puede ver como el resultado global de la predicción es el sumatorio del resultado de un número M de *stacks*, el número de *stacks* se define en los hiperparámetros del modelo. Sigue la arquitectura de DenseNet por la cual todos los *stacks* están conectados, la salida de un *stack* forma parte de la entrada del siguiente y así sucesivamente. Del mismo modo ocurre en los bloques que forman un *stack*, pero además, en este caso, se elimina del entrenamiento la parte que el bloque anterior pueda aproximar correctamente, ya que, cada bloque devuelve un pronóstico parcial reduciendo el tiempo de trabajo y la carga computacional [78].

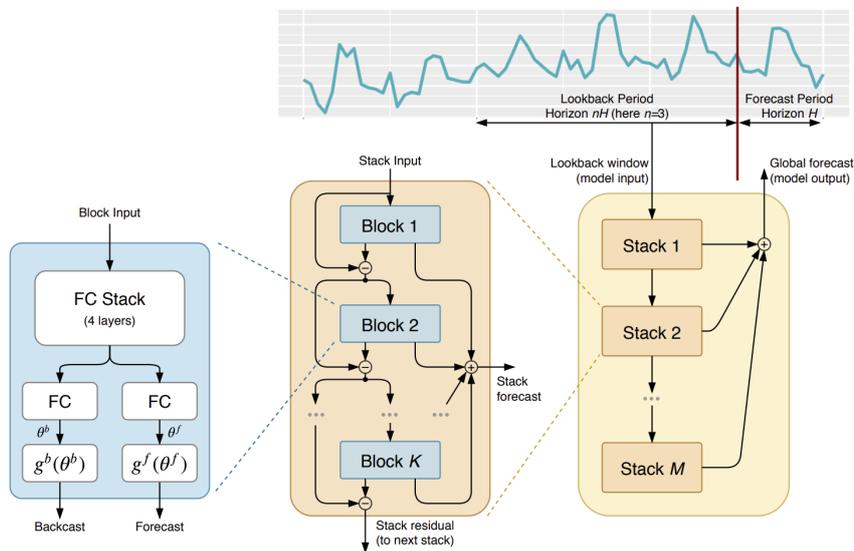


Figura 3.7: Estructura general de N-BEATS. Figura extraída de [78].

En este TFG se utiliza un método, también basado en ANN y que surge como mejora de N-BEATS: N-HiTS que obtiene mejores resultados con menor coste computacional. Para lograrlo, se muestrea la serie temporal a diferentes tasas y así el modelo consigue aprender a corto y lar-

go plazo. Además, al realizar las predicciones, combina los pronósticos de las distintas escalas. Esto se llama interpolación jerárquica y es la mayor mejora que presenta N-HiTS. En cuanto a la arquitectura, como se puede ver en la Figura 3.8, es muy parecida a la de N-BEATS. Divide las secuencias de datos en retrospectiva y prospectiva y funciona por *stacks* que están conectados entre ellos, formados a su vez por bloques. La predicción final, al igual que en el modelo anterior, es la suma de las predicciones de cada *stack* [79]. La primera diferencia es que, dentro de cada bloque hay una capa de *maxpooling*, el *maxpooling* es una técnica para reducir la dimensionalidad producida por las capas de la ANN, pero, manteniendo las características más importantes, para ello, divide los datos en ventanas y se queda con el valor máximo de cada ventana [80]. El tamaño de dichas ventanas es un hiperparámetro del modelo y afecta a la predicción de manera considerable ya que, una ventana muy grande magnifica las características a largo plazo, mientras que uno muy pequeño lo hace a corto plazo. Es por esto, que N-HiTS muestra mejor rendimiento a largo plazo. La otra gran diferencia es que, como dice en su nombre, se emplea interpolación jerárquica. Esto se utiliza para reducir el número de muestras a predecir en la serie temporal, es decir, coge grupos más grandes de muestras pasadas para hacer la predicción, consiste en disponer de distintas frecuencias en los distintos *stacks* [79]. Por ejemplo, en este TFG se tienen muestras cada 5 minutos, si se quiere realizar una predicción a 1 día son 24 horas y 288 muestras (12 muestras cada hora). Si para ello se fija en una ventana pasada también de un día, utilizando la interpolación jerárquica en vez de tomar 288 bloques uno por muestra, puede coger por ejemplo 24 bloques de 12 muestras cada uno. Esto es el resultado de la interpolación jerárquica y la capa de *maxpooling* y gracias a ello se consigue reducir muy considerablemente el coste computacional. En la Figura 3.9 se puede ver como cada *stack* tiene un número de predicciones diferente, esto se debe a la interpolación jerárquica.

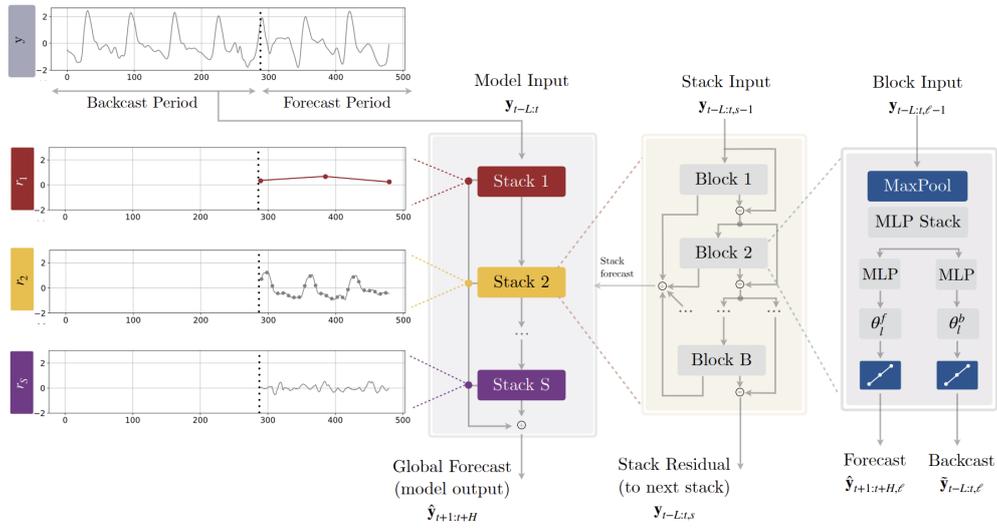


Figura 3.8: Estructura general de N-HiTS. Figura extraída de [79].

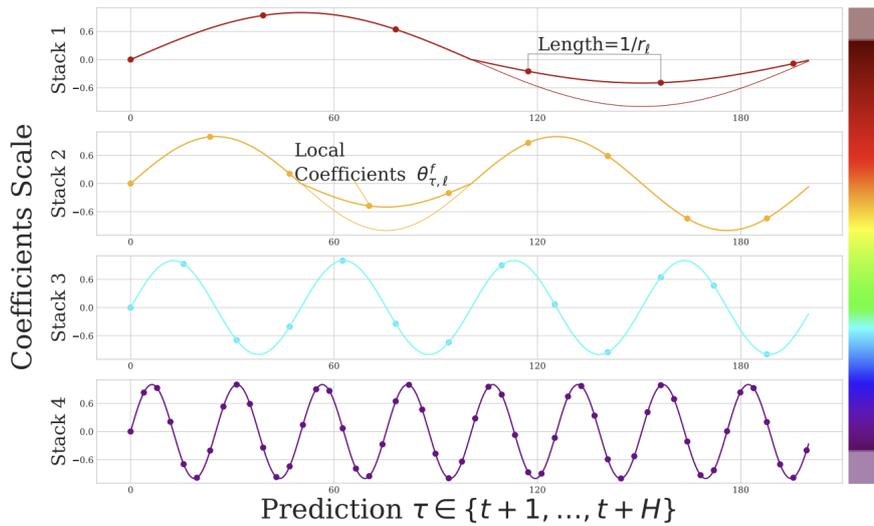


Figura 3.9: Series temporales estimadas por distintos *stacks* y su variación en el número de muestras debido a la interpolación jerárquica. Figura extraída de [79].

3.3.1. Métricas para evaluar predicción en series temporales

Para evaluar las prestaciones de los modelos de predicción en series temporales diferentes figuras de mérito que miden el error del valor estimado y real han sido consideradas. En particular, se usaron errores relativos y errores absolutos [81]. Los errores relativos se basan en

calcular la diferencia entre el valor pronosticado por el modelo y el valor real [81]. En este caso se calcula el error entre el valor de glucosa estimado por N-HiTS y el valor de glucosa real. Mientras que, el error relativo relaciona el error absoluto y el real escalándolos, es decir, dividiéndolo entre el valor real [81]. En concreto se utilizaron: el error cuadrático medio (MSE, del inglés *Mean Squared Error*), el error absoluto medio (MAE, del inglés *Mean Absolute Error*), el error absoluto relativo medio (MRAE, del inglés *Mean Relative Absolute Error*), el *Clarke Error Grid* (CEG) y la norma ISO-15197:2015.

- **MSE.** El MSE está formulado como: $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$. Donde: n es el número total de muestras, en este caso valores de glucosa, y_i es el valor real observado en el instante de tiempo i y \bar{y} es el valor predicho por el modelo en el instante de tiempo i . Es decir, promedio de la diferencia al cuadrado del valor real y el predicho, para todos los instantes de tiempo de la serie temporal. Este error es muy útil en casos en los que hay valores atípicos que es necesario detectar, ya que, al estar elevado al cuadrado resalta el error [82]. Este error es muy representativo en este TFG ya que es muy importante capturar los valores atípicos, estos no se deben a errores, los valores altos o bajos de glucemia representan respectivamente eventos de hiper o hipoglucemia. La predicción será mejor cuanto más cercano a 0 sea el error.
- **MAE.** El MAE, se formula como: $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$. Es decir, consiste en calcular el promedio del valor absoluto de la diferencia entre los valores reales (y_i) y los valores predichos (\hat{y}_i) siendo i cada instante temporal de la serie de tiempo de glucosa. Al contrario del MSE, este error es útil de calcular cuando no se quieren destacar los outliers, cuando estos representan un valor erróneo [82]. La predicción será mejor cuanto más cercano a 0 sea el error.
- **MRAE.** Consiste en relativizar el MAE, es decir, dividirlo por el valor real (y_i) y se formula como $MRAE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}$. Consiste en otra forma de medir la precisión del pronóstico que surge para mejorar la interpretabilidad del funcionamiento del modelo de pronóstico al tratarse de un error relativo. Sin embargo, tiene un gran inconveniente y es que si la muestra real es 0 el error es infinito, aunque existen métodos de filtrado que eliminan los valores extremos para evadir este problema [83]. La predicción será mejor cuanto más cercano a 0 sea el error.
- **CEG.** Se trata de un análisis de error para la DMT1 en el cual se dividen 5 zonas, de la A a la E, en función del riesgo para el paciente en caso de que el dispositivo falle, siendo A la zona de riesgo nulo y E la zona de mayor riesgo. Si el dispositivo registra valores

en la zona E el dispositivo falla y puede dar lugar a consecuencias peligrosas [84]. Se puede ver un ejemplo de CEG en la Figura 3.10. La predicción será mejor cuando mayor número de datos predichos se encuentren en la zona A.

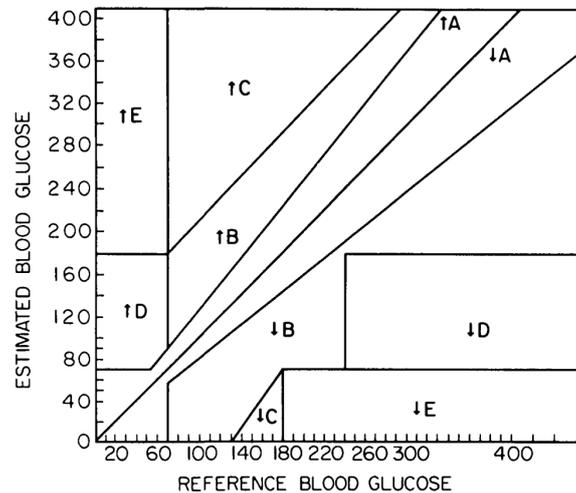


Figura 3.10: Distribución de las zonas de riesgo de CEG. Figura extraída de [85].

- Norma ISO 15197:2015:** Es una norma europea que regula los criterios mínimos que deben cumplir los sistemas de monitorización de glucosa en sangre, hay dos criterios. El primero dice que el 95 % de los valores de glucosa registrados deben estar en el rango de error de $\pm 15 \text{ mg/dL}$ respecto a las mediciones de referencia, cuando el nivel de glucosa es inferior a 100 mg/dL. Cuando es igual o mayor a 100 mg/dL, el error es un porcentaje y debe ser $\pm 15 \%$ [86]. El segundo requisito es, que el 99 % de los valores medidos deben estar en las zonas A y B del CEG [84]. En la Figura 3.11 se pueden ver los rangos permitidos de niveles de glucosa en sangre por la ISO 2003 (amarillo) y la ISO 2015 (rojo), se puede apreciar como las nuevas normas son más restrictivas.

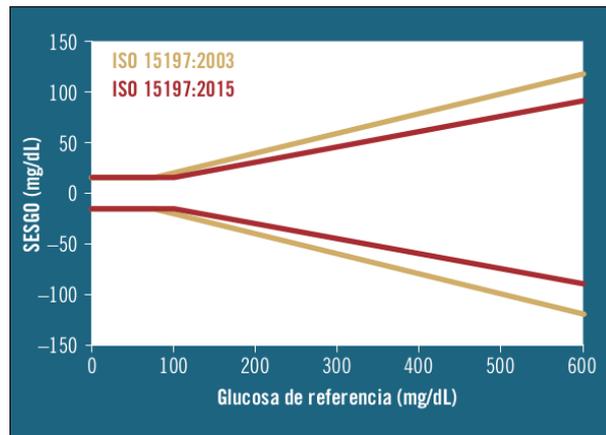


Figura 3.11: Rangos de aceptación para las mediciones de glucosa en sangre por dispositivos CGM según la ISO 2003 y ISO 2015. Figura extraída de [84].

Capítulo 4

Experimentos y resultados

En este capítulo se presentan los experimentos realizados a partir de los conjuntos de datos y los modelos descritos en capítulos previos. La finalidad es evaluar los modelos DGAN y CPAR para la generación de series temporales sintéticas de datos de glucosa.

4.1. Software utilizado

Para la realización de este TFG se ha empleado Python (versión 3.7.13) y en concreto se han utilizado las siguientes librerías: numpy [87] (versión 1.21.6) y pandas [88] (versión 1.5.3) ambas utilizadas para el preprocesado y el análisis descriptivo de las bases de datos, matplotlib [89] (versión 3.6.3), la cual ha sido usada para la visualización de figuras como la evolución de las series temporales, así como seaborn [90] (versión 0.11.2). Además se utiliza la librería SDV, para la implementación del método CPAR [69] (versión 1.1.0) y gretel-synthetics (versión 0.20.0) para la implementación de DGAN [91]. Se utiliza darts (versión 0.24.0) para la implementación de N-HiTS [92]. Por último, se utiliza la librería scikit-learn (versión 1.2.1) para la obtención de figuras de mérito [93].

4.2. Generación de series temporales sintéticas

En esta sección se muestran los resultados obtenidos de la generación de series temporales sintéticas mediante los modelos CPAR y DGAN. Para la evaluación, se comparan ambos modelos empleando las métricas descritas en el Capítulo 3.3.1 y se utilizan las tres bases de datos con datos de glucosa previamente detalladas.

4.2.1. Generación de series temporales sintéticas usando BD1-SEV

En primer lugar, se evalúan los resultados obtenidos tras la generación de series temporales sintéticas para la base de datos BD1-SEV. Principalmente se busca determinar el parecido de las PDFs de datos de glucosa de las series temporales generadas por los modelos CPAR y DGAN, a la serie temporal real. En la Figura 4.1 se puede ver el histograma superpuesto de la distribución de datos serie real y la sintética para un paciente concreto (con identificador PtId=9). En la Figura 4.1 (a) se ve conjuntamente la distribución de glucosa de la serie temporal sintética generada por DGAN y la distribución de glucosa de la serie real. En la Figura 4.1 (b) se muestra el histograma de la distribución de datos de la serie sintética generada por CPAR frente a la real. En el eje de abscisas de los histogramas se muestra los valores de glucosa y en el eje de ordenadas el número de veces que aparece ese valor a lo largo de la serie temporal. Como se observa, el modelo CPAR durante su entrenamiento aprende mejor los valores extremos de glucosa (<50 mg/dL, y >300 mg/dL), sobre todo los mínimos. En general, ambos modelos presentan dificultades para aprender y capturar los valores altos de glucosa. Cabe destacar que la serie sintética generada por DGAN mantiene la media en los mismos rangos que la real [100,150] (mg/dL), mientras que, la generada por CPAR sitúa la media ligeramente más hacia la derecha [150,200] (mg/dL).

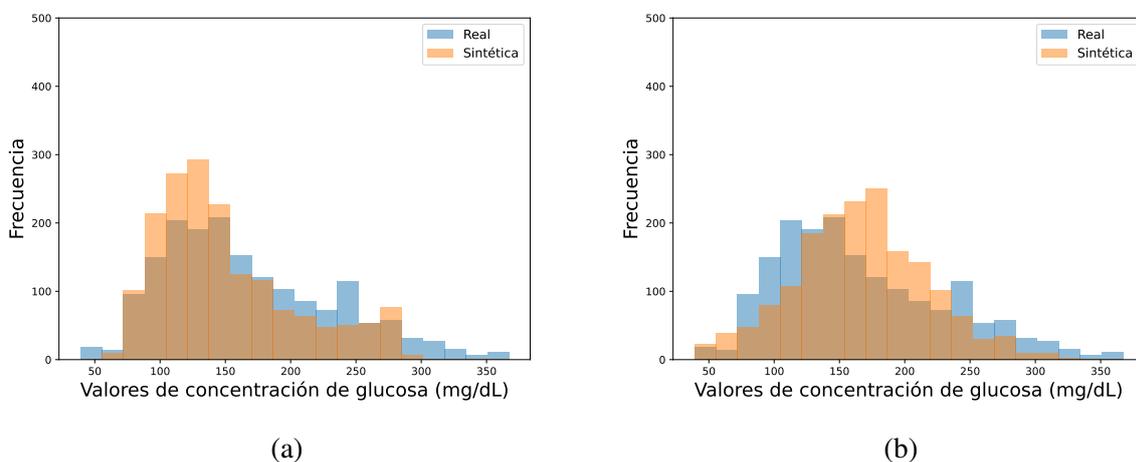


Figura 4.1: Histogramas de la distribución de datos de glucosa de las series temporales reales de un paciente de la BD1-SEV y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

A continuación, se muestra los resultados del análisis de divergencia de *Jensen-Shannon*. Como se explicó previamente, esta divergencia cuantifica la diferencia entre la distribución de

la serie temporal sintética de la real. Se evalúa a lo largo del tiempo, es decir, se toman los valores de cada día y se calcula divergencia. Este proceso es gradual, considerando 2 días, 3, etc. Basándonos en los valores de esta métrica (en torno a 0.15), se puede decir que las series temporales sintéticas son más similares a las reales. En la Figura 4.2 (a) se puede ver los resultados de divergencia de *Jensen-Shannon* para cada serie temporal asociada a cada paciente sintético creado con DGAN, y en la Figura 4.2 (b) las de los pacientes sintéticos creados con CPAR. Pese a que análisis contempla múltiples series temporales con resultados variables, se puede decir que la mayoría de las series temporales toman valores bajos de divergencias. Teniendo en cuenta esto, si se comparan ambos escenarios, se puede observar que son muy similares, aunque, la divergencia es menor para los datos sintéticos generados por el modelo DGAN.

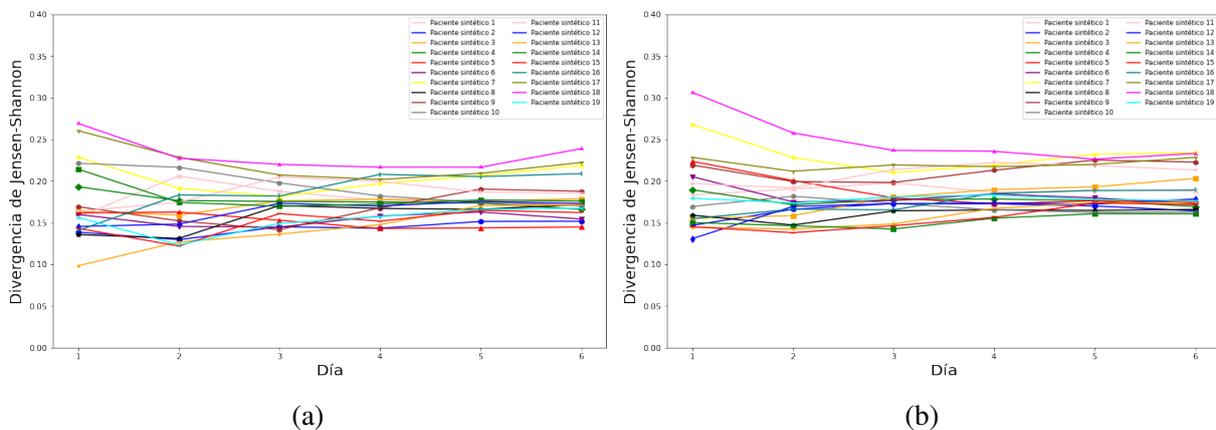


Figura 4.2: Resultados de divergencia de *Jensen-Shannon* entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD1-SEV y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

La Figura 4.3 muestra los resultados de MMD que mide la divergencia (similitud) entre las distribuciones de glucosa de las series reales y sintéticas. Un valor de MMD cercano a 0 indica que la PDF de glucosa de la serie sintética será más similar a la real. En la Figura 4.3 (a) se ve el MMD para las distribuciones de datos de glucosa de las series temporales generadas mediante DGAN, y en la Figura 4.3 (b) para las generadas por CPAR. Para los dos casos, el rango de valores de MMD es cercano a 0, lo que señala que las PDFs son muy similares. Cabe destacar que las series sintéticas generadas con CPAR tienen valores menores, ya que el valor máximo es en torno a 0.02, mientras que una de las series generadas con DGAN alcanza 0.16.

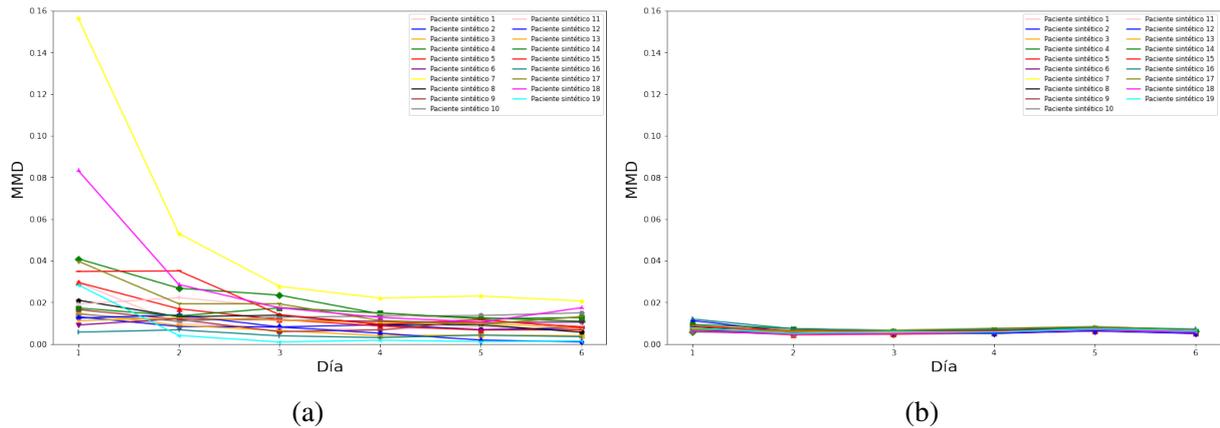


Figura 4.3: Resultados de MMD entre distribución de datos de glucosa de las series temporales reales de la base de datos BD1-SEV y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

A continuación, se aplica el test de *Kolmogorov-Smirnov*, a partir del cual se acepta o rechaza la hipótesis de que las distribuciones de datos de las series sintéticas son similares a las reales. Para validar esta hipótesis se calcula el p_{valor} , siendo la hipótesis aceptada cuando $p_{valor} < 0,05$. Como se puede ver en la Figura 4.4 las PDFs de todas las series sintéticas generadas por los modelos DGAN (Figura 4.4 (a)) obtienen un $p_{valor} < 0.05$ tras aplicar el test. En el caso de las PDFs de las series sintéticas generadas con CPAR (Figura 4.4 (b)), obtienen un p_{valor} mayor en el primer día. Hay incluso un paciente para el cual el $p_{valor} > 0.05$ y por tanto, no es correcto asumir que la serie sintética es similar a la real. Así pues, en todos los casos de DGAN se acepta la hipótesis y, por consiguiente, se puede afirmar que las distribuciones de las series sintéticas son muy similares a las de las reales.

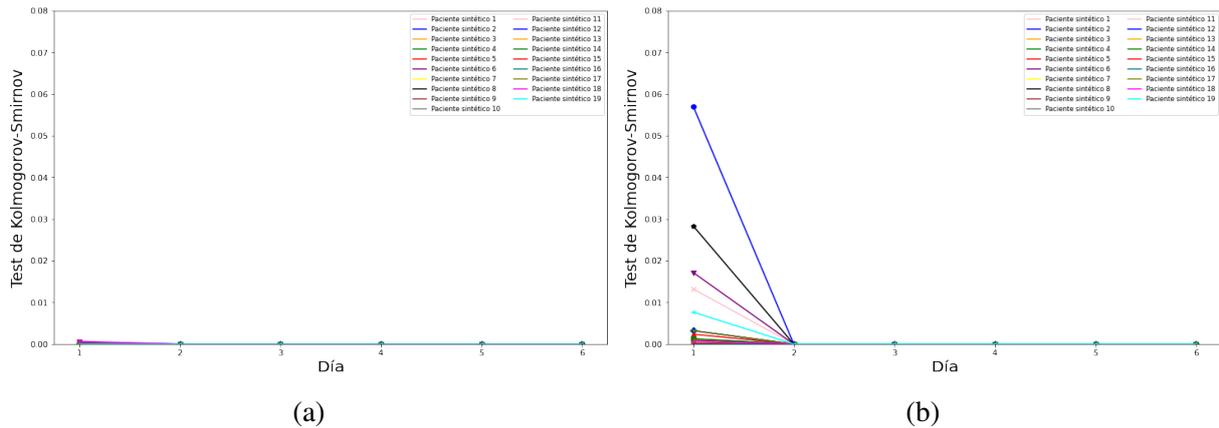


Figura 4.4: p valor de test de *Kolmogorov-Smirnov* entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD1-SEV y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

Por último, se evalúa la calidad de los datos usando la distancia DTW. Esta mide la similitud entre la serie temporal real y sintética a lo largo del tiempo, siendo mínima cuanto más tiempo pasen la serie real y sintética alineadas. Se observa como esta métrica es muy variable según el paciente sobre todo en el caso de las series generadas por DGAN (ver Figura 4.5 (a)). Se puede ver como en los dos casos se mantiene en el mismo rango de valores. El valor de la distancia DTW aumenta a lo largo del tiempo, y esto es lógico por que según se consideran más días de las series temporales, es más probable que difieran una de la otra.

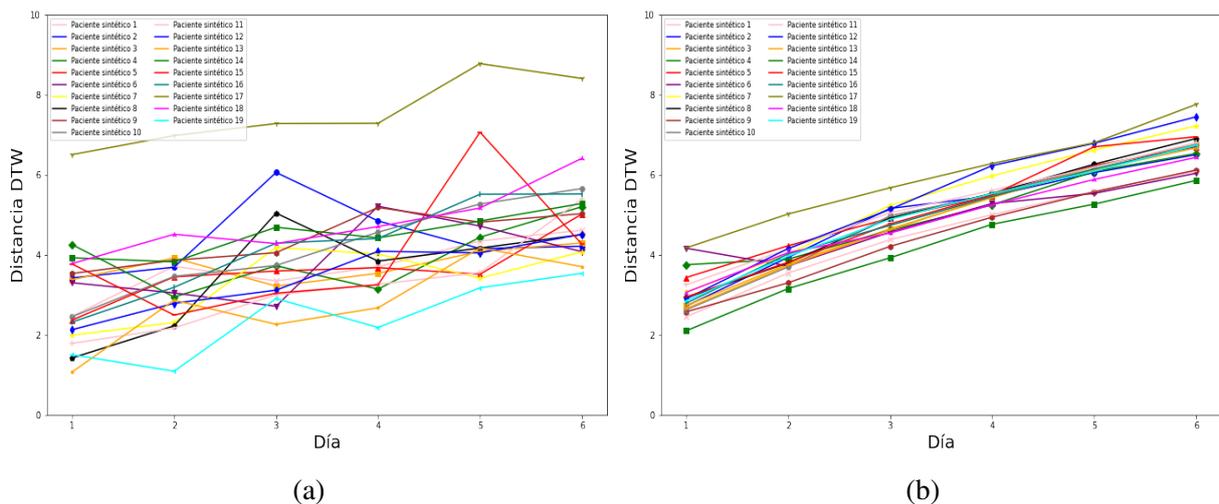


Figura 4.5: Distancia DTW entre series temporales de glucosa reales de la base de datos BD1-SEV y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

4.2.2. Generación de series temporales sintéticas usando BD2-PED

A continuación, se procede a mostrar los resultados de la generación de series temporales sintéticas a partir de la base de datos BD2-PED. Se comparan las series sintéticas generadas por los modelos CPAR y DGAN. En primer lugar, se puede ver en la Figura 4.6 dos histogramas que muestran la distribución de los distintos valores de concentración de glucosa de las series sintéticas respecto a la real para un paciente concreto. En la Figura 4.6 (a) se ve como las distribuciones de las series generadas por el método DGAN tienen una distribución muy similar a la serie real, pero, se ve también como no captura bien los valores máximos (relacionados con el rango de hiperglucemia). Por otro lado, las distribuciones de las series generadas con el modelo CPAR que se muestran en la Figura 4.6 (b) presentan una distribución muy dispar en comparación con la serie real. Comparando los picos más altos de frecuencia en la Figura 4.6 (b) se ve que la media de los datos sintéticos generada por CPAR [100-125] (mg/dL) está desplazada a la derecha con respecto a los valores reales [150-180] (mg/dL) y los datos sintéticos no aprenden bien los valores máximos.

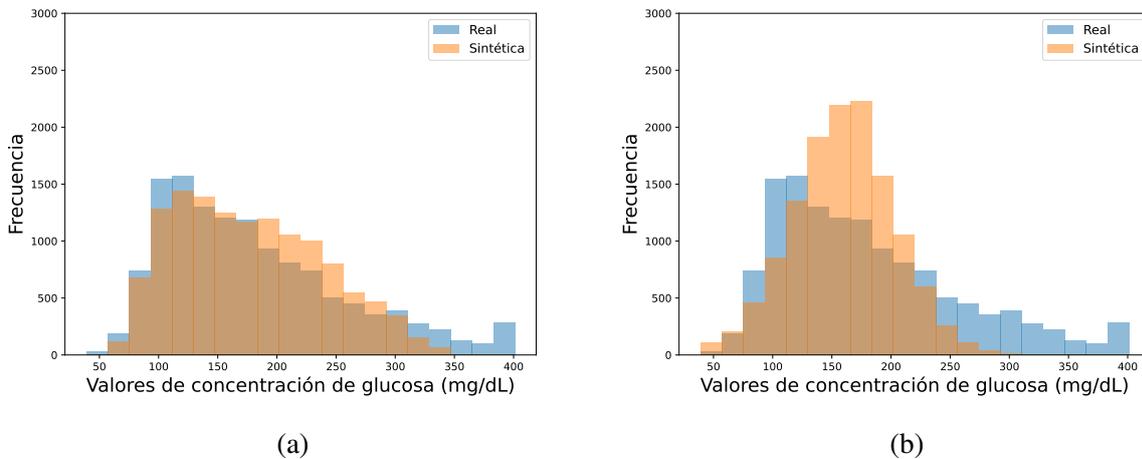


Figura 4.6: Histogramas de la distribución de datos de glucosa de las series temporales de un paciente reales de la BD2-PED y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

Para medir la diferencia entre la distribución de los datos reales y sintéticos la divergencia de *Jensen-Shannon* es calculada. Se puede apreciar como en la Figura 4.7 (a), relativa a la similitud entre PDFs de las series temporales reales y sintéticas generadas por DGAN, el valor de esta métrica es más estable a lo largo del tiempo. En la Figura 4.7, el eje x representa los días que se tienen en cuenta para calcular la divergencia, según se avanza en el eje x se estabiliza el

valor de la divergencia de *Jensen-Shannon* para las distintas PDFs en torno a 0.15. Aun así, el rango de valores es muy similar al que se muestra en la Figura 4.7 (b) para las series generadas con CPAR pero es variable para las distintas PDFs [0.14-0.17].

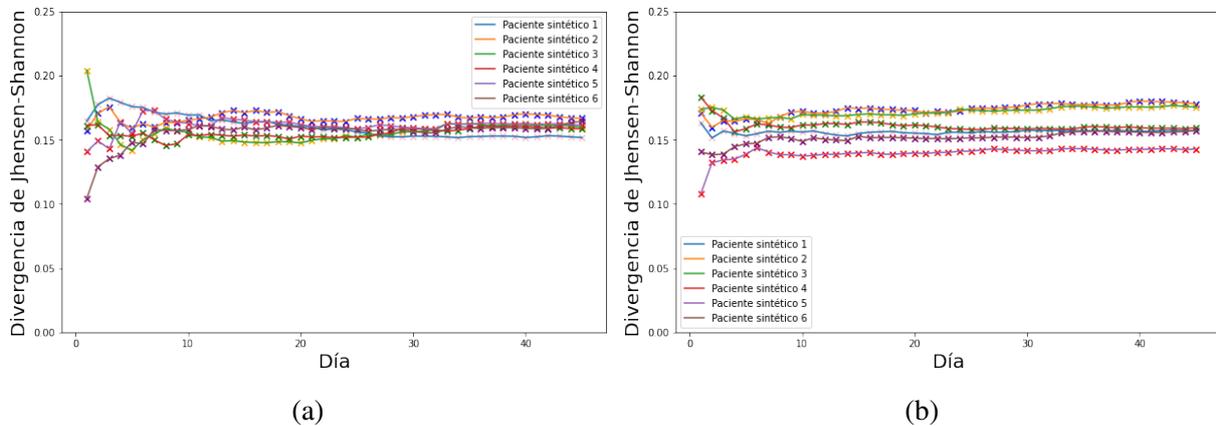


Figura 4.7: Resultados de divergencia de *Jensen-Shannon* entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD2-PED y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

En cuanto al MMD, en este caso se puede apreciar como en la Figura 4.8 (a), correspondiente a las distribuciones de datos de las series temporales generadas mediante DGAN, los valores son cercanos a 0 a lo largo de todas las fechas para la mayoría de los pacientes. Esto indica la gran similitud entre las PDFs asociadas a las series temporales generadas por DGAN y la real. Mientras que, en la Figura 4.8 (b) los valores para esta métrica, pese a ser muy bajos, distan más del 0. Es por ello que en base al MMD, se puede decir, para este caso, que las PDFs de las series temporales generadas con DGAN son mejores que las generadas por CPAR.

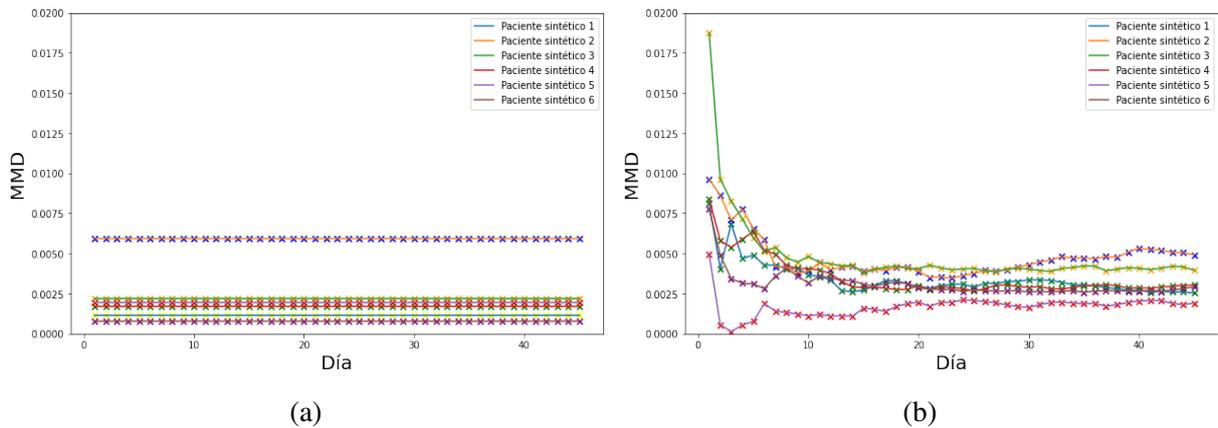


Figura 4.8: Resultados de MMD entre distribución de datos de glucosa de las series temporales reales de la base de datos BD2-PED y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

A continuación, se evalúa si se acepta la H_0 de este TFG mediante el estadístico p_{valor} del test de *Kolmogorov-Smirnov*. El test de *Kolmogorov-Smirnov* está basado en la similitud de las distribuciones, partiendo de la H_0 de este TFG: las series temporales sintéticas son similares a las reales. Se puede ver tanto en la Figura 4.9 (a) como en la Figura 4.9 (b) que la serie sintética 6 tiene instantes en los que no sería posible afirmar que las PDFs son muy similares, ya que el resultado del test p_{valor} es superior a 0.05. En concreto, en la Figura 4.9 (a) se ve como hay un día para este paciente para el cual esta afirmación es falsa, mientras que, en la Figura 4.9 (b) ocurre en dos días diferentes (segundo y cuarto día).

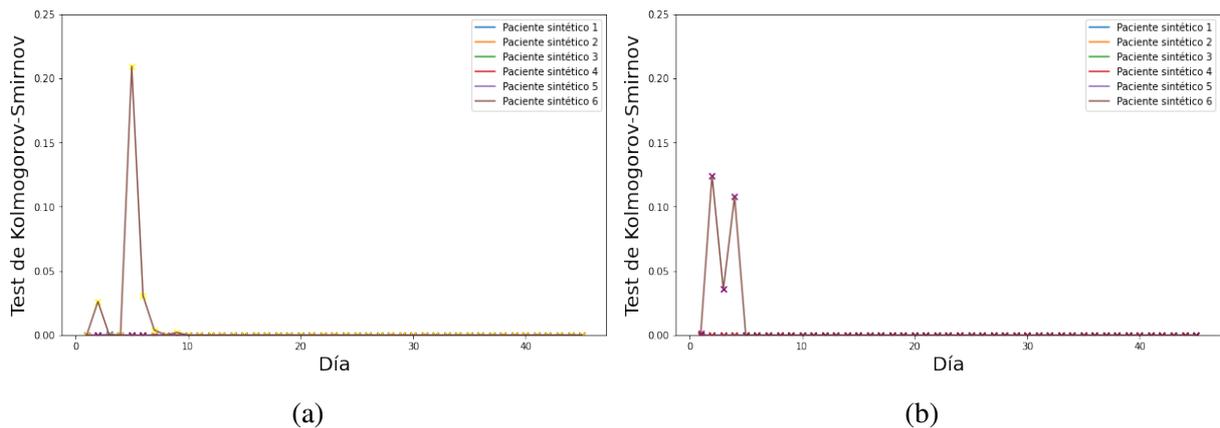


Figura 4.9: p_{valor} de test de *Kolmogorov-Smirnov* entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD2-PED y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

Finalmente, se muestra la distancia DTW, en este caso se ve claramente en la Figura 4.10 (a)

como la distancia entre las series reales y sintéticas generadas por DGAN es mucho menor que para las generadas por CPAR, mostradas en la Figura 4.10 (b). Se puede ver en la Figura 4.10, como el valor de la distancia DTW es equivalente entre DGAN y CPAR durante los cuatro primeros días, a partir de entonces, este valor crece de manera exponencial, siendo mayor el crecimiento para CPAR.

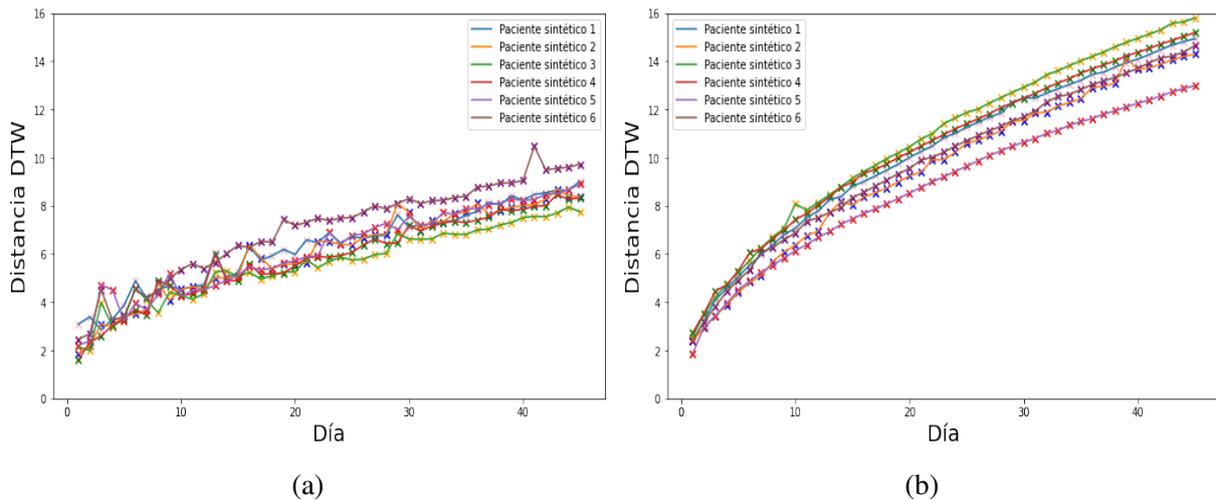


Figura 4.10: Distancia DTW entre las series temporales de glucosa reales de la base de datos BD2-PED y las series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

4.2.3. Generación de series temporales sintéticas usando BD3-SEN

Por último, se analizan los resultados obtenidos de la generación de series temporales sintéticas usando la base de datos BD3-SEN. Para empezar, en la Figura 4.11 (a) se muestra el histograma de la serie temporal sintética generada mediante DGAN superpuesto sobre el histograma de la serie real para un paciente concreto ($PtId=166$). Se puede observar que los histogramas son bastante similares en cuanto a distribución, pero los valores de glucosa por debajo de 100 no son representativos en la serie temporal sintética (véase histograma naranja en Figura 4.11 (a)). A grandes rasgos, se puede deducir que los valores mínimos de concentración de glucosa en sangre no se ‘aprenden’ correctamente. Por otro lado la Figura 4.11 (b) muestra el mismo análisis, pero con los datos sintéticos generados por el modelo CPAR. En este caso los histogramas real y sintético son bastante parecidos aunque CPAR no aprende valores mayores a 275 mg/dL de glucosa.

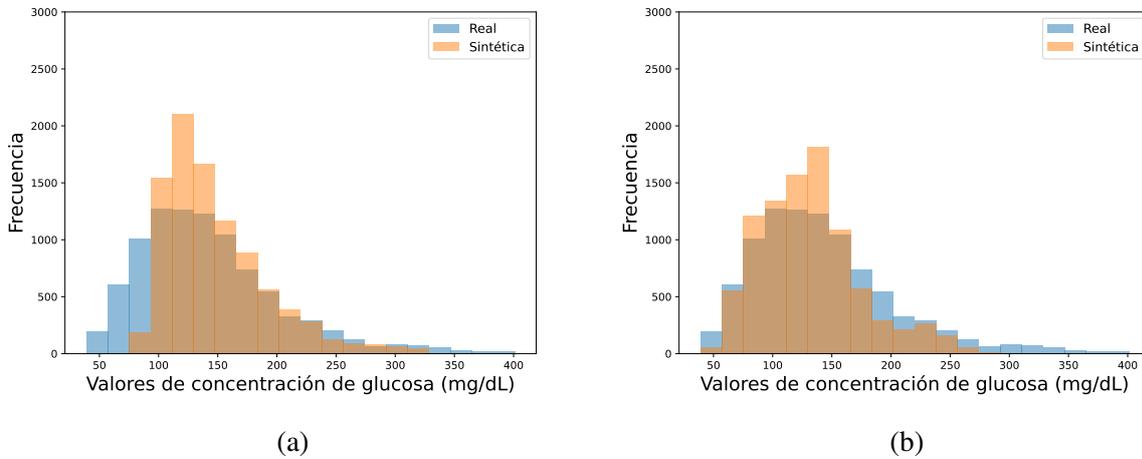


Figura 4.11: Histogramas de la distribución de datos de glucosa de las series temporales de un paciente reales de la BD3-SEN y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

Analizando los valores de divergencia de *Jensen-Shannon* la Figura 4.12 (a) se puede apreciar como en el caso de las series generadas por DGAN la diferencia medida entre las PDFs de las series real y las sintéticas son muy similar entre todas las series asociadas a los distintos pacientes y a lo largo del tiempo, con valores de divergencia en torno a 0.2. Mientras que en el caso de las series generadas con CPAR (ver Figura 4.12 (b)), los valores de divergencia son más variables, aunque se mantienen en los mismos rangos que los generados por DGAN.

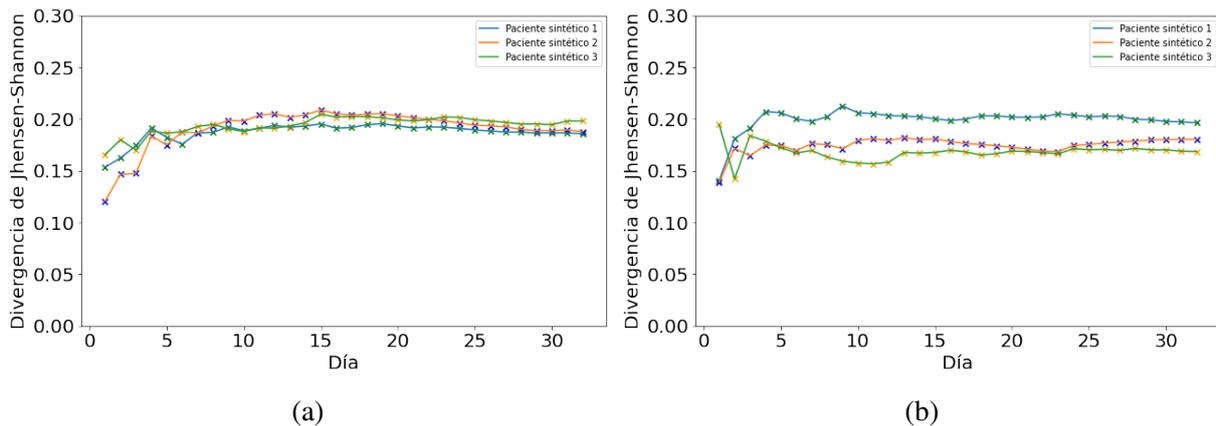


Figura 4.12: Resultados de divergencia de *Jensen-Shannon* entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD3-SEN y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

En cuanto al MMD, se minimiza cuanto más parecidas sean distribuciones reales y sintéti-

cas. En este caso, como se puede ver en la Figura 4.13 (a), asociada a las PDFs de las series sintéticas generadas con DGAN, los valores son muy cercanos a 0, aunque varía ligeramente entre los 3 pacientes. Esto quiere decir, que las distribuciones de los datos de glucosa reales y sintéticos son similares. Por otro lado en la Figura 4.13 (b) se puede ver como el valor de MMD para las series sintéticas generadas por CPAR es muy similar a las de DGAN, pero además, el paciente 1 tiene una distribución más similar al real.

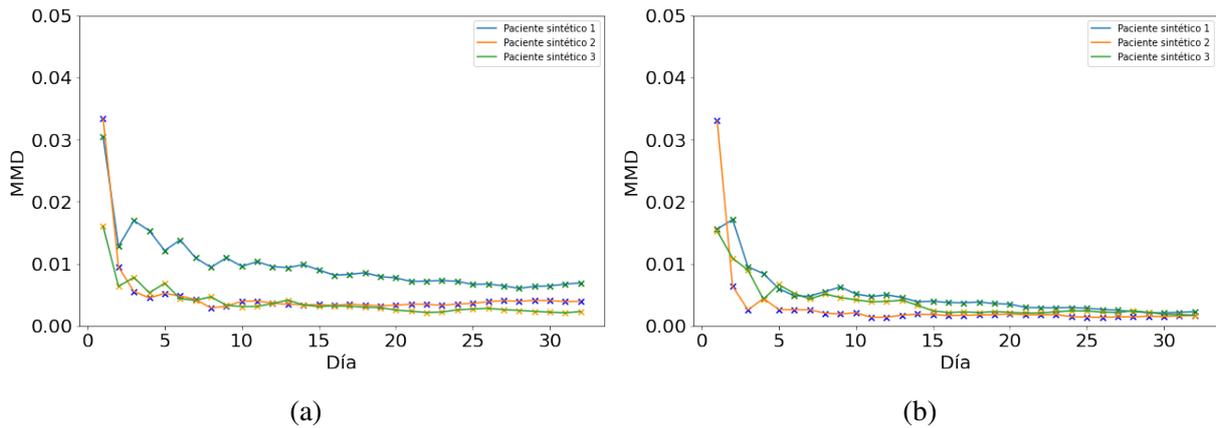


Figura 4.13: Resultados de MMD entre distribución de datos de glucosa de las series temporales reales de la base de datos BD3-SEN y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

Después de calcular el test de *Kolmogorov-Smirnov* entre las distribuciones de las series temporales reales y sintéticas, se obtiene el el p_{valor} para determinar si es posible aceptar la H_0 . Es decir, determinar si las series temporales sintéticas son similares a las reales. Se puede apreciar que en todas las distribuciones, tanto las de las series temporales generadas por DGAN (Figura 4.14 (a)) como en las generadas por CPAR (Figura 4.14 (b)), el p_{valor} es menor a 0.05 y por ello se puede afirmar que todas las series temporales sintéticas son muy similares a las reales.

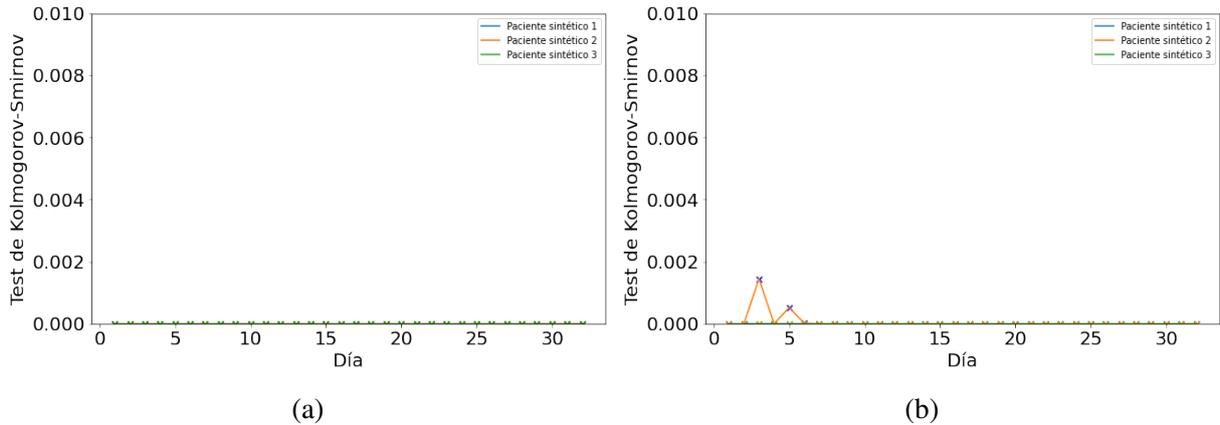


Figura 4.14: p_{valor} de test de *Kolmogorov-Smirnov* entre la distribución de datos de glucosa de las series temporales reales de la base de datos BD3-SEN y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

Por último, se muestra la distancia de DTW. Se puede ver que es más estable para los datos generados con DGAN (véase la Figura 4.15 (a)), que para los datos generados por CPAR (véase la Figura 4.15 (b)) que presentan mayor variabilidad entre las distancias DTW asociadas a las distintas series temporales. Pese a que crece según se avanza en el eje x, en los dos casos el valor máximo alcanzado es 7.

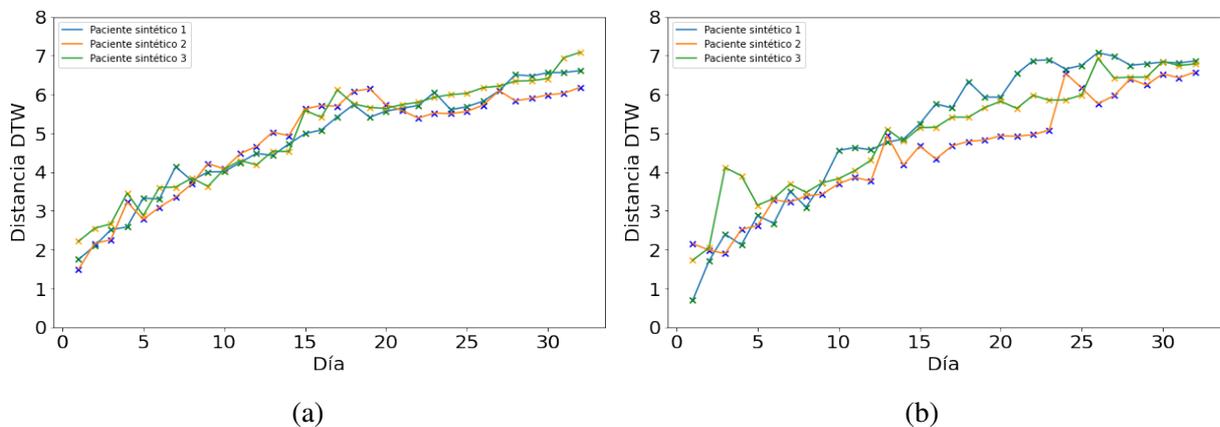


Figura 4.15: Distancia DTW entre las series temporales de glucosa reales de la base de datos BD3-SEN y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

4.3. Predicción de niveles de glucosa

En esta sección se muestra los resultados la predicción de valores de glucosa usando series temporales y el modelo N-HITS. Esto se realiza considerando distintas series temporales, tanto reales como sintéticas generadas por DGAN y CPAR. Se busca evaluar si las capacidades predictivas de las series temporales sintéticas son iguales que las de las series temporales reales. Para esto, se entrena diferentes modelos con N-Hits y series temporales de las bases de datos (BD1-SEV, BD2-PED, BD3-SEN) y se evalúa la capacidad de predecir valores de glucosa en diferentes ventanas (horas). Para medir la capacidad predictiva se consideran las métricas explicadas en el Capítulo 3.3.1.

4.3.1. Predicción de niveles de glucosa usando BD1-SEV

Para la BD1-SEV se realiza una predicción en una ventana de una hora muestreada cada 5 minutos, 12 muestras. En concreto, las últimas 12 muestras de cada serie temporal son utilizadas como conjunto de evaluación (en inglés, *test*) y las 1716 muestras restantes (6 días) son utilizadas como conjunto de entrenamiento (en inglés, *train*). En este caso, se muestran los resultados para el paciente con PtID 9. En primer lugar se analiza la Tabla 4.1, observando la última fila que hace referencia al valor medio entre todos las series temporales asociadas a distintos pacientes para cada error estudiado, se ve como el mínimo MRAE obtenido es con los datos reales, mientras que el método DGAN minimiza el MSE y MAE. Esto puede indicar que el método DGAN tenga buenas capacidades predictivas.

En primer lugar, se evalúa la representación temporal durante un día de las predicciones. Inicialmente, entrenamos el modelo N-HITS con series temporales reales (ver Figura 4.16). En la Figura 4.16 (c) se puede apreciar que el error CEG, en la que todos los valores se ajustan a la línea oblicua y están en la zona A. Esto se comprueba en la Figura 4.16 (d), ya que las diferencias entre valores de glucosa de las estimaciones son bajas y siguen las recomendaciones de la norma ISO.

A continuación, se evalúan las predicciones realizadas con N-HITS sobre las series sintéticas generadas por DGAN. En primer lugar, se evalúa la representación temporal durante un día de las predicciones, en la Figura 4.17 (a) se puede apreciar la evolución temporal de la serie. Aumentando en dicha Figura, se obtiene la Figura 4.17 (b), donde se compara el valor *test* con el valor estimado. En la Figura 4.17 (c) se puede apreciar el error CEG y como el 100% de los valores están en la zona A. Esto se comprueba en la Figura 4.17 (d), ya que todos los valres

PtID	Real			DGAN			CPAR		
	MSE	MAE	MRAE	MSE	MAE	MRAE	MSE	MAE	MRAE
102	0.0548	0.2135	0.3904	0.0005	0.0176	1.7642	0.0195	0.1019	1.1520
125	0.0019	0.0345	0.0681	0.0008	0.0217	2.1752	0.0161	0.1101	0.3245
137	0.0057	0.0203	0.1274	0.0002	0.0111	1.1076	0.0445	0.1837	1.0059
146	0.0044	0.0630	0.2622	0.0007	0.0216	2.1619	0.0260	0.1183	0.3342
163	0.0041	0.0568	0.4252	0.0009	0.0265	2.6519	0.0245	0.1300	0.3025
20	0.0149	0.1134	1.5913	0.0053	0.0698	6.9845	0.0222	0.1066	0.2499
202	0.0002	0.0137	0.0172	0.0008	0.0273	2.7320	0.0110	0.0802	0.1888
21	0.0003	0.0160	0.0678	0.0021	0.0440	4.4039	0.0272	0.1159	0.6407
3	0.0012	0.0307	0.0789	0.0068	0.0709	7.0995	0.0226	0.1427	0.4582
64	0.0042	0.0578	0.1580	0.0047	0.0594	5.9450	0.0457	0.1719	0.9535
72	0.0077	0.0660	0.4387	0.0022	0.0409	4.0859	0.0521	0.1908	4.3423
85	0.0118	0.0994	0.4397	0.0015	0.0343	3.4332	0.0212	0.1251	0.2727
88	0.0019	0.0374	0.0774	0.0025	0.0455	4.5475	0.0227	0.1050	0.2573
9	0.0002	0.0122	0.0204	0.0015	0.0301	3.0098	0.0448	0.1900	0.4454
Media del error	0.0081	0.0596	0.2973	0.0022	0.0372	3.7216	0.0286	0.1337	0.7806

Tabla 4.1: Figuras de mérito obtenidas para los distintos pacientes de la BD1-SEV reales, generados con DGAN y generados con CPAR tras realizar la predicción con el modelo N-HiTS. Se miden: MAE, MSE y MRAE. Se muestra también la media de todos los pacientes para cada error.

están en la zona recomendada por la norma ISO.

Por último, se evalúa la representación temporal durante un día de las predicciones utilizando series temporales generadas con CPAR, mediante el entrenamiento de N-HiTS. Se visualiza, en primer lugar, la evolución de la serie temporal (ver Figura 4.18 (a) haciendo hincapié en la comparación del conjunto *test* y la predicción (ver Figura 4.18) (b). A continuación, en la Figura 4.18 (c) se puede apreciar el CEG, donde se ven varios valores fuera de la zona A. Esto se comprueba en la Figura 4.18 (d), ya que hay escasos valores también en la zona recomendada por la norma ISO.

A continuación, se realiza una comparación entre los resultados obtenidos por la predicción realizada con N-HiTS en la BD1-SEV para los tres tipos distintos de datos: reales, sintéticos generados con DGAN y sintéticos generados con CPAR. Cabe destacar la gran similitud entre la serie temporal real, mostrada en la Figura 4.16 (a), y la serie sintética generada por DGAN, mostrada en la Figura 4.17 (a). Por otro lado, la serie temporal generada por CPAR, Figura 4.18 (a) muestra mucha fluctuación a lo largo del tiempo en los niveles de glucosa. Para ver en mayor detalle la divergencia entre los datos predichos y el conjunto *test* es conveniente

consultar la Figura 4.16 (b) para la predicción en los datos reales, viendo una gran similitud entre ambos conjuntos. Para el caso de los datos generados con DGAN, se aprecia en la Figura 4.17 (b) que también son muy similares. Esto no ocurre con los datos generados con CPAR, en la Figura 4.18 (b) se ve que la predicción difiere más del *test* que en los casos anteriores.

Por último, para tener una idea final de cual de los métodos ha realizado una predicción más certera, se consultan el CEG y la norma ISO. En la Figura 4.16 (c) se ve el CEG de la predicción con datos reales, se observa como las predicciones se ajustan perfectamente a la línea encontrándose todas en la zona A del CEG. Lo mismo ocurre en la Figura 4.17 (c), todos los valores de la predicción en la serie temporal sintética generada con DGAN están en la zona A. Mientras que, para el caso de CPAR, se observa en la Figura 4.18 (c) como un muy bajo porcentaje de las predicciones se encuentra en la zona óptima. Para corroborarlo, se muestra si las observaciones se encuentran en la zona recomendada por la norma ISO. Tanto en el caso de la predicción real, mostrada en la Figura 4.16 (d), como en el caso de la predicción de la serie creada con DGAN, mostrada en la Figura 4.17 (d), más del 90 % de las observaciones están en la zona recomendada. Comparado con la generada por CPAR, en la Figura 4.18 (d) se ve que solo el 17 % de las observaciones está en esa zona.

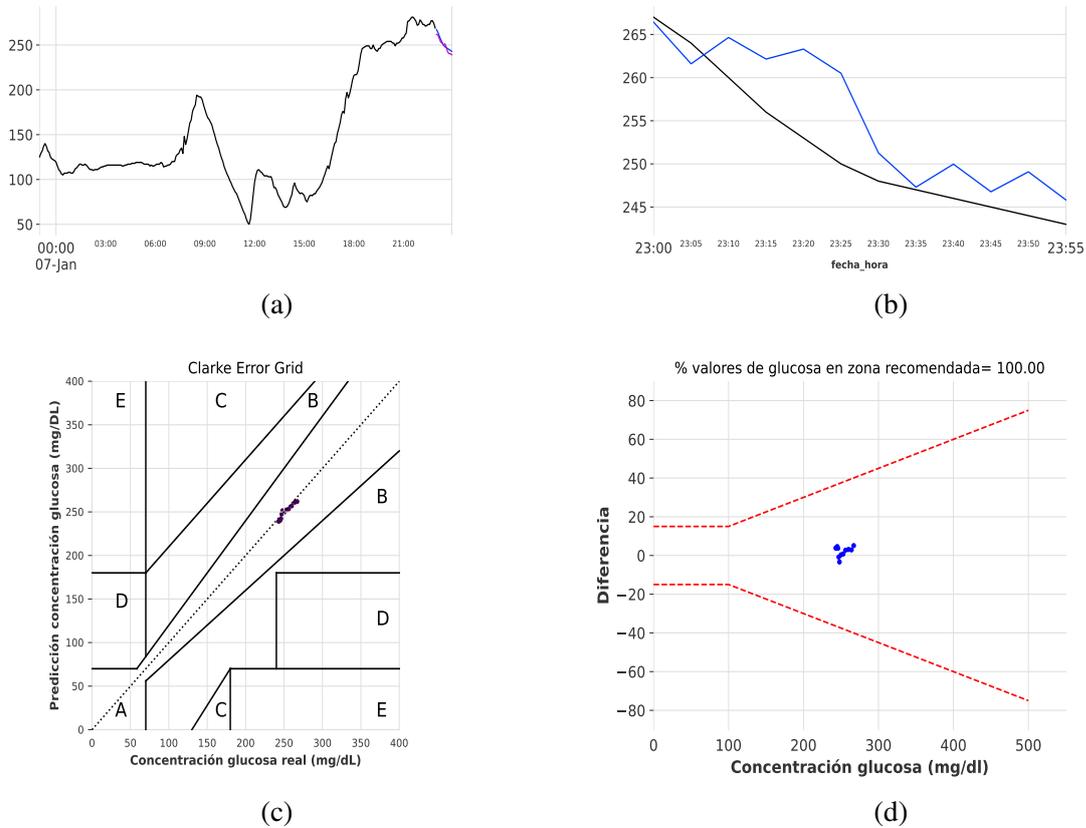


Figura 4.16: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD1-SEV real para el decimonoveno paciente con PtId 9 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 1 día), *test* que se utilizará como referencia visual de la exactitud de la predicción y predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

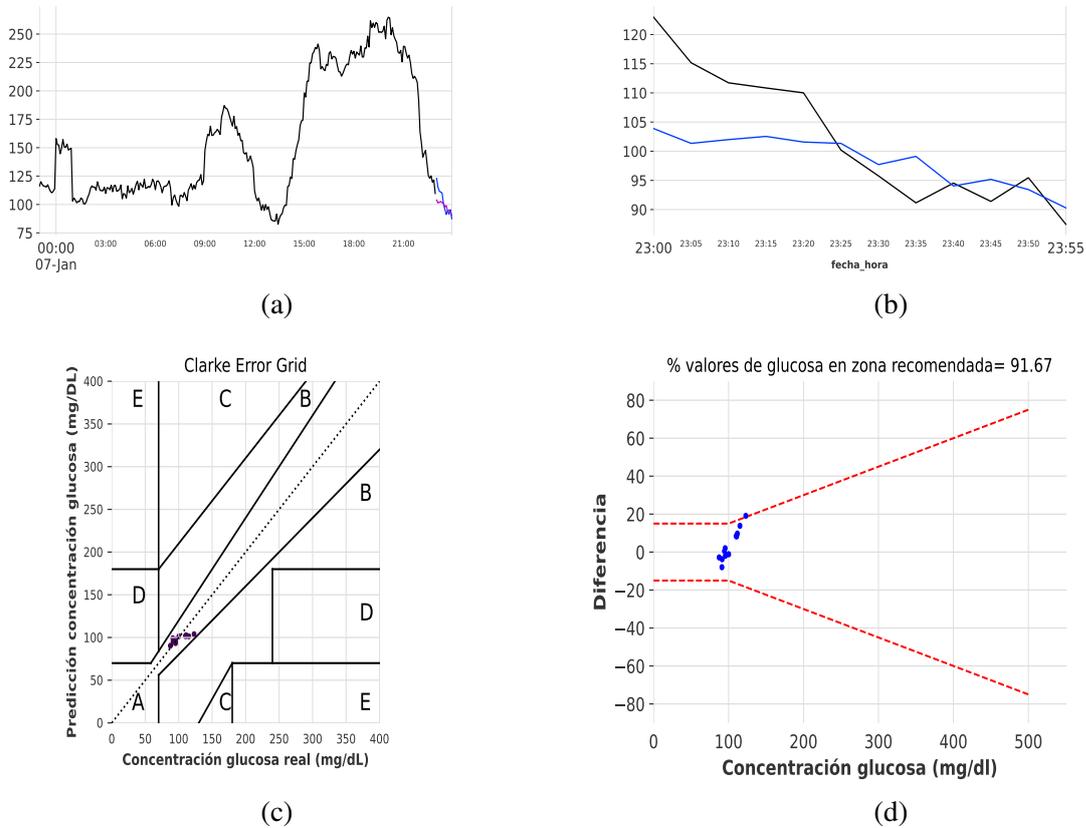


Figura 4.17: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD1-SEV sintética generada por el método DAGN para el decimonoveno paciente con PtId 9 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 1 día), *test* que se utilizará como referencia visual de la exactitud de la predicción y predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

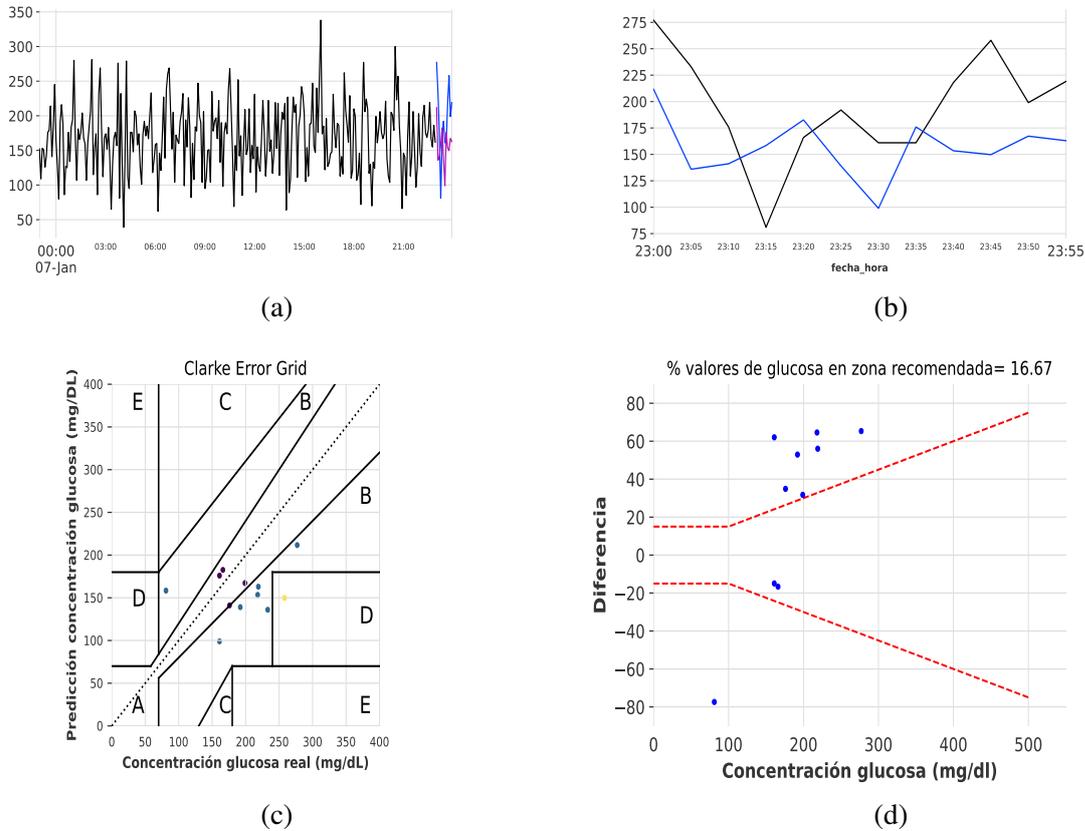


Figura 4.18: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD1-SEV sintética generada por el método CPAR para el PtId=9 con una ventana de predicción de 12 muestras. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 1 día), *test* que se utilizará como referencia visual de la exactitud de la predicción y predicción; (b) Serie temporal de glucosa con los conjuntos *test* y predicción; (c) CEG; (d) ISO.

4.3.2. Predicción de niveles de glucosa usando BD2-PED

Para la BD2-PED se realiza una predicción en una ventana de tres horas muestreada cada 5 minutos, 36 muestras. En concreto, las últimas 36 muestras de cada serie temporal son utilizadas como *test* y las muestras restantes son utilizadas como *train*. En este caso, se muestran los resultados para el paciente con PtID=58. En primer lugar, se muestran en la Tabla 4.2 los errores MSE, MAE y MRAE para cada predicción de cada serie temporal comparando la serie real, la sintética generada por DGAN y la sintética generada por CPAR. Se puede ver como la media del error para todos los errores es mínima para la base de datos real, seguida de los datos sintéticos generados por CPAR y por último los generados por DGAN, aunque los valores son

muy similares.

PtID	Real			DGAN			CPAR		
	MSE	MAE	MRAE	MSE	MAE	MRAE	MSE	MAE	MRAE
48	0.0113	0.0914	0.2814	0.1044	0.3210	1.7913	0.0261	0.1357	1.3393
53	0.0213	0.1245	0.4921	0.0433	0.1804	0.2650	0.0192	0.1134	0.6511
58	0.0285	0.1346	0.3180	0.0058	0.0572	0.1084	0.0291	0.1336	0.4734
61	0.0181	0.1229	0.4137	0.0559	0.2016	0.3270	0.0187	0.1133	0.3580
86	0.0036	0.0462	0.1594	0.0511	0.1851	0.3670	0.0155	0.1009	0.3175
93	0.0322	0.1609	0.2114	0.0955	0.2951	1.0192	0.0216	0.1087	0.4672
Media del error	0.0192	0.1134	0.3127	0.0593	0.2067	0.6462	0.0217	0.1176	0.6011

Tabla 4.2: Figuras de mérito obtenidas para los distintos pacientes de la BD2-PED reales, generados con DGAN y generados con CPAR tras realizar la predicción con el modelo N-HiTS. Se miden: MAE, MSE y MRAE. Se muestra también la media de todos los pacientes para cada error.

En primer lugar, se evalúa la representación temporal durante diez días de los datos *train*, *test* y las predicciones realizadas con N-HiTS sobre las series temporales reales de BD2-PED (véase Figura 4.19 (a)). En la Figura 4.19 (b) se muestra de la Figura anterior únicamente el fragmento en el que se encuentran las predicciones y *test*. En la tercera Figura 4.19 (c) se puede apreciar el error CEG, el cual sirve para valorar la dispersión de los valores predichos. Esto se complementa con la Figura 4.19 (d), en la que se ve cómo se ajustan las predicciones realizadas a la norma ISO.

A continuación, se evalúan las predicciones realizadas con N-HiTS sobre las series sintéticas generadas por DGAN, empezando por visualizar la evolución temporal durante 10 días, en la Figura 4.19 (a). Para mayor detalle sobre la similitud temporal de los conjuntos *test* y predicción, se observa la Figura 4.19 (b). En la Figura 4.19 (c) se puede apreciar el número de predicciones que se encuentra en cada zona descrita por el error CEG y para complementarlo, se muestra el número de valores predichos que cumplen la norma ISO en la Figura 4.19 (d).

Por último, se evalúan series temporales generadas con CPAR, mediante el entrenamiento de N-HiTS. Se visualiza, la evolución de la serie temporal (ver Figura 4.21 (a) haciendo hincapié en la comparación del conjunto *test* y la predicción (ver Figura 4.21) (b). A continuación, en la Figura 4.21 (c) se puede apreciar el CEG, donde se las predicciones en función de las distintas zonas de riesgo. Esto se comprueba en la Figura 4.21 (d), viendo los valores de predicciones que se encuentran en la zona recomendada por la norma ISO.

A continuación, se comparan las 4 figuras descritas para cada tipo de datos: real, sintéti-

co generado por DGAN y sintético generado por CPAR, para determinar qué método presenta mejores capacidades predictivas. Se empieza por visualizar los errores representando la evolución de la serie temporal, para ello, se superponen las predicciones sobre el conjunto *test*. Comparando la Figura 4.19 (a), Figura 4.20 (a), Figura 4.21 (a) se puede ver, respectivamente, la evolución de las series temporales real, generada por DGAN y generada por CPAR. Además en la Figura 4.19 (b), Figura 4.20 (b), Figura 4.21 (b) se muestra una ampliación que capta solo el conjunto de *test* y la predicción. Cabe destacar lo bien que se ajusta a los valores *test* la predicción realizada con la serie generada por DGAN, como se ve en la Figura 4.20 (b). Además, se puede ver en la Figura 4.21 (a) como los registros temporales generados por CPAR no son correctos, muestran mucha más variabilidad de la que se debería.

En cuanto al CEG plot, que se puede ver para la serie temporal generada por DGAN, Figura 4.20 (c), el 100% de las predicciones están en la zona A mientras que para la serie real, Figura 4.19 (c), y sintética generada por CPAR, Figura 4.21 (c), se puede ver como las predicciones son más dispersas y algunas se encuentran en la zona B. Lo mismo se aplica a las representaciones de la norma ISO, con DGAN se ve en la Figura 4.20 (d) como el 80% de los valores están en la zona recomendada, mientras que para la serie real, Figura 4.19 (c), y sintética generada por CPAR, Figura 4.21 (c), el porcentaje de valores en la zona aceptada por la norma ISO es entorno a un 40% menor. Esto puede indicar que las prestaciones predictivas de los datos generados por el modelo DGAN son incluso mejores que las de los datos reales.

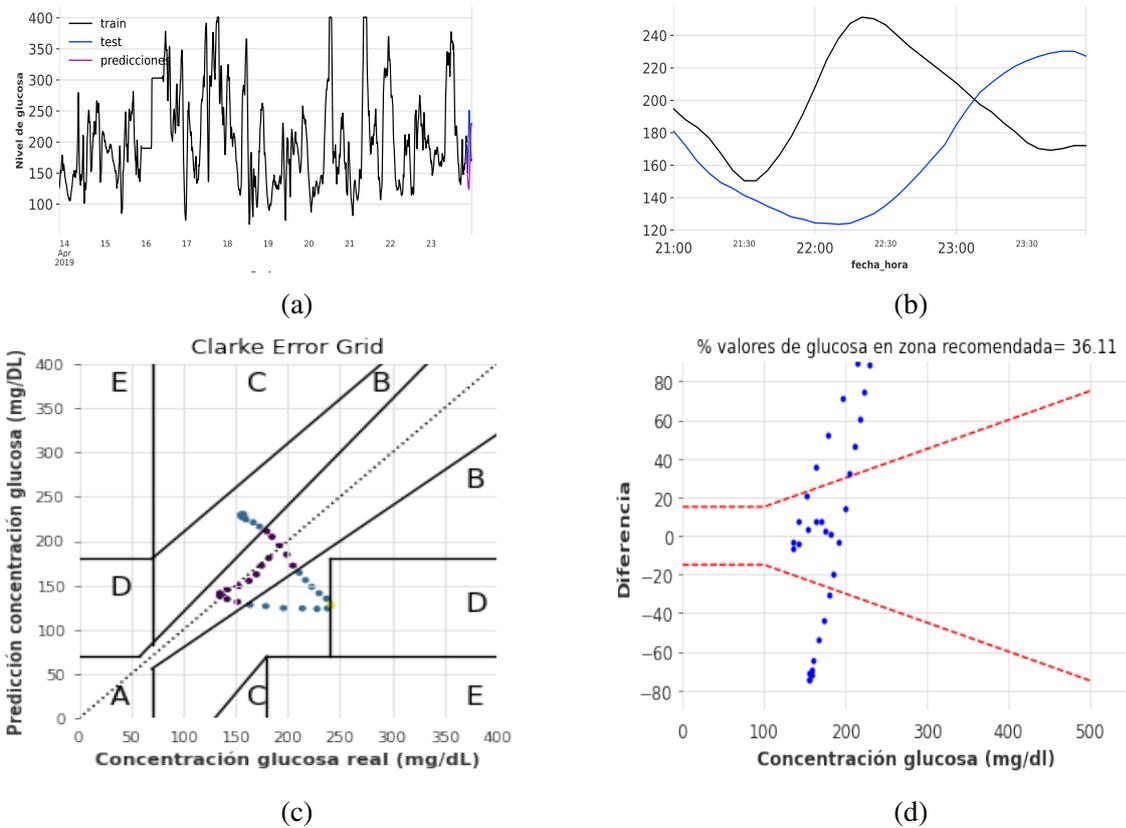


Figura 4.19: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD2-PED real para el tercer paciente con PtId 58 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción (c) CEG, (d) ISO.

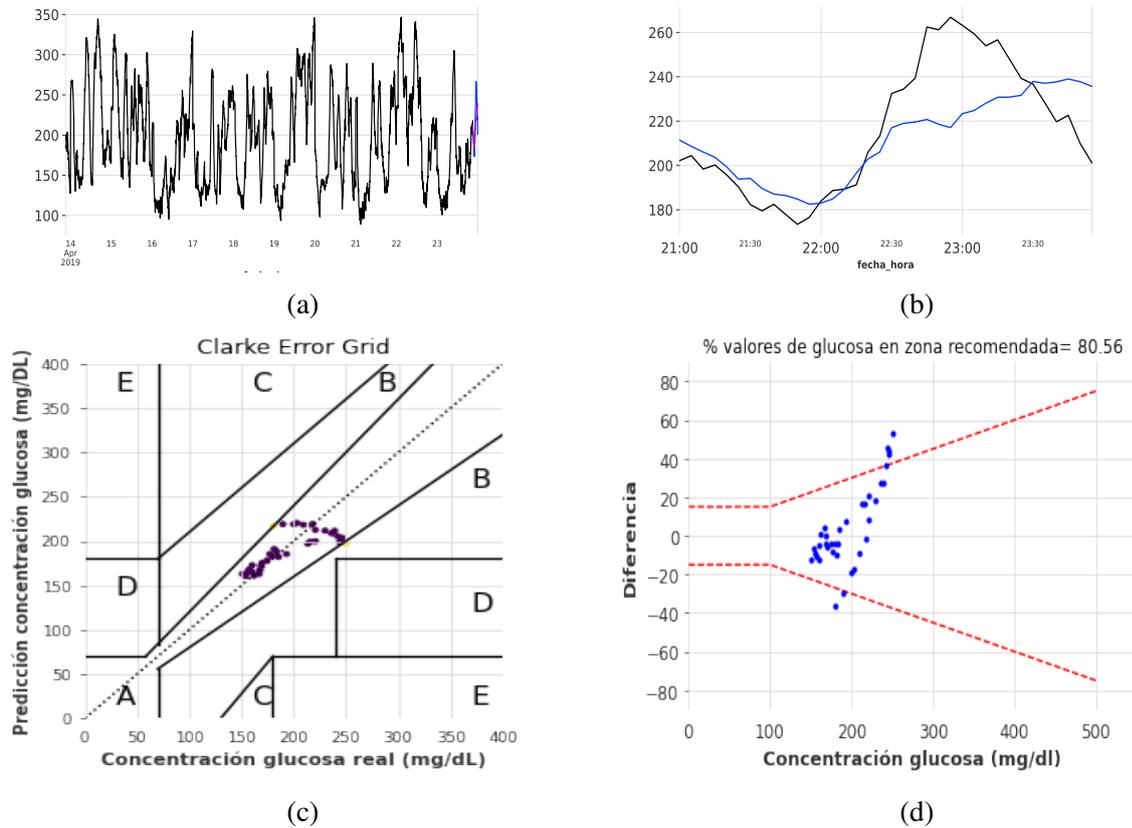


Figura 4.20: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD2-PED sintética generada por el método DAGAN para el tercer paciente con PtId 58 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

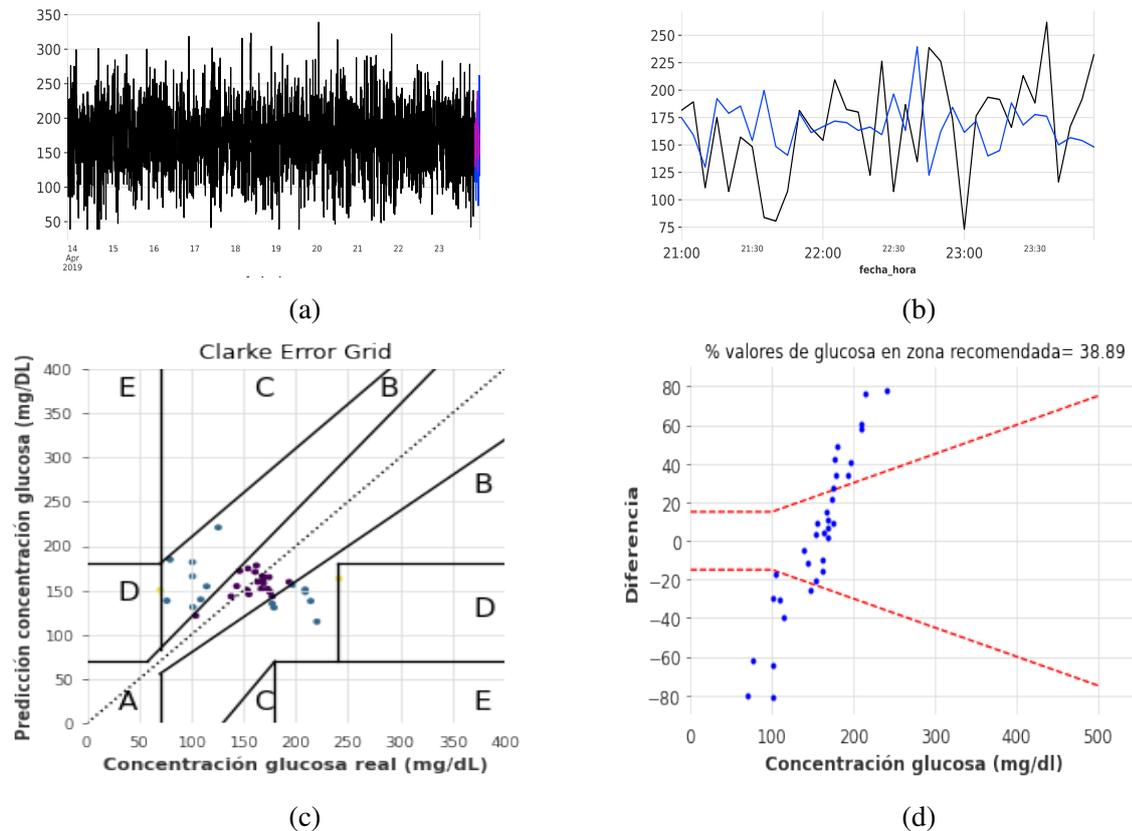


Figura 4.21: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD2-PED sintética generada por el método CPAR para el tercer paciente con PtId 58 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

4.3.3. Predicción de niveles de glucosa usando BD3-SEN

Para la BD3-SEN se realiza una predicción en una ventana de tres horas muestreada cada 5 minutos, 36 muestras. En concreto, las últimas 36 muestras de cada serie temporal son utilizadas como *test* y las muestras restantes son utilizadas como *train*. En este caso, se muestran los resultados para el paciente con PtID 58. En primer lugar, se muestran en la Tabla 4.3 los errores MSE, MAE y MRAE para cada predicción de cada serie temporal comparando la serie real, la sintética generada por DGAN y la sintética generada por CPAR. Se puede ver como para el MAE y MRAE la media de error de los pacientes sintéticos generados por DGAN es incluso menor que la de los reales, mientras que la media de MSE entre los distintos pacientes es menor

para los reales. Se puede ver como el error máximo es siempre para las series generadas por CPAR.

PtID	Real			DGAN			CPAR		
	MSE	MAE	MRAE	MSE	MAE	MRAE	MSE	MAE	MRAE
151	0.0061	0.0593	0.3732	0.0046	0.0617	0.1765	0.0071	0.0618	0.3975
166	0.0014	0.0332	0.2463	0.0009	0.0240	0.1211	0.0060	0.0656	0.4235
57	0.0374	0.1700	0.4437	0.0137	0.1094	0.6359	0.0341	0.1598	0.4181
Media del error	0.0150	0.0875	0.3544	0.0064	0.0650	0.3112	0.0157	0.0957	0.4124

Tabla 4.3: Errores obtenidos para los distintos pacientes de la BD3-SEN reales, generados con DGAN y generados con CPAR tras realizar la predicción con el modelo N-HiTS. Se miden: MAE, MSE y MRAE. Se muestra también la media de todos los pacientes para cada error.

En primer lugar, se evalúa la representación temporal durante diez días de los datos *train*, *test* y las predicciones realizadas con N-HiTS sobre las series temporales reales de BD3-SEN (ver Figura 4.22 (a)). En la Figura 4.22 (b) se hace hincapié en la comparación entre la evolución temporal del conjunto *test* y la predicción. En la Figura 4.22 (c) se puede apreciar el error CEG y determinar a que zona pertenecen cada una de las predicciones. Esto se complementa con la Figura 4.22 (d), en la que se ve cómo se ajustan las predicciones realizadas a la norma ISO.

A continuación, se evalúan las predicciones realizadas con N-HiTS sobre las series sintéticas generadas por DGAN. Primero, se visualiza la evolución temporal durante 10 días, en la Figura 4.23 (a). Se contempla la Figura 4.23 (b) para mayor detalle sobre la comparación temporal entre ambos conjuntos *test* y predicción. En la Figura 4.23 (c) se puede apreciar el número de predicciones en cada zona descrita por el error CEG y se completa con, la Figura 4.23 (d), que muestra el número de valores predichos que cumplen la norma ISO.

Por último, se evalúan series temporales generadas con CPAR, mediante el entrenamiento de N-HiTS. Se visualiza, la evolución de la serie temporal (ver Figura 4.24 (a) haciendo hincapié en la comparación del conjunto *test* y la predicción (ver Figura 4.24) (b). A continuación, en la Figura 4.24 (c) se puede apreciar el CEG, donde se las predicciones en función de las distintas zonas de riesgo. Esto se comprueba en la Figura 4.24 (d), viendo los valores de predicciones que se encuentran en la zona recomendada por la norma ISO.

A continuación, con el fin de comparar los pacientes reales y sintéticos (tanto los generados con DGAN como los generados con CPAR), se estudia la evolución de la serie temporal mostrando la predicción superpuesta sobre el conjunto *test*. Se compara la Figura 4.22 (a) que se corresponde con la evolución de la serie temporal real, con la Figura 4.23 (a) correspondiente a la serie sintética generada con DGAN, y la Figura 4.24 (a) correspondiente con la serie temporal

generada con CPAR. Se puede apreciar en más detalle la superposición de los conjuntos predicción y *test* en la Figura 4.22 (b), Figura 4.23 (b), Figura 4.24 (b) de este modo, se visualiza como la serie temporal real capta algo mejor la tendencia temporal de los datos de glucosa.

Con respecto al CEG, que se puede ver para la serie temporal generada por DGAN, Figura 4.23 (c), el 100% de las predicciones están en la zona A mientras que para la serie real y se ajustan muy bien a la línea oblicua. Esto genera expectativas de que la predicción sea buena. Por otro lado, en el CEG para la predicción de datos reales mostrada en la Figura 4.22 (c), algunos valores invaden la zona B. Por último, en el caso de la serie sintética generada por CPAR, Figura 4.24 (c), se puede ver como la mitad de predicciones están en la zona A y la otra mitad en la zona B. Lo mismo se aplica a las representaciones de la norma ISO, con DGAN se ve en la Figura 4.23 (d) como el 100% de los valores están en la zona recomendada, mientras que para la serie real, Figura 4.22 (c), se encuentran en la zona recomendada el 60% de los valores. Esto decrece aún más para la sintética generada por CPAR, Figura 4.24 (c), en la que el porcentaje de valores en la zona aceptada por la norma ISO es de entrono al 42%.

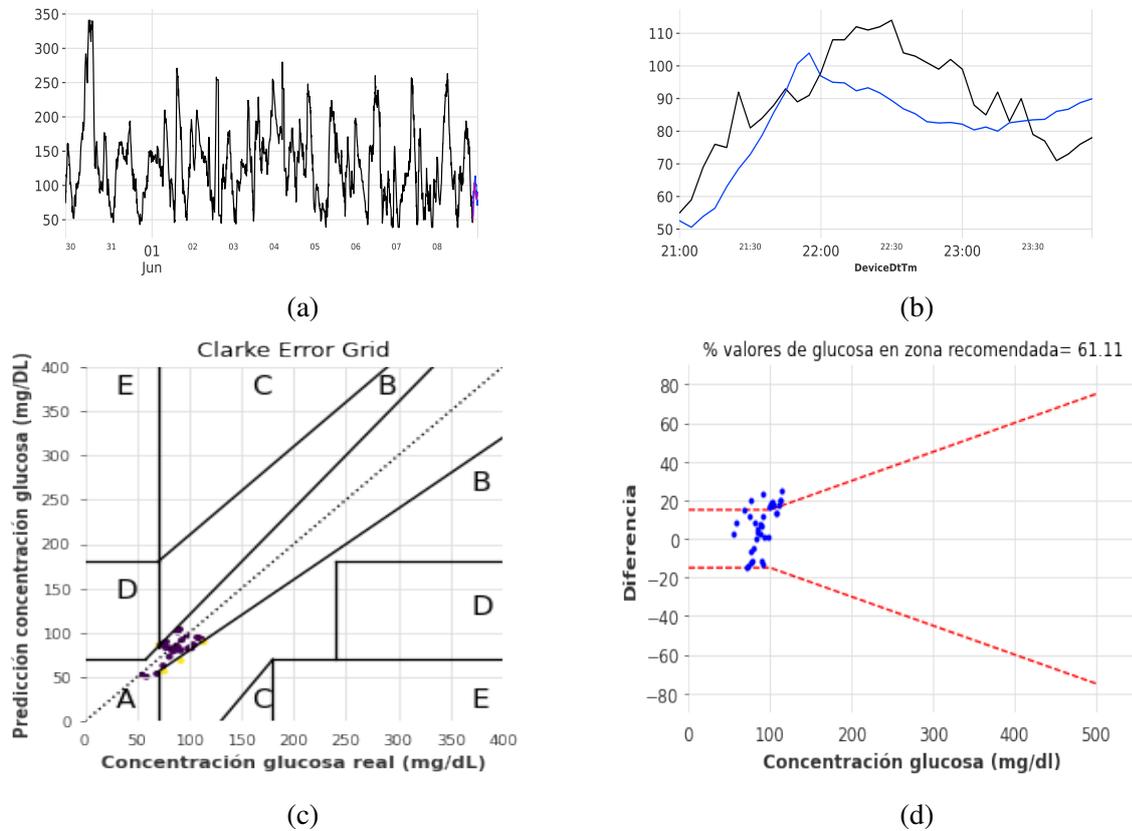


Figura 4.22: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD3-SEN sintética generada por el método CPAR para el segundo paciente con PtId 166 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

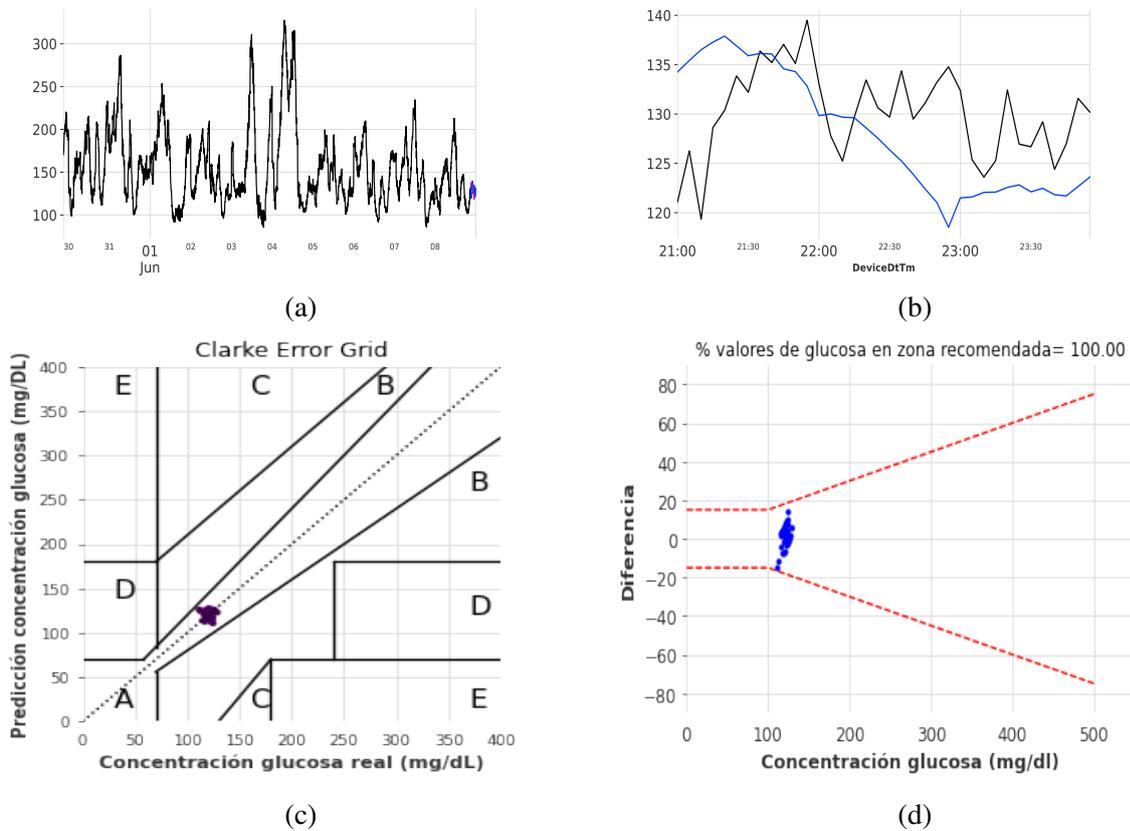


Figura 4.23: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD3-SEN sintética generada por el método DGAN para el segundo paciente con PId 166 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

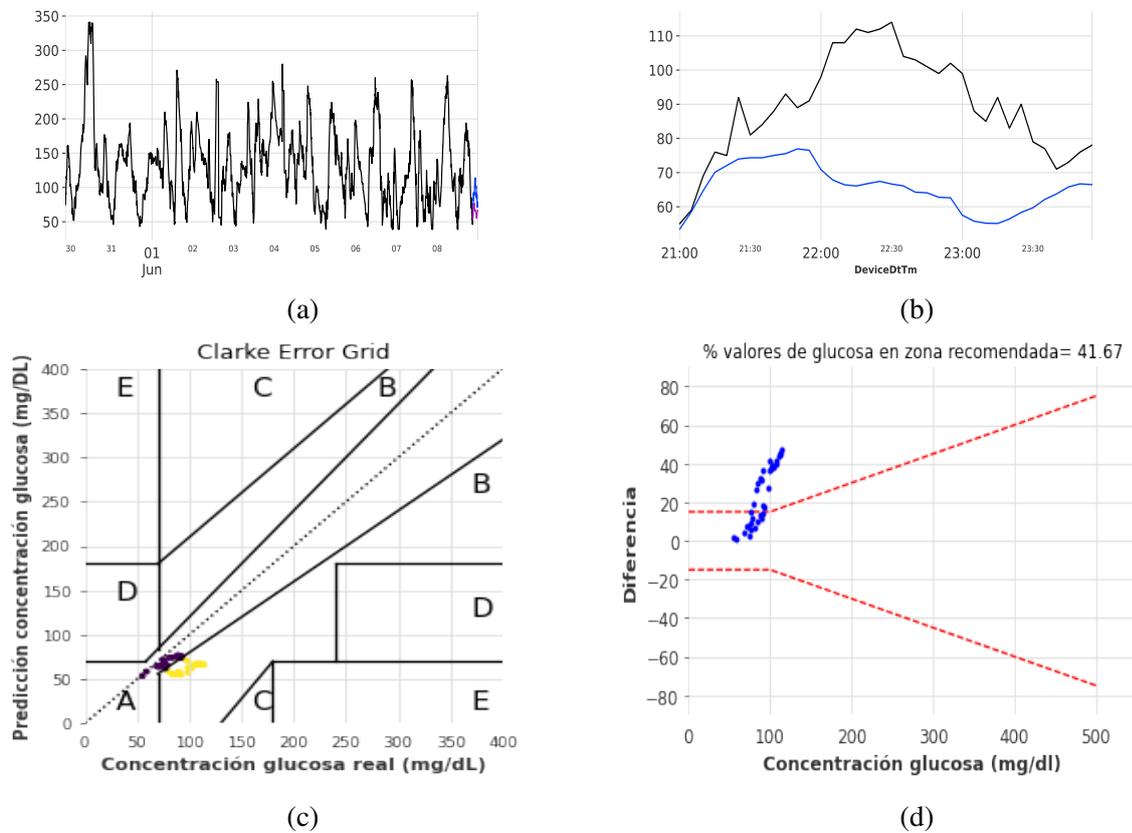


Figura 4.24: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD3-SEN sintética generada por el método CPAR para el segundo paciente con PtId 166 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

Capítulo 5

Conclusiones y líneas futuras

En este capítulo se presentan las conclusiones obtenidas tras la realización del presente TFG y además se incluyen varias líneas futuras de trabajo identificadas.

5.1. Conclusiones

Durante la realización de este TFG se han buscado métodos óptimos de generación de series temporales de glucosa. Para ello, en primer lugar se realizó una búsqueda de bases de datos que recopilaran diversas series temporales de pacientes con DMT1, obteniendo finalmente 3 bases de datos con las cuales se ha validado este estudio. Una vez seleccionadas las bases de datos, se realizó un análisis descriptivo y un extenso preprocesado. Trabajar con series temporales de glucosa con alta cantidad de valores perdidos supuso un gran reto para el desarrollo de este proyecto.

A continuación, se realizó una fase de revisión del estado del arte de generación de series temporales sintéticas, incluyendo las GAN, y de las métricas utilizadas para validar la calidad de las mismas. A raíz de la revisión de la literatura y el estado del arte de estas técnicas de DL, se decidió emplear las métricas mostradas: Divergencia de Jhensen-Shannon, MMD, test de Kolmogorov-Smirnov y distancia DTW. Además de estas métricas, se decidió validar las series temporales sintéticas en función a su capacidad predictiva. Es decir, valorando si los modelos de predicción empleados sobre series temporales reales son reproducibles en las series sintéticas.

Para la parte predictiva, también se llevo a cabo una extensa etapa de revisión del estado del arte y se probaron varios modelos de predicción (N-BEATS, N-HiTS). Finalmente, se elige el N-HiTS por sus buenos resultados y bajo coste computacional. Al tratarse de un método

muy nuevo, la fase inicial en la que se busca comprender el funcionamiento del modelo y encontrar los hiperparámetros óptimos fue de gran importancia. El modelo predictivo se aplicó al escenario real y a los sintéticos con los mismos hiperparámetros y variándolos entre las bases de datos. De este modo, se consigue que cada base de datos tuviera resultados óptimos, pero, comparables entre las series reales y sintéticas. Esta decisión se toma en base a que el fin último de la predicción es validar la bondad de las series temporales sintéticas generadas.

Teniendo en cuenta diversos factores como las métricas aplicadas, la capacidad predictiva, la resolución temporal y el coste computacional para comparar los métodos de generación de series temporales sintéticas se obtiene la conclusión general de este TFG. Tras todos los experimentos y evaluaciones, se podría decir que para el caso particular de generación de series temporales de glucosa el método DGAN ofrece mejores resultados, destacando su gran aptitud predictiva y el bajo coste computacional asociado al entrenamiento de este modelo.

Puesto que el objetivo principal de este TFG es encontrar una herramienta que permita generar series temporales fiables para poder realizar predicciones certeras de las concentraciones de glucosa en sangre, se puede decir que DGAN es un método óptimo que cumple con los requisitos mencionados. Lo cual permite identificar con antelación episodios de hiperglucemia e hipoglucemia en pacientes con DMT1 aunque los datos de los que se disponga sean escasos.

5.2. Líneas futuras

Este TFG supone una pequeña aportación al estado del arte del uso de técnicas de ML y de análisis de datos con el uso de datos clínicos. En concreto, aplica técnicas muy novedosas que ofrecen una solución a uno de los grandes problemas de la investigación con datos clínicos: la falta de datos. No obstante, se trata de un problema complejo y los resultados actuales no pueden ser aplicados. Sin embargo, este TFG puede ser de utilidad hacia otras investigaciones que puedan ir progresando en la misma línea. Entre posibles líneas futuras se encuentran:

- Explorar diferentes métodos de imputación específicos para cada serie temporal basados en los datos registrados en el pasado para ese mismo paciente, al ser posible en una franja horaria similar a la que se encuentra el valor nulo.
- Combinar las series temporales reales y sintéticas de las tres bases de datos para hacer una predicción con un mayor número de pacientes y evaluar si las predicciones mejoran con respecto al escenario en el que se entrenan solo los datos reales.

- Aplicar modelos de optimización (modelos de búsqueda *greedy*, Bayesiana) para automatizar el proceso de selección de hiperparámetros en los modelos de generación de series sintéticas (DGAN, CPAR) y modelo predictivo N-HiTS. Esto garantizará que los parámetros elegidos son los más óptimos para el problema de predicción de glucosa en ventanas temporales (horas).
- Aplicar una ventana de predicción móvil, a lo largo de al menos un día, mediante validación cruzada. De este modo se consigue evitar el sesgo en la predicción producido por seleccionar las últimas muestras de las series temporales de glucosa, ya que, la concentración de glucosa en sangre es muy variable a lo largo de un día.
- Extender el proyecto evaluando más modelos GAN, como los mostrados en la Tabla 3.1 para encontrar el que genere series temporales de glucosa que se ajusten con mayor precisión a las series temporales reales proporcionadas. Cuanto más realistas sean los datos sintéticos generados, más certera será la predicción de valores de glucosa y, por tanto, se podrá determinar con mayor seguridad cuando va a tener lugar un evento de hiperglucemia o hipoglucemia.
- Aplicar las predicciones de concentración de glucosa realizadas con el modelo N-HiTS al fin específico de anticipar los eventos de hiperglucemia e hipoglucemia.
- Evaluar a largo plazo la creación de un sistema de alerta de episodios de hipoglucemia e hiperglucemia basados en los modelos entrenados en el presente TFG. Se necesitaría crear un escenario multi-clase en el que tienes rango de glucemia, hipo, euglucemia y hiperglucemia.

Apéndice A

Evaluación de series temporales sintéticas

En este apéndice se muestran la Figura A.1 y la Figura A.2 relativas a los resultados obtenidos de la generación de series temporales sintéticas para la BD1-SEV. Además, se muestran la Figura A.3 y la Figura A.4, los histogramas de superposición de las distribuciones reales y las obtenidas de la generación de series temporales sintéticas para la BD2-PED. Por último, se muestran más resultados obtenidos para la BD3-SEN en la Figura A.5 y la Figura A.6

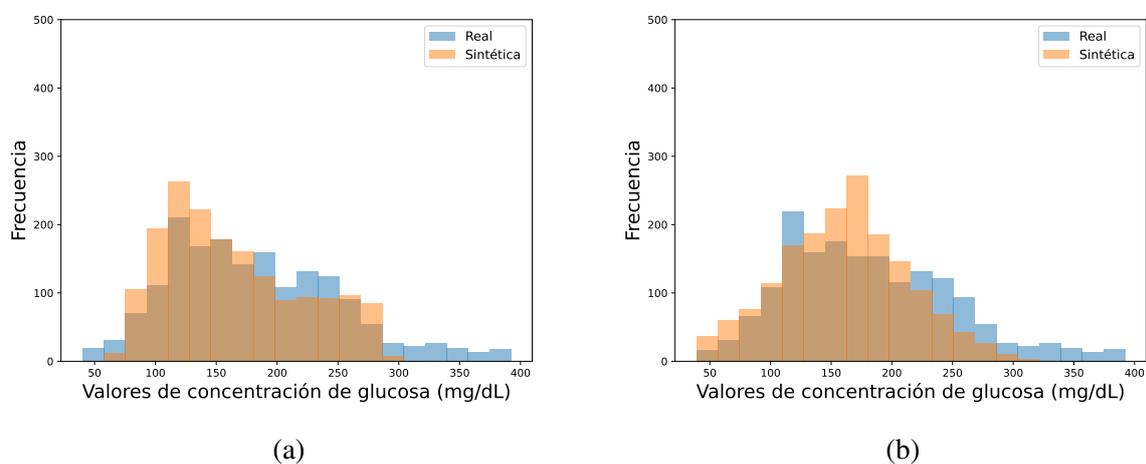


Figura A.1: Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD1-SEV y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

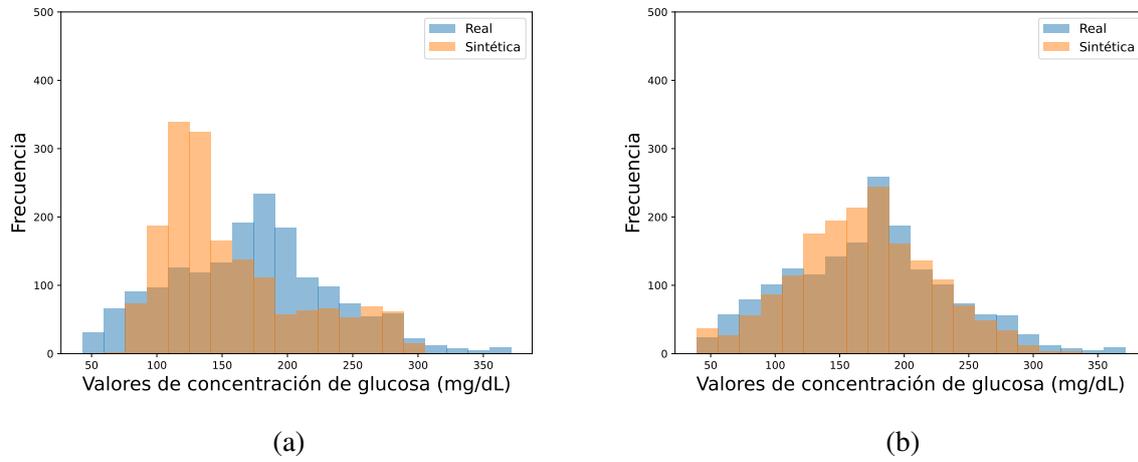


Figura A.2: Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD1-SEV y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

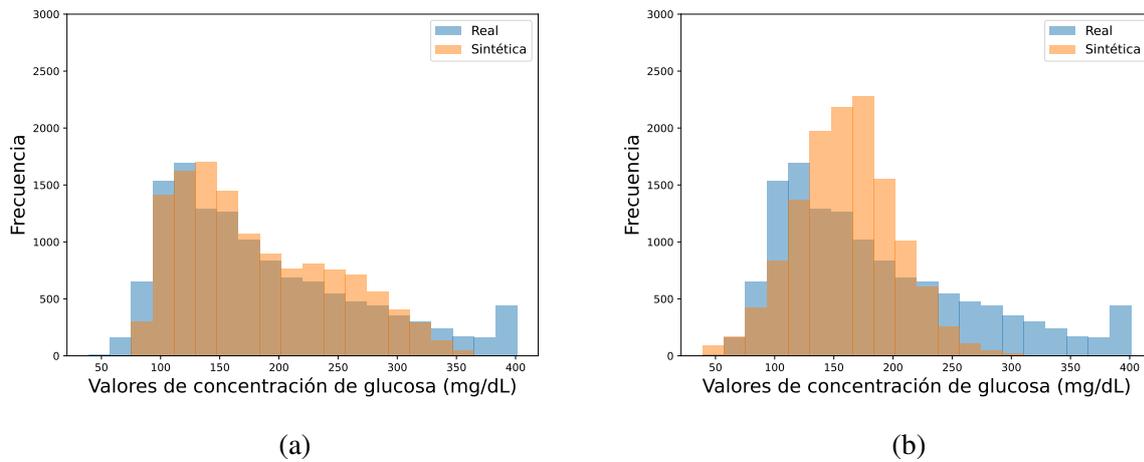


Figura A.3: Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD2-PED y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

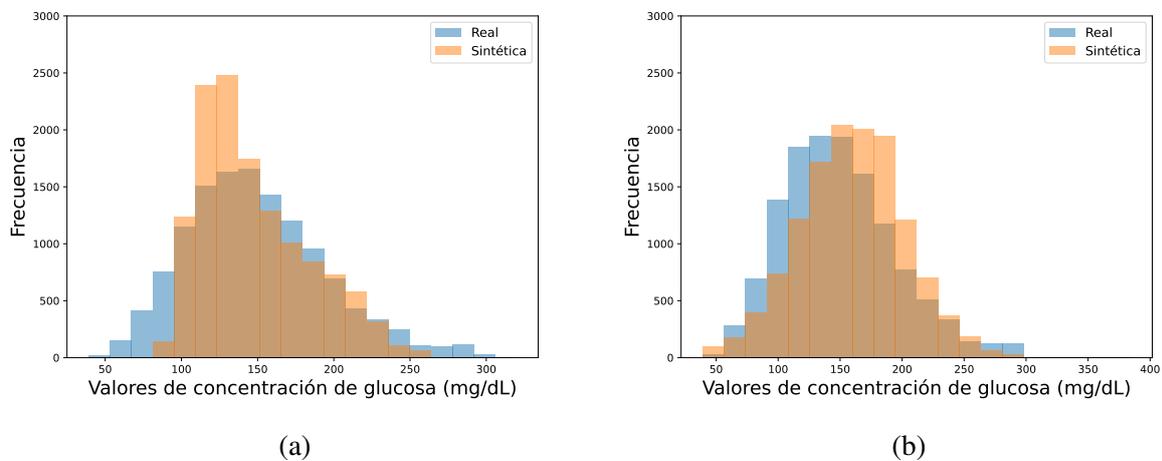


Figura A.4: Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD2-PED y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

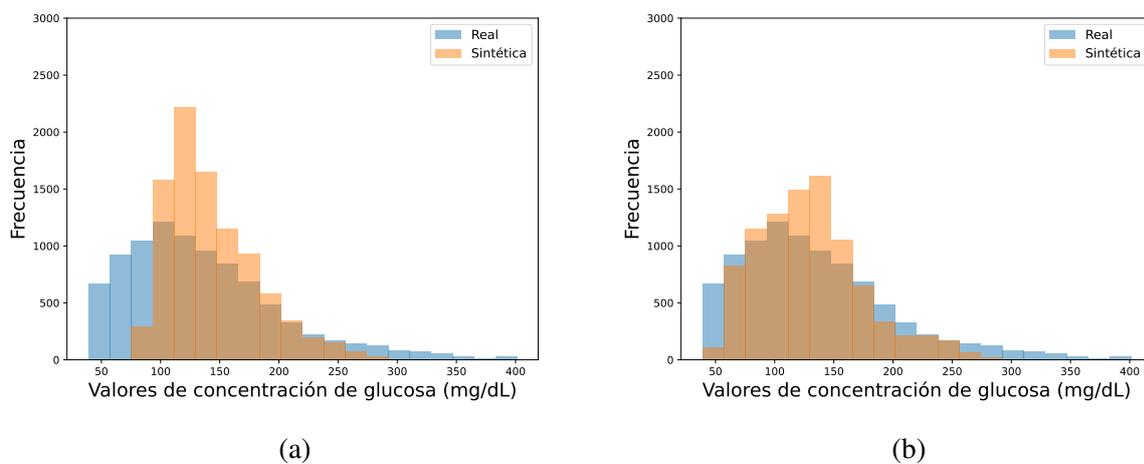


Figura A.5: Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD3-SEN y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

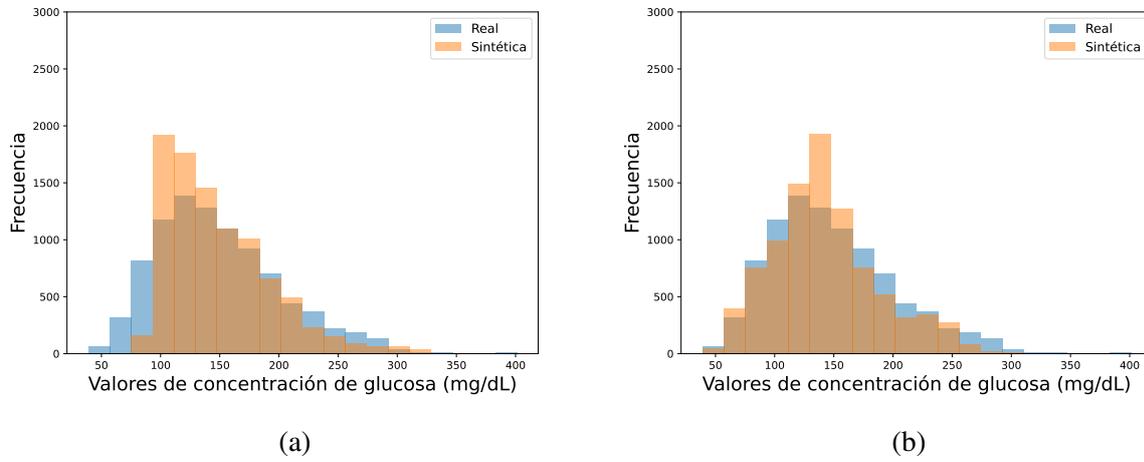


Figura A.6: Histogramas de la distribución de datos de glucosa de las series temporales reales de la BD3-SEN y series sintéticas generadas con DGAN y CPAR. Usando: (a) DGAN; (b) CPAR.

Apéndice B

Evaluación de predicción

En este apéndice se muestran dos resultados de los obtenidos de la predicción en series temporales para la BD1-SEV real en la Figura B.1 y la Figura B.2. Para las series temporales de BD1-SEV generadas con DGAN se ve la Figura B.3 y la Figura B.4. Por último, para ver dos resultados más de la predicción de series generadas con CPAR, se ve la Figura B.5 y la Figura B.6.

Del mismo modo, se muestran otros dos resultados de la BD2-PED real, como se puede ver en la Figura B.7 y la Figura B.8. Para ver los resultados de las series temporales generadas por DGAN se consulta la Figura B.9 y la Figura B.10. Para ver los resultados de predicción en las series temporales generadas por CPAR se consulta la Figura B.11 y la Figura B.12.

Por último, se muestran otros dos resultados de la BD3-SEN real, véase en la Figura B.13 y la Figura B.14. Para contemplar los resultados de las series temporales generadas por DGAN se consulta la Figura B.15 y la Figura B.16. Para ver los resultados de predicción en las series temporales generadas mediante CPAR se consulta la Figura B.17 y la Figura B.18.

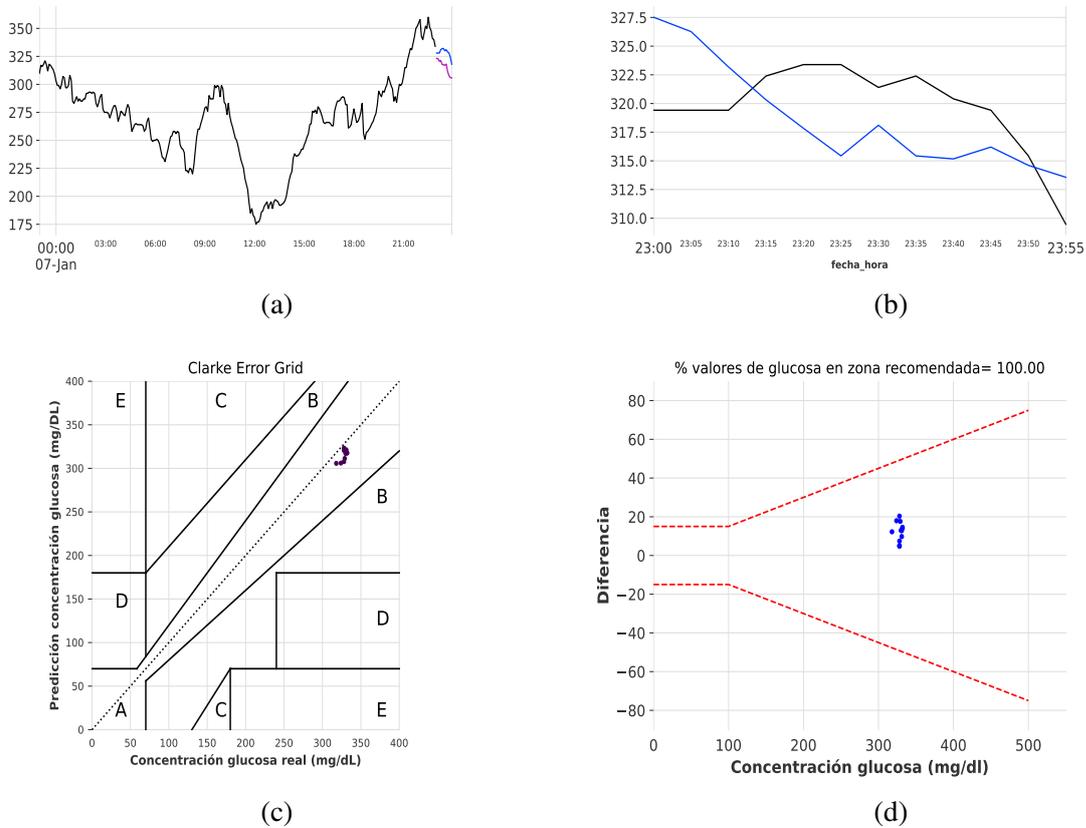


Figura B.1: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD1-SEV real para el paciente con PtId 202 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 1 día), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción (c) CEG, (d) ISO.

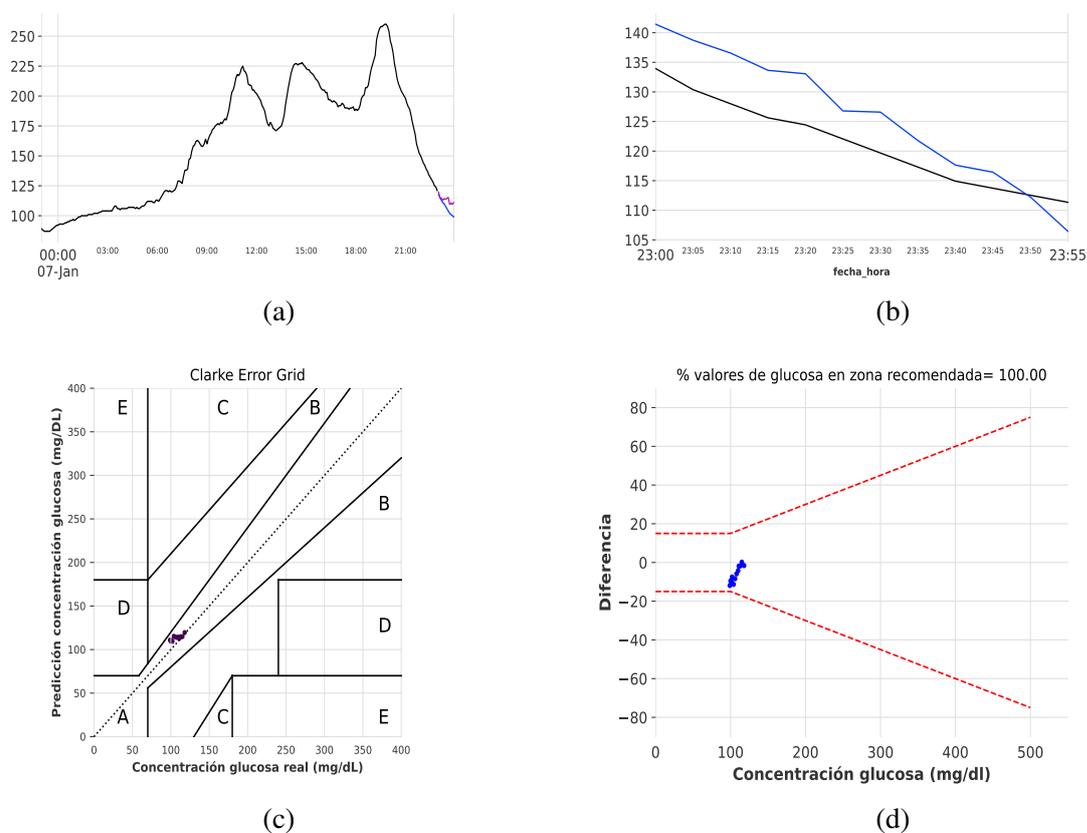


Figura B.2: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD1-SEV real para el paciente con PtId 21 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 1 día), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

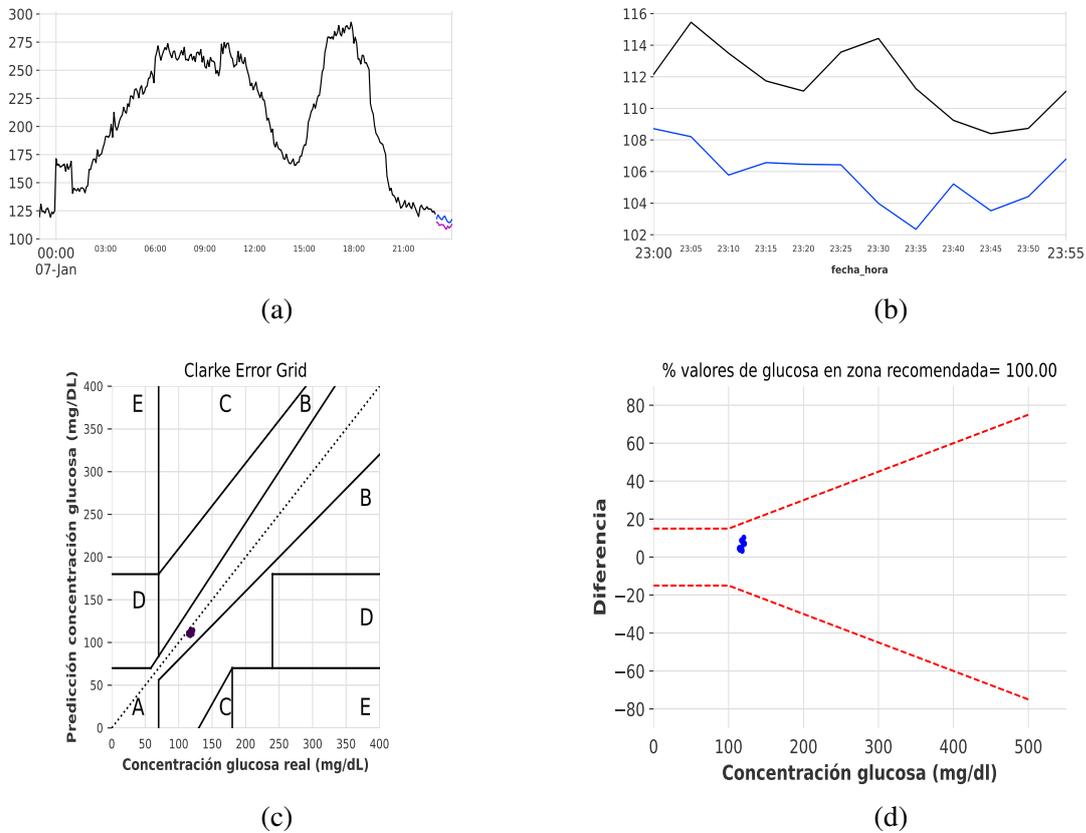


Figura B.3: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD1-SEV sintética generada por el método DGAN para el paciente con PtId 202 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 1 día), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

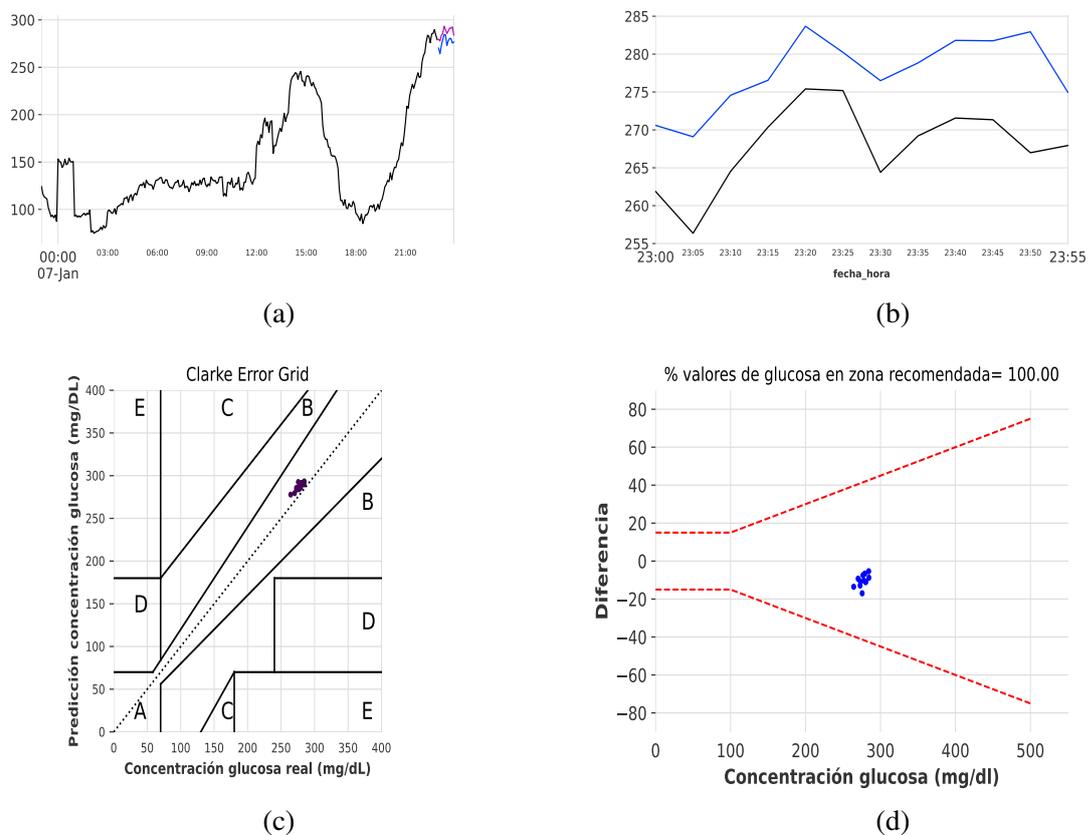


Figura B.4: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD1-SEV sintética generada por el método DGAN para el paciente con PtId 21 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 1 día), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

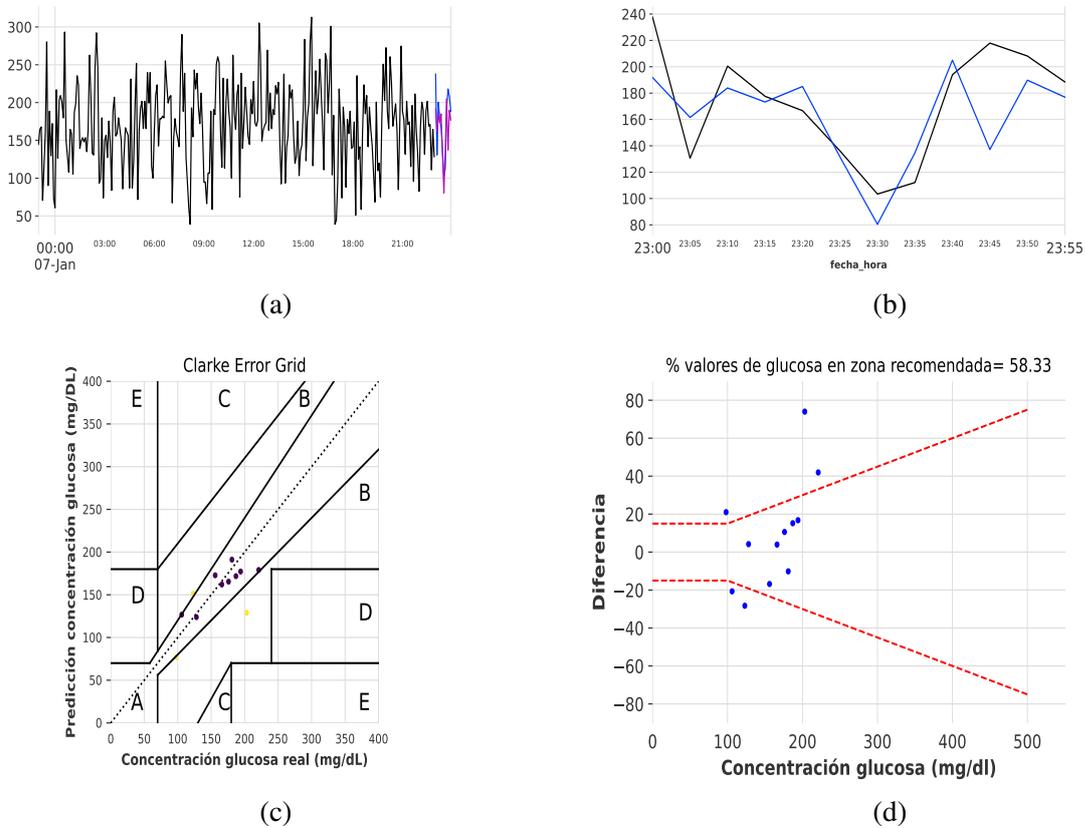


Figura B.5: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD1-SEV sintética generada por el método CPAR para el paciente con PtId 202 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 1 día), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

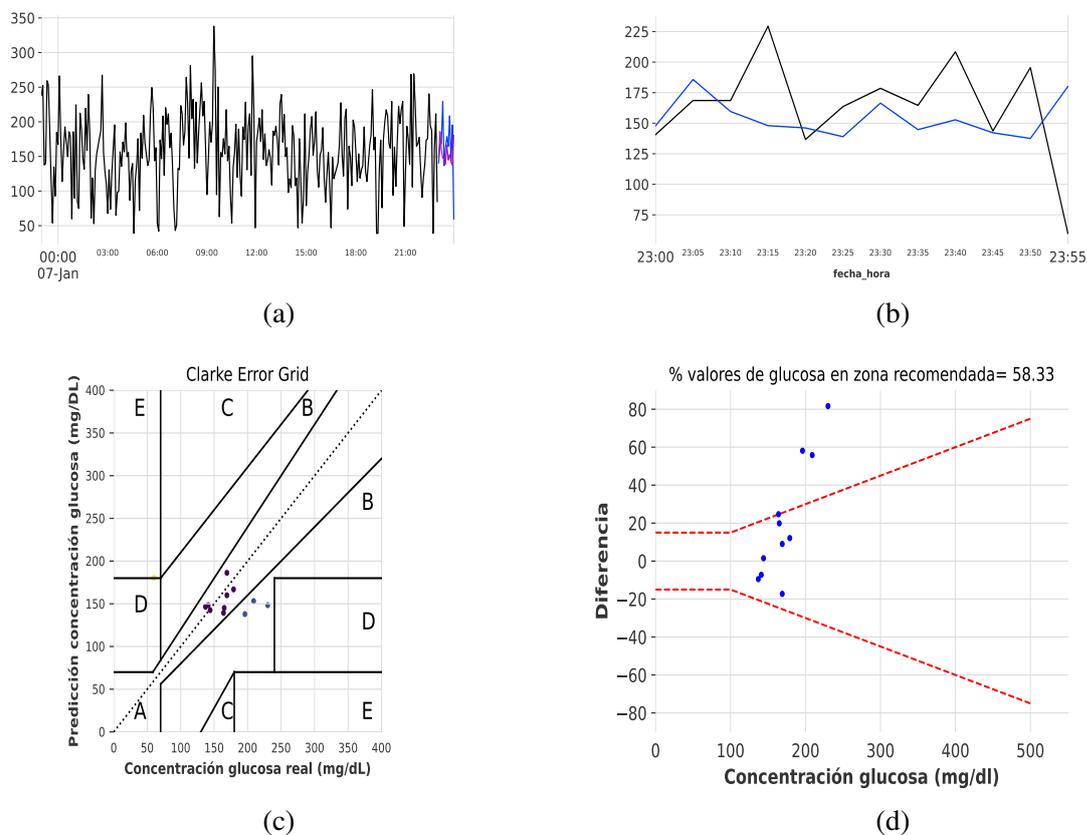
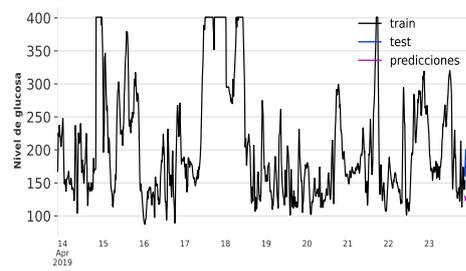


Figura B.6: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD1-SEV sintética generada por el método CPAR para el paciente con PtId 21 con una ventana de predicción de 12 muestras muestreadas cada 5 minutos, equivalente a 1 hora. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 1 día), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

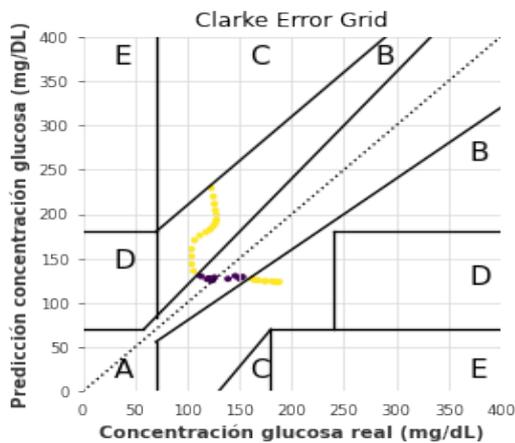
Resultados obtenidos de la predicción en series temporales para la BD2-PED.



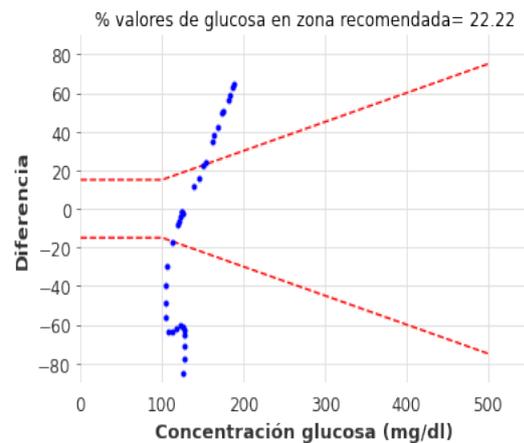
(a)



(b)

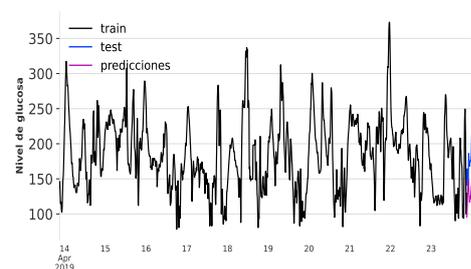


(c)

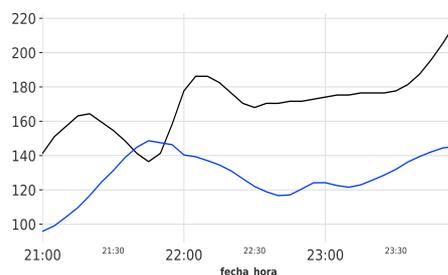


(d)

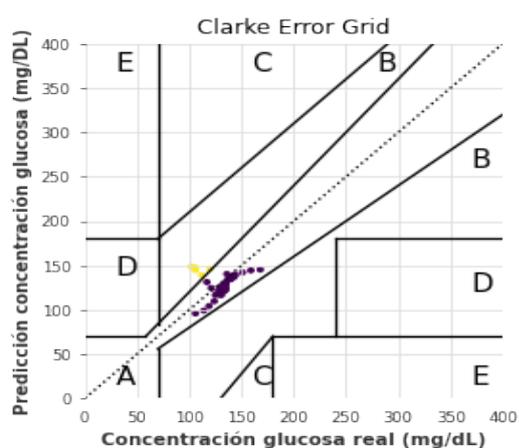
Figura B.7: Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED real para el paciente con PtId 53 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.



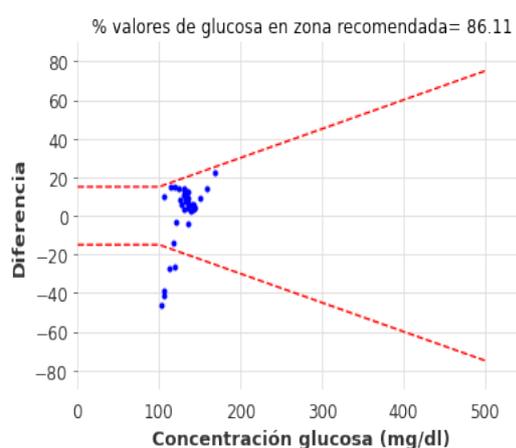
(a)



(b)



(c)



(d)

Figura B.8: Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED real para el paciente con PtId 86 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

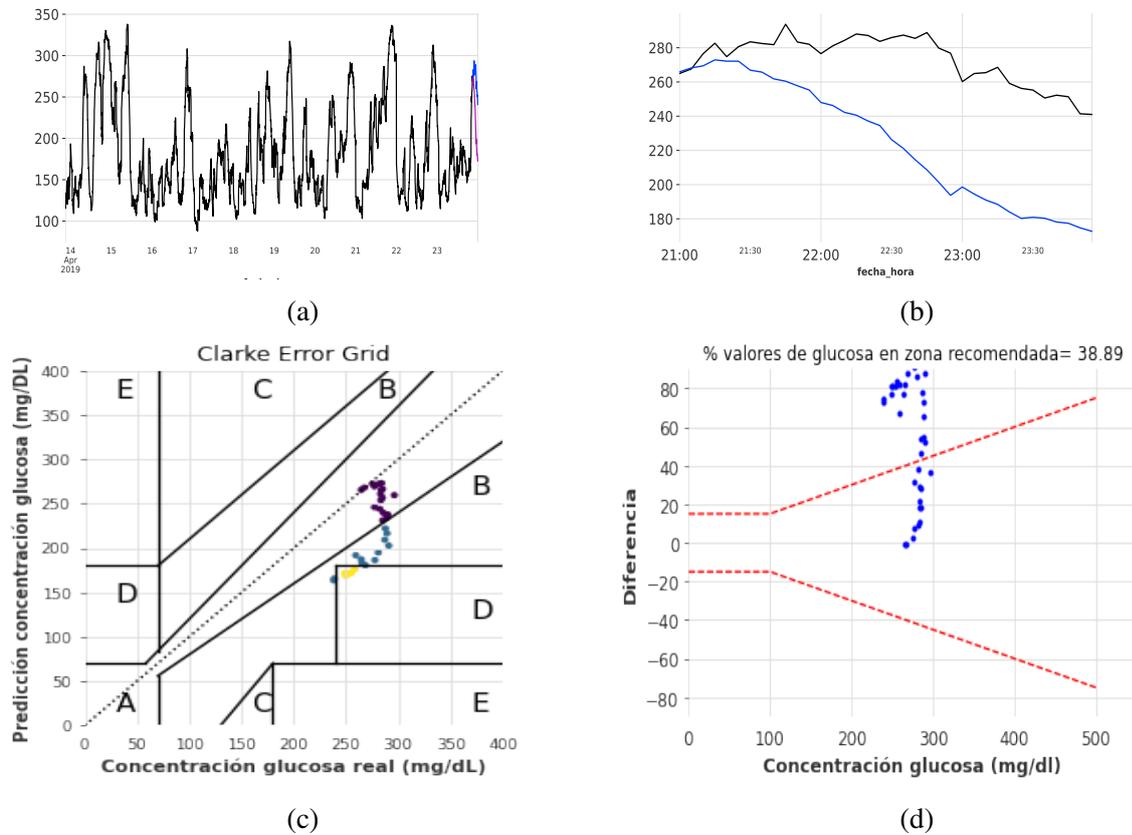


Figura B.9: Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD2-PED sintética generada por el método DAGAN para el paciente con PtId 53 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

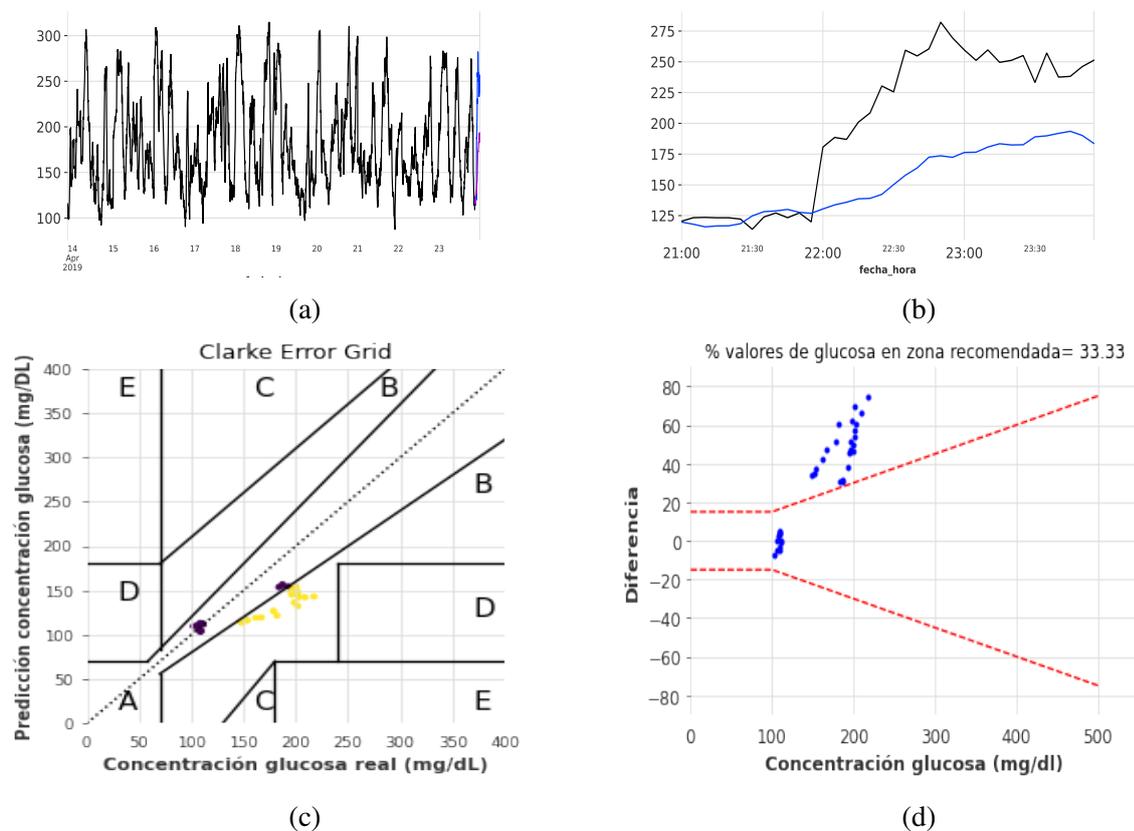


Figura B.10: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD2-PED sintética generada por el método DAGAN para el paciente con PtId 86 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

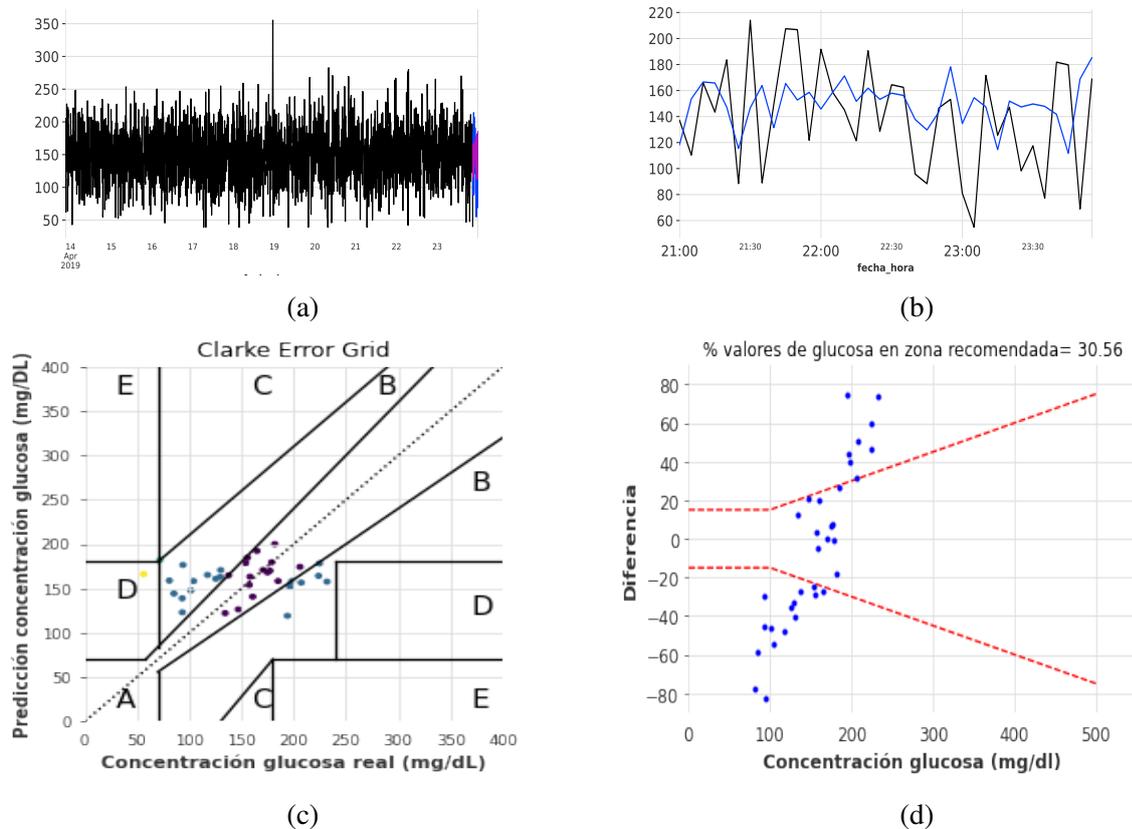


Figura B.11: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD2-PED sintética generada por el método CPAR para el paciente con PtId 53 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

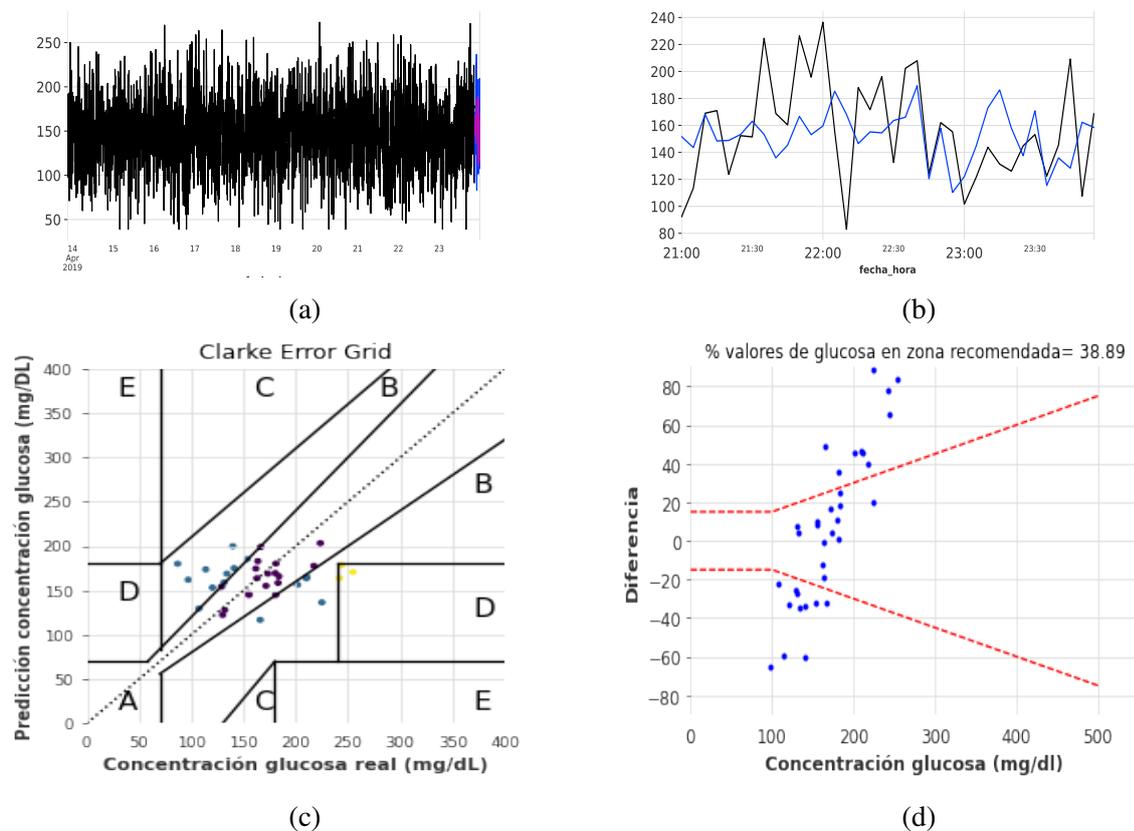


Figura B.12: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD2-PED sintética generada por el método CPAR para el paciente con PtId 86 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

Resultados obtenidos de la predicción en series temporales para la BD3-SEN

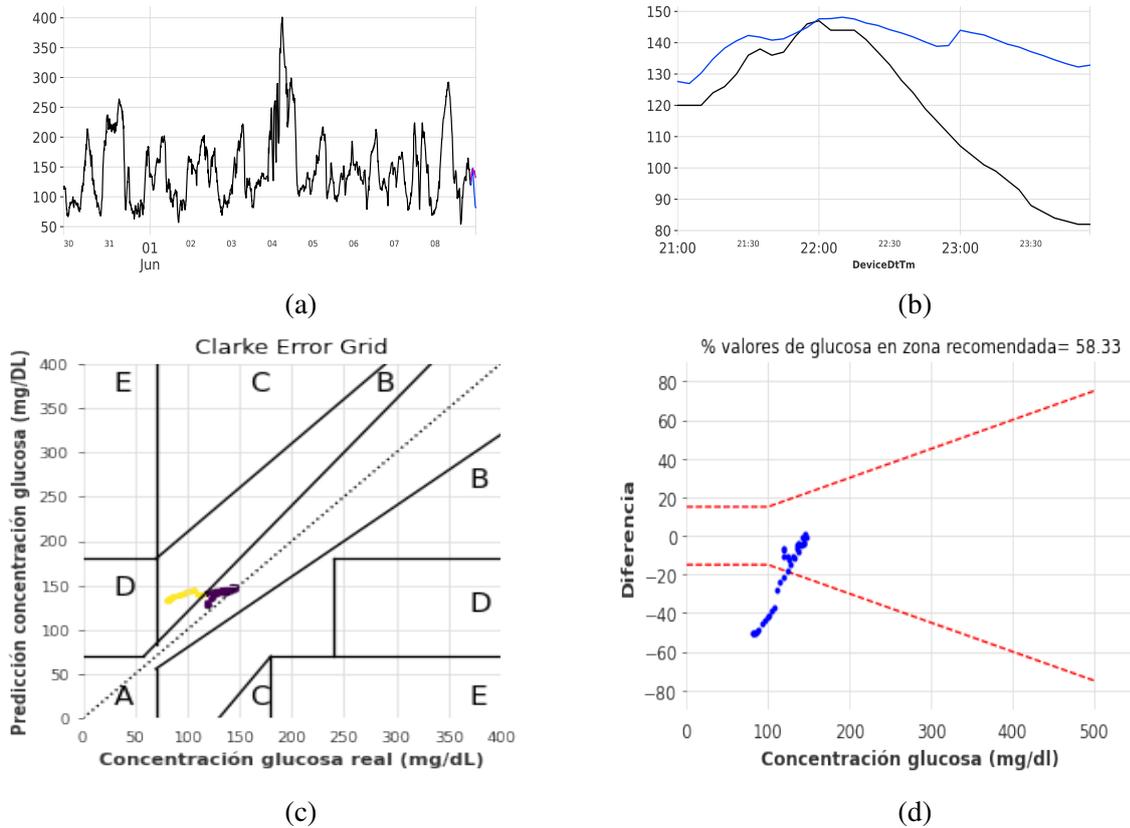


Figura B.13: Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD3-SEN sintética generada por el método CPAR para el paciente con PtId 151 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

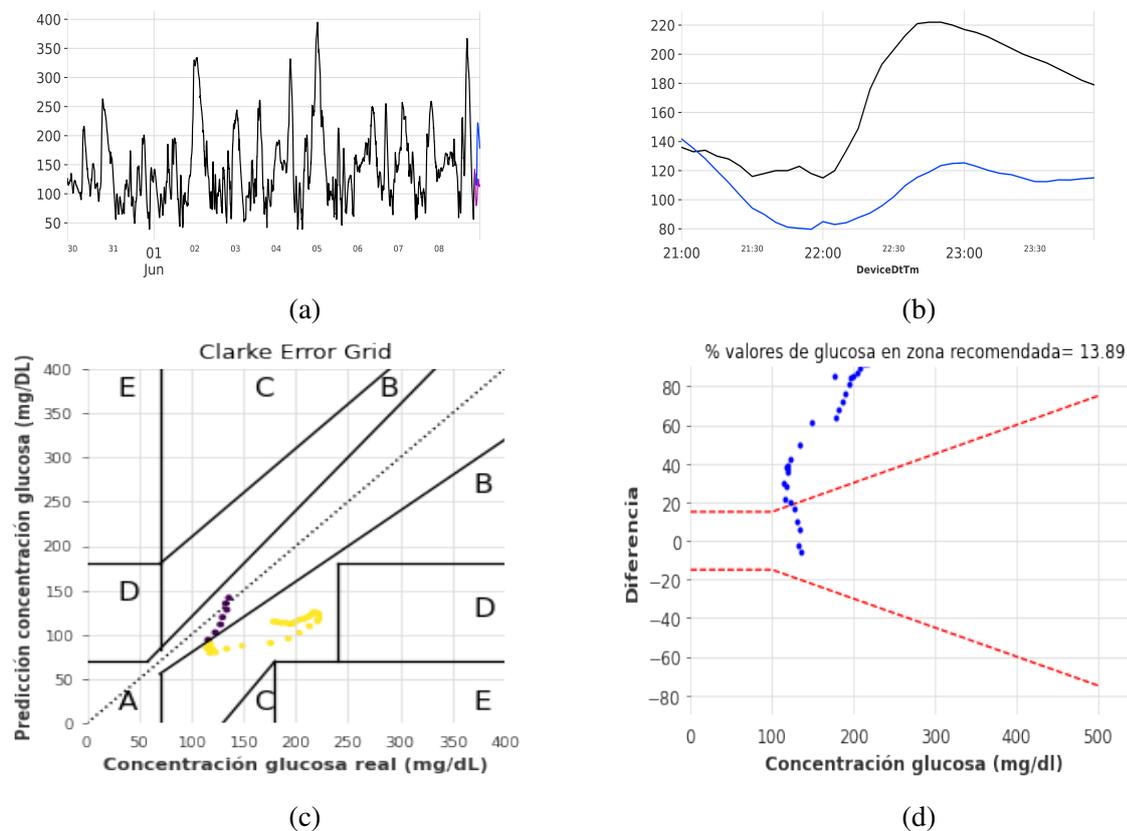


Figura B.14: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD3-SEN sintética generada por el método CPAR para el paciente con PtId 57 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

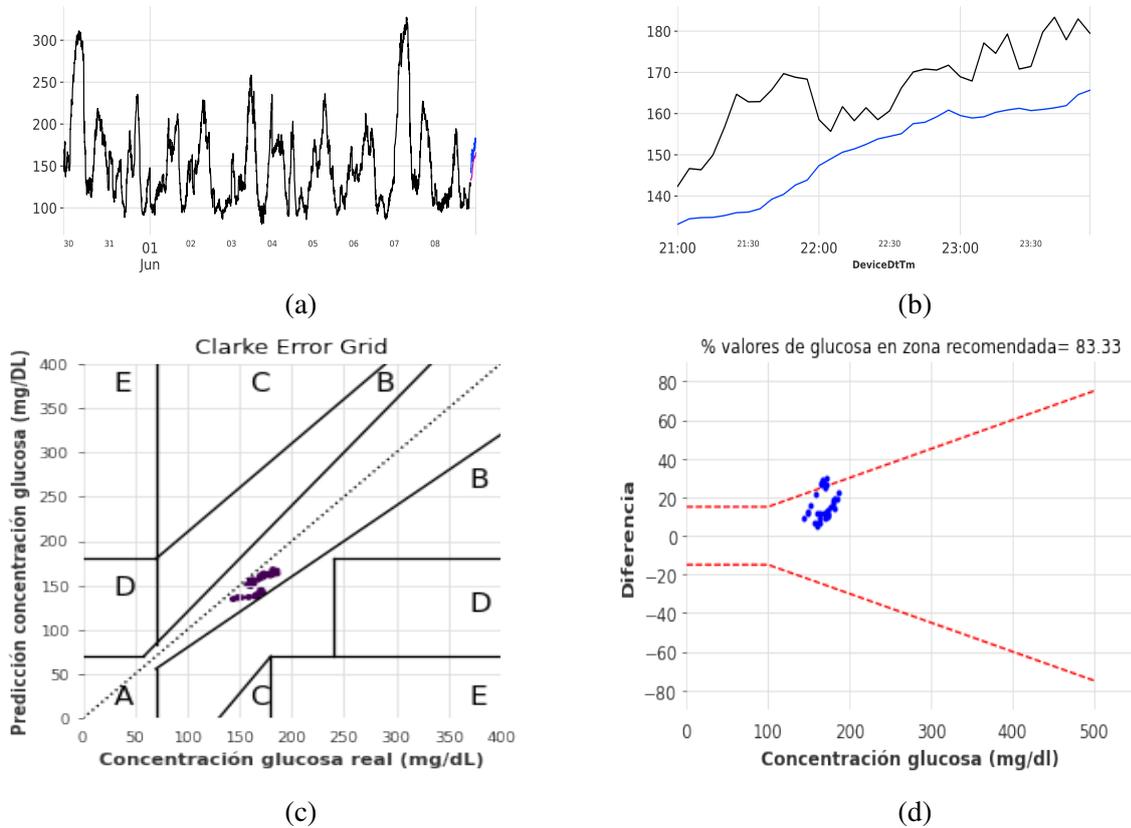


Figura B.15: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD3-SEN sintética generada por el método DGAN para el paciente con PtId 151 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

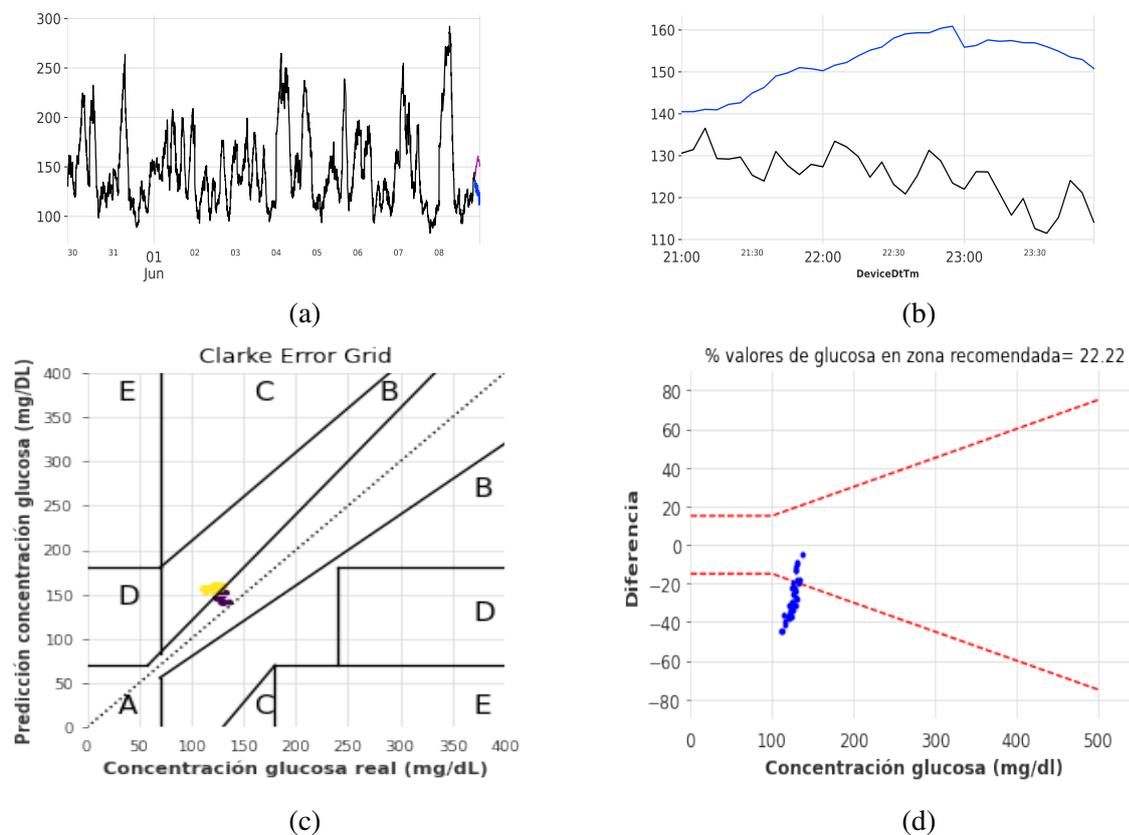


Figura B.16: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD3-SEN sintética generada por el método DGAN para el paciente con PtId 57 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

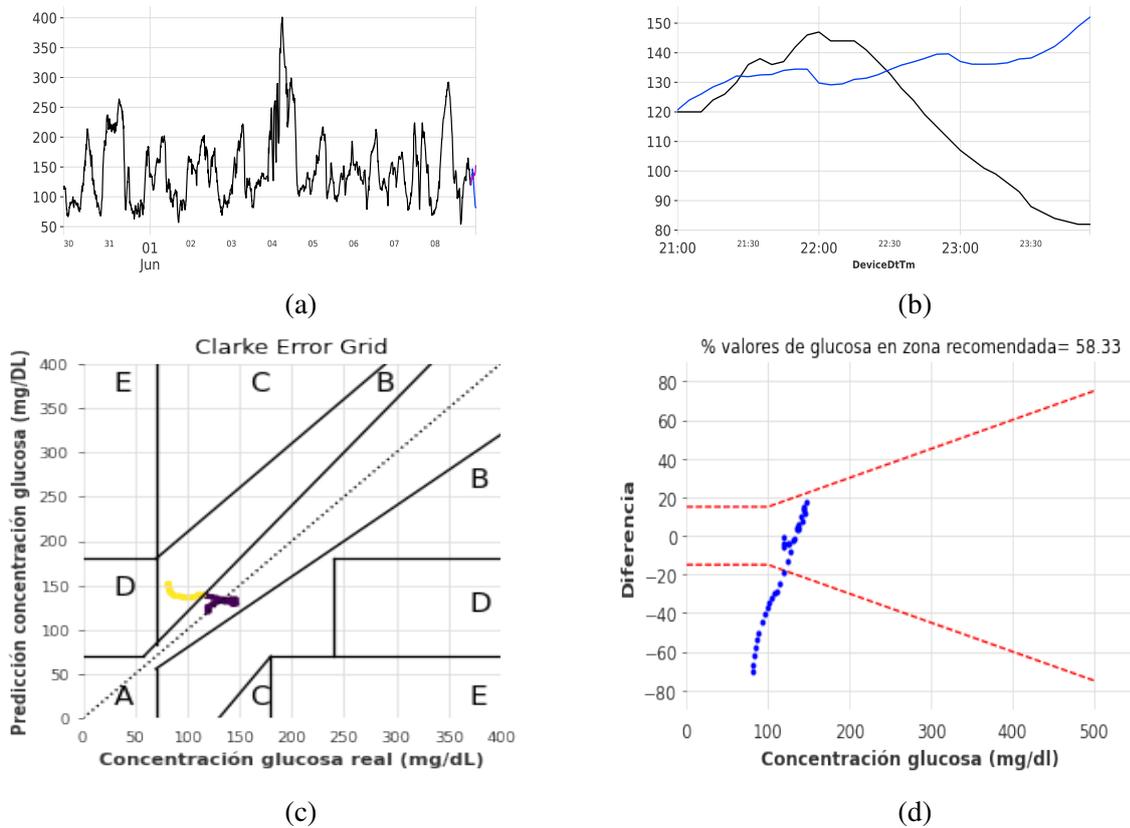


Figura B.17: Resultados obtenidos tras realizar una predicción con el modelo N-HiTS en la BD3-SEN sintética generada por el método CPAR para el paciente con PtId 151 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

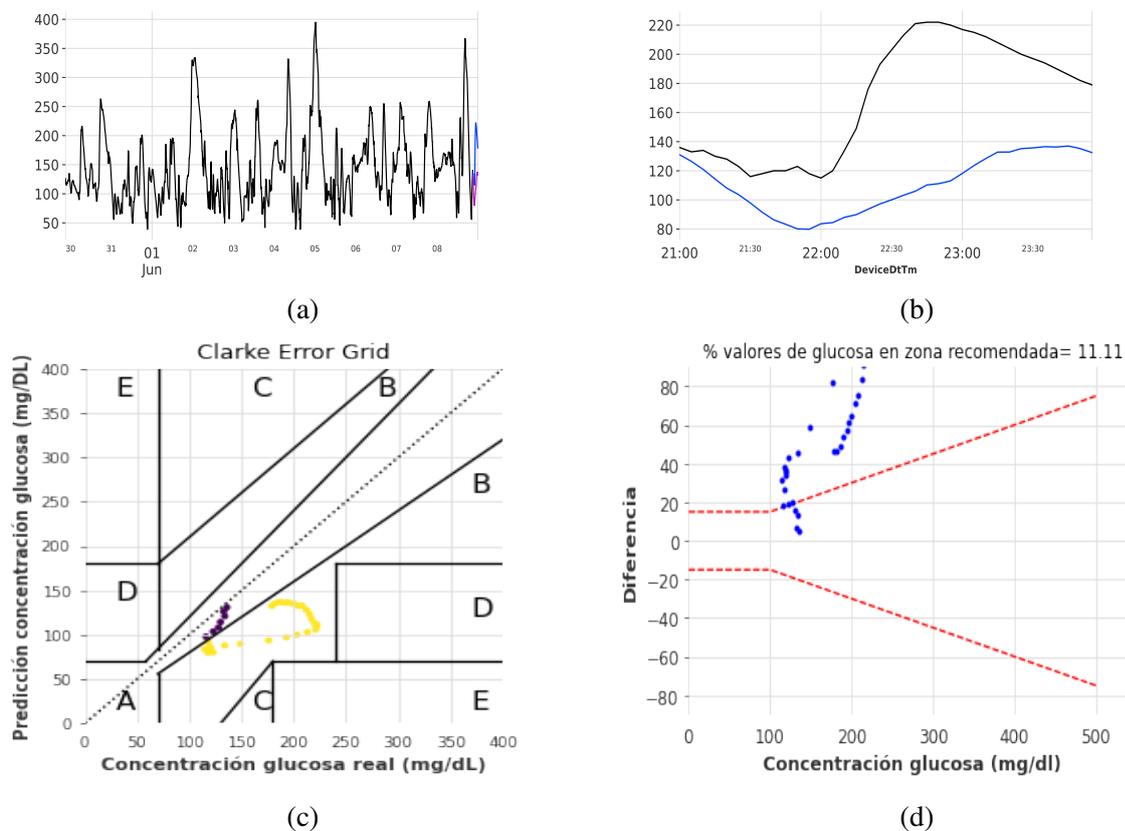


Figura B.18: Resultados obtenidos tras realizar una predicción con el modelo N-HITS en la BD3-SEN 2.2.3 sintética generada por el método CPAR para el paciente con PtId 57 con una ventana de predicción de 36 muestras muestreadas cada 5 minutos, equivalente a 3 horas. Se muestra: (a) Serie temporal de glucosa con los conjuntos *train* (se muestran solo 10 días), *test* que se utilizará como referencia visual de la exactitud de la predicción y (b) Serie temporal de glucosa con los conjuntos *test* y predicción(c) CEG, (d) ISO.

Bibliografía

- [1] Pouya Saeedi, Inga Petersohn, Paraskevi Salpea, Belma Malanda, Suvi Karuranga, Nigel Unwin, Stephen Colagiuri, Leonor Guariguata, Ayesha A Motala, Katherine Ogurtsova, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas. *Diabetes Research and Clinical Practice*, 157:107843, 2019.
- [2] Konstantinos Papatheodorou, Maciej Banach, Eleni Bekiari, Manfredi Rizzo, and Michael Edmonds. *Complications of diabetes 2017*, 2018.
- [3] Majid Mobasseri, Masoud Shirmohammadi, Tarlan Amiri, Nafiseh Vahed, Hossein Hosseini Fard, and Morteza Ghojzadeh. Prevalence and incidence of type 1 diabetes in the world: a systematic review and meta-analysis. *Health Promotion Perspectives*, 10(2):98, 2020.
- [4] National Institute of Diabetes, Digestive, and Kidney Diseases. Qué es la diabetes, 2016. Available at <https://www.niddk.nih.gov/health-information/informacion-de-la-salud/diabetes/informacion-general/que-es>.
- [5] Zhiheng He and George L King. Microvascular complications of diabetes. *Endocrinology and Metabolism Clinics*, 33(1):215–238, 2004.
- [6] Michael J Fowler. Microvascular and macrovascular complications of diabetes. *Clinical Diabetes*, 26(2):77–82, 2008.
- [7] Gita Shafiee, Mohammadreza Mohajeri-Tehrani, Mohammad Pajouhi, and Bagher Larijani. The importance of hypoglycemia in diabetic patients. *Journal of Diabetes & Metabolic Disorders*, 11:1–7, 2012.
- [8] Abu Bakar, Siti Nur Qomariah, Candra Hadi Santoso, Mono Pratiko Gustomi, Yuanita Syaful, and Lilis Fatmawa. Factors the incidence of hypoglycemia in diabetes mellitus patients: A pilot study in the emergency room. *Enfermería Clínica*, 30:46–49, 2020.

- [9] Eleftheria Papachristoforou, Vaia Lambadiari, Eirini Maratou, Konstantinos Makrilakis, et al. Association of glycemic indices (hyperglycemia, glucose variability, and hypoglycemia) with oxidative stress and diabetic complications. *Journal of Diabetes Research*, 2020, 2020.
- [10] Rozalina G McCoy, Jeph Herrin, Rodolfo J Galindo, Kavya Sindhu Swarna, Guillermo E Umpierrez, Sherita Hill Golden, and Patrick J O’connor. Rates of hypoglycemic and hyperglycemic emergencies among us adults with diabetes, 2011–2020. *Diabetes Care*, 46(2):e69–e71, 2023.
- [11] David C Klonoff. Continuous glucose monitoring: roadmap for 21st century diabetes therapy. *Diabetes Care*, 28(5):1231–1239, 2005.
- [12] Jort Kropff, Pratik Choudhary, Sankalpa Neupane, Katharine Barnard, Steve C Bain, Christoph Kapitza, Thomas Forst, Manuela Link, Andrew Dehennis, and J Hans DeVries. Accuracy and longevity of an implantable continuous glucose sensor in the precise study: a 180-day, prospective, multicenter, pivotal trial. *Diabetes Care*, 40(1):63–68, 2017.
- [13] Antonio Rafael Sabino Parmezan, Vinicius MA Souza, and Gustavo EAPA Batista. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information sciences*, 484:302–337, 2019.
- [14] Abdullah Alanazi. Using machine learning for healthcare challenges and opportunities. *Informatics in Medicine Unlocked*, 30:100924, 2022.
- [15] Omer Mujahid, Ivan Contreras, and Josep Vehi. Machine learning techniques for hypoglycemia prediction: trends and challenges. *Sensors*, 21(2):546, 2021.
- [16] Satoru Kodama, Kazuya Fujihara, Haruka Shiozaki, Chika Horikawa, Mayuko Harada Yamada, Takaaki Sato, Yuta Yaguchi, Masahiko Yamamoto, Masaru Kitazawa, Midori Iwanaga, et al. Ability of current machine learning algorithms to predict and detect hypoglycemia in patients with diabetes mellitus: meta-analysis. *JMIR Diabetes*, 6(1):e22458, 2021.
- [17] Rodolfo J Galindo and Grazia Aleppo. Continuous glucose monitoring: the achievement of 100 years of innovation in diabetes technology. *Diabetes Research And Clinical Practice*, 170:108502, 2020.

- [18] Darpit Dave, Daniel J DeSalvo, Balakrishna Haridas, Siripoom McKay, Akhil Shenoy, Chester J Koh, Mark Lawley, and Madhav Erraguntla. Feature-based machine learning model for real-time hypoglycemia prediction. *Journal of Diabetes Science and Technology*, 15(4):842–855, 2021.
- [19] Unión Europea. Reglamento (ue) 2016/679 del parlamento europeo y del consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales ya la libre circulación de estos datos y por el que se deroga la directiva 95/46/ce (reglamento general de protección de datos)(texto pertinente a efectos del eee). 2021.
- [20] Simon Bing, Andrea Dittadi, Stefan Bauer, and Patrick Schwab. Conditional generation of medical time series for extrapolation to underrepresented populations. *PLOS Digital Health*, 1(7):e0000074, 2022.
- [21] Saloni Dash, Andrew Yale, Isabelle Guyon, and Kristin P Bennett. Medical time-series data generation using generative adversarial networks. In *Artificial Intelligence in Medicine: 18th International Conference on Artificial Intelligence in Medicine, AIME 2020, Proceedings 18*, pages 382–391, Minneapolis, MN, USA, 2020. Springer.
- [22] Nita Gandhi Forouhi and Nicholas J Wareham. Epidemiology of diabetes. *Medicine*, 47(1):22–27, 2019.
- [23] Epdata. Evaluación de la prevalencia de diabetes por continente y proyecciones futuras, 2022. Available at <https://www.epdata.es/evolucion-prevalencia-diabetes-continente-proyecciones-futuras/ef2580cf-d887-4e86-a3db-614b75d32216>.
- [24] American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes care*, 37(Supplement_1):S81–S90, 2014.
- [25] Denis Daneman. Type 1 diabetes. *The Lancet*, 367(9513):847–858, 2006.
- [26] Andy Menke, Trevor J Orchard, Giuseppina Imperatore, Kai McKeever Bullard, Elizabeth Mayer-Davis, and Catherine C Cowie. The prevalence of type 1 diabetes in the united states. *Epidemiology (Cambridge, Mass.)*, 24(5):773, 2013.
- [27] Anastasia Katsarou, Soffia Gudbjörnsdottir, Araz Rawshani, Dana Dabelea, Ezio Bonifacio, Barbara J Anderson, Laura M Jacobsen, Desmond A Schatz, and Åke Lernmark. Type 1 diabetes mellitus. *Nature Reviews Disease Primers*, 3(1):1–17, 2017.

- [28] Kellee M Miller, Nicole C Foster, Roy W Beck, Richard M Bergenstal, Stephanie N DuBose, Linda A DiMeglio, David M Maahs, and William V Tamborlane. Current state of type 1 diabetes treatment in the us: updated data from the t1d exchange clinic registry. *Diabetes Care*, 38(6):971–978, 2015.
- [29] John J Mastrototaro. The minimed continuous glucose monitoring system. *Diabetes Technology & Therapeutics*, 2(1, Supplement 1):13–18, 2000.
- [30] World Health Organization. Mean fasting blood glucose, 2016. Available at <https://www.who.int/data/gho/indicator-metadata-registry/imr-details/2380#:~:text=The%20expected%20values%20for%20normal,and%20monitoring%20glycemia%20are%20recommended.>
- [31] MI Mouri and M Badireddy. Hyperglycemia. *StatPearls*, 2023. Updated 2023 Apr 24.
- [32] Cecilia Esparza Catalán. Series temporales. *Laboratorio de Estadística. Consejo Superior de Investigaciones Científicas (CSIC). España*, 54, 2004.
- [33] JAEB. JAEB, 2023. Available at <https://public.jaeb.org/datasets/diabetes>.
- [34] Ruth S Weinstock, Dongyuan Xing, David M Maahs, Aaron Michels, Michael R Rickels, Anne L Peters, Richard M Bergenstal, Breanne Harris, Stephanie N DuBose, Kellee M Miller, et al. Severe hypoglycemia and diabetic ketoacidosis in adults with type 1 diabetes: results from the t1d exchange clinic registry. *The Journal of Clinical Endocrinology & Metabolism*, 98(8):3411–3419, 2013.
- [35] T1D Exchange. The T1D Exchange Registry, 2023. Available at <https://t1dexchange.org/registry/>.
- [36] Yifan Zhang and Peter J Thorburn. A dual-head attention model for time series data imputation. *Computers and Electronics in Agriculture*, 189:106377, 2021.
- [37] Dayal Kumar Behera, Shreela Dash, Ajit Kumar Behera, and CH Sanjeev Kumar Dash. Extreme gradient boosting and soft voting ensemble classifier for diabetes prediction. In *2021 19th OITS International Conference on Information Technology (OCIT)*, pages 191–195, Silicon Institute of Technology, Bhubaneswar, 2021. IEEE.
- [38] Edgar Acuna, Roxana Aparicio, and Velcy Palomino. Analyzing the performance of transformers for the prediction of the blood glucose level considering imputation and smoothing. *Big Data and Cognitive Computing*, 7(1):41, 2023.

- [39] Shah Atiqur Rahman, Yuxiao Huang, Jan Claassen, and Samantha Kleinberg. Imputation of missing values in time series with lagged correlations. In *2014 IEEE International Conference on Data Mining Workshop*, pages 753–762, Shenzhen, China, 2014. IEEE.
- [40] Adam Pantanowitz and Tshilidzi Marwala. Missing data imputation through the use of the random forest algorithm. In *Advances in computational intelligence*, pages 53–62, Berlin, Heidelberg, 2009. Springer.
- [41] Fei Tang and Hemant Ishwaran. Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 10(6):363–377, 2017.
- [42] M AndersonStacey, A BrownSue, E PinskerJordan, ChurchMei Mei, C KudvaYogish, P ForlenzaGregory, WadwaR Paul, J DoyleFrancis III, DeVriesJ Hans, Del FaveroSimone, et al. The international diabetes closed-loop study: testing artificial pancreas component interoperability. *Diabetes technology & therapeutics*, 2019.
- [43] RE Pratley, LG Kanapka, MR Rickels, A Ahmann, G Aleppo, R Beck, A Bhargava, BW Bode, A Carlson, NS Chaytor, et al. Wireless innovation for seniors with diabetes mellitus (wisdm) study group. effect of continuous glucose monitoring on hypoglycemia in older adults with type 1 diabetes: a randomized clinical trial. *Jama*, 323(23):2397–2406, 2020.
- [44] Qifang Bi, Katherine E Goodman, Joshua Kaminsky, and Justin Lessler. What is machine learning? a primer for the epidemiologist. *American Journal of Epidemiology*, 188(12):2222–2239, 2019.
- [45] Jason Bell. What is machine learning? *Machine Learning and the City: Applications in Architecture and Urban Design*, pages 207–216, 2022.
- [46] Ramadass Sathya, Annamma Abraham, et al. Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence*, 2(2):34–38, 2013.
- [47] Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning. *Electronic Markets*, 31(3):685–695, 2021.
- [48] Jianqing Fan, Cong Ma, and Yiqiao Zhong. A selective overview of deep learning. *Statistical Science: A Review Journal of The Institute of Mathematical Statistics*, 36(2):264, 2021.

- [49] Katarzyna Janocha and Wojciech Marian Czarnecki. On loss functions for deep neural networks in classification. *ArXiv Preprint ArXiv*, 1:1702–5659, 2017.
- [50] Ravindra Parmar. Training Deep Neural Networks, 2018. Available at <https://towardsdatascience.com/training-deep-neural-networks-9fdb1964b964>.
- [51] Igor Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1):89–109, 2001.
- [52] José A Castellanos-Garzón, Ernesto Costa, Juan M Corchado, et al. An evolutionary framework for machine learning applied to medical data. *Knowledge-Based Systems*, 185:104982, 2019.
- [53] Roohallah Alizadehsani, Mohamad Roshanzamir, Sadiq Hussain, Abbas Khosravi, Afshaneh Koohestani, Mohammad Hossein Zangoeei, Moloud Abdar, Adham Beykikhoshk, Afshin Shoeibi, Assef Zare, et al. Handling of uncertainty in medical data using machine learning and probability theory techniques: A review of 30 years (1991–2020). *Annals of Operations Research*, 41:1–42, 2021.
- [54] Antreas Antoniou, Amos Storkey, and Harrison Edwards. Data augmentation generative adversarial networks. *ArXiv Preprint ArXiv*, 3:1711–4340, 2017.
- [55] Cristóbal Esteban, Stephanie L Hyland, and Gunnar Rätsch. Real-valued (medical) time series generation with recurrent conditional gans. *ArXiv Preprint ArXiv*, 2:1706–02633, 2017.
- [56] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018.
- [57] Chris Donahue, Julian McAuley, and Miller Puckette. Adversarial audio synthesis. *ArXiv Preprint ArXiv*, 3:1802–04208, 2018.
- [58] Yongjun Hong, Uiwon Hwang, Jaeyoon Yoo, and Sungroh Yoon. How generative adversarial networks and their variants work: An overview. *ACM Computing Surveys (CSUR)*, 52(1):1–43, 2019.
- [59] Simon Bing, Andrea Dittadi, Stefan Bauer, and Patrick Schwab. Conditional generation of medical time series for extrapolation to underrepresented populations. *PLOS Digital Health*, 1(7):e0000074, 2022.

- [60] Yunfei Zhang, Zihua Zhou, Junwei Liu, and Jianjuan Yuan. Data augmentation for improving heating load prediction of heating substation based on timegan. *Energy*, 260:124919, 2022.
- [61] Blanca Pena and Luofeng Huang. Wave-gan: a deep learning approach for the prediction of nonlinear regular wave loads and run-up on a fixed cylinder. *Coastal Engineering*, 167:103902, 2021.
- [62] Zinan Lin, Alankar Jain, Chen Wang, Giulia Fanti, and Vyas Sekar. Using gans for sharing networked time series data: Challenges, initial promise, and open questions. In *Proceedings of the ACM Internet Measurement Conference*, pages 464–483, Pittsburgh, Pennsylvania, USA, 2020.
- [63] Tianlin Xu, Li Kevin Wenliang, Michael Munn, and Beatrice Acciaio. Cot-gan: Generating sequential data via causal optimal transport. *Advances in Neural Information Processing Systems*, 33:8798–8809, 2020.
- [64] Jin Li, Benjamin J Cairns, Jingsong Li, and Tingting Zhu. Generating synthetic mixed-type longitudinal electronic health records for artificial intelligent applications. *NPJ Digital Medicine*, 6(1):98, 2023.
- [65] Abhyuday Desai, Cynthia Freeman, Zuhui Wang, and Ian Beaver. Timevae: A variational auto-encoder for multivariate time series generation. *ArXiv Preprint ArXiv*, 3:2111–08095, 2021.
- [66] Xiaomin Li, Anne Hee Hiong Ngu, and Vangelis Metsis. Tts-cgan: A transformer time-series conditional gan for biosignal data augmentation. *ArXiv Preprint ArXiv*, 1:2206–13676, 2022.
- [67] Paul Jeha, Michael Bohlke-Schneider, Pedro Mercado, Shubham Kapoor, Rajbir Singh Nirwan, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Psa-gan: Progressive self attention gans for synthetic time series. In *International Conference on Learning Representations*, Virtual, 2021.
- [68] Kevin Zhang, Kalyan Veeramachaneni, and Neha Patki. Sequential models in the synthetic data vault. 2022.
- [69] Synthetic Data Vault. SDV, 2023. Available at <https://docs.sdv.dev/sdv/>.

- [70] Jamal Fattah, Latifa Ezzine, Zineb Aman, Haj El Moussami, and Abdeslam Lachhab. Forecasting of demand using arima model. *International Journal of Engineering Business Management*, 10:1847979018808673, 2018.
- [71] Ralf C Staudemeyer and Eric Rothstein Morris. Understanding lstm—a tutorial into long short-term memory recurrent neural networks. *ArXiv Preprint ArXiv*, 1:1909–9586, 2019.
- [72] Pedro W Lamberti, Ana P Majtey, Antoni Borrás, Montserrat Casas, and Angel Plastino. Metric character of the quantum jensen-shannon divergence. *Physical Review A*, 77(5):052311, 2008.
- [73] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.
- [74] Tracy Jo Pasiëka, Robert F Woolson, and Charles Grose. Viral induced fusion and syncytium formation: measurement by the kolmogorov–smirnov statistical test. *Journal of Virological Vethods*, 111(2):157–161, 2003.
- [75] Toni Giorgino. Computing and visualizing dynamic time warping alignments in r: the dtw package. *Journal of Statistical Software*, 31:1–24, 2009.
- [76] Eamonn Keogh and Chotirat Ann Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7:358–386, 2005.
- [77] Bryan Lim and Stefan Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379(2194):1471–2962, 2021.
- [78] Boris N Oreshkin, Dmitri Carпов, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. *ArXiv Preprint ArXiv*, 4:1905–10437, 2019.
- [79] Cristian Challu, Kin G Olivares, Boris N Oreshkin, Federico Garza Ramirez, Max Mergenthaler Canseco, and Artur Dubrawski. Nhits: Neural hierarchical interpolation for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 6989–6997, 2023.
- [80] Huy Phan, Lars Hertel, Marco Maass, and Alfred Mertins. Robust audio event recognition with 1-max pooling convolutional neural networks. *ArXiv Preprint ArXiv*, 2:1604–6338, 2016.

- [81] Manu Joseph. Forecast Error Measures: Scaled, Relative, and other Errors, 2020. Available at <https://deep-and-shallow.com/2020/10/01/forecast-error-measures-scaled-relative-and-other-errors/>.
- [82] Davide Chicco, Matthijs J Warrens, and Giuseppe Jurman. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, 7:e623, 2021.
- [83] Chao Chen, Jamie Twycross, and Jonathan M Garibaldi. A new accuracy measure based on bounded relative error for time series forecasting. *Plos One*, 12(3):e0174202, 2017.
- [84] Óscar Giménez. ISO 15197:2015, la norma de la Comisión Europea que exige una mayor precisión a los sistemas de monitorización de glucosa en sangre, 2017. Available at <https://www.elfarmaceutico.es/uploads/s1/18/02/ef552-legislacion-iso.pdf>.
- [85] David C Klonoff, Courtney Lias, Robert Vigersky, William Clarke, Joan Lee Parkes, David B Sacks, M Sue Kirkman, Boris Kovatchev, and Error Grid Panel. The surveillance error grid. *Journal of Diabetes Science and Technology*, 8(4):658–672, 2014.
- [86] Stefan Pleus, Annette Baumstark, Nina Jendrike, Jochen Mende, Manuela Link, Eva Zschornack, Cornelia Haug, and Guido Freckmann. System accuracy evaluation of 18 ce-marked current-generation blood glucose monitoring systems based on en iso 15197:2015. *BMJ Open Diabetes Research and Care*, 8(1):e001067, 2020.
- [87] Numpy. Numpy, 2021. Available at: <https://numpy.org/doc/1.21/>.
- [88] Pandas. Pandas, 2023. Available at: <https://pandas.pydata.org/docs/whatsnew/v1.5.3.html>.
- [89] Matplotlib. Matplotlib, 2023. Available at: https://matplotlib.org/stable/users/prev_whats_new/github_stats_3.6.3.html.
- [90] Seaborn. Seaborn, 2021. Available at: <https://seaborn.pydata.org/archive/0.11/index.html>.
- [91] Gretel-synthetic. Gretel-synthetic, 2023. Available at: <https://pypi.org/project/gretel-synthetics/>.
- [92] Darts. Darts, 2023. Available at: <https://pypi.org/project/darts/>.

- [93] Scikit-learn. Scikit-learn, 2023. Available at: https://scikit-learn.org/stable/auto_examples/release_highlights/plot_release_highlights_1_2_0.html.